Published in final edited form as:

Int J Parasitol. 2008 December ; 38(14): 1681–1691. doi:10.1016/j.ijpara.2008.05.015.

Multiple near-identical genotypes of *Schistosoma japonicum* can occur in snails and have implications for population-genetic analyses

Mingbo Yin^a, Wei Hu^a, Xiaojin Mo^a, Shengyue Wang^b, Paul J. Brindley^c, Donald P. McManus^d, George M. Davis^c, Zheng Feng^{a,1}, and David Blair^{e,*}

^aNational Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, 207 Rui Jin Rd II, Shanghai 200025, P.R. China

^bChinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, P.R. China

^cDepartment of Microbiology, Immunology & Tropical Medicine, George Washington University Medical Center, Ross Hall, 2300 Eye Street, NW, Washington DC 20037 USA

^dQueensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4029, Australia

^eSchool of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia

Abstract

We genotyped (using 16 or 17 microsatellite loci) numerous adult *Schistosoma japonicum* raised in rabbits exposed to pooled cercariae from small numbers of naturally infected snails from several localities in China. As expected, duplicate multi-locus genotypes (MLGs) were found among these worms. Additionally, many more MLGs, often near-identical, were found than snails used as sources of cercariae. Explanations for these results include i) genotyping errors, ii) development within each infected snail of multiple sibling miracidia and iii) somatic mutation producing genetically varied cercariae from a single miracidium. To control for genotyping errors we re-analysed samples from many individual worms, including repeating the initial PCR. Explanations invoking the development of multiple sibling miracidia within a single snail are not likely to be correct because almost all duplicate MLGs fell within same-sex clusters in a principal coordinates analysis. We would expect both sexes to be represented in a multi-miracidium infection. In addition, we exposed several snails to infection by a single miracidium. One such snail, via an experimentally infected mouse, yielded 48 adult worms. The presence of at least nine near-identical MLGs among these worms was confirmed by re-genotyping. We regard somatic mutation as the most likely explanation for our results. The implications of multiple MLGs for population-genetic studies in *S. japonicum* are discussed.

Keywords

China; Clonal replication; Genotyping; Microsatellites; Multi-locus genotype; *Oncomelania*; *Schistosoma japonicum*; Somatic mutation

*Corresponding author. Tel.: +61 7 4781 4322; fax: +61 7 4725 1570. E-mail address: E-mail: david.blair@jcu.edu.au.

¹Author for communication concerning the China Tropical Medicine Research Center

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The Asian bloodfluke, *Schistosoma japonicum*, causes serious human disease in the Philippines, parts of Indonesia, and in particular in China where more than 30 million people living in the tropical and sub-tropical regions are at risk (Uttinger et al., 2005; Hao et al., 2006). Adult, sexually reproducing worms live in the blood vessels of several mammal species. Their eggs hatch in fresh water yielding free-swimming miracidia capable of penetrating a snail (*Oncomelania hupensis* in this case) within the haemocoel of which cycles of asexual replication produce many cercariae. These swimming infective stages emerge into the water to seek a mammalian host. It is generally assumed that the hundreds or thousands of cercariae potentially derived from a single miracidium will be genetically identical (e.g. Brouwer et al., 2001).

The importance of this parasite is such that the Chinese National Human Genome Center at Shanghai has sequenced the entire genome of *S. japonicum*. This project is almost finished and publication is imminent. Population-genetic approaches will provide insights into any geographic and host structuring that this schistosome may exhibit (e.g. Shrivastava et al., 2005). To this end, we have used the genome resource to identify polymorphic microsatellite loci in *S. japonicum*.

For the study reported here, snails were collected at seven sites in the Yangtze Basin with the intention of genotyping any *S. japonicum* infections present in those. Multilocus genotyping of individual miracidia or cercariae of *S. japonicum* is a challenge yet to be met because of the small sizes of these infective stages (less than one millimetre in length). Consequently, study material was obtained by infecting rabbits with pooled cercariae and harvesting the much larger adult worms. Inspection of the data revealed duplicate multi-locus genotypes (MLGs) in most populations, likely a consequence of clonal replication within snails. Many additional MLGs differed from each other only at one or two alleles across the 17 loci and formed same-sex clusters in principal coordinates analyses. Investigations leading to these findings and explanations for them are presented below. We also discuss some of the implications for population-genetic investigation of *S. japonicum*.

2. Materials and methods

2.1. Sources of schistosomes

Adult worms for the population study were obtained as follows. Snails (*O. hupensis*) infected with *S. japonicum* were collected from a total of seven field sites in five Provinces where the disease is endemic, Jiangxi, Anhui, Hunan, Hubei and Sichuan. Geographic sites from which snails were collected are listed in Table 1 and shown in Fig. 1. Cercariae, to a total of 1,000, were pooled from all infected snails from a single site and used to infect naïve laboratory-raised rabbits from which adult schistosomes were harvested 42 days later. For convenience, we use the term “population” to represent all adult worms derived from cercariae from a single site. The studies were approved by the Institutional Animal Care and Use Committee of the National Institute of Parasitic Diseases, Shanghai.

Snails (*O. hupensis*) were bred and maintained in the laboratories of the Centers for Disease Control, Institute for Parasitic Diseases, in Shanghai. Ninety-eight uninfected snails were each exposed to a single miracidium of the Anhui strain maintained at the National Institute of Parasitic Diseases, Shanghai.

2.2. Genetic markers

A near-complete draft genome (seven-fold coverage) of *S. japonicum* is available as a number of super-contigs at <http://lifecenter.sgsc.cn/sj.do>. We searched this resource for microsatellite

loci containing simple 3-mer repeats (in fact, only TAA repeats were sought and used) flanked by single-copy sequences to which PCR primers could be targeted. To minimise problems due to linkage disequilibrium (LD), no more than one locus was chosen from any super-contig. Ideally, microsatellite loci should be selectively neutral. This might not be the case if they lie within, or close to, a gene. We therefore only chose loci that were not close to known or predicted genes.

Seventeen loci matching these criteria, and consistently yielding PCR products of the anticipated lengths, were eventually used. PCR conditions for each locus were optimised using individual adult worms from several Provinces of China (Table 2).

2.3. DNA extraction, PCR and genotyping

Genomic DNA was extracted from individual worms (48 males and 48 females from each population) using a proteinase K digestion method. After inactivation of proteinase K by incubation at 95°C for 15 min, and centrifugation to pellet debris, the soluble lysate was used directly as the source of template DNA for PCR.

All PCRs were carried out on a GeneAmp PCR System 9700 thermal cycler (ABI). Amplifications were performed in 10 µL reactions containing 10 pm of each primer (labelled with carboxyfluorescein [FAM]), ~25 ng of genomic DNA from a single worm, SBS *Taq* polymerase (2 U), 1.25 mM MgCl₂, 1 µL 10 × reaction buffer and 0.5 uL dNTPs (2.5 mM, TaKaRa). Cycling parameters were: one cycle at 95°C for 5 min; 30 cycles at 94°C for 45 s, 55–65°C for 45 s (depending on locus – see Table 2), 70°C for 45 s; followed by a final cycle at 72°C for 10 min. PCR products were separated using an ABI 3730 XL automated DNA sequencer with ABI GS500 LIZ internal size standards. Results were read in *GENEMAPPER* 4.0 software (Applied Biosystems).

The PCR products of selected loci from the single-miracidium experiment were cloned into the pMD20-T vector (Takara, Japan), after which competent cells were transformed with the ligation products. Recombinant plasmids were isolated from colonies of the transformed cells and the nucleotide sequence of the inserts determined.

2.4. Confirmation of findings from field populations

Although numerous identical MLGs were often identified in worms from a single population, presumably all derived from the same infected snail, some worms had genotypes that differed at one or a very few loci from a frequently observed MLG (near-identical MLG: niMLG). This would not be expected if all these worms had arisen from a single miracidium via clonal replication in the snail. Three classes of explanations were entertained: genotyping error, multiple infections by sibling miracidia within a single snail, and the possibility of somatic mutation arising during cycles of asexual reproduction within the snail. As discussed in Pompanon et al. (2005), there are several sources of genotyping error. These include human error, and PCR errors and artefacts including allelic dropout and machine errors. Any such source of error is of great concern; a low per-locus error rate can lead to a very high rate of error per-genotype for multilocus genotypes (Pompanon et al. 2005).

All traces were re-examined, and those containing data from individuals and loci of interest were scrutinised thoroughly. We regarded any allele as different if it was assigned by the machine a value differing by 2 bp or more from another. Alleles identical in length are not expected to be assigned lengths differing by as much as 1 bp in an ABI 3730 machine. Traces for which we entertained doubts were removed from the analysis. In some cases genotyping, including the initial PCR, was repeated.

2.5. Experimental single miracidium infection of snails

Four of 98 uninfected snails, each exposed to a single miracidium, were subsequently found to be infected. Two of these snails survived to shed cercariae. Pooled cercariae from one of the snails were used to infect a mouse, from which 48 adult worms recovered 42 days p.i. were subsequently genotyped. Checking of data for genotyping errors followed methods outlined in section 2.4. In addition, some PCR products were cloned and sequenced, as above.

2.6. Data analysis

All worms for which one or more loci could not be scored were omitted from further analysis. In addition, amplification of locus Sjp7 often failed in two populations (Changde and Guichi). This locus was therefore omitted from analyses in which data from all populations were combined.

Analyses fell into two categories. One was the exploration of the possible origins of duplicate genotypes and the relationships among similar MLGs. The other included many of the conventional population-genetic analyses and tests that might be applied to organisms such as schistosomes: tests of LD, tests for panmixia, calculation of fixation indices, partitioning of genetic variance and assignment tests.

The number of worms fully genotyped always exceeded the number of infected snails from each site, as did the number of unique MLGs in all but one case (Table 1). Multiple worms with identical MLGs were often found. The program GENCLONE (Arnaud-Haond and Belkhir, 2007) was used to test the hypothesis that such worms had developed from clonally derived sibling cercariae. In each case we could confidently reject the alternative hypothesis that identical MLGs had arisen independently as a consequence of random assortment of alleles via sexual reproduction. For some analyses reported below, duplicate MLGs were therefore removed, leaving a single representative of each. Relationships of individual genotypes to one another were depicted graphically using principal coordinates analyses (PCA) in GENALEX (Peakall and Smouse, 2006). A pairwise co-dominant genotypic distance matrix for individuals within each population was calculated and a PCA plot constructed directly from the distance matrix with data standardisation as explained in the GENALEX documentation.

Linkage (genotypic) disequilibrium, measuring the degree of association between genotypes at pairs of loci in a population, was calculated using FSTAT v2.9.3.2 (Goudet, 2001, FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)), available from <http://www2.unil.ch/popgen/softwares/fstat.htm>. A nominal 5% level of significance was chosen, which dictates the number of permutations used in significance testing by FSTAT. The program applies a Bonferroni correction for multiple tests. Genotypes at linked loci are frequently inherited together because of the rarity of chiasma formation between them. It became obvious that the apparent high degree of linkage between loci in *S. japonicum* (Table 3) would render our data unsuitable for conventional population-genetic analyses, virtually all of which assume little or no LD. However, we did undertake some such analyses to highlight properties of the data and to demonstrate the likely effects of the phenomena we are reporting here. Male and female worms were treated separately in some analyses (i.e. 14 populations instead of seven). Various approaches were used to calculate whether genotype frequencies within each population departed from those expected under panmixia. Tests for significant departures from Hardy-Weinberg Equilibrium (HWE) were performed in GENALEX. The direction of departure from HWE (excess or deficit of heterozygotes) becomes apparent by visual inspection of tabulated observed and expected heterozygosity (calculated in ARLEQUIN; Excoffier et al., 2005). Also useful is the F_{IS} statistic which measures non-random mating of individuals relative to their population. F_{IS} can range in value from -1 to 1; negative values indicate an excess of heterozygotes whereas positive values indicate the opposite. A value of

zero implies that the population is in HWE. Calculations of F_{IS} were made using FSTAT. Significance (set at 5% level) was tested by permuting alleles among individuals within samples. Another F -statistic, F_{ST} , was calculated in GENALEX in an Analysis of Molecular Variance (AMOVA) framework which also allows for significance testing by random permutation (9,999 permutations were used). Input into the AMOVA analysis was a co-dominant allelic distance matrix. Populations were not assigned to regional groups. The F_{ST} statistic measures the proportion of the genetic variance among populations relative to the total variance, giving an indication of the degree of differentiation between populations. GENALEX provides a tool which uses allele frequencies in known populations as the basis for assigning a “new” individual to a population. Males and females from each population were treated as separate “populations” and the program allowed to assign a gender to each worm. Under panmixia, gender assignment should have little accuracy and F_{ST} and F_{IS} values should be close to zero and non-significant.

3. Results

3.1. Checking for errors

All data were carefully checked and re-checked on several occasions. For example, among the 72 fully genotyped worms from Changde (16 loci), there were 55 unique MLGs, 34 of these from female worms (Table 1). MLGs of 27 of the female worms were very similar to one another, alleles differing at no more than one to three loci. Fig. 2 shows the traces for a single variable locus (Sjp4) from several of these worms. This was the most variable locus and was re-genotyped, yielding the same results. Three worms from Changde were also re-genotyped for locus Sjp12, yielding the same variants seen initially.

3.2. Results of single-miracidium experiments

Among the 48 worms produced from one miracidium, initial genotyping (14 loci) yielded 12 MLGs, one occurring in 37 individuals and the remainder singletons. All traces were carefully inspected, followed by re-amplification of variant alleles in unique MLGs. This led to the removal of three singleton MLGs. In one, an unexplained error had occurred at one locus in the first genotyping (confirmed by sequencing of cloned PCR products). In another, an error in data processing had led to an allele length being incorrectly recorded. The third case was due to apparent allelic dropout, as discussed by Pompanon et al. (2005), at one locus. We are confident in our recognition of nine MLGs arising from a single miracidium and differing from one another at a single locus in most cases.

3.3. Clustering of multi-locus genotypes in the population study

Even after checking and removal of incomplete and duplicate MLGs, the numbers of MLGs found in six of the seven populations far exceeded the number of snails used to found that population (Table 1). Many of the MLGs were very similar to one another such that they cluster together in a PCA and, with three exceptions, all individuals in each cluster belonged to the same sex. In the Shashi population, two females and seven males shared one MLG, and one male and one female shared another. In the Changde population, three males and one female shared a MLG. One representative MLG from worms of each sex was retained for analysis in these cases. We cannot say whether the sharing of MLGs between sexes was a consequence of the gender of some worms being mis-identified. The worms had been completely digested during the DNA extraction process so that gender could not be re-confirmed. Fig. 3A shows the PCA of worms from the Guichi population (founded using cercariae from seven snails). The 58 worms (45 unique MLGs) fall into about eight clusters, each consisting of worms of one sex. A similar pattern was seen with worms from Changde (Fig. 3B), but strikingly different results were seen in the Duchang population (Fig. 3C); 62 worms yielded 61 unique genotypes

with typically high pairwise genotypic distances. The PCA in Fig. 3C shows little evidence of clustering of genotypes.

3.4. Population-genetic analyses

All loci deviated significantly from Hardy-Weinberg expectations when sex of worms was disregarded (i.e. seven populations). When populations were subdivided by sex (14 populations), a small number of loci did not deviate from HWE. However, the great majority of loci in all the populations remained out of HWE.

Table 3 and Table 4 present results of further tests, with and without duplicate genotypes. Most striking is the extent to which pairs of loci appear to be in LD. This was the phenomenon that first attracted our attention when we started these analyses. Whether sex was disregarded or not, most pairs of loci were in apparent LD, with values highest for Xichang, Shashi and Tongling. Values were lowest for the Duchang population. Observed heterozygosity was substantial in all populations and can be lower or higher than expected values. There also appeared to be no strong pattern to F_{IS} values; these could be negative (implying a surplus of heterozygotes) or positive. F_{IS} values significantly differed from random expectations in many cases.

F_{ST} values similarly show strong population pairwise differentiation (Table 5 and Table 6). When each population was analysed without regard to sex, all pairwise differences were significant. When sex was taken into account (14 populations), a few pairs of populations were not significantly differentiated.

As shown in Table 1, assignment tests in GENALEX usually had considerable success in assigning worms to the correct gender.

4. Discussion

In each sample of worms obtained by exposing rabbits to pooled cercariae from snails from a particular locality we generally detected much larger numbers of MLGs than of snails used as a source of cercariae. However, many of these MLGs were near-identical, raising the possibility that the slight variants observed were a consequence of genotyping error.

We had advantages over many studies in that we started with high-quality DNA, which is less prone to generation of artefacts (Pompanon et al., 2005), and we knew exactly which individuals and loci to re-check under the informal null hypothesis that those deviating from the state seen in common MLGs were in error. In almost all cases, we refuted this hypothesis. All suspect genotypes were checked by re-examination of trace images and in some cases by re-genotyping. Only genotypes that had passed all tests were included in analyses. All variant genotypes found in the single miracidium experiment were re-genotyped. We were thus able to account for human error and machine artefact.

Another obvious explanation for clusters of niMLGs is that snails had been infected by multiple miracidia and that these may have originated from eggs from the same or related adult pairs in the mammal host. Presence of multiple genotypes of *S. mansoni* within a single snail (*Biomphalaria glabrata*) has been reported by Minchella et al. (1995), Sire et al. (1999), Eppert et al. (2002) and Theron et al. (2004). For the *Schistosoma haematobium* / *Bulinus* spp. system, multiple infections were reported by Dabo et al. (1997) and Davies et al. (1999). Typically, over 50% of infected snails contained multiple genotypes, the maximum reported being nine. All these studies used restriction fragment length polymorphism or random amplified polymorphic DNA approaches, which do not allow easy recognition of closely-related

individuals versus unrelated individuals. All have implicitly assumed that somatic mutation does not occur and that the presence of different genotypes indicates multiple infections.

Mixed-sex infections have been noted in studies in which infrapopulations of *Schistosoma* spp. in individual snails have been genotyped (e.g. Minchella et al., 1995; Dabo et al., 1997; Eppert et al., 2002). If a snail is infected by multiple miracidia originating from a single pair of adults, then clusters of related MLGs might be expected in a PCA. However, random segregation of the sex chromosomes (male schistosomes are homogametic – ZZ and females heterogametic – ZW) should produce both males and females in these clusters. This was rarely observed. In almost all cases, all worms within a cluster were of the same sex. In three cases, worms of both sexes were found in a cluster. The population study involved nearly 700 worms; occasional mis-assignment of gender is certainly possible.

The remaining explanation for niMLGs is that somatic mutations occurred during the phases of life between production of the egg/miracidium and the adult worm, the latter being the stage we studied and genotyped. For such mutations to become fixed within different individuals, they must have occurred during the asexual replication phase within the snail.

Generation of genetic variation in schistosomes within the snail host has been suspected by some researchers. Bayne and Grevelding (2003) detected heterogeneity among sporocysts of *Schistosoma mansoni* cultured in vitro. Working with the same species, Gower et al. (2007) commented on the possibility of mutation occurring in the snail host and said that they had noticed numbers of closely related MLGs suggestive of this in their study. They did not provide any further data or analysis. Semyenova et al. (2005) found some evidence for generation of heterogeneity in another schistosome genus, *Trichobilharzia*. We know of no other explicit reports of somatic mutation in trematodes.

We favour somatic mutations as the explanation for the variant MLGs encountered. Nevertheless, caution prompts us not to completely reject alternatives. It is clear that we are not alone in finding many more MLGs than snails (Shrivastava et al 2005; Gower et al 2007). Consequently, the discussion that follows is relevant for any population-genetic studies in which adult worms are raised from infected snails. The presence of duplicate genotypes and of clusters of niMLGs should lead to deviation from HWE and the appearance of substantial LD. We found both to be the case. High levels of LD, in particular, will confound most forms of conventional population-genetic analysis. Marked deviation from HWE and high levels of LD have not apparently been a problem in studies on *S. mansoni* (e.g. Thiele et al., 2008). However, the only previous study on *S. japonicum* utilising microsatellites (Shrivastava et al., 2005) did record significant deviations from HWE and substantial (but not quantified) LD. They attributed both to the effects of non-random mating, inbreeding and population subdivision. The deficit of heterozygotes they observed was in contrast with the lack of a strong pattern across populations in our study when duplicate genotypes were included in analyses (Table 3). However, when duplicate genotypes were removed (Table 4) most of our populations showed a deficit of heterozygotes. Recent studies (Balloux et al., 2003; Prugnolle et al., 2005; Halkett et al., 2005) have suggested a strongly negative value for F_{IS} is to be expected in populations with high levels of clonal reproduction. We only observed negative F_{IS} values in a few populations, whether duplicates were included or not (Table 3 and Table 4).

Until we have a large sample of unrelated MLGs from a single population, we cannot determine how many of our loci are really in LD. It remains possible that LD is a general phenomenon in *S. japonicum*. Hirai et al. (1996) found very low levels of chiasma formation in *S. japonicum* relative to *S. mansoni* (3 versus 15.3 chiasmata per cell). Consequently, large tracts of genes might appear to be linked in the former but not the latter species. Note that there are only seven pairs of somatic chromosomes, several of them rather small, in *Schistosoma* and

two rather large sex chromosomes (Hirai et al., 2000). The number of possible linkage groups is thus rather small.

The presence of clusters of duplicate MLGs and niMLGs in any population will tend to decrease similarity between that population and others, inflating estimates of population differentiation. The same argument applies to the separate sexes. Generation of heterogeneity in small samples without random segregation of the sexes will tend to produce individual-specific patterns that can be conflated with sex-specific patterns. This is what we saw in our data; high and significant F_{ST} values implying near-universal rejection of the hypothesis of panmixia, even between sexes from the same population (Table 6). Similarly, assignment tests generally assign individuals to the gender of origin (Table 1).

Null (non-amplifying in PCR) alleles are commonly encountered in population genetic studies (Dakin and Avise, 2004). A common cause of non-amplification is mutation in the priming region flanking the microsatellite repeats leading to poor annealing of the primer(s). Somatic mutation might be expected to generate diversity in such flanking regions, thus increasing the chance of alleles becoming non-amplifiable. Methods used for inferring the presence of null alleles (reviewed in Dakin and Avise, 2004) are commonly based on interpretation of heterozygote deficits, a phenomenon that can have numerous causes (Dakin and Avise, 2004). Given the difficulties of detecting null alleles in organisms with “normal” population genetics, we did not investigate the possibility of null alleles being present in our worms.

Any population-genetic studies on schistosomes using worms raised from field-collected snails might encounter the distorting effects of the presence of niMLGs. This can be minimised by ensuring that the number of infected snails (or of miracidia used to infect snails) is far greater than the number of adult worms ultimately genotyped. The study on *S. mansoni* by Thiele et al. (2008) is a good model. They raised adult worms starting by infecting snails with about 150 miracidia per infrapopulation (i.e. a single infected human) and typically obtained MLGs for 20–30 of these worms raised in mice. Similarly, Shrivastava et al. (2005) used cercariae from 50 snails from several populations of *S. japonicum* in China, and genotyped 20 adults from each. Of course, chance and the effects of selection can mean that not all miracidia or cercariae are represented in the sample of adults used for genotyping. This might explain why the 50 infected snails from Tianquan County, Sichuan, ultimately yielded only male worms in the study by Shrivastava et al. (2005).

The problems detailed above can be avoided if genotyping is accomplished using adult worms from naturally infected hosts, or from the miracidia that they produce. Clearly, adult worms cannot be obtained from infected people, but miracidia are accessible. Gower et al. (2007) and Steinauer et al. (2008) were able to obtain MLGs from individual miracidia of *S. mansoni*, but this was technically difficult. Characterisation of MLGs from larval stages of *S. japonicum* has not been reported to date.

Not all of our populations yielded the same picture of clustered MLGs and high levels of LD. The population from Duchang (Jiangxi Province) is of particular interest. Of 62 worms fully genotyped, 61 possessed unique MLGs (Table 1). Fewer individuals were successfully assigned to the correct gender than in most other populations (implying relatively little gender differentiation, as expected in a panmictic population) (Table 1 and pie chart in Fig. 3C) and LD was lower than in any other population (Table 3 and Table 4). As is clear in Fig. 3C, there was little suggestion of clustering of closely related MLGs. All this implies that the Duchang population is tending towards panmixia (but apparently not reaching it: note F_{IS} and F_{ST} values in Table 3, Table 4, Table 6). This population was founded using cercariae from only 11 snails! Many of these presumably had been infected by more than one miracidium. Typical infection levels at the Duchang site are by far the highest of any of our sites (6.5% - Table 1), making it

more plausible that multiple infections of a single snail were common. Two other populations, Shashi and Yueyang, also exhibited some of the properties seen in the Duchang population. In the Shashi population, this can be partly explained by the apparent sharing of some MLGs by both sexes, possibly due to mis-assignment of sex during sorting of specimens.

Where duplicate MLGs are encountered, it is straightforward to remove excess copies after confirming that they are unlikely to have arisen through random sexual recombination. When we removed duplicate MLGs (but not niMLGs clustering with those in PCAs) from the data, population-genetic statistics, such as values for F_{ST} , LD and proportion of individuals correctly assigned to gender, tended towards values expected under panmixia but rarely reached those values (Table 1, Table 5 and Table 6, and compare Table 3 and Table 4). Where clusters of niMLGs appear to be present, it might be appealing to retain for analysis only a single genotype from each cluster. There are two difficulties with this approach. Firstly, there is no clear way to determine whether niMLGs are a consequence of somatic mutation or are actually derived from different, but related, miracidia. Second, sample sizes would be greatly diminished. Fig. 3A provides a basis for discussing this; the 58 worms, developing from cercariae from seven snails from Guichi, fall into about eight groups – three consisting of females and five of males. Thus, the sample size would fall from 58 to eight if a single representative of each group was retained. Of course, this figure might reflect the actual number of miracidia that originally infected the snails. A sample size of eight is not normally regarded as adequate for any population-genetic study.

When appropriate population samples are available, the microsatellite loci presented here are likely to be useful for population-genetic studies on *S. japonicum*. They contain simple trinucleotide repeats that are easy to score and, as far as we know, are not close to any gene that might be subject to selection. Interestingly, none of these loci was consistently homozygous (hemizygous) in females (the heterogametic sex) suggesting that none is on the Z chromosome. This seems odd, given the large size of the sex chromosomes and the large proportion of the genetic material that they presumably constitute (see karyotype in Hirai et al., 2000). Microsatellites on the W chromosome would not amplify in males and would have been rejected in the initial screening for our study.

The phenomenon of multiple niMLGs among worms raised from pooled cercariae creates difficulties for population-genetic analyses. Our preferred explanation of this phenomenon is somatic mutation. If such mutations do indeed arise routinely during development in *Oncomelania* snails, there are fundamental consequences for population-genetic analyses of *S. japonicum* and indeed for schistosomes at large.

Acknowledgments

We gratefully acknowledge support from NIH-NIAID award number P50 AI39461 (the China TMRC entitled “Emerging Helminthiases in China: Genetic Diversity, Transmission Dynamics, and the Impact of Environmental Change”) and from the China National High Tech Program (863) 2007AA02Z153. We thank the Wellcome Trust (UK) and the National Health and Medical Research Council (Australia) for support via an International Collaborative Research Grants Scheme Award.

References

- Arnaud-Haond S, Belkhir K. GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol. Ecol. Notes* 2007;7:15–17.
- Balloux F, Lehmann L, DeMeeus T. The population genetics of clonal and partially clonal diploids. *Genetics* 2003;164:1635–1644. [PubMed: 12930767]
- Bayne CJ, Grevelding CG. Cloning of *Schistosoma mansoni* sporocysts in vitro and detection of genetic heterogeneity among individuals within clones. *J. Parasitol* 2003;89:1056–1060. [PubMed: 14627157]

- Brouwer KC, Ndhlovu P, Munatsi A, Shiff CJ. Genetic diversity of a population of *Schistosoma haematobium* derived from schoolchildren in east central Zimbabwe. *J. Parasitol* 2001;87:762–769. [PubMed: 11534639]
- Dabo A, Durand P, Morand S, Diakite M, Langand J, Imbert ED, Doumbo O, Jourdan J. Distribution and genetic diversity of *Schistosoma haematobium* within its bulinid intermediate hosts in Mali. *Acta Tropica* 1997;66:15–26. [PubMed: 9177092]
- Dakin EE, Avise JC. Microsatellite null alleles in parentage analysis. *Heredity* 2004;93:504–509. [PubMed: 15292911]
- Davies CM, Webster JP, Kruger O, Munatsi A, Ndamba J, Woolhouse MEJ. Host-parasite population genetics: a cross-sectional comparison of *Bulinus globosus* and *Schistosoma haematobium*. *Parasitology* 1999;119:295–302. [PubMed: 10503255]
- Eppert A, Lewis FA, Grzywacz C, Coura-Filho P, Caldas I, Minchella DJ. Distribution of schistosome infections in molluscan hosts at different levels of parasite prevalence. *Journal of Parasitology* 2002;88:232–236. [PubMed: 12053991]
- Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 2005;1:47–50. [PubMed: 19325852]
- Gower CM, Shrivastava J, Lamberton PHL, Rollinson D, Webster BL, Emery A, Kabatereine NB, Webster JP. Development and application of an ethically and epidemiologically advantageous assay for the multi-locus microsatellite analysis of *Schistosoma mansoni*. *Parasitol* 2007;134:523–536.
- Halkett F, Simon JC, Balloux F. Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol. Evol* 2005;20:194–201. [PubMed: 16701368]
- Hao Y, Wu XH, Xia G, Zheng H, Guo JG, Wang LY, Zhou XN. Schistosomiasis situation in People's Republic of China in 2005. *Chin. J. Schistosomiasis Control* 2006;18:321–324. (in Chinese)
- Hirai H, Hirata M, Aoki Y, Tanaka M, Imai HT. Chiasma analyses of the parasite flukes, *Schistosoma* and *Paragonimus* (Trematoda), by using the chiasma distribution graph. *Genes Genet. Syst* 1996;71:181–188.
- Hirai H, Taguchi T, Saitoh Y, Kawanaka M, Sugiyama, Habe S, Okamoto M, Hirata M, Shimada M, Tiu WU, Lai K, Upatham ES, Agatsuma T. Chromosomal differentiation of the *Schistosoma japonicum* complex. *Int. J. Parasitol* 2000;30:441–452. [PubMed: 10731567]
- Minchella DJ, Sollenberger KM, Pereira de Souza C. Distribution of schistosome genetic diversity within molluscan intermediate hosts. *Parasitology* 1995;111:217–220. [PubMed: 7675537]
- Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 2006;6:288–295.
- Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors : causes, consequences and solutions. *Nature Reviews Genetics* 2005;6:847–859.
- Prugnolle F, Roze D, Theron A, De Meeus T. F-statistics under alternation of sexual and asexual reproduction: a model and data from schistosomes (platyhelminth parasites). *Mol. Ecol* 2005;14:1355–1365. [PubMed: 15813776]
- Semyanova SK, Chrisanova GG, Filippova EK, Beer SA, Voronin MV, Ryskov AP. Individual and population variation in cercariae of bird schistosomes of the *Trichobilharzia ocellata* species group as revealed with the polymerase chain reaction. *Russian J. Genetics* 2005;41:12–16.
- Shrivastava J, Qian BZ, McVean G, Webster JP. An insight into the genetic variation of *Schistosoma japonicum* in mainland China using DNA microsatellite markers. *Mol. Ecol* 2005;14:839–849. [PubMed: 15723675]
- Sire C, Langand V, Barral V, Theron A. Parasite (*Schistosoma mansoni*) and host (*Biomphalaria glabrata*) genetic diversity: population structure in a fragmented landscape. *Parasitology* 2001;122:545–554. [PubMed: 11393828]
- Steinauer ML, Agola LE, Mwangi IN, Mkoji GM, Loker ES. Molecular Epidemiology of *Schistosoma mansoni*: A Robust, High-throughput Method to Assess Multiple Microsatellite Markers from Individual Miracidia. *Infection, Genetics and Evolution* 2008;8:68–73.
- Theron A, Sire C, Rognan A, Prugnolle F, Durand P. Molecular ecology of *Schistosoma mansoni* transmission inferred from the genetic composition of larval and adult infrapopulations within intermediate and definitive hosts. *Parasitology* 2004;129:571–585. [PubMed: 15552402]

- Thiele EA, Sorensen RE, Gazzinelli A, Minchella DJ. Schistosome genetic diversity and population structuring in a Brazilian village. *Int. J. Parasitol* 2008;38:389–399. [PubMed: 17825305]
- Utzinger J, Zhou XN, Chen MG, Bergquist R. Conquering schistosomiasis in China: the long march. *Acta Trop* 2005;96:69–96. [PubMed: 16312039]

Fig. 1. Map showing the southern half of China, the course of the Yangtze River and the sampling locations for the population study.

Fig. 2.

Traces from the ABI 3730 XL automated DNA sequencer as read by *GENEMAPPER* for locus Sjp4 from several female worms from Changde that were otherwise near-identical in their multi-locus genotypes (MLGs). Each panel shows the trace for a single worm. Two peaks indicate a heterozygote, one a homozygote. Numbers along the top indicate PCR fragment length; numbers in each panel show the number of TAA repeats in each allele. These traces show that differences at a single locus in otherwise similar or identical MLGs are easily detected and verified.

Fig. 3.

Principal coordinates analysis (first two axes only) based on matrixes of codominant genotypic distances. Ellipses are drawn around some clusters to clarify numbers when points are close together. A) population from Guichi (Anhui Province). The first two axes explain 51% of the total variation. B) Changde in Hunan Province (52%) and C) Duchang in Jiangxi Province (53%). The pie-charts in each panel show the results of an AMOVA analysis in *GENALEX* in which males and females were treated as separate populations. The pie “slices” indicate the amount of variation between (A) and within (W) the sexes.

Table 1

Localities, numbers of infected snails, numbers of worms from each locality completely genotyped and number of unique multi-locus genotypes (MLGs) found. Each geographic population was split into male and female “populations” and an assignment test done in GENALEX, with or without duplicate MLGs included.

Population Province and County of origin (and abbreviation used)	Latitude, longitude	Approx % infection in local snails ^a	Number of infected snails used	Number of worms retained ^b	Number of unique MLGs (male, female)	% males correctly assigned (duplicates present/absent)	% females correctly assigned (duplicates present/absent)
Anhui Province, Tongling County, Laozhou Island, Guanghui Village (TN)	30.94499, 117.76424	3.25%	10	59	23 (12, 11)	91/75	92/72
Anhui Province, Guichi County, Minsheng Village (GC) ^c	30.67025, 117.45383	0.14%	7	58	45 (27, 18)	100/ 100	100/ 100
Jiangxi Province, Duchang County, Tangmei Village (DC)	29.20452, 116.50359	6.5%	11	62	61 (28, 33)	71/ 68	68/ 70
Hunan Province, Changde City, Wuyi Village (CD) ^{c,d}	28.93987, 112.16454	0.93%	21	72	55 (21, 34)	100/ 100	95/ 97
Hunan Province, Yueyang City, Laogang Village (YY)	29.33595, 113.06577	0.02%	29	44	42 (25, 17)	88/ 92	83/ 82
Hubei Province, Shashi City, Maling Village (SH) ^d	30.32112, 112.35380	<0.01%	30	41	29 (15, 14)	81/ 73	55/ 43
Sichuan, Xichang City, Daxing Township, Shian and Jianxin Villages combined (XC)	27.50, 102.2	0.14%	15	67	56 (33, 23)	100/ 100	96/ 96

^aData from Institute for Parasitic Diseases, Shanghai.

^bAfter removal of worms for which any data missing.

^cOnly 16 loci used because of poor scoring success at one locus (Sjp7).

^dMLGs shared between sexes in these populations (see section 3.3).

Table 2

Length range (no. of tandem TAA repeats)				Range observed	
Length range (PCR product – bp)				<i>H</i>	
No. of alleles					
Annealing Temp C					
233–284	17	55	0.69–1.00		

Range observed

H

Annealing
Temp C

No. of alleles

Length range
(PCR product –
bp)

Length range
(no. of tandem
TAA repeats)

0.42–1.00

56

26

252–342

1–31

CCAGTGT
CTTCATT

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)

Length range –
(PCR product –
bp)

No. of alleles

Annealing
Temp C

Range observed

H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
Range observed
H

Range observed

H

Annealing
Temp C

No. of alleles

Length range
(PCR product –
bp)

Length range
(no. of tandem
TAA repeats)

0.78–1.00

63

20

206–266

3–23

Length range
(no. of tandem
TAA repeats)

Length range –
(PCR product –
bp)

No. of alleles

Annealing
Temp C

Range observed

H

Length range
(no. of tandem
TAA repeats)

Length range –
(PCR product –
bp)

No. of alleles

Annealing
Temp C

Range observed

H

0.65–1.00

65

19

226–280

7–25

TTTGTGG
GCTGCTGAA

Range observed

H

Annealing
Temp C

No. of alleles

Length range
(PCR product –
bp)

Length range
(no. of tandem
TAA repeats)

0.73–1.00

63

27

195–276

7–34

GTGTGTG
AAGCGAAT

Length range
(no. of tandem
TAA repeats)
Length range –
(PCR product –
bp)
No. of alleles
Annealing
Temp C
H
Range observed

Sequences of microsatellite loci reported here have been deposited in GenBank with accession numbers [EU262604](#)-[EU262620](#).

Table 3
Observed and expected heterozygosity, F_{IS} values and percentage of pairs of loci in linkage disequilibrium (LD). Duplicate multi-locus genotypes (MLGs) were included in these analyses. Figures are given for each population, and for males (m) and females (f) separately within each population.

Popn.	Mean observed heterozygosity ($\pm$ S.D.)		Mean expected heterozygosity ($\pm$ S.D.)		F_{IS}	Percent of pairs of loci in LD (0.05 level of significance)	
TN	0.825 (0.245)	m 0.853 (0.180) f 0.807 (0.336)	0.805 (0.060)	m 0.858 (0.027) f 0.630 (0.120)	-0.025	m 0.005 f -0.287 ^a	m 100 f 98
	0.776 (0.214)	m 0.831 (0.171) f 0.717 (0.327)	0.858 (0.032)	m 0.862 (0.037) f 0.689 (0.137)	0.097 ^a	m 0.036 f -0.041	m 88 f 88
DC	0.788 (0.137)	m 0.826 (0.135) f 0.757 (0.152)	0.894 (0.025)	m 0.888 (0.041) f 0.891 (0.022)	0.119 ^a	m 0.071 ^a f 0.152 ^a	m 34 f 41
	0.878 (0.169)	m 0.775 (0.275) f 0.945 (0.148)	0.837 (0.030)	m 0.819 (0.061) f 0.744 (0.054)	0.050 ^a	m 0.055 f -0.273 ^a	m 97 f 88
YY	0.767 (0.147)	m 0.788 (0.133) f 0.736 (0.187))	0.896 (0.018)	m 0.881 (0.024) f 0.869 (0.037)	0.145 ^a	m 0.107 ^a f 0.156 ^a	m 76 f 73
	0.838 (0.216)	m 0.833 (0.226) f 0.844 (0.234)	0.809 (0.066)	m 0.770 (0.116) f 0.769 (0.069)	-0.036	m -0.084 f -0.100 ^a	m 58 f 71
XC	0.778 (0.158)	m 0.781 (0.198) f 0.773 (0.218)	0.815 (0.097)	m 0.762 (0.115) f 0.768 (0.070)	0.046 ^a	m 0.025 f -0.007	m 99 f 85

^aSignificant departure from random expectations.

Table 4

Observed and expected heterozygosity, F_{IS} values and percentage of pairs of loci in linkage disequilibrium (LD). Duplicate multi-locus genotypes (MLGs) were removed prior to these analyses. Figures are given for each population, and for males (m) and females (f) separately within each population. Positive values for F_{IS} indicate a deficit of heterozygotes.

Popn.	Mean observed heterozygosity ($\pm$ s.d.)		Mean expected heterozygosity ($\pm$ s.d.)		F_{IS}	Percent of pairs of loci in LD 0.05 level of significance	
TN	0.823	m 0.839 (0.219)	0.889	m 0.898 (0.023)	0.075 ^a	m 0.069	m 0
	(0.202)	f 0.813 (0.232)	(0.029)	f 0.835 (0.063)	f 0.028	f 0	f 0
GC	0.789	m 0.839 (0.166)	0.876	m 0.871 (0.034)	0.100 ^a	m 0.038	m 83
	(0.196)	f 0.720 (0.307)	(0.025)	f 0.728 (0.124)	f 0.011	f 0.011	f 88
DC	0.787	m 0.826 (0.135)	0.895	m 0.888 (0.041)	0.121 ^a	m 0.071*	m 31
	(0.138)	f 0.754 (0.155)	(0.025)	f 0.890 (0.023)	f 0.155*	f 0.155*	f 38
CD	0.881	m 0.769 (0.282)	0.842	m 0.823 (0.056)	-0.046 ^a	m 0.078	m 82
	(0.165)	f 0.945 (0.138)	(0.032)	f 0.761 (0.051)	f -0.246*	f -0.246*	f 88
YY	0.765	m 0.785 (0.131)	0.897	m 0.883 (0.022)	0.149 ^a	m 0.113*	m 72
	(0.143)	f 0.735 (0.180)	(0.019)	f 0.875 (0.035)	f 0.164*	f 0.164*	f 63
SH	0.827	m 0.815 (0.202)	0.835	m 0.833 (0.103)	0.010	m 0.022	m 20
	(0.199)	f 0.839 (0.214)	(0.065)	f 0.818 (0.057)	f -0.027	f -0.027	f 24
XC	0.771	m 0.771 (0.191)	0.818	m 0.761 (0.113)	0.057 ^a	m -0.013	m 95
	(0.149)	f 0.772 (0.212)	(0.088)	f 0.775 (0.069)	f 0.004	f 0.004	f 74

^aIndicates significant departure from random expectations.

Pairwise F_{ST} values from GenAlEx. Values for populations with duplicates included are below diagonal, those without above diagonal. All values are significant below the level of 0.005 (based on 9,999 permutations).

	TN	GC	DC	CD	YY	SH	XC
TN							
GC	0.117	0.055	0.030	0.065	0.037	0.065	0.075
DC	0.076		0.046	0.073	0.043	0.083	0.072
CD	0.110	0.056		0.052	0.019	0.060	0.069
YY	0.092	0.080	0.056		0.056	0.098	0.108
SH	0.116	0.053	0.019	0.060		0.062	0.064
XC	0.129	0.108	0.076	0.115	0.081		0.100
		0.083	0.070	0.114	0.068	0.112	

Table 6

Pairwise F_{ST} values from GenAEx for 14 populations (males and females treated as separate “populations”). Values for populations with duplicates included are below diagonal, those without above diagonal. All values are significant below the level of 0.005 (based on 9,999 permutations) except for the cell marked with superscript 1 in which $P = 0.048$.

	TN(f)	TN(m)	GC(f)	GC(m)	DC(f)	DC(m)	CD(f)	CD(m)	YY(f)	YY(m)	SH(f)	SH(m)	XC(f)	XC(m)
TN(f)		0.050	0.164	0.088	0.050	0.061	0.124	0.110	0.072	0.071	0.104	0.077	0.120	0.131
TN(m)	0.201		0.135	0.057	0.035	0.036	0.111	0.081	0.045	0.048	0.083	0.069	0.103	0.115
GC(f)	0.295	0.180		0.140	0.118	0.130	0.180	0.136	0.142	0.124	0.184	0.164	0.191	0.178
GC(m)	0.213	0.094	0.169		0.051	0.054	0.134	0.093	0.067	0.048	0.093	0.082	0.101	0.100
DC(f)	0.157	0.061	0.143	0.057		0.012	0.100	0.057	0.032	0.032	0.068	0.063	0.091	0.100
DC(m)	0.181	0.066	0.154	0.059	0.011		0.103	0.059	0.023	0.034	0.070	0.070	0.096	0.107
CDF	0.231	0.142	0.207	0.147	0.107	0.113		0.132	0.118	0.105	0.162	0.142	0.175	0.183
CD(m)	0.234	0.116	0.157	0.108	0.068	0.073	0.147		0.070	0.072	0.113	0.102	0.139	0.142
YY(f)	0.206	0.076	0.172	0.075	0.035	0.026	0.133	0.087		0.039	0.078	0.081	0.109	0.108
YY(m)	0.196	0.077	0.147	0.053	0.032	0.036	0.115	0.081	0.044		0.082	0.067	0.095	0.100
SH(f)	0.251	0.132	0.235	0.127	0.097	0.096	0.197	0.152	0.106	0.116		0.022 ¹	0.144	0.138
SH(m)	0.204	0.130	0.217	0.119	0.098	0.106	0.176	0.144	0.123	0.108	0.093		0.130	0.129
XC(f)	0.234	0.135	0.218	0.106	0.096	0.100	0.188	0.149	0.119	0.101	0.172	0.171		0.120
XC(m)	0.253	0.140	0.203	0.102	0.099	0.105	0.192	0.154	0.113	0.102	0.160	0.155	0.124	