



A revision of the interrelationships of *Schistosoma* including the recently described *Schistosoma guineensis*[☆]

Bonnie L. Webster, Vaughan R. Southgate, D. Timothy J. Littlewood *

Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Received 21 December 2005; received in revised form 10 March 2006; accepted 21 March 2006

Abstract

In light of the recently described human schistosome *Schistosoma guineensis* and recent phylogenetic studies of the genus *Schistosoma*, a revision of the interrelationships of the members of this genus is needed. This paper adds to previous phylogenetic studies on the family Schistosomatidae and offers the most up to date and robust phylogeny of the group based on complete small and large nuclear subunit rRNA genes and partial mitochondrial *cox1*, incorporating most of the 21 species of *Schistosoma*. Our findings show that the group retains the same topology as that resolved in previous studies except *Schistosoma margrebowiei* was resolved as the sister taxon to all others in the *Schistosoma haematobium* species group and *S. guineensis* was placed as sister species to both *Schistosoma bovis* and *Schistosoma curassoni*. The *S. haematobium* species group contains eight species of which many are of significant medical and veterinary importance. Additionally, many of these species have been shown to hybridise both in the wild and experimentally, making the correct identification and recognition of species very important. A pairwise comparison of *cox1* among *Schistosoma* species suggests this gene alone would fail as a reliable barcode for species identification. Phylogenetic results clearly treat *Schistosoma intercalatum* and *S. guineensis* as separate taxa with each more closely related evolutionarily to *S. haematobium* than to each other. The study also highlights the problems associated with wrongly attributed sequences on public databases such as GenBank. © 2006 Published by Elsevier Ltd on behalf of Australian Society for Parasitology Inc.

Keywords: Phylogeny; *Schistosoma haematobium* species group; Identification; Africa

1. Introduction

Species concepts, species identification and species delineation have importance far beyond the discipline of taxonomy. Recognition of important parasite species is fundamental in understanding pathogens and, ultimately, controlling them. As Tibayrenc (2006) recently pointed out, identifying and circumscribing species is an on-going process that is constantly revisited for any taxon as more and more comparative information becomes available. Consequently, phylogenetic hypotheses estimating the interrelationships of species also need to be revisited if we are to evaluate and fully utilise the phylogenies. Recent phylogenetic analyses, estimating the interrelationships of mammalian schistosomes in the genus *Schistosoma* (Lockyer et al., 2003b; Morgan et al., 2003), and the recognition that two ‘strains’ of

Schistosoma intercalatum actually represent separate species (Kane et al., 2003; Pagès et al., 2003), make necessary a more up to date assessment of the interrelationships of the genus. Lockyer et al. (2003b) provided an extensive molecular data set providing a relatively robust estimate for relationships within the Schistosomatidae. Here, complementary data are added to the existing molecular data set of the Schistosomatidae. The newly recognised species of human schistosome, *Schistosoma guineensis*, is placed phylogenetically in the *Schistosoma haematobium* species group and considered in the wider context of the genus, with particular reference to species hybridisation known to occur within the group. The review is timely, particularly as greater reliance is being placed on molecular markers for species recognition (diagnostics) and the fact that molecular markers are revealing increasingly more cryptic taxa (e.g. Vilas et al., 2005).

Schistosoma guineensis occurs in Lower Guinea (Cameroon, Equatorial Guinea, Gabon, Nigeria and São Tomé; Wright et al., 1972; Southgate et al., 1994) and shows sufficient differences to be described as a separate species

[☆] Nucleotide sequence data reported in this paper are available in GenBank under the accession numbers: DQ354362, DQ354363, DQ354364.

* Corresponding author. Tel.: +44 20 7942 5742; fax: +44 20 7942 5151.

E-mail address: t.littlewood@nhm.ac.uk (T. Littlewood).

from *S. intercalatum* (Fisher, 1934; Tchuem Tchuente et al., 1997) found only in the Democratic Republic of Congo (DRC) (Pagès et al., 2003). Formal renaming of the species is supported on a number of features including different pre-patent periods in the intermediate and definitive hosts, egg morphology, intermediate host–parasite specificity, an inability of the two ‘strains’ to interbreed beyond the F₂ generation and robust phylogenetic evidence (reviewed in Kane et al., 2003).

In estimating the interrelationships of the Schistosomatidae, Lockyer et al. (2003b) utilised *S. intercalatum* from São Tomé, thus providing a phylogenetic estimate of the family actually using *S. guineensis* and not the true *S. intercalatum*. This was repeated by Morgan et al. (2003) who built on the same data set with addition of *Schistosoma hippopotami* (Thurston, 1963). Kane et al. (2003) used phylogenetic analyses to differentiate *S. guineensis* from *S. intercalatum*, although there is still a need for a comprehensive estimate using all members of *Schistosoma* and, more importantly, all members of the *S. haematobium* species group.

Historically, schistosomes of the genus *Schistosoma* have been grouped according to egg morphology and the genus of the intermediate host in which the parasite develops in nature (see Rollinson and Southgate, 1987). Other features of the worms may also reflect these wider groupings, e.g. position of sensory receptors in cercariae (Bayssade-Dufour, 1982). The groups are named after the most important exemplars of each group: *Schistosoma japonicum* Katsurada, 1904, *Schistosoma indicum* Montgomery, 1906, *Schistosoma mansoni* Sambon, 1907 and *S. haematobium* Bilharz, 1852. These groups are recognised as monophyletic clades with molecular evidence (Barker and Blair, 1996) except that *Schistosoma edwardiense* and *S. hippopotami* (originally members of the *S. mansoni* group) represent a distinct lineage according to Morgan et al. (2003). Additionally, *Schistosoma incognitum* Chandler 1926 falls outside a monophyletic *S. indicum* group. Finally, molecular evidence clearly includes at least one, if not all, members of the genus *Orientobilharzia* such that the original four species groups is now better represented as seven lineages (Lockyer et al., 2003b).

Members of the *S. haematobium* species group include *S. haematobium*, *S. intercalatum*, *S. guineensis*, *S. margrebowiei* Le Roux, 1930, *Schistosoma leiperi* Le Roux 1955, *Schistosoma mattheei* Veglia and Le Roux, 1929, *Schistosoma bovis* Sonsino, 1876 and *Schistosoma curassoni* Brumpt, 1931. The first three are parasites of humans and the others predominantly parasitise livestock (artiodactyls). Consequently, the group as a whole is of immense medical and veterinary importance. This study provides new complementary sequence data from ‘the true’ *S. intercalatum* collected from Kinshasa, the DRC, to estimate better the phylogeny of *Schistosoma*, and in particular all the members of the *S. haematobium* species group. Interrelationships of species are discussed in the light of the need to accurately identify species and the ability for some species to hybridise with one another.

2. Materials and methods

2.1. Source of sequence data

The complete small nuclear subunit ribosomal RNA gene (*ssrDNA*), the complete large nuclear subunit ribosomal RNA gene (*lsrDNA*) and a partial fragment of the mitochondrial cytochrome oxidase subunit I gene (*cox1*) of *S. intercalatum* from DRC (NHM.3367) were PCR amplified, sequenced and added to an existing data set established by Lockyer et al. (2003b). Also included in the analyses were complete *ssrDNA* and partial *cox1* sequences for *S. hippopotami* from Morgan et al. (2003) (AY197343 and AY197346). Sequences from other schistosomatids were used to root the *Schistosoma* tree; namely, *Ornithobilharzia canaliculata* Rudolphi 1819, *Austroilharzia terrigalensis* Johnston, 1917, *Austroilharzia variglandis* Miller and Northrup, 1926, *Heterobilharzia americana* Price 1929, *Schistosomatium douthitti* Cort 1914, *Bilharziella polonica* Kowalewski 1895, *Dendroilharzia pulverulenta* Braun, 1901, *Gigantobilharzia huronensis* Najim 1950, *Trichobilharzia regenti* Horak, Dvorak and Kolarova, 1998, *Trichobilharzia ocellata* La Valette 1855 and *Trichobilharzia szidati* Neuhaus, 1952 (accession numbers are given in Lockyer et al., 2003b). The interrelationships within these bird and mammal schistosomes are relatively poorly resolved and contentious depending on the genes used and taxa sampled (Lockyer et al., 2003b; Morgan et al., 2003; Snyder, 2004) and so the interrelationships of the outgroup clade were not considered in order to focus on the interrelationships of *Schistosoma* spp. For phylogenetic analyses of individual genes, trees of *Schistosoma* (and *Orientobilharzia*) species were rooted against members of the *S. japonicum* species group (*Schistosoma sinensium* Pao 1959; *S. japonicum* Katsurada 1904; *Schistosoma malayensis* Greer, Ow-Yang and Yong, 1988; and *Schistosoma mekongi* Voeg, Bruckner and Bruce, 1978; see Lockyer et al. (2003b).

2.2. DNA extraction, gene amplification and sequencing

Total genomic DNA was extracted from liquid nitrogen preserved adult *S. intercalatum* worms originating from the DRC (NHM.3367), using the DNeasy™ Tissue kit (Qiagen) according to the manufacturer’s protocol. The complete *lsrDNA* was amplified in three overlapping sections using the primer combinations LSU5+LSUD6, U1846+JM28S and 1200F+L2230 (see Table 1). The complete *ssrDNA* was amplified using the primers WA+WB (see Table 1) and amplification of the partial mitochondrial *cox1* was performed using the primers Cox1_Schist_5’ and Cox1_Schist_5’ (see Table 1).

Amplifications were performed in a total reaction volume of 25 µl using Ready-to-go PCR Beads (Amersham Pharmacia Biotech) 0.4 µM of each primer and 2 µl of DNA (~10 ng) according to the

Table 1

Primers used for the amplification and sequencing of the complete *lsrDNA*, *ssrDNA* and partial *cox1* genes

Primers	Forward (F) or reverse (R)	Primer sequence 5'-3'
<i>lsrDNA amplification</i>		
LSU5	F	TAGGTCGACCCGCTGAAYTTAAGCA
LSUD6	R	GGAACCCCTCTCCACTTCAGTC
U1846	R	AGGCCGAAGTGGAGAAGG
JM28S3'	F	TTCTGACTTAGAGGCGTTTCAG
L1200	F	CCCGAAAGATGGTGAACATATGC
L2230	R	AGACCTGCCTGCGGATATGGGT
<i>lsrDNA sequencing</i>		
L300	R	GTTCATGGCACTCCCTTTCAAC
ECD2	R	CTTGGTCCGTGTTTCAAGACGGG
L1642	R	CCAGCGCCATCCATTTTCA
L1600F	F	AGCAGGACGGTGGCCATGGAAG
L2630	R	GGGAATCTCGTTAATCCATTCA
L2229	F	TACCCATATCCGAGCAGGTCT
L4160	R	GTCTAAACCCAGTCCACGTTCCC
L3119	F	TTAAGCAAGAGGTGTCAGAAAAGT
<i>ssrDNA amplification</i>		
WA	F	GCGAATGGCTCATTAATCAG
WB	R	CTTGTTACGACTTTTACTTCC
<i>ssrDNA sequencing</i>		
300F	F	AGGGTTCGATTCCGGAG
300R	R	TCAGGCTCCCTCTCCGGA
930F	F	GCATGGAATAATGGAATAGG
1200F	F	CAGGTCTGTGATGCC
1200R	R	GGGCATCACAGACCTG
<i>cox1 amplification</i>		
Cox1_Schist_5'	F	TCTTTRGATCATAAGCG
Cox1_Schist_3'	R	TAATGCATMGGAATAAACA
<i>cox1 sequencing</i>		
CO1560	F	TTTGATCGGAATTTTGGTAC
CO1800	R	CCAACCATAAACATGTGATG

manufacturer's recommendations. Thermal cycling was performed in a Perkin–Elmer 9600 Thermal Cycler and the PCR conditions used for the *lsrDNA* and *ssrDNA* were: 2 min at 94 °C; 30 cycles of 1 min at 94 °C, 30 s at 52 °C (annealing temp), 2 min at 72 °C; followed by final extension period of 7 min at 72 °C. For the amplification of *cox1*, the annealing temperature was dropped to 49 °C and all the other conditions were the same. All PCR products were purified using Qiagen PCR Purification Kits (Qiagen) according to the manufacturer's protocol. Sequencing was performed using Fluorescent Dye Terminator Sequencing Kits (Applied Biosystems) and the sequencing reactions were run on either an Applied Biosystems 377 or a 373A automated sequencer. A variety of internal primers were used to obtain the full sequence of each fragment from both strands (see Table 1). The sequences were assembled and manually edited using Sequencher ver 4.2 (GeneCodes Corp.) and submitted to EMBL/GenBank (*cox1*: DQ354364; *ssrDNA*: DQ354363; *lsrDNA*: DQ354362). The complete *lsrDNA*, *ssrDNA* and partial *cox1* terminal 5' and 3' ends corresponding to primer sequences were removed prior to phylogenetic analysis.

2.3. Sequence alignment and phylogenetic analyses

The three gene sequences were aligned by eye to the large alignment generated by Lockyer et al. (2003b) for a phylogenetic assessment of the Schistosomatidae. No sequence data were available for the *lsrDNA* of *S. hippopotami*. Positions that could not be unambiguously aligned were removed prior to phylogenetic analysis. No additional phylogenetically informative sites were available, for any of the genes, if only members of *Schistosoma* were considered; i.e. a single alignment of schistosomatids and members of *Schistosoma* only utilised the same sites for analysis.

Phylogenetic analyses were conducted on individual genes and the combined data set using Bayesian inference (BI) and maximum parsimony (MP). The program Modeltest (version 3.6) (Posada and Crandall, 1998) was used to analyse each (gene) data set and select an appropriate model of genetic evolution for Bayesian analysis for each partition, employing the AIC (Akaike Information Content) criterion. The GTR+I+G model (rates set to gamma, with six substitution types) was used for each gene partition and the combined data set. Bayesian inference analyses were conducted using a separate GTR+I+G model for each data partition independently and also for the combined three-gene analysis, thus allowing separate estimates for each model parameter per data set (nst=6, rates=invgamma, ngammacat=4). Bayes analysis was performed using both the Macintosh and the Unix versions of MrBayes version 3.1 (Huelsenbeck and Ronquist, 2001, 2003); MrBayes (Bayesian analysis of phylogeny). Computer program distributed by the authors. The number of generations permitted was 3,000,000 with four chains (temp=0.2), sampled every 1000 generations. One million generations were discarded as burnin (burnin = 1000); likelihood scores had reached a plateau before 100,000 generations for all analyses. The consensus tree was constructed from the non-burn in trees ($n=2000$), and was a 50% majority rule tree. MP analysis was performed using PAUP* (Swofford, 2002. PAUP*. Phylogenetic Analysis Using Parsimony *and other methods). Sinauer Associates, Sunderland, Massachusetts) using a heuristic search strategy with 100 search replicates, random-addition taxon sampling, tree-bisection-reconnection branch-swapping, with all characters run unordered with equal weights and with gaps treated as missing data. Fifty random addition replicates were run to ensure all most parsimonious trees were found. MP nodal support was estimated by bootstrap resampling ($n=2000$). In all analyses, trees were rooted against the outgroup taxa. For the combined analyses *lsrDNA* for *S. hippopotami* was treated as missing and this taxon was excluded in the *lsrDNA* only analyses.

2.4. Gene sequences for *S. intercalatum* on GenBank

In order to assess the potential confusion in taxonomy associated with *S. intercalatum* in GenBank, a search was made to discover which nucleotide and amino acid sequences have been deposited under the name of *S. intercalatum*. In each case, the geographic source was noted from the accession or from the

associated publication in order to establish which sequences are likely to be from *S. guineensis*.

3. Results

The alignment of all three gene partitions was 7034 bp, of which 6729 positions were unambiguously alignable (1122 *cox1*; 1832 *ssrDNA*; 3775 *lsrDNA*), 1945 variable (599 *cox1*; 377 *ssrDNA*; 969 *lsrDNA*) and 1251 informative under the principles of parsimony (502 *cox1*; 182 *ssrDNA*; 567 *lsrDNA*). All the phylogenetic analyses estimated essentially the same tree topology with only minor differences in the placement of *Orientobilharzia* in relation to *S. incognitum* and the interrelationships of the members of the *S. haematobium* species group. These differences are largely due to insufficient phylogenetic resolution for just few of the individual data partitions at two or three nodes. Fig. 1 illustrates the combined data determined by BI, with posterior probabilities and bootstrap support from MP analysis marked on each node; this represents the best estimate reflecting the overall topology

resolved by these data. None of the topologies resolved from MP or BI of individual gene partitions differed significantly from the combined evidence tree shown in Fig. 1 (Templeton's test of data heterogeneity as described by Larson (1994), and implemented in PAUP* under the 'describe trees' option); individual analyses not shown. Differences between the BI topology shown in Fig. 1 and those of single gene analyses and/or MP are minor and none of the differences include nodes supported by high (>70%) bootstrap support, or high (>95%) posterior probabilities.

There is a total of 31 nucleotide and 12 amino acid (protein) accessions found in GenBank under the name of *S. intercalatum* and of these only seven nucleotide and four protein accessions are from specimens collected from the DRC (Table 2). All of the other accessions are from Lower Guinea, and are therefore actually *S. guineensis*. In other words, as a consequence of recent taxonomic changes we believe that over half of the nucleotide and amino acid accessions for *S. intercalatum* on GenBank are currently wrongly attributed.

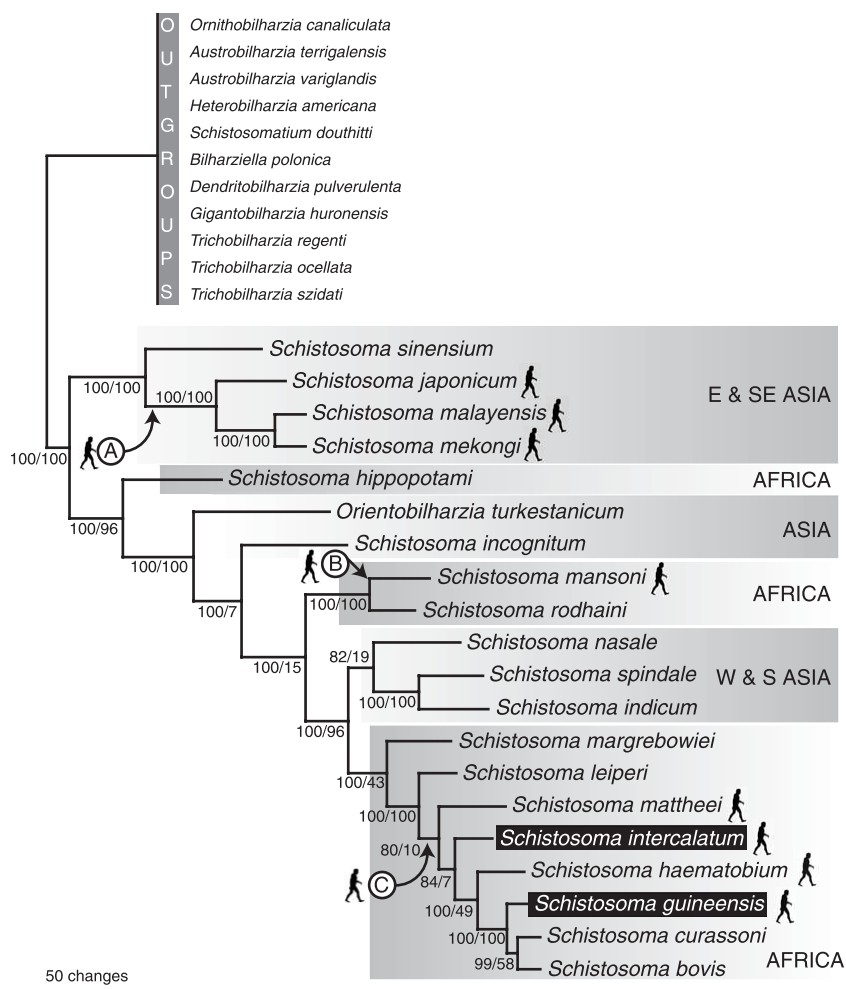


Fig. 1. Interrelationships of member species of *Schistosoma* estimated with a Bayesian analysis of combined partial *lsrDNA*, complete *ssrDNA* and partial *cox1*. Nodal support indicated as posterior probabilities and bootstrap percentages ($n=2000$) from a maximum parsimony analysis. The broad continental distribution is indicated for major lineages and clades, as is the likely evolutionary origin of human schistosomiasis on three separate occasions (A, B and C). The relative position of *Schistosoma intercalatum* and *Schistosoma guineensis* is highlighted to indicate they are neither sister taxa nor share a paraphyletic relationship.

4. Discussion

The phylogenetic information content of each of the three genes varies considerably, but combined, they appear to provide well-supported nodes throughout the genus *Schistosoma* (see also Johnston et al., 1993; Bowles et al., 1995; Littlewood and Johnston, 1995; Lockyer et al., 2003a). The general topology of Fig. 1 is identical to the topology found by Lockyer et al. (2003b) with the addition of *S. hippopotami* resolved as the sister group to the non-Asian *Schistosoma* and with the inclusion of the genuine *S. intercalatum* as sister group to the clade (*S. haematobium* (*S. guineensis* (*S. curassoni*, *S. bovis*))). Confirming recent similar studies, the Asian clade of *Schistosoma* species (the *S. japonicum* species group) was resolved as sister group to all other species (Snyder and Loker, 2000; Lockyer et al., 2003b). Present

results also confirm the position of an African clade represented by *S. hippopotami* (Morgan et al., 2003), the relative positions of *Orientobilharzia* and *S. incognitum*, and the dominant relationships within and between *S. mansoni*, *S. indicum* and *S. haematobium* species groups (Lockyer et al., 2003b; Morgan et al., 2003). Morgan et al. (2003) noted that the placement of an African species (*S. hippopotami*) or clade (when considering the sister group status of *S. edwardiense* and *S. hippopotami*) adds complexity to deciphering the origins of *Schistosoma* and suggested that species from hippopotami may have been instrumental in bridging the gap between Asia and Africa at least during the Cenozoic.

The relatively poor MP bootstrap support for the node leading to *S. incognitum* suggests that this species may still yet be resolved as the sister taxon to *Orientobilharzia* as suggested by *rrnL* data alone (Lockyer et al., 2003b). Greater sampling of

Table 2

GenBank accessions appearing under the name of *Schistosoma intercalatum*; those sequences from the Lower Guinea, West Africa have come from *Schistosoma guineensis* and those from DRC are from the true *S. intercalatum*

Source	Gene(s)	Accession		Original reference
		Nucleotide	Amino acid	
<i>Schistosoma guineensis</i>				
Cameroon	<i>cox1</i>	U22160	AAA84425	Bowles et al. (1995)
São Tomé and Príncipe: San Antonio	<i>cox1</i>	AY157208	AAO18391	Lockyer et al. (2003b)
São Tomé and Príncipe: San Antonio	<i>cox1</i>	AJ519517	CAD58304	Kane et al. (2003)
Cameroon: Edea	<i>cox1</i>	AJ519522	CAD58309	Kane et al. (2003)
Cameroon: Edea	<i>cox1</i>	AJ519523	CAD58310	Kane et al. (2003)
São Tomé and Príncipe: San Antonio	<i>nad6</i>	AJ416896	CAC95184	Kane et al. (2003)
Cameroon: Edea	<i>nad6</i>	AJ416902	CAC95190	Kane et al. (2003)
Cameroon: Edea	<i>nad6</i>	AJ416903	CAC95191	Kane et al. (2003)
Cameroon: Edea	<i>rrnS</i> (12S rRNA)	AJ419787		Kane et al. (2003)
Cameroon: Edea	<i>rrnS</i> (12S rRNA)	AJ419788		Kane et al. (2003)
São Tomé and Príncipe: San Antonio	<i>rrnS</i> (12S rRNA)	AJ419783		Kane et al. (2003)
Cameroon	5.8S, ITS2, 28S rRNA	U22166		Bowles et al. (1995)
São Tomé	ITS1, 5.8S, ITS2	Z21717		Kane and Rollinson (1994)
Cameroon	<i>rrnL</i> (16S rRNA)	L03655		Michot et al. (1993)
Cameroon	<i>lsrRNA</i> (28S rRNA)	U42558		Barker and Blair (1996)
Cameroon: Edea	<i>lsrRNA</i> (28S rRNA)	AJ519525		Kane et al. (2003)
Cameroon: Edea	<i>lsrRNA</i> (28S rRNA)	AJ519528		Kane et al. (2003)
São Tomé and Príncipe: San Antonio	<i>lsrRNA</i> (28S rRNA)	AY157262		Lockyer et al. (2003b)
São Tomé and Príncipe: San Antonio	<i>lsrRNA</i> (28S rRNA)	AJ519526		Kane et al. (2003)
Cameroon	<i>ssrRNA</i> (18S rRNA)	U42564		Barker and Blair (1996)
São Tomé and Príncipe: San Antonio	<i>ssrRNA</i> (18S rRNA)	AY157235		Lockyer et al. (2003b)
São Tomé and Príncipe: Chacara	<i>ssrRNA</i> (18S rRNA)	Z72146		Kaukas and Rollinson (unpublished)
São Tomé	mRNA ^a	AJ223840		Kane and Rollinson (1994)
São Tomé	mRNA ^a	AJ223841		Kane and Rollinson (1994)
<i>Schistosoma intercalatum</i>				
DRC: Kinshasa	<i>cox1</i>	AJ519515	CAD58302	Kane et al. (2003)
DRC: Kinshasa	<i>cox1</i>	AJ519519	CAD58306	Kane et al. (2003)
DRC: NHM3367	<i>cox1</i>	DQ354364		This study
DRC: Kinshasa	<i>nad6</i>	AJ416894	CAC95182	Kane et al. (2003)
DRC	<i>nad6</i>	AJ416899	CAC95187	Kane et al. (2003)
DRC: Kinshasa	<i>rrnS</i> (12S rRNA)	AJ419779		Kane et al. (2003)
DRC: NHM3367	<i>lsrRNA</i> (28S rRNA)	DQ354362		This study
DRC: Kinshasa	<i>lsrRNA</i> (28S rRNA)	AJ519527		Kane et al. (2003)
DRC: Kinshasa	<i>lsrRNA</i> (28S rRNA)	AJ519529		Kane et al. (2003)
DRC: NHM3367	<i>ssrRNA</i> (18S rRNA)	DQ354363		This study

Also indicated are new sequences of *S. intercalatum* (DRC: NHM3367) generated for this study. DRC, Democratic Republic of Congo.

^a With satellites and repeats.

Orientobilharzia species and isolates of the widely distributed *S. incognitum* will likely solve this conundrum.

The inclusion of all members of the *S. haematobium* species group allows for a more robust estimate of interrelationships within this group. Kane et al. (2003) were only able to estimate relationships within the entire group using *cox1* sequences. The present estimate using entire nuclear *ssrDNA* and *lsrDNA* and partial mitochondrial *cox1* does not recognise *S. margrebowiei* and *S. mattheei* as sister taxa as found by Kane et al. (2003). Instead, *S. margrebowiei* is sister to all other species within the group, with lineages leading to *S. leiperi* and *S. mattheei* as basal. Another important difference is the placement of *S. guineensis*. With *cox1* only, Kane et al. (2003) resolved *S. guineensis* as sister to *S. bovis* + *S. curassoni* (as found in the present study), but *cox1* + *nad6* + *rrnS* suggested *S. guineensis* was sister to *S. bovis*, albeit poorly supported. Resolving the relationships within the species group has important consequences when reconstructing the evolutionary association with definitive hosts, and in particular humans.

The new sequences allow us to estimate the number of times schistosomes have entered humans in evolutionary terms. Using parsimony as the criterion for character mapping (Maddison, W.P., Maddison, D.R., 2000. MacClade. Version 4. Sinauer Associates, Sunderland, Massachusetts.), *Schistosoma* has utilised humans as definitive hosts on at least three separate occasions as indicated on Fig. 1: A with the ancestor of *S. japonicum*, *S. malayensis* and *S. mekongi*, B with *S. mansoni*, and C with the ancestor of *S. mattheei*, *S. intercalatum*, *S. haematobium*, *S. guineensis*, *S. curassoni* and *S. bovis*. *Schistosoma mattheei* is usually found in domestic stock and wild ungulates and not normally in humans (e.g. Pitchford, 1959) and there is evidence to suggest that occurrences of *S. mattheei* infections in humans are in fact hybrids between *S. haematobium* and *S. mattheei* (Kruger and Evans, 1990); see also discussion below on hybridisation. Consequently, event C (in Fig. 1) may have been promoted by the domestication of cattle or, at least in an evolutionary and chronological sense, be more accurately placed at the next, more ‘crownward’ node, at the base of *S. intercalatum*, with the move from *S. mattheei* into humans as a subsequent host switch or hybridisation event. In the lineage leading from C, the parasite was lost from humans on the ancestral lineage leading to *S. curassoni* and *S. bovis* (which primarily use domestic stock as definitive hosts). Treating the entries into humans as single events with each species known to occur in humans would require there to be seven separate occasions (or eight if *S. mattheei* is included), and therefore less likely to have been the case under the principle of parsimony. Although, poorly supported, the topology of the *S. haematobium* species group estimated by Kane et al. (2003) using only mitochondrial genes, with *S. guineensis* within the crown of the clade, would suggest that within this group of schistosomes parasitised humans on two separate occasions, instead of once (event C) as suggested with the present data.

The close association between humans and the hosts (often domestic livestock) of *S. bovis* and *S. curassoni* strongly suggests that the ancestor of these two schistosomes speciated

from a predominantly human parasite. Indeed, the radiation of schistosomes in the *S. haematobium* species group seems closely associated not only with the movements of humans but their interaction with other suitable definitive hosts, provided by the animals with which they live in association. Additionally, hybridisation within and between human schistosomes and those infecting livestock demonstrates that these speciation events may be relatively recent (Southgate et al., 1998).

In most cases, the distribution of schistosome species is linked with the distribution of their intermediate and/or definitive hosts and these factors usually prevent interactions between different schistosome species from occurring. However, due to environmental changes areas of sympatry do occur, and as there are no physiological barriers preventing interspecific pairing of adult schistosomes, interspecific interactions can occur when two different species of schistosome are capable of infecting the same definitive host. Fig. 2 explores the radiation of the *S. haematobium* species group in greater detail and summarises published information on the ability of the constituent taxa to hybridise with one another in the laboratory and/or in nature. The result of such interspecific pairings depends upon the phylogenetic distance between the two species involved, and will lead to either hybridisation or parthenogenesis. Most experimental work has been carried out on species within the *S. haematobium* group, and experimentally it has been shown that hybridisation is possible between the different species combinations as shown in Fig. 2. However, in nature some of these hybrids are unlikely to occur because either the different species of schistosomes are allopatric or they may have distinct and different definitive host specificities, hence preventing any chance of interspecific interaction. Nevertheless, there is evidence that certain hybridisation events have taken place. There is a substantial amount of field data demonstrating that *S. guineensis* and *S. haematobium* interact in nature, and the outcome of such interactions is the eventual replacement of *S. guineensis* by *S. haematobium* through introgressive hybridisation (Webster et al., 2005). There is evidence based upon morphology of eggs and isoenzymes that *S. mattheei* occasionally infects man and hybridises with *S. haematobium* (Pitchford, 1961; Wright and Ross, 1980). Rollinson et al. (1990) provided evidence for the hybridisation of *S. bovis* and *S. curassoni* in parts of West Africa. Frandsen (1978) demonstrated that hybrids of *S. intercalatum* (Zaire, now DRC) and *S. intercalatum* (= *S. guineensis*) were only viable to the F₂ generation, and Pagès et al. (2002) confirmed the hybrid breakdown between these two parasites.

A recent review of schistosomes suggests that they retain a plesiomorphic, non-specific immune evasion strategy that allows ‘long-distance’ host switches, i.e. between evolutionarily diverse intermediate and definitive hosts (Brant and Loker, 2005). Even within the closely related *S. haematobium* species group, vertebrate host use is diverse (Fig. 2). Whether this is as a result of sampling bias or a true reflection of host

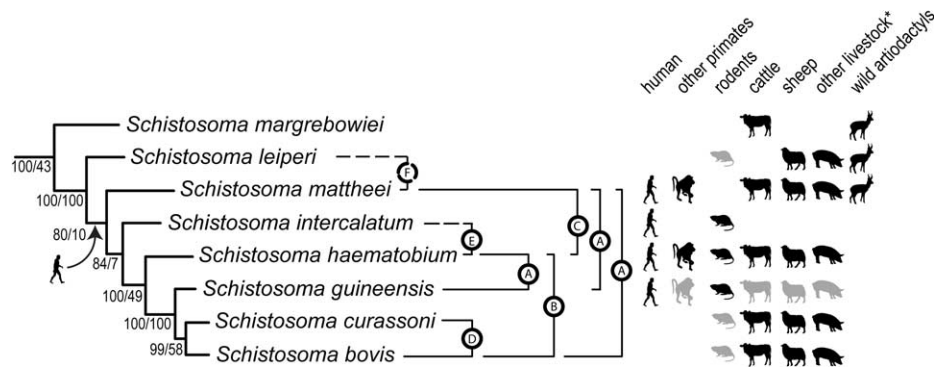


Fig. 2. Interrelationships of members of the *Schistosoma haematobium* species group indicating: (i) the distribution of parasites among different (natural) vertebrate hosts, according to the Natural History Museum (NHM) Host-Parasite Database (grey icons indicate experimental evidence, black icons indicate data from the wild), and (ii) the pairings of species known to hybridise (A–F) in the wild or experimentally (solid lines) or suspected or demonstrated for only limited generations (broken lines). References for known hybridisations include A, Southgate and Rollinson (1987), and Southgate et al. (1998); B, Brémond et al. (1993); C, Tchuem Tchuente et al. (1997), and Rollinson et al., (1990); D, Rollinson et al. (1990); E, Añé et al. (1997), and Webster and Southgate (2003); F, Vercruyse et al. (1994). Vertebrate host groups for each species are also indicated; *other livestock includes goats, pigs, horses and donkeys: these data are from the NHM Host-Parasite Database; <http://www.nhm.ac.uk/research-curation/projects/host-parasites/>

use, these host–parasite associations generally fall under the umbrella of ‘animals that come into close contact with humans’, whether domesticated or wild. Greater sampling seems necessary to fully appreciate the pool of available hosts for these (and other) species of *Schistosoma*. This may also reveal hitherto unrecognised cryptic taxa. Additionally, many older records would benefit from verification using modern molecular methods of identifying the schistosome species. Therefore, and in the light of apparent ease of hybridisation, to understand the epidemiology of human parasites in the *S. haematobium* species group, a far greater understanding of the species in nature is required.

It is clear from the present study that those researchers needing to differentiate between *S. intercalatum* and *S. guineensis* need to be aware of the taxonomic differences and the potential confusion associated with sequences supplied to GenBank (Table 2). Incorrectly attributed sequences on GenBank can and do have important consequences. Not only are there obvious systematic and taxonomic issues but also in the case of human pathogens these incorrect identities may impinge on epidemiological, immunological or other disease related issues. Presently, in spite of the large proportion of wrongly attributed accessions the actual number is relatively small, and with those sequences identified as belonging to *S. guineensis* (Table 2) there is time to correct previous accessions. Sequencing one or more of the *cox1*, *ssrRNA* and *lsrRNA* genes will potentially enable a researcher to run a reasonably swift and accurate identification (through phylogenetic analysis) of any of the members of the *S. haematobium* species group. However, a DNA barcoding approach as espoused by Hebert et al. (2003), utilising only a fragment of *cox1*, is unlikely to provide accurate identification of species of *Schistosoma* through either similarity or phylogenetics based approaches. In the present study, a fragment spanning 375 amino acids of *cox1* was

used. The fragment promoted as the ‘emerging...standard barcode region for higher [sic] animals’ is 216 amino acids (<http://barcoding.si.edu/DNABarCoding.htm>); amino acids 5–201 of the fragments used in this study, excluding regions used by ‘barcoders’ for priming. The pairwise comparison of *Schistosoma* amino acid sequences in this region shows that, some taxa differ by one or two amino acids within the fragment (Table 3). For example, *S. guineensis* differs from *S. bovis* and *S. curassoni* each by only one amino acid. The large number of 1, 2 or 3 amino acid differences between those taxa known to hybridise and even those that do not, suggests there is considerable scope for misidentification of closely related species using *cox1* alone. A number of significant pitfalls have been discussed in a ‘one gene’ approach to barcoding (Moritz and Cicero, 2004), and based on the molecular phylogenetic evidence in the present study, a molecular systematic approach to species identification requires a considerable amount of sequence data (using currently available markers), or a very different diagnostic gene or gene region yet to be found. In addition, as so many closely related members of *Schistosoma* readily hybridise, the utility of a single maternally inherited marker such as *cox1*, or any other mitochondrial marker, might further compound errors in species identification, particularly in hybrid zones. Notwithstanding the additional variation found at the nucleotide level within *Schistosoma*, and in spite of many other real and potential problems with barcoding (Hebert and Gregory, 2005; Will et al., 2005), until different markers are identified, all three genes utilised in the present study should ideally be sampled for optimal molecular identifications. Further studies evaluating geographic variation within these genes remains to be undertaken to fully evaluate them as molecular diagnostic markers for these important parasites.

Table 3
Pairwise differences in *cox1* between species of *Schistosoma*, including *Orientobilharzia turkestanicum* (Skryabin, 1913) of 197 amino acids recently promoted as the region of choice for 'DNA barcoding'.

<i>Schistosoma</i>	Sine	japo	mala	meko	turk	inco	mans	rodh	spin	indi	nasa	marg	leip	guin	haem	cura	matt	bovi	inte	hipp
<i>sinensium</i>	–																			
<i>japonicum</i>	17	–																		
<i>malayensis</i>	22	16	–																	
<i>mekongi</i>	21	17	1	–																
<i>O. turkestanicum</i>	36	34	40	41	–															
<i>incognitum</i>	37	36	36	37	30	–														
<i>mansoni</i>	39	37	41	40	40	34	–													
<i>rodhaini</i>	43	41	41	40	39	37	13	–												
<i>spindale</i>	38	38	36	37	35	37	26	23	–											
<i>indicum</i>	37	38	39	38	35	35	26	29	17	–										
<i>nasale</i>	38	37	39	40	36	34	23	27	14	16	–									
<i>margrebowiei</i>	39	37	38	39	29	34	31	32	15	15	18	–								
<i>leiperi</i>	39	37	38	39	30	33	29	30	14	14	15	3	–							
<i>guineensis</i>	40	37	39	40	30	33	30	31	16	14	17	5	2	–						
<i>haematobium</i>	41	38	39	40	32	34	27	28	15	13	16	8	5	3	–					
<i>curassoni</i>	39	36	38	39	30	32	29	30	15	13	16	6	3	1	2	–				
<i>matthei</i>	41	41	41	40	35	37	28	27	20	21	20	13	10	12	15	13	–			
<i>bovis</i>	39	36	38	39	31	34	30	31	16	15	17	5	2	1	4	2	12	–		
<i>intercalatum</i>	39	36	37	38	32	35	29	28	12	14	13	5	2	4	7	5	10	4	–	
<i>hippopotami</i>	33	35	36	37	37	38	44	42	36	41	41	37	38	39	40	38	40	39	38	–

Shaded cells indicate hybridisations known from the wild; note that many species not yet known to hybridise differ by as few as two amino acids.

Acknowledgements

We would like to thank Prof. L.A. Tchuem Tchuente for isolating *Schistosoma intercalatum* from Kinshasa, Democratic Republic of Congo and our colleagues in Biomedical Parasitology for ongoing support in schistosome phylogenetics. Julia Llewellyn-Hughes and Claire Griffin provided expert technical assistance in operating the gene sequencers. BLW and DTJL were supported by a Wellcome Trust Senior Research Fellowship to DTJL (043965/Z/95/Z).

References

- Añé, B.V., Añé, M.S., Abascal, H.F., Avila, J.P., Viamonte, B.V., 1997. Infection caused by *Schistosoma intercalatum* and probable hybridization with *Schistosoma haematobium* in East Africa. A case report. *Rev. Cub. Medic. Trop.* 49, 215–217.
- Barker, S.C., Blair, D., 1996. Molecular phylogeny of *Schistosoma* species supports traditional groupings within the genus. *J. Parasitol.* 82, 292–298.
- Bayssade-Dufour, C., 1982. Chétotaxies cercariennes comparés de dix espèces de schistosomes. *Annales de Parasitologie* 57, 467–485.
- Bowles, J., Blair, D., McManus, D.P., 1995. A molecular phylogeny of the human schistosomes. *Mol. Phylogenet. Evol.* 4, 103–109.
- Brant, S.V., Loker, E.S., 2005. Can specialized pathogens colonize distantly related hosts? Schistosome evolution as a case study. *PLoS Pathogens* 1, e38.
- Brémond, P., Sellin, B., Sellin, E., Naméoua, B., Labbo, R., Théron, A., Combes, C., 1993. Evidence for the introgression of the human parasite *Schistosoma haematobium* by genes from *S. bovis* in Niger. *Comptes Rendus de L'Académie des Sciences* 316, 667–670 (Serie 3).
- Fisher, A.C., 1934. A study of the schistosomiasis of the Stanleyville district of the Belgium Congo. *Trans R. Soc. Trop. Med. Hyg.* 28, 277–306.
- Frandsen, F., 1978. Hybridization between different strains of *Schistosoma intercalatum* Fisher 1934 from Cameroon and Zaire. *J. Helminthol.* 52, 11–22.
- Hebert, P.D.N., Gregory, T.R., 2005. The promise of DNA barcoding for taxonomy. *Syst. Biol.*, 54.
- Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. Ser. B.* 270, S96–S99.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Johnston, D.A., Kane, R.A., Rollinson, D., 1993. Small subunit (18S) ribosomal RNA gene divergence in the genus *Schistosoma*. *Parasitology* 107, 147–156.
- Kane, R.A., Rollinson, D., 1994. Repetitive sequences in the ribosomal DNA internal transcribed spacer of *Schistosoma haematobium*, *Schistosoma intercalatum* and *Schistosoma matthei*. *Mol. Biochem. Parasitol.* 63, 153–156.
- Kane, R.A., Southgate, V.R., Rollinson, D., Littlewood, D.T.J., Lockyer, A.E., Pagès, J.R., Tchuem Tchuente, L.A., Jourdan, J., 2003. A phylogeny based on three mitochondrial genes supports the division of *Schistosoma intercalatum* into two separate species. *Parasitology* 127, 131–137.
- Kruger, F.J., Evans, A.C., 1990. Do all human urinary infections with *Schistosoma matthei* represent hybridization between *S. haematobium* and *S. matthei*? *J. Helminthol.* 64, 330–332.
- Larson, A., 1994. The comparison of morphological and molecular data in phylogenetic systematics. In: Schierwater, B., Streit, B., Wagner, G.P., De Salle, R. (Eds.), *Molecular Ecology and Evolution: Approaches and Applications*. Birkhäuser Verlag, Basel, pp. 371–390.
- Littlewood, D.T.J., Johnston, D.A., 1995. Molecular phylogenetics of the four *Schistosoma* species groups determined with partial 28S ribosomal RNA gene sequences. *Parasitology* 111, 167–175.

- Lockyer, A.E., Olson, P.D., Littlewood, D.T.J., 2003a. Utility of complete large and small subunit rRNA genes in resolving the phylogeny of the Neodermata (Platyhelminthes): implications and a review of the cercomer theory. *Biol. J. Linn. Soc.* 78, 155–171.
- Lockyer, A.E., Olson, P.D., Østergaard, P., Rollinson, D., Johnston, D.A., Attwood, S.A., Southgate, V.R., Horak, P., Snyder, S.D., Le, T.H., Agatsuma, T., McManus, D.P., Carmichael, A.C., Naem, S., Littlewood, D.T.J., 2003b. The phylogeny of the Schistosomatidae based on three genes with emphasis on the interrelationships of *Schistosoma* Weinland, 1858. *Parasitology* 126, 203–224.
- Michot, B., Després, L., Bonhomme, F., Bachelier, J.-P., 1993. Conserved secondary structures in the ITS2 of trematode pre-rRNA. *Fed. Eur. Biochem. Soc. Lett.* 316, 247–252.
- Morgan, J.A.T., DeJong, R.J., Kazibwe, F., Mkoji, G.M., Loker, E.S., 2003. A newly-identified lineage of *Schistosoma*. *Int. J. Parasit.* 33, 977.
- Moritz, C., Cicero, C., 2004. DNA Barcoding: promise and pitfalls. *PLoS Biol.* 2, e354.
- Pagès, J.R., Southgate, V.R., Tchuem Tchuenté, L.A., Jourdane, J., 2002. Experimental evidence of hybrid breakdown between the two geographical strains of *Schistosoma intercalatum*. *Parasitology* 124, 169–175.
- Pagès, J.R., Jourdane, J., Southgate, V.R., Tchuem Tchuenté, L.A., 2003. Reconnaissance de deux espèces jumelles au sein du taxon *Schistosoma intercalatum* Fisher, 1934, agent de la schistosomose humaine rectale en Afrique. Description de *Schistosoma guineensis* n. sp.. In: Combes, C., Jourdane, J. (Eds.), *Taxonomy, Ecology and Evolution of Metazoan Parasites*, vol. II. Presses Universitaires de Perpignan, Perpignan.
- Pitchford, R.J., 1959. Cattle schistosomiasis in man in the Eastern Transvaal. *Trans. R. Soc. Trop. Med. Hyg.* 53, 285–290.
- Pitchford, R.J., 1961. Observations on a possible hybrid between the two schistosomes *S. haematobium* and *S. matthei*. *Trans. R. Soc. Trop. Med. Hyg.* 55, 44–51.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rollinson, D., Southgate, V.R., 1987. The genus *Schistosoma*: a taxonomic appraisal. In: Rollinson, D., Simpson, A.J.G. (Eds.), *The Biology of Schistosomes. From Genes to Latrines*. Academic Press, London, pp. 1–49.
- Rollinson, D., Southgate, V.R., Vercruyse, J., Moore, P.J., 1990. Observations on natural and experimental interactions between *Schistosoma bovis* and *Schistosoma curassoni* from West Africa. *Acta Trop.* 47, 101–114.
- Snyder, S.D., 2004. Phylogeny and paraphyly among tetrapod blood flukes (Digenea: Schistosomatidae and Spirorchiidae). *Int. J. Parasit.* 34, 1385–1392.
- Snyder, S.D., Loker, E.S., 2000. Evolutionary relationships among the Schistosomatidae (Platyhelminthes: Digenea) and an Asian origin for *Schistosoma*. *J. Parasitol.* 86, 283–288.
- Southgate, V.R., Rollinson, D., 1987. The natural history of transmission and schistosome interactions. In: Rollinson, D., Simpson, A.J.G. (Eds.), *The Biology of Schistosomes: From Genes to Latrines*. Academic Press, London, pp. 347–378.
- Southgate, V.R., Rollinson, D., Kaukas, A., Almeda, J., Sousa, A.M., Castro, F., Soares, E., Corachan, M., 1994. Schistosomiasis in the Republic of São Tomé and Príncipe: characterization of *Schistosoma intercalatum*. *Trans. R. Soc. Trop. Med. Hyg.* 88, 479–486.
- Southgate, V.R., Jourdane, J., Tchuem Tchuenté, L.A., 1998. Recent studies on the reproductive biology of the schistosomes and their relevance to speciation in the Digenea. *Int. J. Parasitol.* 28, 1159–1172.
- Tchuem Tchuenté, L.A., Southgate, V.R., Jourdane, J., Kaukas, A., Vercruyse, J., 1997. Hybridisation between the digeneans *Schistosoma haematobium* and *S. matthei*: viability of hybrids and their development in sheep. *Syst. Parasitol.* 36, 123–131.
- Tibayrenc, M., 2006. The species concept in parasites and other pathogens: a pragmatic approach? *Trends Parasitol.* 22, 66–70.
- Vercruyse, J., Southgate, V.R., Rollinson, D., de Clercq, D., Sacko, M., de Bont, J., Mungomba, L.M., 1994. Studies on transmission and schistosome interactions in Senegal, Mali and Zambia. *Trop. Geog. Med.* 46, 220–226.
- Vilas, R., Criscione, C.D., Blouin, M.S., 2005. A comparison between mitochondrial DNA and the ribosomal internal transcribed regions in prospecting for cryptic species of platyhelminth parasites. *Parasitology* 131, 839–846.
- Webster, B.L., Southgate, V.R., 2003. Isoenzyme analysis of *Schistosoma haematobium*, *S. intercalatum* and their hybrids and occurrences of natural hybridization in Cameroon. *J. Helminthol.* 77, 269–274.
- Webster, B.L., Tchuem Tchuenté, L.A., Jourdane, J., Southgate, V.R., 2005. The interaction of *Schistosoma haematobium* and *S. guineensis* in Cameroon. *J. Helminthol.* 79, 193–197.
- Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54, 844–851.
- Wright, C.A., Ross, G.C., 1980. Hybrids between *Schistosoma haematobium* and *S. matthei* and their identification by isoelectric focusing of enzymes. *Trans. R. Soc. Trop. Med. Hyg.* 74, 326–332.
- Wright, C.A., Southgate, V.R., Knowles, R.J., 1972. What is *Schistosoma intercalatum* Fisher, 1934? *Trans. R. Soc. Trop. Med. Hyg.* 66, 28–64.