*Journal of Helminthology*

**www.cambridge.org/jhl**

# Detailed characterisation of the Co-Smad protein in liver fluke *Fasciola gigantica*

O. Japa[1,2], C. Phuangsri[1], K. Klinbumrung[2], K. Prakhammin[3] and R. J. Flynn[4,5]

[1]Division of Microbiology and Parasitology, School of Medical Sciences, University of Phayao, Phayao, Thailand; [2]Scientific Instrument and Product Standard Quality Inspection Center, University of Phayao, Phayao, Thailand; [3]Department of Applied Statistics, Rajamangala University of Technology Isan, Khon Kaen Campus, Khon Kaen, Thailand; [4]Department of Infection Biology, Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom and [5]Office of the Registrar, Hibernia College, Dublin, Ireland

**Research Paper**

**Cite this article:** Japa O, Phuangsri C, Klinbumrung K, Prakhammin K and Flynn RJ (2024). Detailed characterisation of the Co-Smad protein in liver fluke *Fasciola gigantica*. *Journal of Helminthology*, **98**, e93, 1–12
https://doi.org/10.1017/S0022149X24000841.

Received: 01 June 2024
Revised: 08 November 2024
Accepted: 19 November 2024

**Keywords:**
Co-Smad; *Fasciola gigantica*; FgSmad4; Transforming growth factor beta; mothers against decapentaplegic (Smad); Mothers against decapentaplegic homolog 4 (Smad4)

**Corresponding author:**
O. Japa;
Email: ornampai.ja@up.ac.th

## Abstract

Fasciolosis, caused by the liver flukes *Fasciola hepatica* and *Fasciola gigantica*, is a zoonotic parasitic disease associated with substantial economic losses in livestock. The transforming growth factor-beta signalling pathway is implicated in developmental processes and biological functions throughout the animal kingdom, including the *Fasciola* spp. It may also mediate host–helminth interactions during infection. In this work, we present an exploration of FgSmad4, the sole member of the Co-Smad protein family in *F. gigantica.* The isolated FgSmad4 cDNA was 4,014 bp in length encoding for a protein comprising 771 amino acids. FgSmad4 exhibited typical Co-Smad protein features, including Mad Homology 1 (MH1) and Mad Homology 2 (MH2) domains, a Nuclear Localisation Signal, a DNA-Binding Motif, and a Nuclear Export Signal. Sequence and phylogenetic analyses of FgSmad4 revealed that its MH1 and MH2 sequences are most similar to those of other trematode species. The MH1 domain, in particular, closely resembles the Co-Smad protein in mammalian hosts more than those in cestodes and nematodes. The expression patterns of *FgSmad4* during the liver fluke's developmental stages showed significant variation. Transcript levels were highest at the newly excysted juvenile stage, followed by unembryonated egg, redia, and metacercaria, with the lowest expression in the adult fluke, embryonated egg, and cercaria stages. Our results underscore the conservation and suggest the potential role of FgSmad4, a key transforming growth factor-beta signalling molecule within the liver fluke *F. gigantica.* As Co-Smad is typically involved in several biological pathways, the precise functions and mechanisms of this identified FgSmad4 necessitate further exploration.

## Introduction

Fasciolosis is a zoonotic parasitic disease caused by liver flukes in the genus *Fasciola*, of which *Fasciola hepatica* and *Fasciola gigantica* are the primary agents for this widely neglected parasitic disease (Mas-Coma et al. 2005). These parasites are widespread and have been identified across over 70 countries from all continents (Fürst et al. 2012; Logue et al. 2017). Although *F. hepatica* primarily presents in temperate, tropical, or subtropical regions such as Europe, South America, Middle East, and some part of Asia, *F. gigantica* is predominantly found in tropical climates including Africa, Asia, and the Middle East (Mas-Coma et al. 2009). In livestock, infection can lead to substantial economic losses, arising from factors such as morbidity, mortality, decreased productivity, fertility, and increased susceptibility to secondary infections (Beesley et al. 2018; Mas-Coma et al. 2019).

It is generally recognised that various molecules originating from the host are involved in growth and development of parasites that can influence the progression of parasitic infection (Ednilson Hilário et al. 2022). Numerous studies have demonstrated the existence of diverse growth factor receptors and the conservation of signalling pathways in parasitic nematodes and flatworms. Notably, transforming growth factor-beta (TGF-β), epidermal growth factor, and insulin have been identified (Konrad et al. 2003; Spiliotis et al. 2006; Zavala-Góngora et al. 2006). These pathways are believed to potentially mediate communication between the host and the helminth, influencing the dynamic interactions that unfold during infection and direct parasite intrinsic developmental events. Therefore, investigating these signalling components is important for understanding the complexities of parasitism, parasite development, and the identification of new targets for the development of strategies against these diseases (Salzet et al. 2000; You et al. 2011).

The signalling pathway of the TGF-β/bone morphogenetic protein (BMP) family is implicated in numerous biological processes across the animal kingdom (Huminiecki et al. 2009; Tzavlaki and Moustakas 2020). Signal transduction of all TGF-β proteins is initiated upon the binding of ligands to their cognate receptors and subsequent phosphorylation of the intracellular mediators known as mothers against decapentaplegic (Smads). These phosphorylated Smads then translocate into the nucleus and function as transcription factor regulating transcription of the

downstream target genes (Massagué 2012; Moustakas and Heldin 2009; Tzavlaki and Moustakas 2020).

Within the TGF-β signalling pathway, there are three types of Smad signalling molecules classified based on their functions: receptor-regulated Smads (R-Smads), common-partner Smad (Co-Smad), and inhibitory Smads (Miyazono 2000; Moustakas et al. 2001; Samanta and Datta 2012). Smad4, the Co-Smad present in vertebrates or Medea in *Drosophila* or SMA in *Caenorhabditis elegans* (Lagna et al. 1996; Zhang et al. 1997), acts as a common mediator for both BMP-specific and TGF-β/activins-specific R-Smads. Smad4 possesses a unique characteristic as it is unable to undergo phosphorylation or bind to either TGF-β or BMP receptors. However, it has the capacity to form heteromeric multimeric complexes with almost all activated R-Smads, enabling it to participate in the regulation of TGF-β signalling transduction (Wang et al. 2013).

In our previous works, we described the existence of the TGF-β protein family in liver flukes, specifically FhTLM in *F. hepatica* (Japa et al. 2015) and FgTLM in *F. gigantica* (Japa et al. 2022). These findings strongly suggest the presence of the signal mediator homologue components within the TGF-β pathway of these liver flukes, emphasising the importance of further exploring into this component of signalling molecules. In this study, we present our findings on the identification and characterisation of FgSmad4, the Co-Smad homologue in *F. gigantica*. Additionally, we investigated the gene expression patterns of the FgSmad4 throughout the various developmental stages of *F. gigantica*.

## Material and Methods

### *Ethics approval*

The use of animals in this study was reviewed and approved by the Animal Ethics Committee of the University of Phayao, Thailand, approval number 1-023-65.

### *Preparation of parasites*

#### Adult

Adult *F. gigantica* was obtained from naturally infected buffaloes in Phayao, Thailand, found in local abattoirs (19°11'18.00" N 99° 52'27.59" E). The liver flukes were removed from bile ducts and gall bladder stored in RPMI-1640 supplemented with gentamycin (10 µg/mL). After collection, the flukes underwent multiple washes with sterile phosphate-buffered saline and were then preserved in Trizol for subsequent RNA extraction and preparation of liver fluke eggs.

#### Unembryonated egg

Unembryonated eggs from the uterus were collected from mature flukes. The eggs were thoroughly rinsed with sterile distilled water until the fluid was transparent. The fluke eggs were collected and preserved in Trizol for RNA extraction. Additionally, aliquots of the fluke eggs were reserved and stored in the fridge for subsequent use in inducing the formation of embryonated eggs and miracidia.

#### Embryonated egg

The liver fluke eggs were induced to undergo embryonation through laboratory incubation under dark conditions at 30 °C for 14 days. Following this incubation period, the development of the eggs was observed using a stereo microscope; only those that had undergone embryonation were selected for RNA preparation.

#### Miracidium

The *F. gigantica* miracidia were obtained through egg hatching protocol as outlined by Moxon et al. (2010). In brief, after inducing embryonation in the eggs for 14 days, they were exposed to direct light at room temperature to stimulate miracidia hatching. Following this, the miracidia were examined under a stereo microscope and collected in Trizol for further experimentation.

#### Intra-molluscan larval stages (redia, cercaria) and metacercaria

The larval stages of *F. gigantica*, including rediae, cercariae, and metacercariae, were generated from experimentally infected snails. Specifically, snails of the *Radix* (*Lymnaeae*) *rubiginosa* species were maintained in the laboratory and used to establish *F. gigantica* infection. Laboratory infection of the *F. gigantica* was conducted according to Japa et al. (2022). After approximately 45 days of infection, the infected snails were monitored for cercarial emission under a stereo microscope. The released cercariae were immediately collected and stored in Trizol for RNA extraction.

To prepare metacercaria, a cellophane sheet was placed over the water surface to facilitate cercarial attachment and their subsequent transformation into metacercariae. Following this, the *F. gigantica* infected snails were dissected to retrieve the redial stages. The larval stages of *F. gigantica* acquired from the snails were preserved in Trizol and stored at –20 °C until they were ready for RNA extraction and subsequent analysis.

#### Newly excysted juvenile (NEJ)

The newly excysted juveniles of *F. gigantica* were obtained through an *in vitro* excystation process of the *F. gigantica* metacercariae as outlined by McVeigh et al. (2014) and Japa et al. (2022). Initially, the outer cysts were manually removed from the metacercariae. Afterward, the metacercariae were incubated in a 0.5% bleach solution for 4 min, followed by extensive washing with sterile water (five washes) and transferring to a new Petri dish. They were thoroughly suspended in an excystation solution and incubated for 1 hour at 37 °C. The NEJs were observed under stereo microscope and then collected in RPMI-1640 containing gentamycin (10 µg/mL). The NEJs underwent multiple washes with 1X D-phosphate-buffered saline and were finally preserved in Trizol for RNA extraction.

### *RNA isolation*

Total RNA from each life stage of *F. gigantica* including adult, unembryonated egg, embryonated egg, miracidium, redia, cercaria, metacercaria, and NEJ was isolated using Trizol reagent and the RNeasy Micro Kit (Qiagen, Germany) with some modifications. The parasite was homogenised in 1 mL of Trizol reagent. The homogenate was then centrifuged at 12,000*g* for 5 min. The resulting supernatant was combined with 200 µL of chloroform and incubated at room temperature for 15 min, followed by another centrifugation for 5 min. The clear upper phase containing RNA was thoroughly mixed with an equal volume of 70% ethanol before being transferred to an RNeasy MinElute spin column and centrifuged for 1 min. Then, 350 µL of Buffer RW1 was added to the column before centrifugation as previously described. The flowthrough was discarded, and the column was loaded with PRE buffer and centrifuged. The column was then washed twice with 80% ethanol. The RNA was eluted from the column by 100 µL of sterile diethylpyrocarbonate-treated water. The purified RNA samples from each life stage were stored at –80 °C until further use.

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

### Identification of the Co-Smad in the F. gigantica genome

To identify the Co-Smad sequence within the *F. gigantica* genome, we conducted a tblastn search using mammalian Smad4 protein sequences as the query sequence against the *F. gigantica* genome database available at https://parasite.wormbase.org/. Afterwards, gene-specific primers (GSPs) for the isolation of FgSmad4 via 5′ and 3′ rapid amplification of cDNA ends (RACE) were designed based on partial sequences acquired from tblastn searches against the *F. gigantica* genome/cDNA database.

### Isolation of FgSmad4 cDNA

Total RNA extracted from adult *F. gigantica* specimens was used for the synthesis of first-strand cDNA library using the SMARTer RACE 5′/3′ Kit (Takara, Japan) following the manufacturer's protocol.

The 5′ and 3′ ends of the FgSmad4 cDNA were amplified through two rounds of nested polymerase chain reaction (PCR) using GSP primers (10 µM) and the corresponding primers supplied by the kit. The GSP primer sequences are provided in Table 1. The PCR cycle for the first-round RACE amplification were as follows: pre-denaturation at 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 65 °C for 30 s, 72 °C for 3 min, and a final extension at 72 °C for 10 min. The nested PCR was carried out using 1 µL of the initial PCR product along with nested GSP primers targeting both ends. The nested PCR conditions consisted of 25 cycles following the same program used in the first PCR.

### Cloning and Sequencing

The 5′ and 3′ RACE PCR products were excised and purified from the agarose gel using a gel extraction kit (Qiagen, Germany). The purified PCR products were then cloned into the pRACE vector and transformed into Stellar competent cells. Transformed colonies were selected, and plasmids containing the inserts were isolated using the Wizard Plus SV minipreps DNA purification system (Promega, UK). Subsequent sequencing was carried out bi-directionally using promoter primers, conducted by U2Bio (Korea).

### Bioinformatics analyses

The nucleotide sequence was manually edited and annotated; overlapping sequence of 5′ and 3′ RACE was removed. The complete sequence of FgSmad4 cDNA was deposited in the GenBank database under accession no PP856693. The complete cDNA sequence of FgSmad4 was subsequently translated into the predicted amino acid sequence using the Expasy translation tools (https://web.expasy.org/translate/). Both nucleotide and protein sequences of FgSmad4, were performed homology searches using the blast software available at the National Centre for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST/). The nucleotide sequences were assessed using blastn against entries in the GenBank and WormBase ParaSite databases (https://parasite.wormbase.org/Multi/Tools/Blast). Similarity search of the protein sequence was conducted using PSI-blast and blastp at NCBI.

To determine exon–intron organisation, the cDNA sequences were used for blastn searches against the *F. gigantica* genome. This was carried out through online searches on the WormBaseParaSite database and local blastn analysis within BioEdit (Hall 1999; Hall et al. 2011). A graphical representation of the gene organisation was generated using GeneMaper 2.5.

### In silico analyses of Fgsmad4 protein properties

The FgSmad4 protein sequence was analysed for its properties using available online tools. Conserved domain prediction was conducted using CD-Search at the NCBI website (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The molecular weight and theoretical isoelectric point were predicted using the tool available at ExPASy (https://web.expasy.org/compute_pi/). Other predictions for general physiological and chemical properties, including amino acid composition, estimated instability index, and grand average of hydropathicity, were performed using the ProtParam tool on the ExPASy website (https://web.expasy.org/protparam/).

### Prediction of two-dimensional and three-dimensional structures of FgSmad4

The FgSmad4 model structure was predicted *in silico* using the Phyre2 online tool (intensive method) available at http://www.sbg.bio.ic.ac.uk/~phyre2 (Kelley et al. 2015). Visualisation of the modeled protein was carried out using UCSF Chimera version 1.16 (Pettersen et al. 2004).

### Phylogenetic analysis

Sequences of Co-Smad protein from various parasitic helminths and their mammalian hosts were included into the phylogenetic tree analyses. Multiple alignment of these sequences was conducted using the Muscle method embedded in Seaview (Gouy et al. 2009). All positions with gaps or incomplete data were excluded. Following the removal of informative sites, a total of 65 and 106 amino acids corresponding to conserved sequences in the MH1 and MH2 domains were analysed. The *Drosophila melanogaster* DAD sequence [BAA22841.1] was incorporated as the outgroup for the phylogenetic construction of both MH1 and MH2. The PhyML algorithm implemented in Seaview was utilised to generate the maximum likelihood phylogenetic tree, using the WAG model

**Table 1.** Primer sequences used in 5′/3′ RACE experiments and life stage expression of FgSmad4

| Primer name | | Sequence (5′ – 3′) | bp |
|---|---|---|---|
| GSP primers | GSP_ Fgsmad4F1 | GATTACGCCAAGCTTCATTCCCAGCTCTGGATCCGTTCCAGTT | 43 |
| | GSP _ FgSmad4F2 | GATTACGCCAAGCTTGACTGAGCTGTCTGGACCGACGATTA | 41 |
| | GSP _ FgSmad4R1 | GATTACGCCAAGCTTGACCGGAGAGCTCGGCATAACTGGAA | 41 |
| | GSP _ FgSmad4R2 | GATTACGCCAAGCTTCTGGAACGGATCCAGAGCTGGGAATG | 41 |
| Life stage primers | FgSmad4F | CCAGTTGTCCAACGTGCAT | 19 |
| | FgSmad4R | CAAGTTGGGCTGAATGTGCTAA | 22 |

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

(Whelan and Goldman 2001). The reliability of branching order was assessed by bootstrap analysis consisting of 1,000 replicates. The final tree visualisation was generated using FigTree software version 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

### Reverse transcription PCR

First-strand cDNA was synthesised from RNA extracted at each life cycle stage of *F. gigantica*, including unembryonated egg, embryonated egg, miracidium, redia, cercaria, metacercaria, NEJ, and adult. The synthesis reaction was performed in a 20-µL volume using the GoScript Reverse Transcription System (Promega, UK), following the manufacturer's instructions.

### Quantification of FgSmad4 gene expression levels with real-time PCR

Life stage expression primer for *FgSmad4* was designed from the conserved sequence flanking intron specific for FgSmad4 cDNA to yield PCR products of 275 bp. Standard PCR using Taq DNA polymerase (Qiagen, Germany) and subsequent sequencing were performed to confirm the correct PCR product and to ensure the absence of non-specific PCR product or primer dimers.

The real-time PCR was carried out in a volume of 20 µL using QuantiNova SYBR Green PCR Kit (Qiagen, Germany) consisting of 10 µL of 2x SYBR Green PCR Master mix. The amplification reaction was performed in a real-time PCR (Bio-Rad, CFX96). The PCR condition was used with the following conditions: 2 min at 94 °C for denaturation, 40 cycles of 94 °C for 5 s, 57 °C for 10 s, and 72 °C for 30 s, and 10 min of a final extension step at 72 °C.

To quantify the relative expression level of *FgSmad4* transcript in the life stage of *F. gigantica*, we used $2^{-\Delta CT}$ method as described by Silver et al. (2006). The expression of *FgSmad4* was normalised by *F. gigantica* glyceraldehyde-3-phosphate dehydrogenase (*FgGAPDH*). Each reaction was performed in triplicate (technical replicates), and for each developmental stage, three independent biological replicates were conducted.

Life stage expression data of *FgSmad4* were presented as mean ± standard error of the mean derived from three independent experiments. A bar chart was created using GraphPad Prism 8.0.1 (GraphPad Software, San Diego, CA, USA). Statistical analysis was conducted using one-way ANOVA, with p-values calculated using Tukey's post hoc test, and significance was considered for p-values < .05 in all statistical comparisons.

## Results

### Database search for F. gigantica Co-Smad sequences (FgSmad4)

Nucleotide sequences of Smad4 from mammalian hosts of *F. gigantica*, including human (*Homo sapiens* Smad4 [NP_001393970.1]), cattle (*Bos taurus* Smad4 [NP_001069677.1]), and buffalo (*Bubalus bubalis* Smad4 [NP_001277793.1]), retrieved from the GenBank database, were employed in a tblastn search. The search identified corresponding nucleotides in the genome of *F. gigantica*, PRJNA230515 - Uganda_cow_1. tblastn analysis revealed three matching regions within the FGIG_03163 transcript, located at positions 124-498 nt, 1474-1872 nt, and 2038-2262 nt, with identity scores of 77.6%, 72.2%, and 70.7%, respectively. Diagrams detailing the matching positions are presented in Figure 1.

### Identification and Isolation of FgSmad4 cDNA

The isolated FgSmad4 cDNA consisted of 4,014 bp in length encoding for a protein comprising of 771 amino acids. Within the FgSmad4 cDNA, there was a 5′-untranslated region (UTR) of 51 nt in length, an open reading frame spanning 2,316 nt, and a 3' UTR of 1,647 nt, followed by a poly A tail. The start codon (ATG) and stop codon (TGA) were predicted to be located at nucleotides 52–54 and 2,365–2,367, respectively.

### In silico characterisation of the predicted FgSmad4 protein

The FgSmad4 cDNA encoded a 771 amino acid polypeptide, predominantly composed of glycine (11.3%), serine (11.3%), proline (8.8%), leucine (8.0%), and alanine (6.4%). The protein sequence analyses revealed the conservation of key Smad protein domain features, including the Mad Homology 1 (MH1) domain at the N-terminal and the Mad Homology 2 (MH2) domain at the C-terminal, with FgSmad4 consisting of 125 amino acids in the MH1 domain (33–157 amino acids) and 246 amino acids in the MH2 domain (491–736 amino acids). Additionally, essential structural elements of the Co-Smad protein subfamily were identified, including the Nuclear Localisation Signal (NLS; FARRAIESLVKKLKEKRED) and a DNA-Binding Motif (DBM; RTLDGRMQIAG) in the MH1 domain, along with the Nuclear Export Signal (NES; VDLAALSL) situated at the initiation of the linker region (Figure 2).

Further computational analyses predicted a molecular weight of 82.06 kDa and a theoretical isoelectric point of 6.43 for the translated FgSmad4 protein. Additionally, the deduced FgSmad4 protein was characterised as hydrophilic, with a grand average of hydropathicity value of –0.411. The calculated instability index was 55.14, predicting overall protein instability.

### Sequence comparison of FgSmad4 with other Co-Smad proteins

The FgSmad4 identified in this study was identical to mothers against decapentaplegic of *F. gigantica* in Genbank database [TPP57878.1]. Similarity search by PSI-blast indicated that the entire sequence of FgSmad4 was similar to mothers against decapentaplegic of *Fasciolopsis buski* [KAA0186896.1] and *Schistosoma haematobium* [XP_051073969.1] with similarity of 79% and 53.37%, respectively.

Comparison of the MH1 domain, FgSmad4 exhibited the highest similarity with other Smad4 proteins from trematodes, particularly CsSmad4 and OvSmad4 from *Clonorchis sinensis* [KAG5446872.1] and *Opisthorchis viverrini* [KER20450.1], showing similarities of 84.80%, and 84.00%, respectively. The FgSmad4 protein displayed an average of 80.00% homology in sequence to the MH1 region of Co-Smads from mammals (HsSmad4, BtSmad4, and BbSmad4) but revealed lower identity to cestodes (66.67%) and nematodes (50.00%–62.40%).

In the MH2 domain of FgSmad4, the greatest similarities were observed within platyhelminthes (74.80%–96.77%), with similarities of 85.55%, 86.21%, and 96.77% to *C. sinensis*, *O. viverrini*, and *F. buskii*, respectively. The similarity for cestodes was found to be 74.80%. The MH2 domain showed closer homology to mammals (65.45%) than to nematodes (34.96%–43.90%) (Table 2).

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

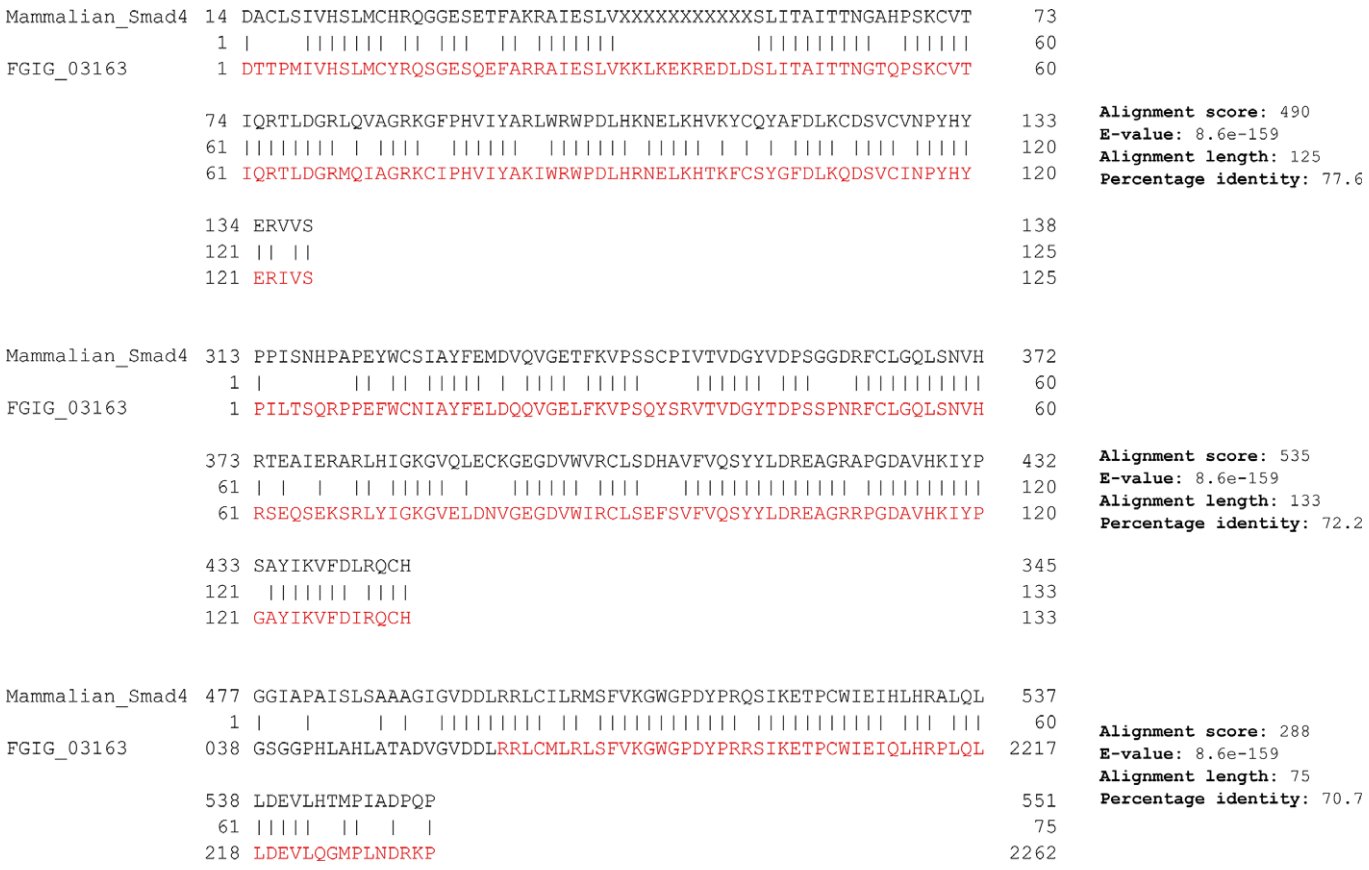


**Figure 1.** The tblastn results illustrate the corresponding positions identified in the *F. gigantica* genome. The red highlighted positions represent matches found through a tblastn search using mammalian Smad4 as the query sequence.

### *Genomic structure of* FgSmad4

The *FgSmad4* gene organisation was determined by blastn search of FgSmad4 cDNA sequences against the *F. gigantica* genome database via online and local analyses. Within the *FgSmad4* gene, there were nine exons, covering a total length of approximately 56.64 kb. The sizes of these nine exons were: 63, 294, 215, 194, 723, 193, 189, 357, and 1,786 bp, respectively (Figure 3).

### *Multiple sequence alignment*

To identify conserved domain sequences between FgSmad4 and related species, multiple sequence alignments were performed on Co-Smad sequences from various organisms. The alignment revealed that FgSmad4 shares 50%–85% amino acid identity in the MH1 domain and 36%–97% in the MH2 domain. Additionally, conserved amino acid domains were also found to be present in other Co-Smad homologues across different organisms (Figure 4).

### *Phylogenetic relationships of FgSmad4*

The phylogenetic relationships of FgSmad4 with other Co-Smad proteins were assessed by constructing a phylogenetic tree based on sequences in the MH1 and MH2 domains. The tree incorporated sequences of FgSmad4 and other Co-Smad proteins from various organisms, including mammalian hosts (human, cattle, buffalo), free-living nematodes (*C. elegans*), parasitic nematodes (*H. contortus*), free-living flatworms (Planaria) and parasitic flatworms (trematode, cestode).

In the MH1 domain analysis, FgSmad4 exhibited its closest relationship with homologues of Smad4 from parasitic trematodes, supported by a high bootstrap value. Notably, in comparison to other organisms, FgSmad4 demonstrated a closer relationship to mammalian hosts such as HsSmad4, BbSmad4, and BtSmad4 than to those of cestodes and nematodes (Figure 5A).

In the phylogenetic analyses of the MH2 region, Co-Smad proteins were categorised into three major branches according to taxonomic classification: Mammalian hosts, Nematodes, and Platyhelminthes. FgSmad4 formed a distinct subgroup within the Platyhelminthes branch, positioning itself with the trematode Smad4 proteins and separating from free-living flatworms and cestodes (Figure 5B).

### *Predicted two-dimensional and three-dimensional structures*

Structural analysis of FgSmad4 protein conducted via the online Phyre2 tool revealed its secondary structure composition, with 14% beta strands, 12% alpha helices, and 2% transmembrane helices. The majority of the core structure was predicted to be disordered.

The three-dimensional model of FgSmad4 (Figure 6) was visualised using Chimera software. This modelled protein was generated based on 216 amino acids (75% identity) from the conserved Smad domain of *H. sapiens* Smad4 (d1dd1a), with a maximum confidence score of 100%.

### *Life stage expression of* FgSmad4

Quantitative reverse transcription was performed to evaluate the expression of the *FgSmad4* gene in different developmental stages

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

```
>FgSmad4_CDS

1     ATG ATT GTC TTA TCC AAC GCC GCT GTG ATC AGC AAC GCT GTC AAC GAC AGT GCA ATG CCG TCT TCT GGG TTT GAA GCC GAT TTT TTT AAC   90
1      M   I   V   L   S   N   A   A   V   I   S   N   A   V   N   D   S   A   M   P   S   S   G   F   E   A   D   F   F   N   30

91    AAC CCG GAT ACT ACC CCT ATG ATT GTA CAC AGT TTA ATG TGC TAT CGT CAA AGT GGC GAG TCT CAG GAA TTT GCC CGG AGG GCC ATT GAA   180
31     N   P   D   T   T   P   M   I   V   H   S   L   M   C   Y   R   Q   S   G   E   S   Q   E   F   A   R   R   A   I   E   60

181   AGT CTG GTA AAA AAG CTC AAG GAA AAA CGG GAG GAT CTT GAT TCA TTA ATT ACT GCG ATC ACC ACT AAT GGA ACT CAA CCC AGC AAA TGT   270
61     S   L   V   K   K   L   K   E   K   R   E   D   L   D   S   L   I   T   A   I   T   T   N   G   T   Q   P   S   K   C   90

271   GTG ACG ATT CAG CGA ACG TTG GAT GGC AGG ATG CAG ATT GCC GGT AGA AAG TGT ATT CCT CAC GTA ATT TAC GCG AAA ATT TGG AGG TGG   360
91     V   T   I   Q   R   T   L   D   G   R   M   Q   I   A   G   R   K   C   I   P   H   V   I   Y   A   K   I   W   R   W   120

361   CCC GAC TTG CAT CGG AAC GAA TTG AAA CAT ACT AAA TTC TGC TCT TAC GGG TTT GAC CTG AAG CAG GAT AGT GTT TGC ATA AAT CCG TAC   450
121    P   D   L   H   R   N   E   L   K   H   T   K   F   C   S   Y   G   F   D   L   K   Q   D   S   V   C   I   N   P   Y   150

451   CAT TAT GAA CGC ATT GTT TCT TCA GTT GAC TTG GCA GCA CTC AGC TTG AGT CCT TGT TTT GAT CAC GAA GCT CAA ACA GAA TAT TCC CGA   540
151    H   Y   E   R   I   V   S   S   V   D   L   A   A   L   S   L   S   P   C   F   D   H   E   A   Q   T   E   Y   S   R   180

541   GAT CAA AGC GAT AGT GAT TCA GAA CTG CCT GAT TCC ACA GCT ACA GCC TCG TTA GAT TTG ACA CCA AAC GAG TCC CAC ACG ACA AAC GCC   630
181    D   Q   S   D   S   D   S   E   L   P   D   S   T   A   T   A   S   L   D   L   T   P   N   E   S   H   T   T   N   A   210

631   GGG GAT ATG ATA TCT ACT CAG CGT CGC CCG CAT CTT ACC GTA CAG GCA CTC CTA GAC GAA TGT GAT TCA TCC ATT CAC AAC GGT TCT GCA   720
211    G   D   M   I   S   T   Q   R   R   P   H   L   T   V   Q   A   L   L   D   E   C   D   S   S   I   H   N   G   S   A   240

721   GGT TCC AGC CCA CAG AAA TGT ATG ACT CCA CCT GAA GAC AAC ATG GGT TGT TTT GCT CAG TCT ACC GGT TAT CCC AAT CAT CTT CAT CCG   810
241    G   S   S   P   Q   K   C   M   T   P   P   E   D   N   M   G   C   F   A   Q   S   T   G   Y   P   N   H   L   H   P   270

811   CTT CTG ACC GAT CGT GTA AAT CCC GCA CCA CAT TTA GAT CGT GGC CGA AAC ACG ATG GCT TGG TCT GGT GTT GCA GTC CCT GGC CTT CAT   900
271    L   L   T   D   R   V   N   P   A   P   H   L   D   R   G   R   N   T   M   A   W   S   G   V   A   V   P   G   L   H   300

901   CAT CCA TCC CAT AAT GGT GTT GGA TTA CCT CAC ATT CCC AGC TCT GGA TCC GTT CCA GTT ATG CCG AGC TCT CCG GTC AGG ACT GAG CTG   990
301    H   P   S   H   N   G   V   G   L   P   H   I   P   S   S   G   S   V   P   V   M   P   S   S   P   V   R   T   E   L   330

991   TCT GGA CCG ACG ATT ACA CCA CGA CCA ACA CAA AAT CCT TAT TCC ACC ATA ATA AAG ACA CTG GTT ATG CCA GAA ACA GAC TTC TTT TCC   1080
331    S   G   P   T   I   T   P   R   P   T   Q   N   P   Y   S   T   I   I   K   T   L   V   M   P   E   T   D   F   F   S   360

1081  GGT TCA TCT TCC AGT GAT TAC AGT TTA TCC TCA GGT ACT CCG ATA ATC GTA CTG AGC GGA AGT GCG AAT CAG CAT GGC GGT GGA GGT GGA   1170
361    G   S   S   S   S   D   Y   S   L   S   S   G   T   P   I   I   V   L   S   G   S   A   N   Q   H   G   G   G   G   G   390

1171  GGT TCA GGA GAC GGA GGA AAC GGT GGC GGC GGT GGC GGT AGC AGT GGT GGT GGT ACC GGT TCT AAC AGC GGG TCT TGT GGG AAT AGC GGA   1260
391    G   S   G   D   G   G   N   G   G   G   G   G   G   S   S   G   G   G   T   G   S   N   S   G   S   C   G   N   S   G   420

1261  CGT AAT GCG GCC GGA TCT AAC TTC ACC GCG AGT ACT TGC GGT GCG AGC GGA GGA AGC GGA GGA GGG GGA GTT GGC GGT GGT GGT CTC CCG   1350
421    R   N   A   A   G   S   N   F   T   A   S   T   C   G   A   S   G   G   S   G   G   G   G   V   G   G   G   G   L   P   450

1351  GCT GGT GCC GGT GGG GGA GGT GGA GGC AAC TGG GGT AAT CGC GGG GGT CCA CCA CAA CCC CCG TTC ACT CCT TCG GGA TCA TTA CTT CAG   1440
451    A   G   A   G   G   G   G   G   G   N   W   G   N   R   G   G   P   P   Q   P   P   F   T   P   S   G   S   L   L   Q   480

1441  CCA GTT CCA ATT TTG ACG TCT CAG AGA CCT CCG GAA TTT TGG TGC AAT ATT GCC TAT TTC GAA CTA GAT CAA CAG GTT GGT GAG CTG TTT   1530
481    P   V   P   I   L   T   S   Q   R   P   P   E   F   W   C   N   I   A   Y   F   E   L   D   Q   Q   V   G   E   L   F   510

1531  AAA GTG CCC AGT CAA TAC TCC CGT GTC ACA GTG GAC GGT TAC ACC GAT CCT TCC AGT CCA AAT AGA TTT TGT CTC GGC CAG TTG TCC AAC   1620
511    K   V   P   S   Q   Y   S   R   V   T   V   D   G   Y   T   D   P   S   S   P   N   R   F   C   L   G   Q   L   S   N   540

1621  GTG CAT CGG TCC GAA CAA TCG GAG AAG TCT CGT CTG TAC ATC GGC AAA GGT GTG GAA TTG GAC AAT GTG GGT GAG GGT GAT GTA TGG ATT   1710
541    V   H   R   S   E   Q   S   E   K   S   R   L   Y   I   G   K   G   V   E   L   D   N   V   G   E   G   D   V   W   I   570

1711  CGC TGC CTG TCC GAA TTT TCT GTG TTT GTA CAA AGC TAC TAT TTG GAT CGA GAA GCT GGC CGA CGA CCG GGT GAT GCA GTG CAC AAA ATT   1800
571    R   C   L   S   E   F   S   V   F   V   Q   S   Y   Y   L   D   R   E   A   G   R   R   P   G   D   A   V   H   K   I   600

1801  TAT CCG GGC GCT TAT ATC AAG GTG TTT GAT ATT CGA CAG TGC CAT GAA CAA ATG AAG CTC TTA GCA CAT TCA GCC CAA CTT GCT GCC GAG   1890
601    Y   P   G   A   Y   I   K   V   F   D   I   R   Q   C   H   E   Q   M   K   L   L   A   H   S   A   Q   L   A   A   E   630

1891  CAC CAG GCC GCA GTT GTG GTA GGC TCT TTA CCT TCT CCG AAT ACC GCC CCA CCT TTG GCC CCA CTG CTT TCC CAA CTC CCG CCC GGC CCG   1980
631    H   Q   A   A   V   V   V   G   S   L   P   S   P   N   T   A   P   P   L   A   P   L   L   S   Q   L   P   P   G   P   660

1981  TCA CAA AGC GGG TCA TTG GCT GCG GGC GCT GGT TCC GGA GGT CCG CAT CTG GCC CAT CTA GCC ACC GCG GAT GTC GGT GTG GAC GAT TTG   2070
661    S   Q   S   G   S   L   A   A   G   A   G   S   G   G   P   H   L   A   H   L   A   T   A   D   V   G   V   D   D   L   690

2071  CGC CGT TTG TGC ATG CTA CGA TTG AGC TTT GTC AAA GGA TGG GGA CCG GAC TAT CCG CGT CGT AGT ATC AAG GAG ACA CCG TGC TGG ATT   2160
691    R   R   L   C   M   L   R   L   S   F   V   K   G   W   G   P   D   Y   P   R   R   S   I   K   E   T   P   C   W   I   720

2161  GAG ATC CAA TTG CAT CGT CCT TTG CAG TTG TTG GAC GAG GTG CTA CAG GGA ATG CCA TTG AAT GAT CGG AAG CCA ACA CGT CAC TTT TTC   2250
721    E   I   Q   L   H   R   P   L   Q   L   L   D   E   V   L   Q   G   M   P   L   N   D   R   K   P   T   R   H   F   F   750

2251  CCA TAT TTC TCC CAA CCA TCG GTG TGT CCT CCG GCC CCC AGG CCA CAA ACC GCC CGA CGC CCT TGA   2316

751    P   Y   F   S   Q   P   S   V   C   P   P   A   P   R   P   Q   T   A   R   R   P   *
```

**Figure 2.** Translated amino acid sequences of FgSmad4 derived from a 2,316-bp open reading frame (ORF). The start codon (ATG) and stop codon (TGA) are indicated in underlined bold text. The conserved MH1 domain is shaded with a grey background, whereas the MH2 domain is highlighted with black letters on a pink background. Nuclear localization signals (NLS) are represented by red letters within a red box, and nuclear export signals (NES) are boxed in green. The DNA binding domain is highlighted within light blue block.

of *F. gigantica*, including unembryonated egg, embryonated egg, miracidium, cercaria, metacercaria, NEJ, and adult. For normalisation, *FgGAPDH* was chosen as a reference gene due to its consistent expression levels throughout the entire lifecycle.

The expression of FgSmad4 was detected in all examined stages of *F. gigantica*. The highest expression was observed in the NEJ stage, followed by unembryonated egg, metacercaria, miracidium, and redia, respectively. In contrast, the expression of *FgSmad4* was

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

**Table 2.** Comparison of FgSmad4 and other Co-Smad homologues from selected organisms

| Organisms | Homologues protein | Accession number | Length (aa) | MH1 homology | MH2 homology |
|---|---|---|---|---|---|
| Mammals | | | | | |
| *Homo sapiens* | Smad4 | [NP_001393970.1] | 552 | 96/120 (80.00%) | 161/246 (65.45%) |
| *Bos taurus* | Smad4 | [NP_001069677.1] | 553 | 96/120 (80.00%) | 161/246 (65.45%) |
| *Bubalus bubalis* | Smad4 | [NP_001277793.1] | 553 | 96/120 (80.00%) | 161/246 (65.45%) |
| Nematodes | | | | | |
| *Caenorhabditis elegans* | SMA4 | [NP_001040864.1] | 565 | 76/125 (60.80%) | 102/246 (41.46%) |
| *Caenorhabditis elegans* | DAF3 | [NP_001300343.1] | 858 | 53/106 (50.00%) | 108/246 (43.90%) |
| *Haemonchus contortus* | SMA4 | [CDJ96373.1] | 589 | 78/125 (62.40%) | 104/246 (42.28%) |
| *Haemonchus contortus* | DAF3 | [QGW58250.1] | 698 | 73/125 (58.40%) | 86/246 (34.96%) |
| Trematodes | | | | | |
| *Clonorchis sinensis* | Smad4 | [KAG5446872.1] | 812 | 106/125 (84.80%) | 225/263 (85.55%) |
| *Opisthorchis viverrini* | Smad4 | [KER20450.1] | 790 | 105/125 (84.00%) | 225/261 (86.21%) |
| *Fasciolopsis buski* | Smad4 | [KAA0186896.1] | 741 | 85/125 (68.00%) | 240/248 (96.77%) |
| *Schistosoma mansoni* | Smad4 | [XP_018648204.1] | 798 | 101/125 (80.80%) | 206/246 (83.73%) |
| Cestodes | | | | | |
| *Echinococcus granulosus* | SmadD | [AEW27102.1] | 719 | 80/120 (66.67%) | 184/246 (74.80%) |
| *Echinococcus multilocularis* | SmadD | [CAK32532.1] | 719 | 80/120 (66.67%) | 184/246 (74.80%) |

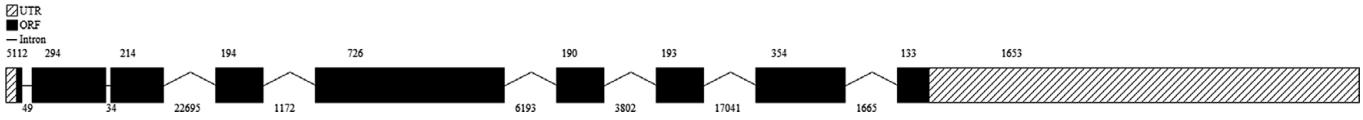


**Figure 3.** Genomic organization of *FgSmad4*. Exons, introns, and UTRs are depicted in shaded boxes, lines, and striped boxes, respectively. Numbers denote the size of exons and introns in base pairs.

lowest in the adult liver fluke. Notably, a significant disparity in expression levels between NEJ and adult was observed, with NEJ exhibiting a 100-fold higher expression. Both miracidium and redia displayed similar levels of *FgSmad4* expression, which were nearly 10 times higher than in adult stage. Among the larval stages of *F. gigantica*, NEJ exhibited the highest level of *FgSmad4* expression, which was approximately 61 times greater than in cercaria (Figure 7).

## Discussion

The Smad proteins are essential components of TGF-β and BMP signalling transductions with Co-Smad being a central mediator in these processes. In the present study, we identified and characterised a Co-Smad homologue from *F. gigantica* termed FgSmad4. From our results, it appears that FgSmad4 represented only one type of the Co-Smad observed in *F. gigantica* genome. This aligned with similar observations in other trematodes, such as *S. mansoni*, where a single Co-Smad has been identified in the genome (Osman et al. 2004). In the case of cestode, *Echinococcus granulosus* and *E. multilocularis* genomes were found to contain only a single Co-Smad known as EgSmadD and EmSmadD, respectively (Zavala-Góngora et al. 2008; Zhang et al. 2014). Our findings are in agreement with the recent study of Wu et al. (2023) who documented a similar finding in the number of Co-Smad in *F. gigantica*. However, we have here provided further details on the characteristics of FgSmad4.

The specific number of Co-Smad proteins and their significance can vary not only between phyla but also within species and genera. Research conducted by Masuyama et al. (1999) on *Xenopus* spp. revealed the existence of two Co-Smads, XSmad4α and XSmad4β, sharing a 70% similarity in amino acid sequences, the author suggested that these proteins may have overlapping but distinct functions. In most mammals, the presence of a single Co-Smad highlights a notable contrast to the diversity observed in *Xenopus* spp., emphasising the variability in Co-Smad configurations across different organisms. Invertebrates, with their simpler anatomical structures compared to mammals, tend to have less complex TGF-β signalling pathways. Notably, both free-living and parasitic flatworms, similar to mammals, possess only one Co-Smad. In contrast, parasitic and free-living nematodes, such as *C. elegans* (Ce-SMA-4 and Ce-DAF-3) and *Haemonchus contortus* (Hc-DAF-3), have been reported to contain multiple Co-Smad molecules, each specialised for their specific biological processes (Di et al. 2019). This diversity may reflect the evolutionary adaptations and functional requirements of various organisms in their respective ecological niches.

The FgSmad4 sequence obtained from 5' and 3' RACE was found to encode 771 amino acids, which aligned closely with the entry sequence FGIG_03163 in the *F. gigantica* genome database (PRJNA230515). Our result suggests that FgSmad4 likely has a

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

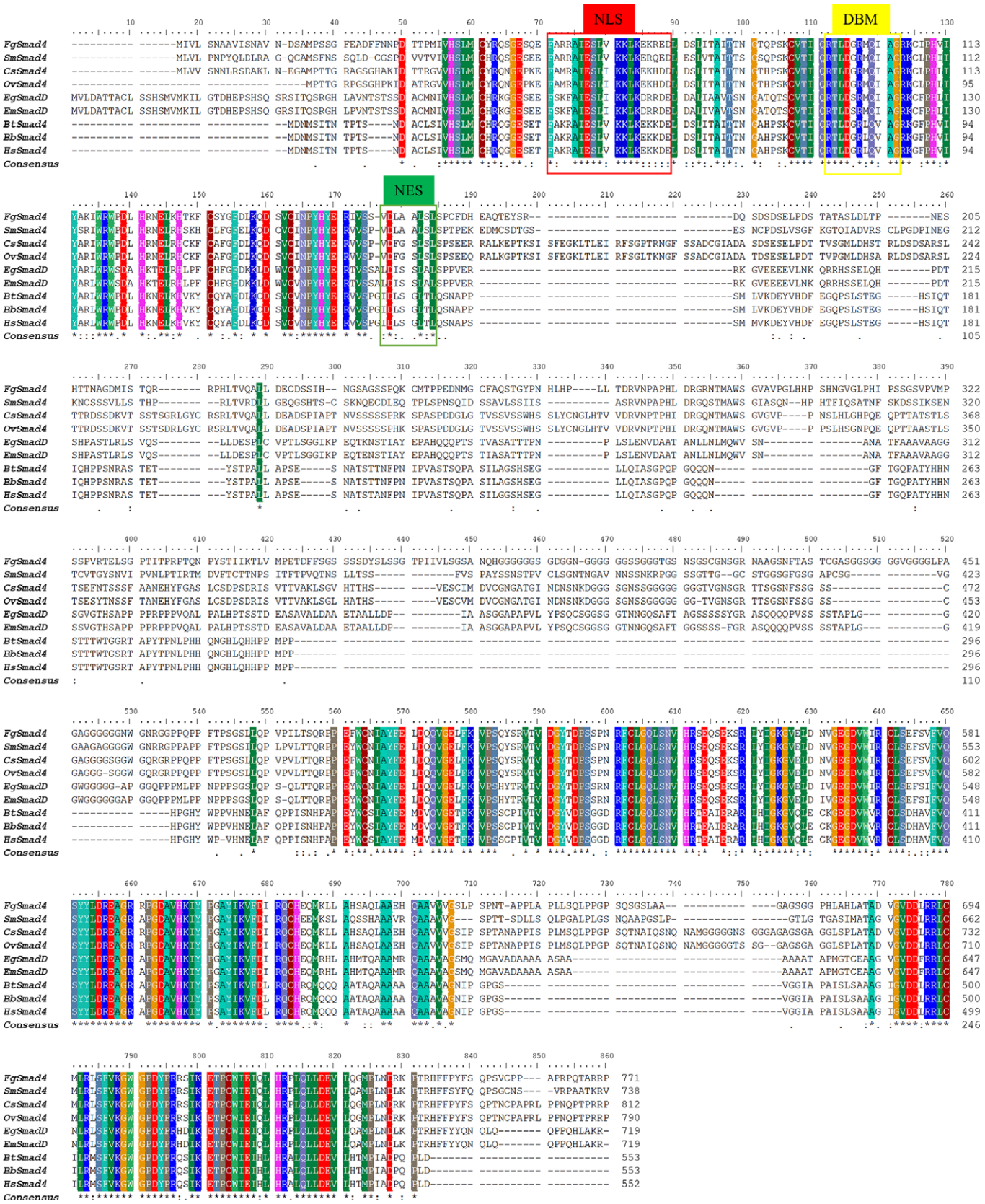


**Figure 4.** Multiple sequence alignment of FgSmad4 protein and other Co-Smad homologues. Conserved amino acid sequences and specific residues are highlighted in vertical columns. Asterisks (*) denote exact matches across all aligned sequences, colons (:) indicate strong similarity, dots (.) signify weak similarity, and the absence of a symbol indicates no similarity among amino acids. The boxed sequences indicate conserved motifs, including the nuclear localisation signal (NLS), the DNA binding motif (DBM), and the nuclear export signal (NES). The sequences included in the alignment are as follows: FgSmad4 (*Fasciola gigantica*), SmSmad4 (*Schistosoma mansoni*), CsSmad4 (*Clonorchis sinensis*), OvSmad4 (*Opisthorchis viverrini*), EgSmadD (*Echinococcus granulosus*), EmSmadD (*Echinococcus multilocularis*), BtSmad4 (*Bos taurus*), BbSmad4 (*Bubalus bubalis*), and HsSmad4 (*Homo sapiens*).

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

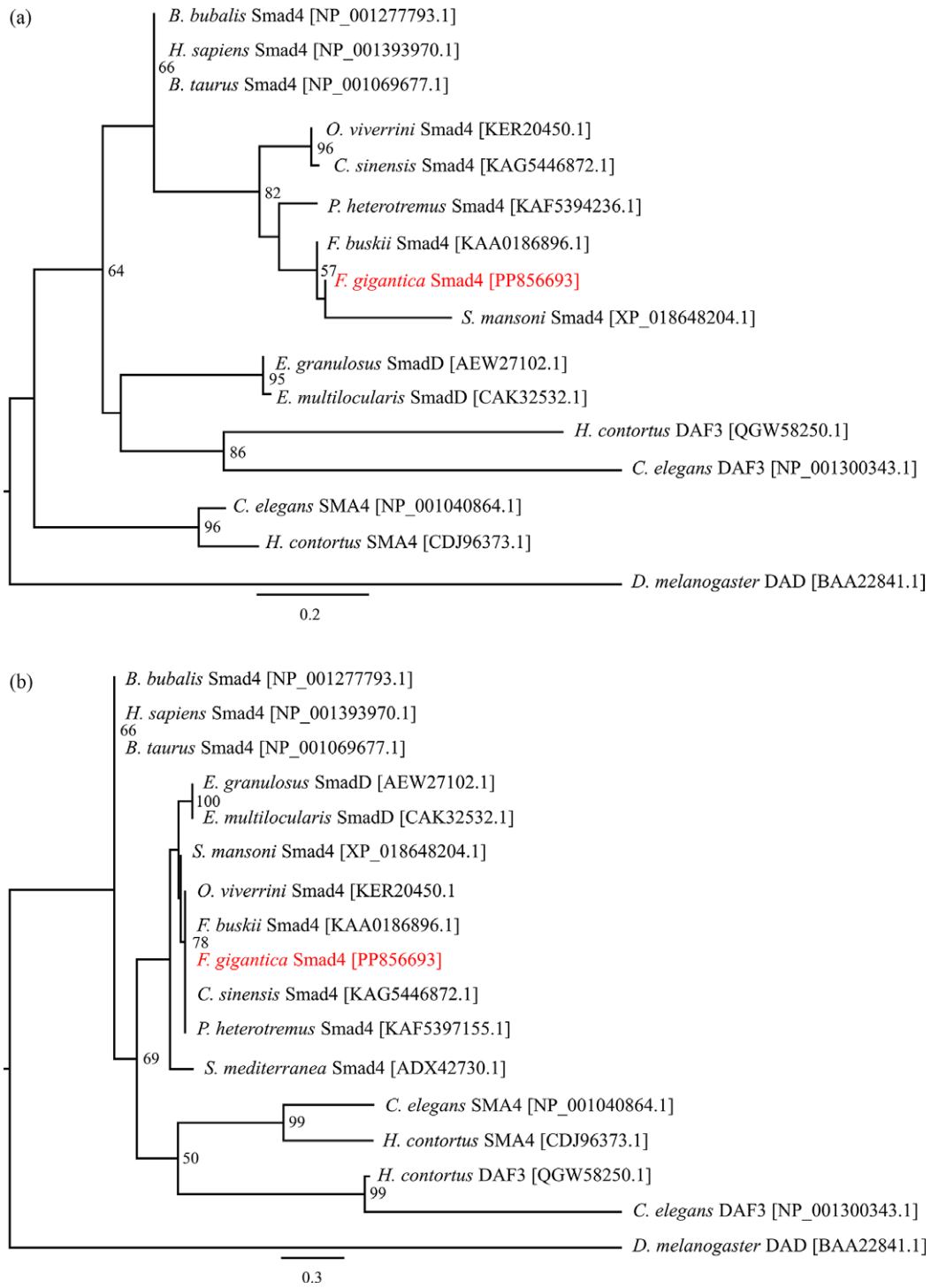


**Figure 5.** Phylogenetic relationships of FgSmad4 and Co-Smad proteins from related organisms based on the MH1 domain (A) and MH2 domain (B). The phylogenetic trees were inferred by maximum likelihood using 65 and 106 amino acid sequences of the MH1 and MH2 regions, respectively. The DAD sequence from *Drosophila melanogaster* [BAA22841.1] served as the outgroup for both MH1 and MH2 phylogenetic analyses. The FgSmad4 sequence identified in our study [PP856693] is highlighted in red bold font. Bootstrap values greater than 50% are displayed at the nodes of the phylogenetic trees (1000 replicates).

single transcript variant, given that both our 5' and 3' RACE results encoded identical protein isoform sizes. Additionally, the exon–intron structure derived from our sequence closely matches the pattern observed in FGIG_03163, which comprises nine exons and eight introns.

The length of the Co-Smad protein appears to be relatively conserved across animal species with the range of 400–500 amino acids (Makkar et al. 2009). In our study, FgSmad4 was approximately 218–219 amino acids longer than the mammalian Co-Smad proteins (553 amino acids in sheep, 552 amino acids in human). Notably, the Smad4 proteins in trematode and cestode exhibit a larger size compared to Smad4 in mammals and other reported animals. This observation aligned with Osman et al. (2004), who noted that the Smad4 homologue of *S. mansoni* (738 amino acids)

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

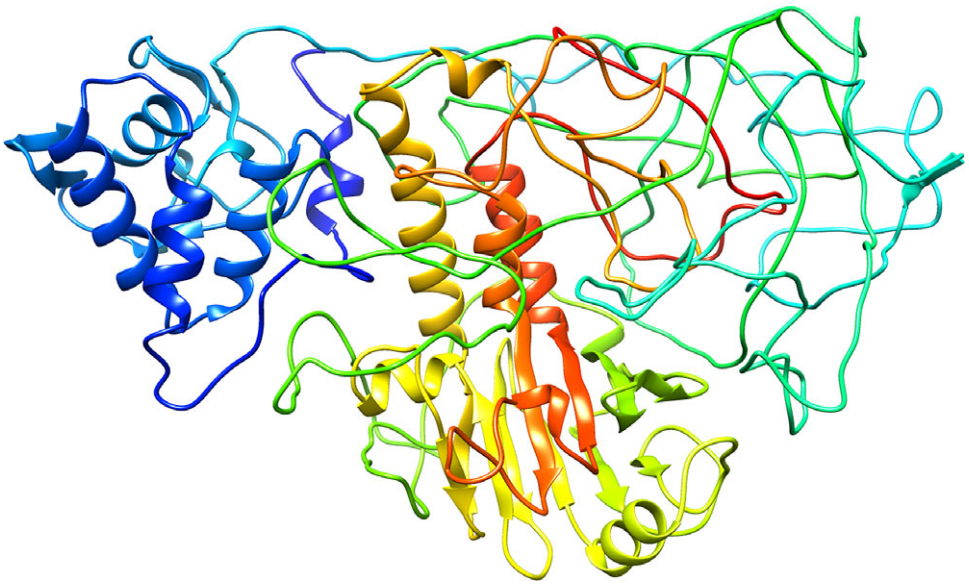

**Figure 6.** Modelled three-dimensional (3D) structure of FgSmad4. The 3D structure prediction was generated using Phyre2 homology modelling with reference to the template model of *Homo sapiens* Smad4 (d1dd1a). Chimera was utilised for visualizing the protein model. The colour scheme in the models indicates blue for the N-terminus and red for the C-terminus.

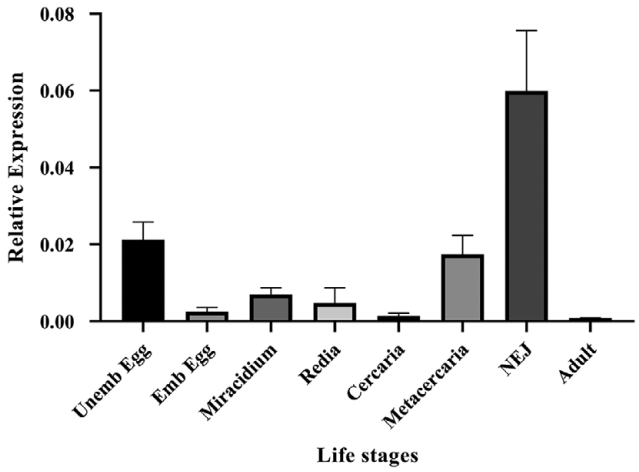


**Figure 7.** Quantitative expression analysis of *FgSmad4* across various life cycle stages of *F. gigantica*. The relative mRNA expression levels of *FgSmad4* throughout the parasite's life stages were determined via quantitative PCR, with *FgGAPDH* serving as the reference gene. The data represent the mean ± standard error of the mean from three independent biological replicates.

possessed an additional 200 amino acids compared to other known Co-Smad proteins. Likewise, Zhang et al. (2014) and Zavala-Góngora et al. (2008) documented that EgSmadD in *E. granulosus* and EmSmadD from *E. multilocularis* comprise 719 amino acids, providing additional evidence that supports the idea of an enlarged size in Co-Smad proteins within parasitic flatworms. The function or relevance of the increased length of these Co-Smad proteins remains unknown.

Classically, Smad proteins are characterised by two key conserved domains: the MH1 situated at the N-terminus, crucial for specific DNA binding, and the MH2 located at the C-terminus, essential for protein–protein interactions. These domains are linked by a proline-rich non-conserved intermediate region (Liu 2003; Liu et al. 1996). FgSmad4 exhibited similarity with Co-Smad proteins from various animal species. Within the MH1 domain, the highest degree of similarity was observed among trematode species, exceeding 80.00%, whereas the observed similarity to nematode and cestode species was comparatively lower at 50.00%–68.80% and 66.67%, respectively. Remarkably, the MH1 sequences of FgSmad4 and those of mammalian hosts exhibited a significant similarity of 80.00%. For the MH2 domain, FgSmad4 displayed significant conservation, aligning well with taxonomic classifications. FgSmad4 exhibited the greatest similarity within the platyhelminth group, particularly showing the most similarity to other trematodes, ranging from 83.73% to 96.77%. The observed similarity extended to 74.80% for cestodes, whereas relatively lower similarities were noted compared to nematodes (34.96%–43.90%) and mammalian hosts (65.45%). Our findings revealed that both MH1 and MH2 sequences share the greatest similarity with trematode species, with MH1 exhibiting a notable similarity to mammalian hosts. This observation was consistent with the findings in SmSmad4, where the MH1 domain showed a higher similarity (81%–87%) to the MH1 domain of Smad4 from mammals, whereas the MH2 domain displayed a lower level of similarity (Osman et al. 2004). These findings underscore the conserved nature of the MH1 domain, particularly among trematodes, and suggest a potential co-evolutionary relationship between host and parasite.

In life stage expression analyses, the presence of the *FgSmad4* transcript was observed in all examined stages of *F. gigantica*. This pattern aligns with the findings in *SmSmad4*, where the *SmSmad4* gene expression was documented throughout the entire life cycle of *S. mansoni* (Osman et al. 2004). Notable variation in the expression levels of *FgSmad4* was observed throughout the life cycle of *F. gigantica*, highlighting a significantly high transcript level in developing life cycle stages, including unembryonated egg, redia, and NEJ. In contrast, embryonated egg, cercaria, and adult exhibited relatively low levels of *FgSmad4* transcript, indicating a potentially diminished significance of this gene in these specific stages.

The transcript level of *FgSmad4* in unembryonated eggs was nearly 10 times higher than in embryonated eggs, underscoring its essential role in initiating developmental processes during early

https://doi.org/10.1017/S0022149X24000841 Published online by Cambridge University Press

embryogenesis in *F. gigantica*. In the NEJ stage, *FgSmad4* expression showed a dramatic increase, exceeding levels in cercaria by more than 60-fold and adult stages by more than 100-fold. This notable increase in expression, along with previously observed high *FgTLM* levels in NEJ (Japa et al. 2022), indicates that *FgSmad4* may be critical for preparing the juvenile parasite to adapt to the host environment, likely by promoting growth, driving essential developmental changes, and supporting survival.

We observed a striking 13-fold increase in *FgSmad4* expression in the metacercaria stage compared to cercaria, suggesting a critical role for FgSmad4 during this transitional phase. This elevated expression aligns with previously reported high levels of *FgTLM* in metacercaria (Japa et al. 2022). Although traditionally considered a dormant stage, metacercaria actively maintains essential biological processes to ensure survival and viability until reaching a mammalian host. Zhang et al. (2019) identified key regulatory functions within the transcriptional profile of *F. gigantica* metacercaria, including gene transcription, protein phosphorylation, and signal transduction, which collectively support critical processes such as metabolic regulation, nucleotide synthesis, pH balance, and endopeptidase activity. Taken together, our findings and Zhang's observations suggest that FgSmad4 likely acts as a key mediator in TGF-β signalling pathways, coordinating a range of biological and metabolic processes essential for the metacercaria's resilience and readiness for host transmission. Further studies are needed to elucidate the specific biological processes involving FgSmad4 at this stage.

In this study, real-time PCR analysis of *FgSmad4* expression across developmental stages revealed stage-specific patterns, indicating distinct regulatory functions at each phase. These functions likely include regulating growth, supporting developmental transitions, and enabling the parasite to adapt to varying environmental conditions. To validate and expand on these findings, further protein-level analyses, such as western blot quantification and localisation studies through immunohistochemistry or *in situ* hybridisation, could provide deeper insights into FgSmad4's role in mediating stage-specific developmental processes and adaptive mechanisms, clarifying its essential role in the biology of *F. gigantica*.

## Conclusion

This study provides a detailed characterisation of FgSmad4, the sole identified Co-Smad protein in *F. gigantica*. FgSmad4 exhibits the typical molecular features of the Co-Smad family, including the defining MH1 and MH2 domains. Expression analysis revealed significant variation in *FgSmad4* transcript levels across developmental stages, with the highest expression observed in the NEJ stage and the lowest in mature flukes. These findings underscore the conserved structure and potentially crucial role of FgSmad4 as a signalling mediator in *F. gigantica*. Given that Co-Smad participates in numerous biological pathways, further exploration of FgSmad4's functions could yield valuable insights into the TGF-β signalling pathway in this important parasite, enhancing our understanding of host-parasite interactions. This signal transduction pathway may represent a promising target for future drug development and vaccine initiatives.

**Acknowledgements.** The authors are grateful for the technical assistance provided by Ms. Wannapha Kanameung and Ms. Thanawadee Chaiphutorn.

**Author contribution.** O.J. and R.J.F. conceived and designed the study. O.J., C.P., and K.K. collected and maintained the parasites. O.J. performed the experiments, data collection, and bioinformatics analyses. O.J. and K.P. analysed and interpreted the data. O.J. and R.J.F. drafted and revised the manuscript. All authors read and approved the final manuscript.

**Financial support.** This research project was supported by the Thailand Science Research and Innovation Fund and the University of Phayao (Grant No. FF66-RIB070).

**Competing interest.** The authors declare that they have no competing interests.

## References

**Beesley NJ, Caminade C, Charlier J, Flynn RJ, Hodgkinson JE, Martinez-Moreno A, Martinez-Valladares M, Perez J, Rinaldi L and Williams DJL** (2018) *Fasciola* and fasciolosis in ruminants in Europe: identifying research needs. *Transboundary and Emerging Diseases* **65**(**S1**), 199–216. https://doi.org/10.1111/tbed.12682.

**Di W, Liu L, Zhang T, Li F, He L, Wang C, Ahmad AA, Hassan M, Fang R and Hu M** (2019) A DAF-3 co-Smad molecule functions in *Haemonchus contortus* development. *Parasites & Vectors* **12**(**1**), 609. https://doi.org/10.1186/s13071-019-3855-3.

**Ednilson Hilário L-J, Rafaella Pontes M, Claudio Romero B and Katia Cristina O** (2022) Perspective chapter: molecular crosstalk and signal transduction between platyhelminths and their hosts. In Jorge M-M, Victor Hugo Del R-A and Romel H-B (eds), *Parasitic Helminths and Zoonoses*. Rijeka: IntechOpen, Ch. 7.

**Fürst T, Keiser J and Utzinger J** (2012) Global burden of human food-borne trematodiasis: a systematic review and meta-analysis. *The Lancet Infectious Diseases* **12**(**3**), 210–221. https://doi.org/10.1016/S1473-3099(11)70294-8.

**Gouy M, Guindon S and Gascuel O** (2009) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**(**2**), 221–224. https://doi.org/10.1093/molbev/msp259.

**Hall T** (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic Acids Symposium Series*.

**Hall T, Biosciences I and Carlsbad CJGBB** (2011) BioEdit: an important software for molecular biology. *GERF Bulletin of Biosciences* **2**(**1**), 60–61.

**Huminiecki L, Goldovsky L, Freilich S, Moustakas A, Ouzounis C and Heldin CH** (2009) Emergence, development and diversification of the TGF-beta signalling pathway within the animal kingdom. *BMC Evolutionary Biology* **9**, 28. https://doi.org/10.1186/1471-2148-9-28.

**Japa O, Hodgkinson JE, Emes RD and Flynn RJ** (2015) TGF-β superfamily members from the helminth *Fasciola hepatica* show intrinsic effects on viability and development. *Veterinary Research* **46**, 29. https://doi.org/10.1186/s13567-015-0167-2.

**Japa O, Prakhammin K and Flynn RJ** (2022) Identification and expression of a transforming growth factor beta (TGF-β) homologue in the tropical liver fluke *Fasciola gigantica*. *Parasitology Research* **121**(**12**), 3547–3559. https://doi.org/10.1007/s00436-022-07679-1.

**Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJE** (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**(**6**), 845–858. https://doi.org/10.1038/nprot.2015.053.

**Konrad C, Kroner A, Spiliotis M, Zavala-Góngora R and Brehm K** (2003) Identification and molecular characterisation of a gene encoding a member of the insulin receptor family in *Echinococcus multilocularis*. *International Journal for Parasitology* **33**(**3**), 301–312. https://doi.org/10.1016/S0020-7519(02)00265-5.

**Lagna G, Hata A, Hemmati-Brivanlou A and Massagué J** (1996) Partnership between DPC4 and SMAD proteins in TGF-β signalling pathways. *Nature* **383**(**6603**), 832–836.

**Liu F** (2003) Receptor-regulated Smads in TGF-beta signaling. *Frontiers in Bioscience* **8**, s1280–1303. https://doi.org/10.2741/1149.

**Liu F, Hata A, Baker JC, Doody J, Cárcamo J, Harland RM and Massagué J** (1996) A human Mad protein acting as a BMP-regulated transcriptional activator. *Nature* **381**(**6583**), 620–623. https://doi.org/10.1038/381620a0.

**Logue CM, Barbieri NL and Nielsen DW** (2017) Chapter Eight - Pathogens of food animals: sources, characteristics, human risk, and methods of detection.

In Toldrá F (ed), *Advances in Food and Nutrition Research*. Academic Press, 277–365.

**Makkar P, Metpally RPR, Sangadala S and Reddy BVB** (2009) Modeling and analysis of MH1 domain of Smads and their interaction with promoter DNA sequence motif. *Journal of Molecular Graphics and Modelling* **27**(**7**), 803–812. https://doi.org/10.1016/j.jmgm.2008.12.003.

**Mas-Coma S, Bargues MD and Valero MA** (2005) Fascioliasis and other plant-borne trematode zoonoses. *International Journal for Parasitology* **35**(**11**–12), 1255–1278. https://doi.org/10.1016/j.ijpara.2005.07.010.

**Mas-Coma S, Valero MA and Bargues MD** (2009) Climate change effects on trematodiases, with emphasis on zoonotic fascioliasis and schistosomiasis. *Veterinary Parasitology* **163**(**4**), 264–280. https://doi.org/10.1016/j.vetpar.2009.03.024.

**Mas-Coma S, Valero MA and Bargues MD** (2019) Fascioliasis. *Advances in Experimental Medicine and Biology* **1154**, 71–103. https://doi.org/10.1007/978-3-030-18616-6_4.

**Massagué J** (2012) TGFβ signalling in context. *Nature Reviews Molecular Cell Biology* **13**(**10**), 616–630. https://doi.org/10.1038/nrm3434.

**Masuyama N, Hanafusa H, Kusakabe M, Shibuya H and Nishida E** (1999) Identification of two Smad4 proteins in *Xenopus*: their common and distinct properties. *Journal of Biological Chemistry* **274**(**17**), 12163–12170. https://doi.org/10.1074/jbc.274.17.12163.

**McVeigh P, McCammick EM, McCusker P, Morphew RM, Mousley A, Abidi A, Saifullah KM, Muthusamy R, Gopalakrishnan R, Spithill TW, Dalton JP, Brophy PM, Marks NJ and Maule AG** (2014) RNAi dynamics in juvenile *Fasciola* spp. liver flukes reveals the persistence of gene silencing in vitro. *PLOS Neglected Tropical Diseases* **8**(**9**), e3185. https://doi.org/10.1371/journal.pntd.0003185.

**Miyazono K** (2000) TGF-β signaling by Smad proteins. *Cytokine & Growth Factor Reviews* **11**(**1**), 15–22. https://doi.org/10.1016/S1359-6101(99)00025-8.

**Moustakas A and Heldin CH** (2009) The regulation of TGFbeta signal transduction. *Development* **136**(**22**), 3699–3714. https://doi.org/10.1242/dev.030338.

**Moustakas A, Souchelnytskyi S and Heldin CH** (2001) Smad regulation in TGF-beta signal transduction. *Journal of Cell Science* **114**(Pt **24**), 4359–4369. https://doi.org/10.1242/jcs.114.24.4359.

**Moxon JV, LaCourse EJ, Wright HA, Perally S, Prescott MC, Gillard JL, Barrett J, Hamilton JV and Brophy PM** (2010) Proteomic analysis of embryonic *Fasciola hepatica*: characterization and antigenic potential of a developmentally regulated heat shock protein. *Veterinary Parasitology* **169**(**1**), 62–75. https://doi.org/10.1016/j.vetpar.2009.12.031.

**Osman A, Niles EG and LoVerde PT** (2004) Expression of functional *Schistosoma mansoni* Smad4: role in ERK-mediated transforming growth factor β (tgf-β) down-regulation. *Journal of Biological Chemistry* **279**(**8**), 6474–6486. https://doi.org/10.1074/jbc.M310949200.

**Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE** (2004) UCSF chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**(**13**), 1605–1612. https://doi.org/10.1002/jcc.20084.

**Salzet M, Capron A and Stefano GB** (2000) Molecular crosstalk in host-parasite relationships: schistosome- and leech-host interactions. *Parasitology Today* **16**(**12**), 536–540. https://doi.org/10.1016/s0169-4758(00)01787-7.

**Samanta D and Datta PK** (2012) Alterations in the Smad pathway in human cancers. *Frontiers in Bioscience (Landmark edition)* **17**(**4**), 1281–1293. https://doi.org/10.2741/3986.

**Silver N, Best S, Jiang J and Thein SL** (2006) Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology* **7**(**1**), 33. https://doi.org/10.1186/1471-2199-7-33.

**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 for Parasitology* **36**(**10**), 1097–1112. https://doi.org/10.1016/j.ijpara.2006.05.008.

**Tzavlaki K and Moustakas A** (2020) TGF-β signaling. *Biomolecules* **10**(**3**). https://doi.org/10.3390/biom10030487.

**Wang G, Li C, Wang Y and Chen G** (2013) Cooperative assembly of Co-Smad4 MH1 with R-Smad1/3 MH1 on DNA: a molecular dynamics simulation study. *PloS One* **8**(**1**), e53841. https://doi.org/10.1371/journal.pone.0053841.

**Whelan S and Goldman N** (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18**(**5**), 691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851 %J Molecular Biology and Evolution.

**Wu D, Kong X, Zhang W and Di W** (2023) Reconstruction of the TGF-β signaling pathway of *Fasciola gigantica*. *Parasitology Research* **123**(**1**), 51. https://doi.org/10.1007/s00436-023-08064-2.

**You H, Gobert GN, Jones MK, Zhang W and McManus DP** (2011) Signalling pathways and the host-parasite relationship: putative targets for control interventions against schistosomiasis: signalling pathways and future anti-schistosome therapies. *Bioessays* **33**(**3**), 203–214. https://doi.org/10.1002/bies.201000077.

**Zavala-Góngora R, Derrer B, Gelmedin V, Knaus P and Brehm K** (2008) Molecular characterisation of a second structurally unusual AR-Smad without an MH1 domain and a Smad4 orthologue from *Echinococcus multilocularis*. *International Journal for Parasitology* **38**(**2**), 161–176. https://doi.org/10.1016/j.ijpara.2007.07.008.

**Zavala-Góngora R, Kroner A, Bernthaler P, Knaus P and Brehm K** (2006) A member of the transforming growth factor-beta receptor family from *Echinococcus multilocularis* is activated by human bone morphogenetic protein 2. *Molecular and Biochemical Parasitology* **146**(**2**), 265–271. https://doi.org/10.1016/j.molbiopara.2005.12.011.

**Zhang C, Wang L, Wang H, Pu H, Yang L, Li J, Wang J, Lü G, Lu X, Zhang W, Vuitton DA, Wen H and Lin R** (2014) Identification and characterization of functional Smad8 and Smad4 homologues from *Echinococcus granulosus*. *Parasitology Research* **113**(**10**), 3745–3757. https://doi.org/10.1007/s00436-014-4040-4.

**Zhang X-X, Cwiklinski K, Hu R-S, Zheng W-B, Sheng Z-A, Zhang F-K, Elsheikha HM, Dalton JP and Zhu X-Q** (2019) Complex and dynamic transcriptional changes allow the helminth *Fasciola gigantica* to adjust to its intermediate snail and definitive mammalian hosts. *BMC Genomics* **20**(**1**), 729. https://doi.org/10.1186/s12864-019-6103-5.

**Zhang Y, Musci T and Derynck R** (1997) The tumor suppressor Smad4/DPC 4 as a central mediator of Smad function. *Current Biology* **7**(**4**), 270–276. https://doi.org/10.1016/S0960-9822(06)00123-0.