Molecular Immunology 83 (2017) 115–126

Contents lists available at ScienceDirect

Molecular Immunology

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

HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographicaly distinct population samples of Brazil and Cyprus

Erick C. Castelli[a,b,*,1], Petroula Gerasimou[c,1], Michelle A. Paz[b], Jaqueline Ramalho[b], Iane O.P. Porto[b], Thálitta H.A. Lima[b], Andréia S. Souza[b], Luciana C. Veiga-Castelli[d], Cristhianna V.A. Collares[d], Eduardo A. Donadi[d], Celso T. Mendes-Junior[e], Paul Costeas[c]

[a] Department of Pathology, School of Medicine, UNESP – Univ. Estadual Paulista, Botucatu, State of São Paulo, Brazil
[b] Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), Sector 5, School of Medicine, UNESP – Univ. Estadual Paulista, Botucatu, State of São Paulo, Brazil
[c] Karaiskakio Foundation, Nicosia, Cyprus
[d] Division of Clinical Immunology, Department of Medicine, School of Medicine of Ribeirão Preto, University of the State of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil
[e] Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil

ARTICLE INFO

Article history:
Received 17 November 2016
Received in revised form 18 January 2017
Accepted 20 January 2017

Keywords:
HLA-G
Next generation sequencing
NGS
Massive parallel sequencing
Brazil
Cyprus
Haplotypes
Variability
Polymorphisms

ABSTRACT

The HLA-G molecule presents immunomodulatory properties that might inhibit immune responses when interacting with specific Natural Killer and T cell receptors, such as KIR2DL4, ILT2 and ILT4. Thus, HLA-G might influence the outcome of situations in which fine immune system modulation is required, such as autoimmune diseases, transplants, cancer and pregnancy. The majority of the studies regarding the *HLA-G* gene variability so far was restricted to a specific gene segment (i.e., promoter, coding or 3′ untranslated region), and was performed by using Sanger sequencing and probabilistic models to infer haplotypes. Here we propose a massively parallel sequencing (NGS) with a bioinformatics strategy to evaluate the entire *HLA-G* regulatory and coding segments, with haplotypes inferred relying more on the straightforward haplotyping capabilities of NGS, and less on probabilistic models. Then, *HLA-G* variability was surveyed in two admixed population samples of distinct geographical regions and demographic backgrounds, Cyprus and Brazil. Most haplotypes (promoters, coding, 3′UTR and extended ones) were detected both in Brazil and Cyprus and were identical to the ones already described by probabilistic models, indicating that these haplotypes are quite old and may be present worldwide.

1. Introduction

The *HLA-G* locus is a non-classical Major Histocompatibility Complex (MHC) class I gene, which presents immunomodulatory properties. Its expression is limited to few tissues in physiological conditions such as cornea (Le Discorde et al., 2003), thymus (Lefebvre et al., 2000) and placenta (Hunt et al., 2006), where it was firstly described (Kovats et al., 1990). HLA-G expression at placenta and trophoblastic tissues is a key feature for pregnancy maintenance and maternal-fetal tolerance, along with the expression of HLA-E and HLA-C (Gregori et al., 2015; Carosella et al., 2003; Flores et al., 2007; Ishitani et al., 2006; Djurisic and Hviid, 2014; Hiby et al., 2010; Chazara et al., 2011). The structure of the *HLA-G* segment encoding the peptide-binding groove is very similar to the one found for HLA-A2 alleles, but its antigen presentation properties are quite limited to self-antigens, in order to promote immune tolerance (Diehl et al., 1996; Munz et al., 1999a; Munz et al., 1999b).

HLA-G might inhibit immune responses when interacting with specific Natural Killer and T cell receptors such as KIR2DL4, ILT2 and ILT4 (Favier et al., 2010; LeMaoult et al., 2005; Rajagopalan and Long, 1999; Shiroishi et al., 2003; Colonna et al., 1997; Kamishikiryo and Maenaka, 2009 Shiroishi et al., 2003; Colonna et al., 1997; Kamishikiryo and Maenaka, 2009). Thus, the HLA-G molecule might influence the outcome of situations in which fine immune system modulation is required, such as during autoimmune diseases

* Corresponding author at: Departamento de Patologia, Faculdade de Medicina, Unesp, Botucatu – SP, CEP 18618970, Brazil.
E-mail address: castelli@fmb.unesp.br (E.C. Castelli).
[1] These authors contributed equally to this study.

http://dx.doi.org/10.1016/j.molimm.2017.01.020
0161-5890/

(Brenol et al., 2012; Aractingi et al., 2001; Rizzo et al., 2008), transplants (Misra et al., 2014; Crispim et al., 2008), cancer (Dunker et al., 2008; Cao et al., 2011; Dong et al., 2012) and pregnancy (Donadi et al., 2011; Larsen et al., 2010; Christiansen et al., 2012; Hylenius et al., 2004; Hviid, 2006; Nilsson et al., 2016). The interaction between HLA-G and KIR2DL4 may up regulate the expression of IFN-γ, which in turn mediates maternal vascular modifications (Goodridge et al., 2009; Tan et al., 2009). The receptors ILT2 and ILT4 are found on the surface of both NK and T cells, and present inhibitory properties (LeMaoult et al., 2005; Lefebvre et al., 2002). These interactions have been responsible for limiting *HLA-G* coding region diversity, leaving a signature of purifying selection (Mendes-Junior et al., 2013).

Due to its immunomodulatory role, the *HLA-G* coding region is highly conserved when compared with the HLA classical counterparts, *HLA-A*, *HLA-B* and *HLA-C*. To date, the IPD and IMGT/HLA database (Robinson et al., 2015) (version 3.26.0) describes 53 coding alleles generating 18 protein variants (and 2 null alleles). In addition, a recent study using data from the *1000 Genomes Project* (Genomes Project et al., 2015) showed that most of the variable sites detected for *HLA-G* are either intronic or synonymous mutations (Castelli et al., 2014a). However, this study was conducted using the *1000 Genomes* data from phase 1, which is characterized by low coverage and does not include many autochthonous and admixed populations. Thus, it is possible that the *HLA-G* variability is greater than our current knowledge.

Besides the *HLA-G* variability assessed at the *1000 Genomes* data (Castelli et al., 2014a; Sabbagh et al., 2014; Gineau et al., 2015), most studies regarding the variability of the *HLA-G* regulatory segments, either promoter or 3'untranslated region (3′UTR), were performed on specific population samples, mainly from USA and China (Tan et al., 2005), Brazil (Castelli et al., 2010; Castelli et al., 2011; Lucena-Silva et al., 2012; Lucena-Silva et al., 2013; Porto et al., 2015; de Albuquerque et al., 2016; Consiglio et al., 2011; Veit et al., 2012; Veit et al., 2014; Zambra et al., 2016; Santos et al., 2013; Catamo et al., 2014; Catamo et al., 2015), France (Martelli-Palomino et al., 2013), Italy (Catamo et al., 2015; Garziera et al., 2015; Sizzano et al., 2012), United Kingdom (Hviid et al., 2006), Denmark (Nilsson et al., 2016) and Africa (Garcia et al., 2013; Courtin et al., 2013; Carlini et al., 2013). Since the IPD-IMGT/HLA, the official database of known HLA alleles (Robinson et al., 2015), does not consider most of nucleotides at these segments, variable sites identified in other populations may still be unknown. With the advent of new sequencing approaches, such as new generation (NGS) or massively parallel sequencing techniques, now we have the opportunity to characterize the entire *HLA-G* segment with little effort.

The aims of this study were to propose a strategy to evaluate the entire *HLA-G* segment by using massively parallel sequencing, and to characterize the *HLA*-G variability of two admixed population samples of distinct geographical regions and demographic backgrounds, Cyprus and Brazil. Cyprus is an island country in the northeastern part of Eastern Mediterranean with an area of 9251 square kilometers (3572 square miles) and, by the end of 2014, Cyprus population was estimated at 847,000 inhabitants (http://www.cystat.gov.cy). Cyprus was settled 3500 years ago by the Mycenaean Greeks and many other cultures joined afterward, including Phoenicians, Assyrians, Egyptians, Ottomans and others, leaving behind a mosaic of different cultures and periods. Cyprus was a British colony until 1960 when it became a politically independent country. In addition, Cyprus has two official languages, Greek and Turkish. Contrastingly, Brazil is the fifth larger country in the world and the largest country in South America, with a population that exceeds 205 million inhabitants. Its official language is Brazilian Portuguese. Brazilians are considered one of the most admixed populations in the world and a great repository of genetic variation because it results from five centuries of ongoing interethnic admixture, mainly composed by Europeans, Africans and Native Americans.

2. Methods

2.1. HLA-G amplification

HLA-G variability was surveyed in 500 samples from the state of São Paulo, Southeastern Brazil, and from Cyprus. The Brazilian sample is composed of a population control group of 315 healthy individuals (80% female), with a mean age of 31.5 years old. According to self-reported ethnicity, individuals were classified as Euro-Brazilians (77.46%), Mulattoes (14.28%), Afro-Brazilians (4.13%) and Asians (3.5%), with 0.63% lacking this information. The Cypriot sample is composed of a random population control group of 185 healthy individuals (42.16% female), recruited from the Cypriot Bone Marrow registry, all Greek-Cypriots. Age was not reported. All Brazilian participants signed an informed consent before blood withdraw and this study protocol was reviewed and approved by the Human Research Ethics Committee from the School of Medicine – Unesp/Brazil (Protocol #24157413.7.0000.5411). Written informed consent was obtained from all participants from Cyprus, and the study was reviewed and approved by the Cyprus National Bioethics Committee (Protocol # ΕΕΒΚ/ΕΠ/2013/20).

The *HLA-G* amplification was performed by using two different approaches. Samples from Brazil were amplified in a single amplicon, comprehending the segment between nucleotides 29,794,114 and 29,799,118 considering the sequence available for chromosome 6 (human genome assembly hg19). Amplification was carried out using primers 5′-ACACTCATAATTCATTCATTCAGC-3′ and 5′-TCTTCTGATAACACAGGAACTTC-3′ in a final volume of 50 μL, containing 2.5 mM of dNTPs (Invitrogen, EUA), 10 pmol of each primer, 1.25 units of DNA polymerase (PrimeStar GXL, Takara) and 1X the PCR buffer supplied with the DNA polymerase and 50 ng of genomic DNA. Cycling conditions followed the recommended for PrimeStar GLX, i.e., 30 cycles of 98 °C for 10 s, annealing at 60 °C for 15 s and extension at 68 °C for 6 min. Amplification was evaluated on 1% agarose gel stained with GelRed® (Biotium™, Hayward, USA).

Samples from Cyprus were amplified as two overlapping amplicons. The first amplicon encompasses nucleotides 29,794,201 to 29,796,012 (human genome draft assembly hg19), using primers 5′-ACATTCTAGAAGCTTCACAAGAATG-3′ and 5′-TGGGCCTTGGTGTTCCGTG-3′. The second amplicon encompasses nucleotides 29,795,807 to 29,798,880 using primers 5′-GGTCGGGCGGGTCTCAA-3′ and 5′-TGGAAAGTTCTCATGTCTTCCA-3′. All reactions were prepared in 25 μL final volume, in the presence of 1X reaction buffer, 200 mM of each dNTP, 1.5-2.0 mM $MgCl_2$, 5U *Taq* DNA polymerase (QIAGEN), 20 pmole of each primer and more than 25 ng of genomic DNA. Reactions for the first amplicon were carried out with an initial denaturation at 95 °C for 5 min, 40 cycles of denaturation at 95 °C for 30 s, annealing at 56 °C for 45 s and extension at 72 °C for 2 min, followed by a final extension step at 72 °C for 10 min. Reactions for the second amplicon were carried out with an initial denaturation at 95 °C for 5 min, 40 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 45 s and extension at 72 °C for 1 min, followed by a final extension step at 72 °C for 10 min.

For both methods the *HLA-G* gene was amplified encompassing the *HLA-G* 5′ promoter segment, the 5′ untranslated region, its complete coding region (including introns), and the 3′ untranslated region up to nucleotide +3273. After amplification, amplicons were purified by using Illustra ExoProStar (GE Healthcare), quantified by using Qubit dsDNA High-Sensitivity Assays (ThermoFisher Sci-