GENETICS

Indigenous people from Amazon show genetic signatures of pathogen-driven selection

Cainã M. Couto-Silva[1]†, Kelly Nunes[1]†, Gabriela Venturini[2,3], Marcos Araújo Castro e Silva[1,4], Lygia V. Pereira[1], David Comas[4], Alexandre Pereira[2,3], Tábita Hünemeier[1,5]*

Ecological conditions in the Amazon rainforests are historically favorable for the transmission of numerous tropical diseases, especially vector-borne diseases. The high diversity of pathogens likely contributes to the strong selective pressures for human survival and reproduction in this region. However, the genetic basis of human adaptation to this complex ecosystem remains unclear. This study investigates the possible footprints of genetic adaptation to the Amazon rainforest environment by analyzing the genomic data of 19 native populations. The results based on genomic and functional analysis showed an intense signal of natural selection in a set of genes related to *Trypanosoma cruzi* infection, which is the pathogen responsible for Chagas disease, a neglected tropical parasitic disease native to the Americas that is currently spreading worldwide.

INTRODUCTION

America presents a wide range of ecoregions that were quickly explored and occupied by the first humans to reach the continent (*1*). This remarkable human migration likely required different genetic and cultural adaptation patterns to ensure successful subsistence (*2*–*5*). The Amazon rainforest is one of the main ecoregions of the American continent, which is well known for its tropical climate and extraordinary biological diversity. Despite being a nutrient-rich environment, it is also hostile, presenting different obstacles to long-term human survival. Several challenges, including the instability of food resources (*6*, *7*), low light penetration (*8*), and high diversity of pathogens (*9*), probably contribute to strong selective pressures for human survival and reproduction in this ecoregion.

The Amazon region, which encompasses the single largest tropical rainforest and nine South American countries, is virtually unrivaled in scale, complexity, and opportunity. It is currently populated by 1 million indigenous people, divided into approximately 300 different ethnic groups (*10*). Ecological conditions in the region have historically been favorable for transmitting numerous tropical diseases, especially vector-borne diseases (*11*–*13*). Although there are numerous studies on postcontact epidemics (*14*) in the Americas, historical data on precontact diseases (i.e., diseases native to the Americas) are inadequate. However, it has been reported that tuberculosis (*Mycobacterium tuberculosis*) (*15*, *16*) and Chagas disease (i.e., American trypanosomiasis) (*17*) were present long before the Europeans arrived. Chagas disease is a vector-borne disease caused by the protozoan *Trypanosoma cruzi*, and it is usually transmitted through different triatomine bugs in endemic areas. The oldest record of *T. cruzi* in South American human archeological remains dates back to 9000 years in mummies from northern Chile and southern Peru (*18*). Human remains infected with *T. cruzi* were also found in Brazil about 7000 years before the present (yr B.P.) (*19*). There are, however, not many studies on the adaptations to the rainforest including Amazonian populations. Most of these are limited to a few individuals from the western Amazonia (*20*, *21*). To date, knowledge regarding genetic adaptations in humans within this complex ecosystem is largely unknown. Motivated by this lack of knowledge, we searched for possible footprints of genetic adaptation to the Amazon rainforest environment by analyzing the genomic data of 118 nonadmixed individuals belonging to 19 native populations (table S1). We were specifically interested in identifying signals for positive natural selection related to tropical diseases.

To search for signals of positive selection, we applied two distinct approaches: (i) Population Branch Statistics (PBS), which identify alleles that have experienced strong changes in frequency in one population relative to two reference populations (*19*), and (ii) Cross-Population Extended Homozygosity Haplotype (XP-EHH) statistics, which contrast the extended haplotype homozygosity within and between populations (*22*). We then explored these results through gene pathway enrichment analyses [the Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS) (*23*), Gene Ontology (GO) (*24*), Kyoto Encyclopedia of Genes and Genomes (KEGG) (*25*), and Reactome (*26*) and the Gene Set Analysis Toolkit (GESTALT) (*27*)]. To formally test whether natural selection underlies the cases of extreme differentiation, rejecting genetic drift as a cause, PBS values were compared against those obtained with neutral coalescent simulations generated according to a plausible demographic scenario for the peopling of South America. Last, we performed a functional follow-up analysis to characterize the role of the putative selected gene.

RESULTS AND DISCUSSION

Genetic adaptation to the Amazon rainforest

The upper distribution of the combined positive selection indices (PBS + XP-EHH) is highly correlated to cardiovascular and metabolic traits (*MTRR*, *DNAJA4*, *KCMA1*, and *MTPN*), immune-related traits (*KCMA1* and *GCA*), and pathogen infection (table

[1]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, 05508090, Brazil. [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [3]Laboratório de Genética e Cardiologia Molecular, Instituto do Coração, Hospital das Clínicas da Faculdade de Medicina da Universidade de Sao Paulo, Sao Paulo, SP, Brazil. [4]Institut de Biologia Evolutiva, Departament de Medicina i Ciències de la Vida, Universitat Pompeu Fabra, Barcelona 08003, Spain. [5]Institut de Biologia Evolutiva (CSIC/Universitat Pompeu Fabra), Barcelona 08003, Spain.
†These authors contributed equally to this work.
*Corresponding author. Email: hunemeier@gmail.com

S2). Among the three genes that were highlighted in our analysis, *PPP3CA* and *DYNC1I1* were suggestively associated with *T. cruzi* seropositivity and immune response (Fig. 1 and table S2) (*28*), while *NOS1AP* was related to the mosquito bite reaction (Fig. 1) (*29*). The strongest selection signature found in Amazonian populations was around the *PPP3CA* gene region (Fig. 1 and figs. S1 and S4). Neutral coalescent simulations indicate that these deviations in allele frequency (rs2659540 G>A) are statistically significant ($P < 0.0075$; fig. S3), which is consistent with the action of ositive selection as opposed to a genetic drift effect in the Amazonian populations. We estimated a selection starting time of 7500 yr B.P. [confidence interval (CI) = 1560 to 12,035] and a selection coefficient of 0.05 (CI = 0.015 to 0.20) (fig. S6). This result suggests that selection started acting on this gene after the split between Amazonian, Coastal Pacific, and Andes populations (*30*). A previous study on ancient tissues of 283 individuals, dating from 9000 to 450 yr B.P., from South America′s coastal area of southern Peru and northern Chile showed a slight increase in the infection rate over time (*18*). These findings support our inference of a selection signal, which is likely exclusive to Amazonian rainforest populations.

The GO enrichment analysis based on both the XP-EHH (95th percentile) and PBS (95th percentile) analysis showed statistically significant enrichment of genes involved in the response to protozoan and eosinophil chemotaxis, which have a known role in parasitic infections (Table 1 and table S3). When FUMA GWAS and Enrichr were applied, associations with other traits were detected, including novelty-seeking behavior, metabolic traits, and systolic/diastolic pressure (tables S4 and S5). Several studies, including Brazilian indigenous populations, have shown high rates of obesity and cardiopathies (*31*, *32*). For instance, in the Xavante population, 66% of individuals suffer from obesity, diabetes, or coronary heart disease (*33*). Bergey *et al.* (*34*) showed a strong signal of convergent selection in heart-related networks in studying Asian and African hunter-gatherers as compensatory adaptations to their short stature, which may possibly also be the case in Native Americans. Furthermore, we also identify associations that could raise a hypothesis for increased novelty-seeking behavior, currently recognized through caffeine and nicotine consumption (tables S4 and S5) (*35*, *36*). Novelty-seeking behaviors may have been important in

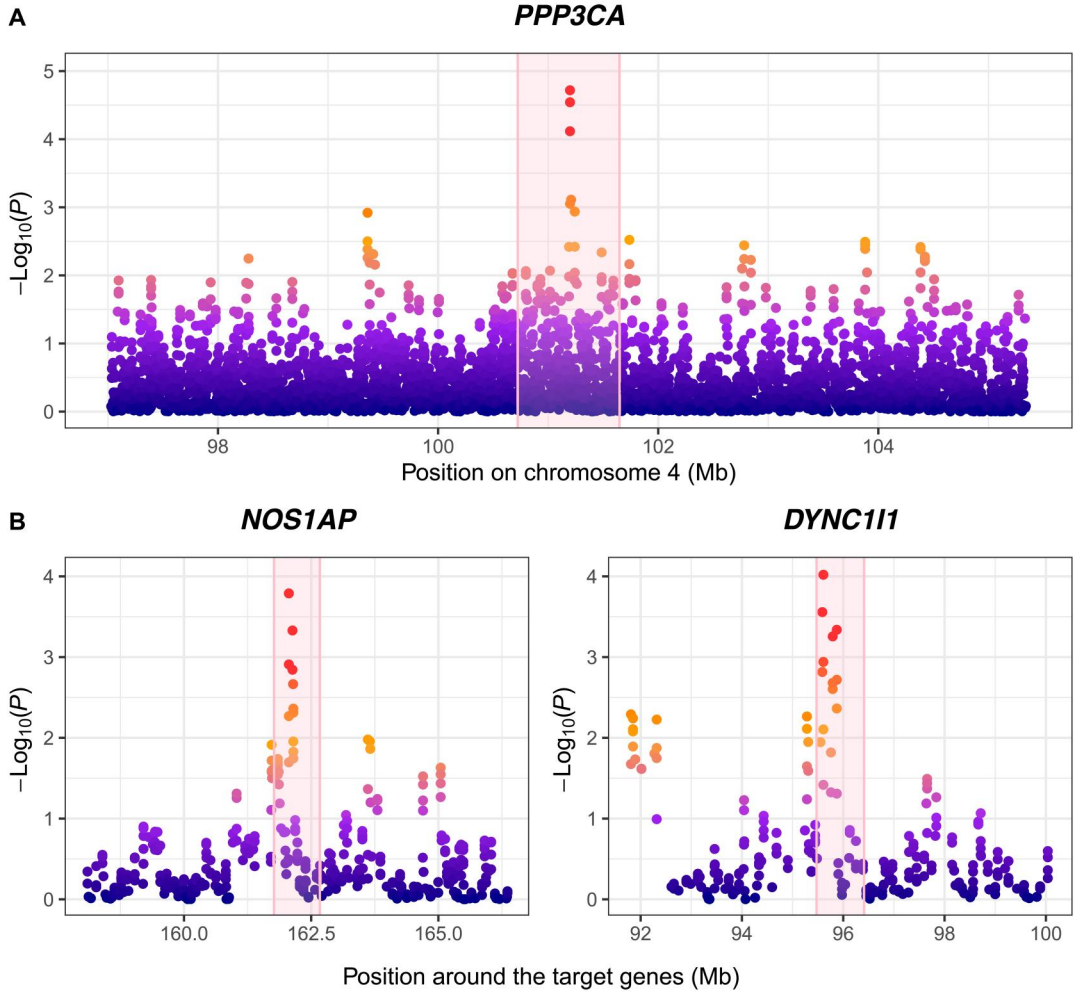

Fig. 1. PBS-windowed scores for target genes. Log_{10} of PBS *P* values around the (**A**) *PPP3CA* and (**B**) *NOS1AP* and *DYNC1I1* genes, chromosomes 1 and 4, respectively. Graphs show up to 4 Mb away from the initial and final gene position. Pink shadow areas correspond to 300 kb of the gene limit description. All these genes have the highest PBS values for their respective chromosomes.