PRIMER NOTE

Microsatellite-enriched genomic libraries as a source of polymorphic loci for *Schistosoma mansoni*

N. B. RODRIGUES,* M. R. SILVA,* M. M. PUCCI,* D. J. MINCHELLA,† R. SORENSEN,‡ P. T. LOVERDE,§ A. J. ROMANHA,* and G. OLIVEIRA*¶

*Centro de Pesquisas René Rachou-FIOCRUZ, Av. Augusto de Lima 1715, 30190-002, Belo Horizonte, MG, Brazil, †Department of Biology, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA, ‡Department of Biology, Minnesota State University at Mankato, MN 56001, USA, §Southwest Foundation for Biomedical Research, PO Box 760549, San Antonio, TX 78245-0549, USA, ¶Programa de Pós-Graduação e Pesquisa, Santa Casa de Belo Horizonte, MG 30150-221, Brazil

Abstract

Microsatellite markers for *Schistosoma mansoni* were developed using four genomic microsatellite-enriched libraries. Microsatellites were observed in 65.4% of all sequences. Primer pairs were designed and tested for 23 loci. Eighteen loci produced amplification products, out of which 11 were polymorphic and were further characterized on 100 individuals of *S. mansoni*. Two to 19 alleles per locus were detected. The average values of expected and observed heterozygosities among the 11 loci were 0.79 and 0.59, respectively.

Keywords: computational biology, genome, genomic library, microsatellites, population genetics, *Schistosoma mansoni*

Received 10 July 2006; revision received 1 September 2006; accepted 1 September 2006

*Schistosoma mansoni* is the main causative agent of schistosomiasis, a human disease affecting over 200 million people (WHO 2002). The development of tools such as polymorphic microsatellite loci for the understanding of genetic variation and population structure of this parasite may improve our knowledge of disease transmission and allow the identification of genes of interest by linkage analysis. Approximately 30 polymorphic microsatellite markers, mostly derived from cDNAs, are available for *S. mansoni* (Durand et al. 2000; Blair et al. 2001; Curtis et al. 2001; Rodrigues et al. 2002). However, selective pressures on expressed genes may decrease microsatellite polymorphism (Stohler et al. 2004). We sequenced clones from genomic microsatellite-enriched libraries and identified new polymorphic microsatellite loci.

Four *S. mansoni* genomic libraries enriched for AAT, CA, GA and TAGA repetitive sequences were constructed and cloned into pUC19 (Genetics Information System). The insert sizes ranged from 300 to 700 bp. Recombinant clones were transformed into DH5α Escherichia coli, and selected on LB agar plates containing 800 µg X-gal, 800 µg IPTG, and 0.1 µg/µL ampicillin (Invitrogen). Selected plasmids were purified from positive white colonies using the R.E.A.L. prep 96 plasmid Kit (QIAGEN). Three hundred and eighty-two clones were sequenced (automated ALF sequencer, Pharmacia) using the Thermo Sequenase Fluorescent Primer Kit (GE Healthcare).

The sequences were clustered with CAP3 (Huang & Madan 1999) and microsatellite repeats were identified using RepeatMasker (http://repeatmasker.org). Microsatellites (di- to hexanucleotides) were observed in 65.4% of the sequences and 31.29% were perfect, 57.31% imperfect and 11.40% compound. Sequences were grouped into 24 clusters containing two to 35 sequences and 71 singlets. A complete description of sequences, BLAST hits, primers and clusters can be found at bioinfo.cpqrr.fiocruz.br/microsat.

GenBank Accession numbers for the sequences are: AAT library, DQ137430 to DQ137590; CA library, DQ137591 to DQ137640; GA library, DQ137641 to DQ137722; and TAGA library, DQ137723 to DQ137796 and DQ514536. Sequences presenting the longest perfect repetitions and flanking regions were selected, and polymerase chain reaction (PCR) primers were designed using oligos version 3.6 (www.biocentre.helsinki.fi/bi/programs/fastpcr.htm).

Correspondence: G. Oliveira. Fax: +55-31-3349-7785; E-mail: oliveira@cpqrr.fiocruz.br

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd
Primer pairs for 23 loci were tested on 13 adult S. mansoni worms of the LE strain. Each PCR was performed in a 10-µL volume containing 1.5 mM MgCl₂, 50 mM KCl and 10 mM Tris/HCl, pH 8.3, 1 pmol of each primer, 200 µM of each dNTP, 0.75 U of Taq DNA polymerase (CENBIOT), and 1 ng of DNA. Thermal cycling was carried out in a PerkinElmer 9600 thermal cycler for 3 min at 95 °C; 35 cycles of 45 s at 95 °C, 1 min at primer-specific annealing temperature (Table 1), 30 s at 72 °C; and a final extension of 5 min at 72 °C. The amplicons were separated on 6% polyacrylamide 7 M urea gels using an ALF sequencer. Allele sizes were assessed using Allelinks (Amersham).

PCR products were observed in 18 of 23 loci tested and 11 were polymorphic. The polymorphic loci were further characterized on 100 S. mansoni worms from field isolates. Loci SBr12–14 were tested on 386 individuals. DNA was extracted from single worms using the phenol/chloroform extraction protocol as described in Rodrigues et al. (2002). Tests for deviations from Hardy–Weinberg equilibrium, observed ($H_O$) and expected ($H_E$) heterozygosities, and linkage disequilibrium were performed using ARLEQUIN 3.0 (Schneider et al. 2000). The loci presented two to 19 alleles per locus, and $H_E$ ranged from 0.58 to 0.90 (Table 1). Loci SBr7 and SBr8 were found to be linked. All loci, but SBr12 and SBr14, showed lower than expected $H_O$ albeit nonsignificant, except for locus SBr7 ($P < 0.05$) (Table 1).

The use of microsatellite-enriched genomic libraries generated a much higher percentage of microsatellite-containing sequences (65.4%) compared to approximately 3% found by Rodrigues et al. (2002) in expressed sequence tags (ESTs) or BAC-end sequences and 1% found by Durand et al. (2000) by screening a partial genomic library using CA and GA oligonucleotide probes. In addition, 25% of the amplified loci were shown to be polymorphic. It is noteworthy that 80% of our sequences are novel in relation to the S. mansoni ESTs, and 91 (24%) were not detected in the TIGR and SANGER genomic databases (El Sayed et al. 2004). New sequences can aid in filling the gaps in the final genome assembly. The major limitation of using microsatellite-enriched genomic libraries is the elevated price of library construction. However, the approach used yielded over 60 times more microsatellite-containing sequences, 25% of which were shown to be polymorphic. The information presented here will be useful for the generation of additional markers for genetic studies, and mapping efforts for genome gap filling.

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd
Acknowledgements

This investigation received financial assistance from WHO/TDR grant A20399, CNPq 521108/01-2, NIH-Fogarty (TW007012-01), NIH (AI042768-05A2) and FAPEMIG (407/02).

References


