Assessment of rDNA IGS as a molecular marker in the
Simulium damnosum complex

R. Morales-Hojas*,†, R. J. Post*, R. A. Cheke† and M. D. Wilson‡
*Department of Entomology, The Natural History Museum, London, U.K., †Natural Resources Institute,
University of Greenwich, Chatham Maritime, Chatham, Kent, U.K. and ‡Noguchi Memorial Institute for Medical Research,
University of Ghana, Accra, Ghana

Abstract. For five cytospecies of the Simulium damnosum Theobald complex of
blackflies (Diptera: Simuliidae) from West Africa, both ends of the intergenic
spacer region (IGS) of the rDNA have been sequenced with the aim of developing
specific molecular markers. No specific differences in these two regions were
detected between Simulium sanctipauli V. & D., Simulium sirbanum V. & D.,
Simulium soubrense V. & D., Simulium squamosum Enderlein and Simulium
yahense V. & D., except in the number of A subrepeats at the 5′ end of the IGS
(two in S. squamosum and four or five in the others) and in position 310 of the
3′ end (a C in S. squamosum and a G in the others). However, genetic distances
within and between species overlapped. These DNA sequences had no strong
phylogenetic signal, and the trees obtained were mostly unresolved. Although
most sequences from S. squamosum clustered together, a few of them were more
similar to those in other cytospecies. These results could be explained either by
hybridization with genetic introgression or by ancestral polymorphism and recent
speciation.

Key words. Simulium damnosum complex, ETS, hybridization, IGS, molecular
markers, recent speciation, ribosomal DNA, Africa.

Introduction

The ability to identify adult individuals of insect vector
species is of obvious importance in epidemiology, to help
understand their different roles in pathogen transmission
dynamics or to reduce disease incidence through selective
vector control. Afrotropical blackflies of the Simulium
damnosum Theobald complex include species responsible
for the transmission of Onchocerca volvulus Leuckart
(Nematoda: Onchocercidae), the filarial nematode that
causes human onchocerciasis in Africa and Yemen. More
than a dozen sibling species of the S. damnosum complex
have been described and are still recognized on the basis of
fixed inversion differences observed in polytene chromo-

omes of the larval silk glands (Boakye, 1993). However,
adult individuals do not have easily workable polytene
chromosomes and are difficult to identify to cytospecies
level.

Different methods have been tested for sibling species
identification of adults, including morphology (Garms,
1978; Garms & Cheke, 1985; Wilson et al., 1993), chromo-
some analysis (Procunier & Post, 1986), cuticular hydro-
carbon analysis (Carlson & Walsh, 1981; Phillips et al.,
1985), allozyme electrophoresis (Meredith & Townson, 1981;
Thomson et al., 1990) and DNA. Previous attempts to use
DNA to identify cytospecies have used differential hybrid-
ization of probes to anonymous middle-repetitive genomic
DNA sequences of different cytospecies of S. damnosum
sensu lato (Post & Flook, 1992). Some differences between
species were found, but they were confounded by intraspecific
variation. A second approach used the PCR to amplify the
intergenic spacer between the histone 3 and 4 genes, which
showed a 10-bp deletion that was unique to S. squamosum
Enderlein (mistakenly labelled S. damnosum Theobald sensu
Two mitochondrial DNA sequences have also been investigated for the identification of cytospecies. The 16S rDNA was too conserved and could not be used as molecular marker for the cytospecies of the *S. damnosum* complex (Tang et al., 1995a; Krüger et al., 2000). However, a directed heteroduplex analysis (DHDA) using a partial sequence of the ND4 gene of mtDNA was described that seemed to identify some of the cytospecies of the complex, but not *S. squamosum* and *S. yahense* Vajime & Dunbar (1975), which showed some identical alleles (Tang et al., 1995a, b). However, the differences remain to be confirmed. The nuclear rDNA ITS has also been tested for the cytospecies identification purpose, but the intraspecific variability was similar to the interspecific one, preventing the use of this sequence as a marker (Tang et al., 1996).

Overall, none of these techniques mentioned above are capable of discriminating all the West African cytospecies by themselves. The most unequivocal and reliable method is the multivariate analysis of morphological characters (Wilson et al., 1993), although this is not capable of discriminating with 100% accuracy the members of the *S. sanctipauli* and *S. damnosum* subcomplexes. Thus, there is still a need for an identification method for the *S. damnosum* complex in both West and East Africa. Furthermore, the characterization of DNA sequences that are suitable for use in identification of sibling species in Simulidae would allow the application of these sequences to other vector sibling species complexes, in South and Central America (Shelley et al., 1997).

The rDNA intergenic spacer (IGS) is known to be a region with a high rate of evolution (Hillis & Dixon, 1991), showing variability in structure and sequence between closely related species of plants and animals (e.g. Bhattacharya et al., 1996; Petersen & Vrain, 1996; Baldwin & Markos, 1998; Kane & Rollinson, 1998). Inter-specific variability has also been observed in the IGS from Diptera species, such as in tsetse flies (*Glossina* spp.) and *Drosophila* spp. (Cross & Dover, 1987; Tautz et al., 1987). In the malaria vector *Anopheles gambiae* Giles complex of mosquitoes, the IGS has been found to show interspecific variability and has been used successfully in the identification of cytospecies members of the complex (Paskewitz & Collins, 1990; Cornel et al., 1997; Favia et al., 1997). Nucleotide differences were observed in the first 400 bp of this region (13% of dissimilarity) between five cytospecies of the *An. gambiae* complex (Scott et al., 1993), and 10 nucleotide positions differed in the first 620 bp of the IGS between two cytoforms of *An. gambiae* s.s. (Favia et al., 2001). Similarly, this region could show differences between cytospecies of *S. damnosum* s.l., which could be used in the identification of the different cytospecies. The complete sequence of the IGS from *S. sanctipauli* Vajime & Dunbar has already been described (Morales-Hojas et al., 2002), and in the present paper the 5' and 3' ends of the IGS region of several cytospecies of the *S. damnosum* complex are compared and their potential as molecular markers is investigated. The 5' and 3' end fragments sequenced in this study together represent approximately one-third of the total length of the IGS (as aligned in Morales-Hojas et al., 2002).

### Materials and methods

#### Blackfly material

The material used in the study, with collection sites and dates and methods of preservation and identification is indicated in Table 1. Larvae were preserved in Carnoy’s fluid (3:1 absolute ethanol: glacial acetic acid) for chromosomal identification, and in absolute ethanol for DNA studies. Adult individuals were collected at human bait or reared from pupae. Samples were preserved dry in gelatine capsules with silica gel. Identification of the individuals was based on morphology and by association with cytological identification of larvae collected at the same site.

#### DNA extraction and amplification of the IGS 5' and 3' ends

DNA from ethanol and dry preserved individual specimens was extracted following the method of Ready et al. (1997). The IGS 5' end of different individuals belonging to the distinct cytospecies was amplified using the primers IGS forw1 5'CTGCTCTAAAGTATACCATGTGAC3', situated 186 bp upstream of the 28S 3' end, and IGSSint rev 5'CAA-GTTATCATACTGTACTGAC3', designed within the IGS region of *S. sanctipauli* (Morales-Hojas et al., 2002).

---

Table 1. Blackfly material used in the study.

<table>
<thead>
<tr>
<th>Species</th>
<th>Life stage</th>
<th>Location</th>
<th>Date</th>
<th>Preservation</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. sanctipauli</em></td>
<td>Larvae</td>
<td>Sutri rapids (R. Tano, SW Ghana, 05°31’N 02°27’W)</td>
<td>April 1998</td>
<td>EtOH; Carnoy’s fluid</td>
<td>Cytotaxonomy</td>
</tr>
<tr>
<td><em>S. squamosum</em></td>
<td>Larvae</td>
<td>Honuta (R. Todze, Ghana, 06°50’N 00°31’E)</td>
<td>April 1998</td>
<td>EtOH; Carnoy’s fluid</td>
<td>Cytotaxonomy</td>
</tr>
<tr>
<td></td>
<td>Adults</td>
<td>Amou-Oblo (R. Amou, Togo, 07°25’N 00°53’E)</td>
<td>March 1990</td>
<td>Dry</td>
<td>Morphology + cytotaxonomy</td>
</tr>
<tr>
<td><em>S. sirbanum</em></td>
<td>Adults</td>
<td>Tienfala (R. Niger, Mali, 12°44’N 08°05’W)</td>
<td>October 1988</td>
<td>Dry</td>
<td>Morphology + cytotaxonomy</td>
</tr>
<tr>
<td><em>S. yahense</em></td>
<td>Adults</td>
<td>Weada (R. Weada, Liberia, 06°53’N 10°22’W)</td>
<td>March 1989</td>
<td>Dry</td>
<td>Morphology + cytotaxonomy</td>
</tr>
<tr>
<td><em>S. soumense</em></td>
<td>Adults</td>
<td>Haindi (R. St. Paul, Liberia, 06°54’N 10°23’W)</td>
<td>March 1989</td>
<td>Dry</td>
<td>Morphology + cytotaxonomy</td>
</tr>
</tbody>
</table>

© 2002 The Royal Entomological Society, *Medical and Veterinary Entomology*, 16, 395–403
The PCR solution consisted of 1 × buffer (Promega, Southampton, U.K.), 2.5 mM MgCl₂ (Promega), 60 μM of each dNTP (PE), 0.2 μM of each primer, 0.5 U of Taq polymerase (Promega) and 1 or 2 μL of DNA template in a final reaction volume of 25 μL. The cycling conditions consisted of an initial denaturation step at 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min, with a final extension time of 10 min at 72 °C.

The IGS 3′ end of different individuals belonging to the distinct cytotypes (except S. yahense) was amplified using the primers ETS forw 5′GTACTCGTTCGCTACTC3′, which was designed within the IGS region of S. sanctipauli, and 185S′R 5′GATTTCACAGTTATCCCAAG3′, located 158 bp downstream from the 5′ end of the 18S rDNA (Morales Hojas et al., 2002). The amplification reactions consisted of 1 × buffer (Promega), 2.5 mM MgCl₂ (Promega), 60 μM of each dNTP (PE), 0.2 μM of each primer, 0.5 U of Taq polymerase (Promega) and 1 μL of DNA template in a final reaction volume of 25 μL. The cycle started with an initial denaturation step at 94 °C for 2 min, followed by 30 cycles of 30 s at 94 °C, 30 s at 55 °C and 1 min at 72 °C, with a final extension of 10 min at 72 °C.

Hot-starts were used in the reactions, with the polymerase being added after the first denaturation step. Negative controls with no DNA were always included. PCR products were run on 1% (w/v) agarose gels and visualized on a UV transilluminator. Sizes of the products were estimated using the 1 kb plus DNA ladder (Life Technologies Ltd, Paisley, U.K.).

**Sequencing of the IGS 5′-and 3′ ends**

Amplification products were extracted from the agarose gel with a genclean kit (Anachem Ltd, Luton, U.K.). They were subsequently cloned using the TOPO TA cloning kit from Invitrogen (Carlsbad, CA, U.S.A.) and recovered with the Hybaid Recovery Plasmid Midi Prep kit. Clones were cycle sequenced using the Big Dye (ABI, Applied Biosystems, Warrington, U.K.) sequencing kit in a Techne thermocycler. The sequencing cycle consisted of 25 cycles of 10 s at 96 °C, 5 s at 50 °C and 4 min at 60 °C, and reactions included 100 ng of insert DNA, 3.2 pmol of sequencing primer, 4 μL of Big Dye reaction mix, and 4 μL of 2.5× dilution buffer, in a total reaction volume of 20 μL. The number of IGS 5′ end sequences obtained were: 17 from seven individual sequences of S. sanctipauli, 18 from five individuals of S. squamosum, 20 from five individuals of S. sirbanum, 19 from five of S. soubrense and four sequences from a single specimen of S. yahense. The number of IGS 3′ end sequences obtained for the different cytotypes were: 16 from seven individuals of S. sanctipauli, 19 sequences from five individuals of S. squamosum, 20 from five specimens of S. sirbanum, and 20 sequences from five individuals of S. soubrense.

**Sequence analyses**

Sequence readings were verified against the chromatograms using Sequencher 3.1. Alignments were made using CLUSTAL W 1.7 and corrected by eye in Sequencer 3.1. Intra- and interspecific p-distances were calculated using MEGA, with alignment gaps treated as missing data and removed from the pair-wise comparisons. To establish the phylogenetic relationships between the species and the intraspecific (intra- and inter-individual) associations of the sequences, unrooted trees were reconstructed with three methods: unweighted pair group method with arithmetic mean (UPGMA), neighbour-joining (NJ) and maximum parsimony (MP), in phylogenetic analysis using parsimony (PAUP* version 4.0b4a, Sinauer Associates, Sunderland, MA, U.S.A.; http://paup.csit.fsu.edu). Evolutionary distances were calculated using the Kimura two-parameter model with the ts/tv set to 2. To obtain the MP trees, heuristic searches, with tree bisection-reconnection as the branch-swapping algorithm, were used. All characters were treated as unordered and had equal weights. Alignment gaps were treated as missing data and removed from the calculations. In the case of the 5′ end, the regions of the alignment spanning from 284 to 541 and from 621 to 694 were removed from the phylogenetic analysis because the alignment was not unequivocal. In total, 608 characters of the IGS 5′ end were included in the analysis. Fifty per cent majority-rule consensus trees were calculated for the three methods. The robustness of the branches was tested with bootstrap analyses using 1000 replications for UPGM and NJ trees, and 500 replications in MP trees.

**Results**

**The 5′ end of the IGS**

The sequences obtained varied in length from 682 to 811 bp, and the alignment was 940 bp long. The stretch of DNA included 187 bp of the 3′ end of the 28S rRNA gene, the A-subrepeats and 370 bp of the unique sequence that follows the A-subrepeats (Morales-Hojas et al., 2002). The number of A-subrepeats found in single arrays ranged between four and five in all cytotypes except in S. squamosum, where only two subrepeat units were found in all the sequences obtained (n = 18). An insertion of 14 bp was observed in all the S. squamosum and S. yahense sequences, and in some S. soubrense, in position 241–254 (immediately before the start of the A-subrepeats).

The A + T content of the sequences in all the cytotypes was on average 65.3% (range 65.1–65.4%). The intra-individual p-distance ranges were very similar to the inter-individual p-distances in all cytotypes, and the intraspecific divergence ranges overlapped with the intraspecific variability (Table 2). Nucleotide variation was observed, but most of these substitutions were present in just a few sequences. No fixed nucleotide differences were found to be diagnostic for a single cytotypes (Table 3).

None of the three methods of tree reconstruction were capable of fully resolving the evolutionary relationships between the cytotypes (MP tree shown in Fig. 1). The MP analysis showed that of 89 variable sites, 25 were parsimony-informative. The number of trees to be retained was set to 100,
and 100 equally parsimonious trees were found. The MP 50% majority-rule consensus tree had a length of 163, a consistency index (CI) of 0.5583, a homoplasy index (HI) of 0.4417, and a retention index (RI) of 0.6814. None of the methods used (UPGMA, NJ, MP) produced fully resolved trees, and the only branch highly supported by bootstrap was that leading to a cluster which consisted of most of the *S. squamosum* sequences. However, there were significant common features in the three consensus trees. Sequences from *S. sanctipauli* and *S. soubrense* were observed in all three methods to cluster together in a branch, although not all the sequences from these cytospecies were included. Probably the most significant feature was a marked lack of structure, such that sequences from the same cytospecies (and the same individual) did not necessarily cluster together.

**The 3′ end of the IGS**

The sequences obtained from the cytospecies *S. squamosum, S. sanctipauli, S. sirbanum* and *S. soubrense* ranged in length from 665 to 683 bp. The alignment of all the sequences was 691 bp, insertion/deletions of short motifs being the reason for the variation in length. The sequences included 158 bp of the 5′ end of the 18S rRNA gene, the entire ETS and 77 bp of the IGS 3′ end preceding the ETS (Morales Hojas et al., 2002). The A + T content of the sequences was on average 62.4% (range 62.2–62.5%).

The intraspecific p-distance ranges calculated can be seen in Table 4. Sequences belonging to the same individual were not necessarily more similar to each other than to sequences from another individual. P-distances estimated between cytospecies ranged from 0 to 2.4% (Table 4), their ranges overlapping with the intraspecific level of variation. Many nucleotide differences were observed, but most were present in few sequences. Nucleotide differences of interest at specific level are shown in Table 5. There was one nucleotide difference fixed in *S. squamosum*. Position 310 was a C in this cytospecies being G in the other cytospecies. There also was a (CT)₉ microsatellite located in position 172–189, which varied in number of repetitions between four and nine. However, there were no alleles specific for any of the cytospecies.

Trees estimated using UPGMA, NJ and MP were mainly unresolved (MP tree shown in Fig. 2). The UPGMA method gave no information whatsoever about the evolutionary relationships of the different cytospecies. The MP analysis was set to retain a maximum number of trees of 400, and the number of equally parsimonious trees saved was 400. Out of 106 variable sites, 20 were parsimony-informative. The MP 50% majority-rule consensus tree had a length of 157, with a CI of 0.7197, an HI of 0.2803 and a RI of 0.6271. Bootstrap support of branches in the NJ and MP trees was low. Although the NJ and MP trees were not fully resolved, there were some common patterns of interest. There was a single branch leading to all the sequences from *S. squamosum*, with about 80% support from bootstrap analysis. However, some sequences belonging to this cytospecies were more similar to other cytospecies. Sequences from the same individual and cytospecies did not necessarily cluster together.

**Discussion**

The low level of differentiation in the IGS between the cytospecies of the *S. damnosum* complex compared to that within cytospecies, was unexpected. This region has been used successfully to differentiate chromosomally distinct taxa in the *An. gambiae* complex (e.g. Paskewitz & Collins, 1990; Favia et al., 1997), in which 51 sites were variable in the first 400 bp of the IGS among five cytospecies (87% similarity)

---

**Table 2.** Comparison of the intra- and inter-individual, and interspecific p-distance ranges calculated in the IGS 5′ end. Distances are given as percentage (%).

<table>
<thead>
<tr>
<th>Species</th>
<th>Inter-individual</th>
<th>Intra-individual</th>
<th>S. sanctipauli</th>
<th>S. squamosum</th>
<th>S. sirbanum</th>
<th>S. yahense</th>
<th>S. soubrense</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. sanctipauli</em></td>
<td>0–1.4</td>
<td>0.3–1.3</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td><em>S. squamosum</em></td>
<td>0–1.3</td>
<td>0–1</td>
<td>0.78–2.54</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td><em>S. sirbanum</em></td>
<td>0.1–2.2</td>
<td>0.1–2.2</td>
<td>0.44–2.60</td>
<td>0.15–2.25</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td><em>S. yahense</em></td>
<td>0.1–1.1</td>
<td>0–1.1</td>
<td>0.65–1.94</td>
<td>0.15–2.03</td>
<td>0.14–2.03</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td><em>S. soubrense</em></td>
<td>0–2.3</td>
<td>0–1.5</td>
<td>0.14–2.18</td>
<td>0.46–2.60</td>
<td>0.45–2.65</td>
<td>0.33–2.11</td>
<td>−</td>
</tr>
</tbody>
</table>

---

**Table 3.** Fixed or almost fixed nucleotide cytospecific differences in the IGS 5′ end. *n* is the number of sequences analysed. When two nucleotides are present in the same position, the second is the alternative, and the number of sequences in which this alternative nucleotide is present is shown between brackets.

<table>
<thead>
<tr>
<th>Species</th>
<th>n</th>
<th>14 bp indel</th>
<th>717</th>
<th>836</th>
<th>902</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. sanctipauli</em></td>
<td>17</td>
<td>−</td>
<td>A</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td><em>S. squamosum</em></td>
<td>18</td>
<td>+</td>
<td>A/G (3)</td>
<td>C</td>
<td>A/T (3)</td>
</tr>
<tr>
<td><em>S. sirbanum</em></td>
<td>20</td>
<td>−</td>
<td>G/A (2)</td>
<td>C/A (3)</td>
<td>T</td>
</tr>
<tr>
<td><em>S. yahense</em></td>
<td>4</td>
<td>+</td>
<td>G/A (1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><em>S. soubrense</em></td>
<td>19</td>
<td>−/+</td>
<td>A</td>
<td>A</td>
<td>T</td>
</tr>
</tbody>
</table>
Fig. 1. Maximum parsimony consensus unrooted cladogram showing the evolutionary relationships between the different IGS 5′ end sequences of *S. sanctipauli* (Sutri), *S. soubrense* (S.sou), *S. squamosum* (Honuta), *S. sirbanum* (S.sb) and *S. yahense* (*S.y*). Numbers by the branches are the nucleotide substitutions. The number in bold and italics is the bootstrap value (only showing those with more than 95% support).

(Scott *et al.*, 1993), and 10 nucleotides differentiated two members (Mopti and Bamako/Savanna) of the *An. gambiae* s.s. within the first 620 bp of the IGS (Favia *et al.*, 2001). As a result, a PCR-based method was developed to differentiate these cytospecies, using nucleotide differences to develop specific primers (e.g. Scott *et al.*, 1993). In *S. damnosum*, however, fewer nucleotide differences have been detected, and only one apparently fixed nucleotide diagnostic for

<table>
<thead>
<tr>
<th>Species</th>
<th>Inter-individual</th>
<th>Intra-individual</th>
<th><em>S. sanctipauli</em></th>
<th><em>S. squamosum</em></th>
<th><em>S. sirbanum</em></th>
<th><em>S. soubrense</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. sanctipauli</em></td>
<td>0–1.6</td>
<td>0–1.6</td>
<td>–</td>
<td>0.15–2.41</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. squamosum</em></td>
<td>0–1.2</td>
<td>0–1.2</td>
<td>0–2.07</td>
<td>0.30–1.81</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. sirbanum</em></td>
<td>0.1–1.3</td>
<td>0–1.2</td>
<td>0–2.07</td>
<td>0.30–1.80</td>
<td>0–1.48</td>
<td>–</td>
</tr>
<tr>
<td><em>S. soubrense</em></td>
<td>0.1–1.3</td>
<td>0.2–1.2</td>
<td>0–2.07</td>
<td>0.30–1.80</td>
<td>0–1.48</td>
<td>–</td>
</tr>
</tbody>
</table>

© 2002 The Royal Entomological Society, *Medical and Veterinary Entomology*, 16, 395–403
Table 5. Fixed or almost fixed nucleotide cytospecific differences in the IGS 3′ end. n is the number of sequences analysed. When two nucleotides are present in the same position, the second is the alternative, and the number of sequences in which this alternative nucleotide present is shown between brackets.

<table>
<thead>
<tr>
<th>Species</th>
<th>n</th>
<th>227</th>
<th>232</th>
<th>233</th>
<th>310</th>
<th>314</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. sanctipauli</td>
<td>16</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>T/C (1)</td>
</tr>
<tr>
<td>S. squamosum</td>
<td>19</td>
<td>A/T (8)</td>
<td>A/C (4)</td>
<td>T/A (4)</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>S. sirbanum</td>
<td>20</td>
<td>T/C (1)</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>S. sobrense</td>
<td>20</td>
<td>T</td>
<td>C</td>
<td>A/G (1)</td>
<td>G</td>
<td>T</td>
</tr>
</tbody>
</table>

Fig. 2. Maximum parsimony consensus unrooted cladogram showing the evolutionary relationships between the different IGS 3′ end sequences of S. sanctipauli (Sutri), S. squamosum (Honuta and S.sq), S. sobrense (S.sou) and S. sirbanum (S.sb). Numbers by the branches are the nucleotide substitutions. No branches have a bootstrap support higher than 95%.
S. squamosum (position 310 of the 3' end), thus precluding the development of different cytotypespecific primers for practical identification. The copy number of A-subrepeats was also consistently different between S. squamosum and the other cytotypes, and it could be used as a diagnostic feature for this species. However, a more extensive sampling is required before asserting the specificity.

The level of interspecific differentiation in S. damnosum s.l. was lower than that between subspecies of the tsetse Glossina morsitans Westwood (Diptera: Glossinidae) (2.1–4.1%), and slightly higher than the variability within subspecies of G. morsitans (0.7%) and within other Glossina species (0.3–0.6%) (Cross & Dover, 1987). In the An. gambiae complex, the genetic differentiation between cytotypes was also higher than that observed in the S. damnosum complex (see above). It is significant that the range of interspecific genetic distances among cytotypes of S. damnosum s.l. overlapped with the intraspecific ranges, in both the 5' and 3' ends of the IGS, and hence some sequences were more similar to sequences from other cytotypes than to their own. This was reflected in a low phylogenetic signal of this region, resulting in mostly unresolved trees. It also rendered this region probably unusable as a cytotypes molecular marker.

The only consistent pattern observed in the trees estimated from both the 5' and 3' ends of the IGS, suggesting a real evolutionary pattern, was the placement of almost all S. squamosum sequences in a differentiated branch. Protein electrophoretic analyses, as well as tree estimation using mitochondrial ND4 gene sequences, suggested that the S. squamosum subcomplex is more differentiated (Meredith & Townsend, 1981; Tang et al., 1995a), and the spacer between the histone 3 and histone 4 genes also showed a deletion unique to S. squamosum (Wilson & Post, 1994). Although the results obtained in this study are partly in accordance with those of the enzymes and mitochondrial ND4 gene, a few sequences indicated that some S. squamosum IGS variants were more similar to the IGS of other cytotypes.

The pattern of low differentiation between cytotypes within the S. damnosum complex has also been observed in other markers. Electrophoretic analysis of 15 enzymes displayed seven that were polymorphic, but with no diagnostic value, and two which showed some alleles specific for S. squamosum and S. yahense (PGM had an allele specific for S. yahense, and Trehalase had one allele diagnostic for both cytotypes together) (Meredith & Townsend, 1981; Thomson et al., 1990). In comparison, the percentage of loci diagnostic between sibling species of the D. willistoni Sturtevant group (for example) was calculated to be 32% (Ayala et al., 1974; Avise, 1976), whereas in the S. damnosum complex only 6% of the isoenzyme loci were diagnostic for a sibling. This percentage of sibling species diagnostic loci in S. damnosum was the same as that calculated between subspecies of the D. willistoni group (Ayala et al., 1974; Avise, 1976). In the D. virilis group of species, 14% (mean value) of larval proteins were unique to individual species (Hubby & Throckmorton, 1965), whereas the estimated percentage of isoenzymes unique to each of the sibling species of the S. damnosum complex was 1%.

In addition to the rDNA IGS reported here, other DNA studies also suggest that the level of differentiation between the cytotypes in West Africa is low. The ITS of the nuclear rDNA revealed a low level of heterogeneity between cytotypes, and the level of intraspecific variability was similar to the interspecific differentiation (Tang et al., 1996). Also, the 16S rDNA of the mitochondrial genome was conserved among the cytotypes, and phylogenetic analyses did not produce well-resolved evolutionary trees (Tang et al., 1995a, b; Krüger et al., 2000). The mitochondrial ND4 gene was reported to differentiate cytotypes using a directed heteroduplex analysis, and produced a tree congruent with cytotaxonomy and the ecological division (Tang et al., 1995a, b). However, subsequent attempts to use the method have not always produced results consistent with the cytotaxonomy, and the method requires more extensive evaluation.

Some of the possible explanations for the low interspecific differentiation observed in the IGS include (1) interspecific hybridization and (2) recent speciation. Hybridization between species can result in symmetrical or asymmetrical genetic introgression, with the gene flow diminishing the allele frequency differentiation between the hybridizing species (Dowling & Secor, 1997). One of the possible outcomes of this horizontal transfer of genetic material would be an increased level of diversity which, given the appropriate ecological and geographical conditions, could end up in the origin of new hybrid taxa (Arnold, 1992; Dowling & Secor, 1997; Arnold et al., 1999). Our results showed high intraspecific variability with respect to the interspecific divergence, in accordance with the above prediction. Reticulate evolution of animal taxa is reflected in the phylogeny by either showing non-concordance in the history of different sequences or by lack of resolution of defined clades (Arnold, 1992). The trees estimated with the sequence data showed a marked lack of resolution, and the fact that some sequences from different cytotypes were more similar to each other than within species could be explained if there had been an exchange of genetic material between the siblings. Indeed, it has been shown by several authors that hybrids exist, and that backcrosses to parental species occur, indicating that hybrids are viable and reproductive (Meredith et al., 1987; Boakye & Mosha, 1988; Boakye & Meredith, 1993; Boakye et al., 2000). The frequency of hybrid individuals in nature was estimated by Post (1984) to be one in 1000, and that of backcrosses one in 10,000. However, this could be an underestimation because of the difficulty in identifying backcrosses (Boakye & Meredith, 1993).

The other possible explanation for the reduced level of interspecific differentiation compared to the intraspecific variation involves ancestral polymorphism and recent speciation. In S. damnosum s.l., cytotypes have been described on the basis of chromosomal inversions. Some of the inversions are not fixed for the different cytotypes but show polymorphism within species (e.g. Traore-Lamizana et al., 2001). Also, at the DNA and protein level, differences between these cytotypes are low. These findings could suggest that speciation has occurred recently with a rapid change at the chromosomal level but no genetic revolution. An example of this
is seen in the superspecies *Spalax ehrenbergi* Nehring (Mammalia: Muridae) (a group of fossorial rodents of the Middle East), in which a complex is formed by several parapatric chromosomal forms with ethological isolation (Nevo, 1969; Nevo et al., 1975), but with little genetic differentiation (Nevo & Shaw, 1972).

There is no possibility of differentiating between these two hypotheses with the present data. In order to be able to distinguish between them, a detailed population genetic study with a more adequate sampling of different geographical populations for each cytotypes, and with a larger number of mitochondrial and nuclear loci surveyed, would be needed.

**Acknowledgements**

R.M.H. was funded by a studentship from The Natural History Museum, London. The research was partly supported by a research development award (reference number 040886/Z/94/Z/14X/REH/CS) by The Wellcome Trust to R.A.C. and M.D.W., for which they are most grateful. Sequences have been submitted to the GenBank and have accession numbers: *S. sanctipauli* IGS 3′ end AF421625–AF421640 + AF421622–AF421624, 5′ end AF421641–AF421657, *S. squamosus* IGS 3′ end AF421566–AF421581, 5′ end AF421658–AF421675; *S. sirbanum* IGS 3′ end AF421582–AF421601, 5′ end AF421676–AF421695; *S. soumbrere* IGS 3′ end AF421602–AF421621, 5′ end AF421700–AF421718; *S. yahense* IGS 5′ end AF421696–AF421699.

**References**


Molecular markers in the Simulium damnosum complex


Accepted 18 June 2002

© 2002 The Royal Entomological Society, Medical and Veterinary Entomology, 16, 395–403