Published in final edited form as:

Exp Parasitol. 2008 October ; 120(2): 208–213. doi:10.1016/j.exppara.2008.07.005.

***Schistosoma mansoni*: Identification of SmNR4A, a member of nuclear receptor subfamily 4**

Wenjie Wu and Philip T. LoVerde[§]

Departments of Biochemistry and Pathology, University of Texas Health Sciences Center, San Antonio, Texas, 78229-3800, USA

Abstract

A cDNA encoding a member of nuclear receptor subfamily 4 (SmNR4A) was isolated from the trematode *Schistosoma mansoni*. The open reading frame (ORF) of SmNR4A cDNA is 2481 base pairs long encoding an 827 amino acid protein. Alignment of the deduced protein sequence showed the DNA binding domain (DBD) of SmNR4A is highly conserved. Like human and *Drosophila* members in NR subfamily 4, SmNR4A possess an atypical ligand binding domain (LBD), the conserved lysine in helix H3 is replaced by a glutamic acid, and three of the four phenylalanines which fill the entire surface of the ligand binding pocket (LBP) are conserved in SmNR4A. A phylogenetic tree of SmNR4A was constructed using the conserved protein sequence of the DBD, the C-terminal-extension of DBD (CTE) and the LBD. The results show that the SmNR4A is a member of NR subfamily 4 from *S. mansoni*. The *SmNR4A* gene contains six exons spanning more than 50 kbp. The relative mRNA expression levels of *SmNR4A* were evaluated in fourteen different developmental stages by quantitative real-time reverse-transcriptase polymerase chain reaction (qPCR). The results demonstrated that *SmNR4A* expression was regulated throughout development. It was highly expressed in daughter sporocysts and 35-day worms, but barely expressed in cercariae and 1-hour and 3-day schistosomules.

Index Descriptors and Abbreviations

Nuclear receptor (NR); Platyhelminthes; *Schistosoma mansoni*; SmNR4A: *S. mansoni* nuclear receptor NR4A; cDNA cloning; gene organization; Quantitative real-time reverse-transcriptase polymerase chain reaction (q-RT-PCR); Phylogenetic analysis

1. Introduction

The nuclear receptor subfamily 4 (NR4) is a small subfamily that only contains one group: NR4A. NR4A group contains four receptors including three vertebrate receptors (NR4A1 (NGFI-B), NR4A2 (Nurr1) and NR4A3 (NOR1)) and one invertebrate member (NR4A4, *Drosophila* HR38 (DHR38)) (Chang et al., 1989; Fisk and Thummel, 1995; Mages et al., 1994; Nuclear Receptor Nomenclature Committee, 1999; Scearce et al., 1993; Sutherland et al., 1995). Like all of the other NRs, NR4A members share a similar protein modular organization: a highly divergent N-terminal A/B domain, a conserved DNA binding domain (DBD), a hinge region and a ligand binding domain (LBD). The conserved DBD region

[§]Corresponding author. Tel: 210-567-3737; Fax: 210-670-3322; Email: loverde@uthscsa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

contains two zinc finger motifs which bind to the DNA in the promoter region of target genes and the LBD which is involved in transcriptional activation of target genes via coactivator and/or ligand binding. The ligand of many NRs have yet to be identified, these NRs with an unknown ligand are called 'orphan' receptors. Among these orphan receptors, the NR4 members which cannot bind a ligand are called true orphans. Crystal structure studies show that the human NGFI-B, Nurr1 and *Drosophila* HR38 (DHR38) lost the ligand binding pocket (LBP) (Baker et al., 2003; Codina et al., 2004; Flaig et al., 2005) as a result of the side chains of four phenylalanines that point into the interior and fill the entire surface of what would be the ligand binding pocket (Baker et al., 2003). NR4 members are classified as early response genes as they are induced by a range of signals such as growth factors, physical stimuli and fatty acids. For example, DHR38 in *Drosophila* is known to mediate the ecdysteroid pathway (Baker et al., 2003; Maxwell and Muscat, 2006).

We previously isolated a short fragment of a cDNA encoding the DBD of a NR4 member from *Schistosoma mansoni* (SmNR4A) (Wu et al., 2006). In this study, a cDNA encoding the entire open reading (ORF) of SmNR4A was isolated. Its gene organization and its mRNA expression throughout the life cycle were determined.

2. Materials and Methods

2.1. cDNA isolation

The cDNA of *S. mansoni* NR4A (*SmNR4A*) was isolated by a PCR strategy using a *S. mansoni* female worm λ ZAPII cDNA library pool as template (Wu et al., 2006; Wu et al., 2008). The specific primers were to a fragment encoding the DBD (Wu et al., 2006). The primer for the 5' end is 5'-ATAATGCAGCATGTCAAC-3' and for the 3' end is 5'-GATTTCAAAAATGTTTAAAAGT-3'. The vector universal primers were M13-Reverse primer: 5'-CAGGAACAGCTATGAC-3' or T3 promoter primer: 5'-AATTAACCCTCACTAAAGGG-3'), and M13-Forward (-20): 5'-GTAAAACGACGGCCAG-3' or T7 promoter primer: 5'-TAATACGACTCACTATAGGG-3'. The PCR products were ligated into pCR2.1 TOPO vector (Invitrogen) after separation on and purification from 1.2% agarose gels and then sequenced. The cDNA was shown to be a single mRNA species by alignment of the cDNA sequence with the genomic sequence which was obtained by blastn against the *S. mansoni* genomic sequence database (<ftp://ftp.sanger.ac.uk/pub/databases/Trematode/S.mansoni/genome>) and also by PCR on a cDNA template using primers located within the 5'UTR and 3'UTR.

2.2. Gene organization

The exon/intron boundaries and the sizes of SmNR4A gene were determined by alignment of the cDNA sequence with the genomic DNA sequence. The sizes of the gaps among DNA contigs were identified by PCR on genomic DNA extracted from cercariae. The exons were translated to determine the coding region for the different protein domains.

2.3. Phylogenetic analysis

A phylogenetic tree was constructed using the deduced protein sequence of the DNA binding domain (DBD), the C-terminal-extension of the DBD (CTE, A-box and T-box) and the ligand binding domain (LBD). The sequences were aligned using ClustalW (<http://www.cf.ac.uk/biosi/research/biosoft/Downloads/clustalw.html>) and the data set was tested by both the Bayesian inference and the Maximum Likelihood (ML) method. The Bayesian inference was carried out under a mixed amino acid replacement model + invgamma rates (Huelsenbeck and Ronquist, 2001) using MrBayes v3.1.1. The trees were run for 3 million generations starting randomly and sampled every 100 generations with four simultaneous

Markov chains. The Bayesian posterior probabilities (PPs) were calculated, with a burn-in value setting at 7,500, using a Markov chain Monte Carlo (MCMC) sampling approach implemented in MrBayes v3.1.1. The maximum likelihood (ML) analysis was carried out using PHYML (v2.4.4) under Jones-Taylor-Thornton (JTT) substitution model (Jones et al., 1992) with a gamma distribution of rates between sites (eight categories, parameter alpha estimated by the program) (Guindon and Gascuel, 2003), 100 replicates were bootstrapped to support the phylogenetic analysis.

2.4. Quantitative real-time RT-PCR

The mRNA expression of *SmNR4A* was tested in 12 different developmental stages including eggs, snail hepatopancreas containing daughter sporocysts, cercariae, 3-hour and 3-day schistosomules, 7, 15, 21, 28 and 35-day worms, 45-day adult female and male worms. Total RNA was extracted using TRIzol reagent (Invitrogen) treated with RNase-free DNaseI (RQ1 DNase; Promega) and reverse transcribed with SuperScript Reverse Transcriptase II and random hexamer (Invitrogen). The primers specific to *SmNR4A* were forward: 5'-CCCCCAAATCTTCCTTA -3' (nt 267 to 284) and reverse: 5'-GTTGTTGTTGTTGTTGTTGACTTATC -3' (nt 327 to 352). *S. mansoni* α -tubulin (GenBank accession no. **M80214**) was used as the reference gene. For the specific primers for *S. mansoni* α -tubulin refer to (Mei and LoVerde, 1997; Oger et al., 2006; Wu et al., 2007b). The PCR amplification was carried out using the reverse-transcribed cDNA as template with the SYBR Green Master Mix® (Invitrogen) using the BIORAD IQ™5 Real-Time PCR Detection System. The normalized gene expression ($\Delta\Delta CT$) (Livak and Schmittgen, 2001) of *SmNR4A* was standardized to the relative quantities of *S. mansoni* α -tubulin with the normalized expression calculations implemented in iQ5 using BioRad IQ™5 Optical System software V1.1. For graphical representation of the expression, the normalized expression was recalculated by dividing the expression level of each tested stage by the highest expression level (35-day worms) with iQ5.

3. Results

3.1. cDNA isolation and sequence analysis

A consensus 2725 bp cDNA of *SmNR4A* was isolated. Alignment of the cDNA and genomic DNA (Contig_0009881+Contig_0009882) shows that there is a 2169 bp genomic sequence upstream of the putative start codon of *SmNR4A* without a stop codon, this region contains several ATGs in the same open reading frame with *SmNR4A* cDNA. The closest ATG was found at -132 to -134 nucleotides upstream of the putative start codon. Nested PCR on a cDNA template was performed to determine whether the isolated cDNA encoded the entire open reading frame. Different 5' primers were designed according to the genomic DNA sequence (upstream of the start codon) and different 3' primers were designed using the cDNA sequence 3' of the first exon. No PCR products were obtained by using any of the primers designed representing the -132 region or upstream of this region. Blast analysis against the NCBI databases revealed no hits using the deduced protein sequence of this upstream genomic DNA fragment as a query. These results demonstrate that the cDNA isolated encodes a complete open reading frame.

The entire open reading frame (ORF) of *SmNR4A* is 2481 bp encoding an 827 amino acid protein. The DBD of *SmNR4A* is 66 amino acids which is identical in size to the human and *Drosophila* NR4 members (Fig. 1). The DBD of *SmNR4A* shows 89.4%, 86.4%, 84.8 and 83.3% identity to that of human NURR1 (hNURR1), sea urchin NR4A (SpNR4A), *Drosophila* HR38 (DHR38) and 83.3% honey bee HR38 (AmHR38), respectively. The P-box (CEGCKG) followed by the amino acid FFKR in the DBD is identical to that of human and *Drosophila* NR4. The C-terminal-extension of the DBD (CTE, A-box and T-box, Fig. 1) in

SmNR4A is also highly conserved, the A-box exhibits 91.6% (11 of 12 amino acids) identity to that of hNURR1, DHR38 and AmNH38, and 75% identity to that of SpNR4, respectively. The T-box shows 85.7% (6 of 7 amino acids) identity to that of hNURR1, DHR38 and AmHR38, and 71.4% (5 of 7 amino acids) identity to that of SpNR4A, respectively (Fig. 1).

Sequence alignment showed that like all of the other isolated *S. mansoni* NRs (de Mendonca et al., 2002; de Mendonca et al., 2000; Freebern et al., 1999a; Freebern et al., 1999b; Hu et al., 2006a; Hu et al., 2006b; Lu et al., 2006; Wu et al., 2006; Wu et al., 2007a; Wu et al., 2007b; Wu et al., 2007c; Wu et al., 2008), helices 1 and 2 of the LBD in SmNR4A are highly diverged. The length of the LBD in SmNR4A is longer than that of human and *Drosophila* NR4A proteins as several fragments have been inserted into helices 1 and 11 and inserted between helices 5 and 6 and helices 9 and 10. In most other NRs, a conserved lysine is found in helix H3. However in members of the NR4A subfamily, the lysine is replaced by glutamic acid. This is also the case for SmNR4A (Fig. 1, **H3 arrowhead**). In human and drosophila NR4A members, four conserved phenylalanines were found in the LBD which filled the entire surface of the ligand binding pocket (LBP) (Baker et al., 2003; Codina et al., 2004; Flaig et al., 2005). Three of them were retained in SmNR4A. This result suggests that the NR4A members in trematodes, arthropods and vertebrates originated from a common ancestor NR whose ligand binding pocket function was lost due to the presence of the phenylalanines. Sequence alignment also showed that helix 12 of SmNR4A was diverged like that of other NR4 proteins.

3.2. Gene organization

Four genomic contigs (Contig_0009881+Contig_0009882, Contig_0009885, Contig_0009889 and Contig_0052937) in scaffold Smp_scaff000069, Smp_scaff001990 and Smp_scaff017649 covering the five prime untranslated region (5'-UTR), the ORF and the three prime untranslated region (3'-UTR) of *SmNR4A* were obtained by blastn against the WTSI genomic sequence database (<ftp://ftp.sanger.ac.uk/pub/databases/Trematode/S.mansoni/genome>) with the *SmNR4A* cDNA as a query (Fig. 2A). The sizes of each gap within the scaffold Smp_scaff000069 and among the genomic contigs were determined by PCR. SmNR4A gene spans more than 50 kbp and contains six exons interrupted by five introns (Fig. 2A). The sizes of the six exons are 839, 130, 434, 170, 524 and 628 bp, respectively (Fig. 2B) interrupted by five introns with the sizes of 17.3, 5, 14.5, 5.9 and 7.4 kb, respectively (Fig. 2A). Translation of the exons shows that the AB domain is encoded by one exon, DBD and the hinge region are encoded by two exons, and the LBD is encoded by three exons, respectively (Fig. 2B).

3.3. Phylogenetic tree construction

A phylogenetic tree was constructed using the protein sequence of the DBD, CTE and LBD (Fig. 3). The results of both the Bayesian inference and the maximum likelihood (ML) analysis show that SmNR4A clusters within the NR4A group. Furthermore, all the vertebrate NR4 members clustered in a group containing three subgroups (NR4A1, NR4A2 and NR4A3). The Deuterostome invertebrate NR4A (sea urchin NR4A) and Protostome NR4A (insect and *Schistosoma* NR4A) members clustered outside of the vertebrate NR group. The result indicates that the SmNR4A is the orthologue of human NR4A and DHR38 proteins.

3.4. mRNA expression

Quantitative real-time RT-PCR (Q-PCR) was performed to evaluate relative mRNA levels of *SmNR4A* in 12 different developmental stages including eggs, daughter sporocysts, cercariae, 1-hour and 3-day schistosomules, 7, 15, 21, 28 and 35-day worms, 45-day adult female worms and male worms (Fig. 4). The results showed that *SmNR4A* is regulated throughout development with different expression levels. It is highly expressed in the endoparasitic stages in both the snail and mammalian hosts, but expressed at the lowest levels in cercariae and 1

hour and 3 day schistosomules. The highest expression was found in the daughter sporocysts and 35-day old worms.

4. Discussion

The nuclear receptors in subfamily 4 are known to be true orphan receptors. They possess an atypical LBD in which the ligand binding pocket is filled by phenylalanines making them unable to bind a ligand (Baker et al., 2003; Codina et al., 2004; Flaig et al., 2005). Protein sequence alignment of SmNR4A with human NGFI-B, Nurrl and NOR1 and *Drosophila* DHR38 showed that three of the four phenylalanines, which fill the surface of the ligand binding pocket, are retained in SmNR4. This result suggests that SmNR4A is an orphan receptor and the LBP of NR4 was lost in a common ancestral NR4 gene before the split of the platyhelminthes, arthropods and vertebrates. Unlike the nuclear receptors in other subfamilies, helix 12 of the LBD in vertebrate and *Drosophila* NR4 proteins is not conserved. In other nuclear receptors, helix 12 in the LBD is highly conserved as it contains an AF2 activating domain core (AF2-AD) with the modular $\Phi\Phi\text{XE}\Phi\Phi$, where Φ denotes a hydrophobic residue. In the NR4A members, the glutamic acid was replaced with other amino acids. Interestingly, this glutamic acid was retained in SmNR4A (Fig. 1). Whether this glutamic acid which is conserved in the AF2-AD of other NRs, is retained in other invertebrate NR4 needs to be evaluated when more data is available. Phylogenetic analysis showed that the Protostome NR4A and Deuterostome invertebrate NR4A clustered outside of the vertebrate NR4A group (NR4A1, NR4A2 and NR4A3), this result suggested that the vertebrate NR4A underwent duplication after the split of the invertebrates and vertebrates.

Although the schistosome genes regulated by SmNR4A are not known, members of the NR4A subfamily in other organisms are known to regulate various genes. For example, in humans, NGFI-b and NURR1 are known to be key regulators of hCYP11B2 expression, and NGFI-b regulates the transcription of 3 β -hydroxysteroid dehydrogenase. Human NOR-1 was shown to enhance the transcription of 3',5'-cyclic adenosine 5'-monophosphate-dependent uncoupling protein-1 gene. In *Drosophila*, DHR38 mediates an atypical ecdysteroid signaling pathway and influences the transcription of the DOPA decarboxylase gene (Baker et al., 2003; Bassett et al., 2004a; Bassett et al., 2004b; Davis et al., 2007; Kumar et al., 2008). The actual role of SmNR4 in *S. mansoni* biology is yet to be determined. However, Quantitative real-time RT-PCR demonstrates that *SmNR4A* is developmentally regulated. Interestingly, the highest level is in 35 day worms which coincides with the beginning of egg deposition (LoVerde et al., 2004) which may suggest a biological function.

Nucleotide sequence accession number

The mRNA sequences reported here were deposited in GenBank under the accession number **AY395064**.

Acknowledgements

This research was supported by NIH grants AI046762 and FIC D43TW6580. We thank the WTSI and TIGR genome sequencing centers for making their data publicly available.

References

- Baker KD, Shewchuk LM, Kozlova T, Makishima M, Hassell A, Wisely B, Caravella JA, Lambert MH, Reinking JL, Krause H, Thummel CS, Willson TM, Mangelsdorf DJ. The *Drosophila* orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. *Cell* 2003;113(6):731–742. [PubMed: 12809604]

- Bassett MH, Suzuki T, Sasano H, De Vries CJ, Jimenez PT, Carr BR, Rainey WE. The orphan nuclear receptor NGFIB regulates transcription of 3 β hydroxysteroid dehydrogenase. implications for the control of adrenal functional zonation. *The Journal of Biological Chemistry* 2004a;279(36):37622–37630. [PubMed: 15208301]
- Bassett MH, Suzuki T, Sasano H, White PC, Rainey WE. The orphan nuclear receptors NURR1 and NGFIB regulate adrenal aldosterone production. *Molecular Endocrinology* (Baltimore, Md) 2004b; 18(2):279–290.
- Chang C, Kokontis J, Liao SS, Chang Y. Isolation and characterization of human TR3 receptor: a member of steroid receptor superfamily. *Journal of Steroid Biochemistry* 1989;34(1–6):391–395. [PubMed: 2626032]
- Codina A, Benoit G, Gooch JT, Neuhaus D, Perlmann T, Schwabe JW. Identification of a novel co-regulator interaction surface on the ligand binding domain of Nurr1 using NMR footprinting. *The Journal of Biological Chemistry* 2004;279(51):53338–53345. [PubMed: 15456745]
- Davis MM, Yang P, Chen L, O'Keefe SL, Hodgetts RB. The orphan nuclear receptor DHR38 influences transcription of the DOPA decarboxylase gene in epidermal and neural tissues of *Drosophila melanogaster*. *Genome* 2007;50(11):1049–1060. [PubMed: 18059550]
- de Mendonca RL, Bouton D, Bertin B, Escriva H, Noel C, Vanacker JM, Cornette J, Laudet V, Pierce RJ. A functionally conserved member of the FTZ-F1 nuclear receptor family from *Schistosoma mansoni*. *European Journal of Biochemistry* 2002;269(22):5700–5711. [PubMed: 12423370]
- de Mendonca RL, Escriva H, Bouton D, Zelus D, Vanacker JM, Bonnelye E, Cornette J, Pierce RJ, Laudet V. Structural and functional divergence of a nuclear receptor of the RXR family from the trematode parasite *Schistosoma mansoni*. *European Journal of Biochemistry* 2000;267(11):3208–3219. [PubMed: 10824105]
- Fisk GJ, Thummel CS. Isolation, regulation, and DNA-binding properties of three *Drosophila* nuclear hormone receptor superfamily members. *Proceedings of the National Academy of Sciences of the United States of America* 1995;92(23):10604–10608. [PubMed: 7479849]
- Flaig R, Greschik H, Peluso-Iltis C, Moras D. Structural basis for the cell-specific activities of the NGFI-B and the Nurr1 ligand-binding domain. *The Journal of Biological Chemistry* 2005;280(19):19250–19258. [PubMed: 15716272]
- Freebern WJ, Niles EG, LoVerde PT. RXR-2, a member of the retinoid x receptor family in *Schistosoma mansoni*. *Gene* 1999a;233(1–2):33–38. [PubMed: 10375618]
- Freebern WJ, Osman A, Niles EG, Christen L, LoVerde PT. Identification of a cDNA encoding a retinoid X receptor homologue from *Schistosoma mansoni*. Evidence for a role in female-specific gene expression. *The Journal of Biological Chemistry* 1999b;274(8):4577–4585. [PubMed: 9988692]
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 2003;52(5):696–704. [PubMed: 14530136]
- Hu R, Wu W, Niles EG, LoVerde PT. Isolation and characterization of *Schistosoma mansoni* constitutive androstane receptor. *Molecular and Biochemical Parasitology* 2006a;148(1):31–43. [PubMed: 16603257]
- Hu R, Wu W, Niles EG, Loverde PT. SmTR2/4, a *Schistosoma mansoni* homologue of TR2/TR4 orphan nuclear receptor. *International Journal for Parasitology* 2006b;36:1113–1122. [PubMed: 16839558]
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001;17(8):754–755. [PubMed: 11524383]
- Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 1992;8(3):275–282. [PubMed: 1633570]
- Kumar N, Liu D, Wang H, Robidoux J, Collins S. Orphan nuclear receptor NOR-1 enhances 3',5'-cyclic adenosine 5'-monophosphate-dependent uncoupling protein-1 gene transcription. *Molecular Endocrinology* (Baltimore, Md) 2008;22(5):1057–1064.
- Livak K, Schmittgen T. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C_T$) Method. *Methods* 2001;25(4):402–408. [PubMed: 11846609]
- LoVerde PT, Niles EG, Osman A, Wu WJ. *Schistosoma mansoni* male-female interactions. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* 2004;82(2):357–374.

- Lu C, Wu W, Niles EG, Loverde PT. Identification and characterization of a novel fushi tarazu factor 1 (FTZ-F1) nuclear receptor in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology* 2006;150(1):25–36. [PubMed: 16870276]
- Mages HW, Rilke O, Bravo R, Senger G, Kroczeck RA. NOT, a human immediate-early response gene closely related to the steroid/thyroid hormone receptor NAK1/TR3. *Molecular endocrinology* (Baltimore, Md) 1994;8(11):1583–1591.
- Maxwell MA, Muscat GE. The NR4A subgroup: immediate early response genes with pleiotropic physiological roles. *Nuclear Receptor Signaling* 2006;4:e002. [PubMed: 16604165]
- Mei H, LoVerde PT. *Schistosoma mansoni*: the developmental regulation and immunolocalization of antioxidant enzymes. *Experimental Parasitology* 1997;86(1):69–78. [PubMed: 9149242]
- Nuclear Receptor Nomenclature Committee. A unified nomenclature system for the nuclear receptor superfamily. *Cell* 1999;97:161–163. [PubMed: 10219237]
- Oger F, Bertin B, Caby S, Dalia-Cornette J, Adams M, Vicogne J, Capron M, Pierce RJ. Molecular cloning and characterization of *Schistosoma mansoni* Ftz-F1 interacting protein-1 (SmFIP-1), a novel corepressor of the nuclear receptor SmFtz-F1. *Molecular and Biochemical Parasitology* 2006;148(1):10–23. [PubMed: 16574255]
- Scearce LM, Laz TM, Hazel TG, Lau LF, Taub R. RNR-1, a nuclear receptor in the NGFI-B/Nur77 family that is rapidly induced in regenerating liver. *The Journal of Biological Chemistry* 1993;268(12):8855–8861. [PubMed: 8473329]
- Sutherland JD, Kozlova T, Tzertzinis G, Kafatos FC. *Drosophila* hormone receptor 38: a second partner for *Drosophila* USP suggests an unexpected role for nuclear receptors of the nerve growth factor-induced protein B type. *Proceedings of the National Academy of Sciences of the United States of America* 1995;92(17):7966–7970. [PubMed: 7644522]
- Wu W, Niles EG, El-Sayed N, Berriman M, LoVerde PT. *Schistosoma mansoni* (Platyhelminthes, Trematoda) nuclear receptors: Sixteen new members and a novel subfamily. *Gene* 2006;366(2):303–315. [PubMed: 16406405]
- Wu W, Niles EG, Hirai H, LoVerde PT. Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. *BMC Evolutionary Biology* 2007a;7:27. [PubMed: 17319953]
- Wu W, Niles EG, Hirai H, LoVerde PT. Identification and characterization of a nuclear receptor subfamily I member in the Platyhelminth *Schistosoma mansoni* (SmNR1). *The FEBS Journal* 2007b;274(2):390–405. [PubMed: 17173548]
- Wu W, Niles EG, LoVerde PT. Thyroid hormone receptor orthologues from invertebrate species with emphasis on *Schistosoma mansoni*. *BMC Evolutionary Biology* 2007c;7:150. [PubMed: 17727708]
- Wu W, Tak EY, Loverde PT. *Schistosoma mansoni*: SmE78, a nuclear receptor orthologue of *Drosophila* ecdysone-induced protein 78. *Experimental Parasitology* 2008;119(2):313–318. [PubMed: 18430421]

Fig. 1. Sequence alignment of DNA binding domain, hinge and ligand bind domain

Stars indicate the conserved cysteine residues that comprise the two zinc fingers of the DBD. The P-Box, D-Box of DBD, A-Box and T-Box in C-terminal-extension (CTE) and signature sequence of LBD (Tτ) are identified under their respective sequences. H1, H3-H12 indicates the helices of LBD according to that of the human NGFI-B (Codina et al., 2004; Flaig et al., 2005). Helix 2 in NR4 members has not been clearly identified and is not indicated in the figure. The conserved residues are highlighted with grey. The number at the end of each line indicates the residue position in the original sequence. The four conserved phenylalanines which fill the surface of the ligand binding pocket are highlighted with black. The glutamic acid in helix 12 which was replaced by another residue in human and *Drosophila* NR4A members but retained

in SmNR4A, is indicated with an arrow. The arrow head denotes the glutamic acid that replaces the conserved lysine in members of other NR subfamilies. Sequences used for the alignment are: *Drosophila* HR38 (DHR38): [CAA61534](#); Bee NR38 (*Am*NR38): [XP_623628](#); Human Nurr1 (hNURR1): [CAA53518](#) and Sea Urchin NR4A (SpNR4A): [XP_786266](#).

Fig. 2. Gene structure of SmNR4A

A) Identifies the exons and the size of introns; roman numerals indicate exons. The three scaffolds (scaffold Smp_scaff000069, and overlapped Smp_scaff001990 and Smp_scaff017649) and four genomic contigs obtained in WTSI genomic sequence database (Contig_0009881+Contig_0009882, Contig_0009885, Contig_0009889, Contig_0052937) are shown in the figure, the size of the gaps among the contigs was determined by PCR. **B)** Shows size of exons and their correspondence to the different NR domains. A/B: A/B domain; DBD: DNA binding domain; LBD: ligand binding domain, Tr: signature of LBD.

NR38): **AF165528**; **Mosquito (*Anopheles*) NR38** (*Anopheles gambiae* NR438); **XP 317123**; **Rat NR4A1** (*Rattus norvegicus* (Norway rat) NR4A1): **AAA42058**; **Rat NR4A3** (*R. norvegicus* NR4A3): **NP 113816** and **Sea urchin NR4A** (*Strongylocentrotus purpuratus* (purple sea urchin) NR4A): **XP 786266**.

Fig. 4. Quantitative real-time RT-PCR shows mRNA expression of *SmNR4A*

Normalized gene expression ($\Delta\Delta CT$) of *SmNR4A* was standardized to the relative quantities of *S. mansoni* α -tubulin. For graphical representation of expression, the normalized expression was recalculated by dividing the expression level for each stage by the lowest expression level (in 1-hour schistosomule). uninfect: uninfected *Biomphalaria glabrata* hepatopancreas served as a negative control, inf: infected *Biomphalaria glabrata* hepatopancreas containing daughter sporocysts, Cer: cercariae, 1h: 1-hour schistosomules, 3d: 3-day schistosomules, 7d: 7-day worms, 15d: 15-day worms, 21d: 21-day worms, 28-d: 28-day worms, 35d: 35-day worms, F: 42–45 day female worms and M: 45-day male worms.