Renan Barbosa Lemes

INBREEDING STUDIES IN A *QUILOMBO* ISOLATE FROM THE STATE OF SÃO PAULO

ESTUDOS SOBRE ENDOCRUZAMENTO EM UM ISOLADO QUILOMBOLA DO ESTADO DE SÃO PAULO

São Paulo

2017

Renan Barbosa Lemes

INBREEDING STUDIES IN A *QUILOMBO* ISOLATE FROM THE STATE OF SÃO PAULO

ESTUDOS SOBRE ENDOCRUZAMENTO EM UM ISOLADO QUILOMBOLA DO ESTADO DE SÃO PAULO

Dissertation Presented in a Partial Fulfillment of the Requirements for the Degree of Doctor of Science (Biology/Genetics) in the Institute of Biosciences of the University of São Paulo.

Advisor: Prof. Dr. Paulo A. Otto

São Paulo

2017

Catalog Record

Lemes, Renan Barbosa

Inbreeding Studies in a Quilombo Isolate from the State of São Paulo
78 pages

PhD Dissertation - Institute of Biosciences of the University of São Paulo. Department of Genetics and Evolutionary Biology.

Palavras-chave: Endocruzamento; Coeficiente de endocruzamento; Índice de fixação de Wright; *Runs of Homozygosity*; Isolado populacional.

Keywords: Inbreeding; Inbreeding coefficient; Wright's fixation index; Runs of homozygosity; Population isolate.

I. University of São Paulo. Institute of Biosciences. Department of Genetics and Evolutionary Biology.

Examination board:

_____ _____

Prof. Dr. Prof. Dr.

_____ _____

Prof. Dr. Prof. Dr.

_____

Prof. Dr.
Paulo Alberto Otto
Advisor

To Juliana Carnavalli and to my family for their support.

"All we have to decide is what to do with the time that is given to us."

J. R. R. Tolkien

Acknowledgements

To my advisor Prof. Dr. Paulo A. Otto for guiding me safely through the path of science during the last eight years.

To Professors Diogo Meyer and Regina Célia Mingroni Netto for their valuable collaboration.

To Juliana Carnavalli, Kelly Nunes, and Lilian Kimura for the assistance in the sample preparation and for their friendship.

To Professor Fabio M. do Nascimento for his trust and encouragement.

To my laboratory colleagues and friends Adriano, Alex, Allysson, Ana Carol, André, Darine, Dayane, Esther, Gustavo, Leandro, Luis Gustavo, Maria, Rodrigo Barbosa, Rodrigo Salazar, Talita, Thaise, Uirá, and Vinicius; and to Professors Ana Krepischi, Angela Morgante, Carla Rosenberg, Tábita Hünemeier for the many stimulating discussions.

To the head of the Department and direction of the Institute for the infrastructure facilities provided.

To CAPES, FAPESP, and CNPq for financial support.

To friends Fátima, Israel, Mara, Maraisa, Paulo Rogério, and Silvia for technical support.

To City Hall members of the municipalities of Eldorado, Iporanga, and Barra do Turvo for their support, sisters Angela Biagioni and Maria Sueli Berlanga and Antônio Carlos Nicomedes for their assistance.

The inhabitants of quilombo communities.

Finally, to my friends and members of my family.

I.GENERAL INTRODUCTION

This dissertation deals with issues related to the estimation of inbreeding levels and substructure levels, as well as with demographic inferences from a Brazilian population *quilombo* isolate. The document is structured in five sections: (1) this general introduction, where basic concepts related to inbreeding are reviewed; (2) chapter 1, dealing with the estimation of inbreeding and substructure levels in a *quilombo* population; (3) chapter 2, in which a simplified method is presented to estimate the variance of inbreeding coefficient; (4) chapter 3, containing results from inbreeding and demographic analyses performed in the *quilombo* isolate by means of the information of hundreds of thousands of biallelic markers; and (5) a final section with general conclusions. Demographic, historical, and geographical details about the quilombo studied here are exhaustively presented on pages 276-277 of the published article attached to Chapter 1.

I.1. Inbreeding coefficient (Wright's fixation index)

Inbreeding is a non-random mating system in which the choice of mate is influenced or directed by the degree of biological relationship between individuals that mate (Crow and Felsenstein, 1968; Lewontin *et al.*, 1968). Since relatives have one or more ancestors in common, the proportion of alleles identical by descent (IBD) in the genome of their offspring is associated to the amount of ancestry that is shared by their parents.

Endogamy levels are usually estimated by the inbreeding coefficient, which can be defined in terms of correlation as well as of probability (Templeton, 2006; Hartl and Clark, 2007).

Inbreeding coefficient **f** can first be understood as the population correlation coefficient between gametes that come together to generate a zygote (Wright, 1922) and that estimates the deviation λ (covariance among uniting gametes) from genotype frequencies in Hardy-Weinberg (HW) proportions. Considering this parameter, the expected genotype proportions from a biallelic locus (**A**, **a**) can be written down as

$\{d = P(AA) = p^2 + \lambda,\ h = P(Aa) = 2pq - 2\lambda,$ and $r = P(aa) = q^2 + \lambda\}$,

where $p = P(A)$ and $q = 1-p = P(a)$.

The inbreeding coefficient **f** is then defined as the correlation coefficient $f = \rho_{x,y} = \frac{\sigma^2_{x,y}}{\sqrt{\sigma^2_x \sigma^2_y}} = \frac{\lambda}{pq}$, from where we obtain $\lambda = fpq$. In the equation for **f**, **x** and **y** are dummy variables that take the value **1** when the gamete is **A** and **0** otherwise, when the gamete is **a**.

Replacing λ by **fpq** in the genotype proportions **{d, h, r}** above we immediately obtain the usual formulation for genotype frequencies under inbreeding:

$\{d = p^2 + fpq,\ h = 2pq(1 - f),\ r = q^2 + fpq\}$,

from which the value of the inbreeding coefficient can be directly estimated: $f = 1 - \frac{h}{2pq}$.

An alternative approach to estimate the inbreeding coefficient, referred here as **F**, takes into account the probability that two alleles segregating at an autosomal locus are IBD. **F** is usually estimated from genealogies and can be interpreted as the genomic proportion of an individual that is IBD (Haldane and Moshinsky, 1939; Cotterman, 1940; Malécot, 1948). Since the inbreeding coefficient of an individual is the probability that any pair of his homologous genes are identical

by descent, its value coincides with the probability (coefficient of consanguinity) that two homologous genes drawn randomly, one from each individual, are identical. Thus, the inbreeding coefficient $\mathbf{F_k}$ of the individual **k** is also the coefficient of consanguinity $\mathbf{F_{ij}}$ of his/her parents **i** and **j**.

The correct estimation of **F** depends however on the existence of an arbitrary founder population completely unrelated. It is, therefore, very difficult or even impossible to trace back the reliable ancestry information from more ancient generations, which rarely includes relationships more remote than third cousins (Cavalli-Sforza and Bodmer, 1971; Speed and Balding, 2015).

Conceptually, **f** and **F** as defined above are different both biologically as well as mathematically, since **F** is a probability (belonging to the domain $0 \leq \mathbf{F} \leq 1$) that estimates the amount of identity by descent for an individual, while **f** is a coefficient of correlation (belonging to the domain $-1 \leq \mathbf{f} \leq 1$) that measures the population proportions of genotypes above or below the ones randomly expected.

I.2. Hierarchical structure of a population

Natural populations frequently are aggregates formed by partially isolated subpopulations within which mating preferentially occurs. Given the reduced subpopulation sizes, the consequence of substructure is an increase of homozygous levels within the population considered as a whole even if mating within subpopulations takes place randomly, due to changes in allelic frequencies secondarily to genetic drift within subpopulations (Crow and Kimura, 1970).

Hierarchically structured populations were first considered by Wright (1951), who defined three different types of fixation indices: $\mathbf{f_{IS}}$ (fixation index due to inbreeding within each subpopulation), $\mathbf{f_{ST}}$ (fixation index due to genetic drift responsible for differences in allele frequencies among subpopulations), and $\mathbf{f_{IT}}$ (fixation index due to the combined effects of inbreeding and genetic drift), related by the following equations:

$$\mathbf{f_{IT} = f_{ST} + f_{IS} - f_{IS}f_{ST} = 1 - \frac{P(Aa)}{2pq}} \quad ;$$

$$\mathbf{f_{ST} = \frac{f_{IT} - f_{IF}}{1 - f_{IS}} = \frac{var(p)}{pq}} \quad ;$$

$$\mathbf{f_{IS} = \frac{f_{IT} - f_{ST}}{1 - f_{ST}}} \quad ,$$

where **p**, **q**, **P(Aa)**, and **2pq** are respectively the estimated allelic frequencies of alleles **A** and **a**, and the directly observed frequency and the expected panmictic proportion of heterozygous individuals in the whole population. As Chakraborty (2016) noticed, these indices have been conceptually defined in several ways: Wright (1943, 1951) defined them in terms of correlations between uniting gametes, Nei (1973, 1977) and Nei and Chesser (1986) defined them as functions of heterozygotes and differences from their respective expectations under HW equilibrium proportions, while Cockerham (1969, 1973), Weir and Cockerham (1984) and Long (1986) formulated them in terms of functions of parameters of components of nested analysis of variance.

I.3. Consequences of inbreeding

The immediate consequence of inbreeding ($\mathbf{f} > 0$) is the increase in the frequency of homozygotes in the population, which favors the expression of deleterious recessive alleles previously hidden in

heterozygous state. Inbreeding usually leads also to other harmful effects (inbreeding depression), such as the decrease in size, fertility, vigor, yield and fitness, as described for the first time with experimental accuracy by Darwin, who observed its effects in cultivated plants (Fisher, 1949; Crow and Kimura, 1970; Hartl and Clark, 2007).

The effects of endogamy in humans are, in general, more difficult to detect when compared to other species, since the inbreeding levels are usually low and methods that can be developed easily in experimental populations cannot be applied to humans. Empirical studies as well as theoretical risks based on realistic population genetic models show that the chances of affected progeny are largely increased in the offspring of consanguineous marriages (Otto *et al.*, 2007). Strategies based on homozygous mapping (Lander and Botstein, 1987) were developed recently to detect deleterious variants and have been successfully used (1) in identifying new variants related to many disorders of Mendelian recessive inheritance (Lander and Botstein, 1987; Sheffield *et al.*, 1994; Christodoulou *et al.*, 1997; Parvari *et al.*, 1998; Winick *et al.*, 1999; Abou Jamra *et al.*, 2011; Alkuraya, 2013; Ghadami *et al.*, 2015); and (2) in determining susceptibility genes associated with polygenic or complex diseases (Lencz *et al.*, 2007; Nalls *et al.*, 2009; Yang *et al.*, 2012).

I.4. Consanguinity in humans

In humans, consanguineous marriages are still today a relatively common practice, being regarded as customary in many countries throughout the world, because of its traditional status in some cultures. The highest inbreeding levels are found in populations of

the Middle East, Central South Asia and the Americas (Leutenegger *et al.*, 2011).

During the last decades, empirical estimates of consanguinity levels were grossly obtained for many populations over the world, by censoring the frequencies of marriages between second cousins and more closely related pairs of individuals; the information was assembled into a database (consang.net) by Bittles and Black (2015). Despite the very low values (much less than 1% on average) observed for most urbanized populations, the prevalence of consanguineous marriages for the global human population was estimated in about 10%, reaching values above 50% in some extremely inbred populations (Bittles, 2002; Bittles and Black, 2010; Hina and Malik, 2015; Ahmad *et al.*, 2016; Riaz *et al.*, 2016).

I.5. Population isolates

Among humans (and other organisms as well), individuals are, in general, heterogeneously distributed in the population territory, tending to form clusters called population isolates, that can be defined as sets of individuals with imprecise boundaries of different natures: geographical, religious, social, ethnic, political, and so on. (Salzano and Freire-Maia, 1967).

Population isolates offer many advantages to medical and evolutionary studies, mainly when isolates have well documented pedigrees, high prevalence of individuals affected by rare genetic conditions, a high degree of inbreeding due to cultural practices or limited population size, and demographic history of foundation consisting in a bottleneck followed by a founder effect (Arcos-Burgos and Muenke, 2002).

Inbreeding and demographic analyses have been the focus of many studies developed in isolates with different ancestries, with the aim (1) to establish relationships among socio-cultural factors and individual homozygous proportions, (2) to provide demographic information for complementing historical records, and (3) to explain in some extent differences in the prevalence of diseases among different populations (Carothers *et al.*, 2006; McQuillan *et al.*, 2008; Lemes *et al.*, 2014; Abdellaoui *et al.*, 2015; Ben Halim *et al.*, 2015; Jalkh *et al.*, 2015; Karafet *et al.*, 2015).

I.6. Runs of Homozygosity

As known from basic population genetic theory, when two individuals are related in some degree, they share segments that are identical by descent (IBD), that is, autozygous. The offspring of biologically related individuals inherit these segments from both parents, which explains the presence, in them, of long stretches of consecutive homozygosity, called runs of homozygosity (ROH). Broman and Weber (1999) were the first to point out the obvious fact that ROH could be identified by means of the occurrence in homozygous state of a large number of contiguous markers detected by molecular analysis.

Individuals may inherit identical chromosomal segments even when the biological relationship between their parents is very distant. Since elapsed time is positively correlated with the event of recombination occurrence responsible for the breaking up of previously existing segments, ROH from more ancient origin tend to be shorter, while those from recent origin tend to be longer (Kirin *et al.*, 2010).

Recently more precise identification of ROH has been greatly enhanced by the use of genomic data. The inbreeding coefficient,

referred here as $\mathbf{F_{ROH}}$, can be directly estimated from the proportion of the genome composed of these long tracts in homozygous state (McQuillan *et al.*, 2008). $\mathbf{F_{ROH}}$ is very similar to that directly obtained from pedigree analyses, but much more conservative, since it also takes reliable information from ancient and cryptic inbreeding.

Recent studies of ROH data performed in the worldwide human population detected high levels of autozigosity even in cosmopolitan non-inbred populations. It revealed an increment of endogamy levels and a reduction of genetic diversity according to the population distance from African ones, as expected by the out-of-Africa model of modern human migration. The differences have been explained by the occurrence of small and medium ROH resulting from background relatedness, which also enables the use of ROH to obtain reliable information about demographic and evolutionary events (Kirin *et al.*, 2010; Pemberton *et al.*, 2012).

I.7. General Objective

The aim of this work is to obtain reliable estimates of the average inbreeding coefficient using data obtained from a traditional Brazilian tri-hybrid quilombo population. To achieve this, we used different alternative methods, some of them adapted by us for the specific task of dealing with such a genetically complex population aggregate.

We also tried to establish demographic inferences about the foundation of this population isolate.

The specific objectives are presented in the sections labeled as chapters 1 to 3.

GENERAL DISCUSSION AND CONCLUSIONS

This dissertation dealt with issues related to the estimation of average population inbreeding levels and includes two manuscripts already published in specialized international journals and another one yet to be submitted.

Chapter 1 shows how the inbreeding coefficient is estimated by using genealogical and marker (molecular) information. The genealogical (direct) estimation of inbreeding coefficient **F** is complicated due to the usual lack of complete pedigree information and to the arbitrary choice of the number of generations to take into account in its estimation. In spite of these limitations, **F**-values so estimated are used to make valid comparisons of autozygous levels among populations.

Quilombo **F**-values were obtained using all available pedigree information and averaging the individual inbreeding coefficients from all individuals. The values thus obtained were compared with others estimates from the literature (Table 4, Chapter 1). Quilombo **F**-values (and the frequencies of consanguineous marriages) showed to be significantly lower than the values obtained for most isolates from the literature, except in relation to a Brazilian Jewish isolate (Freire-Maia and Krieger, 1963). In any case, the value we estimated is about three times higher than the corresponding one from the Brazilian population (**F** = 0.00088; Freire-Maia, 1990).

As to the quilombo **f** (molecular) estimates of Chapter 1, we used a highly heterogeneous set of 30 molecular markers (14 biallelic SNPs and 16 multiallelic microsatellites). Seven SNPs markers were obtained

from a sample of 700 individuals in an association study of hypertension (Kimura *et al.*, 2012) and another seven SNPs from 400 sampled individuals in an obesity association study (Angeli *et al.*, 2011); the remaining 16 microsatellites were genotyped from a sample of 300 individuals especially selected for the study described in Chapter 1.

We analyzed SNPs and microsatellites data separately and together (Tables 5 and 6, Chapter 1), obtaining average population **f**-values by weighing **f** estimates from each community by the reciprocal of the corresponding variances. Given that the sample sizes required to obtain **f**-values significantly different from zero are extremely high (Figure 2, Chapter 1) in tests that verify departures from HW proportions, no **f** estimate obtained from SNP markers was significantly different from zero. In two instances of microsatellite markers we found **f**-values significantly lower than zero, a result that might result from the combination of small sample sizes and multiallelic nature of these markers.

Historical records collected by members of Dr. Regina Mingroni's laboratory account for the presence of intense migration among all subpopulations analyzed, indicating an absence of genetic isolation. Using our molecular markers data, we estimated Wright's fixation indexes. The estimates of $\mathbf{f_{ST}}$ values obtained were in general lower than 5%, which is according to results previously obtained from the analysis of INDEL markers data for the same subpopulations (Kimura *et al.*, 2013). These results indicate, as expected from the historical records mentioned, the absence of significant population substructure levels in the whole quilombo aggregate.

The second article presented in Chapter 2 dealt with the estimation of **var(f)**. The very simple approximation we provided could be applied to a locus with any number of alleles, producing estimates very similar to those obtained using simulations or approximations already known in the literature for two allele case (Fyfe and Bailey, 1951; Curie-Cohen, 1982). Given that the formal estimation of **var(f)** is (mathematically) a very complicated issue, our work resulted in a very simple and efficient method to obtain reliable f-variance estimates.

The third chapter is represented by an umpublished manuscript dealing with the estimation of the coefficient **f** (in the same quilombo population) using high density SNP array data and presenting a new manner to estimate the index, by using the joint information from two sets of markers (complete and no-LD datasets).

It is known from population genetics theory that the unbiased estimation of the average inbreeding coefficient $\bar{\mathbf{f}}$ should consider only completely independent loci, that is, loci with no linkage disequilibrium. The main problem in excluding linked data is the drastic reduction of dataset information.

With the aim of seeking for markers with more reliable information, we considered in our analysis both datasets (complete and no-LD), observing that: (1) markers with MAF < 0.3 introduced a bias underestimating the average **f**-values, since they might include data with errors in genotype determination that resisted to the filtering process; (2) no statistically significant difference between the **f** average estimates from both datasets was found, since their 95% confidence intervals overlapped.

We made also some inferences from the quilombo demographic history, as we were dealing with a highly admixed tri-hybrid population with a complex foundation history. Both the total ROHs lengths and the $\mathbf{F_{ROH}}$ values were lower in the quilombo than in the European and Asian population datasets and a bit higher than in the African one selected for comparison. The results we obtained suggest that the patterns of ROH and $\mathbf{F_{ROH}}$ of an admixed population such as the quilombo reported here should be somehow proportional to the contribution of the parental (stock) populations, but lower, given that the admixture process inserted some degree of variability in the gene pool of the hybrid population.

ABSTRACT

Endogamy levels are usually estimated using genealogical or molecular markers data. By means of both type of data from a traditional Brazilian tri-hybrid quilombo population aggregate (located at the Ribeira River Valley in the State of São Paulo), the aim of this work, using different methods, was to obtain reliable estimates of its average inbreeding coefficient, as well as to establish pertinent demographic inferences.

The results we obtained are presented in three chapters.

The first one, represented by the offprint of a published paper, deals with the estimation of the inbreeding coefficient using both a complete genealogical and comprehensive molecular information. **F**-values were estimated for each community using all available *pedigree* information and averaging the inbreeding coefficients from all individuals represented in the genealogies. Molecular **f**-values were estimated from the analysis of 30 highly heterogenous sets of molecular markers (14 biallelic SNPs and 16 multiallelic microsatellites), genotyped in different groups of individuals from the population.

The second chapter (a research paper already published), presents a simplified method to estimate the variance of the inbreeding coefficient. The simple approximations we provided can be applied to a locus with any number of alleles, producing estimates fully validated by computer simulations.

The last chapter is a manuscript yet to be published that deals with inbreeding and demographic inferences, obtained from the information of hundreds of thousands of biallelic SNP markers. A new

manner to obtain estimates of Wright's fixation index **f** is presented, consisting in the use of the joint information of two sets of markers (one complete and another excluding markers in patent linkage disequilibrium). Quilombo demographic inferences were obtained by means of ROHs analyses, which were adapted to cope with a highly admixed population with a complex foundation history.

RESUMO

Os níveis de endogamia de uma população são comumente estimados por meio do coeficiente de endocruzamento, que pode ser obtido de dados genealógicos (**F**) ou dados provenientes da análise de marcadores moleculares (**f**).

O objetivo do trabalho foi obter estimativas confiáveis do coeficiente de endocruzamento populacional, bem como realizar inferências demográficas, usando dados de um agregado populacional quilombola miscigenado com ancestralidade complexa tri-híbrida, localizado no Vale do Rio Ribeira, na região sul do estado de São Paulo.

No trabalho é apresentado em três capítulos. No primeiro (um trabalho já publicado), estimamos o coeficiente de endocruzamento usando dados genealógicos e moleculares. As estimativas genealógicas de **F** foram obtidas para cada comunidade por meio da média dos coeficientes individuais de todos os indivíduos representados nas genealogias da população. Os valores de **f** foram estimados por meio dos dados de 30 marcadores moleculares altamente heterogêneos (14 SNPs e 16 microssatélites), genotipados em diferentes grupos de indivíduos com diferentes finalidades.

O segundo capítulo, representado por um trabalho também já publicado, apresenta um método simples para estimar a variância do coeficiente de endocruzamento **f**. As aproximações obtidas, validadas devidamente por simulações em computador, podem ser aplicadas a lóci multialélicos, produzindo estimativas que não diferem significativamente de outras aproximações complicadas descritas na literatura.

O último capítulo (um manuscrito a ser submetido para publicação) apresenta inferências a respeito dos processos de endogamia e demografia no isolado quilombola, utilizando a informação de centenas de milhares de marcadores moleculares bialélicos. É apresentada uma nova maneira de se estimar o índice de fixação **f** de Wright, usando a informação combinada de dois conjuntos de marcadores (o conjunto completo de marcadores e um outro contendo apenas marcadores não ligados significativamente entre si). Também foram feitas inferências sobre a história demográfica do isolado por meio do estudo das regiões genômicas em homozigose (ROHs), uma contribuição inédita e importante do trabalho, adaptada à análise de um isolado populacional altamente miscigenado com contribuição tri-híbrida e uma história de fundação complexa.

REFERENCES

Abdellaoui A; Hottenga JJ; Willemsen G; Bartels M; van Beijsterveldt T; Ehli EA; Davies GE; Brooks A; Sullivan PF; Penninx BWJH; Geus EJ; Boomsma DI. Educational attainment influences levels of homozygosity through migration and assortative mating. **PLoS One. 10: e0118935, 2015.**

Abney M; McPeek MS; Ober C. Estimation of variance components of quantitative traits in inbred populations. **American Journal of Human Genetics. 66:629–650, 2000.**

Abou Jamra R; Wohlfart S; Zweier M; Uebe S; Priebe L; Ekici A; Giesebrecht S; Abboud A; Al Khateeb MA; Fakher M; Hamdan S; Ismael A; Muhammad S; Nöthen MM; Schumacher J; Reis A. Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. **European Journal Human Genetics. 19: 1161-1166, 2011.**

Ahmad B; Rehman AU; Malik S. Consanguinity and Inbreeding Coefficient in Tribal Pashtuns Inhabiting the Turbulent and War-Affected Territory of Bajaur Agency, North-West Pakistan. **Journal of Biosocial Science. 48: 113-128, 2016.**

Alkuraya FS. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. **Human Genetics. 132: 1197-1211, 2013.**

Angeli CB; Capelli LP; Auricchio MTBM; Vianna-Morgante AM; Mingroni-Netto RC; Leal-Mesquita ER; Ribeiro-dos-Santos AKC; Ferrari I; Oliveira SF; Klatau-Guimarães MN. AGG interspersion patterns in the CGG repeat of the FMR1 gene and linked DXS548/FRAXAC1 haplotypes in Brazilian populations. **American Journal of Medical Genetics. 132A: 210–214, 2005.**

Angeli CB; Kimura L; Auricchio MT; Vicente JP; Mattevi VS; Zembrzuski VM; Hutz MH; Pereira AC; Pereira TV; Mingroni-Netto RC. Multilocus analyses of seven candidate genes suggest interacting pathways for obesity-related traits in Brazilian populations. **Obesity. 19: 1244-1251, 2011.**

Arcos-Burgos M; Muenke M. Genetics of population isolates. **Clinical Genetics. 61: 233-247, 2002.**

Auricchio MTBM; Vicente JP; Meyer D; Mingroni-Netto RC. Frequency and origins of hemoglobin S mutation in African-derived Brazilian populations. **Human Biology. 79: 667-677, 2007.**

Bhatia G; Patterson N; Sankararaman S; Price AL. Estimating and interpreting FST: the impact of rare variants. **Genome Research. 23: 1514-1521, 2013.**

Ben Halim N; Nagara M; Regnault B; Hsouna S; Lasram K; Kefi R; Azaiez H; Khemira L; Saidane R; Ammar SB; Besbes G; Weil D; Petit C; Abdelhak S; Romdhane L. Estimation of Recent and Ancient Inbreeding in a Small Endogamous Tunisian Community Through Genomic Runs of Homozygosity. **Annals of Human Genetics. 79: 402-417, 2015.**

Bittles AH. Endogamy, consanguinity and community genetics. **Journal Genetics. 81: 91-98, 2002.**

Bittles AH; Black ML. Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex

diseases. **Proceedings of the National Academy of Sciences USA. 107: 1779-1786, 2010.**

Bittles AH; Black ML. Global patterns & tables of consanguinity. **URL http://consang.net, 2015.**

Broman KW; Weber JL. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. **American Journal of Human Genetics. 65: 1493-1500, 1999.**

Cannings C; Edwards AW. 1969. Expected genotypic frequencies in a small sample: Deviation from Hardy-Weinberg equilibrium. **American Journal of Human. Genetics. 21: 245–247, 1969.**

Carothers AD; Rudan I; Kolcic I; Polasek O; Hayward C; Wright AF; Campbell H; Teague P; Hastie ND; Weber JL. Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. **Annals of Human Genetics. 70: 666-676, 2006.**

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

Chakraborty, R. Comments on 'A note on the variance of the estimate of the fixation index F'. **Journal of genetics, 95: 229-230, 2016.**

Christodoulou K; Tsingis M; Deymeer F; Serdaroglu P; Ozdemir C; Al-Shehab A; Bairactaris C; Mavromatis I; Mylonas I; Evoli A; Kyriallis K; Middleton LT. Mapping of the familial infantile myasthenia (congenital myasthenic syndrome type Ia) gene to chromosome 17p with evidence of genetic homogeneity. **Human Molecular Genetics. 6: 635-640, 1997.**

Cockerham CC. Variance of gene frequencies. **Evolution. 23: 72–84, 1969.**

Cockerham CC. Analyses of gene frequencies. **Genetics. 74: 679–700, 1973.**

Cotrim NH; Auricchio MT; Vicente JP; Otto PA; Mingroni-Netto RC. Polymorphic Alu insertion in six brazilian african-derived populations. **American Journal of Human Biology. 16: 264-277, 2004.**

Cotterman CW. **A Calculus for Statistico-genetics.** Unpublished thesis, Ohio State University, Columbus, Ohio. 190

Crow JF; Felsenstein J. The effect of assortative mating on the genetic composition of a population. **Eugenics Quarterly. 15: 85-97, 1968.**

Crow JF, Kimura M. **An introduction population genetics theory.** Alpha Editions, Madison, 1970.

Curie-Cohen M. Estimates of inbreeding in a natural population: a comparison of sampling properties. **Genetics. 100: 339-358, 1982.**

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

Ellis WS; Starmer WT. Inbreeding as measured by isonymy, pedigrees, and population size in Törbel, Switzerland. **American Journal of Human Genetics. 30: 366-376, 1978.**

Fraley C; Raftery AE; Scrucca L. Normal mixture modeling for model-based clustering, classification, and density estimation. **Department of Statistics, University of Washington, 23, 2012.**

Fisher RA. **The theory of inbreeding.** Oliver and Boya; London, 1949.

Freire-Maia N. Inbreeding in Brazil. **American Journal of Human Genetics. 9: 284–298, 1957.**

Freire-Maia N. Genetic efffects in Brazilian populations due to consanguineous marriages. **American Journal of Medical Genetics, 35:115–117, 1990.**

Freire-Maia N; Krieger H. A Jewish isolate in southern Brazil. **Annals of Human Genetics. 27:31–39, 1963.**

Fyfe JL; Bailey NTJ. Plant breeding studies in leguminous forage crops I. Natural cross-breeding in winter beans. **The Journal of Agricultural Science. 41: 371, 1951.**

Ghadami S; Mohammadi HM; Malbin J; Masoodifard M; Sarhaddi AB; Tavakkoly-Bazzaz J; Zeinali S. Frequencies of Six (Five Novel) STR Markers Linked to TUSC3 (MRT7) or NSUN2 (MRT5) Genes Used for Homozygosity Mapping of Recessive Intellectual Disability. **Clinical Laboratory. 61: 925-932, 2015.**

Gogarten SM; Bhangale T; Conomos MP; Laurie CA; McHugh CP; Painter I; Zheng X; Crosslin DR; Levine D; Lumley T; Nelson SC; Rice K; Shen J; Swarnkar R; Weir BS; Laurie CC. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. **Bioinformatics, 28: 3329-3331, 2012.**

Hamamy H; Jamhawi L; Al-Darawsheh J; Ajlouni K. Consanguineous marriages in Jordan: why is the rate changing with time? **Clinical Genetics. 67: 511–516, 2005.**

Haldane JBS; Moshinsky P. Inbreeding in mendelian populations with special reference to human cousin marriage. **Annals of Eugenics. 9: 321-340, 1939.**

Hartl DL; Clark AG. **Principles of Population Genetics.** Sinauer Associates, Inc, Sunderland, MA, 2007.

Hina S; Malik S. Pattern of Consanguinity and Inbreeding Coefficient in Sargodha District, Punjab, Pakistan. **Journal of Biosocial Science. 47: 803-811, 2015.**

Jackson CE, Symon WE, Pruden EL, Kaehr IM, Mann JD. Consanguinity and Blood Group Distribution in an Amish Isolate. **American Journal of Human Genetics. 20: 522-527, 1968.**

Jalkh N; Sahbatou M; Chouery E; Megarbane A; Leutenegger AL; Serre JL. Genome-wide inbreeding estimation within Lebanese communities using SNP arrays. **European Journal of Human Genetics. 23: 1364-1369, 2015.**

Karafet TM; Bulayeva KB; Bulayev OA; Gurgenova F; Omarova J; Yepiskoposyan L; Savina OV; Veeramah KR; Hammer MF. Extensive genome-wide autozygosity in the population isolates of Daghestan. **European Journal of Human Genetics. 23: 1405-1412, 2015.**

Kimura L; Angeli CB; Auricchio MT; Fernandes GR; Pereira AC; Vicente JP; Pereira TV; Mingroni-Netto RC. Multilocus family-based association analysis of seven candidate polymorphisms with essential hypertension in an african-derived semi-isolated brazilian population. **International Journal of Hypertension. 2012: 859219, 2012.**

Kimura L; Ribeiro-Rodrigues EM; De Mello Auricchio MT; Vicente JP; Batista Santos SE; Mingroni-Netto RC. Genomic ancestry of rural African-derived populations from Southeastern Brazil. **American Journal of Human Biology. 25: 35-41, 2013.**

Kirin M; McQuillan R; Franklin CS; Campbell H; McKeigue PM; Wilson JF. Genomic runs of homozygosity record population history and consanguinity. **PLoS One. 5: e13996, 2010.**

Lander E; Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. **Science. 236: 1567-1570, 1987.**

Lemes RB; Nunes K; Meyer D; Mingroni-Netto RC; Otto PA. Estimation of inbreeding and substructure levels in african-derived brazilian quilombo populations. **Human Biology. 86: 276-288, 2014.**

Lencz T; Lambert C; DeRosse P; Burdick KE; Morgan TV; Kane JM; Kucherlapati R; Malhotra AK. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. **Proceedings of the National Academy of Sciences USA. 104: 19942-19947, 2007.**

Leutenegger AL; Sahbatou M; Gazal S; Cann H; Genin E. Consanguinity around the world: what do the genomic data of the HGDP-CEPH diversity panel tell us? **European Journal of Human Genetics. 19: 583-587, 2011.**

Lewontin R; Kirk D; Crow J. Selective mating, assortative mating, and inbreeding: definitions and implications. **Eugenenics Quarterly. 15: 141-143, 1968.**

Long JC. The allelic correlation structure of Gainj- and Kaam-speaking people. I. The estimation and interpretation of Wright's F-statistics. **Genetics. 112: 629–647, 1986.**

Malécot G. **Les mathématiques de l'hérédité.** Masson et Cie, Paris, 1948.

McQuillan R; Leutenegger AL; Abdel-Rahman R; Franklin CS; Pericic M; Barac-Lauc L; Smolej-Narancic N; Janicijevic B; Polasek O; Tenesa A; MacLeod AK; Farrington SM; Rudan P; Hayward C; Vitart V; Rudan I; Wild SH; Dunlop MG; Wright AF; Campbell H; Wilson JF. Runs of homozygosity in European populations. **American Journal of Human Genetics. 83: 359-372, 2008.**

Mingroni-Netto RC; Angeli CB; Kimura L; Auricchio MTBM; Vicente JP. Doenças modernas nos antigos quilombos: a obesidade e a hipertensão no Vale do Ribeira. In: Volochko A; Batista LE. **Saúde nos quilombos.** Instituto da Saúde, São Paulo, 2009a, pp. 179-191.

Mingroni-Netto RC; Auricchio MTBM; Vicente JP. Importância da pesquisa do traço e da anemia falciforme nos remanescentes de quilombos do Vale do Ribeira-SP. In: Volochko A; Batista LE. **Saúde nos quilombos.** Instituto da Saúde, São Paulo, 2009b, pp. 169-177.

Nalls MA; Guerreiro RJ; Simon-Sanchez J; Bras JT; Traynor BJ; Gibbs JR; Launer L; Hardy J; Singleton AB. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. **Neurogenetics. 10: 183-190, 2009.**

Nei M. Analysis of gene diversity in subdivided populations. **Proceedings of the National Academy of Sciences USA. 70: 3321–3323, 1973.**

Nei M. F-statistics and analysis of gene diversity in subdivided populations. **Annals of Human Genetics. 41: 225–233, 1977.**

Nei M; Chesser R. K. Estimation of fixation indices and gene diversities. **Annals of Human Genetics. 47: 253–259, 1983.**

Nunes K; Zheng X; Torres M; Moraes ME; Piovezan BZ; Pontes GN; Kimura L; Carnavalli JEP; Mingroni-Netto RC; Meyer D. HLA imputation in an admixed population: An assessment of the 1000 Genomes data as a training set. **Human immunology. 77: 307-312, 2016.**

Otto PA; Frota-Pessoa O. Genetic risks of consanguineous marriages. In: Mayo O; Leach C (Orgs.). **Fifty years of human genetics: A Festschrift and liber amicorum to celebrate the life and work of**

George Robert Fraser. Adelaide: Wakefield Press, Australia, 2007, pp. 436-442.

Otto PA; Lemes RB. A note on the variance of the estimate of the fixation index F. **Journal of Genetics. 94: 759-763, 2015.**

Pasinato R; Rettl KI. Desenvolvimento local sustentável: a contribuição das comunidades quilombolas do Vale do Ribeira. In: Volochko A; Batista LE. (Orgs.). **Saúde nos quilombos.** Instituto da Saúde, São Paulo, 2009, pp. 43-56.

Parvari R; Hershkovitz E; Kanis A; Gorodischer R; Shalitin S; Sheffield VC; Carmi R. Homozygosity and linkage-disequilibrium mapping of the syndrome of congenital hypoparathyroidism, growth and mental retardation, and dysmorphism to a 1-cM interval on chromosome 1q42-43. **American Journal of Human Genetics. 63: 163-169, 1998.**

Pemberton TJ; Absher D; Feldman MW; Myers RM; Rosenberg NA; Li JZ. Genomic patterns of homozygosity in worldwide human populations. **American Journal of Human Genetics. 91: 275-292, 2012.**

Pemberton TJ; Rosenberg NA. Population-genetic influences on genomic estimates of the inbreeding coefficient: a global perspective. **Human Heredity. 77: 37-48, 2014.**

Purcell S; Neale B; Todd-Brown K; Thomas L; Ferreira MA; Bender D; Maller J; Sklar P; de Bakker PI; Daly MJ; Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics. 81: 559-575, 2007.**

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. **URL https://www.R-project.org/. 2016.**

Riaz HF; Mannan S; Malik S. Consanguinity and its socio-biological parameters in Rahim Yar Khan District, Southern Punjab, Pakistan. **Journal of Health, Population and Nutrition. 35: 14, 2016.**

Robertson A; Hill W. Deviations from Hardy-Weinberg proportions: sampling variances and use in the estimation of inbreeding coefficients. **Genetics. 107, 703–718, 1984.**

Rosenberg NA; Pemberton TJ; Li JZ; Belmont JW. Runs of homozygosity and parental relatedness. **Genetics in Medicine. 15: 753-754, 2013.**

Rozen S; Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S; Misener S (Eds.). **Bioinformatics Methods and Protocols: Methods in Molecular Biology.** Humana Press, Totowa, NJ, 2000, pp. 365-386.

Salzano FM, Freire-Maia N. **Populações brasileiras: Aspectos demográficos, genéticos e antropológicos.** Editora da Universidade de São Paulo, São Paulo, 1967.

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

Sheffield VC; Carml R; Kwltek-Black A; Rokhlina T; Nishlmura D; Duyk GM; Elbedour K; Sunden SL; Stone EM. Identification of a Bardet-Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. **Human Molecular Genetics. 3: 1331-1335, 1994.**

Souza IR; Culpi L. Valongo, an isolated Brazilian Black community. I. Structure of the population. **Brazilian Journal of Genetics. 15: 439-447, 1992.**

Speed D; Balding DJ. Relatedness in the post-genomic era: is it still useful? **Nature Reviews Genetics. 16: 33-44, 2015.**

Templeton AR. **Population genetics and microevolutionary theory**. John Wiley & Sons, Inc, Hoboken, New Jersey, 2006.

Teo SM; Ku CS; Salim A; Naidoo N; Chia KS; Pawitan Y. Regions of homozygosity in three Southeast Asian populations. **Journal of Human Genetics. 57: 101-108, 2012.**

Wang S; Haynes C; Barany F; Ott J. Genome-wide autozygosity mapping in human populations. **Genetic Epidemiology. 33: 172-180, 2009.**

Weir BS. Genetic data analysis II. **Sinauer Associates Inc, Sunderland, MA, 1996.**

Weir BS. Interpreting Whole-Genome Marker Data. **Statistics in Biosciences. 5: 2013.**

Weir BS; Cockerham CC. Estimating F-statistics for the analysis of population structure. **Evolution. 38: 1358–1370, 1984.**

Wigginton JE; Cutler DJ; Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. **American Journal of Human Genetics. 76: 887-893, 2005.**

Winick JD; Blundell ML; Galke BL; Salam AA; Leal SM; Karayiorgou M. Homozygosity mapping of the Achromatopsia locus in the Pingelapese. **American Journal of Human Genetics. 64: 1679-1685, 1999.**

Wright S. Coefficients of inbreeding and relationship. **American Naturalist. 56: 330-338, 1922.**

Wright S. Isolation by distance. **Genetics. 28: 114–138, 1943.**

Wright S. The genetical structure of populations. **Annals of Eugenics. 15:323-354, 1951.**

Yang HC; Chang LC; Liang YJ; Lin CH; Wang PL. A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. **PLoS One. 7: e34840, 2012.**

Yeh E; Kimura L; Errera FI; Angeli CB; Mingroni-Netto RC; Silva ME; Canani LH; Passos-Bueno MR. 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. **Brazilian Journal of Medical and Biological Research. 41: 468-472, 2008.**