Supporting Information

Rougeron et al. 10.1073/pnas.0904420106

SI Text

F Statistics for Measuring Inbreeding (see Fig. S1). We can define 6 fixation indices: F_{IS} (consanguinity of individuals relative to subpopulations), F_{SA} (of subpopulations relative to archipelagos), $F_{IA} = 1- (1 - F_{IS})(1 - F_{SA})$ (of individuals relative to archipelagos), F_{AT} (of archipelagos relative to the total population), $F_{ST} = 1- (1 - F_{SA})(1 - F_{AT})$ (of subpopulations relative to the total), and $F_{IT} = 1- (1 - F_{IS})(1 - F_{ST})$ (of individuals relative to the total population. A general expression for these indices can also be expressed (1, 2):

$$\begin{cases} F_{IS} = \dfrac{Q_i - Q_s}{1 - Q_s} \\ F_{SA} = \dfrac{Q_s - Q_a}{1 - Q_a} \\ F_{IA} = \dfrac{Q_i - Q_a}{1 - Q_a} \\ F_{AT} = \dfrac{Q_a - Q_T}{1 - Q_T} \\ F_{ST} = \dfrac{Q_s - Q_T}{1 - Q_T} \\ F_{IT} = \dfrac{Q_i - Q_T}{1 - Q_T} \end{cases} \quad [1]$$

If we focus on indices that measure the nested contribution to genetic differentiation at the different geographical levels (between subpopulations within archipelagos and between archipelagos), then only F_{SA} and F_{AT} are of interest. If we assume that Q_a is very small, it can be seen from Eq. **1** that F_{SA} will be big ($\approx Q_s$ if $Q_a \approx 0$). This is what is expected if the migration between subpopulations from the same archipelago is very small (≈ 0). If so, it is reasonable to assume that migration between archipelagos is not higher, and even lower, than between subpopulations from the same archipelago, and so $Q_T \approx 0$ and $F_{AT} \approx Q_a$. From this, it is easy to deduce that when differentiation is strong between subpopulations, it will necessarily be weak between higher levels. This is because most information of the apportionment of genetic information is contained in subpopulations. Biologically speaking, if migration is weak at one nested level, it is necessarily weak between higher levels also. This also can be seen by the fact that in the example described above, $F_{ST} \approx F_{SA}$.

Role of Gene Conversion (see Fig. S2). Gene conversion generates a transition from the heterozygous stage to the homozygous stage, and thus may result in large heterozygote deficiency (3). To test this hypothesis for our microsatellite loci, we used a generalized linear model to analyze the influence of length difference in number of bases between alleles (Δ) and subsamples (*subsample*) on the number of heterozygous loci (N_{Hz}). We tested this hypothesis with the software R (R Development Core Team 2008) using the following model: $N_{Hz} \approx subsample + \Delta$ with a generalized linear model (GLiM) with an expected Poisson distribution and a log link function. The GLiM method yields a significant overall statistical effect for only one locus, E11 ($P = 0.0001$). As it can be seen from Fig. 1, there is a negative significant relationship between the number of heterozygous loci (N_{Hz}) and the length difference in number of bases between alleles (Δ). E11 is located 60 bp before the gene of the trifunctional enzyme alpha subunit mitochondrial precursor-like, which could account for the significant effect observed for locus E11 because this locus is situated near a gene that could be subject to gene conversion. Our analyses without the locus E11 yield virtually the same population genetics results (such as important heterozygote deficit F_{IS} and significant linkage disequilibria) for all 12 loci.

Distribution of Heterozygous Loci Among Individuals in *Leishmania braziliensis* (see Fig. S3). Our findings indicate a large heterozygote deficiency, even within the clusters, found by BAPS. These heterozygote deficiencies can be explained by a mixed-mating regime with clonal fission and sexual recombination between individuals from the same (selfing) or different (outcrossing) strains (4). It can also be explained by a remaining Wahlund effect that BAPS cannot detect further because of sample sizes. Under the first hypothesis, we expect the distribution of heterozygous loci in each individual to be random (i.e., Poisson distribution), which is indeed the case (multinomial exact test, $P = 0.6$). According to a short simulation study undertaken with software EASYPOP 2.0.1 (5), our data are compatible with a population of 100 demes of 100 individuals with 0.008 migration rate, 50% of clonality, 47% of selfing, and 3% of outcrossing (other parameters available on request). Under the Wahlund effect hypothesis, we would have expected more skewed distributions with a different kurtosis, especially with high rates of clonality that are not compatible with the number of homozygous loci. However, very small subunits below the reach of BAPS, with much lower clonal rates, cannot be excluded.

1. Cockerham CC (1969) Variance of gene frequencies. *Evolution* 23:72–84.
2. Cockerham CC (1973) Analysis of gene frequencies. *Genetics* 74:679–700.
3. Raymond M, Rousset F (1995) Genepop (version 1.2). Population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249.
4. Regis-da-Silva CG, et al. (2006) Characterization of the *Trypanosoma cruzi* Rad51 gene and its role in recombination events associated with the parasite resistance to ionizing. *Mol Biochem Parasit* 149:191–200.
5. Balloux F (2001) EASYPOP (version 1.7): A computer program for population genetics simulations. *J Hered* 92:301–302.

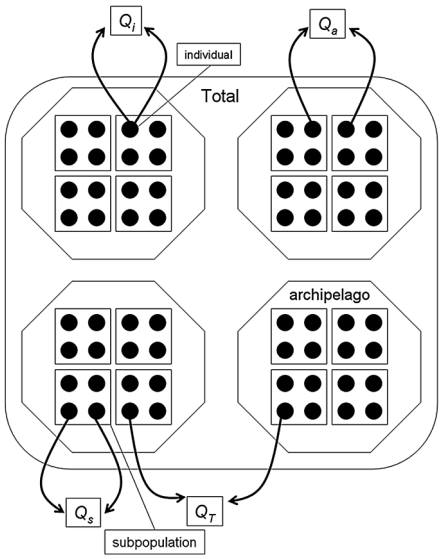

Fig. S1. In a 4-level structure (individuals, subpopulations, archipelagos, and total), 4 identities can be defined: Q_i, the probability that 2 alleles at one locus are identical within a randomly chosen individual; Q_s, the probability that 2 alleles at one locus, chosen from 2 random individuals from the same subpopulation, are identical; Q_a, the probability that 2 alleles from a given locus from 2 randomly chosen individuals from 2 randomly chosen subpopulations from one (random) archipelago are identical; and Q_T, the probability that 2 alleles at one locus from 2 randomly chosen individuals from 2 different subpopulations and from 2 different archipelagos are identical.

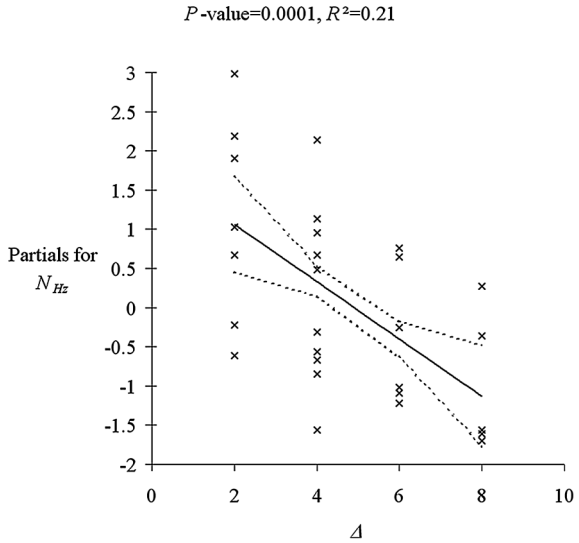

Fig. S2. Partial residuals for the number of heterozygous loci (N_{Hz}) as a function of length difference in number of bases between alleles (Δ). *P* represents the significance of Δ in the GLiM; R^2 represents its contribution to the total (null) residual deviation [Raymond M, Rousset F (1995) Genepop (version 1.2). Population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249].

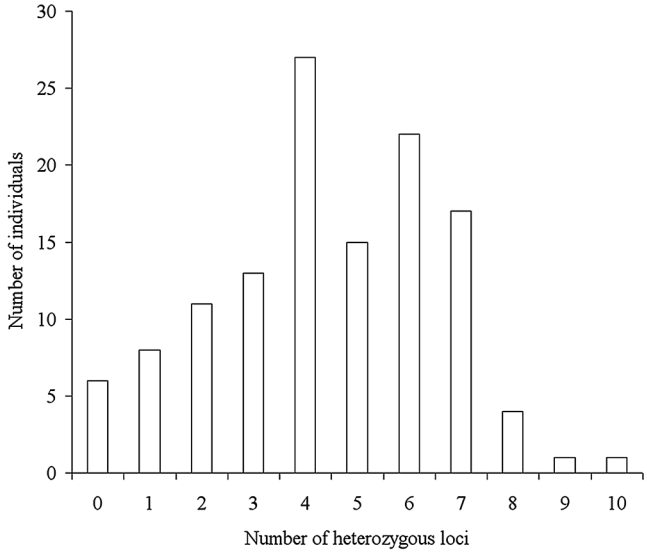

Fig. S3. Multinomial exact test undertaken with the program multinom [Regis-da-Silva CG, et al. (2006) Characterization of the *Trypanosoma cruzi* Rad51 gene and its role in recombination events associated with the parasite resistance to ionizing. *Mol Biochem Parasit* 149:191–200].