Environmental Toxicology and Pharmacology 62 (2018) 30–39

Contents lists available at ScienceDirect

Environmental Toxicology and Pharmacology

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

Unraveling *CYP2E1* haplotypes in alcoholics from Central Brazil: A comparative study with 1000 genomes population

Jheneffer Sonara Aguiar Ramos[a,b], Leandro Prado Felício[c], Alessandro Arruda Alves[a,b], Mariana Paiva Lopes[a,b], Thannya Nascimento Soares[a,d], Daniela de Melo e Silva[a,b,*]

[a] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal de Goiás, Goiânia, GO, Brazil
[b] Laboratório de Mutagênese, Universidade Federal de Goiás, Goiânia, GO, Brazil
[c] Laboratório de Farmacologia Bioquímica e Molecular, Universidade Federal de Goiás, Goiânia, GO, Brazil
[d] Laboratório de Genética e Biodiversidade, Universidade Federal de Goiás, Goiânia, GO, Brazil

ARTICLE INFO

Keywords:
Ethanol
Linkage disequilibrium
1000 Genomes Project
Genetic diversity

ABSTRACT

We evaluated genetic variability of single nucleotide polymorphisms (SNPs) situated in the *CYP2E1* gene promoter in alcoholics. We also compared 1000 Genomes Project of *CYP2E1* polymorphisms with frequencies of genotypes and haplotypes. Eight variation points were exclusively found in Brazilians. The allelic distributions of the rs3813867, rs2031920 and rs2031921 polymorphisms in the *CYP2E1* showed that the wild alleles (G, C, T, respectively) had higher frequencies in both groups, alcoholic (96%, 96%, 96%) and a control group (95.8%, 94.9%, 94.9%), when compared to the mutated allele (C, T, C, respectively). The variation points, rs3813867, rs2031920 and rs2031921 showed strong linkage disequilibrium (LOD ≥ 2, D ' = 1). South Asian populations presented larger LD blocks compared to the other populations. Our results showed that the allelic frequencies were markedly different among ethnicities and have contributed to the knowledge regarding the distribution among ethnic groups, being associated to alcohol consumption worldwide.

1. Introduction

The harmful use of alcohol is an important public health problem and an estimated 3.3 million (5.9%) people died from alcohol related conditions in the world annually (World Health Organization, 2014). The complex process of gene-environment interaction could result in an individual response due to the use of alcohol on the organism (Zakhari, 2006; Wang et al., 2016). The *CYP2E1* gene (OMIM no. 124040) is located on chromosome 10q26.3. The human *CYP2E1* gene is mainly expressed in the liver and extrahepatic tissues, spans over 11 kb, and presents nine exons which codes for a protein of 493 amino acid residues with a molecular weight of ~ 57 kDa. CYP2E1 is induced by chronic alcohol consumption and assumes an important role in metabolizing of alcohol into acetaldehyde at elevated ethanol concentrations. Previous studies have shown that the activity of CYP2E1 is increased up to 10-fold after chronic alcohol consumption or after long-term ethanol intake (Ramchandani et al., 2013; Wang et al., 2016).

Besides generation of acetaldehyde, the ethanol oxidation by CYP2E1 also produces oxygen-containing molecules called free radicals or reactive oxygen species (ROS) which contribute to liver injury due to alcohol consumption (Koop, 2006; Plemenitas et al., 2015).

There is also a high individual variation in ethanol metabolism and differences in accumulation of acetaldehyde and ROS production, mainly due to some genetic polymorphisms in the main ethanol and acetaldehyde metabolizing enzymes (Khan et al., 2009; Wang et al., 2016).

Genetic polymorphisms in the *CYP2E1* promoter are located at a putative binding site for the transcription factor hepatocyte nuclear factor (HNF-1). Two of the promoter polymorphisms, rs2031920 (−1055 C > T transition) and rs3813867 (−1295 G > C

Abbreviations: CYP2E1, cytochrome P450, family 2, subfamily E, polypeptide 1; ROS, reactive oxygen species; HNF-1, transcription factor hepatocyte nuclear factor; SNP, single nucleotide polymorphism; YRI, Yoruba in Ibadan, Nigeria; LWK, Luhya in Webuye, Kenya; GWD, Gambian in Western Divisions in the Gambia; MSL, Mende in Sierra Leone; ESN, Esan in Nigeria; TSI, Toscani in Italia; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian Population in Spain; CHB, Han Chinese in Bejing, China; JPT, Japanese in Tokyo, Japan; CHS, Southern Han Chinese; CDX, Chinese Dai in Xishuangbanna, China; KHV, Kinh in Ho Chi Minh City, Vietnam; PJL, Punjabi from Lahore, Pakistan; BEB, Bengali from Bangladesh; STU, Tamil from the UK; ITU, Indian Telugu from the UK; CEU, Utah Residents with Northern and Western Ancestry; ASW, Americans of African Ancestry in SW USA; ACB, African Caribbeans in Barbados; MXL, Mexican Ancestry from Los Angeles USA; PUR, Puerto Ricans from Puerto Rico; CLM, Colombians from Medellin, Colombia; PEL, Peruvians from Lima, Peru; GIH, Gujarati Indian from Houston, Texas; PCR, polymerase chain reaction; LD, linkage disequilibrium; MAF, minimum allele frequency

* Corresponding author at: Estrada do Campus, s/n, Instituto de Ciências Biológicas, Departamento de Genética, Universidade Federal de Goiás, Bairro, Campus Universitário, CEP: 74690-900, Goiânia, GO, Brazil.
E-mail address: danielamelosilva@ufg.br (D. de Melo e Silva).

https://doi.org/10.1016/j.etap.2018.06.004
Received 27 February 2018; Accepted 18 June 2018
Available online 19 June 2018
1382-6689/

transversion) which are in complete linkage disequilibrium, had been associated with altered of transcriptional activity of CYP2E1 (Ramchandani et al., 2013). Increased transcriptional activity is associated with an increase in acetaldehyde level in alcoholics that contributes to the genotoxic effect of alcohol consumption. Different frequency of these genetic variants has been reported in distinct human populations and various ethnic groups (Khan et al., 2009; Tang et al., 2010). Further studies in several populations are necessary to understand the regulation mechanisms of the CYP2E1 gene and the effects of CYP2E1 variants on metabolism of ethanol (Cederbaum, 2008; Ramchandani et al., 2013). It is well established the influence of the ethnic component in the distribution of *CYP2E1* genetic polymorphism especially in those with admixed ancestry, such as the Brazilian population with the contribution of native Amerindians, European colonizers and Africans arrived during slavery provide immense opportunity in this regard. Therefore, the aim of this study was to evaluate the genetic variability of single nucleotide polymorphisms (SNPs) in *CYP2E1* gene promoter in alcoholic individuals from the city of Goiânia, Goiás, Brazil. We also compared 1000 Genomes Project data of *CYP2E1* polymorphisms with frequencies of genotypes and haplotypes in our study population.

2. Material and methods

2.1. Subjects and sample collection

Eighty-four (84) individuals who were diagnosed at the Psychosocial Care Center-Alcohol and Drugs (CAPS A/D) as alcoholics were included in this study. The control group consisted of ninety-six (96) individuals with a similar mean age to the case group. The individuals of both groups lived in Goiânia, Goiás, Brazil. For the composition and classification of the case and control groups, individuals older than 18 years who volunteered to participate in this study were considered. Data regarding medical conditions and lifestyle of both groups were collected via administration of systematic questionnaires. The Research Ethics Committee (CEP number 476938) of the Federal University of Goiás (UFG) approved the study protocol and all participants provided written informed consent. After the questionnaire and informed consent forms. 10 ml of peripheral blood was collected from all individuals. DNA was extracted using the extraction kit Blood ReliaPrep ™ Miniprep System (Promega®, Fitchburg, Wisconsin, USA), according to the manufacturer's instructions. The concentration and purity of the DNA were quantified in NanoVue Plus™ spectrophotometer (GE Healthcare, Little Chalfont, UK) and determined using a 260/280 nm optic density ratio.

2.2. Polymerase chain reaction

Fragment of *CYP2E1* promoter was amplified by PCR reaction, with the following primers: 5′ CCGTGAGCCAGTCGAGTCTA 3′- and 5′ GGA AAGAGTGAGTCAACCAATTCTG 3′. The volume used for each PCR was 25 µL containing DNA 100 ng/µL, 1.5 mM magnesium chloride, 1 mM of primers, 0.2 mM dNTPs, 1X buffer and 1 U of Taq Platinum. The PCR reaction was performed on a Veriti® 96-Well Thermal Cycler thermocycler (Applied Biosystems ™, Foster City, California, USA) and consisted of 95 °C for 1 min and 40 cycles of 95 °C for 10 s, 61 °C for 20 s, 72 °C for 25 s and then a final extension at 72 °C for 5 min. The amplified fragments were visualized on 2% agarose gel stained with ethidium bromide

2.3. DNA sequencing

Seven microliters of the DNA were incubated with 0.5 U/µl of exonuclease I and 0.5 U/µl alkaline phosphatase (USB Corporation) for 90 min at 37 °C followed by 80 °C for 20 min. The amplified PCR product was sequenced using the Sanger methodology on an ABI PRISM® 3500 Genetic Analyzer automatic DNA sequencer (Applied Biosystems ™, Foster City, California, USA), using the BigDye® Terminator Cycle Sequencing Kit Standard, version 3.1 (Applied Biosystems ™, Foster City, California, USA). Sequences were analyzed using the CLC Sequence Viewer version 7.6.1 (CLC bio, Aarhus, Dinamarca) and BioEdit Sequence Alignment Editor version 7.2.5 (Hall, 1999) or SeqScape 6.0 (Applied Biosystems ™, Foster City, California, USA).

2.4. Statistical analysis

This study also used data from the promoter region of the *CYP2E1* gene obtained from the official Website of the 1000 Genomes Project (www.1000genomes.org) (The 1000 Genomes Project Consortium, 2012; The 1000 Genomes Project Consortium, 2015), which assessed 2504 individuals from 26 populations distributed in five regions: Africa (AFR), Europe (EUR), East Asia (EAS), South Asia (SAS) and the Americas (AMR). The bioinformatics analysis methods were defined using the same method described by (Felício et al., 2014). The promoter region of the *CYP2E1* gene was filtered from the downloaded files in. vcf format of the official Website of the 1000 Genomes Project and was concatenated with the data from the samples obtained for this study (alcoholics and controls). The resulting. vcf file was converted to the Genepop format using PGDSpider software, version 2.0.19 (Lischer and Excoffier, 2012).

For variations (rs3813867, rs532607936, rs6413422, rs569232477, rs2031920, rs2031921, rs181373843, rs147346339, rs3813870, rs577399320, rs546046390, rs556363528, rs576417537, rs2031922, rs368294078, rs527327421, rs2854143) found in one analysis, but not in the other, for example, the variation points occurring exclusively on the analysis of the Brazilian samples and not in the populations of the 1000 Genomes Project and vice versa were considered monomorphic in the group where they do not varied, and the base was considered to be present in the standard sequence used in the analysis (NM_000773.3). The allelic and genotypic frequencies in each of the studied locus were obtained by direct counting and the adherences of the genotype frequencies to expectations under Hardy-Weinberg equilibrium were tested by the exact test of Guo and Thompson (Guo and Thompson, 1992). All statistical analysis was performed using GenePop software version 4.0 (Raymond and Rousset, 1995). The construction of the UPGMA dendrogram was performed using R software (R Development Core Team R (computer software), 2011), and the allele frequencies of rs3813867, rs2031920 and rs2031921 polymorphisms of the *CYP2E1* gene were used for the analysis of proximity between 26 populations from the 1000 Genomes Project and the case and control groups. The validation of the groups was determined by the cophenetic correlation coefficient (Sokal and Rohlf, 1962) using R software (R Development Core Team R (computer software), 2011). For the analysis of possible linkage disequilibrium (LD), the. vcf file was converted into the Linkage ped format, and inference pattern linkage disequilibrium (LD) in the region was determined using Haploview® software, version 4.2 (Barrett et al., 2005). The LD images were generated using Haploview SNPs with the least frequent allele having at least a minimum allelic frequency (MAF) of 0.01. The regions with high LD (segregation blocks) were inferred by the method of confidence intervals (Gabriel et al., 2002).

The region evaluated here, i.e. the *CYP2E1* promoter, were downloaded and concatenated with our data from Brazil. Allele and genotype frequencies were calculated with GenePop 4.1 (Rousset, 2008). The association between each variable site was evaluated by Haploview 4.2 (Barrett et al., 2005). Linkage disequilibrium (LD) plots were generated using variable sites with a minimum allele frequency (MAF) of 0.01 and segregation blocks were inferred by the confidence interval method (Gabriel et al., 2002). The haplotypes were inferred using the Bayesian method implemented in PHASE (Stephens et al., 2001) and the maximum-likelihood Expectation-Maximization (EM) algorithm implemented in Partition-Ligation-Expectation-Maximization (PL-EM) (Qin et al., 2002). For this, a script called HaploRunner (available at