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
Available online 27 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/© 2017 Elsevier Ltd. All rights reserved.

(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 # EEBK/EΠ/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-

Table 1
HLA-G diversity indices detected both in Brazil and Cyprus.

Summary [a]	***Entire HLA-G***	**Promoter and 5'UTR**	**Coding Region**	**3'UTR Segment** [b]
Heterozygous genotypes directly phased by GATK	78.65%	88.90%	72.06%	74.03%
Missing alleles (%) before imputation	1.21%	0.03%	2.00%	0.04%
Nucleotide diversity	0.00775 ± 0.00373	0.00577 ± 0.00296	0.00635 ± 0.00311	0.03040 ± 0.01546
Nucleotide diversity in Brazil	0.00779 ± 0.00376	0.00587 ± 0.00301	0.00639 ± 0.00314	0.03019 ± 0.01537
Nucleotide diversity in Cyprus	0.00763 ± 0.00368	0.00554 ± 0.00286	0.00623 ± 0.00306	0.03070 ± 0.01564
Gene diversity	0.9040 ± 0.0051	0.8048 ± 0.0074	0.8786 ± 0.0053	0.8138 ± 0.0059
Gene diversity in Brazil	0.9006 ± 0.0069	0.8167 ± 0.0089	0.8760 ± 0.0068	0.8115 ± 0.0074
Gene diversity in Cyprus	0.9037 ± 0.0072	0.7805 ± 0.0136	0.8780 ± 0.0080	0.8168 ± 0.0102
Tajima's D [c]	2.41132, *P*=0.0134	1.99464, *P*=0.0293	2.67735, *P*=0.0067	0.83866, *P*=0.1693
Tajima's D in Brazil [c]	2.42413, *P*=0.0107	1.84295, *P*=0.0368	2.68924, *P*=0.0069	1.16837, *P*=0.1108
Tajima's D in Cyprus [c]	2.53007, *P*=0.0078	2.01736, *P*=0.0251	2.83267, *P*=0.0030	0.93511, *P*=0.1589

[a] When the population is not specified (e.g., Brazil or Cyprus), this data considers both samples pooled together. [b] In order to compare the sequences detected here with previously published haplotypes, this segment considers only the last *HLA-G* exon. However, there are at least 28 nucleotides in a previous exon that is also part of the 3′UTR, but no variable sites were detected there. [c] A p-value was computed by the comparison of the estimated statistic to a distribution of estimates computed for 10,000 random samples of the same sample size and level of polymorphism as the observed data, and represents the probability of obtaining a simulated Tajima's D larger than the observed.

entific, USA) and normalized to the proper concentration prior to library preparation.

2.2. Library preparation and sequencing

Amplicons were sequenced at the MiSeq Platform (Illumina, Inc.). Sequencing libraries were prepared by using Nextera XT Library Preparation Kit, multiplexed with the Nextera XT Index Kit (both from Illumina, Inc.). Libraries were quantified by qPCR with Kapa (Kapa Biosystems, Wilmington, USA), and normalized to the recommended concentration. The library fragmentation pattern was evaluated by using High-Sensitivity DNA BioAnalyzer chips (Agilent Technologies, CA, USA) [samples from Brazil] or BIO-RAD Experion^{TM} Automated Electrophoresis System [samples from Cyprus]. Sequencing was performed with the MiSeq Reagent Kit (V2, 500 cycles, 2 × 250 bp). The paired-end sequencing data was evaluated using freely available and locally developed software, as described below.

2.3. Raw data processing (mapping)

Prior to mapping, all DNA segments produced by NGS, referenced to as 'reads', were trimmed on both ends for primer and adapter sequences. In addition, both read ends were trimmed for low quality sequences using seqtk trimfq, with parameter "q" set to 0.01. A major goal when performing HLA sequencing by NGS procedures is to get a reliable mapping of the produced reads, as previously discussed elsewhere (Lima et al., 2016; Castelli et al., 2015; Brandt et al., 2015). This is particularly challenging for HLA genes because of the high level of sequence similarity among most of the HLA genes and the high level of polymorphisms, which might bias the read mapping procedure (Lima et al., 2016; Castelli et al., 2015; Brandt et al., 2015). When HLA-related reads are directly mapped against a genome reference, a great number of misaligned reads is usually expected. This low mapping accuracy is observed when several genes are sequenced together (as for the Brazilian data, since other HLA loci were amplified together with *HLA-G*), or even when only a specific gene is sequenced (as for the Cypriot data). To circumvent this issue, we used *hla-mapper* (version 0.6, database version 001.6, function *regressive* using default parameters) to produce *HLA-G*-specific files (BAM format), with reads aligned to the reference genome version hg19. This software is available at www.castelli-lab.net/apps/hla-mapper and was previously introduced elsewhere (Lima et al., 2016; Castelli et al., 2015).

2.4. Genotype calling and processing

The Genome Analysis Toolkit (GATK, version 3.6) HaplotypeCaller algorithm was used to infer genotypes using the GVCF mode and a VCF (variant call format) file was generated concatenating all samples together with the GATK GenotypeGVCFs algorithm (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013). To infer genotypes using GATK, we used the chromosome 6 sequence (hg19) as reference. The original VCF file was then processed by *vcfx*, function checkpl (with the minimum genotype likelihood set to 99.9%), which guarantees that only high quality genotypes continue to a further imputation step. *vcfx* is available at www.castelli-lab.net/apps/vcfx. The use of vcfx introduced missing alleles in a proportion of about 1.21%, considering all samples together and the entire *HLA-G* segment (Table 1). These missing alleles were introduced mainly in the coding region, as discussed further.

We have detected bias in the template fragmentation pattern produced by Nextera (Illumina, Inc.), especially around two *HLA-G* segments, between 29,795,707 and 29,795,800 (intron 1) and between 29,796,229 and 29,796,319 (intron 2). Some variable sites detected here lay within these segments, including the ones at positions +99, +126, +130, +147, +613, +636, +644 and +685 (Table S1). Apparently, this bias is characteristic of CG-rich regions (Wang et al., 2011) and did inflate the number of missing alleles at the coding segment. Those segments presenting low coverage would be prone to genotyping errors, especially when a homozygous genotype is inferred at a position presenting low coverage. This issue was addressed by using *vcfx checkpl*, and also by the imputation steps that were performed as addressed further. Due to possible genotype errors in these two low-coverage segments, approaches such as this one should be considered when analyzing *HLA* genes by NGS and when using enzymatic fragmentation kits such as Nextera.

Phases of neighboring variable sites were determined by the GATK routine ReadBackedPhasing using minimum base quality set to 20 and phaseQualityThresh set to 2000. All the attempts to set lower phaseQualityThresh scores produced inaccurate double recombinants, in which both chromosomes of the same individual presented a recombination of common extended haplotypes

Table 2
HLA-G 5′ promoter haplotypes detected in Brazil and Cyprus.

	***HLA-G* Promoter Haplotypes** [a]																																		
HLA-G Promoter Haplotypes	-1406	-1377	-1305	-1179	-1155	-1140	-1138	-1121	-1098	-964	-922	-810	-762	-725	-716	-689	-666	-646	-633	-546	-541	-540	-509	-486	-483	-477	-443	-400	-399	-391	-369	-355	-297	-256	-201	-56
010101a	AGT	T	G	A	G	A	A	C	G	G	C	C	C	C	T	A	G	A	G	A	GA	A	C	A	A	C	G	G	G	G	C	G	G	TC	G	C
010101b	AGT	T	G	A	G	A	A	C	G	G	C	C	C	G	T	A	G	A	G	A	GA	A	C	A	A	C	G	G	G	G	C	G	G	TC	G	C
010101c	AGT	T	G	A	G	A	A	T	G	G	C	C	C	G	T	A	G	A	G	A	GA	A	C	A	A	C	G	G	G	G	C	G	G	TC	G	C
010101d	AGT	T	G	A	G	A	A	C	G	G	C	C	C	C	T	A	G	A	G	A	GA	A	C	A	G	C	G	G	G	G	C	G	G	TC	G	C
010101f	AGT	T	G	A	G	A	A	C	G	G	C	C	C	C	T	A	G	A	G	A	G	*	C	A	A	0	G	G	G	G	C	G	G	TC	G	C
010101g	AGT	T	G	A	G	A	A	C	G	G	C	C	C	C	T	A	G	A	G	A	GA	A	C	A	A	C	G	G	G	G	C	G	A	TC	G	C
010101h	AGT	G	G	A	G	A	A	C	G	G	C	C	C	C	T	A	G	A	G	A	GA	A	C	A	A	C	G	G	G	G	C	G	G	TC	G	C
010101i	A	T	G	A	G	A	A	C	G	G	C	C	C	C	T	A	G	A	G	A	GA	A	C	A	A	C	G	G	G	G	C	G	G	TC	G	C
010101j	AGT	T	G	A	G	A	A	C	G	G	C	C	C	C	T	A	G	A	G	A	GA	A	C	A	A	C	G	G	A	G	C	G	G	TC	G	C
010102a	AGT	T	A	G	G	T	A	C	G	A	C	C	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	G	G	G	G	A	G	G	TC	A	C
010102b	AGT	T	A	G	G	T	A	C	G	A	C	T	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	G	G	G	G	A	G	G	TC	A	C
010102c	AGT	T	A	G	G	T	A	C	G	A	A	C	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	G	G	G	G	A	G	G	TC	A	C
010102d	AGT	T	A	G	G	T	A	C	A	A	C	C	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	G	G	G	G	A	G	G	TC	A	C
010102e	AGT	T	G	G	G	T	A	C	G	A	C	C	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	G	G	G	G	A	G	G	TC	A	C
0103a	AGT	T	G	G	G	A	G	C	G	G	C	C	C	T	T	A	G	A	G	AG	GA	A	G	A	A	G	G	A	G	A	A	G	G	TC	G	T
0103c	AGT	T	G	G	G	A	G	C	G	G	C	C	C	T	T	A	G	G	G	A	GA	A	G	A	A	G	G	G	G	G	A	G	G	TC	G	T
0103d	AGT	T	G	G	G	A	G	C	G	G	C	C	C	T	T	A	G	A	G	AG	GA	A	C	A	A	G	G	A	G	A	A	G	G	TC	G	T
0103e	AGT	T	G	G	G	A	G	C	G	G	C	C	C	T	T	A	G	G	G	A	GA	A	G	A	A	G	G	A	G	A	A	G	G	TC	G	T
0103f	AGT	T	G	G	G	A	G	C	G	G	C	C	C	T	T	A	G	G	G	A	GA	A	G	A	A	G	G	A	G	A	A	A	G	TC	G	T
0103g	AGT	T	G	G	G	A	G	C	G	G	C	C	C	T	T	A	G	A	G	AG	GA	G	C	A	A	G	G	A	G	A	A	G	G	TC	G	T
0104a	AGT	T	A	G	A	A	A	C	G	A	C	C	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	G	G	G	G	A	G	G	TC	A	C
0104b	AGT	T	A	G	A	A	A	C	G	A	C	C	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	A	G	G	G	A	G	G	TC	A	C
0104c	AGT	T	A	G	A	A	A	C	G	A	C	C	T	C	G	G	T	A	A	A	GA	A	C	C	A	G	G	G	G	G	A	G	G	T	A	C

[a] The positions given are based on the IMGT/HLA database, considering the Adenine of the first translated ATG as +1. The list of genomic positions is depicted at Table S1. Haplotypes were named according to previous studies (Nilsson et al., 2016; Gineau et al., 2015; Tan et al., 2005; Castelli et al., 2011; Castelli et al., 2014b). New haplotypes were named taking into account the relatedness with known haplotypes. The alternative alleles regarding the human genome draft version hg19 are marked in shades of gray.

and the break point was between distant variable sites, which is highly unlikely. This assures that only alleles from closely located variable sites, which are within a same read, are directly phased. Considering that variable sites may be quite distant from each other in a same sample and that ReadBackedPhasing does not perform phase inference on indels and multi-allelic loci, not all variable sites were straightforwardly phased. In the present series, and considering all samples together, 78.65% of the heterozygous sites were directly phased by using GATK (Table 1). The remaining 21.35% were inferred by using the PHASE algorithm (Stephens et al., 2001; Stephens and Donnelly, 2003 Stephens and Donnelly, 2003), which also imputed the 1.21% of missing alleles observed after the *vcfx* treatment. It should be mentioned that the *HLA-G* locus present several indels and at least one tri-allelic site, which are not supported by the ReadBackedPhasing algorithm and thus were not straightforwardly phased by GATK.

To proceed with the PHASE algorithm analysis, the partially GATK-phased VCF file (obtained with the ReadBackedPhasing algorithm) was converted into an input file for PHASE and into an accessory file containing the known phases between variable sites. For many individuals, blocks of variable sites with known phases were generated, but with unknown phase between blocks. The PHASE algorithm was used to infer haplotypes for each sample, performing several different runs. For each run, a single block of variable sites with known phases was fixed. In addition, a final run without considering the GATK information was also performed. We accepted the samples in which all runs revealed the same pair of haplotypes as most probable, as long as the pair of haplotypes was in agreement with the known phases detected by GATK (with no exceptions). The most probable haplotype pair of each sample was then used to build a phased VCF file. It should be highlighted that all singletons were removed to perform the haplotyping step using the PHASE algorithm. Singletons were properly included in the phased VCF file manually, but only when they met the following criteria: (a) no samples presented missing allele at this particular position, and (b) there was a clear relationship between the singleton and the next heterozygous site visually defined using the BAM file of that sample or it was the only heterozygous site detected in that particular sample.

This phased VCF file was converted into complete *HLA-G* sequences and CDS sequences using the hg19 reference sequence as a draft and replacing the correct nucleotide in each position, two sequences per samples, by using application *vcfx* function *fasta* (www.castelli-lab.net/apps/vcfx). By using a local BLAST server with databases containing all known HLA alleles described so far, either for genomic sequences and CDS sequences, downloaded from the IPD-IMGT/HLA database (Robinson et al., 2015), version 3.26.0, the closest known *HLA-G* coding allele was defined for each haplotype. For the promoter and 3′UTR segments, haplotypes were named according to a previously described nomenclature (Nilsson et al., 2016; Castelli et al., 2014a; Sabbagh et al., 2014; Gineau et al., 2015; Tan et al., 2005; Castelli et al., 2010; Castelli et al., 2011; Castelli et al., 2014b).

2.5. *Other analyses*

The frequency of each *HLA-G* haplotype was computed by the direct counting method. Nucleotide diversity, gene diversity and Tajima's *D* were calculated using Arlequin 3.5 (Excoffier and Lischer, 2010). The Tajima's D p-value was computed by the comparison of the estimated statistic to a distribution of estimates computed for 10,000 random samples of the same sample size and level of polymorphism as the observed data, and represents the probability of obtaining a simulated Tajima's D larger than the observed. The network was created using PopArt (http://popart.otago.ac.nz) and the Median-Joining algorithm. The synonymous and nonsynonymous nucleotide substitution test, which evaluates the relative abundance of synonymous substitutions and nonsynonymous substitutions that occurred in the gene sequences, was carried out using the Nei-Gojobori method (Nei and Gojobori, 1986) implemented in the MEGA 7 program (Kumar et al., 2016).

3. Results

The strategy proposed in the previous section revealed the presence of 120 variable sites at the evaluated *HLA-G* segment considering all samples together. Table S1 lists all these 120 variable sites, together with their chromosome positions, SNPid, IMGT/HLA relative position and frequencies for the reference allele. Altogether, 88 out of the 120 variable sites (73.3%) did present a general frequency higher than 1% for the minor allele (Table S1). The nucleotide diversity for the entire *HLA-G* segment was slightly higher in Brazil than in Cyprus (Table 1). Twelve variable sites

Table 3
Frequency of each *HLA-G* 5′ promoter haplotype detected in Brazil and Cyprus.

***HLA-G* promoter Haplotype** [a]	**Cyprus (2*n*=370)**	**Brazil** [b] **(2*n*=630)**	**All samples (2*n*=1000)**
PROMO-010101a	0.2297	0.2587	0.2480
PROMO-010101b	0.0568	0.0508	0.0530
PROMO-010101c	0.0324	0.0476	0.0420
PROMO-010101d	0.0027	0.0254	0.0170
PROMO-010101f	0.0486	0.0476	0.0480
PROMO-010101g	0.0135	0.0016	0.0060
PROMO-010101h	0.0027	0.0032	0.0030
PROMO-010101i	0.0054	0.0079	0.0070
PROMO-010101j	0.0054	0.0032	0.0040
PROMO-010102a	0.3676	0.2952	0.3220
PROMO-010102b	0.0027	0.0016	0.0020
PROMO-010102c	0.0027	0.0032	0.0030
PROMO-010102d	-	0.0111	0.0070
PROMO-010102e	0.0108	0.0048	0.0070
PROMO-0103a	0.0027	0.0397	0.0260
PROMO-0103c	-	0.0032	0.0020
PROMO-0103d	0.0432	0.0159	0.0260
PROMO-0103e	0.0162	0.0254	0.0220
PROMO-0103f	-	0.0016	0.0010
PROMO-0103g	-	0.0016	0.0010
PROMO-0104a	0.1568	0.1413	0.1470
PROMO-0104b	-	0.0079	0.0050
PROMO-0104c	-	0.0016	0.0010

[a] Information regarding the sequence of each haplotype is presented at Table 2. Haplotypes were named according to previous studies (Nilsson et al., 2016; Gineau et al., 2015; Tan et al., 2005; Castelli et al., 2011; Castelli et al., 2014b). New haplotypes were named taking into account the relatedness with known haplotypes. [b] Brazilians from the State of São Paulo, Southeast Brazil.

detected here are not currently described at the IPD-IMGT/HLA database (Table S1).

Considering all samples, the variable sites were arranged into 58 different extended haplotypes. In order to characterize these extended haplotypes, the variability and haplotypes of each *HLA-G* segment will be presented separately, starting from the variable sites detected upstream of the first translated ATG and considered as the promoter segment, from position −1406 to −1 (Tables 2 and 3); the variability and haplotypes laying in the segment that is tracked by the IMGT/HLA database version 3.26.0, i.e., from −300 to +2838 (Table 4); the variability and haplotypes at the last *HLA-G* exon, which constitute most of the *HLA-G* 3′ untranslated region (Table 5); and finally, the extended set of haplotypes (Table 6).

The segment upstream the first translated ATG, from nucleotide −1406 to −1, presented 37 variable sites (Table 2). These 37 variable sites were arranged into 23 different promoter haplotypes, named accordingly to previous studies as described in the methods section. Some of these haplotypes have never been detected before, thus, they were named following the pattern previously established (Table 2). Most of the haplotypes detected here were identical to the ones already described in different samples (Nilsson et al., 2016; Castelli et al., 2014a; Gineau et al., 2015; Tan et al., 2005; Castelli et al., 2011; Santos et al., 2013; Castelli et al., 2014b). In addition, some of new ones are actually derived from these well-documented haplotypes, but carrying additional variable sites (e.g., positions −1406, −1377, −399 and −297).

Although most of the haplotypes were detected in both populations, some haplotypes, such as 0104b and 0104c, were detected only in Brazil (Table 3). In addition, some closely related haplotypes do present very different frequencies between both populations, i.e., the promoter lineage 0103 in Cyprus is mostly represented by the haplotypes 0103d, while in Brazil this lineage is mostly represented by haplotype 0103a and 0103e. The two most frequent haplotypes (and also the most divergent ones) were 010101a and 010102b, with a summed frequency of 55.39% and 59.73% in Brazil and Cyprus, respectively.

We detected 38 different coding sequences (coding haplotypes) considering both Brazil and Cyprus (Table 4). Of those, at least 20 are identical to known *HLA-G* alleles already described at the IPD-IMGT/HLA database version 3.26.0. This includes 12 coding alleles with no divergence from a known allele and 8 marked as compatibles, since the sequence detected here is identical to partial sequences already described by the IMGT/HLA database. It should be noted that the two versions of the G*01:11 alleles we have found do present compatibility with the partial sequence described by the IMGT/HLA database, but they present an intron mutation splitting them into two different alleles. The 18 remaining coding sequences may be considered new *HLA-G* coding alleles. These new *HLA-G* alleles carry either one or more of the new variable sites (not already described in the IPD-IMGT/HLA database) detected here, or different variable sites at already well-documented positions. Many of these new variable sites and new coding alleles were detected either in Brazil or Cyprus, and at least 4 presented a global frequency higher than 1% (Table 4).

Although the coding alleles detected here would inflate the IPD-IMGT/HLA database version 3.26.0 in 33.96%, and in spite of the presence of several new variable sites, only five variations represent non-synonymous substitutions, including position +292 (rs41551813, allele G*01:03), position +293 (rs72558173, allele G*01:11), position +755 (rs12722477, allele G*01:04), position +813 (rs41557518, allele G*01:05N) and position +1799 (rs12722482, allele G*01:06), all described at Table S1. All other variable sites detected in the coding region are synonymous substitutions or mutations in introns.

The segment encompassing the last *HLA-G* exon, which encoded most of the *HLA-G* 3′ untranslated region, presented 16 different haplotypes by the combination of 14 variable sites (Table 5). Fif-

Table 4
HLA-G coding alleles detected in Brazil and Cyprus.

HLA-G Coding Haplotype [a]	Cyprus (2*n* = 370)	Brazil [b] (2*n* = 630)	All samples (2*n* = 1000)
G*01:01:01:01	0.2243	0.2508	0.2410
G*01:01:01:01 $^{(+1147T,+2412A)}$	–	0.0016	0.0010
G*01:01:01:01 $^{(+1328C)}$	–	0.0016	0.0010
G*01:01:01:01 $^{(+2412A)}$	0.0027	0.0175	0.0120
G*01:01:01:01 $^{(+2756C)}$	0.0135	0.0079	0.0100
G*01:01:01:01 $^{(+755A)}$	0.0027	0.0016	0.0020
G*01:01:01:01 $^{(-297A)}$	0.0135	0.0016	0.0060
G*01:01:01:04	0.0162	0.0444	0.0340
G*01:01:01:04 $^{(+1078T)}$	0.0324	0.0032	0.0140
G*01:01:01:05	0.0892	0.0921	0.0910
G*01:01:01:05 $^{(+99G,+1147C,+2412A)}$	–	0.0063	0.0040
G*01:01:01:06	0.0027	0.0079	0.0060
G*01:01:02:01	0.1486	0.1619	0.1570
G*01:01:02:01 $^{(+1237C,+2018C,+2798G)}$	0.0027	–	0.0010
G*01:01:02:01 $^{(+2798G)}$	–	0.0016	0.0010
G*01:01:02:01 $^{(+482T,+494A)}$	0.0027	–	0.0010
G*01:01:02:02	–	0.0079	0.0050
G*01:01:03:03	0.0784	0.0571	0.0650
G*01:01:03:03 $^{(+2374C)}$	–	0.0016	0.0010
G*01:01:09 $^{(\text{no deletion at }+615)}$	–	0.0032	0.0020
G*01:01:11$^{\text{compatible}}$	–	0.0016	0.0010
G*01:01:12 $^{(+324G)}$	0.0459	0.0032	0.0190
G*01:01:14 $^{\text{compatible}}$	0.0054	0.0063	0.0060
G*01:01:15 $^{\text{compatible}}$	–	0.0048	0.0030
G*01:01:17 $^{\text{compatible}}$	–	0.0016	0.0010
G*01:01:19 $^{\text{compatible}}$	–	0.0016	0.0010
G*01:03:01:02	0.0595	0.0873	0.0770
G*01:03:01:02 $^{(+769G)}$	0.0027	–	0.0010
G*01:04:01	0.1378	0.0825	0.1030
G*01:04:01 $^{(+2190T)}$	0.0027	–	0.0010
G*01:04:01 $^{(+498A,+755C,+1104G,+2412A)}$	0.0108	0.0048	0.0070
G*01:04:01 $^{(+531G)}$	0.0027	–	0.0010
G*01:04:04	0.0135	0.0635	0.0450
G*01:04:05 $^{\text{compatible}}$	–	0.0016	0.0010
G*01:05N	0.0216	0.0222	0.0220
G*01:06	0.0676	0.0460	0.0540
G*01:11 $^{\text{compatible-1}}$	–	0.0016	0.0010
G*01:11 $^{\text{compatible-2}}$	–	0.0016	0.0010

[a] Coding haplotypes were named with the closest known IPD-IMGT/HLA allele followed by any divergences eventually detected. The notation "compatible" is used when the sequence of this allele is not complete at the IPD-IMGT/HLA database. For sequences that are not identical to one already described (IMGT/HLA database version 3.26.0), the closest known allele is informed followed by the divergences eventually observed.

[b] Brazilians from the State of São Paulo, Southeast Brazil.

teen of those haplotypes were previously detected and reported (Nilsson et al., 2016; Castelli et al., 2014a; Sabbagh et al., 2014; Castelli et al., 2010; Lucena-Silva et al., 2012; Santos et al., 2013; Martelli-Palomino et al., 2013; Castelli et al., 2014b), and one (named UTR-46) characterizes a new haplotype. Haplotypes UTR-1, UTR-2, UTR-3, UTR-4, UTR-5, UTR-6, UTR-7, UTR-13, UTR-18 and UTR-44 were detected in both samples.

Table 6 presents the extended haplotypes that were detected at least twice in the present evaluation. In general, one promoter lineage is usually associated with a unique coding allele (and their derivatives) and a specific 3′UTR haplotype. For example, alleles G*01:01:01:01 and derivatives [such as G*01:01:01:01 $^{(+2756C)}$] are associated with promoters of the lineage 010101, mainly 010101a, and the 3′UTR haplotype known as UTR-1, and all copies of the G*01:01:01:04 allele are associated with the promoter 010101f and the 3′UTR haplotypes UTR-6 or UTR-18. Although most of these extended haplotypes were detected in both samples, some were detected in high frequency in just one sample (e.g., 010101 g/G*01:01:01:01 $^{(-297A)}$/UTR-1 in Cyprus) and some extended haplotypes detected in both samples presented very different frequencies, e.g., 010101f/G*01:01:01:04/UTR-18 and 0103a/G*01:03:01:02/UTR-5 are quite frequent in Brazil but not in Cyprus, while the opposite is observed for 010102a/G*01:01:12 $^{(+324G)}$/UTR-02 and others.

The complete sequences of each individual (all polled together), representing each extended *HLA-G* haplotype were used to infer a network to evaluate the relationship among these haplotypes (Fig. 1). These sequences are available upon request. This network reveals the presence of at least three great *HLA-G* haplotype groups. The first group is composed by promoters 0103, coding alleles G*01:03 and 3′UTR haplotypes UTR-5 and UTR-17; the second is composed by promoters 010101, coding alleles mainly from the G*01:01:01 group and 3′UTR haplotypes UTR-1, −4, −6 and −27; and the third is composed by three distinct subgroups. These subgroups are (a) promoters 0104, coding alleles from the G*01:04 group and 3′UTR haplotypes UTR-3, −13 and −46, (b) promoters 010102, coding alleles G*01:01:03 and UTR-7, and (c) promoters 010102, coding alleles mainly from the G*01:01:02 group (including derivatives such as G*01:05N and G*01:06) and UTR-2 (Fig. 1).

Some of the new *HLA-G* coding alleles were detected associated with a specific extended haplotype and with a frequency higher than 1% in at least one population sample, such as the extended haplotypes 010101a/G*01:01:01:01 $^{(2756C)}$/UTR-1 and 010101d/G*01:01:01:01 $^{(2412A)}$/UTR-1.

4. Discussion

Next generation sequencing (NGS) or massively parallel sequencing procedures are useful to characterize DNA variability and haplotypes, but its use is challenging when *HLA* genes are the main focus, as discusses elsewhere (Lima et al., 2016). It is particularly difficult to get reliable read mappings when more than one

Table 5
HLA-G 3′UTR haplotypes detected in Brazil and Cyprus.

HLA-G 3'UTR Haplotype [a]	14bp	+3001	+3003	+3010	+3027	+3032	+3035	+3044	+3092	+3121	+3142	+3187	+3196	+3227	Cyprus (2*n*=370)	Brazil [b] (2*n*=630)	All samples (2*n*=1000)
UTR-01	G	C	T	G	C	G	C	A	G	T	C	G	C	G	0.2486	0.2857	0.2720
UTR-02	GATTTGTTCATGCCT	C	T	C	C	G	C	A	G	T	G	A	G	G	0.2892	0.2524	0.2660
UTR-03	G	C	T	C	C	G	C	A	G	T	G	A	C	G	0.1405	0.1476	0.1450
UTR-04	G	C	C	G	C	G	C	A	G	T	C	A	C	G	0.0892	0.1048	0.0990
UTR-05	GATTTGTTCATGCCT	C	T	C	C	G	T	A	G	T	G	A	C	G	0.0622	0.0825	0.0750
UTR-06	G	C	T	G	C	G	C	A	G	T	C	A	C	G	0.0405	0.0175	0.0260
UTR-07	GATTTGTTCATGCCT	C	T	C	A	G	T	A	G	T	G	A	C	G	0.0784	0.0587	0.0660
UTR-08	GATTTGTTCATGCCT	C	T	G	C	G	C	A	G	T	G	A	G	G	0.0027	-	0.0010
UTR-10	G	C	T	C	C	G	C	A	G	T	G	A	G	G	0.0027	-	0.0010
UTR-13	G	C	T	C	C	G	T	A	G	T	G	A	C	G	0.0135	0.0032	0.0070
UTR-17	GATTTGTTCATGCCT	T	T	C	C	G	T	A	G	T	G	A	C	G	-	0.0048	0.0030
UTR-18	G	C	T	G	C	G	C	A	G	T	C	A	C	A	0.0162	0.0317	0.0260
UTR-20	G	C	T	G	C	C	C	A	G	T	C	A	C	G	-	0.0063	0.0040
UTR-27	G	C	C	G	C	G	C	A	T	T	C	A	C	G	0.0027	-	0.0010
UTR-44	GATTTGTTCATGCCT	C	T	C	C	G	T	A	G	C	G	A	C	G	0.0108	0.0048	0.0070
UTR-46	G	C	T	C	C	G	C	T	G	T	G	A	C	G	0.0027	-	0.0010

[a]The list of genomic positions is depicted at Table S1. Haplotypes were named according to previous studies (Nilsson et al., 2016; Sabbagh et al., 2014; Gineau et al., 2015; Castelli et al., 2010; Castelli et al., 2011; Lucena-Silva et al., 2012; de Albuquerque et al., 2016; Santos et al., 2013; Castelli et al., 2014b). New haplotypes were named after the known haplotypes. [b]Brazilians from the State of São Paulo, Southeast Brazil.

HLA gene is processed together, mainly because of their polymorphic nature, and an unreliable read mapping would bias genotype inference. Although one may expect that this is not the case for *HLA-G* (and other non-classical genes), since they are much less variable than their classical counterparts, this same issue arises when *HLA-G* sequences are mapped against the reference genome or more than one HLA gene is sequenced together.

As previously introduced elsewhere (Lima et al., 2016; Castelli et al., 2015), to circumvent the issues, the raw sequencing data from both Brazil and Cyprus were processed using ***hla-mapper***, which applies a series of filters to address each sequence (read) to its proper gene (Lima et al., 2016; Castelli et al., 2015). Then, we used HaplotypeCaller from the GATK package (McKenna et al., 2010; Van der Auwera et al., 2013) to infer genotypes, coupled with *vcfx checkpl* to introduce missing alleles on uncertain genotypes. This strategy, implemented here for the first time, improved genotype inference and better downstream imputation procedures.

The strategy used here to infer haplotypes took advantage of the straightforward phase observed at many NGS reads, since 78.65% of all heterozygous sites were directly phased by the GATK ReadBackedPhasing algorithm (Table 1). The heterozygoses that were not straightforwardly phased were dealt with the PHASE method (Stephens et al., 2001), which is suitable for HLA genotypes because it can manage multi-allelic loci and impute missing alleles. Both strategies (GATK + PHASE) were combined in order to produce reliable haplotypes, in which both fragment-inferred and probabilistic-inferred haplotypes were in agreement. Since most of the previous studies regarding *HLA-G* promoter, coding and 3′UTR variability did use only probabilistic models to infer haplotypes, we anticipated that the present data would confirm whether the previously defined haplotypes were correct or not.

In fact, the promoter haplotypes detected here, in which most of the variable sites were straightforwardly phased, are generally identical to the ones already described using probabilistic models or cloning (Nilsson et al., 2016; Castelli et al., 2014a; Gineau et al., 2015; Tan et al., 2005; Castelli et al., 2011; Castelli et al., 2014b), with new haplotypes mainly representing previous known haplotypes carrying additional mutations. This demonstrates that (a) the haplotypes already characterized for the *HLA-G* promoter are indeed correct, (b) they represent the most frequent haplotypes found worldwide and (c) the NGS strategy proposed here is suitable to characterize such promoter variability.

Most of the promoter haplotypes were detected both in Brazil and Cyprus but some of them presented very different frequencies between both populations. Haplotype 0103a, for instance, was quite frequent in this Brazilian sample, but rare in Cyprus, while 0103d is three times more frequent in Cyprus then in Brazil (Table 3).

It should be mentioned that the *HLA-G* structure considered by the IPD-IMGT/HLA database does not match the transcripts described in either the hg19 and hg38 human genome draft. In this matter, part of the promoter segment considered here is identified at hg19 as two exons (and an intron between) that encode the *HLA-G* 5′ untranslated region [refer to (Castelli et al., 2014b) for details regarding this issue]. Thus, considering the hg19 or hg38 transcripts, only the segment upstream position −866 would be consider as the actual promoter segment. Even thought, the promoter haplotypes would still be represented by four distinct lineages or haplotypes groups, i.e., the promoter groups 010101, 010102, 0103 and 0104. It is not clear yet whether the variable sites upstream the first translated ATG do influence *HLA-G* expression or not, but evidence of balancing selection increasing the heterozygosis of this segment was detected in many population samples (Gineau et al., 2015; Tan et al., 2005; Castelli et al., 2011; Santos et al., 2013), including the two samples studied here (Tajima's *D*

Table 6
HLA-G extended haplotypes detected in Brazil and Cyprus. Only haplotypes that were found at least twice are represented above.

Promoter haplotype [a]	Coding allele or coding haplotype [b]	3′UTR haplotype [c]	Cyprus ($2n = 370$)	Brazil [d] ($2n = 630$)	All samples ($2n = 1000$)
PROMO-010101a	G*01:01:01:01	UTR-01	0.2054	0.2365	0.2250
PROMO-010101a	G*01:01:01:01	UTR-06	0.0081	0.0032	0.0050
PROMO-010101a	G*01:01:01:01 $^{(+2756C)}$	UTR-01	0.0135	0.0079	0.0100
PROMO-010101a	G*01:01:01:01 $^{(+755A)}$	UTR-01	0.0027	0.0016	0.0020
PROMO-010101a	G*01:01:01:06	UTR-04	–	0.0079	0.0050
PROMO-010101b	G*01:01:01:05	UTR-04	0.0568	0.0460	0.0500
PROMO-010101b	G*01:01:15 $^{\text{compatible}}$	UTR-06	–	0.0048	0.0030
PROMO-010101c	G*01:01:01:05	UTR-04	0.0324	0.0460	0.0410
PROMO-010101d	G*01:01:01:01 $^{(+2412A)}$	UTR-01	0.0027	0.0175	0.0120
PROMO-010101d	G*01:01:01:05 $^{(+99G,+1147C,+2412A)}$	UTR-01	–	0.0063	0.0040
PROMO-010101f	G*01:01:01:04	UTR-06	–	0.0063	0.0040
PROMO-010101f	G*01:01:01:04	UTR-18	0.0162	0.0317	0.0260
PROMO-010101f	G*01:01:01:04	UTR-20	–	0.0063	0.0040
PROMO-010101f	G*01:01:01:04 $^{(+1078T)}$	UTR-06	0.0324	0.0032	0.0140
PROMO-010101g	G*01:01:01:01 $^{(-297A)}$	UTR-01	0.0135	0.0016	0.0060
PROMO-010101h	G*01:01:09 $^{(\text{nodel}+615)}$	UTR-04	–	0.0032	0.0020
PROMO-010101i	G*01:01:01:01	UTR-01	0.0054	0.0079	0.0070
PROMO-010101j	G*01:01:01:01	UTR-01	0.0054	0.0032	0.0040
PROMO-010102a	G*01:01:02:01	UTR-02	0.1405	0.1460	0.1440
PROMO-010102a	G*01:01:02:02	UTR-02	–	0.0079	0.0050
PROMO-010102a	G*01:01:03:03	UTR-07	0.0784	0.0571	0.0650
PROMO-010102a	G*01:01:12 $^{(+324G)}$	UTR-02	0.0459	0.0032	0.0190
PROMO-010102a	G*01:01:14 $^{\text{compatible}}$	UTR-02	0.0054	0.0063	0.0060
PROMO-010102a	G*01:05N	UTR-02	0.0216	0.0222	0.0220
PROMO-010102a	G*01:06	UTR-02	0.0676	0.0460	0.0540
PROMO-010102b	G*01:01:02:01	UTR-02	0.0027	0.0016	0.0020
PROMO-010102c	G*01:01:02:01	UTR-02	0.0027	0.0032	0.0030
PROMO-010102d	G*01:01:02:01	UTR-02	–	0.0111	0.0070
PROMO-010102e	G*01:04:01 $^{(+498A,+755C,+1104G,+2412A)}$	UTR-44	0.0108	0.0048	0.0070
PROMO-0103a	G*01:03:01:02	UTR-05	0.0027	0.0397	0.0260
PROMO-0103c	G*01:03:01:02	UTR-05	–	0.0032	0.0020
PROMO-0103d	G*01:03:01:02	UTR-05	0.0432	0.0159	0.0260
PROMO-0103e	G*01:03:01:02	UTR-05	0.0135	0.0206	0.0180
PROMO-0103e	G*01:03:01:02	UTR-17	–	0.0048	0.0030
PROMO-0104a	G*01:04:01	UTR-03	0.1216	0.0714	0.0900
PROMO-0104a	G*01:04:01	UTR-13	0.0135	0.0032	0.0070
PROMO-0104a	G*01:04:04	UTR-03	0.0135	0.0635	0.0450
PROMO-0104b	G*01:04:01	UTR-03	–	0.0079	0.0050
Others	–	–	0.0216	0.0190	0.0200

[a] The complete list of *HLA-G* promoter haplotypes is presented at Table 2.
[b] The complete list of *HLA-G* coding alleles is presented at Table 4.
[c] The complete list of *HLA-G* 3′UTR haplotypes is presented at Table 5.
[d] Brazilians from the State of São Paulo, Southeast Brazil.

value greater than 1.8 for the promoter segment for both samples, Table 1).

The *HLA-G* coding segment presented many different haplotypes, all defined according to the closest known *HLA-G* allele as described by the IPD-IMGT/HLA database version 3.26.0 (Robinson et al., 2015), followed by any divergences that have occurred (Table 4). Although several new *HLA-G* coding alleles were detected, the summed frequency of all coding sequences that were identical to any one already recognized by the IPD-IMGT/HLA database was 91.32%. The remaining alleles are new coding sequences that carry either new mutations or alleles with different combinations of well-documented mutations, or both. Six of these new alleles were frequent (MAF > 0.01) in at least one population sample. This demonstrates that the NGS strategy proposed here is suitable to detect the coding variability and haplotypes, and also that the *HLA-G* coding variability may be higher than our current awareness.

It can be observed at Table 4 that all coding alleles presenting a general frequency higher than 1% in at least one population were detected both in Brazil and Cyprus, even the new coding alleles. Nonetheless, their frequency varies between both samples. Besides the presence of several new *HLA-G* coding alleles, none of these sequences would encode a different HLA-G molecule because the variable sites associated with them are either intronic mutations or synonymous substitutions. The mechanisms underlying this lack of protein variability detected for *HLA-G* are not well understood, mainly because *HLA-G* is located within the most variable genes of the human genome. This phenomenon is also observed for other non-classical *HLA* genes such as *HLA-E* and *HLA-F*, which present tolerogenic and immune modulatory properties (Rizzo et al., 2008; Donadi et al., 2011; Nilsson et al., 2016; Mendes-Junior et al., 2013; Castelli et al., 2011; Zambra et al., 2016; Hviid et al., 2006; Lima et al., 2016; Castelli et al., 2015; Castelli et al., 2014b; Veiga-Castelli et al., 2016; Dias et al., 2015; Felicio et al., 2014; Veiga-Castelli et al., 2012; Castelli et al., 2007; Hviid et al., 1997; Hviid et al., 1999; Hviid et al., 2001; Hviid et al., 2003; Carosella, 2011; Carosella and LeMaoult, 2011; Arnaiz-Villena et al., 2007a; Arnaiz-Villena et al., 2007b; Alegre et al., 2007; Moscoso et al., 2006; Carvalho dos Santos et al., 2013; Liu et al., 2012). Considering the 500 samples studied here, only six different protein molecules would be encoded by the *HLA-G* sequences observed: G*01:01 (69.3%), G*01:03 (7.8%), G*01:04 (15.10%), G*01:05N (2.2%), G*01:06 (5.4%) and G*01:11 (0.2%).

Interestingly, evidence of balancing selection were also detected for the coding segment in both samples studied here (Tajima's *D* higher than 2.65 for both), but considering the low HLA-G protein variability, it is possible that this is a hitchhiking effect because of the Linkage Disequilibrium (LD) between the promoter and the coding segments. This became evident by evaluating the extended

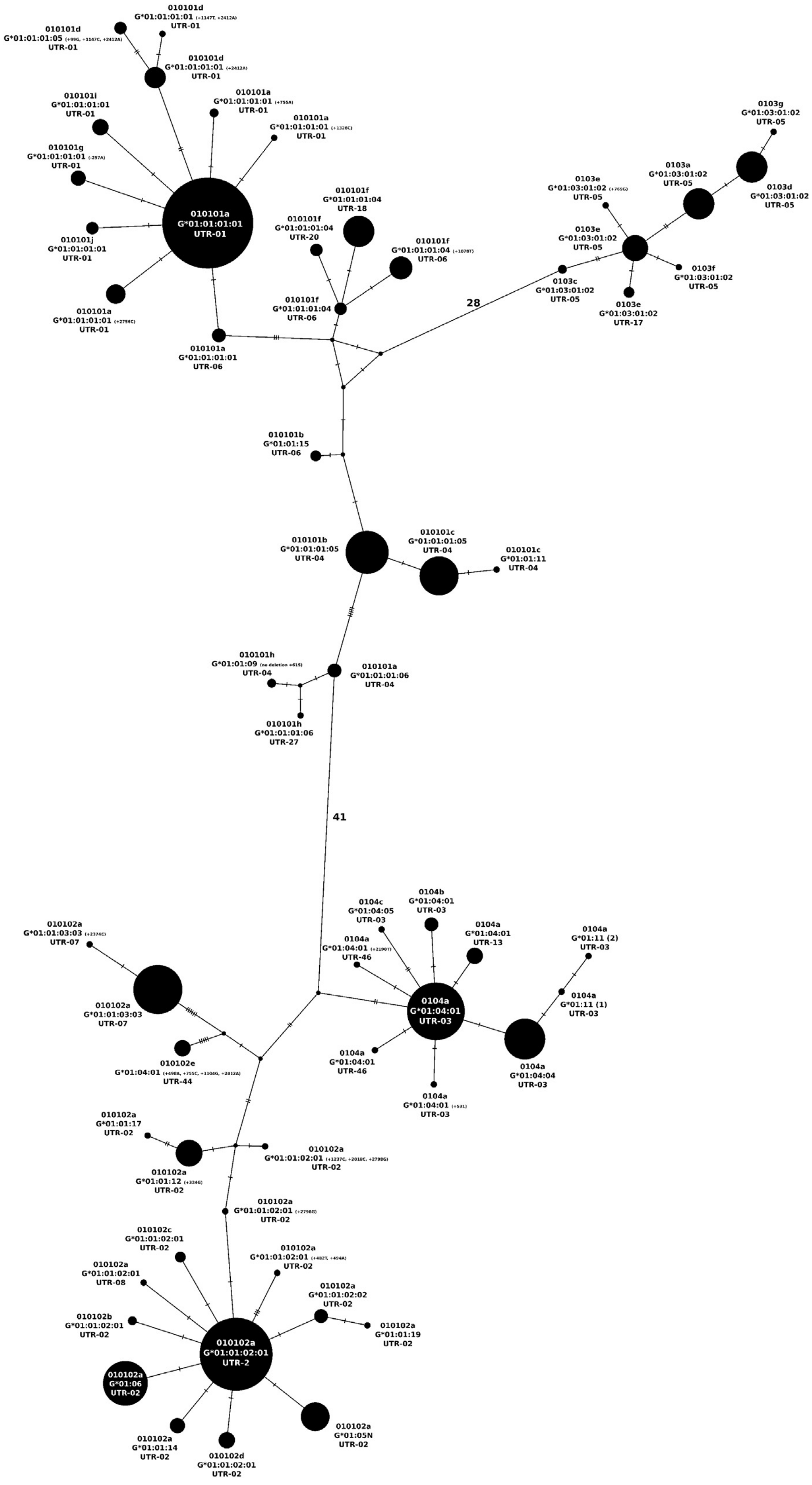
010101d G*01:01:01:01 (+1147T, +2412A) UTR-01
010101d G*01:01:01:05 (+99G, +1147C, +2412A) UTR-01
010101d G*01:01:01:01 (+2412A) UTR-01
010101i G*01:01:01:01 UTR-01
010101a G*01:01:01:01 (+755A) UTR-01
010101a G*01:01:01:01 (+1328C) UTR-01
010101g G*01:01:01:01 (-297A) UTR-01
010101a G*01:01:01:01 UTR-01
010101j G*01:01:01:01 UTR-01
010101a G*01:01:01:01 (+2756C) UTR-01
010101a G*01:01:01:01 UTR-06
010101f G*01:01:01:04 UTR-18
010101f G*01:01:01:04 UTR-20
010101f G*01:01:01:04 (+1078T) UTR-06
010101f G*01:01:01:04 UTR-06
28
0103g G*01:03:01:02 UTR-05
0103a G*01:03:01:02 UTR-05
0103d G*01:03:01:02 UTR-05
0103e G*01:03:01:02 (+769G) UTR-05
0103e G*01:03:01:02 UTR-05
0103c G*01:03:01:02 UTR-05
0103f G*01:03:01:02 UTR-05
0103e G*01:03:01:02 UTR-17
010101b G*01:01:15 UTR-06
010101b G*01:01:01:05 UTR-04
010101c G*01:01:01:05 UTR-04
010101c G*01:01:11 UTR-04
010101h G*01:01:09 (no deletion +615) UTR-04
010101a G*01:01:01:06 UTR-04
010101h G*01:01:01:06 UTR-27
41
010102a G*01:01:03:03 (+2374C) UTR-07
010102a G*01:01:03:03 UTR-07
010102e G*01:04:01 (+490A, +755C, +1104G, +2412A) UTR-44
0104c G*01:04:05 UTR-03
0104b G*01:04:01 UTR-03
0104a G*01:04:01 (+2190T) UTR-46
0104a G*01:04:01 UTR-13
0104a G*01:11 (2) UTR-03
0104a G*01:04:01 UTR-03
0104a G*01:11 (1) UTR-03
0104a G*01:04:01 UTR-46
0104a G*01:04:04 UTR-03
0104a G*01:04:01 (+531) UTR-03
010102a G*01:01:17 UTR-02
010102a G*01:01:02:01 (+1237C, +2018C, +2798G) UTR-02
010102a G*01:01:12 (+324G) UTR-02
010102a G*01:01:02:01 (+2798G) UTR-02
010102c G*01:01:02:01 UTR-02
010102a G*01:01:02:01 (+482T, +494A) UTR-02
010102a G*01:01:02:01 UTR-08
010102a G*01:01:02:02 UTR-02
010102b G*01:01:02:01 UTR-02
010102a G*01:01:19 UTR-02
010102a G*01:01:02:01 UTR-2
010102a G*01:06 UTR-02
010102a G*01:05N UTR-02
010102a G*01:01:14 UTR-02
010102d G*01:01:02:01 UTR-02

Fig. 1. Haplotype network illustrating the similarity among the *HLA-G* extended haplotypes found in two geographically and demographically distinct population samples, Brazil and Cyprus. Each mutation is indicated as a hatch mark. When more than five mutations are detected between two haplotypes, the total number of mutations is indicated. This haplotype network was constructed using the median-joining algorithm of the PopArt program.

haplotypes, in which one promoter haplotype is usually associated with a unique coding allele, and this association is not population-dependent since the same pattern was detected in both samples (Table 6). However, the highest Tajima's *D* values here observed for the coding segment (Table 1) and not observed for the regulatory segments do not support this reasoning. In addition, a purifying selection signature was detected when the *HLA-G* CDS is considered and both samples pooled together (H_A: dn < ds, dn-ds = 2.7404, P = 0.0035), reinforcing what was previously observed for worldwide *HLA-G* sequences retrieved from the IMGT/HLA database (Mendes-Junior et al., 2013).

The 3′ untranslated region evaluation revealed the presence of many haplotypes, but they have already been well documented and inferred by using probabilistic models (Donadi et al., 2011; Nilsson et al., 2016; Sabbagh et al., 2014; Castelli et al., 2010; Castelli et al., 2011; Lucena-Silva et al., 2012; Porto et al., 2015; Santos et al., 2013; Castelli et al., 2014b). The present 3′UTR analysis was conducted with a small number of missing alleles (only 0.04%) and with 74.03% of the heterozygosis phased directly by the GATK ReadBackedPhasing. Even thought, the same well-documented haplotypes were detected, with a summed general frequency of 98.70% (UTR-1 to UTR-18). Likewise the promoter segment, all frequent haplotypes were detected both in Brazil and Cyprus, and in many populations studied so far (Nilsson et al., 2016; Sabbagh et al., 2014; Gineau et al., 2015; Castelli et al., 2010; Castelli et al., 2011; de Albuquerque et al., 2016; Zambra et al., 2016; Santos et al., 2013; Castelli et al., 2014b; Sabbagh et al., 2013), which may indicate that they might have arisen in Africa before human dispersion. In addition, the *HLA-G* 3′UTR presented a higher nucleotide diversity when compared to other *HLA-G* segments, for both samples (Table 1).

The extended haplotype pattern observed for *HLA-G* is mostly identical to the ones already described by using probabilistic models (Donadi et al., 2011; Nilsson et al., 2016; Gineau et al., 2015; Castelli et al., 2010; Castelli et al., 2011; Lucena-Silva et al., 2012; Santos et al., 2013; Castelli et al., 2014b), but now they were confirmed by a different haplotyping approach that do not rely solely on statistical inference. In general, each coding allele is associated with a specific promoter and 3′UTR haplotype (Table 6). This supports the high LD detected for *HLA-G* in previous studies (Santos et al., 2013). The haplotypes detected in Brazil and Cyprus do indicate that the *HLA-G* variability might have arisen in Africa before human dispersion, because almost all haplotypes were detected in both these geographically and demographically distinct samples. In fact, this is also supported by the *1000 Genomes* data, which indicated that most of the *HLA-G* haplotypes are indeed detected worldwide (Castelli et al., 2014a; Gineau et al., 2015; Castelli et al., 2014b). In addition, unlike Cyprus, Brazil presents a great African and Amerindian background, but the summed frequency of all haplotypes detected exclusively in Brazil is quite low, which reinforces the evidence that the haplotypes detected here are quite old and might be found worldwide.

The frequency of extended haplotypes varies between samples, but in general, the two most frequent haplotypes, i.e., 010101a/G*01:01:01:01/UTR-1 and 010102a/G*01:01:02:01/UTR-2, are present in both samples and they represent the most divergent ones in terms of the number of variable sites. However, they are functionally different only in the regulatory segments, since they encode the same HLA-G molecule. As already addressed in other studies, it is possible that the *HLA-G* locus is under balancing selection, which would increase the frequency of haplotypes with differential regulation profiles within a population (Sabbagh et al., 2014; Gineau et al., 2015; Tan et al., 2005; Castelli et al., 2011; Veit et al., 2012). In the present study, both Brazil and Cyprus did present a Tajima's *D* value higher than 2.4, reinforcing the aforementioned balancing selection evidence.

To the best of our knowledge, this is the first study to fully characterize the *HLA-G* gene variability considering both regulatory segments and the entire coding sequence in Cyprus. In addition, this is the first study to fully characterize the *HLA-G* variability using NGS procedures. The methodologies proposed here were able to reliably detect the *HLA-G* variability considering the promoter, complete coding and 3′ untranslated region segments, and even as extended haplotypes, relying more on the straightforward haplotyping capabilities of NGS procedures, and less on probabilistic models. Besides the possible imputation errors that might happen in the two short segments presenting a fragmentation bias because of the use of Nextera kits, we detected the presence of many well-documented and new *HLA-G* coding alleles in Brazil and Cyprus, indicating that the *HLA-G* coding variability is much higher than our currently knowledge. Finally, most of the haplotypes described here (and all the frequent ones) were detected both in Brazil and Cyprus, which are geographically distant and present a completely different demographic history, indicating that these haplotypes are quite old and may be present worldwide.

Acknowledgements

This work was supported by grants obtained from the São Paulo Research Foundation (FAPESP/Brazil – grants # 2013/17084-2 and # 2014/18730-8), and from CNPq/Brazil (grants # 304471/2013-5, 304931/2014-4 and 309572/2014-2).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.molimm.2017.01.020.

References

Alegre, R., Moscoso, J., Martinez-Laso, J., Martin-Villa, M., Suarez, J., Moreno, A., Serrano-Vela, J.I., Vargas-Alarcon, G., Pacheco, R., Arnaiz-Villena, A., 2007. HLA genes in Cubans and the detection of Amerindian alleles. Mol. Immunol. 44, 2426–2435.

Aractingi, S., Briand, N., Le Danff, C., Viguier, M., Bachelez, H., Michel, L., Dubertret, L., Carosella, E.D., 2001. HLA-G and NK receptor are expressed in psoriatic skin: a possible pathway for regulating infiltrating T cells? Am. J. Pathol. 159, 71–77.

Arnaiz-Villena, A., Vargas-Alarcon, G., Serrano-Vela, J.I., Reguera, R., Martinez-Laso, J., Silvera-Redondo, C., Granados, J., Moscoso, J., 2007a. HLA-E polymorphism in amerindians from Mexico (Mazatecans), Colombia (Wayu) and Chile (Mapuches): evolution of MHC-E gene. Tissue Antigens 69 (Suppl 1), 132–135.

Arnaiz-Villena, A., Martinez-Laso, J., Serrano-Vela, J.I., Reguera, R., Moscoso, J., 2007b. HLA-G polymorphism and evolution. Tissue Antigens 69 (Suppl 1), 156–159.

Brandt, D.Y., Aguiar, V.R., Bitarello, B.D., Nunes, K., Goudet, J., Meyer, D., 2015. Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. G3 5, 931–941.

Brenol, C.V., Veit, T.D., Chies, J.A., Xavier, R.M., 2012. The role of the HLA-G gene and molecule on the clinical expression of rheumatologic diseases. Rev Bras Reumatol 52, 82–91.

Cao, M., Yie, S.M., Liu, J., Ye, S.R., Xia, D., Gao, E., 2011. Plasma soluble HLA-G is a potential biomarker for diagnosis of colorectal gastric, esophageal and lung cancer. Tissue Antigens 78, 120–128.

Carlini, F., Traore, K., Cherouat, N., Roubertoux, P., Buhler, S., Cortey, M., Simon, S., Doumbo, O., Chiaroni, J., Picard, C., Di Cristofaro, J., 2013. HLA-G UTR haplotype conservation in the Malian population: association with soluble HLA-G. PLoS One 8, e82517.

Carosella, E.D., LeMaoult, J., 2011. HLA-G: a look back a look forward. Cell. Mol. Life Sci. 68, 337–340.

Carosella, E.D., Moreau, P., Le Maoult, J., Le Discorde, M., Dausset, J., Rouas-Freiss, N., 2003. HLA-G molecules: from maternal-fetal tolerance to tissue acceptance. Adv. Immunol. 81, 199–252.

Carosella, E.D., 2011. The tolerogenic molecule HLA-G. Immunol. Lett. 138, 22–24.

Carvalho dos Santos, L., Tureck, L.V., Wowk, P.F., Mattar, S.B., Gelmini, G.F., Magalhaes, J.C., Bicalho Mda, G., Roxo, V.M., 2013. HLA-E polymorphisms in an afro-descendant southern brazilian population. Hum. Immunol. 74, 199–202.

Castelli, E.C., Mendes-Junior, C.T., Donadi, E.A., 2007. HLA-G alleles and HLA-G 14 bp polymorphisms in a Brazilian population. Tissue Antigens 70, 62–68.

Castelli, E.C., Mendes-Junior, C.T., Deghaide, N.H., de Albuquerque, R.S., Muniz, Y.C., Simoes, R.T., Carosella, E.D., Moreau, P., Donadi, E.A., 2010. The genetic structure of 3′untranslated region of the HLA-G gene: polymorphisms and haplotypes. Genes Immun. 11, 134–141.

Castelli, E.C., Mendes-Junior, C.T., Veiga-Castelli, L.C., Roger, M., Moreau, P., Donadi, E.A., 2011. A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. Mol. Biol. Evol. 28, 3069–3086.

Castelli, E.C., Veiga-Castelli, L.C., Yaghi, L., Moreau, P., Donadi, E.A., 2014a. Transcriptional and posttranscriptional regulations of the HLA-G gene. J. Immunol. Res. 2014, 734068.

Castelli, E.C., Ramalho, J., Porto, I.O., Lima, T.H., Felicio, L.P., Sabbagh, A., Donadi, E.A., Mendes-Junior, C.T., 2014b. Insights into HLA-G genetics provided by worldwide haplotype diversity. Front. Immunol. 5, 476.

Castelli, E.C., Mendes-Junior, C.T., Sabbagh, A., Porto, I.O., Garcia, A., Ramalho, J., Lima, T.H., Massaro, J.D., Dias, F.C., Collares, C.V., Jamonneau, V., Bucheton, B., Camara, M., Donadi, E.A., 2015. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. Hum. Immunol. 76, 945–953.

Catamo, E., Addobbati, C., Segat, L., Sotero Fragoso, T., Domingues Barbosa, A., Tavares Dantas, A., de Ataide Mariz, H., da Rocha L, J.F., Branco Pinto Duarte, A.L., Monasta, L., Sandrin-Garcia, P., Crovella, S., 2014. HLA-G gene polymorphisms associated with susceptibility to rheumatoid arthritis disease and its severity in Brazilian patients. Tissue Antigens 84, 308–315.

Catamo, E., Addobbati, C., Segat, L., Sotero Fragoso, T., Tavares Dantas, A., de Ataide Mariz, H., Ferreira da Rocha Junior, L., Branco PintoDuarte, A.L., Coelho, A.V., de Moura, R.R., Polesello, V., Crovella, S., Sandrin Garcia, P., 2015. Comprehensive analysis of polymorphisms in the HLA-G 5' upstream regulatory and 3' untranslated regions in Brazilian patients with systemic lupus erythematosus. Tissue Antigens 85, 458–465.

Chazara, O., Xiong, S., Moffett, A., 2011. Maternal KIR and fetal HLA-C: a fine balance. J. Leukoc. Biol. 90, 703–716.

Christiansen, O.B., Kolte, A.M., Dahl, M., Larsen, E.C., Steffensen, R., Nielsen, H.S., Hviid, T.V., 2012. Maternal homozygocity for a 14 base pair insertion in exon 8 of the HLA-G gene and carriage of HLA class II alleles restricting HY immunity predispose to unexplained secondary recurrent miscarriage and low birth weight in children born to these patients. Hum. Immunol. 73, 699–705.

Colonna, M., Navarro, F., Bellon, T., Llano, M., Garcia, P., Samaridis, J., Angman, L., Cella, M., Lopez-Botet, M., 1997. A common inhibitory receptor for major histocompatibility complex class I molecules on human lymphoid and myelomonocytic cells. J. Exp. Med. 186, 1809–1818.

Consiglio, C.R., Veit, T.D., Monticielo, O.A., Mucenic, T., Xavier, R.M., Brenol, J.C., Chies, J.A., 2011. Association of the HLA-G gene +3142C > G polymorphism with systemic lupus erythematosus. Tissue Antigens 77 (6), 540–545, http://dx.doi.org/10.1111/j.1399-0039.2011.01635.x.

Courtin, D., Milet, J., Sabbagh, A., Massaro, J.D., Castelli, E.C., Jamonneau, V., Bucheton, B., Sese, C., Favier, B., Rouas-Freiss, N., Moreau, P., Donadi, E.A., Garcia, A., 2013. HLA-G 3' UTR-2 haplotype is associated with Human African trypanosomiasis susceptibility. Infect. Genet. Evol. 17, 1–7.

Crispim, J.C., Duarte, R.A., Soares, C.P., Costa, R., Silva, J.S., Mendes-Junior, C.T., Wastowski, I.J., Faggioni, L.P., Saber, L.T., Donadi, E.A., 2008. Human leukocyte antigen-G expression after kidney transplantation is associated with a reduced incidence of rejection. Transpl. Immunol. 18, 361–367.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

Dias, F.C., Castelli, E.C., Collares, C.V., Moreau, P., Donadi, E.A., 2015. The role of HLA-G molecule and HLA-G gene polymorphisms in tumors viral hepatitis, and parasitic diseases. Front. Immunol. 6, 9.

Diehl, M., Munz, C., Keilholz, W., Stevanovic, S., Holmes, N., Loke, Y.W., Rammensee, H.G., 1996. Nonclassical HLA-G molecules are classical peptide presenters. Curr. Biol. 6, 305–314.

Djurisic, S., Hviid, T.V., 2014. HLA class ib molecules and immune cells in pregnancy and preeclampsia. Front. Immunol. 5, 652.

Donadi, E.A., Castelli, E.C., Arnaiz-Villena, A., Roger, M., Rey, D., Moreau, P., 2011. Implications of the polymorphism of HLA-G on its function regulation, evolution and disease association. Cell. Mol. Life Sci. 68, 369–395.

Dong, D.D., Yie, S.M., Li, K., Li, F., Xu, Y., Xu, G., Song, L., Yang, H., 2012. Importance of HLA-G expression and tumor infiltrating lymphocytes in molecular subtypes of breast cancer. Hum. Immunol. 73, 998–1004.

Dunker, K., Schlaf, G., Bukur, J., Altermann, W.W., Handke, D., Seliger, B., 2008. Expression and regulation of non-classical HLA-G in renal cell carcinoma. Tissue Antigens 72, 137–148.

Excoffier, L., Lischer, H.E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resources 10, 564–567.

Favier, B., Lemaoult, J., Lesport, E., Carosella, E.D., 2010. ILT2/HLA-G interaction impairs NK-cell functions through the inhibition of the late but not the early events of the NK-cell activating synapse. FASEB J. 24, 689–699.

Felicio, L.P., Porto, I.O., Mendes-Junior, C.T., Veiga-Castelli, L.C., Santos, K.E., Vianello-Brondani, R.P., Sabbagh, A., Moreau, P., Donadi, E.A., Castelli, E.C., 2014. Worldwide HLA-E nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. Tissue Antigens 83, 82–93.

Flores, A.C., Marcos, C.Y., Paladino, N., Arruvito, L., Williams, F., Middleton, D., Fainboim, L., 2007. KIR receptors and HLA-C in the maintenance of pregnancy. Tissue Antigens 69 (Suppl 1), 112–113.

Garcia, A., Milet, J., Courtin, D., Sabbagh, A., Massaro, J.D., Castelli, E.C., Migot-Nabias, F., Favier, B., Rouas-Freiss, N., Donadi, E.A., Moreau, P., 2013. Association of HLA-G 3'UTR polymorphisms with response to malaria infection: a first insight. Infect. Genet. Evol. 16, 263–269.

Garziera, M., Bidoli, E., Cecchin, E., Mini, E., Nobili, S., Lonardi, S., Buonadonna, A., Errante, D., Pella, N., D'Andrea, M., De Marchi, F., De Paoli, A., Zanusso, C., De Mattia, E., Tassi, R., Toffoli, G., 2015. HLA-G 3'UTR polymorphisms impact the prognosis of stage II-III CRC patients in fluoropyrimidine-based treatment. PLoS One 10, e0144000.

Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R., 2015. A global reference for human genetic variation. Nature 526, 68–74.

Gineau, L., Luisi, P., Castelli, E.C., Milet, J., Courtin, D., Cagnin, N., Patillon, B., Laayouni, H., Moreau, P., Donadi, E.A., Garcia, A., Sabbagh, A., 2015. Balancing immunity and tolerance: genetic footprint of natural selection in the transcriptional regulatory region of HLA-G. Genes Immun. 16, 57–70.

Goodridge, J.P., Lathbury, L.J., John, E., Charles, A.K., Christiansen, F.T., Witt, C.S., 2009. The genotype of the NK cell receptor, KIR2DL4, influences INFgamma secretion by decidual natural killer cells. Mol. Hum. Reprod. 15, 489–497.

Gregori, S., Amodio, G., Quattrone, F., Panina-Bordignon, P., 2015. HLA-G orchestrates the early interaction of human trophoblasts with the maternal niche. Front. Immunol. 6, 128.

Hiby, S.E., Apps, R., Sharkey, A.M., Farrell, L.E., Gardner, L., Mulder, A., Claas, F.H., Walker, J.J., Redman, C.W., Morgan, L., Tower, C., Regan, L., Moore, G.E., Carrington, M., Moffett, A., 2010. Maternal activating KIRs protect against human reproductive failure mediated by fetal HLA-C2. J. Clin. Invest. 120, 4102–4110.

Hunt, J.S., Langat, D.K., McIntire, R.H., Morales, P.J., 2006. The role of HLA-G in human pregnancy. Reprod. Biol. Endocrinol. 4 (Suppl 1), S10.

Hviid, T.V., Meldgaard, M., Sorensen, S., Morling, N., 1997. Polymorphism of exon 3 of the HLA-G gene. J. Reprod. Immunol. 35, 31–42.

Hviid, T.V., Sorensen, S., Morling, N., 1999. Polymorphism in the regulatory region located more than 1.1 kilobases 5' to the start site of transcription, the promoter region, and exon 1 of the HLA-G gene. Hum. Immunol. 60, 1237–1244.

Hviid, T.V., Christiansen, O.B., Johansen, J.K., Hviid, U.R., Lundegaard, C., Moller, C., Morling, N., 2001. Characterization of a new HLA-G allele encoding a nonconservative amino acid substitution in the alpha3 domain (exon 4) and its relevance to certain complications in pregnancy. Immunogenetics 53, 48–53.

Hviid, T.V., Hylenius, S., Rorbye, C., Nielsen, L.G., 2003. HLA-G allelic variants are associated with differences in the HLA-G mRNA isoform profile and HLA-G mRNA levels. Immunogenetics 55, 63–79.

Hviid, T.V., Rizzo, R., Melchiorri, L., Stignani, M., Baricordi, O.R., 2006. Polymorphism in the 5' upstream regulatory and 3' untranslated regions of the HLA-G gene in relation to soluble HLA-G and IL-10 expression. Hum. Immunol. 67, 53–62.

Hviid, T.V., 2006. HLA-G in human reproduction: aspects of genetics, function and pregnancy complications. Hum. Reprod. Update 12, 209–232.

Hylenius, S., Andersen, A.M., Melbye, M., Hviid, T.V., 2004. Association between HLA-G genotype and risk of pre-eclampsia: a case-control study using family triads. Mol. Hum. Reprod. 10, 237–246.

Ishitani, A., Sageshima, N., Hatake, K., 2006. The involvement of HLA-E and −F in pregnancy. J. Reprod. Immunol. 69, 101–113.

Kamishikiryo, J., Maenaka, K., 2009. HLA-G molecule. Curr. Pharm. Des. 15, 3318–3324.

Kovats, S., Main, E.K., Librach, C., Stubblebine, M., Fisher, S.J., DeMars, R., 1990. A class I antigen HLA-G, expressed in human trophoblasts. Science 248, 220–223.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Larsen, M.H., Hylenius, S., Andersen, A.M., Hviid, T.V., 2010. The 3'-untranslated region of the HLA-G gene in relation to pre-eclampsia: revisited. Tissue Antigens 75, 253–261.

Le Discorde, M., Moreau, P., Sabatier, P., Legeais, J.M., Carosella, E.D., 2003. Expression of HLA-G in human cornea, an immune-privileged tissue. Hum. Immunol. 64, 1039–1044.

LeMaoult, J., Zafaranloo, K., Le Danff, C., Carosella, E.D., 2005. HLA-G up-regulates ILT2, ILT3, ILT4, and KIR2DL4 in antigen presenting cells NK cells, and T cells. FASEB J. 19, 662–664.

Lefebvre, S., Adrian, F., Moreau, P., Gourand, L., Dausset, J., Berrih-Aknin, S., Carosella, E.D., Paul, P., 2000. Modulation of HLA-G expression in human thymic and amniotic epithelial cells. Hum. Immunol. 61, 1095–1101.

Lefebvre, S., Antoine, M., Uzan, S., McMaster, M., Dausset, J., Carosella, E.D., Paul, P., 2002. Specific activation of the non-classical class I histocompatibility HLA-G antigen and expression of the ILT2 inhibitory receptor in human breast cancer. J. Pathol. 196, 266–274.

Lima, T.H., Buttura, R.V., Donadi, E.A., Veiga-Castelli, L.C., Mendes-Junior, C.T., Castelli, E.C., 2016. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. Hum. Immunol. 77, 841–853.

Liu, X.X., Pan, F.H., Tian, W., 2012. Characterization of HLA-E polymorphism in four distinct populations in Mainland China. Tissue Antigens 80, 26–35.

Lucena-Silva, N., Monteiro, A.R., de Albuquerque, R.S., Gomes, R.G., Mendes-Junior, C.T., Castelli, E.C., Donadi, E.A., 2012. Haplotype frequencies based on eight polymorphic sites at the 3' untranslated region of the HLA-G gene in individuals from two different geographical regions of Brazil. Tissue Antigens 79, 272–278.

Lucena-Silva, N., de Souza, V.S., Gomes, R.G., Fantinatti, A., Muniz, Y.C., de Albuquerque, R.S., Monteiro, A.L., Diniz, G.T., Coelho, M.R., Mendes-Junior, C.T., Castelli, E.D., Donadi, E.A., 2013. HLA-G 3' untranslated region polymorphisms are associated with systemic lupus erythematosus in 2 brazilian populations. J. Rheumatol. 120814, http://dx.doi.org/10.3899/jrheum.120814 (DOI jrheum.120814 [pii].).

Martelli-Palomino, G., Pancotto, J.A., Muniz, Y.C., Mendes-Junior, C.T., Castelli, E.C., Massaro, J.D., Krawice-Radanne, I., Poras, I., Rebmann, V., Carosella, E.D., Rouas-Freiss, N., Moreau, P., Donadi, E.A., 2013. Polymorphic sites at the 3' untranslated region of the HLA-G gene are associated with differential hla-g soluble levels in the Brazilian and French population. PLoS One 8, e71742.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Mendes-Junior, C.T., Castelli, E.C., Meyer, D., Simoes, A.L., Donadi, E.A., 2013. Genetic diversity of the HLA-G coding region in Amerindian populations from the Brazilian Amazon: a possible role of natural selection. Genes Immun. 14, 518–526.

Misra, M.K., Pandey, S.K., Kapoor, R., Sharma, R.K., Kapoor, R., Prakash, S., Agrawal, S., 2014. HLA-G gene expression influenced at allelic level in association with end stage renal disease and acute allograft rejection. Hum. Immunol. 75, 833–839.

Moscoso, J., Serrano-Vela, J.I., Pacheco, R., Arnaiz-Villena, A., 2006. HLA-G, −E and −F: allelism, function and evolution. Transpl. Immunol. 17, 61–64.

Munz, C., Nickolaus, P., Lammert, E., Pascolo, S., Stevanovic, S., Rammensee, H.G., 1999a. The role of peptide presentation in the physiological function of HLA-G. Semin. Cancer Biol. 9, 47–54.

Munz, C., Stevanovic, S., Rammensee, H.G., 1999b. Peptide presentation and NK inhibition by HLA-G. J. Reprod. Immunol. 43, 139–155.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nilsson, L.L., Djurisic, S., Andersen, A.M., Melbye, M., Bjerre, D., Ferrero-Miliani, L., Hackmon, R., Geraghty, D.E., Hviid, T.V., 2016. Distribution of HLA-G extended haplotypes and one HLA-E polymorphism in a large-scale study of mother-child dyads with and without severe preeclampsia and eclampsia. HLA: Immune Response Genet. 88, 172–186.

Porto, I.O., Mendes-Junior, C.T., Felicio, L.P., Georg, R.C., Moreau, P., Donadi, E.A., Chies, J.A., Castelli, E.C., 2015. MicroRNAs targeting the immunomodulatory HLA-G gene: a new survey searching for microRNAs with potential to regulate HLA-G. Mol. Immunol. 65, 230–241.

Rajagopalan, S., Long, E.O., 1999. A human histocompatibility leukocyte antigen (HLA)-G-specific receptor expressed on all natural killer cells. J. Exp. Med. 189, 1093–1100.

Rizzo, R., Hviid, T.V., Govoni, M., Padovan, M., Rubini, M., Melchiorri, L., Stignani, M., Carturan, S., Grappa, M.T., Fotinidi, M., Ferretti, S., Voss, A., Laustrup, H., Junker, P., Trotta, F., Baricordi, O.R., 2008. HLA-G genotype and HLA-G expression in systemic lupus erythematosus: HLA-G as a putative susceptibility gene in systemic lupus erythematosus. Tissue Antigens 71, 520–529.

Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P., Marsh, S.G., 2015. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 43, D423–431.

Sabbagh, A., Courtin, D., Milet, J., Massaro, J.D., Castelli, E.C., Migot-Nabias, F., Favier, B., Rouas-Freiss, N., Moreau, P., Garcia, A., Donadi, E.A., 2013. Association of HLA-G 3' untranslated region polymorphisms with antibody response against Plasmodium falciparum antigens: preliminary results. Tissue Antigens 82, 53–58.

Sabbagh, A., Luisi, P., Castelli, E.C., Gineau, L., Courtin, D., Milet, J., Massaro, J.D., Laayouni, H., Moreau, P., Donadi, E.A., Garcia, A., 2014. Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity. Genes Immun. 15, 95–106.

Santos, K.E., Lima, T.H., Felicio, L.P., Massaro, J.D., Palomino, G.M., Silva, A.C., Oliveira, S.F., Sabbagh, A., Garcia, A., Moreau, P., Donadi, E.A., Mendes-Junior, C.T., Castelli, E.C., 2013. Insights on the HLA-G evolutionary history provided by a nearby Alu insertion. Mol. Biol. Evol. 30 (11), 2423–2434, http://dx.doi.org/10.1093/molbev/mst142 (DOI mst142 [pii].).

Shiroishi, M., Tsumoto, K., Amano, K., Shirakihara, Y., Colonna, M., Braud, V.M., Allan, D.S., Makadzange, A., Rowland-Jones, S., Willcox, B., Jones, E.Y., van der Merwe, P.A., Kumagai, I., Maenaka, K., 2003. Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8 for MHC class I binding and bind preferentially to HLA-G. Proc. Natl. Acad. Sci. U. S. A. 100, 8856–8861.

Sizzano, F., Testi, M., Zito, L., Crocchiolo, R., Troiano, M., Mazzi, B., Turchiano, G., Torchio, M., Pultrone, C., Gregori, S., Chiesa, R., Gaziev, J., Sodani, P., Marktel, S., Amoroso, A., Roncarolo, M.G., Lucarelli, G., Ciceri, F., Andreani, M., Fleischhauer, K., 2012. Genotypes and haplotypes in the 3' untranslated region of the HLA-G gene and their association with clinical outcome of hematopoietic stem cell transplantation for beta-thalassemia. Tissue Antigens 79, 326–332.

Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73, 1162–1169.

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989.

Tan, Z., Shon, A.M., Ober, C., 2005. Evidence of balancing selection at the HLA-G promoter region. Hum. Mol. Genet. 14, 3619–3628.

Tan, C.Y., Chong, Y.S., Loganath, A., Chan, Y.H., Ravichandran, J., Lee, C.G., Chong, S.S., 2009. Possible gene–gene interaction of KIR2DL4 with its cognate ligand HLA-G in modulating risk for preeclampsia. Reprod. Sci. 16, 1135–1143.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinform./Edit. Board 11 (1110) (Andreas D. Baxevanis . . . [et al.] 11.10.1–11.10.33.

Veiga-Castelli, L.C., Castelli, E.C., Mendes Jr., C.T., da Silva Jr., W.A., Faucher, M.C., Beauchemin, K., Roger, M., Moreau, P., Donadi, E.A., 2012. Non-classical HLA-E gene variability in Brazilians: a nearly invariable locus surrounded by the most variable genes in the human genome. Tissue Antigens 79, 15–24.

Veiga-Castelli, L.C., Bertuol, J.M., Castelli, E.C., Donadi, E.A., 2016. Low variability at the HLA-E promoter region in the Brazilian population. Hum. Immunol. 77, 172–175.

Veit, T.D., Cazarolli, J., Salzano, F.M., Schiengold, M., Chies, J.A., 2012. New evidence for balancing selection at the HLA-G locus in South Amerindians. Genet. Mol. Biol. 35, 919–923.

Veit, T.D., de Lima, C.P., Cavalheiro, L.C., Callegari-Jacques, S.M., Brenol, C.V., Brenol, J.C., Xavier, R.M., 2014. M.F. da Cunha Sauma, E.J. dos Santos, J.A., Chies, HLA-G +3142 polymorphism as a susceptibility marker in two rheumatoid arthritis populations in Brazil. Tissue Antigens 83, 260–266.

Wang, W., Wei, Z., Lam, T.W., Wang, J., 2011. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. Sci. Rep. 1, 55.

Zambra, F.M., Biolchi, V., de Cerqueira, C.C., Brum, I.S., Castelli, E.C., Chies, J.A., 2016. Immunogenetics of prostate cancer and benign hyperplasia – the potential use of an HLA-G variant as a tag SNP for prostate cancer risk. HLA: Immune Response Genet. 87, 79–88.

de Albuquerque, R.S., Mendes-Junior, C.T., Lucena-Silva, N., da Silva, C.L., Rassi, D.M., Veiga-Castelli, L.C., Foss-Freitas, M.C., Foss, M.C., Deghaide, N.H., Moreau, P., Gregori, S., Castelli, E.C., Donadi, E.A., 2016. Association of HLA-G 3' untranslated region variants with type 1 diabetes mellitus. Hum. Immunol. 77, 358–364.