Human Immunology 71 (2010) 922–928

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Balanced polymorphism in bottlenecked populations: The case of the *CCR5* 5′ cis-regulatory region in Amazonian Amerindians

Rodrigo F. Ramalho [a,*], Eduardo J.M. Santos [b], João F. Guerreiro [b], Diogo Meyer [a]

[a] *Universidade de São Paulo, Instituto de Biociências, Departamento de Genética e Biologia Evolutiva, São Paulo, SP, Brasil*
[b] *Universidade Federal do Pará, Centro de Ciências Biológicas, Laboratório de Genética Humana e Médica, Belém, PA, Brasil*

ARTICLE INFO

Article history:
Received 5 January 2010
Accepted 26 May 2010
Available online 9 June 2010

Keywords:
5′ cis-regulatory region of *CCR5* gene
Amerindians
Balancing selection
Demography
Genetic drift

ABSTRACT

The 5′ cis-regulatory region of the *CCR5* gene exhibits a strong signature of balancing selection in several human populations. Here we analyze the polymorphism of this region in Amerindians from Amazonia, who have a complex demographic history, including recent bottlenecks that are known to reduce genetic variability. Amerindians show high nucleotide diversity (π = 0.27%) and significantly positive Tajima's *D*, and carry haplotypes associated with weak and strong gene expression. To evaluate whether these signatures of balancing selection could be explained by demography, we perform neutrality tests based on empiric and simulated data. The observed Tajima's *D* was higher than that of other world populations; higher than that found for 18 noncoding regions of South Amerindians, and higher than 99.6% of simulated genealogies, which assume nonequilibrium conditions. Moreover, comparing Amerindians and Asians, the Fst for *CCR5* cis-regulatory region was unusually low, in relation to neutral markers. These findings indicate that, despite their complex demographic history, South Amerindians carry a detectable signature of selection on the *CCR5* cis-regulatory region.

1. Introduction

The CC-chemokine 5 receptor (CCR5) binds the β-chemokines CCL2, CCL3, CCL4, and CCL5, playing a fundamental role in the innate immune response and inflammatory processes [1]. It is expressed not only in cells and organs involved in immunity, but also in cells of the central nervous system and on the epithelium and endothelium of other tissues [2]. In addition to its constitutive function, CCR5 is used by the human immunodeficiency virus (HIV) as a co-receptor when entering a cell [3]. This explains why individuals homozygous for a 32-bp deletion (delta32) in the open reading frame (ORF) of the gene are relatively resistant to infection by HIV and heterozygous individuals show a delay in disease progression [4].

Great variability in the expression of CCR5 has been observed in T cells from homozygous individuals for the delta32 wild-type allele [5]. However, the polymorphisms found in the coding region of the *CCR5* gene could not explain the variation in disease risk among individuals and populations because they are, in most cases, found in low frequencies [6–8]. As an exception, Boldt et al. [9] described two new nonsynonymous mutations in Amerindians, one of them being relatively common (~10%).

These findings motivated the study of the promoter region of *CCR5*, revealing that specific haplotypes were associated with increased or decreased rates of disease progression in HIV-infected individuals [10].

The *CCR5* gene has a complex structure, with two promoters and an alternatively spliced cis-regulatory region located in the 5′ UTR of the primary transcript [11]. These features suggest that the 5′ UTR plays an important role in pre- and post-transcriptional control of *CCR5* expression and therefore could be an important target for natural selection.

Balancing selection upon the 5′ UTR region of *CCR5* is supported by several lines of evidence [12]: (a) its diversity is fourfold higher than that of other noncoding regions of the human genome [13]; (b) population differentiation, as measured by Fst, is significantly lower than for other markers; (c) there are more polymorphic sites at intermediate frequency than expected under neutral-equilibrium conditions, resulting in positive and significant Tajima's *D* value; and (d) haplotypes are connected by a network exhibiting two main clusters, separated by a relatively high level of divergence. Bamshad et al. [12], showed that the high genetic diversity observed for *CCR5* cis-regulatory region could not be accounted for by a high recombination or mutation rate, and that the low differentiation among populations was not an artifact of the sampling strategy used, strengthening the claim that this region is under balancing selection.

Interestingly, several examples of balancing selection upon regulatory regions of genes involved in immunity have recently been found, including *HLA-G* [14], *HLA-DPA1* and *DPB1* [15]. Specifically

* Corresponding author.
E-mail address: rfr@ib.usp.br (R.F. Ramalho).

0198-8859/10/$32.00 - see front matter © 2010 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.humimm.2010.05.022

for Amerindians, Mendes-Junior et al. [16] found a slight trend of balancing selection, analyzing a 14-bp insertion in the 3′ UTR of *HLA-G* gene. Moreover, in a survey of SNPs from 481 genes, Hughes et al. [17] found that 5′ UTRs of immune genes show significantly higher heterozygosity than those of nonimmune genes, among a total of 481 loci. Fumagalli et al. [18] also described evidence of balancing selection upon cis-regulatory regions of genes involved in innate immune response, including *CD55*, *CD151*, and *FUT2*. Overall, it appears that regulatory regions of genes involved in immunity may commonly possess signatures of balancing selection. One hypothesis to explain these observations is that genetic diversity of regulatory regions can expand the breadth of expression patterns, a trait that has the potential to enhance the fitness of individual, by allowing appropriate temporal and tissue-specific responses to be initiated [19].

Studies that investigate natural selection on candidate loci face an important challenge, which is the fact that genetic diversity is shaped by both selective and demographic processes. Therefore, considering the demographic history is an essential step in the study of genes that are candidates for natural selection [20]. The effect of demography can be dealt with by comparing the genes that are putatively selected to other regions of the genome, expected to reflect the effect of demographic history [21]. This approach allows the interaction between demographic and selective processes to be explicitly accounted for, and has been successfully used to identify genes under both balancing and directional selection [*e.g.*, 22].

Native American populations represent a particularly interesting dataset from the standpoint of how natural selection and demographic history interact. First, these populations occupied a region of the world with markedly different environments from those found in northeastern Asia, their likely place of origin, resulting in new modes of selection. Second, they represent an extreme in demography among human populations, with exceptionally low levels of polymorphism, a finding supported by analyses of a diverse set of markers (autossomes [23], mitochondrial DNA [24,25], X chromosome [26], and the Y chromosome [27–31]). In addition, Native Americans also show high levels of interpopulation differentiation, with a recent study of 29 populations, based on microsatellites, showing that in Amerindians, 8.1% of the total variance is due to differences among populations, compared with 3% in Africa, 1.4% in Asia, and 0.8% in Europe [23].

The low genetic variability of Amerindians results from the history of serial founder events that occurred as human populations migrated toward the Americas, resulting in a cumulative loss of genetic variation, whereas the high Fst is likely to result from the increased rates of genetic drift in Amerindian populations, a consequence of their reduced effective population sizes [31]. Moreover, diversity in the Americas has regional differences: eastern South America, which includes the Amazonian region, has the lowest heterozygosities, when compared with western South America, Central America, or North America [23].

Whereas the demographic features of the colonization process lead to loss of diversity, balancing selection can be expected to counter this effect, maintaining highly divergent lineages in the population, increasing heterozygosity beyond what would be expected in the absence of selection, and decreasing interpopulation differentiation [21,32,33]. In the present study, we examine genetic variation in the *CCR5* cis-regulatory region in Amazonian Amerindians. By studying a locus under balancing selection in a set of populations that have experienced increased rates of genetic drift, we are able to investigate how the joint effects of selection and demography combine in human populations and to examine whether one of these process can override the other. To attain this goal, we use neutrality tests based on empiric and simulation data to study the role of selection on *CCR5* cis-regulatory region of Amerindians while accounting for their demographic history.

2. Subjects and methods

2.1. Samples

We analyzed data from 62 individuals (2N = 124 chromosomes, where 2N is the number of chromosomes) from seven Native American populations from the Brazilian Amazonian region: Arara do Laranjal (2N = 26), Arara do Iriri (2N = 26), Kayapó (2N = 24), Xikrin (2N = 26), Mapuera-Katuena (2N = 4), Parakanã (2N = 6) and Waiãpí (2N = 12). Arara do Laranjal, Arara do Iriri, and Mapuera-Katuena belong to the Karib linguistic group; Kayapó and Xikrin belong to the Gê linguistic group; and Parakanã and Waiãpi belong to the Tupi linguistic group. The Arara do Iriri is a highly inbred group, probably recently founded by a single couple of siblings [34], and therefore, unless otherwise stated, were not included in the analyses. Blood samples were collected during health expeditions organized in 1990–1992 by Fundação Nacional do Índio (FUNAI) agents and researchers from Laboratory of Human and Medical Genetics (LGHM) of Federal University of Pará [UFPA]). Appropriate consent was obtained from tribes to use the DNA in genetic and anthropological studies. DNA was extracted by standard phenol/chloroform method described elsewhere [35].

2.2. Molecular biology techniques

To increase the amount of template DNA, we used whole-genome amplification via multiple displacement amplification, as implemented by the GenomiPhi v. 2 kit (GE Healthcare, Chalfont St Giles, Buckinghamshire, England), following the manufacturer's protocol. To assure ourselves that the whole-genome amplification protocol was not introducing spurious variation (*e.g.*, because of mutations associated with replication), we analyzed 122 pairs of reads from both genomic and whole-genome amplified DNA, distributed among 37 individuals and using seven different primers. We did not find any mismatch between genomic and whole-genome amplified sequences when considering the high quality base calls (Phred score ≥50), which are those that we used for our analyses. We concluded that Genomiphi shows high fidelity for our samples and did not introduce spurious mutations.

All samples were amplified by polymerase chain reaction (PCR) using primers *CCR5*L2 and *CCR5*R2b (5′ CTCATAGCTTCAGATAGATTAT 3′), designed to span 1009 bp of the cis-regulatory region of the *CCR5* gene. Standard amplification conditions were 50 ng of genomic DNA, 1.5 mmol/l of $MgCl_2$, 0.2 mmol/l of each dNTP, 0.2 μmol/l of each primer, and 0.5 U of Platinum Taq DNA polymerase (Invitrogen). PCR amplification was carried out in a 50-μl final volume in an MJ PTC-200 thermocycler (MJ Research, Inc.). Cycling conditions included an initial denaturation at 96°C for 3 minutes, 40 cycles of 96°C for 25 seconds, 54°C for 60 seconds, 72°C for 2 minutes, and a final extension step of 72°C for 8 minutes. PCR products were purified with ExoSAP-IT (GE Healthcare, Chalfont St Giles, Buckinghamshire, England). Sequencing reactions were carried out in a 10-μl final volume by using BigDye (Applied Biosystems) or ET terminators chemistry (GE HealthCare, Chalfont St Giles, Buckinghamshire, England) with the following internal primers: *CCR5*L2b (5′ TATGAGCACTTGGTGTTTGCC 3′), *CCR5*D, *CCR5*E, *CCR5*A, *CCR5*B, and *CCR5*I. All primers used in this study are described in Bamshad et al. [12], except *CCR5*R2b and *CCR5*L2b, which were designed by us. Sequencing reactions used 30 ng of purified PCR product, 2 μl of BigDye version 2.1 (Applied Biosystems) or 2 μl of DYEnamic ET terminator premix (GE HealthCare, Chalfont St Giles, Buckinghamshire, England), 1 μmol/l of sequencing primer. The sequencing reactions were submitted to 40 cycles of 95°C for 20 seconds, 56.5°C for 15 seconds, and 60°C for 1 minute. Before automated DNA analysis, sequencing reactions were precipi-