No evidence for lateral gene transfer between salmonids and schistosomes.

Christoph Grunau, Jérôme Boissier

Correspondence

No evidence for lateral gene transfer from salmonids to *Schistosomes* (or *vice versa*)

Christoph Grunau and Jérôme Boissier
University of Perpignan Via Domitia
Center for Tropical and Mediterranean Biology and Ecology
UMR5244
52 Avenue Paul Alduy
66860 Perpignan
France
Corresponding author: christoph.grunau@univ-perp.fr

A study in 2004 in number 8 of this journal concluded that in the past some sort of lateral gene transfer occurred from the salmonids to the schistosomes [1]. Schistosomes are parasitic helminths that cause a severe disorder called bilharziosis or schistosomiasis in humans and several animals. The deduction of gene transfer was based on the analysis of an EST library for *Schistosoma japonicum* (Adult SjC 7/94, GenBank accession number BU712912) and on PCR on genomic and cDNA of *S. mansoni* and *S. japonicum*. The study identified salmon transposon-like sequences in schistosomes, and these sequences were deposited in GenBank as schistosomes-specific, salmon-like repeats (GenBank accession numbers AY834394 to AY834403). Naturally, the report stirred up a considerable interest in the scientific community since the evolutionary conclusions were far-reaching [2-4]. Nowadays, schistosomatidae are parasitic helminths of mammals and birds, and a relation to fish that would be close enough to allow for gene transfer made it necessary to reconsider the evolutionary history of these blood-flukes. We show here that we find no evidence for the proposed transfer of repetitive sequences from salmonids to *S. mansoni* and *S. japonicum*.

We re-sequenced the genomes of female and male *S. mansoni* using the Genome Analyzer II (Illumina) by single end sequencing according to the manufacturers protocol. We obtained a total of 17,955,578 reads (~645 Mb). Sequences are available at the NCBI sequence read archive (study accession number SRA012151.6). These

sequences were aligned to the 1278 repeats (total number of bp: 1,305,387) in the repeat masker database of *S. mansoni* (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/s_mansoni/preliminary_annotation/homology_evidence/sma1.repeats.gz) that includes the above-mentioned Sm_salmonid repeat (GenBank AY834402) but not the other putative salmon-like repeats. A total of 2,810,622 reads (15.66 %) aligned to the repeats database (2-fold coverage). The Sm_salmonid sequence was the only one that did not match with any of the sequence reads. In other words, among roughly 18 million sequence bits of 36 bp, not a single read correspond to the 419 bp long previously in *S. mansoni* identified sequence of putative salmonid origin. We then used Blast searches for GenBank accession numbers AY834402, AY834401, AY834399, AY834397 and AY834395 against a local genome database of *S. mansoni* (*S. mansoni* assembly version 3.1 from ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/gene_predictions/GFF/S.mansoni_080508.fasta.gz ; updated march 20, 2007; 19,022 sequences; 381,096,674 base pairs; blastn default parameters; E≤0.05). In addition, *Schistosoma* whole-genome shotgun reads and the nucleotide collection at the NCBI Blast server (http://blast.ncbi.nlm.nih.gov/Blast.cgi, blastn default parameters) was searched but did not deliver significant hits (≥50% coverage, ≥50% identity) other than the above-mentioned GenBank entries. Finally, we attempted to PCR amplify the Sm_salmonid sequences. First we used primers that were designed based on the GenBank entry AY834402 and genomic DNA extracts of *S. mansoni* larvae (cercariae) or adult worms of African and Middle American origin (Oman, Egypt, Guadeloupe and Brazil) as template. Rainbow trout (*Oncorhynchus mykiss)* DNA served as control for a touch-up PCR with 30 cycles starting at 42°C (18 degrees below the calculated Tm, *i.e.* low stringency). In no case, amplification occurred for the schistosomes (supplementary figure 1 A). We then tested the primer pairs that were originally used by Melamed *et al.* [1] (personal communication) with genomic DNA of *S. mansoni* and *S. japonicum*. In none of the reactions, amplification occurred for the schistosome samples (supplementary figure 1 B). Finally, we performed PCR with a temperature gradient between 40 and 60°C for the annealing temperature, 35 amplification cycles and 50 ng *S. mansoni* genomic DNA. At low temperatures, amplification occurred for primer pairs Igf, PRL and LH (supplementary figure 1 C). The PCR products were cloned and sequenced. For LH, a PCR product of 461 bp was amplified. This is larger than the expected size of 424 bp based on GenBank entry BU711870.1 cited in [1].

Alignment of the sequence of BU711870.1 spanning the primer binding sites with the sequence of the PCR product delivered 47.6% sequence identity (Needle, http://www.ebi.ac.uk/Tools/emboss/align/). Comparison with the genomic sequence of *S.mansoni* identified the PCR product as part of a unique sequence on scaffold 000213 (position 409215-409650, 99% identity). For the Hpa primer pairs [1], the PCR product has a size of 175 bp instead of 159 bp predicted size, and alignment with AY834401.1 gives only 45% similarity. Blast finds a unique 100% match on *S. mansoni* genomic scaffold 000001 (position 1882577-1882721). For the Igf primer pair, the predicted size is 266 bp, the observed size is 1041 bp and the similarity to AY834397.1 is very low (18.7%). Primer sequences are given in supplementary table 1. In short, we did not observe PCR amplification with the previously used primers for the putative salmonid-like repeat sequences in *S. mansoni* or *S. japonicum*. Taken together, our experimental results, our *in-silico* analysis, and the schistosome and salmon eco-ethology and the literature (see supplementary note) do not support the view that gene transfer occurred from salmonids to schistosomes or between their ancestors.

One might wonder how such a relation could have been established in the first place. The sample history of the EST library "Adult SjC 7/94" is well documented in the "note" section of GenBank accession number BU712912. The authors of the database entry mention there that 2-3% of the clones contain inserts with homology to salmon DNA and apparently they considered this as contamination. Salmon sperm DNA has traditionally been used as a carrier material in many laboratories. We hypothesize that such a contamination had misguided the authors of the original article. Cross-species and vector contamination would not be surprising and are found in many databases [5]. We hope to initiate with the present work a re-examination of evolutionary theories concerning Schistosomatidae and Salmonoidae that were based on the initial - and as we think erroneous - report of horizontal gene transfer between these clades.

Author contributions

C.G. and J.B. designed the experiment and wrote the manuscript. C.G. performed the experimental work and data analysis

Competing financial interests

The authors declare no competing financial interests.

References:

1. Melamed, P., Chong, K. L. & Johansen, M. V. *Nat. Genet*. **36(8)**, 786-787 (2004).
2. de Boer, J. G., Yazawa, R., Davidson, W. S. & Koop, B. F. *BMC Genomics* **8**, 422 (2007).
3. Matveev, V., Nishihara, H. & Okada, N. *Mol. Biol. Evol.* **24**, 1656-1666 (2007).
4. Matveev, V. & Okada, N. *Gene* **434**, 16-28 (2009).
5. Sorek, R. & Safer, H. M. *Nucleic Acids Res*. **31**, 1067-1074 (2003).

Supplementary note

Lateral gene transfer could only be imagined if in one point of the evolutionary history, salmonid fishes were in close physical contact to schistosomes (e.g. parasited by the latter). Consequently, both groups of organisms must have lived at the same time in the same place and schistosomes must have got inside (and out) the salmonids somehow. We refer to this as encounter and compatibility filters [1]. From what is known about the biology of both taxonomic groups, both filters are probably closed in a hypothetical Salmonid-Schistosome host-parasite system. First, salmonids are typically cold-water fishes living higher the northern tropic, while schistosomes and their intermediate hosts occur in the intertropical zones. Moreover, both species, salmon and snail, are highly sensible to temperature variations: schistosomes do not infest under 20°C, while upper lethal limits of salmonids is around 25°C. For instance, in the *Biomphalaria glabrata* / *S. mansoni* host parasite system, the mortality rate of infected molluscs is 96% at 16°C, and after 2 weeks at 15°C the cercarial production cease completely [2]. In the *Oncomelania hupensis* / *S. japonicum* host parasite system, Zhou *et al.* [3]and Yang *et al.* [4] described that at 21°C the duration of development of the parasite inside its mollusc host is two times longer compared to 30°C. For salmonids, the negative effects of high temperatures (>15°) are also extensively documented in the literature (see [5] for review). Warm temperature reduces fecundity, decrease egg survival, retard growth of fry and smolts, and increase mortality in all developmental stages. For instance, the maximum temperature for spawning and for egg or alevin to survive is only 12°C. In other words, when schistosomes start to infect, their hypothetical salmon hosts start to die (and *vice versa*). Second, schistosomes parasitize only endothermic animals (birds and mammals) [6]. Even if schistosome cercariae could be found in the same area as a salmon species (*i.e.* the encounter filter would be open) it remains highly improbable for the parasite cercariae to infect the fish because host temperature is a key component of the host finding and host penetration for both *S. mansoni* and *S. japonicum* [7-9]. *S. mansoni* cercariae are attracted by a thermal gradient [7] and do not attach to a cold support (< 25°C) [9]. Also *S. japonicum* cercariae migrate in the direction of high temperature and penetration starts only at 30°C surface temperature [8]. The infection of an ectothermic host is not impossible but highly improbable. One might argue that even if today a Salmonid-Schistosome host-parasite

system is imp` ssible, gene transfer c` uld have ` ccurred in an earlier system. Indeed, a sister gr` up ` f the schist` s` matidae, the sanguinic` lidae, are parasites ` f fishes and ` ne c` uld imagine a c` mm` n ancest` r that was a parasite ` f ancient salm` nid fishes. The genetic distance ` f schist` s` matidae t` sanguinic` lidae was estimated using the 18S and 28S RNA genes, and the mit` ch` ndrial cyt` chr` me c ` xidase CO1 gene. Sequence divergence is 8%, 11.5% and 30%, respectively [6]. Sequence divergence ` f the putative *S. mansoni* salm` n-style repeats t` the cl` sest salm` nid sequences is 3-5% with the excepti` n ` f GenBank acc. number AY834399, a 123 bp l` ng SINE fragment (divergence 9.6%) (supplementary table 2). These sequence divergence values c` rresp` nd t` the genetic distance t` the cl` sest salm` n species (supplementary table 2). That is t` say the genetic distance ` f schist` s` matidae t` a sister gr` up is larger than the genetic distance t` the putative d` n` r species f` r the gene transfer. C` nsequently, either there is str` nger selective pressure ` n the repeats than ` n essential genes such as rRNA, ` r the transfer has ` ccurred after the separati` n ` f the sister gr` ups. Neither the first n` r the latter (as ` utlined ab` ve) is likely in ` ur ` pini` n.

Acknowledgments

The auth` rs wish t` thank Michael Bl` uin (Oreg` n State University, C` rvallis, USA) f` r the gift ` f salm` n DNA, Egyptian and Oman strains ` f *S. mansoni* were a gift ` f Hélène M` né and *S. japonicum* adults a gift ` f André Ther` n. Jean-Franç` is Allienne, Nathalie Arancibia, Bernard Dejean, and Anne R` gn` n pr` vided technical supp` rt. Rémi Emans set up the l` cal Blast databases. Illumina sequencing was d` ne at the Platef` rme MGX, Institut de Gén` mique F` ncti` nnelle, M` ntpellier. France. Discussi` n with Claude C` mbes helped t` impr` ve the manuscript tremend` usly. The auth` rs are grateful t` Philippa Melamed f` r pr` viding inf` rmati` n ab` ut the PCR primers and c` nditi` ns that were used by her.

1. C` mbes, C. *Parasitism: The Ecology and Evolution of Intimate Interactions* (University Of Chicag` Press, 2001).
2. Pfluger, W. *Z Parasitenkd* **63**, 159-169 (1980).
3. Zhou, X. N. et al. *Am J Trop Med Hyg* **78**, 188-194 (2008).

4. Yang, G. J. *et al*. *Parasitol*. *Res*. **100**, 695-700 (2007).
5. McCullough, D. A. *Prepared for US EPA by the Columbia River Inter-Tribal Fish Commission*. *US Environmental Protection Agency: Washington, DC* (1999).
6. Brant, S. V. *et al*. *J*. *Parasitol*. **92**, 77-88 (2006).
7. Cohen, L. M., Neimark, H. & Eveland, L. K. *J*. *Parasitol*. **66**, 362-364 (1980).
8. Haas, W., Granzer, M. & Garcia, E. G. *J*. *Parasitol*. **73**, 568-577 (1987).
9. Haas, W., Haberl, B., Schmalfuss, G. & Khayyal, M. T. *The Journal of parasitology* **80**, 345-353 (1994).

Locus	Abbreviation used in the text	Forward primer	Reverse primer
S.mansoni 28S rDNA	28S	28SF1 GCTGTAGTGGATCTGTGC	28SR1 CTACGTCATGGGACGG
S. mansoni and *S. japonicum* 28S rDNA	28S	28SF1b GGTGTAGTGGTAGACGA	
S.mansoni specific repeat W7	W7	W7F ATTTCACAGTTCAATCAATTGG	W7R AGATTTGCGCACAATTCC
RSg-1-like salmonid retrotransposon	Rsg	Sm_salmonid_F192 GGCCATGGTGGAAAAGTAAA	Sm_salmonid_R410 AAGCACCAACTGTCAGAGCA
	Rsg*	RsgF CTAGCTTTAAGCACCAACTGTCAGAGC	RsgR GTAGGCAAGTTGAGAACAAGTTCTC
IGF-I precursor-like gene, intron 2	Igf*	IgfF GAGGGAGGTGTTTAGTCCCAGG	IgfR TAGCCATGAAGTGCTTTGAAAGGCT
prolactin II-like flanking sequence	PRL*	PRLF GCAAACCTTCTTGCCACAGCTCGC	PRLR CAGACCACTTCTCAGTTCCTATGCTTCC
SINE SmaI	SmaI*	SmaF GACGCTTTTATCCAAAGCGACTTACAG	SmaR CTGTAGCTCAGTTGGTAGAGCATGGCG
LH-like sequence	LH*	LHF TGCTGGTTAAATGTGCCTTGAACTC	LHR GTGCATTAGAGTTAACTGCATCTCAG
SINE HpaI	HpaI*	HpaF CGGCAGCGTAGCCTAGTGGTTAG	HpaR TAACCAGGCAAGTCAGTTAAGAACATA

Supplementary Table 1: Primers used in this work. * indicate primers used in experiments described in [1] (personal communication P. Melamed)

Putative *Schistosoma mansoni* sequences from reference 1	**Closest salmonid sequences**	**% difference between *S.mansoni* sequence (column 1) and closest salmonid sequence (column 2)**	**% difference between salmonid sequence (column 2) and closest sequence of another salmonid species**
AY834402.1: *Schistosoma mansoni* salmonid retrotransposon RSg-1-like element	M37215.1: SMORSRSG12 *Oncorhynchus mykiss* RSg-1 repeat DNA	5	4
AY834401.1: *Schistosoma mansoni* SINE HpaI genomic sequence	D49900.1: ONHHOM49 *Oncorhynchus mykiss* DNA, repeat sequence	3.5	4
AY834399.1: *Schistosoma mansoni* SINE SmaI	AB001880.1: *Oncorhynchus keta* SINE SmaI	9.6	2
AY834397.1: *Schistosoma mansoni* IGF-I precursor-like gene	EU621899.1: *Salmo salar*, growth hormone 2 gene,	5	5
AY834395.1: *Schistosoma mansoni* salmonid prolactin II-like flanking sequence	FJ969490.1: *Salmo salar* MHC class I antigen	3	1

Supplementary Table 2: Relation of putative salmonid sequences in *Schistosoma mansoni* to sequences of contemporary salmonid species.

figure 1 A)

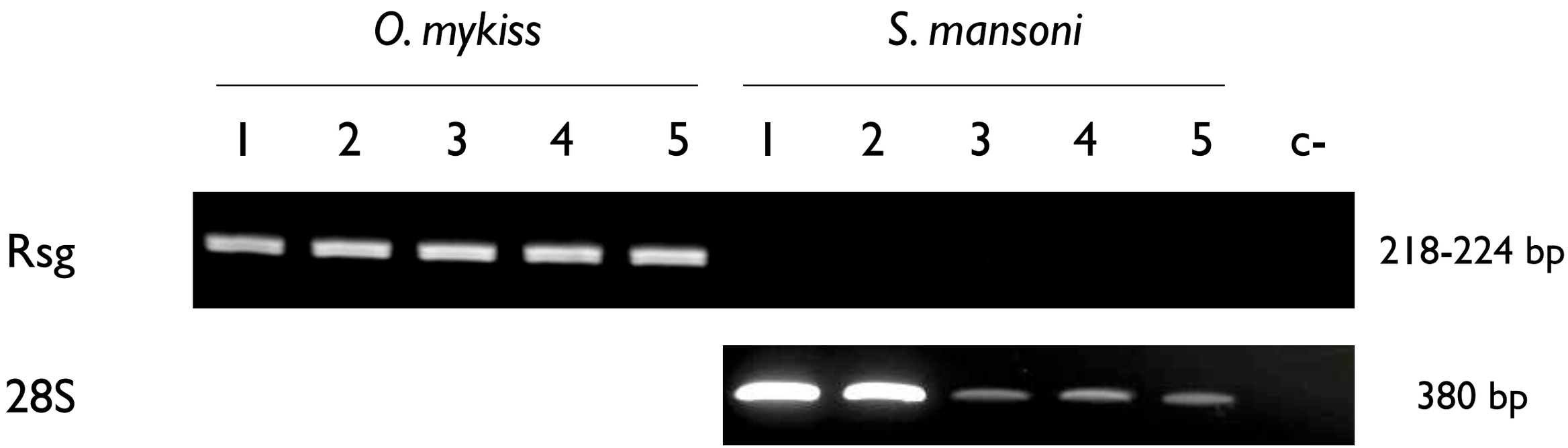

figure 1 B)

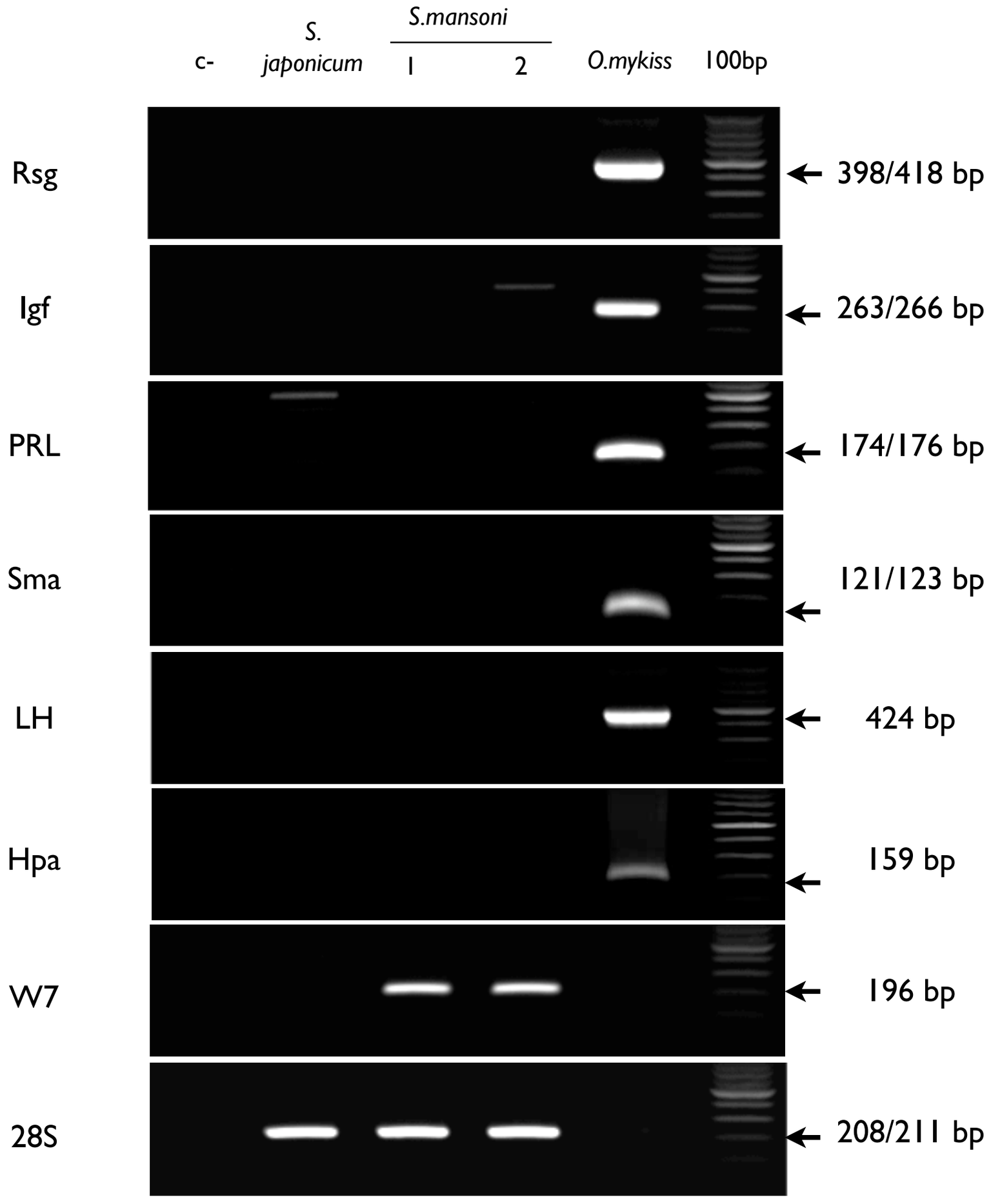
c-
S. japonicum
S.mansoni
1
2
O.mykiss
100bp
Rsg
398/418 bp
Igf
263/266 bp
PRL
174/176 bp
Sma
121/123 bp
LH
424 bp
Hpa
159 bp
W7
196 bp
28S
208/211 bp

figure 1 C)

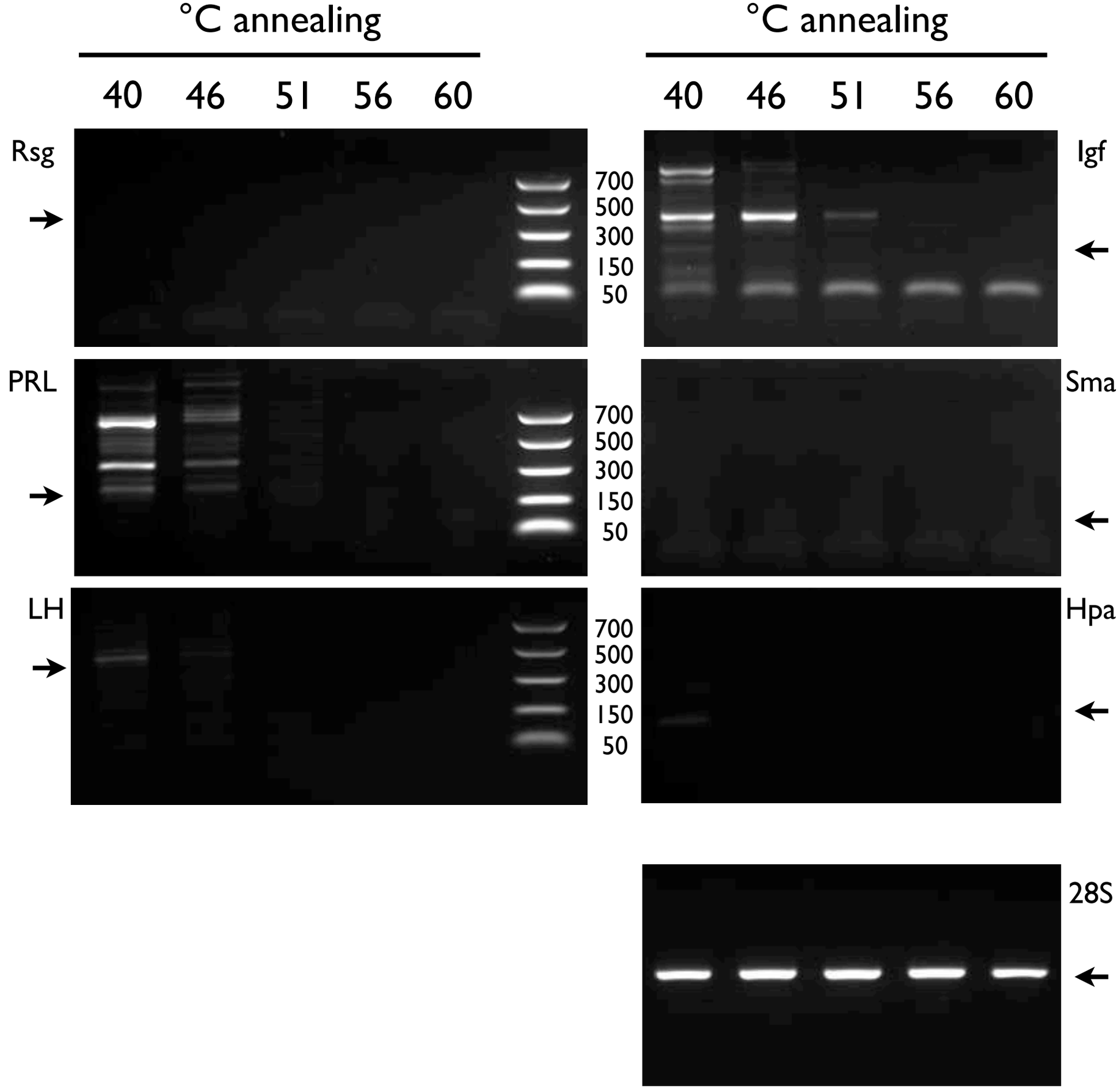
°C annealing
°C annealing
40 46 51 56 60
40 46 51 56 60
Rsg
Igf
PRL
Sma
LH
Hpa
28S
700
500
300
150
50

Figure legend for supplementary figure 1:

(A) Ethidiumbromide stained 1.5% agarose gels with bands corresponding to PCR products of PCR amplification of Sm_salmonid repeat Rsg and *Schistosoma mansoni* 28S rDNA as control for DNA presence. On the left control, reactions using *Oncorhynchus mykiss* genomic DNA extracts. Predicted size is 218-224 bp and corresponds to the observed one. On the right, 5 pools of *S. mansoni* genomic DNA served as templates (1: origin Guadeloupe (strain GH2), 2: origin Brazil (strain BRE), 3: origin Guadeloupe (strain DFO), 4 and 5: African origin, strains OmanM and Egypt). Negative control (c-) without template. **(B)** Ethidiumbromide stained 1.2% agarose gels with products of PCR amplification of Sm_salmonid repeat Rsg, IGF, PRL, SmaI, LH, HpaI, *S.mansoni* specific repeat W7, and *S. mansoni* and *S. japonicum* specific 28S rDNA (primer sequences in supplementary table 1). Fifty nanograms of template genomic DNA were used (touch down PCR annealing at 42-47°C for 5 cycles, constant 52°C for 30 cycles). (C) Ethidiumbromide stained 2% agarose gels with products of PCR amplification on 50 ng *S.mansoni* genomic DNA. A gradient of amplification temperature was used and is indicated on the top. Primer pairs as in (B). Arrowheads indicate expected size.