IMMUNOLOGY   ORIGINAL ARTICLE

Gene conversion of the major histocompatibility complex class I Caja-G in common marmosets (*Callithrix jacchus*)

Anna-Lena Neehus,[1]
Joachim Wistuba,[2]
Nektarios Ladas,[1]
Britta Eiz-Vesper,[1]
Stefan Schlatt[2] and
Thomas Müller[1,3]

[1]*Institute for Transfusion Medicine, Hannover Medical School, Hannover,* [2]*Institute of Reproductive and Regenerative Biology, Centre of Reproductive Medicine and Andrology, University Münster, Münster, and* [3]*Synlab Medical Care Centre Weiden Ltd, Weiden, Germany*

Summary

Currently, the amount of sequenced and classified MHC class I genes of the common marmoset is limited, in spite of the wide use of this species as an animal model for biomedical research. In this study, 480 clones of MHC class I G locus (Caja-G) cDNA sequences were obtained from 21 common marmosets. Up to 10 different alleles were detected in each common marmoset, leading to the assumption that the Caja-G loci duplicated in the marmoset genome. In the investigated population, four alleles occurred more often, giving evidence for higher immunological advantage of these alleles. In contrast to the human non-classical MHC class I genes, Caja-G shows high rates of polymorphism at the relevant peptide-binding sites, despite its phylogenetic relationship to the non-classical HLA-G. Our results provide information for better understanding of the immunological properties of the common marmoset and confirm the theory of a gene conversion of the Caja-G due to its detected plasticity and the absence of any known HLA-A equivalent.

Keywords: Caja-G; common marmoset; major histocompatibility complex; new world primate.

Introduction

The common marmoset (*Callithrix jacchus*) is a small New World primate, which is a well-established pre-clinical animal model in the field of stem cell research and regenerative, reproductive and neural medicine.[1–5] Its advantages compared with other non-human primates are the small size and weight, easy handling and breeding, as well as low diet and housing costs.[6] Interestingly, marmoset twins are bone marrow chimeras by nature, resulting from placental anastomoses through which haematopoietic cells exchange during pregnancy.[7] As a result of this chimeric character, genetically distinct cells can be found in one animal.[8]

The MHC is a varied gene family that encodes for immunoregulatory proteins playing a major role in autoimmunity as well as the adaptive immune system.[9]

The classical MHC class I molecules are expressed on every nucleated cell, presenting degraded antigen peptides to CD8+ T cells. MHC molecules are peptide receptors that bind predominantly peptides that are nine residues in length by interaction with at least three of their six binding pockets (A–F). The N-terminal residues are bound by the A pocket, whereas the C-terminal anchor side chains are accepted by the F pocket, formed by 10 amino acids in the α1 and α2 domains. Also crucial for the specificity of peptide-binding is the B pocket, which is located on the α1 domain of the MHC molecules.[10,11] In humans, the classical human leucocyte antigens HLA-A, HLA-B and HLA-C are highly polymorphic, whereas the non-classical HLA-E, HLA-F and HLA-G show lower levels of polymorphism and their expression is limited to some tissues.[12,13] HLA-G proteins are known to play an important role in induction of immune tolerance between

---

Abbreviations: CeRa, Centre of Reproductive Medicine and Andrology; d_N, non-synonymous substitutions; d_S, synonymous substitutions; HLA, human leucocyte antigen; MHC, major histocompatibility complex; RT-PCR, reverse transcription polymerase chain reaction

the mother and the partially allogeneic fetus during pregnancy.[14,15] For investigation of the immunological functions of the MHC system, the classical mouse model cannot be applied because the non-classical MHC molecules of mouse and human are not homologous. Orthologues of the non-classical genes can be found within, but not outside, phylogenetic families, wherefore a primate model is necessary for more precise immunological analysis.[16]

The MHC class I genes of New World primates (Caja) seem to show higher sequence homologies with the non-classical human HLA-G gene than with the classical MHC class I genes.[8] For the common marmoset, no transcription of HLA-A-like proteins could be detected but they seem to have an intact B/C segment as well as MHC-E and MHC-G transcripts.[17] As marmosets can give birth to dizygotic twins, the role of the Caja-G, which seems to be orthologous to the HLA-G loci,[18] has to be investigated for better understanding of the immunity of the common marmoset. The animal could be regarded as a natural model of cellular tolerance and induction of chimerism, which is of foremost interest in future cell and organ transplantation for regenerative medicine in the human. In contrast, transplantation studies using the marmoset need to be closely verified in terms of true third-party allogenicity. In each case, the distribution of MHC alleles in the target tissue must be verified before transplantation procedures. However, despite the wide use of the marmoset as a transplantation model, the amount of analysed and classified HLA-G related genes of the common marmoset is limited. In this study, transcripts of the MHC class I G locus (Caja-G) were analysed in terms of phylogenetic relationship and level of polymorphism to explore the functional position in the common marmoset and understand better the immunological properties of this important animal model.

Material and methods

Receipt of cell and tissue samples

Twenty-one animals out of three families (Fig. 1) were selected for sequence analysis, in cooperation with the Centre of Reproductive Medicine and Andrology (CeRA) Münster, Germany. The monkeys from the breeding colony at CeRA were kept in groups under a 12-hr light : 12-hr darkness regimen and fed pelleted food from Altromin (Lage, Germany) composed for marmosets together with beef or chicken meat and a daily supplement of fresh fruits and vegetables. They had unlimited access to tap water. The experimental work was performed in accordance with the German Federal Law on the Care and Use of Laboratory Animals (LANUV NRW Licences 8·87-50·10·46·09·018, 84-02·05·20·12·0·18, 8·87-51·05·20·10·154). Blood and spleen samples were used for analyses. To avoid cross-contamination, samples were taken at different time-points and sterile single-use scalpels were used for spleen preparation. Blood samples were obtained using single-use syringes.

RNA extraction and RT-PCR

RNA extraction from tissue and cell samples was performed using a peqGOLD Total RNA Kit (Peqlab GmbH, Erlangen, Germany) following the manufacturer's instructions. Contaminant DNA was removed by a DNA Removing Column. Eluted RNA was dissolved in 50 μl sterile RNase-free dH$_2$O and final concentration was
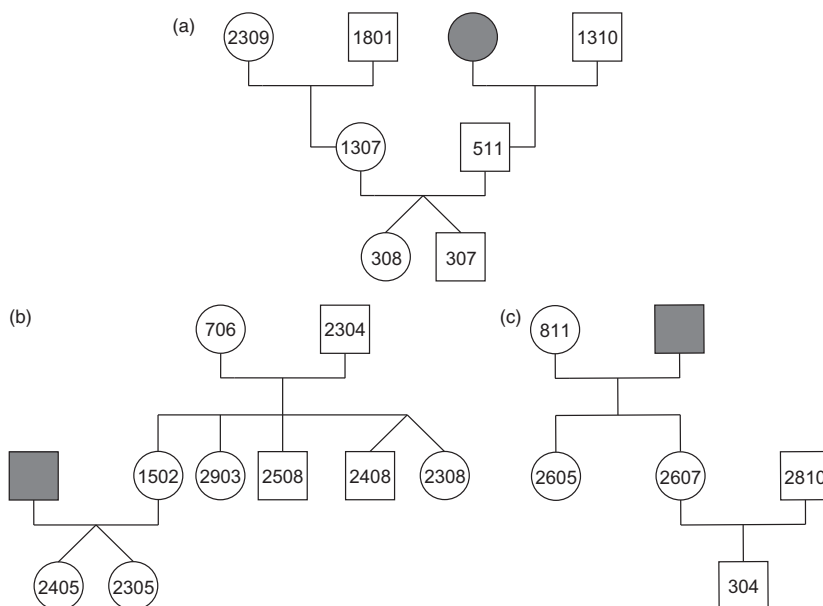

Figure 1. Relationship of animals from three families (a–c) used in this study. Squares represent male animals; circles represent female animals. Symbols coloured grey represent animals that were not included in this study.

measured with a NanoDrop photometer ND-1000 (Thermo Scientific, Waltham, MA). One microgram of extracted RNA was transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Life Technologies GmbH, Darmstadt, Germany). For additional control of genomic DNA contamination, the same solution was used for PCR but without the reverse transcriptase (−RT). By adding oligo(dT) primers (TIB Molbiol, Berlin, Germany) only mRNA was transcribed. Amplification of Caja-G alleles was obtained by RT-PCR with specific oligonucleotides (5′-GAG ATG GGC G TC ATG GCG CCC-3′ and 5′-CCA TCT CAG GGT GA G GGG-3′) in a reaction of 30 μl per sample, set up as follows: 24 μl dH$_2$O, 3 μl of 10× PCR buffer (NEB, Frankfurt, Germany), 0·5 Units *Taq* polymerase (NEB), 100 mM dNTPs (Fermentas, St Leon-Rot, Germany), 20 pmol/μl of each primer and 1 μg cDNA. The PCR was performed under the following conditions: initial denaturation for 3 min at 95° followed by 35 cycles of denaturation at 95° for 45 seconds, annealing at 62° for 45 seconds, elongation at 72° for 90 seconds with a final extension step at 72° for 10 min.

Cloning and sequencing of PCR fragments

The RT-PCR amplification products were visualized on an agarose gel (1·5%) and fragments of 900 bp were extracted using a peqGOLD Gel Extraction Kit (Peqlab). Afterwards, the products were cloned using the TA Cloning Kit (Life Technologies, Darmstadt, Germany) and One Shot Top 10 Competent Cells (Life Technologies). After 12 hr of incubation on ampicillin-containing agar, colonies were picked and expanded in overnight culture at 37° for plasmid isolation with the peqGOLD Mini Prep Kit (Peqlab). Sequencing reactions were performed with the BigDyeTM Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems v1·1) according to the manufacturer's instructions in 96-well PCR-plates (Kisker, Steinfurt, Germany) in a C1000 Thermal Cycler (BioRad, Munich, Germany). To avoid sequencing errors each PCR fragment was sequenced in both directions.

Sequence analysis

The resulting sequences were analysed using the SEQMAN PRO and MEGAALIGN (DNASTAR, Madison, WI) software and compared with available Caja-G alleles via NCBI Nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Identical sequences from at least three different clones were considered as novel alleles and were submitted to the NCBI database. Nomenclatures were allocated by the Immuno Polymorphism Database for Major Histocompatibility Complex genes of the Non Human Primate (IPD-MHC NHP).[19]

For phylogenetic analysis of Caja-G alleles, the MEGA6 software[20] was used for comparison of exons 1–4. Ancillary exon sequences of the classical and non-classical human MHC class I genes as well as pseudogenes were downloaded for comparison from the IMGT/HLA Database (http://www.ebi.ac.uk/ipd/imgt/hla/). Sequences of common chimpanzees (Patr), rhesus macaques (Mamu), cotton-top tamarins (Saoe) and golden lion tamarins (Lero) were obtained from the IPD-MHC NHP.

Biochemical influences of amino acid substitutions on protein formation were rated using Risler's amino acid substitution matrix.[21] To identify specific codons that are likely to be under selection in the MHC class I complex of the common marmoset, a Wu–Kabat variability analysis was performed on the amino acid alignment of all known 74 Caja-G alleles. Wu–Kabat variability was calculated by dividing the number of different amino acids at a position by the frequency of the most common amino acid at this position. Highly polymorphic sites are defined by having a Wu–Kabat score more than twice the average score for all sites in the evaluated sequences.[22] Selection at the amino acid level was measured by estimating the rates of non-synonymous (d_N) and synonymous (d_S) substitutions by using DNASP5·10. For HLA-G and Caja-G alleles the mean number of non-synonymous and synonymous nucleotide substitutions for the peptide binding region (PBR), α1–α2 remainder and α3 was calculated pursuant to the Nei and Gojobori method with the Jukes and Cantor corrections.[23] For additional comparison, d_N and d_S value calculations of HLA-A, -B and -C of Zemmour and Parham were appropriated.[24] Identification of positive and negative selected sites within the Caja-G alleles was performed using the single likelihood ancestor counting (SLAC) method of www.datamonkey.org.[25]
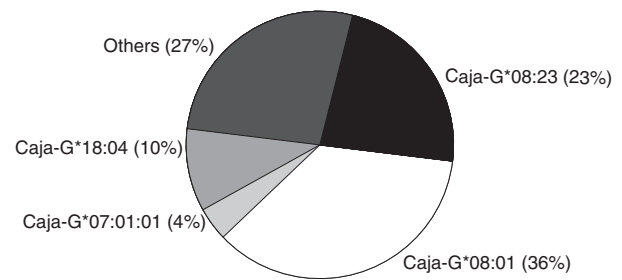
Results

Identification of Caja-G alleles

In our study, 480 sequences of 21 common marmosets were analysed and classified as members of the Caja-G family. Comparison with already confirmed Caja-G sequences[26–28] resulted in 34 different alleles in our investigated group of animals. Four of the detected alleles have been published before[26,27] and a total of 30 new alleles, concerning the sequence of exon 1–4, were uploaded to the NCBI database and are presented in Table 1 with GenBank accession numbers, official IPD-MHC nomenclature and number of animals that shared the correspondent alleles. Of 34 MHC class I sequences most alleles were found in one (15) or two (10) marmosets, whereas other alleles like *Caja-G*08:23* occurred in 11 different marmosets, that originated from different families (compare Fig. 1, Table 1). *Caja-G*08:01* was most frequently detected and was found to be shared by 18 out of 21

Table 1. List of detected Caja-G alleles in 21 common marmoset monkeys from three different families

Allele	Accession number	No. of animals shared the alleles	
Caja-G*07:01:01	HE962239	6/21	Novel
Caja-G*08:01	U59640.1	18/21	
Caja-G*08:08	KC903165	1/21	Novel
Caja-G*08:09	KC862260	3/21	Novel
Caja-G*08:10	KC903141	1/21	Novel
Caja-G*08:11	KC903143	1/21	Novel
Caja-G*08:12	KC903144	1/21	Novel
Caja-G*08:13	KC903148	2/21	Novel
Caja-G*08:14	KC903150	1/21	Novel
Caja-G*08:15	KC903153	1/21	Novel
Caja-G*08:16	KC903154	1/21	Novel
Caja-G*08:17	KC903156	1/21	Novel
Caja-G*08:18	KC903157	4/21	Novel
Caja-G*08:19	KC903158	1/21	Novel
Caja-G*08:20	KC903159	1/21	Novel
Caja-G*08:21	KC903161	2/21	Novel
Caja-G*08:22	KC903162	2/21	Novel
Caja-G*08:23	KC862259	11/21	Novel
Caja-G*09:01	KC903151	2/21	
Caja-G*09:08	KC903155	1/21	Novel
Caja-G*10:01	U59639.1	1/21	
Caja-G*10:03:01	HE962238	2/21	Novel
Caja-G*10:04	KC903142	2/21	Novel
Caja-G*12:03:01	KC862261	8/21	
Caja-G*12:03:02	HE962237	2/21	Novel
Caja-G*18:03	HE962240	5/21	Novel
Caja-G*18:04	HE962236	8/21	Novel
Caja-G*18:06	KC903147	1/21	Novel
Caja-G*18:07	KC903152	1/21	Novel
Caja-G*18:08	KC903160	4/21	Novel
Caja-G*18:09	KC903164	2/21	Novel
Caja-G*23:01	KC903145	2/21	Novel
Caja-G*24:01	KC903146	2/21	Novel
Caja-G*25:01	KC903149	1/21	Novel

Figure 2. Distribution of sequenced Caja-G alleles in all 480 sequenced clones concerning Caja-G*08:01, Caja-G*08:23, Caja-G*18:04 and Caja-G*07:01:01. The fraction entitled 'others' combines a group of 31 alleles with < 4% frequency for a better visualization.

Table 2. Distribution of Caja-G alleles in the investigated group of animals

Family	Animal	Caja-G*
A	2309	07:01:01, 10:03:01, 12:03:01, 12:03:02, 18:03
	1801	07:01:01, 08:10, 10:03:01, 10:02, 12:03:01, 12:03:02, 18:03
	1310	07:01:01, 08:01, 08:23, 10:01, 12:03:01, 18:03
	1307	08:01,08:09,08:23, 10:04, 18:03, 18:04, 18:06, 25:01
	511	08:01, 08:09, 08:23, 12:03:01, 18:04, 18:08, 23:01
	308	08:01, 18:09
	307	18:04
B	706	07:01:01, 08:01, 08:23, 12:03:01, 18:04, 18:08, 18:09
	2304	08:01, 08:11, 18:04, 23:01
	1502	08:01, 08:18 08:21, 08:22, 08:23
	2903	08:01, 08:23, 18:04
	2508	07:01:01, 08:01
	2408	08:01, 08:12, 08:13, 08:18, 08:23, 09:01, 18:04, 18:07, 18:08, 24:01
	2308	08:01, 08:09, 08:14, 08:19, 08:20, 08:22, 08:23, 18:03, 18:08, 24:01
	2405	08:01, 08:17, 08:18, 08:21, 08:23
	2305	08:01, 08:18, 08:21, 08:23, 09:01, 12:03:01
C	811	08:01
	2605	08:01, 18:04
	2607	08:01, 12:03:01
	2810	08:01, 08:15, 08:16
	304	07:01:01, 08:01, 08:08, 08:13, 08:23, 09:08, 12:03:01

individuals. In contrast to former publications[27,29] the observed allocation of the Caja-G alleles was not linear. In 73·34% of all analysed sequences, representing 307 investigated clones, at least one of the alleles Caja-G*08:01 (36·46%), Caja-G*18:04 (9·79%), Caja-G*07:01:01 (4·38%) and Caja-G*08:23 (22·71%) were found (Fig. 2).

For each animal, one to ten distinct Caja-G alleles were detected (Tables 1 and 2). For instance, the twins 2308 and 2408 both carried 10 putative MHC class I sequences, whereas animal number 307 just carried the Caja-G*08:01 allele in each of the 30 sequenced clones. For most examined animals, a number of two to seven different alleles was detected. The examination of the Caja-G distribution in the three common marmoset families indicates that some combinations of alleles seem to segregate on a haplotype. Caja-G*08:01, Caja-G*08:18 and 08:23 often occur

at the same time, as well as the combination of Caja-G*12:03:01 and 07:01:01, even in animals from different families (Table 2). Therefore, this leads to the assumption of at least two Caja-G loci on a haplotype.

Phylogenetic relationship of MHC class I alleles

Phylogenetic analysis of MHC class I alleles of the common marmoset was performed by using the neighbour-joining method for creating the phylogenetic tree of exons 1–4. The analysis comprised sequences of the HLA alleles HLA-A, -B, -C, -E, -F, -G, -H, -J and -K, MHC

class I sequences of rhesus monkey (Mamu), common chimpanzee (Patr), cotton-top tamarin (Saoe), golden lion tamarin (Lero) and the Caja-G alleles that were investigated in this study, all encoding functional MHC Class I proteins.

The identified MHC class I alleles of the common marmoset cluster closer to the non-classical HLA-G, HLA-F and HLA-E branch than to the classical HLA-A, HLA-B or HLA-C alleles, indicating their phylogenetic relationship (Fig. 3a,b). The MHC class I lineage of the investigated New World monkeys is more closely related to

HLA-G than to the other known non-classical human MHC class I genes (bootstrap value 60%). Figure 3(b) also constitutes the close relationship between the cotton-top tamarins Saoe-G and common marmoset Caja-G (bootstrap support of 79%). Clustering of the Callitrichidae species sequences with a high bootstrap support of 99% is leading to the assumption of an orthologous relationship between the MHC-G genes.

Compared with the oligomorphic HLA-G, the Caja-G alleles cluster in different lineages and therefore lineage numbers Caja-G*07 to Caja-G*25 (Table 1, Fig. 3) have
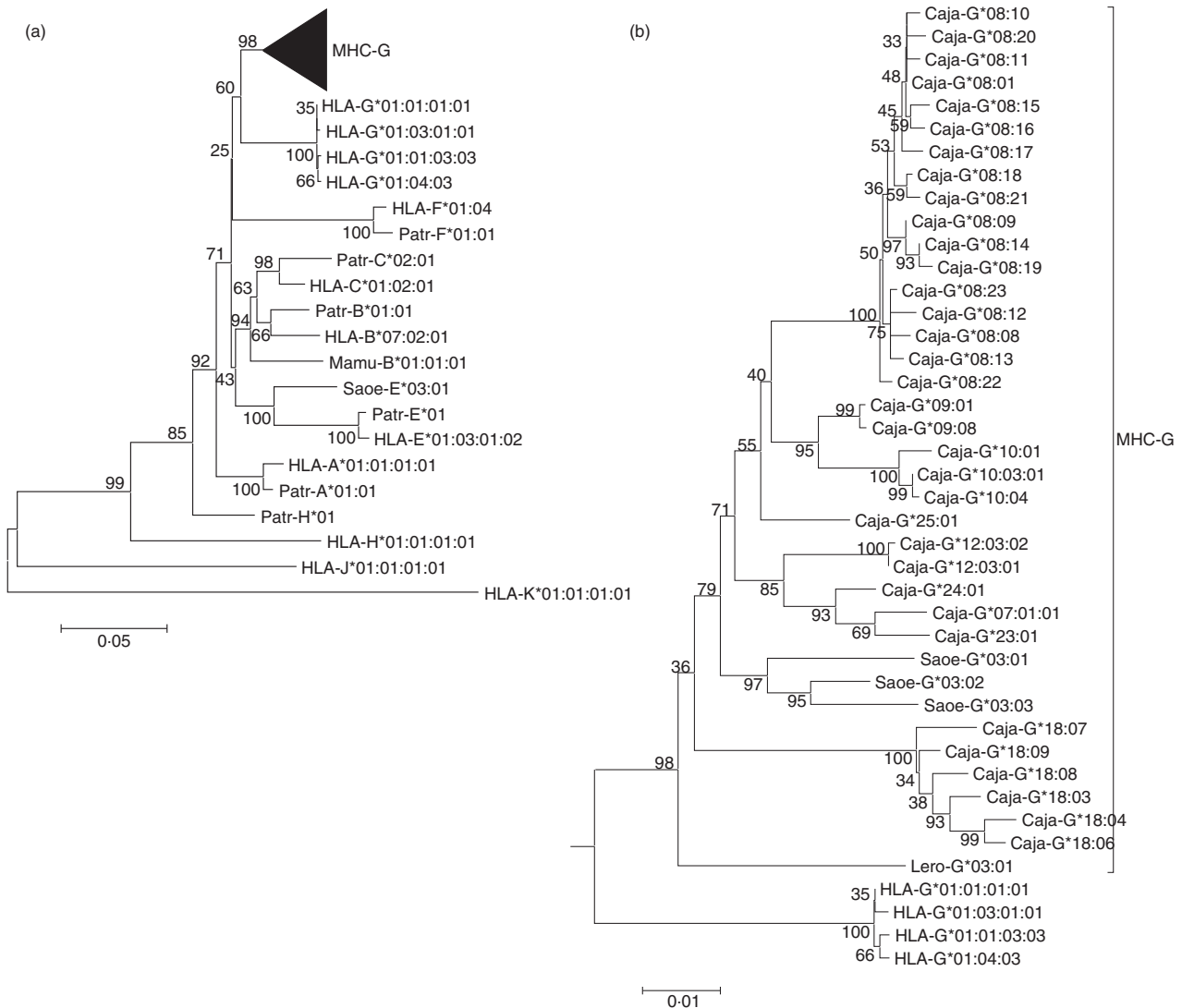

Figure 3. Phylogenetic tree of MHC class I exons 1–4. Evolutionary analysis of phylogenetic relationship utilizing MEGA6 software including MHC class I sequences of humans, common chimpanzees, rhesus macaques, cotton-top tamarins, common marmosets and golden lion tamarins cDNA sequences inferring the neighbour-joining method. The percentages of replicate trees with the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is scaled with branch lengths with the same units as those resembling the evolutionary distances, which were computed using the maximum composite likelihood method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 852 positions in the final data set of 56 analysed sequences. (a) Phylogenetic tree of MHC class I exons 1–4 with compressed subtree of MHC-G; (b) expanded MHC-G subtree of exons 1–4.

been introduced by the IPD-MHC nomenclature committee. The different lineages show abundant rates of polymorphism, which corresponds with the results of former studies.[27–29] Interestingly, the four alleles with the highest occurrence are clustered into three different lineages (Caja-G*07, Caja-G*18, Caja-G*08) leading to the assumption of a high variability on amino acid and therefore also on the nucleotide level.

B/F antigen-binding pocket polymorphism

The Caja-G heavy chain can be divided into five domains with different functions: three extracellular domains forming the α1, α2, α3 chains, one transmembrane domain and one intracellular domain. Crucial for antigen interaction are six binding pockets located in the α1 and α2 chains and transcribed by exons 2 and 3. For binding and anchoring of antigen fragments, the B pocket, located in the α1 domain, and the F pocket, located in α1 and α2, are most important.[11] Comparison of amino acid substitutions in exon 2 and exon 3 revealed abundant polymorphisms not only in positions relevant for the B and F pockets (Fig. 4). Forty-three amino acid substitutions have been detected in exon 2, six of which may influence B pocket and four F pocket formation. In contrast, in exon 3, a lower number of substitutions could be detected and just one out of 22 substitutions may affect the conformation of the F pocket. Every position for B pocket formation in exon 3 shows no diversity at the amino acid level. In conclusion, diversity in the α1 and α2 chains could be detected, whereas the α2 chain is less

subjected to variation than the α1 chain. The observed diversity also affects positions for B and F pocket formation, crucial for peptide-binding (Table 3).

The B and F pockets of the four alleles with the highest occurrence (*Caja-G*07:01:01*, *Caja-G*08:01*, *Caja-G*08:23*, *Caja-G*18:04*) also showed a high number of variations, but only between alleles of different lineages. Comparison of *Caja-G*08:01* and *Caja-G*08:23* showed no amino acid substitution influencing peptide-binding pocket formation, in contrast to comparison of the other most frequently occurring lineages. In total, for the B pocket, five substitutions were observed and for the F pocket, four exchanges at the amino acid level. Amino acid substitutions leading to structural or functional changes were analysed using Risler's amino acids substitution matrix (Table 4) on polymorphic positions.[21] A high Risler score indicates relatively rare substitutions. For example, the change from tyrosine (Tyr) to phenylalanine (Phe) at position 38 has been reported to occur comparatively often leading to a score of four from Risler's matrix. The variation between the four different alleles of the three different MHC class I lineages affects the F pocket because the exchanges lead to higher Risler scores compared with substitutions on B pocket positions. Changes in amino acid composition may lead to conformational changes and variant binding specificity. Although the four alleles are predominant in our study, high variability at a nucleotide level, as well as an amino acid level, can be observed. The alleles exhibit high diversity among themselves, also concerning peptide-binding pocket formation.
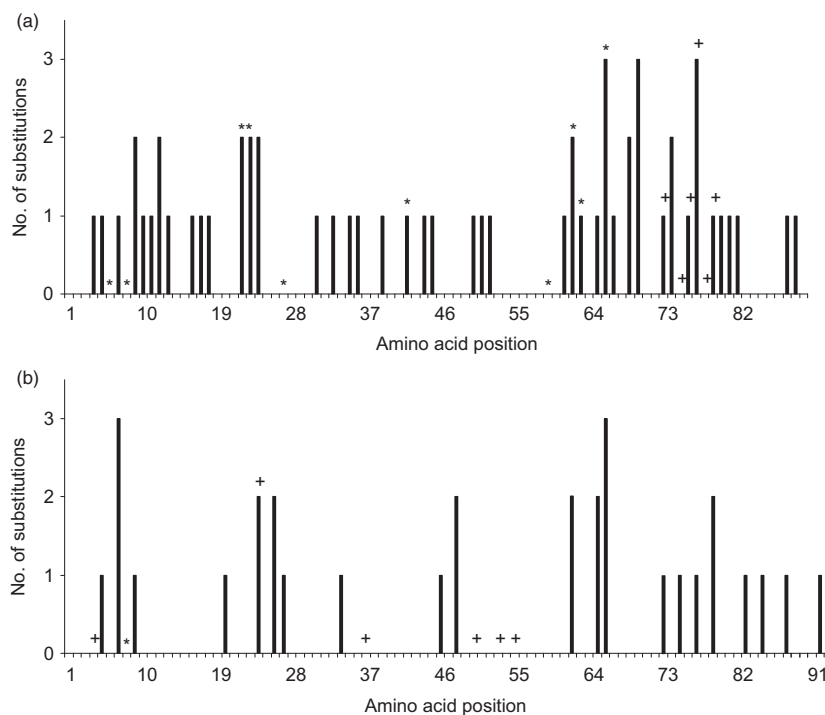

Figure 4. Amino acid substitutions of Caja-G exon 2 (a) and exon 3 (b) encoded protein sequences. Comparison between all 34 detected Caja-G alleles. Positions that form the B antigen-binding pocket are marked (*) and positions that form the F antigen-binding pocket are marked (+). Positions that form the B antigen-binding pocket that are marked (**) indicates that the amino acids at position 22&23 are crucial for peptide-binding.

Table 3. Comparison of Risler index amino acid substitutions *of Caja-G*08:01, Caja-G*08:23, Caja-G*07:01:01, Caja-G*18:04*

Peptide-binding pocket	AA position	AA exchange	Risler index
B pocket	38	*Tyr − Phe*	4
	58	*Asn − Ser*	7
	78	*Arg − Glu*	8
	79	*Gln − Glu*	2
	82	*Ile − Arg*	20
F pocket	89	*Ile − Thr*	16
	92	*Val − Gly*	52
	93	*Asp − Asn*	36
	130	*Asn − Arg*	24

Amino acids (AA) are presented in three letter code. AA position identifies position in the 852 bp fragment of exon 1–4.

For identification of amino acid variability, the Wu–Kabat plot was used with all known Caja-G sequences (Fig. 5). The average Wu–Kabat score for all sites is 1·84. There are 20 highly polymorphic sites (Wu–Kabat score > 4) in the whole sequence: two of those are located in the B pocket (codon position 38 and 82), and in the F pocket two positions are highly variable (codon 93 and 130), whereas other residues have limited polymorphism or are completely conserved. The highest variability can be observed at codon position 130 (Wu–Kabat score 13·036). For the highly polymorphic sites of the antigen-binding positions that were identified, the average Wu–Kabat score is 6·83 for the B pocket, respectively 9·33 for the F pocket.

Patterns of nucleotide substitutions

The frequency of replacement (non-synonymous) and silent (synonymous) nucleotide substitutions of Caja-G and HLA-G sequences was determined by the method of Nei and Gojobori (Table 4).[23] In a previous study of Zemmour and Parham analysis of HLA-A, -B and -C alleles, replacement substitutions were shown to be more likely to occur on amino acid residues forming the antigen recognition site. For the 34 analysed Caja-G sequences a similar conclusion can be made, considering

that the d_N value is 1·6 times greater than the d_S value when calculated for the antigen recognition site. In contrast, for the non-classsical HLA-G, in accordance with former studies,[30] no priority of non-synonymous substitutions was observed, which indicates the low variety of HLA-G in peptide-binding positions. A d_N/d_S ratio > 1 as seen in the classical MHC class I and Caja-G indicates positive Darwinian selection acting to increase the diversity of the peptide-binding region. At non-functional positions, the distribution of synonymous and non-synonymous substitutions for Caja-G is similar to those observed in the classical MHC class I sequences. All the d_S values are greater than the d_N values in the α1 and α2 remainder and the α3 regions.

Additionally, the selective force at single amino acid residues in the Caja-G was detected using the datamonkey SLAC analysis tool.[25] Of the 54 antigen recognition sites, two amino acids at positions 95* and 171*, were inferred as positively selected and one residue at position 24** was concluded as negatively selected (significance: *P < 0·05, **P < 0·01). Of the remaining 220 positions not coding for peptide-binding positions, one position was detected as positive selected (40) and 13 were inferred as negatively selected sites (9**, 16**, 37, 110*, 127**, 129**, 132, 134**, 197**, 216*, 251*, 279**, 280). In comparison to studies on the HLA genes (HLA-A, HLA-B and HLA-C), the number of positively selected sites in the antigen recognition site is much lower but the predominance of negatively selected sites in the remaining amino acid positions is similar.[31]

Discussion

In the studied group of marmosets, 30 new Caja-G alleles were detected with a high amount of variant alleles from different phylogenetic lineages, revealing high diversity at the nucleotide as well as the amino acid level. As a result of naturally occurring bone marrow chimerism, the presence of up to 10 functional Caja-G alleles and the similar appearance of some allele combinations indicate that the Caja-G locus may have been subjected to expansion, which corresponds with recently published data.[27,29] To

Table 4. Mean number of nucleotide substitutions per 100 synonymous (d_S ± SEM) and per 100 non-synonymous sites (d_N ± SEM)

	Antigen recognition site (*n* = 54)		α1−α2 (remainder) (*n* = 128)		α3 (*n* = 92)	
	d_N	d_S	d_N	d_S	d_N	d_S
HLA-A (31)*	12·5 ± 1·6	3·8 ± 1·1	1·3 ± 0·4	3·3 ± 1·0	1·5 ± 0·5	6·6 ± 1·9
HLA-B (31)*	15·1 ± 1·6	4·8 ± 2·0	1·8 ± 0·4	4·8 ± 1·2	0·2 ± 0·2	2·1 ± 1·1
HLA-C (21)*	6·9 ± 1·1	2·3 ± 1·4	1·8 ± 0·4	5·2 ± 1·2	1·2 ± 0·4	3·4 ± 1·3
Caja-G (34)	10·5 ± 1·2	6·5 ± 1·0	1·7 ± 0·2	6·4 ± 1·0	2·3 ± 1·0	5·2 ± 0·9
HLA-G (26)	0·0 ± 0·0	0·02 ± 0·0	0·2 ± 0·2	1·4 ± 1·4	0·1 ± 0·1	0·4 ± 0·2

n = Number of codons compared. d values are estimated using Nei and Gojobori method.[23]
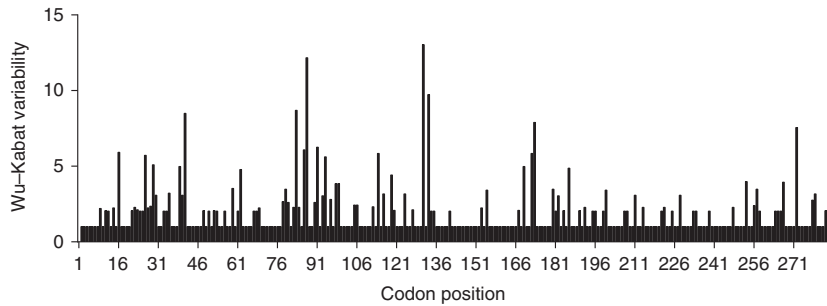*The analysis of Zemmour and Parham.[24]

Figure 5. Wu–Kabat plot of amino acid variability from a 852 bp fragment of Caja-G exon 1–4 ($n = 74$).

limit the risk of cross-contamination, samples were taken at different time-points and by the use of sterile single-use syringes. The group of Cao et al. even detected up to 11 different alleles in a single common marmoset, which indicates that the variety of different alleles per animal may not be due to contamination but rather to the natural genotype.[28] Duplication of the MHC class I Caja-B segment has already been observed, supporting the possibility of locus duplication, although further studies are required for verification. Contrary to previous publications,[27–29] the Caja-G alleles *Caja-G*07:01:01*, *Caja-G*08:01*, *Caja-G*08:23* and *Caja-G*18:04* were present in the majority of animals, seeming to have priority over others because of their high occurrence in individual unrelated animals. It can be speculated that 40 years of breeding that started from a relatively small group of breeding pairs in Europe might propagate such a phenomenon as major alleles. However, in contrast to this hypothesis, the same high-frequency alleles were detected in embryonic stem cells obtained from the Central Institute of Experimental Animals in Kawasaki, Japan (data not shown). Furthermore, due to the low number of sequenced cDNA clones it is also possible that some alleles were missed or that high-frequency alleles have true immunological advantages in the breeding group. Further investigation with more animals from different origins is necessary to make more precise assertions about the allelic distribution.

Analysis of phylogenetic relationships of classical and non-classical human MHC class I alleles as well as pseudogenes with the Caja-G revealed locus-specific clustering and a closer relationship to the non-classical HLA-G. In contrast, analysis of other platyrrhinid species on MHC class I loci did not show specific assembly but a single cluster composed of HLA-A, -G, -H, -K and -J.[32] While HLA-G, -K and -J have the same ancestral proto-G gene and can be found in both Old World and New World monkeys, HLA-A and HLA-H, which also seem to share a common ancestral gene, have not been found in New World monkeys.[33] As a result, a phylogenetic relationship between Caja-G and HLA-A seems to be unlikely and, in contrast, other members of the Callitrichidae, such as cotton-top tamarins, also show locus-specific clustering to HLA-G. To exclude convergent evolution leading to a homologous relationship, van der Wiel et al. analysed intron sequences of Caja-G and HLA genes, confirming the analysis based on exon sequences.[27]

Whereas only limited variability of the HLA-G has been observed,[13,34,35] in contrast to classical MHC class I genes, the number of recently discovered variant Caja-G alleles from different lineages is still growing.[18,27–29] Even though the phylogenetic relationship is closer to non-classical MHC class I, it can be speculated that the Caja-G might have taken over the immunological function of a classical MHC class I molecule. This is supported by analysis of cotton-top tamarin MHC-G, showing high polymorphism at T-cell receptor, natural killer cell receptor and antigen-binding sites. Due to the close evolutionary relationship of marmosets and tamarins, it is likely that the MHC-G function in this branch of New World monkeys is similar to classical HLA class I antigen-presenting molecules.

Comparison of antigen-specific binding pockets B and F also indicates strong polymorphism leading to conformational variability and, therefore, different specificity of antigen presentation, increasing the number of parasite antigens to which T cells can respond.[9] Because of the chimeric character of the common marmoset, the widespread immune response may be beneficial for viral defence, but could potentially lead to higher risk of autoimmune diseases and problems in pregnancies. As observed in a previous publication, a limited MHC class I polymorphism may not be important in the establishment of bone marrow chimerism, but a limited variation of the MHC class I molecules.[8] To get an idea of the biochemical variability of the four predominant Caja-G alleles, B and F pocket amino acid substitutions were ranked using Risler's amino acid substitution matrix.[36] The polymorphic residues of the MHC class I molecules determine the size and stereochemistry of the different binding pockets leading to individual interaction between an allele and a peptide.[37] Analysis showed that the substitutions might have more influence on the F pocket formation than on the B pocket. The F pocket is crucial for binding the C-terminus of antigen peptides, whereas the B pocket determines the peptide-binding sequence specificity. Wu–Kabat analysis of all known Caja-G alleles also revealed highly variable residues in B and F pocket positions.

Although both binding pockets concern two highly polymorphic sites, the F pocket amino acid positions show higher average variability. Testing the ratio of non-synonymous compared with synonymous amino acid substitutions also revealed higher amounts of coding substitutions influencing the antigen recognition site of the Caja-G alleles and increasing the variety of peptide-binding positions. The high d_N/d_S value also indicates that replacement substitutions are not randomly focused on the antigen recognition site and are providing evidence for selection for the polymorphism. As shown in Table 4, the non-classical HLA-G showed no obvious variability, characterized by amino acid substitutions and comparable to the low substitution levels of other non-classical HLA-E and HLA-F genes.[38] Additionally, two positively selected sites could be inferred in the Caja-G alleles. Compared with classical MHC class I genes,[31] the total amount of positive selected sites influencing the peptide-binding region is much lower in the Caja-G but the amount of investigated sequences is still limited and more positions may be significantly identified by the detection of more variant alleles in future studies. For the remaining amino acids, non-coding for the antigen recognition site, a high number of negatively selected positions was documented, similar to previous studies.[31]

Hence, it appears that the MHC-G locus of the common marmoset shares features with the classical MHC class I loci, as human classical HLA-A, HLA-B and HLA-C genes show high variability and diversity in their peptide-binding regions.[39] Despite the phylogenetic relationship of Caja-G and HLA-G alleles, the highly variable positions at the antigen recognition site, supported by Wu–Kabat plot, d_N/d_S ratio and Risler index, together with the diversity of different detected Caja-G lineages, advocates the idea of a gene conversion.

Conclusion

In this study, we described 34 different MHC class I Caja-G alleles, in which 30 alleles were novel. New alleles were uploaded to the NCBI database and renamed by the MHC NHP Immuno Polymorphism Database. The presence of one to ten different Caja-G alleles per animal suggests that the Caja-G loci duplicated in the genome of the marmosets. Phylogenetic analysis showed closer relationship to the non-classical HLA-G than to classical MHC loci, whereas a high degree of polymorphism and variability was detected, which is unusual for non-classical MHC loci, leading to the idea of a gene conversion of the Caja-G from a non-classical to a classical MHC molecule.

Acknowledgements

Author contributions

Anna-Lena Neehus performed molecular genetic studies, implemented bioinformatic analysis and drafted the manuscript. Joachim Wistuba prepared and provided blood and tissue samples of the common marmosets for the study. Nektarios Ladas assisted in bioinformatic analysis and helped to revise the manuscript. Britta Eiz-Vesper and Stefan Schlatt designed and coordinated the study. Thomas Müller designed and coordinated the study, and participated in data analysis and manuscript writing.

Disclosures

The authors have declared no conflicting interests.

References

1 Adams AP, Aronson JF, Tardif SD, Patterson JL, Paige A, Adams JF et al. Common marmosets (*Callithrix jacchus*) as a nonhuman primate model to assess the virulence of eastern equine encephalitis virus strains. *J Virol* 2008; **82**:9035–42.

2 Wiedemann A, Hemmer K, Beremann I, Göhring G, Pogozhykh O, Figueiredo C et al. Induced pluripotent stem cells generated from adult bone marrow-derived cells of the nonhuman primate (*Callithrix jacchus*) using a novel quad-cistronic and excisable lentiviral vector. *Cell Reprogram* 2012; **14**:485–96.

3 Pogozhyk O, Pogozhyk D, Neehus AL, Hoffmann A, Blasczyk R, Müller T. Molecular and cellular characteristics of human and non-human primate multipotent stromal cells from the amnion and bone marrow during long term culture. *Stem Cell Res Ther* 2015; **6**:150.

4 Müller T, Fleischmann G, Eildermann K, Mätz-Rensing K, Horn P, Sasaki E et al. A novel embryonic stem cell line derived from the common marmoset monkey (*Callithrix jacchus*) exhibiting germ cell-like characteristics. *Hum Reprod* 2009; **6**:1359–72.

5 Albert S, Ehmcke J, Wistuba J, Eildermann K, Behr R, Schlatt S et al. Germ cell dynamics in the testis of the postnatal common marmoset monkey (*Callithrix jacchus*). *Reproduction* 2010; **140**:733–42.

6 Dell'Mour V, Range F, Huber L. Social learning and mother's behaviour in manipulative tasks in infant marmosets. *Am J Primatol* 2009; **71**:503–9.

7 Benirschke K, Anderson JM, Brownhill LE. Marrow chimerism in marmosets. *Science* 1962; **138**:513–5.

8 Watkins DI, Letvin NL. Limited MHC class I polymorphism is not essential for bone marrow chimerism in New World primates. *Immunogenetics* 1991; **3**:194–7.

9 Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science* 1996; **272**:67–74.

10 Diehl M, Münz C, Keilholz W, Stevanovic S, Holmes N, Loke YW et al. Nonclassical HLA-G molecules are classical peptide presenters. *Curr Biol* 1996; **6**:305–14.

11 Garstka MA, Fish A, Celie PHN, Joosten RP, Janssen GMC, Berlin I et al. The first step of peptide selection in antigen presentation by MHC class I molecules. *Proc Natl Acad Sci USA* 2015; **112**:1505–10.

12 Trowsdale J. The MHC, disease and selection. *Immunol Lett* 2011; **137**:1–8.

13 Ishitani A, Kishida M, Sageshima N, Yashiki S, Sonoda S, Hayami M et al. Re-examination of HLA-G polymorphism in African Americans. *Immunogenetics* 1999; **49**:808–11.

14 Kovats S, Main EK, Librach C, Stubblebine M, Fisher SJ, DeMars R. A class I antigen, HLA-G, expressed in human trophoblasts. *Science* 1990; **248**:220–3.

15 Parham P, Norman PJ, Abi-Rached L, Hilton HG, Guethlein LA. Review: immunogenetics of human placentation. *Placenta* 2012; **33**:S71–80.

16 Shawar S, Vyas JM, Rodgers JR, Rich RR. Antigen presentation by major histocompatibility complex class I-B molecules. *Annu Rev Immunol* 1994; **12**:839–80.

17 Rölleke U, Flügge G, Plehm S, Schlumbohm C, Armstrong VW, Dressel R *et al.* Differential expression of major histocompatibility complex class I molecules in the brain of a New World monkey, the common marmoset (*Callithrix jacchus*). *J Neuroimmunol* 2006; **176**:39–50.

18 Kono A, Brameier M, Roos C, Suzuki S, Shigenari A, Kametani Y *et al.* Genomic sequence analysis of the MHC class I G/F segment in common marmoset (*Callithrix jacchus*). *J Immunol* 2014; **192**:3236–46.

19 de Groot NG, Otting N, Robinson J, Blancher A, Lafont BAP, Marsh SGE *et al.* Nomenclature report on the major histocompatibility complex genes and alleles of great ape, old and new world monkey species. *Immunogenetics* 2012; **64**:615–31.

20 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 2013; **30**:2725–9.

21 Risler JL, Delorme MO, Delacroix H, Henaut A. Amino acid substitutions in structurally related proteins a pattern recognition approach: determination of a new and efficient scoring matrix. *J Mol Biol* 1988; **204**:1019–29.

22 Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* 1970; **132**:211–50.

23 Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986; **3**:418–26.

24 Zemmour J, Parham P. Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. *J Exp Med* 1992; **176**:937–50.

25 Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 2005; **22**:1208–22.

26 Cadavid LF, Shufflebotham C, Ruiz FJ, Yeager M, Hughes AL, Watkins DI. Evolutionary instability of the major histocompatibility complex class I loci in New World primates. *Proc Natl Acad Sci USA* 1997; **94**:14536–41.

27 van der Wiel MK, Otting N, de Groot NG, Doxiadis GGM, Bontrop RE. The repertoire of MHC class I genes in the common marmoset: evidence for functional plasticity. *Immunogenetics* 2013; **65**:841–9.

28 Cao YH, Fan JW, Li AX, Liu HF, Li LR, Zhang CL *et al.* Identification of MHC I class genes in two Platyrrhini species. *Am J Primatol* 2015; **77**:527–34.

29 Li T, Xu Y, Yin S, Liu B, Zhu S, Wang W *et al.* Characterization of major histocompatibility complex class I allele polymorphisms in common marmosets. *Tissue Antigens* 2014; **84**:568–73.

30 Suarez CF, Cardenas PP, Llanos-Ballestas EJ, Martinez P, Obregon M, Patarroyo ME *et al.* α1 and α2 domains of Aotus MHC class I and Catarrhini MHC class Ia share similar characteristics. *Tissue Antigens* 2003; **61**:362–73.

31 Suzuki Y, Gojobori T. A method for detecting positive selection at single amino acid sites. *Mol Biol Evol* 1999; **16**:1315–28.

32 Sawai H, Kawamoto Y, Takahata N, Satta Y. Evolutionary relationships of major histocompatibility complex class I genes in simian primates. *Genetics* 2004; **166**:1897–907.

33 Adams EJ, Parham P. Species-specific evolution of MHC class I genes in the higher primates. *Immunol Rev* 2001; **183**:41–64.

34 Yamashita T, Fujii T, Watanabe Y, Tokunaga K, Tadokoro K, Juji T *et al.* HLA-G gene polymorphism in a Japanese population. *Immunogenetics* 1996; **44**:186–91.

35 Suarez MB, Morales P, Castro MJ, Fernandez V, Varela P, Alvarez M *et al.* A new HLA-G allele (*HLAG *0105N*) and its distribution in the Spanish population. *Immunogenetics* 1997; **45**:464–5.

36 Elsner HA, Blasczyk R. Sequence similarity matching: proposal of a structure-based rating system for bone marrow transplantation. *Eur J Immunogenet* 2002; **29**:229–36.

37 Zeng ZH, Castaño AR, Segelke BW, Stura EA, Peterson PA, Wilson IA. Crystal structure of mouse CD1: an MHC-like fold with a large hydrophobic binding groove. *Science* 1997; **277**:339–45.

38 Knapp LA, Cadavid LF, Watkins DI. The MHC-E locus is the most well conserved of all known primate class I histocompatibility genes. *J Immunol* 1998; **160**:189–96.

39 Bjorkman PJ, Parham P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev Biochem* 1990; **59**:253–88.