CrossMark

Expression of the histone chaperone SET/TAF-Iβ during the strobilation process of *Mesocestoides corti* (Platyhelminthes, Cestoda)

CAROLINE B. COSTA[1], KARINA M. MONTEIRO[1,2], ALINE TEICHMANN[1], EDILEUZA D. DA SILVA[1], KARINA R. LORENZATTO[1], MARTÍN CANCELA[1], JÉSSICA A. PAES[1], ANDRÉ DE N. D. BENITZ[1]†, ESTELA CASTILLO[3], ROGÉRIO MARGIS[4], ARNALDO ZAHA[1,2] *and* HENRIQUE B. FERREIRA[1,2]*

[1] *Laboratório de Genômica Estrutural e Funcional, Laboratório de Biologia Molecular de Cestódeos, and Programa de Pós-Graduação em Biologia Celular e Molecular (PPGBCM), Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil*
[2] *Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, UFRGS, Porto Alegre, RS, Brazil*
[3] *Sección Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay*
[4] *Laboratório de Genomas e Populações de Plantas, and PPGBCM, Centro de Biotecnologia, UFRGS, Porto Alegre, RS, Brazil*

(*Received 20 January 2015; revised 19 February 2015; accepted 28 February 2015; first published online 31 March 2015*)

SUMMARY

The histone chaperone SET/TAF-Iβ is implicated in processes of chromatin remodelling and gene expression regulation. It has been associated with the control of developmental processes, but little is known about its function in helminth parasites. In *Mesocestoides corti*, a partial cDNA sequence related to SET/TAF-Iβ was isolated in a screening for genes differentially expressed in larvae (tetrathyridia) and adult worms. Here, the full-length coding sequence of the *M. corti* SET/TAF-Iβ gene was analysed and the encoded protein (McSET/TAF) was compared with orthologous sequences, showing that McSET/TAF can be regarded as a SET/TAF-Iβ family member, with a typical nucleosome-assembly protein (NAP) domain and an acidic tail. The expression patterns of the McSET/TAF gene and protein were investigated during the strobilation process by RT-qPCR, using a set of five reference genes, and by immunoblot and immunofluorescence, using monospecific polyclonal antibodies. A gradual increase in McSET/TAF transcripts and McSET/TAF protein was observed upon development induction by trypsin, demonstrating McSET/TAF differential expression during strobilation. These results provided the first evidence for the involvement of a protein from the NAP family of epigenetic effectors in the regulation of cestode development.

Key words: SET/TAF-Iβ, *Mesocestoides corti*, cestode development, strobilation process.

INTRODUCTION

Mesocestoides corti is an endoparasitic platyhelminth used as experimental model to study the class Cestoda (Markoski *et al*. 2003), which also harbours helminths of medical and veterinary importance, such as those from the genera *Echinococcus* and *Taenia*. Despite of several studies in the field of cestode biology, little is known about the molecular mechanisms underpinning developmental processes of these organisms, such as the strobilation process. This process comprises cestode development from the larval stage to adult and includes proglottization (generation of serially repeated reproductive organs) and total or partial segmentation (Olson *et al*. 2001).

* Corresponding author: LGEF, Centro de Biotecnologia, UFRGS, Caixa Postal 15005, 91501-970, Porto Alegre, RS, Brazil. E-mail: henrique@cbiot.ufrgs.br
† Current address: Laboratório de Neurobiologia do Estresse, Departamento de Bioquímica, Instituto de Ciências Básicas da Saúde, UFRGS, Porto Alegre, RS, Brazil.

Several studies have been conducted to characterize genes and proteins involved in cestode development, but the current knowledge is still limited to allow the definition of developmental pathways involved in strobilation and other typical cestode developmental processes (Olson *et al*. 2012). However, difficulties to obtain parasite samples and safety issues regarding manipulation of pathogenic forms have imposed limitations for molecular studies. *Mesocestoides corti*, which is easily cultivated and is regarded as non-infective for humans, has allowed us to circumvent some of these experimental limitations, especially those concerning the availability of biological material and safety (Markoski *et al*. 2003). Moreover, the *in vitro M. corti* culture system allows monitoring of the whole strobilation process under controlled experimental conditions (Markoski *et al*. 2003, 2006).

In a previous study conducted by our group, subtracted cDNA libraries enriched in sequences differentially expressed in *M. corti* larvae (tetrathyridia) or

Parasitology (2015), **142**, 1171–1182. © Cambridge University Press 2015
doi:10.1017/S003118201500030X

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

strobilated adult worms were used to provide the first report on differential gene expression between the larval and adult stages of this organism (Bizarro *et al*. 2005). We also performed a proteomic analysis of strobilating tetrathyridia in order to identify proteins differentially represented in early strobilation events (Laschuk *et al*. 2011). These studies revealed several genes and proteins with apparent differential expression between tetrathyridia and adult worms or upon strobilation induction, including a putative *M. corti* orthologue of SET/TAF-Iβ proteins (McSET/TAF).

SET/TAF-Iβ proteins form a subfamily of the nucleosome-assembly protein (NAP) family. They consist of an N-terminal NAP domain, which is conserved among NAP-family members, and a C-terminal acidic tail, implicated in histone binding and chromatin modification (Akey and Luger, 2003). SET/TAF-Iβ orthologues of different species have been described as inhibitors of PP2A protein (Li *et al*. 1996) and as part of inhibitor of acetyltransferases (INHAT) complexes (Kim *et al*. 2012). As chromatin regulators, they have been implicated in several important cellular processes, including DNA replication (Nagata *et al*. 1995), transcriptional co-activation (Kato *et al*. 2007), cell differentiation (Kim *et al*. 2010) and apoptosis induction (Fan *et al*. 2003).

In this work, the full length McSET/TAF encoding gene (*McSET/TAF*), retrieved from the *M. corti* unannotated genome draft, was analysed regarding its exon–intron structure. The McSET/TAF deduced amino acid sequence was also analysed and the protein was established as a true member of the SET/TAF-Iβ subfamily of NAP proteins. We also investigated the expression of *McSET/TAF* transcripts and of the McSET/TAF protein in tetrathyridia and in three time points after strobilation induction, demonstrating the progressive increase of McSET/TAF expression during strobilar development and providing the first evidence for the involvement of a protein from the SET/TAF-Iβ family in the regulation of cestode development.

MATERIALS AND METHODS

Parasite material

Mesocestoides corti tetrathyridia were maintained by intraperitoneal infection of 3-month-old female Balb/c mice. Parasites collected from mice were used for *in vitro* cultures in modified RPMI medium (Gibco), as previously described (Markoski *et al*. 2003). Briefly, tetrathyridia were collected from euthanized mice by peritoneal aspiration and washed three times in the culture medium prior to culture. Approximately 5000 tetrathyridia were collected from each infected mouse and cultures were carried out into six-well plates, with about 50 tetrathyridia in 3 mL of medium per well, at 37 °C and 5% CO_2 for 2 days prior to strobilation induction. All experimental procedures for *in vivo* maintenance of *M. corti* tetrathyridia in mice hosts were approved by the Ethical Committee of the Universidade Federal do Rio Grande do Sul (UFRGS) (Project no. 21625).

Strobilar induction

Tetrathyridia strobilation was performed as described (Markoski *et al*. 2003). Parasites were pre-incubated in modified RPMI media supplemented with 20% fetal bovine serum (FBS, Cultilab, Brazil) containing 0·662% (w/v) trypsin (Sigma), equivalent to 10^5 Na-benzoyl-l-arginine ethyl ester (BAEE) units ml^{-1}, at 39°C for 24 h. After 24 h induction, the parasites were incubated with 3 ml per well of the fresh modified RPMI media supplemented with 20% fetal bovine serum (FBS, Cultilab, Brazil) and maintained at 39°C under 5% CO_2 for at least 7 days. Fully strobilated parasites were selected in stereo microscope. All cultures were performed in quadruplicates and strobilation efficiency upon induction was 70–80%. In control cultures (not induced by trypsin treatment) strobilation rate was under 60%.

Four developmental stages of *M. corti* were selected during strobilar induction for preparing RNA and protein extracts: *bona fide* tetrathyridia (TT), tetrathyridia 24 h post strobilation induction (24 h-PI), strobilating worms 72 h post-induction (72 h-PI) and fully strobilated worms (SW).

Protein extracts

All samples (TT, 24 h-PI, 72 h-PI and SW) were washed ten times with 40 mM Tris–HCl buffer (pH 7·2) and homogenized using a glass tissue grinder in an ice bath. The parasite lysate was clarified by centrifugation at 4 °C (15 000 ***g*** for 30 min). The protein content of the soluble supernatant was estimated using a Qubit™ quantitation fluorometer (Invitrogen) and qualitatively assessed by SDS–PAGE. Aliquots were stored at −20 °C until use.

McSET/TAF *gene prediction, amino acid sequence analysis and phylogeny*

A previously isolated partial *McSET/TAF* cDNA (Bizarro *et al*. 2005) was used as a query against the *M. corti* genomic contigs deposited at the Sanger Institute (ftp://ftp.sanger.ac.uk/pub/project/pathogens/HGI/) using the BLAST tool (http://www.ncbi.nlm.nih.gov/BLAST/). The *McSET/TAF* gene structure was predicted using the GeneMark (Besemer and Borodovsky, 2005), FGENESH (Salamov and Solovyev, 2000) and GENESCAN algorithms (Burge and Karlin,

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

1998). The amino acid sequence deduced from the full-length coding region of the *McSET/TAF* gene (corresponding to the McSET/TAF protein) was analysed using Pfam Ver. 27·0 (http://pfam.xfam.org/), and ProtParam tool (www.expasy.org) in order to determine protein domains and molecular masses/isoelectric point, respectively. Phylogeny analysis was performed using SET/TAF-Iβ orthologous proteins recovered from the GeneDB (http://www.genedb.org) (Logan-Klumpler *et al*. 2012) and NCBI (http://www.ncbi.nlm.nih.gov/) databases using the BLAST tool. Protein sequences were aligned using ClustalW2 and visualized using GeneDoc ver. 2·6·002. Phylogenetic trees were constructed by the Neighbour-Joining method (Saitou and Nei, 1987) and all evolutionary analyses were conducted using the MEGA 5·2 tool (Tamura *et al*. 2011). The bootstrap consensus tree inferred from 10 000 replicates was taken to represent the evolutionary history of the analysed taxa. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved the full-length McSET/TAF deduced amino acid sequence and orthologous proteins from *Danio rerio* (NCBI no. NP_958876·1), *Drosophila melanogaster* (NCBI no. NT_033777·2), *Echinococcus granulosus* (GeneDB no. EgrG_000465500·1), *Echinococcus multilocularis* (GeneDB no. EmuJ_000465500·1), *Homo sapiens* (NCBI no. NP_003002·2), *Mus musculus* (NCBI no. NP_001191804·1), *Schistosoma japonicum* (GeneDB no. Sjp_0002330·1), *Schistosoma mansoni* (GeneDB no. Smp_155060·2) and *Taenia solium* (GeneDB no. TsM_000881900·1). All positions containing gaps and missing data were eliminated. There were a total of 232 positions in the final dataset.

RT-qPCR

Total RNA was extracted from *M. corti* TT, 24 h-PI, 72 h-PI and SW samples (with ~150 parasites each) using TRIzol® Reagent (Invitrogen) following the manufacturer's instructions. Isolated RNA was quantified by Qubit® 2·0 Fluorometer (Qubit® RNA Assay Kit), treated with RNase-free DNaseI (Fermentas) and used for reverse transcription using M-MuLV Reverse Transcriptase (Fermentas) with Oligo(dT)$_{18}$ Primer (Fermentas). RT-qPCRs were performed in a 7500 Real-Time PCR System (Applied Biosystems) in a 96-well microtiter plate format. Each reaction was prepared with 10 μL of cDNA (diluted 1:100), SYBR Green (1X) (Invitrogen), 5 pmol of gene-specific forward and reverse primers, 5 mM dNTPs, 0·24 U Platinum Taq DNA Polymerase (Invitrogen) and RNase/DNase free water for a final volume of 20 μL. The applied cycling conditions were as follows: 5 min DNA polymerase activation at 95 °C, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing at 60 °C for 10 s, and elongation at 72 °C for 15 s. Negative control reactions (without RT enzyme) were carried out to exclude the possibility of amplification due to the presence of contaminating DNA. All samples and corresponding negative controls were amplified in triplicates.

Based on previous studies (Bizarro *et al*. 2005; Koziol *et al*. 2011), the *M. corti* orthologues of the receptor-mediated endocytosis-8 (ARME-8), chorea-acanthocytosis (*CHOREIN*), receptor-mediated endocytosis-8 (*CRME-8*), lipopolysaccharide (LPS)-responsive and beige-like anchor (*LRBA*), eyelid/OSA brahma complex gene (*OSA*), poly(A)-binding protein (*PAPB*), suvar3-complex (*SBF1*), programmed cell death 4 (*PDCD4*), splicing factor, arginine/serine-rich 6 (*SR6*) and tropomyosin (*TROPO*) genes were selected as potential references genes for *M. corti* RT-qPCR data normalization. The primers used for the amplification of sequences from *McSET/TAF* and selected reference genes are shown in the Appendix (Table A1). The geNorm software (Vandesompele *et al*. 2002) was used for determination of reference genes and normalization of real-time PCR expression data, based on the expression levels of target genes in the TT stage. The calculations of the relative expression of each gene were made by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Statistical analyses were performed using the Duncan's test at 5% significance, using SPSS 20 software (IBM Corp).

McSET/TAF partial cDNA cloning

Total RNA from *M. corti* TT stage was extracted and used for cDNA synthesis as described above. The cDNA synthesis was followed by PCR amplification with High Fidelity DNA Polymerase (Fermentas) using gene-specific primers for *McSET/TAF* (5′-ATCTTGGAGAGCGTGGACT-3 and 5′-GCCAGACGTAATATCTTCG-3′). The primers were designed according to the previously published partial *McSET/TAF* cDNA sequence (Bizarro *et al*. 2005), which codes for 93 amino acids near the N-terminal end of McSET/TAF (corresponding to amino acids 40–132 of the protein) (GenBank Acc: KJ472143). The amplified coding sequence (*McSET/TAF*$_{42–132}$), with 273 pb, codes for McSET/TAF amino acids 42–132 (McSET/TAF$_{42–132}$). In the first PCR amplification, 24 nt recombination tags FrecI (5′-TATTTTCAGGGAGAATTCCCGGGT-3′) and RrecI (5′-GCGAGGCAGATCGTCAGTCAGTCA-3′), matching the cloning vector, were added to the ends of the 5′ gene-specific primers. Secondary PCR amplification was performed using the primary reaction as template and the primers FrecII

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

(5′-TGGTTCCGCGTGGATCTGAAAACCTG TATTTTCAGGGAGAATTCCCGGGT-3′) and RrecII (5′-GGTTTTCACCGTCATCACCGAAA CGCGCGAGGCAGATCGTCAGTCAGTCA-3′). The two rounds of PCR amplification resulted in a *McSET/TAF*$_{42–132}$ amplicon tagged with 50 bp matching pGEX-TEV in its 5′ and 3′ ends.

The final amplification product was then cloned into the pGEX-TEV plasmid, a version of pGEX-4-T1 (GE Healthcare) modified in our laboratory to include the sequence coding for a cleavage site for tobacco etch virus (TEV) protease into the vector polycloning site (data not show). Cloning was performed by *in vivo* homologous recombination (Parrish *et al*. 2004), with minor modifications. In frame cloning of the McSET/TAF$_{42–132}$ coding sequence with the vector encoded glutathione S-transferase (GST) was confirmed by sequencing using the Dyenamic ET Dye Terminator Cycle Sequencing kit (GE Healthcare) in a MEGABACE 1000 Sequencing System.

Expression and purification of recombinant McSET/TAF$_{42–132}$

Recombinant McSET/TAF$_{42–132}$ expression in *Escherichia coli* and purification were performed essentially as described by Moitinho-Silva *et al*. (2012), with some modifications. Briefly, recombinant plasmids carrying the coding sequence for the McSET/TAF$_{42–132}$ were transformed into BL21-CodonPlus-RIL (Stratagene) and cultures were induced with 0·1 mM IPTG for 3 h at 37 °C. After induction, cells were harvested and lysed, and 0·5% sarkosyl was used for protein solubilization. The GST-tagged recombinant protein was purified by affinity chromatography on Glutathione-Sepharose 4B (GE Healthcare) and cleaved with TEV protease for 16 h at 34 °C to be released from the GST moiety. Eluted protein concentration was measured using a Qubit™ quantitation fluorometer and Quant-iT™ reagents (Invitrogen) and purity was assessed by SDS–PAGE.

Antiserum production and antibody purification

For specific polyclonal antiserum production, a rabbit was immunized by subcutaneous injection using 150 μg of recombinant protein in complete Freund's adjuvant (Sigma). Immunization was followed by three boosters of 150 μg of protein in incomplete Freund's adjuvant (Sigma) every two weeks. Blood sample was collected one week after the last immunization and the serum was separated by centrifugation. IgG antibody purification from non-immune control serum and from polyclonal serum from the immunized animal was carried out in HiTrap Protein G HP columns (GE Healthcare), according to the manufacturer's protocol. All experimental procedures for polyclonal antiserum production in rabbits were approved by the Ethical Committee of the UFRGS (Project no. 21625).

Immunoblots

Twenty micrograms of protein extracts from *M. corti* developmental stages TT, 24 h-PI, 72 h-PI and SW were resolved by 12% SDS–PAGE and transferred to polyvinylidene fluoride membranes (PVDF) (Hybond™-ECL™, GE Healthcare) as described by Monteiro *et al*. (2010). Rabbit polyclonal antiserum anti-McSET/TAF$_{42–132}$ (diluted 1:20 000,v/v) was used as primary antibody, and horseradish peroxidase (HRP)-labelled anti-rabbit IgG (ECL™, GE Healthcare) was used as secondary antibody (1:9000 v/v dilution). Antigen–antibody complexes were detected using ECL detection reagent (GE Healthcare) and imaged using a VersaDoc imaging system (Bio-Rad). All experiments were performed in duplicate, with protein samples corresponding to biological replicates (from independent parasite cultures). The ImageJ software (NIH, Maryland, USA) was used for expression relative quantifications.

Immunohistochemistry

Samples of each *M. corti* developmental stage (TT, 24 h-PI, 72 h-PI and SW) were fixed, dehydrated and embedded in paraffin as described by Koziol *et al*. (2010). Sections 8 μm thick were mounted, rehydrated and blocked with 1% bovine serum albumin, 0·05% Tween in PBS for 1 h at 37 °C. Processed sections were then incubated in a humid chamber for 1 h at 37 °C with purified anti-McSET/TAF$_{42–132}$ antibodies, diluted 1:200 (v/v) in blocking solution, followed by six washes for 20 min each with PBS. Sections were then incubated with goat anti-rabbit IgG conjugated with Alexa Fluor 488 (Invitrogen), diluted 1:500 (v/v) in blocking solution, in a humid chamber for 1 h at 37 °C. After three additional washes in PBS, sections were incubated with 50 μM DAPI for 20 min at 37 °C, mounted with Fluoromount (Sigma) and observed under a confocal microscope (Olympus FluoView™ 1000), available at the Centro de Microscopia Eletrônica of the UFRGS (CME-UFRGS). Images were digitally captured and processed using the software Olympus Fluoview-F1000.

RESULTS

McSET/TAF *gene prediction, amino acid sequence analysis and phylogeny*

In order to determine the full-length coding sequence of the *M. corti* SET/TAF-Iβ orthologous gene and protein (*McSET/TAF* and McSET/

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

TAF, respectively), we performed a blastN analysis using the previously available *McSET/TAF* partial cDNA sequence (Bizarro *et al.* 2005) as query against the *M. corti* draft genome sequence available at Sanger Institute (ftp://ftp.sanger.ac.uk/pub/project/pathogens/HGI/). A single contig (MCOS.contig.01068·30933 contig), containing only one identifiable gene with homology to the query sequence (e-value of 5·2e–57), was retrieved, indicating that McSET/TAF is encoded by a single copy gene in the *M. corti* genome. No evidence of paralogue sequences were found in the performed searches.

The MCOS.contig.01068·30933 contig sequence was analysed with three gene prediction algorithms, and the predicted exon–intron structures for the identified *McSET/TAF* gene are shown in the Appendix (Table A2). The longest predicted coding DNA sequence (that predicted by the GeneMark algorithm), which also resulted in the highest similarity with flatworm orthologues, was used for subsequent *in silico* analyses. Based on that, the *McSET/TAF* gene is composed of four exons and three introns, with a coding region of 792 bp encoding a 263 aa-long protein (Fig. 1A). The analysis of the deduced amino acid sequence revealed the presence of both a NAP domain and a C-terminal acidic tail (Fig. 1B), typical of NAPs of the SET/TAF-I*β* subfamily.

Multiple sequence alignments of McSET/TAF and its orthologues from a wide range of eukaryotic species, including other helminths and model organisms, are shown in the Appendix (Fig. A1). Regarding other Platyhelminthes, the *M. corti* McSET/TAF amino acid sequence displayed 68% identity and 76% similarity with the orthologous sequences from the cestodes *E. granulosus* and *E. multilocularis*, 71% identity and 79% similarity with *T. solium* sequence, 47% identity and 63% similarity with the trematode *S. japonicum* and 48% identity and 63% similarity with *S. mansoni*. Comparisons with *D. melanogaster* sequence showed identity/similarity levels of 38%/56%. With mammalian sequences, levels of identity/similarity were 40/57% and 39/57% for mouse and human orthologues, respectively. With *D. rerio*, the levels of identity/similarity were 41/57%.

The McSET/TAF deduced amino acid sequence and the previously aligned orthologous sequences were also used to infer a phylogenetic tree (Fig. 1C). In this tree, McSET/TAF appears in a 'more primitive' branch, along with orthologues from other helminths, while orthologues from *D. melanogaster* and vertebrates formed another major branch. Taken together, our phylogenetic analysis, the overall identities and similarities with other eukaryotic SET/TAF-I*β* proteins, and the identification of typical NAP domain and acidic tail clearly point out that McSET/TAF can be considered a true SET/TAF-I*β* orthologue.

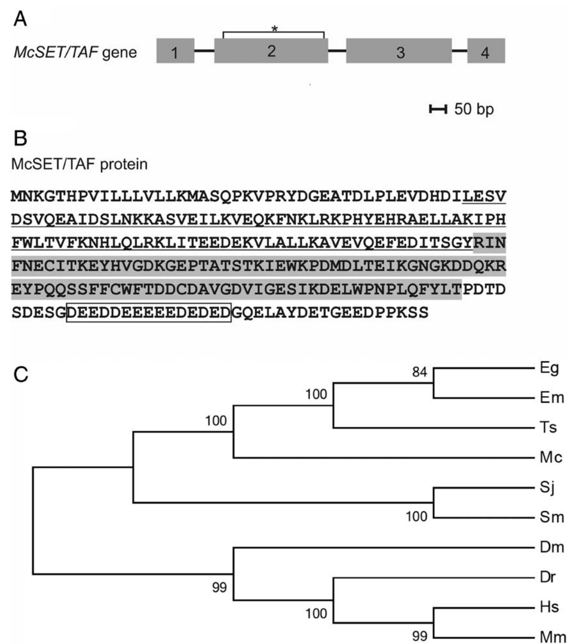

Fig. 1. Structure of *McSET/TAF* gene, deduced amino acid sequence of the encoded protein and phylogenetic relationships. (A) Exon–intron structure as predicted *in silico*. Numbered grey bars represent exons, lines represent introns, and the cloned region, coding for McSET/TAF amino acids 42–132 ($McSET/TAF_{42–132}$), is indicated (*); (B) McSET/TAF deduced amino acid sequence and predicted domains, which grey-shadowed amino acids belong to the NAP domain and the C-terminal acidic tail is boxed. The cloned and expressed segment of the protein is underlined; (C) Phylogenetic analysis of McSET/TAF inferred using the Neighbour-Joining method. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The percentages of replicate trees in which the associated proteins clustered together in the bootstrap test (10 000 replicates) are shown next to the branches. The phylogenetic tree was generated with the full-length McSET/TAF deduced amino acid sequence and orthologous proteins from *D. rerio* (Dr), *D. melanogaster* (Dm), *E. granulosus* (Eg), *E. multilocularis* (Em), *H. sapiens* (Hs), *M. musculus* (Mm), *S. japonicum* (Sj), *S. mansoni* (Sm) and *T. solium* (Ts).

RT-qPCR standardization and assessment of McSET/TAF *mRNA levels during strobilation*

For RT-qPCR standardization, 10 *M. corti* genes (listed in Table A1, Appendix) were evaluated as potential normalizers for *M. corti* transcriptional data. Based on the expression stability (M) values obtained in TT, 24 h-PI, 72 h-PI and SW samples (Fig. A2, Appendix), five genes were selected as the more stable ones (with the lowest M values), namely *PDCD4* ($M = 0·5$), *PAPB* ($M = 0·53$), *SR6* ($M = 0·57$), *CHOREIN* ($M = 0·72$) and *TROPO* ($M = 0·8$). Using the established normalizing genes

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

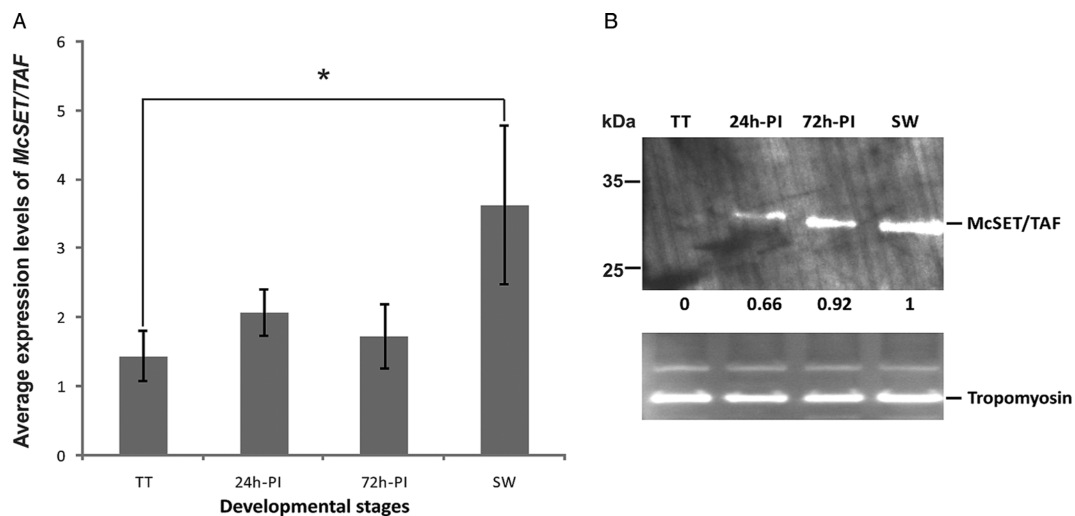

Fig. 2. *McSET/TAF* gene and protein expression levels in different *M. corti* developmental stages. (A) The levels of *McSET/TAF* gene transcripts were normalized to the expression level of *PDCD4, PAPB, SR6, CHOREIN* and *TROPO* gene transcripts. Normalized data were submitted to analysis of variance and differences with >5% significance by Duncan test are indicated (*). The bars indicate standard deviation between replicates; (B) McSET/TAF protein levels assessed by immunoblot using anti-McSET/TAF purified IgG. All lanes were loaded with 20 µg of each protein extract and relative McSET/TAF expression levels were normalized against the *M. corti* tropomyosin band for each sample. In the comparisons between McSET/TAF bands from different samples, the SW band was used as reference, and the relative expression values for all samples are indicated. Expression data were normalized and compared using the IMAGE J software.

and standardized RT-qPCR conditions, we then assessed the *McSET/TAF* mRNA levels in TT, 24 h-PI, 72 h-PI and SW. The RT-qPCR results, summarized in Fig. 2A, revealed that the *McSET/TAF* mRNA was expressed in all tested parasite stages, with significantly different levels only between TT and SW, and higher expression level in the SW stage.

Temporal and spatial expression pattern of McSET/TAF during M. corti *development*

The McSET/TAF protein expression in TT, 24 h-PI, 72 h-PI and SW stages was assessed both by immunoblot and immunofluorescence using polyclonal antibodies generated in rabbit against the recombinant polypeptide $McSET/TAF_{42-132}$. In immunoblot experiments (Fig. 2B), the anti-$McSET/TAF_{42-132}$ antibodies detected a band corresponding to a ~30 kDa protein, in accordance with the expected molecular mass of McSET/TAF (see Appendix, Table A2). A faint McSET/TAF band was detected in the TT protein extract, and, in the other analysed developmental stages, the McSET/TAF band showed a gradual increase in its intensity, from 24 h-PI to SW. Apparent small variations in McSET/TAF size between the 24 h-PI, 72 h-PI, and SW extracts could be noticed.

By immunofluorescence (Fig. 3), McSET/TAF expression showed a pattern essentially in line with that observed in the immunoblot experiment described above. McSET/TAF expression was also very weak in the TT stage, and, upon strobilation induction, it increased. Actual differences in McSET/TAF expression between 24 h-PI, 72 h-PI and SW sections are not easy to establish, due to differences in cell densities and distribution in the different developmental stages. The protein showed a uniform pattern of distribution in inner tissues for all samples, as expected for typical SET/TAF-Iβ proteins. In 24 h-PI e 72 h-PI sections, there was a strong staining in the sub-tegumental region, corresponding to sub-tegumental muscle cells or tegumental cells (also known as perinuclear cell bodies of the tegumental syncytium).

DISCUSSION

The characterization of differentially expressed proteins during development of *M. corti* is important for a better understanding of basic biological aspects of the parasite, such as the molecular mechanisms involved during its development process. To date, only a few cestode genes involved in processes of cell proliferation and differentiation, such as those encoding the protein-kinases Ras, Raf and MPK1 in *E. multilocularis* (Spiliotis *et al*. 2005, 2006) and the transcription factor Mvlim in *M. corti* (Lalanne *et al*. 2004). The potential involvement of these genes in developmental processes has been suggested, but further studies are needed to demonstrate it. Recently, the genome sequences of some cestode species were published (Tsai *et al*. 2013; Zheng *et al*. 2013), which will facilitate the

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

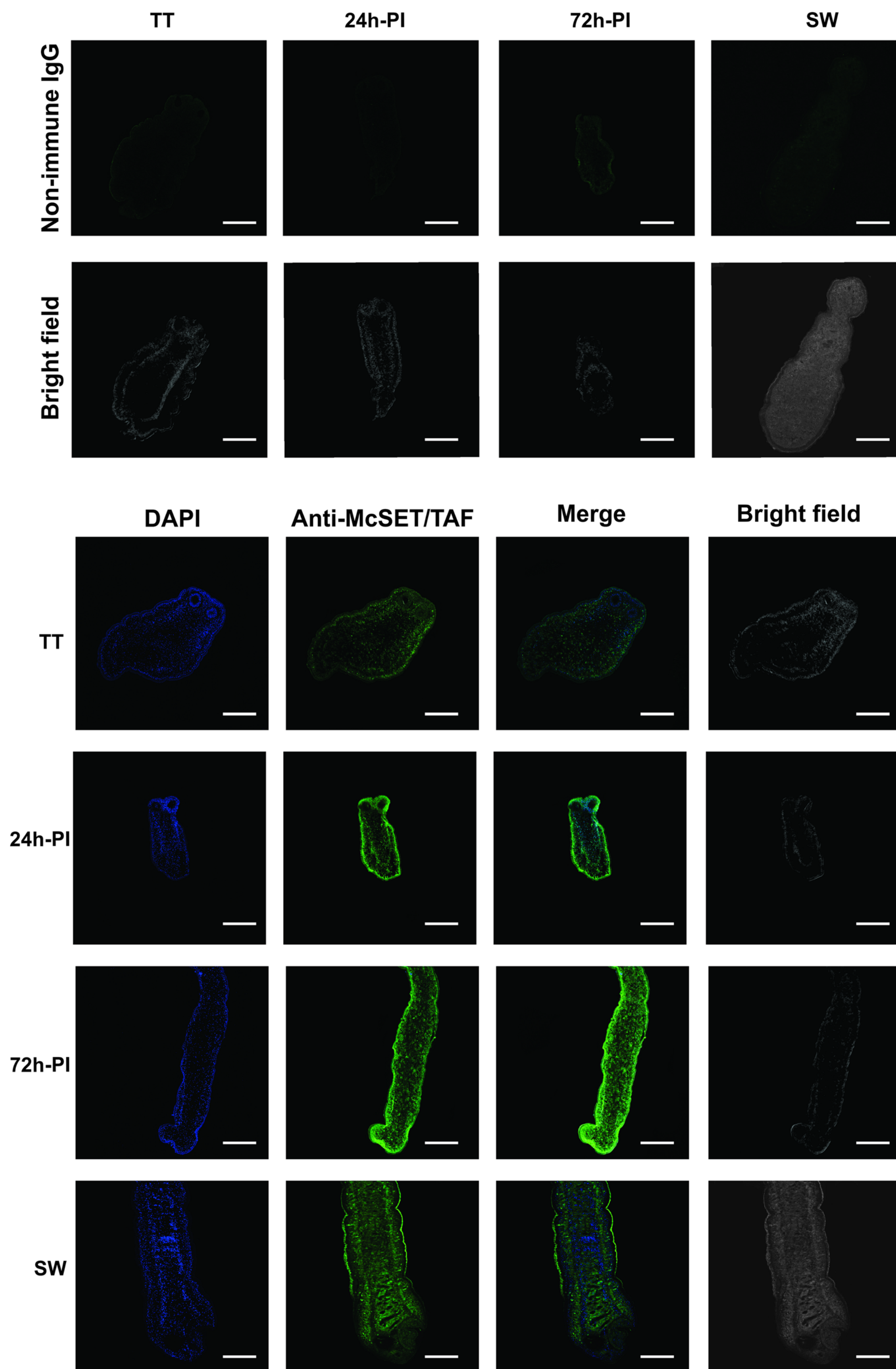

Fig. 3. Immunolocalization of McSET/TAF in different *M. corti* developmental stages. Parasite sections (8 μm) were incubated with purified non-immune IgG (upper panel) and purified anti-McSET/TAF$_{42-132}$ IgG. Recognition of immune complexes was achieved using Alexa 488-conjugated secondary antibodies. DAPI nuclei staining; antibody staining; merged DAPI and antibody images; and bright field images are shown, from left to right, in the first, second, third and fourth columns, respectively. Scale bars: TT images, 200 μm; 24 h-PI and 72 h-PI, and SW images 50 μm.

prospection of developmental genes and proteins for this taxon of parasites.

The *M. corti* SET/TAF-Iβ related sequence studied here was originally isolated as a partial cDNA enriched in SW libraries (Bizarro *et al.* 2005). Here, we were able to characterize the whole gene sequence *in silico* and establish the identity and expression pattern of the encoded protein in undifferentiated larvae and in three stages of the strobilation process. Based on a phylogenetic tree,

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

the McSET/TAF protein appears in a 'more primitive' branch, along with orthologues from other helminths, while orthologues from *D. melanogaster* and vertebrates formed another major branch. This phylogenetic tree, along with the overall identities and similarities of McSET/TAF with other eukaryotic SET/TAF-Iβ proteins and the identification of typical NAP domain and acidic tail clearly point out that McSET/TAF is a member of the SET/TAF-Iβ subfamily of NAP proteins.

McSET/TAF gene expression revealed that it is transcribed in all four tested parasite stages, which represent the undifferentiated larval stage (TT) and the early (24 h-PI), intermediate (72 h-PI) and end (SW) stages of strobilation. A significant higher level of expression in SW in comparison with the TT stage was demonstrated. This corroborated the preliminary transcriptomic data reported by Bizarro *et al.* (2005) for *M. corti*, which suggested that the *McSET/TAF* gene was differentially transcribed between TT and SW stages, with higher expression in the SW stage. Although not significantly different, the *McSET/TAF* expression was also consistently higher from 24 h-PI to 72 h-PI to SW samples in all experiments performed with independent cultures, indicating a progressive increase in the transcription levels of this gene from TT to SW. During this developmental transition, the variation in McSET/TAF protein size observed in the immunoblot experiments raises the possibility of differential regulation at a post-translational level. This could be due to post-translational modifications, such as phosphorylation, known to occur in SET/TAF-Iβ proteins (Irie *et al.* 2012). Other alternatives to explain possible McSET/TAF variants would be the use of alternative translational initiation codons or alternative splicing. SET/TAF-Iβ variants generated by alternative splicing have been described in humans (Nagata *et al.* 1995), but such post-transcriptional events remain to be investigated in *M. corti*.

The expression levels of the *McSET/TAF* gene and the McSET/TAF protein are very low in TT, where lower levels of overall expression are expected, since the parasite is in an undifferentiated stage, in which only vegetative growth and asexual reproduction take place (Markoski *et al.* 2003). In 24 h-PI, 72 h-PI and SW stages, there was an increase in the levels of McSET/TAF, suggesting a greater involvement of this protein in both initial and later strobilar development. Therefore, McSET/TAF expression level increase in parallel with the activation and functioning of processes of growth, differentiation and cell death during strobilation indicates its possible involvement with the cellular machinery that regulates them. Besides, the expression pattern found for SET/TAF-Iβ proteins in different eukaryotic organisms, such as *Saccharomyces cerevisiae, D. melanogaster* and mouse, is ubiquitous, with the protein being found in both the nucleus and cytoplasm of the cells (Nagata *et al.* 1998). This is in line with the ubiquitous and uniform pattern found for McSET/TAF in our immunofluorescence experiments.

SET/TAF-Iβ proteins are known to act both in the nucleus and in the cytoplasm regulating several cellular processes, from DNA replication and gene expression to cell differentiation and apoptosis (Nagata *et al.* 1995; Fan *et al.* 2003; Kim *et al.* 2010). It is then interesting to speculate the possible functional involvement of McSET/TAF in these and other cellular processes. As a putative histone chaperone of the SET/TAF-Iβ subfamily of NAPs, as inferred from the performed alignment and phylogenetic analyses, McSET/TAF may act as a transcriptional co-activator (Kato *et al.* 2007) and, as such, regulate developmental genes important for *M. corti* strobilation. Furthermore, in immunofluorescence assays, SW sections showed an apparent staining in the tegument surface, which could indicate an ectopic localization of McSET/TAF at this stage, whose tegument can present differences in structure and/or composition (Markoski *et al.* 2003). The presence of typical intracellular proteins in the tegument or among excretion/secretion products of cestodes has been described and associated with potential moonlighting functions (Monteiro *et al.* 2010; Lorenzatto *et al.* 2012).

Concluding remarks

The *McSET/TAF* gene codes for a SET/TAF-Iβ protein of the NAP family. Based on a set of five *M. corti* reference genes standardized as normalizers, it was shown that *McSET/TAF* transcription level is higher in SW than in TT. It was also shown that the encoded McSET/TAF protein progressively increases during *M. corti* strobilation, providing the first evidence for the involvement of a protein from the SET/TAF-Iβ family of epigenetic effectors in the regulation of cestode development. McSET/TAF can now be further studied in order to characterize its activities, functions and/or interactions relevant for cestode developmental pathways.

ACKNOWLEDGEMENTS

The authors thank the *Centro de Microscopia Eletrônica* (UFRGS), for technical support with the confocal microscopy.

FINANCIAL SUPPORT

This work was supported by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) (AUX-PE-PARASITOLOGIA 1278/2011), in Brazil. C.B.C was a recipient of a CAPES M.Sc. fellowship.

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

K.M.M. was a recipient of a CAPES postdoctoral fellowship. E.D.S. and J.A.P. were recipients of *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) M.Sc. fellowships. M.C. is a recipient of a CAPES postdoctoral fellowship. A.T. was a recipient of a CAPES PhD fellowship. K.R.L. was a recipient of a CNPq PhD fellowship.

REFERENCES

Akey, C. W. and Luger, K. (2003). Histone chaperones and nucleosome assembly. *Current Opinion in Structural Biology* **13**, 6–14.

Besemer, J. and Borodovsky, M. (2005). GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* **33**, W451–W454.

Bizarro, C. V., Bengtson, M. H., Ricachenevsky, F. K., Zaha, A., Sogayar, M. C. and Ferreira, H. B. (2005). Differentially expressed sequences from a cestode parasite reveals conserved developmental genes in platyhelminthes. *Molecular and Biochemical Parasitology* **144**, 114–118.

Burge, C. B. and Karlin, S. (1998). Finding the genes in genomic DNA. *Current Opinion in Structural Biology* **8**, 346–354.

Fan, Z., Beresford, P. J., Oh, D. Y., Zhang, D. and Lieberman, J. (2003). Tumor suppressor NM23-H1 is a granzyme A-activated DNase during CTL-mediated apoptosis, and the nucleosome assembly protein SET is its inhibitor. *Cell* **112**, 659–672.

Irie, A., Harada, K., Araki, N. and Nishimura, Y. (2012). Phosphorylation of SET protein at Ser171 by protein kinase D2 diminishes its inhibitory effect on protein phosphatase 2A. *PLoS ONE* **7**, e51242.

Kato, K., Miyaji-Yamaguchi, M., Okuwaki, M. and Nagata, K. (2007). Histone acetylation-independent transcription stimulation by a histone chaperone. *Nucleic Acids Research* **35**, 705–715.

Kim, D. W., Kim, K. B., Kim, J. Y., Lee, K. S. and Seo, S. B. (2010). Negative regulation of neuronal cell differentiation by INHAT subunit SET/TAF-Iβ. *Biochemical and Biophysical Research Communication* **400**, 419–425.

Kim, J. Y., Kim, K. B., Son, H. J., Chae, Y. C., Oh, S. T., Kim, D. W., Pak, J. H. and Seo, S. B. (2012). H3K27 methylation and H3S28 phosphorylation-dependent transcriptional regulation by INHAT subunit SET/TAF-Iβ. *FEBS Letters* **586**, 3159–3165.

Koziol, U., Domínguez, M. F., Marín, M., Kun, A. and Castillo, E. (2010). Stem cell proliferation during *in vitro* development of the model cestode *Mesocestoides corti* from larva to adult worm. *Frontier in Zoology* **7**, 22.

Koziol, U., Costábile, A., Domínguez, M. F., Iriarte, A., Alvite, G., Kun, A. and Castillo, E. (2011). Developmental expression of high molecular weight tropomyosin isoforms in *Mesocestoides corti*. *Molecular and Biochemical Parasitology* **175**, 181–91.

Lalanne, A. I., Britos, L., Ehrlich, R. and Castillo, E. (2004). *Mesocestoides corti*: a LIM-homeobox gene upregulated during strobilar development. *Experimental Parasitology* **108**, 169–175.

Laschuk, A., Monteiro, K. M., Vidal, N. M., Pinto, P. M., Duran, R., Cerveñanski, C., Zaha, A. and Ferreira, H. B. (2011). Proteomic survey of the cestode *Mesocestoides corti* during the first 24 hours of strobilar development. *Parasitology Research* **108**, 645–656.

Li, M., Makkinje, A. and Damuni, Z. (1996). The myeloid leukemia-associated protein SET is a potent inhibitor of protein phosphatase 2A. *Journal of Biological Chemistry* **271**, 11059–11062.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408.

Logan-Klumpler, F. J., De Silva, N., Boehme, U., Rogers, M. B., Velarde, G., McQuillan, J. A., Carver, T., Aslett, M., Olsen, C., Subramanian, S., Phan, I., Farris, C., Mitra, S., Ramasamy, G., Wang, H., Tivey, A., Jackson, A., Houston, R., Parkhill, J., Holden, M., Harb, O. S., Brunk, B. P., Myler, P. J., Roos, D., Carrington, M., Smith, D. F., Hertz-Fowler, C. and Berriman, M. (2012). GeneDB – an annotation database for pathogens. *Nucleic Acids Research* **40**, D98–108.

Lorenzatto, K. R., Monteiro, K. M., Paredes, R., Paludo, G. P., da Fonsêca, M. M., Galanti, N., Zaha, A. and Ferreira, H. B. (2012). Fructose-bisphosphate aldolase and enolase from *Echinococcus granulosus*: genes, expression patterns and protein interactions of two potential moonlighting proteins. *Gene* **506**, 76–84.

Markoski, M. M., Bizarro, C. V., Farias, S., Espinoza, I., Galanti, N., Zaha, A. and Ferreira, H. B. (2003). *In vitro* segmentation induction of *Mesocestoides corti* (Cestoda) tetrathyridia. *Journal of Parasitology* **89**, 27–34.

Markoski, M. M., Trindade, E. S., Cabrera, G., Laschuk, A., Galanti, N., Zaha, A., Nader, H. B. and Ferreira, H. B. (2006). Praziquantel and albendazole damaging action on *in vitro* developing *Mesocestoides corti* (Platyhelminthes: Cestoda). *Parasitology International* **55**, 51–61.

Moitinho-Silva, L., Heineck, B. L., Reolon, L. A., Paes, J. A., Klein, C. S., Rebelatto, R., Schrank, I. S., Zaha, A. and Ferreira, H. B. (2012). *Mycoplasma hyopneumoniae* type I signal peptidase: expression and evaluation of its diagnostic potential. *Veterinary Microbiology* **154**, 282–291.

Monteiro, K. M., de Carvalho, M. O., Zaha, A. and Ferreira, H. B. (2010). Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. *Proteomics* **10**, 1985–1999.

Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi, Y., Okuda, A., Kikuchi, A. and Matsumoto, K. (1995). Replication factor encoded by a putative oncogene, set, associated with myeloid leukemogenesis. *Proceedings of National Academy of Science of the United States of America* **92**, 4279–4283.

Nagata, K., Saito, S., Okuwaki, M., Kawase, H., Furuya, A., Kusano, A., Hanai, N., Okuda, A. and Kikuchi, A. (1998). Cellular localization and expression of template-activating factor I in different cell types. *Experimental Cell Research* **240**, 274–281.

Olson, P. D., Littlewood, D. T., Bray, R. A. and Mariaux, J. (2001). Interrelationships and evolution of the tapeworms (Platyhelminthes: Cestoda). *Molecular Phylogenetics Evolution* **19**, 443–467.

Olson, P. D., Zarowiecki, M., Kiss, F. and Brehm, K. (2012). Cestode genomics – progress and prospects for advancing basic and applied aspects of flatworm biology. *Parasite Immunology* **34**, 130–150.

Parrish, J., Limjindaporn, T., Hines, J., Liu, J., Liu, G. and Finley, R. J. (2004). High-throughput cloning of *Campylobacter jejuni* ORFs by *in vivo* recombination in *Escherichia coli*. *Journal of Proteome Research* **3**, 582–586.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Salamov, A. A. and Solovyev, V. V. (2000). Ab initio gene finding in *Drosophila* genomic DNA. *Genome Research* **10**, 516–522.

Spiliotis, M., Tappe, D., Brückner, S., Mösch, H. U. and Brehm, K. (2005). Molecular cloning and characterization of Ras- and Raf-homologues from the fox-tapeworm *Echinococcus multilocularis*. *Molecular and Biochemical Parasitology* **139**, 225–237.

Spiliotis, M., Konrad, C., Gelmedin, V., Tappe, D., Brückner, S., Mösch, H. U. and Brehm, K. (2006). Characterisation of EmMPK1, an ERK-like MAP kinase from *Echinococcus multilocularis* which is activated in response to human epidermal growth factor. *International Journal of Parasitology* **36**, 1097–1112.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biological and Evolution* **28**, 2731–2739.

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K. L., Tracey, A., Bobes, R. J., Fragoso, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H. M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T. A., Deplazes, P., Estrada, K., Fernández, C., Holland, P. W., Hou, J., Hu, S., Huckvale, T., Hung, S. S., Kamenetzky, L., Keane, J. A., Kiss, F. *et al*. (2013). The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* **496**, 57–63.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, RESEARCH0034.

Zheng, H., Zhang, W., Zhang, L., Zhang, Z., Li, J., Lu, G., Zhu, Y., Wang, Y., Huang, Y., Liu, J., Kang, H., Chen, J., Wang, L., Chen, A., Yu, S., Gao, Z., Jin, L., Gu, W., Wang, Z., Zhao, L., Shi, B., Wen, H., Lin, R., Jones, M. K., Brejova, B., Vinar, T., Zhao, G., McManus, D. P., Chen, Z., Zhou, Y. *et al*. (2013). The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nature Genetics* **45**, 1168–1175.

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

APPENDIX

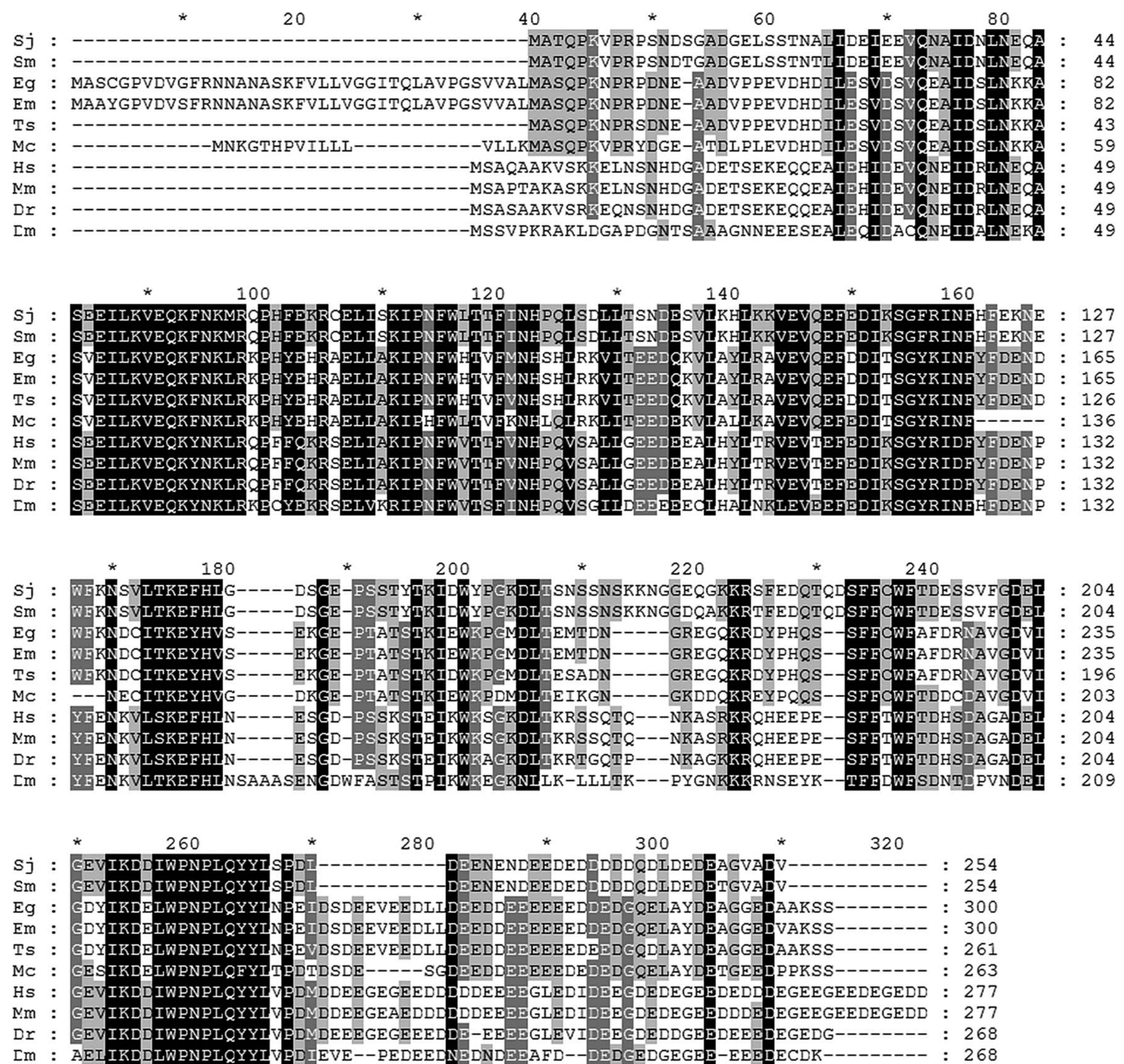

Fig. A1. Alignment of SET/TAF deduced amino acid sequences from *M. corti*, selected helminths and model organisms. For each alignment, amino acids which are conserved in all the aligned sequences are printed in white on a black background; amino acids conserved in at least five of the aligned sequences are printed in black on gray background; and amino acids conserved in at least eight of the aligned sequences are printed in white on gray background. The following sequences were used: *M. corti* (Mc) McSET/TAF amino acid sequence deduced from contig MCOS.contig.01068·30933; *Danio rerio* (Dr) NCBI no. NP_958876·1; *Drosophila melanogaster* (Dm) NCBI no. NT_033777·2; *Echinococcus granulosus* (Eg) GeneDB no. EgrG_000465500·1; *Echinococcus multilocularis* (Em) GeneDB no. EmuJ_000465500·1; *Homo sapiens* (Hs) NCBI no. NP_003002·2; *Mus musculus* (Mm) NCBI no. NP_001191804·1; *Schistosoma japonicum* (Sj) GeneDB no. Sjp_0002330·1; *Schistosoma mansoni* (Sm) GeneDB no. Smp_155060·2; and *Taenia solium* (Ts) GeneDB no. TsM_000881900·1.

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

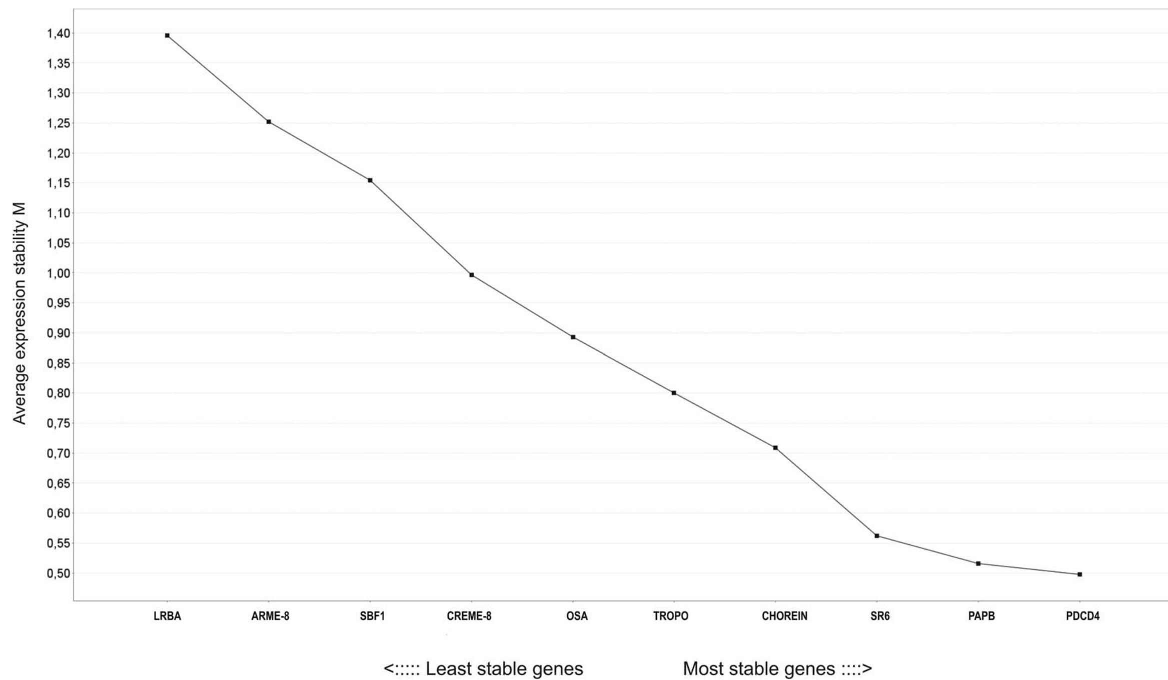

Fig. A2. Analysis of expression stability of potential *M. corti* reference genes by geNorm program. *M* values obtained for each of the genes analysed are indicated on the figure. Among the ten genes analysed, those with the lowest *M* values are *PDCD4* ($M = 0{\cdot}5$), *PAPB* ($M = 0{\cdot}53$), *SR6* ($M = 0{\cdot}57$), *CHOREIN* ($M = 0{\cdot}72$) and *TROPO* ($M = 0{\cdot}8$). The *LRBA* gene ($M = 1{\cdot}40$) was considered the less stable.

Table A1. Primer sequences used in RT-qPCR experiments, with corresponding Tm and amplicon sizes

Gene[a]	Primer sequences[b]	Tm (°C)	Amplicon size (bp)	Ref.
ARME-8	FW 5′-AGGATATTGCTCCCGGTGTA-3′	59·31	148	Livak and Schmittgen (2001)
	REV 5′-ACAGCTCCTCCTGACTTTCG-3′	58·95		
CHOREIN	FW 5′-CGGGCATTCAAATCCAGTAT-3′	59·78	130	Livak and Schmittgen (2001)
	REV 5′-GACTGTGGAAAAGGTGCAAAC-3′	59·63		
CRME-8	FW 5′-ACAGCCAGTGGTGTTGAAGA-3′	59·41	147	Livak and Schmittgen (2001)
	REV 5′-CCGCATGTCTGTTAGCAAGT-3′	59·60		
LRBA	FW 5′-CCGAACCGGTTGTTCTACTC-3′	59·59	140	Livak and Schmittgen (2001)
	REV 5′-AGACTGCCCAGGTCGTAGTC-3′	59·33		
OSA	FW 5′-AAAGTGTGTGTGTGCGATTG-3′	58·17	148	Livak and Schmittgen (2001)
	REV 5′-AGCTGAGGACTTCGCACITC-3′	59·75		
PABP	FW 5′-GGGATGCTTAITGGAATGGA-3′	59·72	131	Livak and Schmittgen (2001)
	REV 5′-ACGGAACCATTTGGAGGAT-3′	59·20		
PDCD4	FW 5′-GTTCTATCGGAGCTGCCTTG-3′	59·98	147	Livak and Schmittgen (2001)
	REV 5′-TTGCTCATGTCCTTCTGGAG-3′	58·96		
SBF1	FW 5′-TCTCACCAGTTTTCCTGCAA-3′	59·41	148	Livak and Schmittgen (2001)
	REV 5′- TCCGAGTCGTGTTTGAAGTC-3′	58·85		
SR6	FW 5′-ACCTGTCGTCTCGCGTTAGT-3′	59·94	145	Livak and Schmittgen (2001)
	REV 5′-TCAATGGCAGCITTCATGTC-3′	59·81		
TROPO	FW 5′-CTGAAGCTGAAGITGCAGCCATGA-3′	60·00	247	Koziol *et al.* (2011)
	REV5′-TCGTACTTACGTTCAGCATCCTCC-3′	60·00		
McSET/ TAF	FW 5′-GGAGAGCGTGGACTCTGTCCAA-3′	60·00	132	This work
	REV 5′-GCCAACAGTTCGGCACGAT-3′	60·00		

FW, forward; REV, reverse.

[a] *ARME-8*, receptor-mediated endocytosis-8 gene; *CHOREIN*, Chorea-acanthocytosis gene; *CRME-8*, receptor-mediated endocytosis-8 gene; *LRBA*, lipopolysaccharide-responsive and beige-like Anchor gene; *OSA*, Eyelid/OSA brahma complex gene; *PAPB*, Poly (A)-binding protein gene; *PDCD4*, Programmed cell death 4 gene; *SBF1*, Suvar3-complex gene; *SR6*, Splicing factor arginine/serine-rich 6 gene; *TROPO*, tropomyosin gene; *McSET/TAF, M. corti* SET/TAF-Iβ gene.

[b] Designed with Vector NTI Advance software v.10.

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press

Table A2. Structure and characteristics of McSET/TAF gene and encoded protein predicted by different algorithms

Gene prediction algorithm[a]	Protein	Contig[b]	Gene length (bp)[c]	CDS[d]	E1 (bp)	I1 (bp)	E2 (bp)	I2 (bp)	E3 (bp)	I3 (bp)	E4 (bp)	Protein length (aa)[e]	Molecular mass (Da)	pI
Gene Mark	McSET/ TAF	MCOS. contig.01068·30933	976	792	106	69	302	64	283	51	101	263	30205·4	4·37
FGENESH	McSET/ TAF	MCOS. contig.01068·30933	921	741	58	65	302	64	283	51	98	246	28337·0	4·29
GENSCAN	McSET/ TAF	MCOS. contig.01068·30933	814	738	318	37	310	39	110	–	–	245	28532·1	4·23

E#, exon number; I#, intron number.

[a] Gene prediction algorithms: GeneMark (http://exon.gatech.edu/GeneMark/); FGENESH (http://nhjy.hzau.edu.cn/kech/swxxx/jakj/dianzi/Bioinf6/GeneFinding/GeneFinding2.htm); GENSCAN (http://genes.mit.edu/GENSCAN.html).

[b] Contig number as recorded at the Sanger Institute (ftp://ftp.sanger.ac.uk/pub/project/pathogens/HGI/).

[c] Sum of exons and introns.

[d] Sum of exons.

[e] Length of the deduced amino acid sequence.

https://doi.org/10.1017/S003118201500030X Published online by Cambridge University Press