Genetic ancestry effects on the distribution of toll-like receptors (TLRs) gene polymorphisms in a population of the Atlantic Forest, São Paulo, Brazil

Lilian O. Guimarães, Miklos Maximiliano Bajay, Eliana F. Monteiro, Gerhard Wunderlich, Sidney E. Santos, Karin Kirchgatter

ABSTRACT

The innate immune system governed by toll-like receptors (TLRs) provides the first line of defense against pathogens. Surface-localized TLR1 and TLR6 are known to detect parasite components. TLR encoding genes were shown to display signatures of recent positive selection in Europeans and might be involved in local adaptation at immune-related genes. To verify the influence of Brazilian population admixture on the distribution of polymorphisms in TLRs, we analyzed the genotype frequencies of 24 polymorphisms distributed across five TLR genes in a Southeastern Brazilian population where autochthonous cases of malaria occur in small foci of transmission. The estimation of ancestry showed mainly European ancestry (63%) followed by African ancestry (22%). Mean proportions of European ancestry differed significantly between the genotypes of the TLR1 (I602S) gene and in the TLR6 (P249S) gene. The chance of having the G allele in TLR1 gene increases as European ancestry increases as well as the chance of having the T allele in the TLR6 gene. The 602S allele is related to a ‘hypo-responsiveness’ possibly explaining the high prevalence of asymptomatic malaria cases in areas of Southeastern Brazil. Our results underline the necessity to include informative ancestry markers in genetic association studies in order to avoid biased results.

1. Introduction

The first line of defense against pathogens is provided by the innate immune system including toll-like receptors (TLRs) [1]. Ten TLR encoding genes have been described in humans dispersed all over the genome [1]. TLR1, TLR6, and TLR10 occur on the cell surface and are known to be activated by bacterial, fungal, and parasite components including flagellin and glycolipids [1]. TLRs are highly conserved, although cell-surface TLRs are less strictly constrained than intracellular TLRs [2]. In fact, the region including the TLR1/TLR6/TLR10 genes is under natural selection [3–5] and has been shown to have signatures of recent positive selection in Europeans [2,6] suggesting that they might be involved in local adaptation.

Single nucleotide polymorphisms (SNPs) are present in variable frequencies in different populations. They play an important role in different diseases to which populations are differently exposed, therefore, their analysis is important for genetic studies in humans. SNPs in TLR genes have been associated with susceptibility to various infectious diseases [1], including malaria [7–11]. Analysis of the genetic structure of human populations has become an important practical issue for the study of complex genetic disorders. Recent studies have shown that genetic ancestry and natural selection drive population differences in immune responses to pathogens [12]. Differences in frequency of alleles between ethnic groups and subgroups and admixture between different ethnic groups may result in frequent false-positive results or reduced power in genetic studies [13]. Thus, with the prospect of future genetic Association studies using SNPs in TLRs and malaria susceptibility, it is first necessary to characterize the genetic structure of the study population in order to avoid biases and falsely interpreted results.

Currently, one of the most diverse populations in the world is the Brazilian population, which resulted from five hundred years of interethnic admixture among European, African and Amerindian ancestors. Estimates of ancestry results produced by genetic studies have shown that the Brazilian population has a major contribution of European ancestry (around 75–77%) followed by African and Amerindian contributions [14,15]. The population from the southern region has...
displayed the highest proportion of European ancestry (87.7%) and a lower degree of variance of the estimates of individual admixture causing a significant genetic differentiation from the southeast, northeast, center-west and northern regions [15].

Considering the relevance of SNPs in important TLRs and their relationship with the European population coupled with the fact that the Southeastern Brazilian population has the second highest proportion of European ancestry (79.9%), we set out to analyze the connection between genomic ancestry in a population from Juquitiba region, a typical town in Southeastern Brazil in an area of Atlantic Forest with the second highest number of autochthonous malaria cases in São Paulo State [16]. For this, we used ancestry informative markers (AIMs) which are better indicators of genetic ancestry than physical appearances, such as skin color, used in self-declaration of ethnicity [17–20]. Our results analyze (i) the genotype frequencies of 24 polymorphisms distributed across five TLR genes (TLR1, TLR2, TLR4, TLR6 and TLR9) in a Brazilian population of the Atlantic Forest; (ii) the influence of Brazilian population admixture on the distribution of these polymorphisms; and (iii) the comparison of the allele and genotypic frequencies found in this population with those observed in African, European, and American populations from the 1000 Genomes Project obtained by International Genome Sample Resource (IGSR) by dbSNP database [21].

2. Materials and methods

2.1. Samples

The study was conducted in a population living along the Palestina road in Juquitiba (São Paulo State) (23°55′55" S; 47°04′04" W), situated in the Brazilian Atlantic Forest region (Fig. 1). Juquitiba County is located in the metropolitan area of São Paulo and covers an area of 521.6 km² with a population of 28 thousand inhabitants of whom 22.6% live in the rural area [22]. It has a significant area of preserved Atlantic Forest in the Metropolitan Region of São Paulo.

The sample of this study was composed of 195 individuals (96 men and 99 women) aged from fourteen to eighty-one years. All the participants signed informed consent forms. This project was approved by the Ethics Committee of Institute of Biomedical Sciences, University of São Paulo.

DNA was extracted from peripheral blood samples using the Illustra™ Blood Genomic Prep Mini Spin Kit (GE Healthcare, UK), according to the manufacturer’s instructions.

2.2. SNP genotyping

We analyzed the presence of 24 polymorphisms: (i) 20 non-synonymous SNPs: TLR1 I602S, TLR2 (P631H, R650Q, L658, P669L, R677W, F679I, I693T, E694D, Y715N, F722L, R723C, N729S, E738Q, P746S, R748H, R753Q), TLR4 (D299G, T399I) and TLR6 P249S; (ii) the two promoter SNPs at TLR9 in positions –1486 T.C and –1237 T.C as well as one SNP in the intron region at position 1174 A.G; and (iii) a GT dinucleotide repeat that varies by approximately 12–30 repeats (GTn) that is present within the second intron, approximately 100 bp upstream of the translational start site of TLR2. PCR-restriction fragment length polymorphism approaches were used to determine genotypes of five SNPs: TLR1 I602S, TLR4 D299G, TLR6 P249S, TLR9 –1486 T.C and TLR9 –1237 T.C. The amplified PCR fragments were digested with appropriate restriction endonucleases and then resolved on 3% agarose gels. To ensure the validity of our genotyping methods, a known genotype for each SNP was used as a control. PCR primers, amplification conditions and the restriction enzymes used for each polymorphism are described in Supplementary Table 1. Eighteen SNPs were analyzed by PCR and sequencing. PCR primers and amplification conditions used for this analysis are also listed in Supplementary Table 1. TLR2 GTn genotypes were based on size discrimination in electropherograms were determined using the open-access STRand software [23]. The number of GTn repeats was categorized to short (S) (≤16 repeats or 96 bp), medium (M) (17–22 repeats or 98–108 bp) and large (L) (≥23 repeats or ≥110 bp) as described [24].

2.3. Ancestry informative marker analysis

The evaluation of genomic ancestry was conducted as described [25], using forty-eight biallelic ancestry informative markers (AIMs) and insertions-deletions (INDELS), from autosomal chromosomes. Estimation of the parental ancestry of the Brazilian samples was performed considering three parental populations: Africans (from Angola, Mozambique, Zaire, Cameroon, and the Ivory Coast), Europeans (mainly Portuguese), and Native Americans (individuals from indigenous tribes of the Brazilian Amazon region) [25]. The genotypes of the sample population from Juquitiba, São Paulo, Brazil, and parental populations (Europeans, Africans, and Amerindians) were analyzed together, assuming three main clusters (K = 3).

2.4. Statistical analysis

We used the Structure v.2.3.4 program for Bayesian model-based estimates of the proportion of ancestry for each subject [26–29]. The major populations of origin in Brazil are European, African and Amerindian, and hence three subpopulations (k = 3) were modeled. Quantitative genetic ancestry was estimated for individuals by including contemporary descendants of approximate populations of origin (pseudo-ancestors) [28]. The admixture model was employed assuming correlated population allele frequencies with a burn-in period of 100,000 and 1,000,000 iterations. All other analyses were conducted using packages contained in the R environment, version 3.3.2 [30].

The Genetics package was used to obtain: (i) the allele and genotype frequencies for each variant; (ii) deviations from Hardy-Weinberg equilibrium by the Chi-square test; and (iii) the linkage disequilibrium (LD) between pairs of loci using parameter D’ [31]. Linkage disequilibrium was evaluated between the pairs of SNPs in the TLR1 (I602S) and TLR6 (P249S) genes located in the chromosome 4p14 region, TLR2 (P631H and R753Q) gene located in the chromosome 4q31.3 region, TLR4 (D299G and T399I) gene located in the chromosome 9q32–33 region and TLR9 (–1237, –1486 and 1174) gene.