# HYPOTHETICAL 'NO MATCH' GENES IN *SCHISTOSOMA MANSONI* ARE EXPRESSED IN ADULT WORMS

**Dr. Sergio Verjovski-Almeida**

Universidade de São Paulo, Instituto de Química, Brazil

An assembly of all publicly available mRNA and EST sequence data for *Schistosoma mansoni* has resulted in 13,355 contigs or singlets for which a putative encoded protein could be determined, plus 14,213 “no match” contig or singlet sequences, for which no ortholog was found in GenBank. Based on this information, we designed in our group a new *S. mansoni* 60-mer oligonucleotide microarray platform with approximately 44,000 probes. The large number of probes combined with the extensive sequence annotation available allowed a comprehensive approach, where most of the *S. mansoni* transcriptome is represented in our oligoarray. Sequences for which a putative encoded protein had been determined had a known protein-coding strand, and we were able to design single-stranded unique oligonucleotide probes for 11,107 of them, representing an efficiency of 83%. For “no match” sequences, where the coding strand is unknown, we designed probes for both strands whenever possible; we have successfully found unique probes for each of the sense and antisense strands of 13,045 “no match” sequences; for an additional 1,168 “no match” sequences a probe for only one strand was found. Although not every sequence could successfully be represented by a probe, the overall efficiency of the probe design process was high (76%). In addition to the probes designed directly from *S. mansoni* genes, we selected *S. japonicum* transcripts from GenBank without correspondence to any sequence in the *S. mansoni* transcriptome assembly, but with similarity to some genomic *locus* in the *S. mansoni* genome. In this case, probes were designed from the *S. mansoni* genomic region having similarity to the *S. japonicum* transcript. As a result 3,344 additional probes were designed, representing 2,727 putative novel *S. mansoni* gene orthologs, that were not represented in the *S. mansoni* EST and mRNA GenBank databases. Overall, we obtained 39,342 probes representing 19,907 different putative unique genes. Hybridization of the oligoarrays with adult worm RNA pointed to a set of genes transcriptionally active in this stage of the parasite’s life cycle. Interestingly, a large proportion (43%) of genes for which transcription was detected in adults is comprised of “no match” genes, i.e. *S. mansoni* genes with no identifiable orthologs in GenBank. Although some of these “no match” genes may represent 5’ and 3’ non-coding regions of known genes, it is expected that the majority of these “no match” genes encodes Schistosome specific proteins of unknown function. Analysis of the set of “no match” genes for which we have designed probes for both strands shows that both sense and antisense messages were detected for a considerable fraction (6.8%) of the active genes in adults. A wide range of sense-antisense expression ratios has been found. It is apparent that bi-directional transcription is widespread in *S. mansoni*, and further characterization of its function is warranted.