12-3-2014

Estimation of Inbreeding and Substructure Levels in African-Derived Brazilian Quilombo Populations

Renan B. Lemes
Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de São Paulo

Kelly Nunes
Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de São Paulo

Diogo Meyer
Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de São Paulo

Regina Célia Mingroni-Netto
Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de São Paulo

Paulo A. Otto
Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de São Paulo, otto@usp.br

Estimation of Inbreeding and Substructure Levels in African-Derived Brazilian Quilombo Populations

Renan B. Lemes[1], Kelly Nunes[1], Diogo Meyer[1], Regina Célia Mingroni-Netto[1], Paulo A. Otto[1]

[1]Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de São Paulo

Correspondence to: Paulo A. Otto, Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de São Paulo, Caixa Postal (POBox) 11.461, 05422-970 São Paulo SP Brazil. E-mail: otto@usp.br.

Key words: inbreeding, population isolates, quilombo remnants, substructure analysis

Abstract

The present paper deals with the estimation of inbreeding and substructure levels in a set of ten (later regrouped as eight) African-derived quilombo communities from the Ribeira River Valley in the southern portion of the state of São Paulo, Brazil. Inbreeding levels were assessed through F values estimated from the direct

analysis of genealogical data and from the statistical analysis of a large set of 30 molecular markers. The levels of population substructure found were modest, as well as the degree of inbreeding: in the set of all communities considered together, F values ranged from 0.00136 to 0.00248, when using raw and corrected data from their complete genealogical structures, respectively, to 0.027 to 0.036, when using the information taken from the statistical analysis of all 30 loci and of 14 loci of SNPs respectively. The overall frequency of consanguineous marriages in the set of all communities considered together was around 2%. Although modest, the values of the estimated parameters are much larger than those obtained for the overall Brazilian population and in general much smaller than the ones recorded for other Brazilian isolates. To circumvent problems related to heterogeneity sampling and virtual absence of reliable records of biological relationships we had to develop or adapt several methods for making valid estimates of the prescribed parameters.

Over three million Africans were brought to Brazil as slaves over a period of three hundred years. Runaway, abandoned, and freed slaves created small communities known as *quilombos*, the remnants of which in the state of São Paulo are confined to its southern border along the Ribeira River Valley (Figure 1). The region's relief afforded these communities a certain degree of geographical isolation. These settlements became traditional rural communities surviving on subsistence agriculture for many decades. Some drastic recent changes have taken place in the lifestyle of their inhabitants, traditional agriculture having been replaced by the cultivation of more commercially valuable products. This nutritional transition process has resulted in the high rates, among its inhabitants, of multifactorial (complex) diseases such as essential hypertension and obesity (Santos and Tatto, 2008; Pasinato and Rettl, 2009; Angeli *et al.*, 2011; Kimura *et al.*, 2012).

Quilombos have long been the subject of interest for population and evolutionary geneticists. They usually originate from a relatively small number of individuals (founder effect) and remain isolated over several generations, thus being subjected to the classical process of micro-differentiation due mainly to random genetic drift.

Many (but not all) isolates studied in Brazil and elsewhere (see Table 4 of section Results and Discussion) show detectable levels of inbreeding. This is measured by the average inbreeding coefficient **F** of its individuals or, as usually happens, using simplified methods that weigh the various inbreeding coefficients

of the progenies corresponding to the different types of marriages occurring in the population. As Cavalli-Sforza and Bodmer (1971, page 352) point out, "these inbreeding estimates take into account only easily detectable consanguinity, which rarely includes relationships more remote than third cousins." Therefore genealogical estimates of the mean inbreeding coefficient, in spite of being able to demonstrate the presence of consanguinity even at very modest rates, clearly constitute an underestimate of the real parameter value. More realistic estimates of consanguinity rates can be inferred from the population analysis of genetic markers (classical or molecular). The main problem with this strategy is that incredibly large samples are required in order to reveal statistically significant departures from p^2:2pq:q^2 Hardy-Weinberg equilibrium rates, as Figure 2 clearly shows. For instance, a sample size of about 1,500 individuals is necessary to detect a significant value of the inbreeding coefficient in an inbred population having a parameter value of **F = 0.05**. Another problem with **F** coefficients so estimated is that they should be differentiated from similar coefficients that might be spuriously interpreted as indicative of inbreeding and that commonly arise when the populations under study are hierarchically stratified (Wahlund's effect).

The primary objective of this paper is to provide estimates of inbreeding and of substructure levels from a set of ten quilombo communities. In order to circumvent problems related to the paucity of written and oral historical records and those related to heterogeneous molecular sampling (both detailed in the

section *subjects and methods* and also discussed in the *results* section) we had to develop or adapt several methods for obtaining reliable estimates of the prescribed parameters of inbreeding and population substructure. The presentation of these methodological variations is an important contribution of this report.

Subjects and Methods

1) Populations and subjects

Like most other quilombos in Brazil, the communities here presented were founded, in the last decades of the 19th century, by a relatively small number of runaway, abandoned, and freed African-derived slaves. Over the years the communities grew to include individuals from different ancestries (most of them African-derived, but also some Amerindians and admixed individuals with African and European ancestry). Given their proximity (most communities of the Ribeira River Valley are contiguous and within walking distance), relatively high levels of gene flow are expected to have occurred among the communities over the next five or six generations that have elapsed since their foundation. Taking all this into account, a relatively high degree of homogeneity is expected to be found among them, as well as a relatively low inbreeding level within them. Table 1 lists the present number of living individuals in each community and the corresponding numbers of individuals interviewed for assessing genealogical data

(per community) and of individuals molecularly genotyped (per locus and community).

The data from two pairs of communities (Galvão + São Pedro and Maria Rosa + Pilões) were grouped and analyzed together since they occupy adjacent territories, being basically formed by the same family groups.

This study was approved by the ethics committee of the Instituto de Ciências Biomédicas, Universidade de São Paulo. Informed consent was obtained from all participants in the study.

2) Genotype determination

Molecular (DNA markers) and genealogical data from the eight communities were obtained in different surveys organized and performed by members of the Laboratory of Human Genetics of our Department and partly reported in the following papers: Mingroni-Netto *et al.*, 2009a, 2009b; Cotrim *et al.*, 2004; Angeli *et al.*, 2005, 2011; Auricchio *et al.*, 2007; Yeh *et al.*, 2008; Kimura *et al.*, 2012, 2013.

Our analyses used data from 14 autosomal SNPs previously genotyped in our laboratory (for details on methodology, see Angeli *et al.*, 2011 and Kimura *et al.*, 2012): *ACE* (rs1799752), *NOS3* (rs1799983), *GNB3* (rs5443), *GNB3* (rs5441), *AGT* (rs669), *ADD2* (rs3755351), *GRK4* (rs1801058), *PLIN1* (rs2289487),

INSIG2 (rs7566605), *LEP* (rs2167270), *LEPR* (rs1137101), *ADRB2* (rs1042713), *PPARG* (rs1801282), and *RETN* (rs1862513).

Using DNA samples from some 300 individuals of the communities, we determined the genotypes of the following 16 autosomal microsatellite loci: *D1S551*, *D4S3248*, *D5S816*, *D6S1040*, *D7S821*, *D7S3061*, *D8S2324*, *D9S301*, *D9S922*, *D10S1426*, *D13S317*, *D16S539*, *D18S535*, *D19S559*, *D20S482*, and *D21S1437*. The primer sequences were generated using software *Primer3* (Rozen and Scalestsky, 2000) and the forward sequences were marked with fluorescence (Supplementary Table 1). Microsatellite genotypes were determined by polymerase chain reaction (PCR) in four multiplex systems submitted to capillary electrophoresis on *ABI 3730 DNA analyzer* (*Applied Biosystems,* Foster City, USA). All analyses were carried out using the *Peak Scanner™ v1.0* software (also from *Applied Biosystems*).

Different groups of individuals were selected for determination of molecular markers on different occasions with distinct purposes: the first set of seven SNP markers out of the 14 listed above were used primarily in association studies with arterial hypertension and the last seven in association studies with obesity. As a result, data for each set of marker only partially overlaps, introducing an additional source of variation, leading us to expect to find a significant degree of heterogeneity among loci and populations.

3) Genealogical data

Genealogical analysis of data based on detailed interviews provided information for about 2,000 individuals, which allowed us to estimate a mean inbreeding coefficient or fixation index ($\mathbf{F_G}$) for each community and in the set of all communities.

Our analysis included all living individuals who were born in a given community. We also considered as belonging to a given community migrant individuals who had offspring with native quilombo individuals from that community. Information from deceased individuals was used only to assess biological relationships among individuals within communities.

The total number of inhabitants and individuals interviewed (2641 and 1879 respectively) varied from 573 to 184 and 364 to 148 per community; the total number of genotype determinations varied from 788 to 207 in relation to different loci in the total population (see Table 1).

The quilombo communities here studied were isolated for a long period of time with paucity of historical records (written or oral) of biological relationships. In order to correct or decrease this bias, average inbreeding coefficients (per community and for the set of all communities grouped together), in addition to being estimated using all available information, were assessed just from individuals that possessed double-checked information on his ascendants over at least two generations. From the total of 3,959 individuals represented in the

genealogies, 2,171 provided complete information on their ascendants over at least two generations; just 794 among them had reliable information (in order to establish the presence of eventual biological relationships) on at least half of their great-grandparents; and less than 100 individuals had reliable information for all their great-grandparents.

4) Quantitative analyses

4.1) Genealogical analysis

Genealogical estimates of the mean inbreeding coefficient (fixation index $\mathbf{F_G}$) for each community and in the set of all communities were obtained by averaging the individual inbreeding coefficients ($\mathbf{f_G}$) from all individuals represented in the genealogies and from a subsample of individuals that possessed information on their ascendants over at least two generations. The values of each $\mathbf{f_G}$ were obtained by the usual Wright's (1922) formula $\mathbf{f_G = \Sigma[1/2^n.(1+f_A)]}$, in which **n** is the number of individuals between the parental pair and the common ancestor, including these three individuals, and $\mathbf{f_A}$ is the inbreeding coefficient of the common ancestor of the parental pair.

4.2) Molecular markers data analysis

Reliable estimates of genotype and allele frequencies and of the average inbreeding coefficient (Wright's fixation index) $\mathbf{F = 1-\Sigma P(a_i a_j)/(2\Sigma p_i p_j)}$, that

reduces to **F = 1 - P(Aa)/(2pq)** in the 2-allele case, were obtained through programs developed in a Windows-based structured BASIC dialect (Liberty BASIC v4.04, © Shoptalk Systems) and using the package of mathematical routines Mathematica V. 8.0.4.0 (© Wolfram Research). By means of chi-squared tests and bootstrap simulation techniques, these programs test the samples for departures of Hardy-Weinberg ratios, estimate their corresponding fixation index values, construct "exact" confidence intervals for them, and perform appropriate substructure analyses.

Mean values of **F** for the whole population in relation to each locus were obtained by adding together the corresponding data of all communities; in the case of the set of all loci per population or in the set of all populations, average figures of **F** were estimated by the usual method of combining them by the reciprocal values of their corresponding variances:

$$\mathbf{F = \sum [F_i/var(F_i)] \,/\, \sum [1/var(F_i)]} \ ,$$

with i varying from 1 to the number of different loci.

The appropriate estimation of the variance of the inbreeding coefficient **var(F)** is a complicated issue and the formula derived by Fyfe and Bailey (1951) for the case of 2 autosomal alleles is generally used:

$$\mathbf{var(F) = (1-F)^2(1-2F)/N + F(1-F)(2-F)/[2Np(1-p)]} \ ,$$

in which $\mathbf{p = P(A) = [2N(AA)+N(Aa)]/2N}$, $\mathbf{F = 1 - [N(Aa)/N]/[2p(1-p)]}$, $\mathbf{N = N(AA) + N(Aa) + N(aa)}$, and **A** and **a** are a pair of alleles segregating in an autosomal locus.

We were able to derive a different formula for the variance of **F** whose numerical values for the two-allele case are virtually the same as those obtained using either the formula proposed by Fyfe and Bailey (1951) or the average population values estimated by simulations using bootstrapping techniques. Our formula is expressed in the two-allele case by the equation

$$\mathbf{var(F) = N_1.N_2.N_3/[(Npq)^2.(N_2.N_3+4.N_1.N_3+N_1.N_2)]}$$

$$\mathbf{= (1-F)(p+qF)(q+pF)/[Npq(1+F)]} \ ,$$

where $\mathbf{N_1 = N(AA)}$, $\mathbf{N_2 = N(Aa)}$, $\mathbf{N_3 = N(aa)}$, $\mathbf{N = N_1+N_2+N_3}$, $\mathbf{p = 1-q = (2N_1+N_2)/2N}$, and $\mathbf{F = 1 - N_2/2pq}$.

It is possible, unlike what happens to Fyfe and Bailey's formula, to adapt it to the generalized case of any number of alleles segregating at an autosomal locus. The subject has theoretical interest; mathematical details about its derivation and properties will be published and discussed elsewhere.

In order to determine which values of **F** could be considered as outliers and should be excluded from a global analysis, we proceeded as follows: on the long run the various per locus estimates of **F** inside a same community are expected to be normally distributed around the average **F** value for that community, so that the outlier values should be outside the usual 95% range

$\mathbf{F \pm 1.96 \sqrt{[var(F)]}}$,

where $\mathbf{F = \Sigma x_i F_i}$, $\mathbf{var(F) = \Sigma x_i F_i^2 - F^2}$ and

$\mathbf{x_i = var^{-1}(F_i)/\Sigma_{(j=1,n)} var^{-1}(F_j)}$.

"Exact" 95% confidence intervals for the estimated values of the mean inbreeding coefficient (fixation index) **F** were obtained for each combination locus/community through 1,000 computer-assisted bootstrap simulations of samples, each of them having the same size and genotypic proportions observed in the actual one. A similar approach with variations was used to construct the confidence intervals of Wright's substructure indexes $\mathbf{F_{ST}}$, $\mathbf{F_{IT}}$ and $\mathbf{F_{IS}}$.

For the substructure analysis, we recoded the microsatellite markers as biallelic, where the first allele corresponds to the allele with the highest frequency in the population and the second allele as being equivalent to the total of the remaining alleles.

In order to circumvent problems related to heterogeneous sampling of loci and communities, besides performing the analyses detailed above in the whole data set (considering all genotyped individuals), we repeated the procedures using

a sub-sample containing only individuals genotyped for all loci. Since with this strategy the sample size dropped to only 87 individuals (Table S2), we also used a sub-sample containing all individuals who were genotyped for at least 27 out of the 30 marker systems, resulting in a sample of 207 individuals (Table S3). To take into account the issue of the different nature of the sets of molecular markers used, we estimated all parameters in relation to SNPs and microsatellites separately.

Results and Discussion

1) Genealogical analysis

Table 2 lists the estimated values of the inbreeding coefficient ($\mathbf{F_G}$) from the genealogical analysis of the eight communities considered separately and together, taking into account the data from all 3,959 individuals with genealogical information. Table 3 lists the same values estimated from the set of 2,171 individuals who had complete information about his ascendants over at least two generations. Unlike other estimates derived from genealogical analysis, that calculate the population **F** value weighing the different **F** values by the mean sizes of the sibships from which they were estimated, our **F** estimate is the average value of the parameter estimated for each living individual of the population.

Before applying our methodology to the quilombos reported here, we tested its performance by applying it to the published genealogical structure of the

quilombo isolate of Valongo (Souza and Culpi, 1992) in the southern state of Santa Catarina (Figure S1 supplementary), founded by just four couples and where the frequency of consanguineous unions is 85%. We obtained the estimate $\mathbf{F_G = 0.0457}$ for the whole community, a value that is not significantly different from the estimate of **0.0477** obtained by Souza and Culpi (1992) using the formula $\mathbf{F = 2(N_r-1)/[2N_e-(2N_e-1)(1-m_e)^2]}$, where $\mathbf{N_r}$ is the breeding population size, $\mathbf{N_e = 2(N_r-1)/(k-1+\sigma_k^2/k)}$ is the effective population size, $\mathbf{m_e}$ is the effective migration rate, and **k** is the average offspring size in the breeding population.

The estimated values of **F** for the set of all communities grouped together range from **0.00136** (considering all individuals) to **0.00248** (considering only the subset of 2,171 individuals with more reliable information). These values are approximately 1.5 to 3 times higher than the corresponding estimate for the total Brazilian population (**F = 0.00088**) and about 2 to 4 times higher than the estimate for the population of the state of São Paulo (**F = 0.00067**) (Freire-Maia, 1957; 1990). As Tables 2 and 3 clearly shows, the community values of **F** ranged from zero in two aggregates to **0.00344** or **0.00699** in the population of Abobral (AB).

As already commented, the values of $\mathbf{F_G}$ in the quilombos reported here surely are underestimates of the true values due to many factors, such as lack of information on many branches of the genealogies and generalized absence of reliable records as to the origin of the populations as well as to biological relationships among their members. In any case, the strategy of reassessing the

parameter in the subsample containing only individuals with more reliable information was able to partially eliminate this bias.

Table 4 compares our estimates of both inbreeding coefficient and the frequency of consanguineous marriages with the results from isolate surveys in the literature. With the exception of the Brazilian Jewish isolate studied by Freire-Maia and Krieger (1963), all other communities listed in this table show relatively large values of **F,** almost always associated with substantial levels of consanguineous unions, unlike our results shown in Tables 2 and 3.

The strikingly high inbreeding levels of Valongo quilombo are perfectly compatible with the fact that the community is presently composed by less than 100 individuals, all originated from only four founding couples. Unlike this community, the whole isolate of the Ribeira River Valley has more than 2,500 adult individuals. Its size, together with other factors already referred to on section "Subjects and methods", probably account for the unusually low inbreeding levels detected in the isolate here reported.

2) Molecular marker analysis

Our analysis of a set of independent autosomal loci provided us with estimates of both mean **F** values for the individual quilombo communities as well as all of them together, in relation to each locus and for the set of all loci considered

together. Outlier values, determined using the method described in section Subjects and Methods, were not considered for any calculations.

Considering the frequency of **P** values less than the critical figure of 0.05, only in six out of a total of 239 combinations (around 2.5 per cent) of locus/community was the hypothesis of $p^2 : 2pq : q^2$ ratios of Hardy-Weinberg equilibrium rejected, which is slightly less than the expected proportion by chance in the long run. When all quilombo communities were considered together, the genotype frequencies at two out of 30 *loci* (around 6.7 per cent) deviated significantly from Hardy-Weinberg ratios at the same rejection level of 5%, which clearly indicates just a non-significant excess of positive results. Including the data obtained from pooling, per locus, all communities together, a total of approximately 250 tests for verifying the hypothesis **F = 0** were performed. A Bonferroni-type correction of our data will show that none of the tests produced a significant **P** value.

Table 5 summarizes the results for each isolate and for the set of all communities considered together, in relation to (1) the set of 16 microsatellite markers, (2) the set of 14 SNPs, and (3) all loci considered together. Table 6 shows the results for the analysis of a dataset containing all individuals that were genotyped for at least 27 out of the 30 markers. Unlike what happens when only the SNPs are used, the average **F** estimates using microsatellite data have negative values for practically all communities. This is especially noted when the sample

sizes are drastically reduced in order to minimize data heterogeneity (Table 6), and it is known from sampling theory that small sized samples favor the occurrence of heterozygous individuals (see Cannings and Edwards, 1969). This should be critical when the number of segregating alleles is high, a situation in which most sampled individuals will be heterozygous even under panmictic expectations. Summing it up, the estimates using biallelic markers such as autosomal SNPs seem to be more reliable than the ones using microsatellites or the set of all markers. Therefore, our analysis using adequate molecular markers (SNPs) indicate average figures of the mean inbreeding coefficient ranging from about **0.036** (using data from all sampled individuals) to **0.055** (using the more homogeneous data from individuals that were genotyped for at least 27 different markers).

3) Population substructure analysis

Genealogical relations among individuals from different quilombo communities of the Ribeira Valley exist to a certain degree, since the founders of some of these population aggregates are likely to be the same, as indicated by the sharing of some common surnames. This fact and the physical proximity of the different communities (as Figure 1 shows, most are contiguous, within walking distance, the furthest away lies less than 20km apart) suggest *a priori* a modest level of substructure among these communities.

Table 7 presents the values of the fixation indexes ($\mathbf{F_{IT}}$, $\mathbf{F_{ST}}$, and $\mathbf{F_{IS}}$) obtained from all 30 loci for the set of all quilombo communities. Simulations by means of bootstrap techniques, using all data (but also excluding outliers), generated reliable estimates of the 95% confidence interval for each one of these fixation indexes. When the lower and upper limits of a 95% confidence interval of $\mathbf{F_{IT}}$ or $\mathbf{F_{IS}}$ thus constructed have different signs it is assumed that the corresponding fixation indexes are not significantly different from zero at the rejection level of 5%. Since $\mathbf{F_{ST}}$ indexes are always obtained from the relation **var(p)/(pq)** and all three quantities in the formula belong to the domain of positive numbers, the numerical value of the parameter as well as all the values contained in its corresponding confidence interval will be positive. Inferences regarding the significance of $\mathbf{F_{ST}}$ (is $\mathbf{F_{ST}}$ significantly different from zero?) are then obtained indirectly from the behavior of the corresponding confidence intervals of both $\mathbf{F_{IT}}$ and $\mathbf{F_{IS}}$: in all instances in which $\mathbf{F_{IS}}$ is not different from zero, $\mathbf{F_{IT}}$ is not different from $\mathbf{F_{ST}}$; therefore, in all cases in which both $\mathbf{F_{IT}}$ and $\mathbf{F_{IS}}$ are not different from zero, $\mathbf{F_{ST}}$ is also not statistically different from zero. The very few instances in which this did not take place are indicated by $\mathbf{F_{ST}}$ values in bold face on Table 7 and should be interpreted as cases in which we can assume unambiguously that the index is different from zero.

The $\mathbf{F_{ST}}$ values were in general very small, a finding already detected for these same populations in a study using INDEL molecular markers by Kimura *et*

al. (2013). This suggests the existence of a significant amount of gene flow or recent shared ancestry, with little time for differentiation between the subpopulations.

What is important and immediately assumed from the mere inspection of Table 7 is that, with exception of locus *ACE* (rs1799752), in the few instances in which the $\mathbf{F_{ST}}$ was significantly different from zero, the proportionate contribution of $\mathbf{F_{ST}}$ to the $\mathbf{F_{IT}}$ index was always much smaller than the one for $\mathbf{F_{IS}}$. The dubious results obtained in relation to locus *PLIN1* were caused by extremely high **F** values in three out of the seven communities that resisted to the process of outlier cleaning, a behavior for which we have no logical explanation.

In spite of the difficulties brought about by the sets of genealogical as well as molecular data, our results indicate that the levels of substructure among the quilombo communities are negligible or at least very small, probably a consequence of gene flow and shared history among communities. This finding legitimizes the genealogical and molecular estimations of the fixation index we performed by considering the set of communities as a whole.

Acknowledgements

This work was partially supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq), Brazil. The comments, suggestions, and corrections from three anonymous referees from the editorial board of Human Biology are deeply acknowledged. Many of them were included in the final (corrected) version of this paper. We also thank Maria Teresa B. M. Auricchio for technical support and Ms. Lilian Dluhosch for reading critically the manuscript.

Literature Cited

Abney M, McPeek MS, Ober C. 2000. Estimation of variance components of quantitative traits in inbred populations. *Am. J. Hum. Genet.* 66: 629–650.

Angeli CB, Capelli LP, Auricchio MTBM et al. 2005. AGG interspersion patterns in the CGG repeat of the FMR1 gene and linked DXS548/FRAXAC1 haplotypes in Brazilian populations. *Am. J. Med. Genet.* 132A: 210-214.

Angeli CB, Kimura L, Auricchio MTBM et al. 2011. Multilocus analyses of seven candidate genes suggest interacting pathways for obesity-related traits in Brazilian populations. *Obesity.* 19: 1244–1251.

Auricchio MTBM, Vicente JP, Meyer D et al. 2007. Frequency and origins of hemoglobin S mutation in African-derived Brazilian populations. *Hum. Biol.* 79: 667–678.

Bittles AH. 2002. Endogamy, consanguinity and community genetics. *J. Genet.* 81: 91–98.

Cannings C, Edwards AW. 1969. Expected genotypic frequencies in a small sample: deviation from Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* 21: 245–247.

Cavalli-Sforza LL, Bodmer WF. 1971. *The Genetics of Human Populations.* W. H. Freeman, San Francisco.

Cotrim NH, Auricchio MT, Vicente JP et al. 2004. Polymorphic ALU insertion in six Brazilian African-derived populations. *Am. J. Hum. Biol.* 16: 264–277.

Dorsten LE, Hotchkiss L, King TM. 1999. The effect of inbreeding on early childhood mortality: twelve generations of an Amish settlement. *Demography.* 36: 263–271.

Ellis WS, Starmer WT. 1978. Inbreeding as measured by isonymy, pedigrees, and population size in Törbel, Switzerland. *Am. J. Hum. Genet.* 30: 366–376.

Freire-Maia N. 1957. Inbreeding in Brazil. *Am. J. Hum. Genet.* 9: 284–298.

Freire-Maia N. 1990. Genetic effects in Brazilian populations due to consanguineous marriages. *Am. J. Med. Genet.* 35: 115–117.

Freire-Maia N, Krieger H. 1963. A Jewish isolate in southern Brazil. *Ann. Hum. Genet. Lond.* 27: 31–39.

Fyfe JL, Bailey NT. 1951. Plant breeding studies in leguminous forage crops I. Natural cross-breeding in winter beans. *J. Agric. Sci.* 41: 371–378.

Hamamy H, Jamhawi L, Al-Darawsheh J et al. 2005. Consanguineous marriages in Jordan: why is the rate changing with time? *Clin. Genet.* 67: 511–516.

Jackson CE, Symon WE, Pruden EL et al. 1968. Consanguinity and Blood Group Distribution in an Amish Isolate. *Am. J. Hum. Genet.* 20: 522–527.

Kimura L, Angeli CB, Auricchio MTBM et al. 2012. Multilocus family-based association analysis of seven candidate polymorphisms with essential hypertension in an African-derived semi-isolated Brazilian population. *Am. J. Hypertens.* 2012: 859219.

Kimura L, Ribeiro-Rodrigues EM, Auricchio MTBM et al. 2013. Genomic ancestry of rural African-derived populations from southeastern Brazil. *Am. J. Hum. Genet.* 25: 35–41.

Mingroni-Netto RC, Angeli CB, Kimura L et al. 2009a. Doenças modernas nos antigos quilombos: a obesidade e a hipertensão no Vale do Ribeira. In *Saúde nos quilombos*, Volochko A and Batista LE, orgs. Instituto da Saúde, São Paulo, Brazil.

Mingroni-Netto RC, Auricchio MTBM, Vicente JP. 2009b. Importância da pesquisa do traço e da anemia falciforme nos remanescentes de quilombos do Vale do Ribeira-SP. In *Saúde nos quilombos*, Volochko A and Batista LE, orgs. Instituto da Saúde, São Paulo, Brazil.

Pasinato R, Rettl KI. 2009. Desenvolvimento local sustentável: a contribuição das comunidades quilombolas do Vale do Ribeira. In *Saúde nos quilombos*, Volochko A and Batista LE, orgs. Instituto da Saúde, São Paulo, Brazil.

Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Krawetz S and Misener S, eds. Humana Press, Totowa, NJ.

Santos KMP, Tatto N, eds. 2008. *Agenda socioambiental de comunidades quilombolas do Vale do Ribeira.* Ipsis Gráfica e Editora, São Paulo.

Souza IR, Culpi L. 1992. Valongo, an isolated Brazilian Black community. I. Structure of the population. *Brazil. J. Genet.* 15: 439–447.

Wright S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* 56: 330–338.

Yeh E, Kimura L, Errera FI et al. 2008. Association of polymorphisms at the ADIPOR1 regulatory region with type-2 diabetes and body mass index in a Brazilian population with European or African ancestry. *Braz. J. Med. Biol. Res.* 41: 468–472.

Table 1: N = estimated number of adult individuals (Auricchio *et al.*, 2007), NG = number of individuals interviewed for gathering genealogical data. The other cells of the table show the numbers of genotyped individuals for each molecular marker (identified at the leftmost column) at a given locality.

	AB	**AN**	**GA/SP**	**IV**	**MR/PS**	**NH**	**PC**	**TU**	**Total**
N	573	320	266	270	184	447	286	295	2641
NG	364	247	224	217	148	237	263	179	1879
***ACE* (rs1799752)**	96	86	99	77	55	89	78	56	636
***NOS3* (rs1799983)**	59	79	92	76	30	67	78	56	537
***GNB3* (rs5443)**	95	78	98	77	39	67	76	56	586
***GNB3* (rs5441)**	93	65	94	62	54	84	77	66	595
***AGT* (rs669)**	58	48	92	76	30	63	78	56	501
***ADD2* (rs3755351)**	92	75	90	76	45	48	73	59	558
***GRK4* (rs1801058)**	91	85	97	75	52	86	77	72	635

***PLIN1* (rs2289487)**	93	108	115	128	64	109	93	78	788
***INSIG2* (rs7566605)**	93	103	112	125	65	102	93	79	772
***LEP* (rs2167270)**	94	106	114	116	61	109	92	80	772
***LEPR* (rs1137101)**	94	107	115	116	60	109	91	79	771
***ADRB2* (rs1042713)**	95	102	111	110	61	104	91	78	752
***PPARG* (rs1801282)**	93	103	115	102	61	106	93	80	753
***RETN* (rs1862513)**	89	105	113	126	65	104	91	76	769
D1S551	36	24	34	51	37	41	39	28	290
D4S3248	36	24	34	50	37	41	39	28	289
D5S816	36	25	34	51	37	41	39	28	291

Table 1 (Contd.)**:**

	AB	**AN**	**GA/SP**	**IV**	**MR/PS**	**NH**	**PC**	**TU**	**Total**
D6S1040	35	22	33	52	37	43	39	31	292
D7S821	36	24	34	51	37	41	39	28	290
D7S3061	36	24	34	51	37	41	39	28	290
D8S2324	36	22	34	52	37	43	39	31	294
D9S301	37	23	34	52	37	43	39	31	296
D9S922	36	24	34	51	36	41	39	28	289
D10S1426	29	18	30	49	34	39	38	25	262
D13S317	37	22	34	52	37	43	39	31	295
D16S539	36	24	34	51	37	41	39	28	290
D18S535	37	23	34	52	37	43	39	31	296
D19S559	36	22	34	52	37	43	39	31	294

D20S482	37	23	34	52	37	43	39	31	296
D21S1437	32	19	22	38	26	34	25	11	207

Table 2: Estimated values of F obtained through genealogical analysis. N: number of individuals included in the analyses; F_G: estimated value of the inbreeding coefficient; %cm: observed frequencies of consanguineous marriages (in percentages); AB,...,TU: identification of communities.

Community	**N**	**F_G**	**%cm**
AB	773	0.00344	3.63
AN	567	0.00245	2.31
GA/SP	446	0.00070	1.72
IV	575	0.00033	0.63
MR/PS	324	0.00024	0.88
NH	434	0.00176	5.26
PC	368	0	0
TU	472	0	0
Total	**3959**	**0.00136**	**1.87**

Table 3: Estimated values of F obtained through genealogical analysis. N: number of individuals who had complete information about his ascendants over at least two generations; F_G: estimated value of the inbreeding coefficient; %cm: observed frequencies of consanguineous marriages (in percentages); AB,...,TU: identification of communities.

Community	**N**	**F_G**	**%cm**
AB	380	0.00699	8.18
AN	383	0.00363	5.68
GA/SP	235	0.00133	4.76
IV	288	0.00065	1.47
MR/PS	152	0.00052	2.22
NH	221	0.00346	13.95
PC	368	0	0
TU	472	0	0
Total	**2171**	**0.00248**	**4.58**

Table 4: Estimates of the mean inbreeding coefficient (**F**) and percentage of consanguineous marriages (%cm) from several isolates reported in the literature.

Population	**F**	**%cm**	**Reference**
Jewish isolate from Curitiba (Brazil)	0.0013	4.0	Freire-Maia and Krieger (1963)
Amish of Adams county (USA)	0.0195	66.5	Jackson *et al.* (1968)
Törbel (Switzerland)	0.0058	-	Ellis and Starmer (1978)
Quilombo of Valongo (Brazil)	0.0477	85.0	Souza and Culpi (1992)
Amish of Lancaster (USA)	0.0166	-	Dorsten *et al.* (1999)
Hutterites of South Dakota (USA)	0.0340	-	Abney *et al.* (2000)
India	0.0075	11.9	Bittles (2002)
South of India	0.0212	31.0	Bittles (2002)
Amman (Jordan)	0.0142	28.4	Hamamy *et al.* (2005)
Quilombo of Ribeira River Valley (Brazil)	0.0025	4.6	Present study

Table 5: Average F values and corresponding 95% confidence intervals (per community and in the total population, considering all genotyped individuals) in relation to microsatellites, SNPs and all markers together. AB,...,TU: identification of communities.

Community	**Microsatellites**	**SNPs**	**All markers**
AB	-0.010 (-0.104, 0.085)	0.020 (-0.151, 0.192)	0.011 (-0.149, 0.171)
AN	-0.042 (-0.244, 0.160)	0.003 (-0.113, 0.119)	-0.002 (-0.132, 0.129)
GA/SP	-0.138 (-0.225, -0.052)	0.045 (-0.145, 0.235)	-0.057 (-0.226, 0.112)
IV	-0.051 (-0.176, 0.074)	-0.006 (-0.249, 0.236)	-0.014 (-0.239, 0.211)
MR/PS	-0.036 (-0.157, 0.086)	0.060 (-0.247, 0.366)	0.031 (-0.246, 0.309)
NH	-0.064 (-0.117, -0.010)	-0.051 (-0.206, 0.105)	-0.059 (-0.169, 0.052)
PC	-0.041 (-0.060, -0.021)	-0.037 (-0.180, 0.106)	-0.035 (-0.117, 0.047)
TU	-0.028 (-0.149, 0.094)	0.001 (-0.231, 0.232)	-0.002 (-0.223, 0.218)
Total	**-0.002 (-0.064, 0.060)**	**0.036 (-0.049, 0.121)**	**0.022 (-0.050, 0.093)**

Table 6: Average F values and corresponding 95% confidence intervals (per community and in the total population), considering only individuals genotyped as to at least 90% of all markers. AB,...,TU: identification of communities.

Community	**Microsatellites**	**SNPs**	**All markers**
AB	-0.071 (-0.101, -0.042)	-0.013 (-0.166, 0.140)	-0,057 (-0.140, 0.026)
AN	-0.049 (-0.272, 0.175)	-0.035 (-0.323, 0.253)	-0.039 (-0.309, 0.230)
GA/SP	-0.065 (-0.138, 0.009)	0.017 (-0.183, 0.216)	-0.078 (-0.249, 0.093)
IV	-0.031 (-0.105, 0.043)	-0.045 (-0.288, 0.198)	-0.013 (-0.195, 0.170)
MR/PS	-0.057 (-0.151, 0.038)	-0.069 (-0.348, 0.209)	-0.038 (-0.273, 0.197)
NH	-0.089 (-0.227, 0.050)	0.059 (-0.286, 0.404)	-0.053 (-0.238, 0.133)
PC	-0.104 (-0.204, -0.005)	0.011 (-0.298, 0.321)	-0.065 (-0.242, 0.111)
TU	-0.049 (-0.224, 0.127)	0.005 (-0.322, 0.332)	0.001 (-0.277, 0.280)
Total	**-0.024 (-0.467, 0.419)**	**0.055 (-0.464, 0.575)**	**0.013 (-0.167, 0.192)**

Table 7: Estimates of fixation indexes (F_{IT}, F_{ST} and F_{IS}) and corresponding 95% confidence intervals.

	F_{IT}		F_{ST}		F_{IS}	
***ACE* (rs1799752)**	**0.097**	(0.014, 0.179)	**0.045**	(0.029, 0.076)	0.054	(-0.032, 0.128)
***NOS3* (rs1799983)**	0.054	(-0.048, 0.163)	0.021	(0.011, 0.051)	0.033	(-0.067, 0.132)
***GNB3* (rs5443)**	0.030	(-0.058, 0.110)	0.037	(0.022, 0.067)	-0.007	(-0.096, 0.063)
***GNB3* (rs5441)**	0.085	(-0.013, 0.175)	0.025	(0.011, 0.057)	0.062	(-0.046, 0.151)
***AGT* (rs669)**	-0.028	(-0.118, 0.069)	0.013	(0.005, 0.039)	-0.041	(-0.137, 0.052)
***ADD2* (rs3755351)**	0.062	(-0.027, 0.147)	0.020	(0.011, 0.047)	0.043	(-0.053, 0.118)
***GRK4* (rs1801058)**	0.018	(-0.061, 0.102)	0.015	(0.008, 0.038)	0.003	(-0.082, 0.083)
***PLIN1* (rs2289487)**	***0.104***	(0.026, 0.172)	***0.031***	(0.018, 0.056)	***0.075***	(-0.006, 0.139)
***INSIG2* (rs7566605)**	0.002	(-0.077, 0.076)	0.153	(0.008, 0.036)	-0.014	(-0.099, 0.058)
***LEP* (rs2167270)**	0.017	(-0.058, 0.089)	0.023	(0.012, 0.045)	- 0.006	(-0.082, 0.064)
***LEPR* (rs1137101)**	0.001	(-0.063, 0.068)	0.032	(0.021, 0.055)	-0.033	(-0.103, 0.031)
***ADRB2* (rs1042713)**	-0.034	(-0.113, 0.046)	0.027	(0.014, 0.053)	-0.063	(-0.152, 0.014)

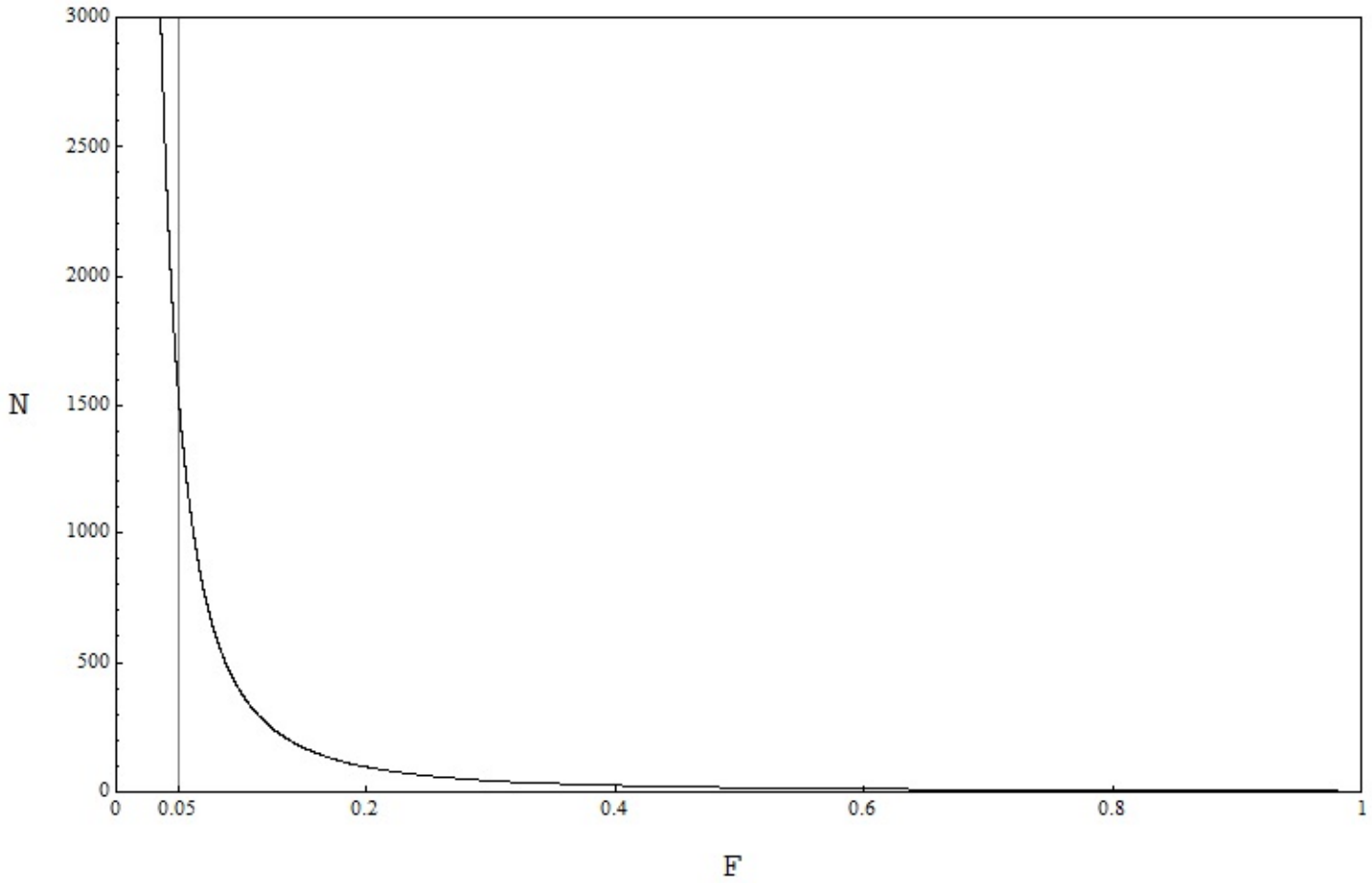
***PPARG* (rs1801282)**	0.056	(-0.013, 0.140)	0.061	(0.037, 0.103)	-0.002	(-0.074, 0.065)
***RETN* (rs1862513)**	-0.004	(-0.071, 0.065)	0.015	(0.009, 0.034)	-0.019	(-0.092, 0.046)
D5S816	***-0.122***	(-0.219, -0.029)	***0.001***	(0.003, 0.028)	***-0.123***	(-0.231, -0.041)
D1S551	0.097	(-0.014, 0.207)	0.024	(0.014, 0.068)	0.075	(-0.049, 0.174)
D7S3061	0.092	(-0.030, 0.209)	0.007	(0.005, 0.045)	0.086	(-0.046, 0.190)
D4S3248	0.067	(-0.056, 0.186)	0.012	(0.007, 0.049)	0.056	(-0.081, 0.160)
D16S539	-0.015	(-0.122, 0.098)	0.011	(0.006, 0.047)	-0.026	(-0.149, 0.073)
D9S922	-0.062	(-0.182, 0.045)	0.018	(0.010, 0.057)	-0.082	(-0.215, 0.013)
D10S1426	0.047	(-0.102, 0.180)	0.054	(0.030, 0.115)	-0.007	(-0.168, 0.118)

Table 7 (Contd.)**:**

	F_{IT}		F_{ST}		F_{IS}	
D7S821	-0.087	(-0.195, 0.023)	0.011	(0.006, 0.046)	**-0.099**	(-0.220, -0.009)
D13S317	0.017	(-0.089, 0.131)	0.033	(0.021, 0.078)	-0.016	(-0.140, 0.089)
D8S2324	0.106	(-0.032, 0.251)	0.013	(0.006, 0.054)	0.095	(-0.058, 0.230)
D19S559	-0.007	(-0.131, 0.112)	0.018	(0.009, 0.057)	-0.026	(-0.164, 0.083)
D6S1040	-0.077	(-0.202, 0.039)	0.006	(0.004, 0.036)	-0.084	(-0.218, 0.018)
D20S482	0.111	(-0.012, 0.229)	0.022	(0.010, 0.074)	0.090	(-0.048, 0.195)
D21S1437	***0.197***	(0.015, 0.347)	***0.026***	(0.010, 0.097)	*0.175*	(-0.017, 0.324)
D9S301	-0.023	(-0.139, 0.080)	0.035	(0.021, 0.081)	-0.061	(-0.188, 0.035)
D18S535	-0.021	(-0.140, 0.092)	0.007	(0.005, 0.038)	-0.028	(-0.158, 0.072)

Figure 1: (A) State of São Paulo highlighted within the Brazilian territory; (B) location of both Ribeira Valley region in São Paulo (gray area) and (in black) the municipalities of Eldorado (EL) and Iporanga (IP), in which territory the ten quilombo communities shown in C are located (from Kimura *et al.*, 2013): Abobral (AB), Maria Rosa (MR), Pilões (PS), Galvão (GA), São Pedro (SP), Pedro Cubas (PC), Ivaporanduva (IV), Sapatu (TU), André Lopes (AN), and Nhunguara (NH).

Figure 2. Sample sizes (ordinate axis) required for obtaining statistical significance of **F** values (abscissa axis) at the rejection level of 5%. The gray line corresponds to an **F** value of 0.05.



SUPPLEMENTARY MATERIAL

Figure S1: Genealogy of quilombo from Valongo located in the state of Santa Catarina, Brazil (from Souza and Culpi, 1992).

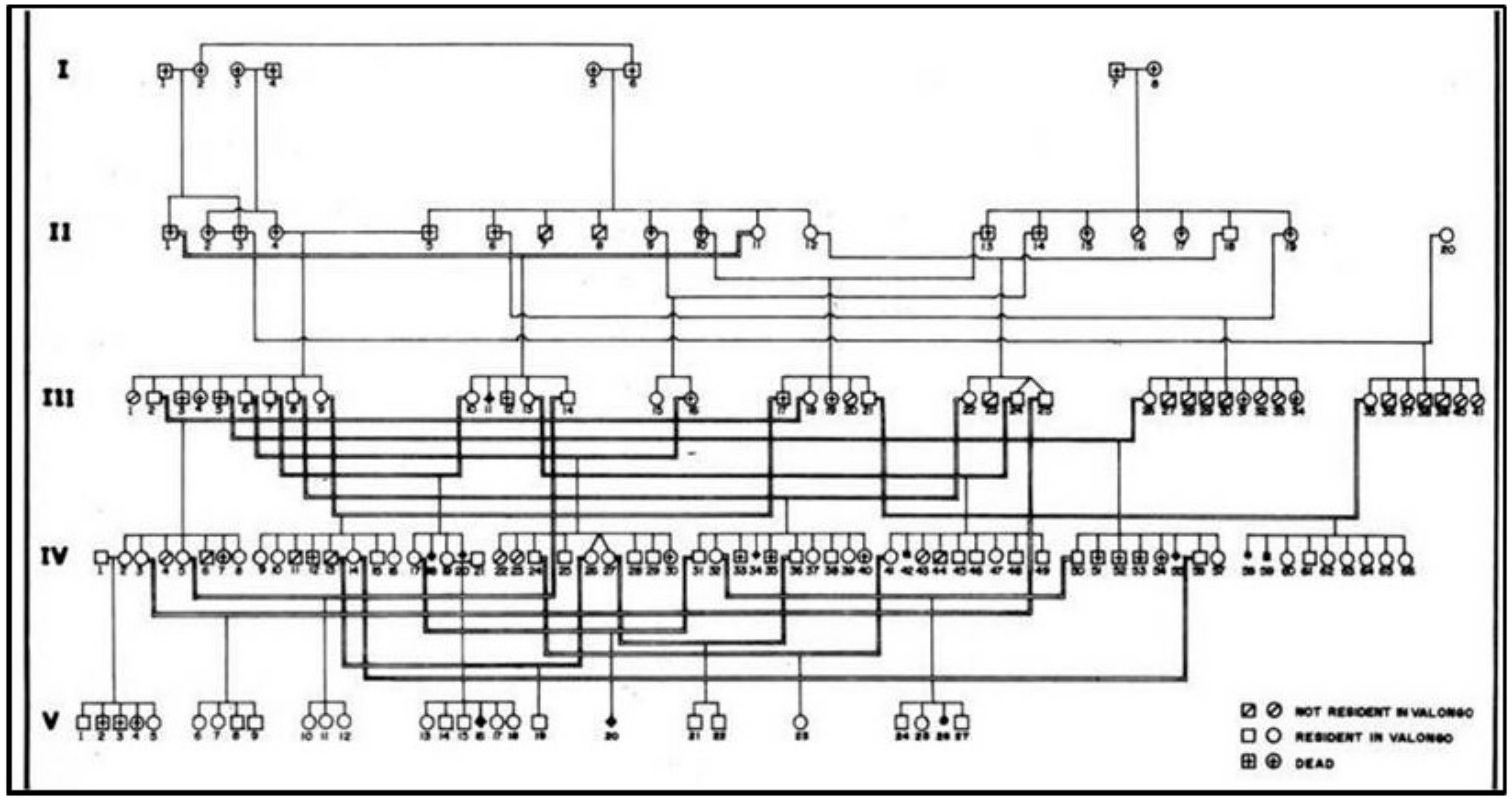


Table S1: Primer sequences and fluorescence types of all microsatellite loci.

Locus	Chromosome	Primer F 5'- 3'	Primer R 5'- 3'	Fluorescence	Multiplex
D1S551	1	TGGTGATCTGCCCCTATTCTA	TGGGAGTGTGCTCATTTTTAAC	FAM	II
D4S3248	4	CACACAGACAGAAAGCGTTACA	AATGCAGTGGGCCTATGTATCTA	FAM	II
D5S816	5	GAGCTATTGCCACTGAAAATCA	CTACTTGGCATCCCTGATGG	FAM	II
D6S1040	6	ATTGGATGAGGCTGGTGAGA	GGAAATGGCCAGAAAATCAG	FAM	IV
D7S821	7	TTTAAGATGGTGTGTGAAGCAGTAG	GGGGCAATAGGTAGGGAACTATAA	HEX	I
D7S3061	7	CCTGGCCTACTATAGGATTTTATCA	GGAAGAGTGGGTGAGGAAAGTA	FAM	II
D8S2324	8	GCAGGTGTTCCTGTCCATAATC	TGACGGAATGAGACTCCATCTAA	FAM	IV
D9S922	9	GAATTCACTCACGGAGCATACA	TCACAGCCACACAAGGACATA	HEX	I
D9S301	9	TTCAAGACAGACAGGCAGACA	GGAAGGTGTGCAAGGATGTT	HEX	III
D10S1426	10	TTTGCTTGGCACCAACTATTC	GTTGAAAACAGGGGCCTACAC	HEX	I
D13S317	13	GAAGTCTGGGATGTGGAGGA	TCCTTCAACTTGGGTTGAGC	FAM	IV
D16S539	16	CAAGCTCTTCCTCTTCCCTAGAT	GTGTGTGCATCTGTAAGCATGTAT	HEX	I
D18S535	18	GACAAAAGCCACACCCATAACT	GCAGTTCCTTTCCTGGGATAAT	HEX	III
D19S559	19	ACCAGCCTGACCAACATAGTG	GGAGGTCGATTTGGGACATA	FAM	IV

D20S482	20	ATCAGAGGACAGCCTCCATATC	CAGAGACACCGAACCAATAAGA	HEX	III
D21S1437	21	GGTTGATTCCATGTCTTTGCT	TGAGGTGCTCCCAAACTCTT	HEX	III

Table S2: Number of genotyped individuals [NG] (in relation to the total number of inhabitants [N] of each community) as to all 30 loci. The last column of the table [RF] lists the corresponding proportions of genotyped individuals per community.

Community	**NG**	**N**	**RF**
AB	17	573	0.0297
AN	8	320	0.0250
GA/SP	16	266	0.0602
IV	9	270	0.0333
MR/PS	8	184	0.0435
NH	7	447	0.0157
PC	16	286	0.0599
TU	6	295	0.0203
Total	**87**	**2641**	**0.0329**

Table S3: Number of genotyped individuals [NG] (in relation to the total number of inhabitants [N] of each community) as to at least 27 out of all 30 loci. The last column of the table [RF] lists the corresponding proportions of genotyped individuals per community.

Community	**NG**	**N**	**RF**
AB	26	573	0.0454
AN	20	320	0.6250
GA/SP	31	266	0.1165
IV	35	270	0.1296
MR/PS	25	184	0.1359
NH	24	447	0.0537
PC	29	286	0.1014
TU	17	295	0.0576
Total	**207**	**2641**	**0.0784**