The effect of sample size on estimates of genetic differentiation and effective population size for *Schistosoma mansoni* populations

Lúcio M. Barbosa [a,b], Bruna C. Barros [b], Moreno de Souza Rodrigues [a], Luciano K. Silva [a], Mitermayer G. Reis [a,b], Ronald E. Blanton [c,*]

[a] Gonçalo Moniz Research Centre, Oswaldo Cruz Foundation, Rua Waldemar Falcão, 121 Brotas, Salvador, Bahia 40296-710, Brazil
[b] Bahiana School of Medicine and Public Health, Av. Silveira Martins, n° 3386, Salvador, Bahia 41150-100, Brazil
[c] Center for Global Health and Diseases, Case Western Reserve University, Biomedical Research Building, 2109 Adelbert Rd., Cleveland, OH 44106, USA

ABSTRACT

Eradication or local extinction of the human parasite *Schistosoma mansoni* is a goal for many control programs. Population genetic analyses are helping to evaluate and guide these efforts, yet what to sample, how to sample and how densely to sample is not well established. We determined the *S. mansoni* allele frequency profile of nearly all infected inhabitants in two small Brazilian communities and created sub-samples representing 5–50% of all detected human infections (infrapopulations). Samples were selected at random with replacement, and each size class was replicated 100 times. Mean pairwise differentiation for all infrapopulations (Di) and the variance effective population size (Ne) were calculated for each sample. Prior to community-wide treatment, the true mean Di was moderate (0.095–0.123) and Ne large (>30,000). Most samples of <50% of those infected produced estimates outside of 5% of the true value. For estimates within 10%, sample sizes of >15% of all infrapopulations were required. At the 3 year follow-up after treatment, the Di increased and Ne was reduced by >15 fold. At this time sampling of >30–45% was needed to achieve the same accuracy. Following a second treatment and 4 years from baseline, the Di further increased and Ne decreased with little change in the sampling effort required. Extensive sampling is required for accurate estimates of these important population parameters. Characteristics such as population census size, infection prevalence, the community's treatment history and the degree of infrapopulation differentiation should be taken into account. The intensity of infection was weakly correlated with the ability of a single infrapopulation to represent the component population (Dic), indicating a tendency toward random acquisition of parasite genotypes. This also suggests that targeted sampling from those most heavily infected will better represent the genetic diversity of the whole community than a random sample of infrapopulations.

© 2018 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

In a perfect epidemiologic study, every individual in a population would be enrolled in order to answer a given research question. This is not practical or perhaps even advisable for large populations. In response, studies resort to statistical analyses of a representative portion of the whole. A sample, however, may be biased and not represent the population, is susceptible to winner's curse (Kraft, 2008; Xiao and Boehnke, 2009), may identify trivial effect sizes (Biau et al., 2008; Button et al., 2013). One of the greatest dangers is if too small, it may lack power to detect differences between groups. The problem of small sample sizes is common. Moher et al. (1994) indicated that only 36% of the clinical trials between 1975 and 1990 with negative results had statistical power to detect differences of 50% between groups. By contrast, a very large sample size can be complicated by the cost, time to complete enrollment and potentially subjecting individuals unnecessarily to risk in clinical trials. In population genetics, no single sample size will serve to answer all questions for all populations, since there can be differences in frequency and distribution of variation, the resolution and accuracy desired, in the amount of uncertainty in the estimates and basic biology among populations and species. Nevertheless, estimates of the minimum sample size required in general or for specific species have been calculated to be from

* Corresponding author at: Center for Global Health and Diseases, Case Western Reserve University, Cleveland, Wolstein Research Building, 2103 Cornell Rd., Cleveland, OH 44106, USA. Fax: +1 216 368 4825.
E-mail address: reb6@case.edu (R.E. Blanton).

two (Nazareno et al., 2017) to greater than 300 (B-Rao, 2001; Miyamoto et al., 2008; Whiteley et al., 2012).

In studies of population genetics involving parasites such as *Schistosoma mansoni*, determining adequate sample size and selecting a sampling approach are further complicated by the biology of the microorganism itself. During their life cycle, schistosomes employ asexual and sexual reproduction in different host species. Its sexual stage is distributed as genetically heterogenous aggregates (Curtis et al., 2002; Thiele et al., 2008; Agola et al., 2009) that can be viewed from the perspective of the individual host (infrapopulation) or as an overall community of hosts (component population) (Bush et al., 1997). In addition to geography, component populations can be defined by time or host characteristics (Barbosa et al., 2013).

Hardy–Weinberg equilibrium is an assumption of many indices in population genetics. Infrapopulations formed by schistosomes are intrinsically different from typical Hardy–Weinberg populations in that all of the adults are acquired by immigration and all of the progeny emigrate. In addition, the assumption of random mating cannot be met. Those genotypes only return to the host by chance and not because they are retained at or return to the same site as the adults. Analyses of infrapopulations based on assumption of Hardy–Weinberg proportions are suspect, but not necessarily wrong. The component population, however, is likely to more closely conform to Hardy–Weinberg conditions by reflecting the whole parasite community's reproductive activity and retain a set of shared traits and loci. The inbreeding characteristics of the population are observed at the level of the component population.

Schistosoma mansoni adult worms reside in the human blood stream and are inaccessible for all practical purposes. Thus, the parasite's eggs, excreted in faeces, are the sample commonly collected for genetic studies. Eggs also have the advantage of representing a single generation, an important assumption for many analyses. Humans with light infections can excrete over 5000 genetically diverse eggs per day. When genotyping eggs or miracidia, all (Van den Broeck et al., 2014) or most (Curtis et al., 2002; Prugnolle et al., 2004a) are unique multilocus genotypes. Prugnolle et al. (2004a) also found that clonal mating and mating of highly related individuals was unusual in adult worm pairs from naturally infected wild rats. Most of these observations would seem to dictate the necessity for extensive sampling for studies of schistosome population genetics. There have been few estimates for the minimum sample size required to provide sample size estimates for *S. mansoni* (Jarne and Theron, 2001; Sorensen et al., 2006; French et al., 2013).

One approach for estimating allele frequencies is to disaggregate and isolate individual eggs from faeces (or miracidia hatched from eggs) and obtain discrete genotypes that are then pooled to estimate population allele frequencies (Stohler et al., 2004; Steinauer et al., 2009; Norton et al., 2010). Our group has taken advantage of the naturally aggregated nature of faecal eggs to collect them from whole stools and determine allele frequencies directly based on this pool. This facilitated examination of every stool egg-positive individual in a small population of 849 in 2009 (Barbosa et al., 2016). With this approach, there is no question of how many eggs or miracidia to sample nor how many hosts. In much larger populations, however, even with this approach it may not be possible to examine all infected individuals, necessitating taking a sample. The characteristics of the population will also determine how extensive sampling should be, since greater diversity with uneven distribution should require larger samples to be representative. Due to the aggregated nature of infections, however, a population with low heterogeneity may show high differentiation between infrapopulations, or conversely, a highly heterogeneous population might still have similar representation among infrapopulations (Blank et al., 2009). We calculated how a genetic differentiation index (Jost's D) and effective population size (Ne) would vary for *S. mansoni* populations with resampling different samples sizes of infrapopulations from human infections. Allele frequencies were obtained at different periods from a cohort study in Bahia state, Brazil where nearly all those infected within a discrete transmission zone were enrolled. The parasite population was exposed to repeated treatment with praziquantel, permitting evaluation of how different population structures affected results from different sample sizes.

2. Materials and methods

2.1. Study sites and database

Between 2009 and 2013, two rural communities (Jenipapo and Volta do Rio) in the state of Bahia, Brazil were surveyed and treated twice for *S. mansoni* infections (Barbosa et al., 2016). These villages are 6 km apart on the Jiquiriçá River and have a total population of 849. An average of 85% of the population participated in all three parasitological surveys where three stools were collected on different days, examined for infection and parasite eggs collected from all positive stools. All infected individuals were treated with success rates ranging from 86% to 98%. All those found to be infected 6 weeks post-treatment were re-treated (Fig. 1). No community-wide treatment was carried out in either village prior to 2009.

2.2. Genotyping

Stool eggs from every infected individual were concentrated by sieving and sedimentation, for DNA extraction and genotyping for at least 11 microsatellite markers (Barbosa et al., 2016). A sample was considered successfully genotyped and included in the analysis when at least eight of the markers successfully amplified. Samples that did not amplify tended to be from individuals with low egg counts and consequently lower concentrations of parasite DNA (Blanton et al., 2011). PCR success rates were similar throughout the period (range 58.8–66%). In Jenipapo, total numbers of successfully genotyped infrapopulations were 140, 70 and 48 for 2009, 2012 and 2013, respectively. The total numbers in Volta do Rio were 81 and 16 for 2009 and 2012.

2.3. Sampling scheme, differentiation and Ne

The sample size categories represented proportions of the total number of those infected (5, 10, 20, 30, 40, 50%), since the number
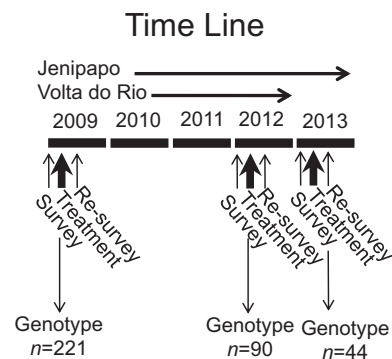

Fig. 1. Study design. Stool from nearly all members (89–98%) of two communities (Jenipapo and Volta do Rio, Brazil) were examined on three different days and *Schistosoma mansoni* eggs collected and genotyped. All those positive were re-examined 4–6 weeks post-treatment and re-treated if persistently positive. Only Jenipapo was surveyed in 2013.

infected differed by year. For each of six size categories, 100 replicate samples of infrapopulations were selected at random with replacement using the open source software R 3.3.2 (R Development Core Team 2016). The R script is available upon request. Allele frequencies were then used to estimate genetic differentiation (Josts D) between infrapopulations (Di) using the software SPADE (http://chao.stat.nthu.edu.tw). SPADE also calculates *P* values and the 95% confidence interval for each estimate by bootstrap methods.

Variance Ne was calculated with the program MLNe (Wang and Whitlock, 2003) that uses maximum-likelihood to jointly calculate Ne and the migration rate. The program parameters were the assumption that the populations were not in drift equilibrium, maximum Ne value of 35,000, source populations from 2009 and focal populations from 2012. The proportion of samples falling outside of 5% and 10% of the mean values for the total population was plotted. The temporal method uses the allele frequency variance over generations. Minimum generation time for *S. mansoni* was estimated to be 12 weeks, so a maximum of 10.8 generations passed between the first treatment period and the resurvey in 2012. While adults represent overlapping generations and, therefore, complicate estimates of Ne, the eggs are of the same generation.

To further assess sampling strategies, we calculated the differentiation between each infrapopulation and the component populations (Dic). Correlation between the Dic and host characteristics (age, sex, intensity of infection) was assessed by multiple linear regression using Excel (Microsoft, Redmond, CA, USA) after log transformation of the Dic and stool egg counts.

3. Results

3.1. Infrapopulation differentiation

Changes in differentiation and effective population size between sampling periods allowed analysis of sample size under differing population structure. In both communities, the pairwise Di rose significantly and the Ne decreased more than 10 fold over the period of study (Barbosa et al., 2016). Prior to community-wide treatment in 2009, most of the subsamples produced results for mean Di within 5% of the true value only when sampling >40% of the infected population (Fig. 2A and B). Most results were within 10% of the true value with samples representing $\geq$15% of the total population. In 2012, 3 years after the first community-wide treatment, no subsample category was on average within 5% of the true value, and sampling at least 30% in Jenipapo and 45% in Volta do Rio was required to place most estimates within 10% of the true value (Fig. 2C and D). In 2013, when only Jenipapo was surveyed, and 1 year following the second round of community-wide treatment, most samples were within 10% of the true value only when sampling at least 35% of the infected population (Fig. 2E). The S.D. of these estimates (Fig. 2F) increased with smaller sample sizes and as the prevalence of infection decreased. The S.D., nevertheless, remained relatively small, since thousands of eggs in aggregate were used in the calculations.

3.2. Ne

Ne was calculated between 2009 and 2012 for both communities with an estimated 10.8 generations between these years. Ne is dependent on sample size, and this was well reflected in estimates based on subsamples from 2012. All samples less than 100% produced lower estimates of Ne (Fig. 3) for Volta do Rio and plateaued between 50% and 5%, but for Jenipapo the decline
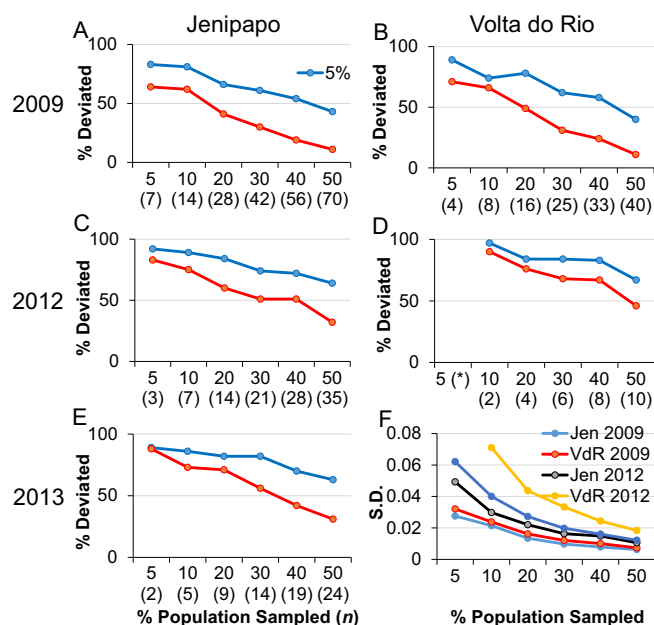

Fig. 2. Mean differentiation between *Schistosoma mansoni* infrapopulations (Di) (A–E) and S.D. (F) for 100 subsamples. Di based on the total population or subsamples of different sizes formed by resampling with replacement. The percent of samples that gave values 5% or 10% outside of the true value (based on all samples). Samples obtained from Jenipapo, Brazil in 2009 (A), 2012 (C) and 2013 (E). Samples obtained from Volta do Rio, Brazil in 2009 (B) and 2012 (D). The *X*-axis represents the percent (and number) of all infrapopulations in the subsamples. The *Y*-axis is the proportion of samples that were 5% or 10% above or below the mean for all infrapopulations. (F) Mean S.D. for each sample size, community and year of collection. Jen, Jenipapo; VdR, Volta do Rio.
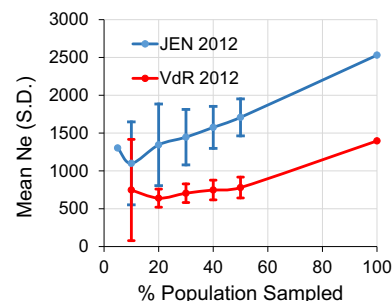

Fig. 3. Mean Ne for 100 *Schistosoma mansoni* infrapopulation subsamples from Brazil. Variance Ne based on the total 2012 population or subsamples of different sizes formed by resampling with replacement. The total population from 2009 was the source population. Jen, Jenipapo; VdR, Volta do Rio.

continued as the size of the subsamples decreased. Ne was not corrected for migration rate since this was estimated to be negligible.

3.3. Dic

How closely any individual infrapopulation reflected the composition of the component population was measured by comparing differentiation between the infrapopulation and the component population. This is the Dic. By multiple linear regression, age and sex were unrelated. The log normalised Dics from 2009 showed weak correlation with intensity of infection in Jenipapo, but no correlation in Volta do Rio (r^2 = 0.05 and 6 × 10^{-6}, respectively, Supplementary Fig. S1). The correlation with egg counts increased for both communities in subsequent years following treatment. In Jenipapo for 2012 r^2 = 0.18 and for 2013 r^2 = 0.32. In Volta do Rio for 2012 r^2 = 0.11. Residuals were normally distributed.

4. Discussion

We have approached the sampling problem by empirically determining the outcome of using subsamples of a population where essentially every known human infection was analysed, and a large number of eggs produced from each infection was used in aggregate to determine differentiation and diversity. We found that a large percentage of all infections have to be examined in order to achieve good accuracy and precision. We also observed, as have others, that the precise number depends on the structure of the population and the question being asked. As population structure changed over time, indicated by changes in Di and Ne, sampling requirements also changed relative to the percentage of infections that needed to be included. For infrapopulation differentiation, sample size greatly affected the accuracy of estimates. As *S. mansoni* prevalence decreased and the degree of genetic difference increased between parasites from individuals, a larger proportion of infrapopulations needed to be sampled to be representative (Fig. 2).

Ne was most sensitive to sample size when prevalence and diversity were highest. Ne, which correlates with diversity, required larger sample sizes than differentiation for accuracy. Some authors suggest 10% of the census population (Hare et al., 2011; Costa et al., 2015), while others recommend a number of the census population equivalent to 10% of the effective population size (Palstra and Ruzzante, 2008). In a study of ash trees, >300 adult samples were needed for accurate estimates of diversity, whereas smaller samples were sufficient for expected heterozygosity (Miyamoto et al., 2008). When seeds were sampled, 50% more seeds needed to be sampled than adults, but the resultant diversity measures were identical to those of adults. In a simulation study, a sample size of 2000 was needed for accurate estimates when Ne was 8000 (Ovenden et al., 2007). Ackerman et al. (2016), investigating Ne as the number of breeders in populations of steelhead trout recommended the equivalent of >100% of the effective population. Of note, the relationship between prevalence and diversity here is influenced by the study design, since the baseline infections were the result of decades of acquisition in some cases. By contrast, in subsequent years the diversity represented the acquisition of parasites by individuals over only 1–3 years.

Identifying the optimal sample size has been an active area of discussion in population genetics for years. This is primarily because so many answers are possible depending on the research question, the type of analysis to be applied and the structure and diversity of the population to be studied. Hale et al. (2012) appear to suggest that 25–30 individuals is a sufficient size to provide accurate allele frequencies and F_{ST} analysis for a variety of organisms. Two years later, Fung and Keenan (2014) in the same journal argued that under a variety of scenarios samples sizes of 49–285 could be required to ensure narrow confidence intervals. Some studies (Willing et al., 2012; Nazareno et al., 2017) indicate that using very high single nucleotide polymorphism densities in the thousands obtained from sequencing yields good population level information from just 2–8 individuals. By contrast, Fumagalli (2013) in simulation studies found that large sample sizes were more important for accuracy of allele frequencies and estimation of population genetic parameters than sequencing depth, despite the uncertainty for allele calling when sequencing depth is low.

There have been other attempts at validating parasite sample sizes in studies of schistosome population genetics. Sorensen et al. (2006) suggested using the results of genotyping worms derived from eggs from five human infections and passaged through laboratory snails and definitive hosts to estimate the sampling required for the rest of the population. It was not clear, how-ever, how this would translate into a specific number for the whole population. Jarne and Theron (2001) suggested 20 parasites from five hosts located at 10 different dispersal sites as being practical and providing statistical power to detect structure. French et al. (2013) empirically addressed the sampling problem of directly genotyping eggs collected from infected humans. Using data from a Tanzanian study of school-aged children, their conclusion was that only 1–5 miracidia from 50 individuals would be sufficient to reproduce their study with confidence. Their conclusion assumes a large degree of genetic redundancy in the egg population and that infections in their 115 children between the ages of 7–11 represent the whole community. Since the total sample in their study was 115, their recommendation could be taken to mean miracidia should be collected from 50% of all those infected. The authors make clear that a different site with different conditions is likely to require modification of their estimates.

There are potential errors in our analysis. The Kato-Katz method for stool examination is particularly insensitive for low intensity infections, so especially in years following treatment there might be an increasing number of cryptic infections. The allele frequencies derived from pooled samples is subject to 5–10% error in its measurement (Silva et al., 2006; Hanelt et al., 2009). Null alleles, heterozygosity and some F statistics cannot be measured in pooled samples. By contrast, the use of pooled samples has the potential to identify alleles not captured by sampling small numbers of eggs, to avoid the error associated with small samples of the infrapopulation, biased hatching of eggs or sampling eggs from a small region of the stool. Analysis of pools might not estimate the number of breeding adults due to variation in family size (Steinauer et al., 2013), but direct measurement made in laboratory (Blank et al., 2010) or natural animal infections (Shrivastava et al., 2005) indicate that egg allele frequencies do approach the frequencies found in the adults. In any case, the egg population itself is a valid and useful population for analysis.

Variation in family size means that some individuals will contribute more offspring to the next generation than others. This is not always an error to be corrected for differentiation, but is a characteristic of the population, and to simulate it away is to describe some population other than the one being examined. The effect on Ne is generally small (Frankham, 1995; Waples and Anderson, 2017). Some authors suggest that the effect of siblings should be removed from population genetics analysis, since they carry redundant genetic information and are not independent samples (Steinauer et al., 2013). This is a notion that comes primarily from statistical genetics where inclusion of relatives in association studies leads to bias (Anderson et al., 2010). Siblings are a natural part of a population. Their removal risks the combined errors of reduced sample size, statistical power and precision, and incorrect assignment (Waples and Anderson, 2017). In studies of parasites, siblings or highly related individuals are estimated from genetic marker data alone, and there is often unknown error with these estimation algorithms (Wang, 2014; Taylor, 2015). What Allendorf et al. (1982) and others (Whiteley et al., 2012) point to as an error when sampling fish populations is the likelihood in a limited spatial collection of capturing a sample enriched for siblings and not representative of the population over a larger area. The error is in the non-random collection of relatives and not that they are related per se. If the problem is the sampling approach, the best solution is not to remove siblings, but to perform better sampling.

While the location of adult worms and the distribution of parasites in infrapopulations that are not Hardy–Weinberg populations make sampling complex, one aspect of parasite biology simplifies sampling. The good news is that parasite genotypes are acquired relatively randomly from the environment, and within the defini-

tive host, clonal siblings (Curtis et al., 2002) and brother:sister matings are not common (Prugnolle et al., 2004b). Parasites also do not select for host features (Barbosa et al., 2013). Schistosomes are acquired from the local environment at random, as indicated by the inverse correlation between Dic and intensity of infection with reinfection. As we show here and as suggested by others (Sorensen et al., 2006), the more heavily infected, the better the individual infrapopulation approximates the component population. This is consistent with genotypes being acquired at random in a frequency-dependent manner. In contrast to our site in Brazil and others (Betson et al., 2013), one study in Senegal did find parasite differentiation by age but no other characteristic (Van den Broeck et al., 2014). When demographic features such as age are associated with parasite genotypes, some significant change in the parasite population, such as migration or widespread treatment, should be suspected. The contribution of rodents to the persistence of the parasite is unknown, but the presence of rodent infection should not affect the relationship of infection intensity to the Dic. A random acquisition of parasites is a representative model and mathematically the most simple. For the parameters that most affect public health decisions, i.e. parasite differentiation and diversity, analyses need not be overly complex, but sampling has to be extensive or one has to accept a large degree of error.

Acknowledgement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2018.10.001.

References

Ackerman, M.W., Hand, B.K., Waples, R.K., Luikart, G., Waples, R.S., Steele, C.A., Garner, B.A., McCane, J., Campbell, M.R., 2016. Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. Evol. Appl. 10, 146–160.

Agola, L.E., Steinauer, M.L., Mburu, D.N., Mungai, B.N., Mwangi, I.N., Magoma, G.N., Loker, E.S., Mkoji, G.M., 2009. Genetic diversity and population structure of Schistosoma mansoni within human infrapopulations in Mwea, central Kenya assessed by microsatellite markers. Acta Trop. 111, 219–225.

Allendorf, F.W., Knudsen, K.L., Phelps, S.R., 1982. Identification of a gene regulating the tissue expression of a phosphoglucomutase locus in rainbow trout. Genetics 102, 259–268.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., Zondervan, K.T., 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–1573.

B-Rao, C., 2001. Sample size considerations in genetic polymorphism studies. Hum. Hered. 52, 191–200.

Barbosa, L.M., Reis, E.A., Dos Santos, C.R., Costa, J.M., Carmo, T.M., Aminu, P.T., Pitanga, T.N., Ponce-Terashima, R., Blank, W.A., Silva, L.K., Reis, M.G., Blanton, R.E., 2016. Repeated praziquantel treatments remodel the genetic and spatial landscape of schistosomiasis risk and transmission. Int. J. Parasitol. 46, 343–350.

Barbosa, L.M., Silva, L.K., Reis, E.A., Azevedo, T.M., Costa, J.M., Blank, W.A., Reis, M.G., Blanton, R.E., 2013. Characteristics of the human host have little influence on which local Schistosoma mansoni populations are acquired. PLoS Negl. Trop. Dis. 7, e2572.

Betson, M., Sousa-Figueiredo, J.C., Kabatereine, N.B., Stothard, J.R., 2013. New insights into the molecular epidemiology and population genetics of Schistosoma mansoni in Ugandan pre-school children and mothers. PLoS Negl. Trop. Dis. 7, e2561.

Biau, D.J., Kerneis, S., Porcher, R., 2008. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. Clin. Orthop. Relat. Res. 466, 2282–2288.

Blank, W.A., Reis, E.A., Thiong'o, F.W., Braghiroli, J.F., Santos, J.M., Melo, P.R., Guimaraes, I.C., Silva, L.K., Carmo, T.M., Reis, M.G., Blanton, R.E., 2009. Analysis of Schistosoma mansoni population structure using total fecal egg sampling. J. Parasitol. 95, 881–889.

Blank, W.A., Test, M.R., Liu, S.F., Lewis, F.A., Blanton, R.E., 2010. Long-term genetic stability and population dynamics of laboratory strains of Schistosoma mansoni. J. Parasitol. 96, 900–907.

Blanton, R.E., Blank, W.A., Costa, J.M., Carmo, T.M., Reis, E.A., Silva, L.K., Barbosa, L.M., Test, M.R., Reis, M.G., 2011. Schistosoma mansoni population structure and persistence after praziquantel treatment in two villages of Bahia, Brazil. Int. J. Parasitol. 41, 1093–1099.

Bush, A.O., Lafferty, K.D., Lotz, J.M., Shostak, A.W., 1997. Parasitology meets ecology on its own terms: Margolis et al. revisited. J. Parasitol. 83, 575–583.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafo, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376.

Costa, L.S., Corneleo, N.S., Stefenon, V.M., 2015. Conservation of forest biodiversity: how sample size affects the estimation of genetic parameters. An. Acad. Bras. Cienc. 87, 1095–1100.

Curtis, J., Sorensen, R.E., Minchella, D.J., 2002. Schistosome genetic diversity: the implications of population structure as detected with microsatellite markers. Parasitology 125. Suppl S51-59.

Frankham, R., 1995. Effective population size/adult population size ratios in wildlife: a review. Genet. Res. 66, 95–107.

French, M.D., Churcher, T.S., Basanez, M.G., Norton, A.J., Lwambo, N.J., Webster, J.P., 2013. Reductions in genetic diversity of Schistosoma mansoni populations under chemotherapeutic pressure: the effect of sampling approach and parasite population definition. Acta Trop. 128, 196–205.

Fumagalli, M., 2013. Assessing the effect of sequencing depth and sample size in population genetics inferences. PLoS ONE 8, e79667.

Fung, T., Keenan, K., 2014. Confidence intervals for population allele frequencies: the general case of sampling from a finite diploid population of any size. PLoS ONE 9, e85925.

Hale, M.L., Burg, T.M., Steeves, T.E., 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS ONE 7, e45170.

Hanelt, B., Steinauer, M.L., Mwangi, I.N., Maina, G.M., Agola, L.E., Mkoji, G.M., Loker, E.S., 2009. A new approach to characterize populations of Schistosoma mansoni from humans: development and assessment of microsatellite analysis of pooled miracidia. Trop. Med. Int. Health 14, 322–331.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R.S., Ruegg, K., Palstra, F., 2011. Understanding and estimating effective population size for practical application in marine species management. Conserv. Biol. 25, 438–449.

Jarne, P., Theron, A., 2001. Genetic structure in natural populations of flukes and snails: a practical approach and review. Parasitology 123 (Suppl), S27–S40.

Kraft, P., 2008. Curses–winner's and otherwise–in genetic epidemiology. Epidemiology 19, 649–651.

Miyamoto, N., Fernandez-Manjarres, J.F., Morand-prieur, M.-E., Miyamoto, Naoko, Bertolino, P., Frascaria-Lacoste, N., 2008. What sampling is needed for reliable estimations of genetic diversity in Fraxinus excelsior L. (Oleaceae)? Ann. For. Sci. 65, 403.

Moher, D., Dulberg, C.S., Wells, G.A., 1994. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 272, 122–124.

Nazareno, A.G., Bemmels, J.B., Dick, C.W., Lohmann, L.G., 2017. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Mol. Ecol. Resour. 17, 1136–1147.

Norton, A.J., Gower, C.M., Lamberton, P.H., Webster, B.L., Lwambo, N.J., Blair, L., Fenwick, A., Webster, J.P., 2010. Genetic consequences of mass human chemotherapy for Schistosoma mansoni: population structure pre- and post-praziquantel treatment in Tanzania. Am. J. Trop. Med. Hyg. 83, 951–957.

Ovenden, J.R., Peel, D., Street, R., Courtney, A.J., Hoyle, S.D., Peel, S.L., Podlich, H., 2007. The genetic effective and adult census size of an Australian population of tiger prawns (Penaeus esculentus). Mol. Ecol. 16, 127–138.

Palstra, F.P., Ruzzante, D.E., 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Mol. Ecol. 17, 3428–3447.

Prugnolle, F., Choisy, M., Theron, A., Durand, P., De Meeus, T., 2004a. Sex-specific correlation between heterozygosity and clone size in the trematode Schistosoma mansoni. Mol. Ecol. 13, 2859–2864.

Prugnolle, F., Theron, A., Durand, P., De Meeus, T., 2004b. Test of pangamy by genetic analysis of Schistosoma mansoni pairs within its natural murine host in Guadeloupe. J. Parasitol. 90, 507–509.

Shrivastava, J., Gower, C.M., Balolong Jr., E., Wang, T.P., Qian, B.Z., Webster, J.P., 2005. Population genetics of multi-host parasites–the case for molecular epidemiological studies of Schistosoma japonicum using larval stages from naturally infected hosts. Parasitology 131, 617–626.

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

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

Steinauer, M.L., Christie, M.R., Blouin, M.S., Agola, L.E., Mwangi, I.N., Maina, G.M., Mutuku, M.W., Kinuthia, J.M., Mkoji, G.M., Loker, E.S., 2013. Non-invasive sampling of schistosomes from humans requires correcting for family structure. PLoS Negl. Trop. Dis. 7, e2456.

Steinauer, M.L., Hanelt, B., Agola, L.E., Mkoji, G.M., Loker, E.S., 2009. Genetic structure of Schistosoma mansoni in western Kenya: the effects of geography and host sharing. Int. J. Parasitol. 39, 1353–1362.

Stohler, R.A., Curtis, J., Minchella, D.J., 2004. A comparison of microsatellite polymorphism and heterozygosity among field and laboratory populations of Schistosoma mansoni. Int. J. Parasitol. 34, 595–601.

Taylor, H.R., 2015. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. Ecol. Evol. 5, 3140–3150.

Thiele, E.A., Sorensen, R.E., Gazzinelli, A., Minchella, D.J., 2008. Genetic diversity and population structuring of *Schistosoma mansoni* in a Brazilian village. Int. J. Parasitol. 38, 389–399.

Van den Broeck, F., Meurs, L., Raeymaekers, J.A., Boon, N., Dieye, T.N., Volckaert, F.A., Polman, K., Huyse, T., 2014. Inbreeding within human *Schistosoma mansoni*: do host-specific factors shape the genetic composition of parasite populations? Heredity (Edinb) 113, 32–41.

Wang, J., 2014. Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. J. Evol. Biol. 27, 518–530.

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

Waples, R.S., Anderson, E.C., 2017. Purging putative siblings from population genetic data sets: a cautionary view. Mol. Ecol. 26, 1211–1224.

Whiteley, A.R., Coombs, J.A., Hudy, M., Robinson, Z., Nislow, K.H., Letcher, B.H., 2012. Sampling strategies for estimating brook trout effective population size. Conserv. Genet. 13, 625–637.

Willing, E.M., Dreyer, C., van Oosterhout, C., 2012. Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. PLoS ONE 7, e42649.

Xiao, R., Boehnke, M., 2009. Quantifying and correcting for the winner's curse in genetic association studies. Genet. Epidemiol. 33, 453–462.