Analysis of Schistosoma mansoni Population Structure Using Total Fecal Egg Sampling

Authors: Blank, Walter A., Reis, Eliana A. G., Thiong'o, Fred W., Braghiroli, Joao F., Santos, Jarbas M., et al.

Source: Journal of Parasitology, 95(4) : 881-889

Published By: American Society of Parasitologists

URL: https://doi.org/10.1645/GE-1895.1

J. Parasitol., 95(4), 2009, pp. 881–889

ANALYSIS OF *SCHISTOSOMA MANSONI* POPULATION STRUCTURE USING TOTAL FECAL EGG SAMPLING

Walter A. Blank, Eliana A. G. Reis*, Fred W. Thiong'o†‡, Joao F. Braghiroli*, Jarbas M. Santos*, Paulo R. S. Melo*, Isabel C. S. Guimarães§, Luciano K. Silva*, Theomira M. A. Carmo*, Mitermeyer G. Reis*, and Ronald E. Blanton‖
Center for Global Health and Diseases, Case Western Reserve University, Cleveland, Ohio 44106. *e-mail: reb6@case.edu*

ABSTRACT: Many parasite populations are difficult to sample because they are not uniformly distributed between several host species and are often not easily collected from the living host, thereby limiting sample size and possibly distorting the representation of the population. For the parasite *Schistosoma mansoni*, we investigated the use of eggs, in aggregate, from the stools of infected individuals as a simple and representative sample. Previously, we demonstrated that microsatellite allele frequencies can be accurately estimated from pooled DNA of cloned *S. mansoni* adults. Here, we show that genotyping of parasite populations from reproductively isolated laboratory strains can be used to identify these specific populations based on characteristic patterns of allele frequencies, as observed by polyacrylamide gel electrophoresis and automated sequencer analysis of fluorescently labeled PCR products. Microsatellites used to genotype aggregates of eggs collected from stools of infected individuals produced results consistent with the geographic distribution of the samples. Preferential amplification of smaller alleles, and stutter PCR products, had negligible effect on measurement of genetic differentiation. Direct analysis of total stool eggs can be an important approach to questions of population genetics for this parasite by increasing the sample size to thousands per infected individual and by reducing bias.

Parasite population structure may influence transmission dynamics, host preference, and virulence, and all of these factors could significantly influence control strategies and drug design. Uncontrolled, schistosomiasis is one of the most common threats to health in the developing world (King and Dangerfield-Cha, 2008). While schistosomiasis mansoni contributes to anemia and linear growth deficits (Assis et al., 2004; Brito et al., 2006), the most serious outcome of infection is the portal hypertension that results from the host's response to eggs trapped in the liver.

Schistosoma mansoni is a multicellular parasite with a complex developmental life cycle distributed between 2 hosts. Adult male and female worms produce eggs in the mesenteric veins that drain the intestines of mammalian hosts. Approximately half of these eggs pass out of the body in the stool, hatching as miracidia when they reach freshwater. The miracidia infect snails and undergo asexual reproduction, resulting in the release of cercariae that penetrate the skin of humans who come in contact with the contaminated water. Because of the way they are distributed, a community of schistosomes is, therefore, a collection of discrete infrapopulations that together form a component population in a community of human or snail hosts. The infection within an individual host is not clonal, but it is also not a complete population in the genetic sense, since the viable members of the next generation are exported. This type of distribution is likely to obscure the true population structure of *S. mansoni* unless the infrapopulations are adequately sampled.

The development of 2 elements is critical for understanding schistosome infra- and component-population structure, i.e., a set of polymorphic neutral markers and an appropriate sampling strategy. Fortunately, microsatellite markers are common in schistosomes (Rodrigues et al., 2002), and their discovery has been aided by the *S. mansoni* Genome Project. Optimizing a sampling strategy has been more problematic. The exact numbers and distribution of snails is rarely, if ever, known, and their short life spans and low rates of active shedding also mean that the snails are usually only carrying the parasites transmitted in the preceding few months by a limited segment of the human population, likely underestimating the underlying variation across all human hosts. Human activity, however, is the major determinant for the distribution of the parasite. It is simpler and more accurate to sample all, or nearly all, of the human population in a community, since most individuals can be accounted for, and their probability for infection can be accurately estimated. The most accessible and most representative parasite sample that can be obtained from humans is the eggs discharged in stool.

One approach to sampling has been to produce cercariae or adult worms in the laboratory from eggs collected from stools. This is a time-consuming process that samples relatively few organisms and introduces significant bias due to selection at the point of egg hatching, infection of laboratory snails, and infection of typical laboratory animals (LoVerde et al., 1985; Dalton et al., 1997; Sorensen et al., 2006). More recently, methods for assaying individual miracidia (obtained from eggs hatched in vitro) with multiple microsatellite markers have been described (Gower et al., 2007; Steinauer et al., 2008), but these studies are still limited in the number of individuals that may be tested within an infrapopulation, and they introduce the potential for bias toward eggs capable of hatching in vitro. By a recent estimate, a moderately infected individual with a typical parasite burden of 80–100 worm pairs would excrete 12,000–15,000 eggs daily (Wilson et al., 2006). Adequate sample size has yet to be determined, but larger numbers are more likely to be representative of the infrapopulation.

The estimation of allele frequencies by microsatellite analysis in pooled DNA samples has been performed successfully in the past (Shaw et al., 1998; Collins et al., 2000; Schnack et al., 2004). We previously demonstrated (Silva et al., 2006) that microsatellite allele frequencies can be accurately estimated from the pooled DNA of *S. mansoni* clones. We show here how the same principles may be applied to naturally aggregated samples by using thousands of parasite eggs isolated from individual human infections to study the population structure of *S. mansoni*.

MATERIALS AND METHODS

Schistosoma mansoni laboratory strains were named according to the laboratory in which they were maintained and are represented by 300–500

Received 29 September 2008; revised 12 December 2008; accepted 17 February 2009.
* Oswaldo Cruz Foundation, R. Valdemar Falcão 121, 40295-001 Salvador, Bahia, Brazil.
† University of Nairobi, P.O. Box 30197-00100, Nairobi, Kenya.
‡ Deceased.
§ Federal University of Bahia, Salvador, Bahia, Brazil.
‖ To whom correspondence should be addressed.
DOI: 10.1645/GE-1895.1

adult worms from each. The laboratory strains were: CWRU (Case Western Reserve University, Cleveland, OH); BRI-1, BRI-2, BRI-3, and BRI-4 (Biomedical Research Institute, Gaithersburg, Maryland); CPGM (Centro de Pesquisa Gonçalo Moniz – Oswaldo Cruz Foundation, Salvador, Bahia, Brazil); IPR (Institute of Primate Research, Nairobi, Kenya); TBI (Theodor Bilhartz Institute, Cairo, Egypt); and York (University of York, York, U.K.). The CWRU and BRI-1 strains were derived from the same founder population of Puerto Rican origin (Naval Medical Research Institute, NMRI strain) and have been separated for at least 30 yr. BRI-4 may have also originated from the NMRI strain, while the BRI-2, BRI-3, and the York strains are of Puerto Rican origin, but are not from the NMRI strain. All strains were maintained in mice except for the IPR strain, which represented a laboratory infection in baboons from wild-caught snails. BRI-3 (Richards and Shade, 1987) and TBI are distinguished by having a more restricted snail host range. *Schistosoma japonicum* and *S. hematobium* laboratory strains were obtained from Biomedical Research Institute (NIAID Contract N01-A1-30026). Four hookworm samples (designated HK1-HK4) were obtained from eggs isolated from the stool of individual Kenyans with hookworm infections. Clonal populations of adult worms derived from the BRI-1 strain of *S. mansoni* were created as described earlier (Silva et al., 2006).

For the isolation of *S. mansoni* eggs, total morning stools were collected from children and adults as a part of studies on schistosome morbidity in the communities of Itaquara and Salvador, Brazil, and Katheka, Kenya. In the Kenyan community and in Salvador, previous chemotherapy was uncommon, while Itaquara had undergone several campaigns of widespread treatment with praziquantel. Intensity of infection was quantified by the Kato-Katz method (Peters et al., 1980). A protocol based on published sieving methods (Dresden and Payne, 1981), and on density gradients (Baltz et al., 1982; Dalton et al., 1997) for isolating eggs from the liver, was used to isolate eggs from human feces. After homogenizing 200 g of feces in a large volume of cold 2% saline and a series of gravity sedimentations, the material was passed in succession through 420-μm mesh, 177-μm mesh, 107-μm mesh, and 50-μm mesh sieves. The eggs (~100 μm × 70 μm) were washed off of the final 2 sieves. Some of the Brazilian samples were further purified over a 10%:70% Percoll-saline step gradient centrifuged at 400 *g* for 10–20 min. After pelleting to the bottom of the tube, the eggs were washed with saline and a final egg count was made. Because little difference in the quality and quantity of DNA was noted from eggs subjected to this additional purification step, the process was discontinued for subsequent samples.

DNA was isolated from samples by proteinase K digestion followed by phenol:chloroform extraction (Sambrook and Russell, 2001). A further purification step using QIAamp tissue kit spin columns (Qiagen, Inc., Valencia, California) was performed on DNA extracted from stool egg samples to remove PCR inhibitors (Verweij et al., 2004).

Polymerase chain reaction (PCR) conditions and primer sequences for microsatellite loci SMMS 2 (*S. mansoni* microsatellite 2), SMMS 3, SMMS 13, SMMS 16, SMMS 17, SMMS 18, and SMMS 21 were as previously described (Silva et al., 2006). Primers for the human chemokine coreceptor gene CCR5 (Salkowitz et al., 2003) were used to test for human DNA contamination in DNA extracted from *S. mansoni* eggs. Primers previously used for randomly amplified polymorphic DNA (RAPD) analysis (Dias Neto et al., 1993) were used as controls for amplification. DNA samples from laboratory strains of *S. mansoni*, and from eggs isolated from feces, were genotyped by PCR amplification of the microsatellite loci and separated by polyacrylamide gel electrophoresis (PAGE) on 8% nondenaturing gels along with a 50 bp molecular weight marker (Fermentas, Glen Burnie, Maryland). Gels were stained with ethidium bromide, and digital images were generated with a Doc-It gel documentation system and software (UVP, Upland, California). To verify the identity of the presumed microsatellite amplicons from stool eggs, PCR products of some egg samples were purified from polyacrylamide gels using the QIAEX II kit (Qiagen, Inc.) and then sequenced.

We also evaluated the use of an automated sequencer for microsatellite analysis. The 11 Kenyan samples were amplified with fluorescently labeled forward primers SMMS 3, SMMS 16, SMMS 18, and SMMS 21 (using the dyes FAM, HEX, JOE, and ROX, respectively). The reaction products and the GS500LIZ size standard (Applied Biosystems, Foster City, California) were separated on an Applied Biosystems 3730 genetic analyzer. DNA samples from 15 *S. mansoni* BRI-1 clones were pooled by combining equal volumes of each, and equal volumes of the individual and pooled samples were amplified with the 4 primer sets as well as with fluorescently labeled SMMS 2 (FAM), SMMS 13 (ROX), and SMMS 17 (HEX) primers, and were analyzed using the GS600LIZ size standard (Applied Biosystems).

For PAGE-separated PCR reaction products, DNA band intensities and migration were measured from digitized gel images using SigmaGel (Jandel Scientific, San Rafael, California) as previously described (Silva et al., 2006). Relative population allele frequencies were calculated by dividing the measured intensity of a particular band (representing an allele) by the summed intensities of all bands measured at that locus. For samples run on the AB 3730, output files were analyzed using Peak Scanner software v1.0 (Applied Biosystems) using a minimum peak height cutoff of 50, and relative allele frequencies were calculated using peak areas in a manner similar to band intensities from PAGE. Calculations were made using all observed alleles, and then repeated using only the alleles representing greater than 5% of the total allele intensities in each sample. Raw allele frequencies for the individually analyzed *S. mansoni* BRI-1 clones were obtained by summing the peak areas for each allele and dividing the sum by the total area of all the alleles at that locus, and the mean ratio of allele intensities for heterozygotes at each locus was used to adjust for unequal allelic amplification, as described earlier (Le Hellard et al., 2002; Silva et al., 2006).

F_{ST} values were calculated with the program Arlequin 3.11 (Excoffier et al., 2005) using analysis of molecular variance (AMOVA), inputting the allele frequencies as relative proportions, and using worm or final egg counts for sample sizes. The analysis was carried out with 16,000 permutations. Weighted average estimators for F_{ST} values were calculated across *m* loci (Reynolds et al., 1983) using the following formula:

$$\text{Weighted } F_{ST} = \left(\sum_{l=1}^{m} Va_l\right) \Big/ \sum_{l=1}^{m} (Va_l + Vb_l)$$

where Va and Vb are, respectively, the among-population and within-population variance components determined by Arlequin at each locus *l*. Weighted averages have been shown to perform better than unweighted averages when combining F_{ST} values over multiple loci (Reynolds et al., 1983), although both were calculated for pairwise comparison of the laboratory strains. F_{ST} values are generally interpreted as follows: <0.05, little genetic differentiation; 0.05–0.15, moderate differentiation; 0.15–0.25, great differentiation; >0.25, very great differentiation (Hartl and Clark, 1997). Effective population sizes (N_e) were calculated from allele frequencies and infrapopulation sizes by a maximum likelihood method using the program MLNE (Wang and Whitlock, 2003). Other calculations were performed in Excel (Microsoft, Redmond, Washington).

RESULTS

Microsatellite amplification of the *S. mansoni* laboratory strains

DNA extracted from *S. mansoni* populations, maintained in several laboratories in the United States and abroad, was amplified using the SMMS primer sets (Fig. 1, lanes 1–9). SMMS 2 primarily amplified 2 bands of approximately 228 and 261 bp, although the IPR strain exhibited several smaller bands that more closely resembled the stepwise variation expected for microsatellites. The SMMS 3 amplification showed a greater degree of heterogeneity, as well as a greater differentiation between the strains, than did SMMS 2. By PAGE, a number of the laboratory strains appeared to be mono-allelic at several loci, yet polymorphic for the IPR strain (Fig. 1, lane 7).

Band sizes and relative intensities were measured, and pairwise F_{ST} values were calculated between strains. The results indicated a high degree of differentiation between all laboratory strains (Table I); the weighted F_{ST} for all strains, over all loci, was 0.651, with the pairwise F_{ST} values ranging from 0.289–0.955.

Specificity of the SMMS primers

The specificity of the primers for the 7 SMMS loci against other DNA templates likely to be encountered in stool samples (human,

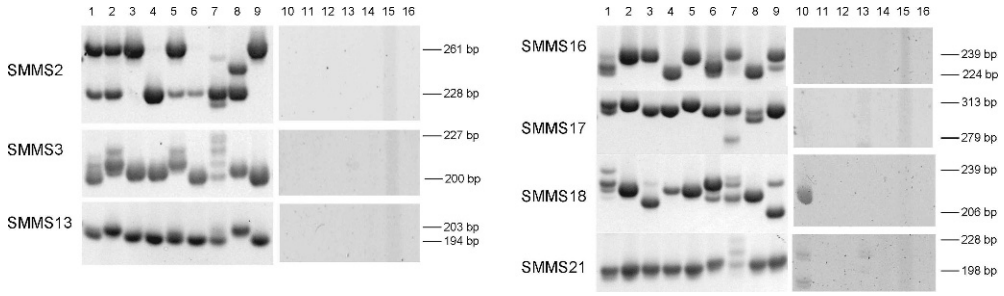

FIGURE 1. *Schistosoma mansoni* microsatellite specificity. Microsatellite DNA amplification of laboratory strains of *S. mansoni* and other organisms. DNA templates are as follows: lanes 1–9, *S. mansoni* strains CWRU, BRI-1, BRI-2, BRI-3, BRI-4, CPGM, IPR, TBI, York; lane 10, human; lanes 11–14, hookworm isolates 1–4; lane 15, *S. japonicum*; lane 16, *S. haematobium*.

hookworm, *S. japonicum*, and *S. haematobium*) were used in PCR reactions with the 7 SMMS primer sets (Fig. 1, lanes 10–16). All these samples could be amplified by RAPD primers (data not shown), indicating that the DNA was of sufficient quality and concentration for PCR. In Figure 1, the hookworm sample in lane 13 shows very weak bands for primers SMMS 3 and 21. Their intensity was much lower than that found in the *S. mansoni* PCR reactions and was generally too low to register above background levels. Amplification did not occur with DNA template from *S. japonicum* and *S. haematobium*, demonstrating the specificity of the primers for *S. mansoni*. Given that the more closely related schistosome species did not amplify with the *S. mansoni*-specific primers, it is likely that the individual from whom the hookworm eggs were isolated also harbored a low-level infection with *S. mansoni* not detected by the stool examination.

Amplified bands for the SMMS 18 and 21 primer sets in the expected size range of the microsatellite amplicons were noted for human DNA. However, PCR, using primers specific for a 312-bp fragment of the human CCR5 gene, did not amplify DNA extracted from *S. mansoni* eggs (data not shown), indicating that the levels of human DNA were low, or that methods used to isolate the eggs from fecal material were sufficient to minimize human DNA contamination.

Microsatellite amplification and analysis of Kenyan and Brazilian *S. mansoni* populations

DNA was extracted from *S. mansoni* eggs collected directly from stool samples of infected persons in Brazil and Kenya. Egg counts in stools ranged from 60 eggs per g (epg) to greater than 3,000 epg. The results of PCR amplification of these samples, using the SMMS primer sets and PAGE analysis of the products, are shown in Figure 2. The total eggs isolated from 2 individuals (Kenyan sample 1 and Brazilian sample 4) amplified poorly with most primer sets. For SMMS2, the samples from individual Brazilian infections amplified 2 alleles (228 and 261 bp) almost exclusively. This was consistent with the pattern of amplification with this marker for all of the laboratory strains originating from the Western Hemisphere. By contrast, populations from individual infections in Kenya amplified these 2 along with several additional alleles also observed in IPR, the only laboratory strain of Kenyan origin. Common allele sizes between the Kenyan and Brazilian populations were seen at all loci. Only 1 allele was observed by PAGE for loci SMMS 13 and 21 in the Brazilian samples, and in all but 1 of the Brazilian samples for locus 17. Geographic populations in Table II indicate the observed number and size of alleles amplified and the F_{ST} calculated using relative allele frequencies for each locus and egg counts of the samples. Division of the Brazilian population into its 2 component locales yielded F_{ST} values of 0.031 in Itaquara and 0.066 in Salvador. The effective population-size likelihood point estimate for the Kenyan population (N_e = 840) was approximately twice that of the Brazilians sampled (N_e = 410).

Bands corresponding to several putative alleles for each locus were purified from gels and sequenced. In each case, sequences flanking the microsatellite regions from the field samples were identical to those obtained from GenBank Release 164 (data not shown). The single exception was the 19 bp indel in SMMS 2, which was noted previously (Silva et al., 2006) in cloned parasites from the CWRU life-cycle and was also seen in the field samples.

TABLE I. Average weighted and unweighted pairwise F_{ST} values* for *S. mansoni* laboratory strains.

	Case	BRI-1	BRI-2	BRI-3	BRI-4	CPGM	IPR	TBI	York
Case	–	0.366	0.420	0.553	0.267	0.387	0.303	0.449	0.348
BRI-1	0.435	–	0.579	0.734	0.122	0.769	0.425	0.652	0.597
BRI-2	0.498	0.789	–	0.545	0.443	0.433	0.369	0.718	0.135
BRI-3	0.596	0.876	0.955	–	0.662	0.323	0.507	0.754	0.500
BRI-4	0.340	0.289	0.741	0.843	–	0.696	0.359	0.648	0.464
CPGM	0.429	0.785	0.765	0.758	0.745	–	0.361	0.660	0.361
IPR	0.327	0.537	0.523	0.642	0.486	0.425	–	0.465	0.382
TBI	0.510	0.745	0.827	0.844	0.736	0.718	0.520	–	0.671
York	0.396	0.764	0.524	0.874	0.681	0.659	0.472	0.771	–

* Unweighted values are presented above the diagonal.

Automated sequencer of total stool egg sample microsatellites

PCR products of Kenyan samples at loci SMMS 3, 16, 18, and 21, generated with fluorescently labeled primers, were also analyzed with an AB automated sequencer. This method was able to resolve a greater number of separate alleles at each locus, as compared to PAGE (Fig. 3; Table III), and was also more sensitive, providing data at locus SMMS 18 from one sample that did not produce visible bands by PAGE. Loci that appeared to be monomorphic in some samples by PAGE, e.g., SMMS 18 and SMMS 21 (Fig. 3), were observed to be polymorphic by sequencer genotyping. Additionally, as determined by the 2 methods, sizes of the PCR products were not in complete agreement. Despite this, the F_{ST} values calculated by using peak heights as a measure

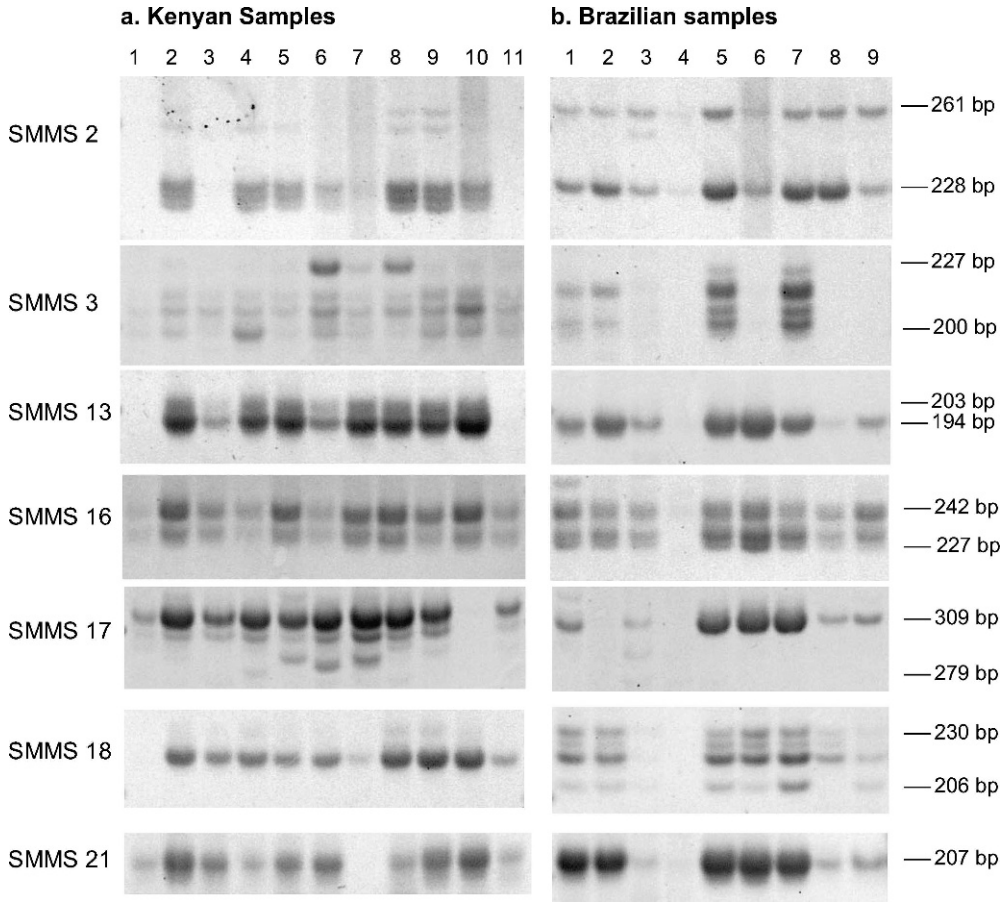

FIGURE 2. Microsatellite amplification of DNA extracted from *S. mansoni* total stool eggs from individual human infections analyzed on non-denaturing polyacrylamide gels. (**a**) Kenyan samples 1–11. (**b**) Brazilian samples 1–9.

of allele frequency (Table III) do not differ greatly from those calculated using PAGE band intensities. Calculation of F_{ST} values, disregarding minor alleles, i.e., those contributing <5% of the summed total of peak heights in any 1 sample, also correlated well with values generated using all measurable peaks. Analysis of serial dilutions of sampled DNA showed that allele frequencies were maintained (as indicated by F_{ST} values ≤0.01 between dilution pairs) to dilutions of 1:25 and, in some cases, 1:125 (data not shown).

The effect of amplification bias and microsatellite stuttering

The preferential PCR amplification of certain alleles in a mixture at some microsatellite loci is a well-known phenomenon (Daniels et al., 1998) and may potentially result in the overestimation of the frequency of certain alleles, along with a corresponding underestimation of that of other alleles. To quantify this effect, DNA was isolated from 15 clonal populations of adult worms from the BRI-1 strain of *S. mansoni* and pooled. Equal amounts of the individual clones and the pool were assayed with fluorescently labeled primers for all 7 microsatellite loci and analyzed on the automated sequencer (Figs. 4a, b). Most loci showed 2–5 alleles among the 15 clones; however, all of the clones were mono-allelic for SMMS 21, which was subsequently excluded from analysis. Allele frequencies of the individually analyzed clones, and the pool at the other 6 loci, are given in Table IV. In heterozygous clones, the smaller allele was observed to have a peak area greater than the larger allele by a factor of 1.03–1.29, depending upon the pairing of alleles (Fig. 4a). An adjusted score for the individually analyzed clones was obtained by multiplying the area of the underrepresented peaks by that factor and including the adjusted values in the summed peak areas. Pairwise F_{ST} values were generated between the pool and the summed individually analyzed clones at each locus (Table IV). Whether raw or adjusted clone allele frequencies were used, the F_{ST} between the pool and individuals was negligible, indicating that the F_{ST} measurement is robust to the slight variations in allele frequencies which may be observed when analyzing aggregated samples. Similarly, MLNE estimates using the summed (raw), summed (adjusted), and pooled clone allele frequencies with a sample size of 3,000 gave N_e point estimate values of 23, 24, and 24, respectively.

Stutter products, a consequence of template slippage during the PCR reaction, may increase the apparent number of alleles present in a population or may artificially inflate the apparent frequency of shorter alleles. Stutter peaks were readily identifiable in the genotyping trace files of the individual clones as being shorter than the true allele peaks by a single repeat subunit (Fig. 4a). The intensity of stutter peaks among the individual clones varied by locus, with the average area of stutter peaks

TABLE II. Microsatellite results summary for field egg populations.

	Kenya				Brazil				Combined	
Locus	Samples analyzed	Alleles observed	Band sizes*	F_{ST}	Samples analyzed	Alleles observed	Band sizes*	F_{ST}	Common alleles	F_{ST}
SMMS 2	7	5	220, 224, 228, 251, 261	0.004	8	3	228, 251, 261	0.126	3	0.317
SMMS 3	11	4	200, 209, 215, 227	0.043	4	6	200, 203, 209, 215, 218, 224	0.043	3	0.153
SMMS 13	9	2	194, 203	0.006	8	1	194	N.D.†	1	0.627
SMMS 16	11	2	230, 242	0.004	8	5	227, 230, 239, 242, 251	0.044	2	0.217
SMMS 17	10	6	279, 282, 285, 292, 298, 309	0.030	7	2	285, 309	0.349	2	0.369
SMMS 18	10	4	218, 224, 227, 230	0.096	7	4	206, 218, 224, 230	0.020	3	0.247
SMMS 21	10	2	204, 207	0.049	9	1	207	N.D.†	1	0.434
Weighted F_{ST}				0.029				0.057		0.240

* PCR product sizes as estimated by SigmaGel software.
† N.D., F_{ST} not determined at monoallelic loci.

ranging from 2% of the true allele peak area for SMMS 2 to 11% for SMMS 3. Raw allele frequencies for the summed individual clones were calculated using only the true allele peaks for each clone, while adjusted frequencies counted the area of each observed stutter peak toward the sum for its respective true allele. At all 6 loci, pairwise F_{ST} values of the pool, versus both raw and adjusted summed individual clones, were less than 0.01, indicating very little difference.

DISCUSSION

In this work, we demonstrate approaches for the investigation of *S. mansoni* population structure by microsatellite analysis of DNA, from all eggs isolated from the stool of a single infected individual in a single sample, as an enhanced sampling technique. Although no pooling is involved, the determination of allele frequencies in this sample type is done in a manner similar to that which we have previously successfully demonstrated with artificial pools. This is likely to better represent the true infrapopulation composition and allow for an accurate determination of allele frequencies and a subsequent calculation of the F_{ST}. Simple calculations suggest the potential magnitude of the sample size: mean stool weight may range from 200–300 g, so for a moderate infection of 100 eggs per g, there will be 20,000–30,000 eggs/stool. Final recovery of only 10% of these eggs will still yield a 2,000–3,000-egg sample from each infected individual.

We find that DNA from the total aggregate of eggs recovered from individual infections amplified as well as adult worm DNA to produce clear, interpretable patterns with microsatellite markers. Interpretability is improved when samples are genotyped using an automated sequencer as compared to PAGE. We observed a discrepancy in PCR product sizes between automated sequencer separations and PAGE (Tables II, III; Fig. 3). This is most likely due to the anomalous migration of double-stranded DNA, which has been observed in native polyacrylamide gels (Stellwagen, 1983). However, the major alleles can still be identified and measured with both techniques, and the F_{ST} values for the Kenyan samples, calculated from values based on both methods, did not differ greatly, particularly when weighted across multiple loci. Thus, good approximations can be made in resource-limited laboratories without access to automated sequencers. Redman et al. (2008) have similarly demonstrated that they were able to determine microsatellite allele frequencies by aggregate analysis of laboratory isolates of the nematode *Haemonchus contortus*, which are also maintained as mixed populations. They note that measuring frequencies of rare alleles is less accurate, although they conclude that the major alleles contribute to the bulk of the genetic information for a population.

It has been suggested that there is a low severity of error in determination of population structure measurements (including F_{ST} and migration rates) introduced by genotyping errors (Taberlet et al., 1999). This appears to be confirmed by our own results. Our analysis of pooled DNA from individual clones indicates that PCR artifacts, such as stuttering and unequal amplification, have a minimal effect on calculated F_{ST}. As a limitation of available laboratory strains, the clones we examined showed fewer alleles than would be found in field samples of *S. mansoni*. However, many of the alleles in field samples would be expected to be rare and, therefore, have little effect on F_{ST}. The tri- and tetranucleotide microsatellites used in this study displayed

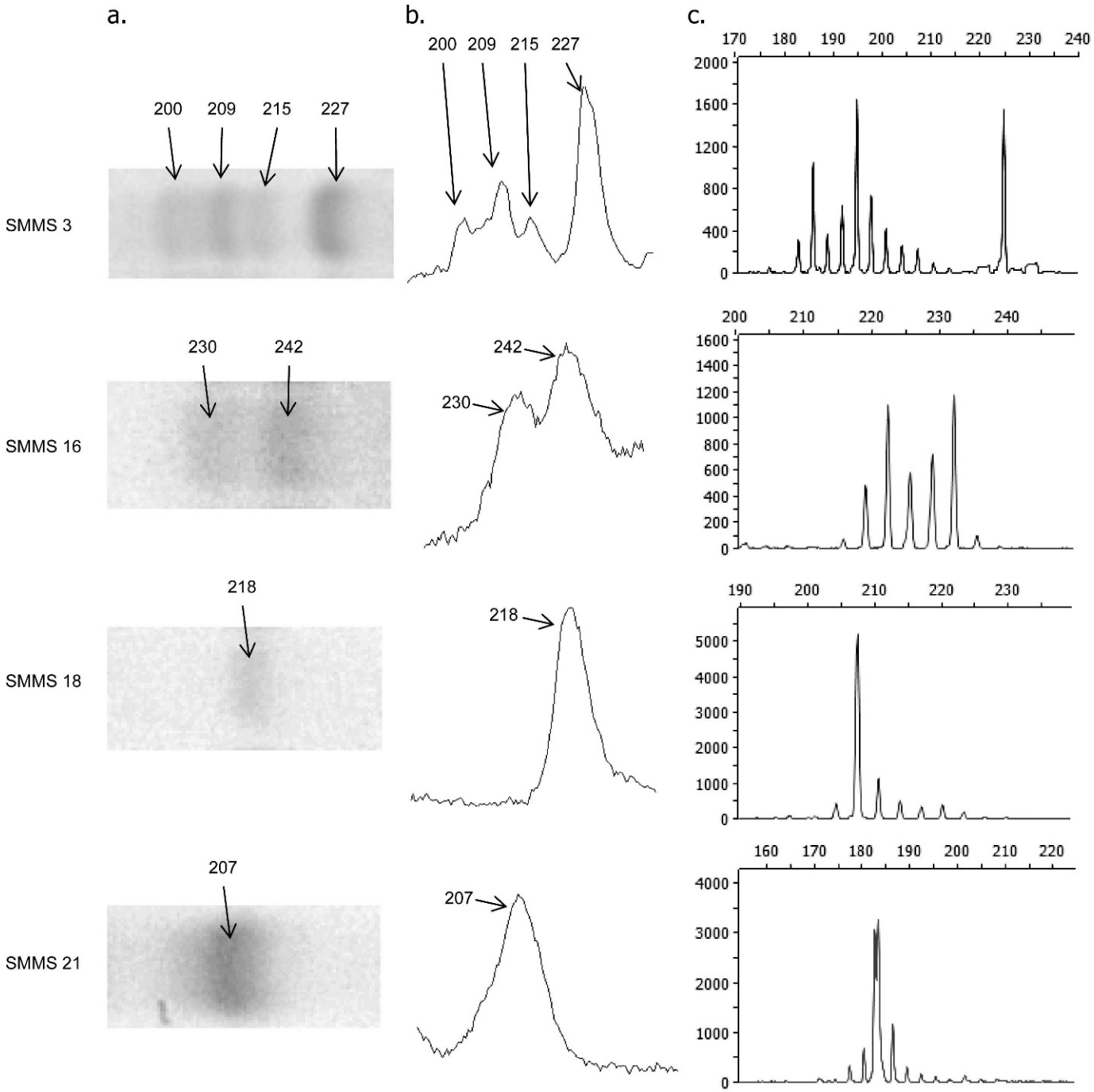

FIGURE 3. Microsatellite analysis of stool egg samples by automated sequencer analysis. (**a**) Inverted digital image of EtBr-stained polyacrylamide gel. (**b**) Band intensity peaks as interpreted by gel analysis software. Allele size in bp as calculated by SigmaGel software in (a) and (b). (**c**) Automated sequencer analysis of fluorescently labeled products. Vertical scale, fluorescence intensity; horizontal scale, size in bp.

relatively little stuttering in PCR, a finding which can be expected in comparison to dinucleotide repeats (Sheffield et al., 1995). Microsatellite loci should be adequately characterized for stuttering, and for preferential amplification of shorter alleles, prior to inclusion in studies utilizing this method.

The egg samples produced F_{ST} results consistent with geographic location and population history, but we observed that each of the laboratory strains had a unique pattern of allele frequencies, some despite having common origins. *Schistosoma mansoni* laboratory strains are closed populations of worms maintained by passage between *Biomphalaria glabrata* snails and mammalian hosts, usually mice. Given their reproductive isolation, the potential for bottlenecks, and selection in different laboratory hosts, the pairwise F_{ST} results of the laboratory strains

TABLE III. Microsatellite results summary for Kenya samples as determined by automated sequencer analysis.

		All peaks				Peaks ≥5% of total			
Locus	Samples analyzed	Alleles observed	Peak sizes	F_{ST}	Alleles observed	Peak sizes	F_{ST}
SMMS 3	11	12	183, 186, 189, 192, 195, 198, 201, 204, 206, 209, 212, 215	0.013	9	183, 186, 189, 192, 195, 198, 201, 204, 206	0.013
SMMS 16	11	9	211, 214, 217, 220, 223, 226, 229, 233, 236	0.006	7	211, 214, 217, 220, 223, 226, 229	0.006
SMMS 18	11	12	194, 197, 201, 204, 208, 211, 214, 217, 220, 224, 230, 240	0.090	6	204, 208, 211, 214, 217, 220	0.084
SMMS 21	10	10	178, 181, 184, 187, 190, 193, 196, 198, 201, 204	0.043	6	178, 181, 184, 187, 190, 193	0.043
Weighted F_{ST}				0.028			0.027

reflected that significant divergence and reduction in allelic diversity has occurred for strains derived from a common source, obscuring the original relationships among strains.

Although we effectively sampled many thousands of eggs, we examined only a small number of human infections, so it would be premature to draw firm conclusions about schistosome population structure in the locations sampled, although general statements may be made from the infrapopulation results. It is likely that Brazilian populations emanated from West Africa and, as such, might be considered a founder population representing only a fraction of the original population. Indeed, mitochondrial DNA data have been used to demonstrate how New World *S. mansoni* populations were closely related to West African isolates, and the low number of New World haplotypes was hypothesized to be due to a founder effect (Morgan et al., 2005). However, if we suppose that allelic diversity is similar in East and West Africa, decreased allelic diversity was not observed in our Brazilian samples, since there was a total of 25 alleles in the Kenyan versus 21 in the Brazilian field isolates over the 7 loci, and 15 alleles were common to both. The difference between our findings and those of Morgan et al. (2005) may be due to the nature of the mutational mechanisms in microsatellites and to differences in selection pressure between the 2 types of markers (Jarne and Lagoda, 1996). The distribution of alleles for certain markers did demonstrate strong regional patterns. SMMS 2, for example, showed primarily 2 alleles representing the indel of 19 bp in American laboratory strains as well as in the Brazilian field population, while the Kenyan samples primarily showed variation in microsatellite repeat lengths (Figs. 1, 2).

As with all methods, there are weaknesses and appropriate uses for our approach. Analysis of aggregated samples is not appropriate for certain types of population analysis (Sorensen et al., 2006; Redman et al., 2008), as heterozygosity and Hardy-Weinberg proportions cannot be measured. Allelic diversity, on the other hand, is perhaps better measured because of the greater number of individuals sampled. F_{ST}, a very effective measurement of genetic differentiation, can be calculated from allele frequency data, and estimates of effective population size (N_e), migratory fraction, and isolation by distance may also be determined (Slatkin, 1993; Wang and Whitlock, 2003).

The number of eggs appearing in the stool on any given day is highly variable, and there may be variation in the relative contribution of each worm pair to the day's output. This may not be significant over many eggs, but because of it, the effect could be minimized by collecting stools over several days. This type of variation would be a valid concern for any population sample, but an assay of a small numbers of eggs would be more susceptible to bias. Importantly, the eggs in stools represent the propagation of future generations. Some eggs will remain trapped in host tissues, but the entrapment of eggs is more a consequence of the host's immune response than of an inherited characteristic, and it is not likely to contribute significantly to continuation of the species. A degree of clonality and inbreeding is a characteristic of schistosome populations (Prugnolle et al., 2005), due to asexual reproduction in the snail host and to the large number of siblings generated in each infrapopulation. Sampling the full complement of *S. mansoni* eggs from the stools of a large number of infected humans will provide an indirect, yet powerful, assessment of the composite worm population.

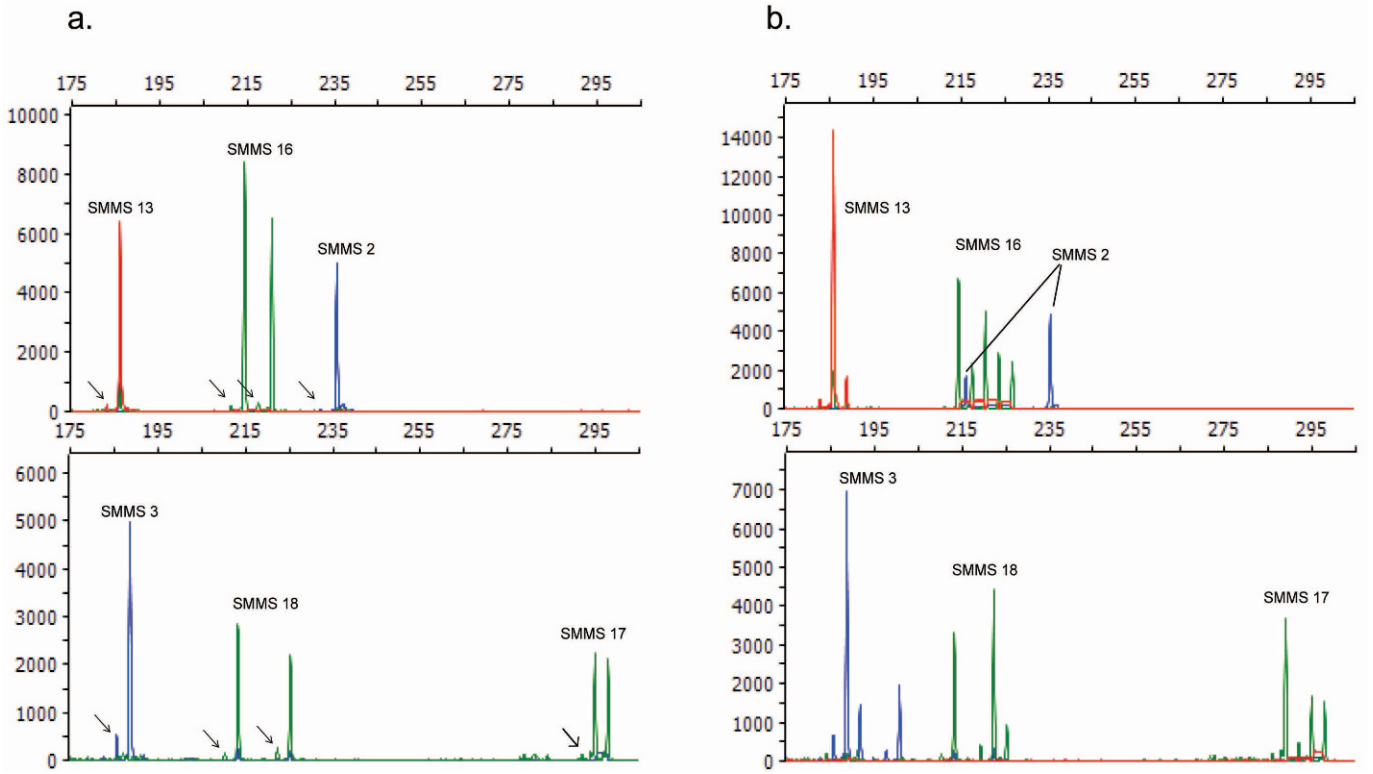

FIGURE 4. Microsatellite analysis of a single *S. mansoni* clone (**a**) and pooled clones (**b**) by automated sequencer analysis. Arrows indicate locations of stutter peaks. Vertical scale, fluorescence intensity; horizontal scale, size in bp.

TABLE IV. Allele frequencies with and without adjustment for unequal amplification: comparison of F_{ST}.

		Measured allele frequencies*			Pairwise F_{ST} of pool vs.	
Locus	Allele size	Individuals (raw)	Individuals (adjusted)†	Pool (raw)	Individuals (raw)	Individuals (adjusted)
SMMS2	216	0.232	0.226	0.276	0.002	0.003
	235	0.768	0.774	0.724		
SMMS3	189	0.664	0.637	0.675		
	192	0.119	0.130	0.135	<0.001	<0.001
	201	0.217	0.234	0.190		
SMMS13	186	0.896	0.870	0.902	<0.001	0.002
	189	0.104	0.130	0.098		
SMMS16	215	0.388	0.359	0.345		
	218	0.115	0.107	0.123		
	221	0.257	0.295	0.258	<0.001	<0.001
	224	0.129	0.120	0.151		
	227	0.111	0.120	0.123		
SMMS17	289	0.522	0.482	0.502	<0.001	<0.001
	292	0.044	0.040	0.063		
	295	0.216	0.225	0.230		
	298	0.218	0.253	0.205		
SMMS18	213	0.379	0.331	0.389		
	222	0.510	0.544	0.506	<0.001	0.001
	225	0.111	0.125	0.105		

* Allele frequencies and sizes for 15 *S. mansoni* BRI-1 individual and pooled clones from peak heights, as determined by automated sequencer analysis.
† Adjustment for unequal allelic amplification, as described (Le Hellard et al., 2002; Silva et al., 2006).

ACKNOWLEDGMENTS

We wish to thank Fred Lewis at Biomedical Research Institute, and Alan Wilson at York University, for samples and information on strains. This work was supported by NIH grant AI41680 and by CNPq grants 304713/2002-3 and 479839/2007-7.

LITERATURE CITED

Assis, A. M. O., M. S. Prado, M. L. Barreto, M. G. Reis, S. M. C. Pinheiro, I. M. Parraga, and R. E. Blanton. 2004. Childhood stunting in Northeast Brazil: The role of *Schistosoma mansoni* infection and inadequate dietary intake. European Journal of Clinical Nutrition **58:** 1022–1029.

Baltz, T., I. Lacassie, J. Tribouley-Duret, and J. Tribouley. 1982. Density-gradient separation of *Schistosoma mansoni* eggs. Journal of Parasitology **68:** 963–965.

Brito, L. L., M. L. Barreto, and R. S. De Cassia. 2006. Moderate- and low-intensity co-infections by intestinal helminths and *Schitosoma mansoni*, dietary iron intake, and anemia in Brazilian children. The American Journal of Tropical Medicine and Hygiene **75:** 939.

Collins, H. E., H. Li, S. E. Inda, J. Anderson, K. Laiho, J. Tuomilehto, and M. F. Seldin. 2000. A simple and accurate method for determination of microsatellite total allele content differences between DNA pools. Human Genetics **106:** 218–226.

Dalton, J. P., S. R. Day, A. C. Drew, and P. J. Brindley. 1997. A method for the isolation of schistosome eggs and miracidia free of contaminating host tissues. Parasitology **115:** 29–35.

Daniels, J., P. Holmans, N. Williams, D. Turic, P. McGuffin, R. Plomin, and M. J. Owen. 1998. A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. The American Journal of Human Genetics **62:** 1189–1197.

Dias Neto, E., C. P. de Souza, D. Rollinson, N. Katz, S. D. Pena, and A. J. Simpson. 1993. The random amplification of polymorphic DNA allows the identification of strains and species of schistosome. Molecular and Biochemical Parasitology **57:** 83–88.

Dresden M. H., and D. C. Payne. 1981. A sieving method for the collection of schistosome eggs from mouse intestines. Journal of Parasitology **67:** 450–452.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online **1:** 47–50.

Gower, C., J. Shrivastava, P. Lamberton, D. Rollinson, B. Webster, A. Emery, N. Kabatereine, and J. Webster. 2007. Development and application of an ethically and epidemiologically advantageous assay for the multi-locus microsatellite analysis of *Schistosoma mansoni*. Parasitology **134:** 523–536.

Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics. Sinauer Associates, Sunderland, Massachusetts, 519 p.

Jarne, P., and P. Lagoda. 1996. Microsatellites, from molecules to populations and back. Trends in Ecology & Evolution **11:** 424–429.

King, C. H., and M. Dangerfield-Cha. 2008. The unacknowledged impact of chronic schistosomiasis. Chronic Illness **4:** 65–79.

Le Hellard, S., S. J. Ballereau, P. M. Visscher, H. S. Torrance, J. Pinson, S. W. Morris, M. L. Thomson, C. A. M. Semple, W. J. Muir, D. H. R. Blackwood et al. 2002. SNP genotyping on pooled DNAs: Comparison of genotyping technologies and a semi automated method for data storage and analysis. Nucleic Acids Research **30:** e74.

Loverde, P. T., J. Dewald, D. J. Minchella, S. C. Bosshardt, and R. T. Damian. 1985. Evidence for host-induced selection in *Schistosoma mansoni*. Journal of Parasitology **71:** 297–301.

Morgan, J. A. T., R. J. Dejong, G. O. Adeoye, E. D. O. Ansa, C. S. Barbosa, P. Bremond, I. M. Cesari, N. Charbonnel, L. R. Correa, and G. Coulibaly. 2005. Origin and diversification of the human parasite *Schistosoma mansoni*. Molecular Ecology **14:** 3889–3902.

Peters, P. A., M. El Alamy, K. S. Warren, and A. A. Mahmoud. 1980. Quick Kato smear for field quantitation of *Schistosoma mansoni* eggs. American Journal of Tropical Medicine and Hygiene **29:** 217–219.

Prugnolle, F., D. Roze, A. Theron, and T. De Meeus. 2005. F-statistics under alternation of sexual and asexual reproduction: A model and data from schistosomes (platyhelminth parasites). Molecular Ecology **14:** 1355–1365.

Redman, E., E. Packard, V. Grillo, J. Smith, F. Jackson, and J. S. Gilleard. 2008. Microsatellite analysis reveals marked genetic differentiation between *Haemonchus contortus* laboratory isolates and provides a rapid system of genetic fingerprinting. International Journal for Parasitology **38:** 111–122.

Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics **105:** 767–779.

Richards, C. S., and P. C. Shade. 1987. The genetic variation of compatibility in *Biomphalaria glabrata* and *Schistosoma mansoni*. Journal of Parasitology **73:** 1146–1151.

Rodrigues, N. B., P. T. Loverde, A. J. Romanha, and G. Oliveira. 2002. Characterization of new Schistosoma mansoni microsatellite loci in sequences obtained from public DNA databases and microsatellite enriched genomic libraries. Memórias do Instituto Oswaldo Cruz **97:** S1 71–75.

Salkowitz, J. R., S. E. Bruse, H. Meyerson, H. Valdez, D. E. Mosier, C. V. Harding, P. A. Zimmerman, and M. M. Lederman. 2003. CCR5 promoter polymorphism determines macrophage CCR5 density and magnitude of HIV-1 propagation in vitro. Clinical Immunology **108:** 234–240.

Sambrook, J., and D. W. Russell. 2001. Molecular cloning: A laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, New York, 999 p.

Schnack, H. G., S. C. Bakker, R. van't Slot, B. M. Groot, R. J. Sinke, R. S. Kahn, and P. L. Pearson. 2004. Accurate determination of microsatellite allele frequencies in pooled DNA samples. European Journal of Human Genetics **12:** 925–934.

Shaw, S., M. Carrasquillo, C. Kashuk, E. Puffenberger, and A. Chakravarti. 1998. Allele frequency distribution in pooled DNA samples: Applications to mapping complex disease genes. Genome Research **8:** 111–123.

Sheffield, V. C., J. L. Weber, K. H. Buetow, J. C. Murray, D. A. Even, K. Wiles, J. M. Gastier, J. C. Pulido, C. Yandava, and S. L. Sunden. 1995. A collection of tri-and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Human Molecular Genetics **4:** 1837–1844.

Silva, L. K., S. Liu, and R. E. Blanton. 2006. Microsatellite analysis of pooled *Schistosoma mansoni* DNA: An approach for studies of parasite populations. Parasitology **132:** 331–338.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution **47:** 264–279.

Sorensen, R. E., N. B. Rodrigues, G. Oliveira, A. J. Romanha, and D. J. Minchella. 2006. Genetic filtering and optimal sampling of *Schistosoma mansoni* populations. Parasitology **133:** 443–451.

Steinauer, M. L., L. E. Agola, I. N. Mwangi, G. M. Mkoji, and E. S. Loker. 2008. Molecular epidemiology of *Schistosoma mansoni*: A robust, high-throughput method to assess multiple microsatellite markers from individual miracidia. Infection, Genetics, and Evolution **8:** 68–73.

Stellwagen, N. C. 1983. Anomalous electrophoresis of deoxyribonucleic acid restriction fragments on polyacrylamide gels. Biochemistry **22:** 6186–6193.

Taberlet, P., L. P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling: Look before you leap. Trends in Ecology & Evolution **14:** 323–327.

Verweij, J. J., R. A. Blange, K. Templeton, J. Schinkel, E. A. Brienen, M. A. van Rooyen, L. van Lieshout, and A. M. Polderman. 2004. Simultaneous detection of *Entamoeba histolytica*, *Giardia lamblia*, and *Cryptosporidium parvum* in fecal samples by using multiplex real-time PCR. Journal of Clinical Microbiology **42:** 1220–1223.

Wang, J., and M. C. Whitlock. 2003. Estimating effective population size and migration rates from genetic samples over space and time. Genetics **163:** 429–446.

Wilson, R. A., G. J. Van Dam, T. M. Kariuki, I. O. Farah, A. M. Deelder, and P. S. Coulson. 2006. The detection limits for estimates of infection intensity in schistosomiasis mansoni established by a study in non-human primates. International Journal for Parasitology **36:** 1241–1244.