

## Research

### Corresponding author

Suraksha Agrawal, PhD

Professor

Department of Medical Genetics  
Sanjay Gandhi Post Graduate Institute  
of Medical Sciences (SGPGIMS)  
Raebareli Road, Lucknow  
UP 226014, India

Tel. 091-522-2668004-8

Fax: 091-522-26680973/6680017

E-mail: [sur\\_ksha\\_agrawal@yahoo.co.in](mailto:sur_ksha_agrawal@yahoo.co.in)

Volume 2 : Issue 1

Article Ref. #: 1000ANTPOJ2106

### Article History

Received: September 13<sup>th</sup>, 2016

Accepted: September 29<sup>th</sup>, 2016

Published: October 5<sup>th</sup>, 2016

### Citation

Prakash S, Misra MK, Agrawal S. Non-classical human leukocyte antigen-G allelic diversity among North Indians. *Anthropol Open J.* 2016; 2(1): 1-9. doi: [10.17140/ANTPOJ-2-106](https://doi.org/10.17140/ANTPOJ-2-106)

# Non-Classical Human Leukocyte Antigen-G Allelic Diversity Among North Indians

Swayam Prakash, PhD<sup>1</sup>; Maneesh Kumar Misra, PhD<sup>2</sup>; Suraksha Agrawal, PhD<sup>2\*</sup>

<sup>1</sup>Department of Nephrology, Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow, UP 226014, India

<sup>2</sup>Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow, UP 226014, India

## ABSTRACT

**Objectives:** Throughout evolution high degree of polymorphism of human leukocyte antigens (HLA) has been observed which ensured accommodation and presentation of all possible antigens. Non-classical human leukocyte antigens-G (HLA-G) region shows polymorphism at 5'-upstream regulatory region (5'-URR) and in the 3'-untranslated region (3'-UTR). These polymorphisms have not been explored extensively in spite of the fact that they have been associated with immune-modulation. This study explores the HLA-G allele diversity and their plausible role in reshaping the global population structure.

**Methods:** Three hundred North Indians were screened using HLA-G 5'URR region by sequencing. The 14 bp insertion deletion polymorphism was evaluated by polymerase chain reaction (PCR) followed by polyacrylamide gel electrophoresis (PAGE). Detailed statistical analysis was done to compare the present population with allele frequency data of 12 global populations. The number of populations was restricted as only sequence based typing based studies were considered.

**Results:** The genetic distance, phylogenetic analysis and principal component analysis revealed genetic similarity of North Indians with populations of West Eurasian descent from HLA-G allelic diversity view point. Mongoloid, African, and South American populations revealed less genetic similarity with North Indians.

**Conclusions:** Indian subcontinent has undergone many historical migrations and invasions hence it is intriguing to observe how the conserved HLA-G has behaved from the population diversity view point throughout these years. Our results revealed less genetic distance with Caucasians suggesting an evolutionary link among them.

**KEYWORDS:** Gene diversity; Human leukocyte antigen-G; North India; Phylogenetic analysis.

**ABBREVIATIONS:** HLA: Human Leukocyte Antigens; 5'URR: 5'-upstream regulatory region; HLA-G: Human Leukocyte Antigens G; 3'-UTR: 3'-untranslated region; PCR: Polymerase Chain Reaction; PAGE: Polyacrylamide gel electrophoresis; IMGT/HLA: International ImmunoGeneTics project/human major histocompatibility complex; SNPs: Single Nucleotide Polymorphisms; IEC: Institutional Ethical Committee; EDTA: Ethylenediaminetetraacetic acid; HWE: Hardy Weinberg Equilibrium; DNA: Deoxyribonucleic acid; PCA: Principal Component Analysis; mRNA: messenger RNA.

## INTRODUCTION

India being a land rich of ethnic diversity with more than 2,000 ethnic groups draws attention of anthropologists and population geneticists while studying the pattern of human migration. India showcases extensive social, cultural, linguistic, and biological variation. About 60,000 years ago the South Asian peninsula served as an important corridor for the dispersal of human beings from Africa. During the course of migrations people of India must have undergone differential selective pressures. Such phenomena might reshape the existing population.

## Copyright

©2016 Agrawal S. This is an open access article distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Over the period of evolution, high degree of polymorphism has been observed for human leukocyte antigens (HLA). The HLA play a key role in modulating immune responses. Classical HLA antigens have been studied in different Indian populations for clinical and diversity perspective. However, non-classical class-I antigens have not been explored much in Indian subcontinent.

Human leukocyte antigens G (HLA-G) belongs to non-classical HLA class I antigens. It shows low allelic variation, a restricted peptide repertoire,<sup>1</sup> an unusually high diversity of molecular structures due to alternative splicing of the primary transcript,<sup>2</sup> and a restricted expression under physiological conditions, which can be up-regulated in various situations. HLA-G was originally discovered on the extravillous cytotrophoblast at the maternal-fetal interface.<sup>3</sup> It has 7 different isoforms that have been created due to alternative splicing. Four of them are membrane-bound (HLA-G1, -G2, -G3, and -G4) while other 3 are found in secreted form (HLA-G5, -G6, and -G7). The secreted isoforms lack transmembrane region. HLA-G binds to inhibitory receptors such as leukocyte immunoglobulin-like receptors B1 (LILRB1/ILT2) and LILRB2 (ILT4), which are expressed on immune cells, to suppress a broad range of immune responses. Expression of HLA-G in placenta provides immune tolerance to the fetus. There are many studies showing its association with renal transplant rejection,<sup>4</sup> auto-immune diseases,<sup>5</sup> diabetes<sup>6</sup> and many other conditions including different malignancies.<sup>7</sup>

As of now, 52 HLA-G alleles/haplotypes are officially recognized by the International ImMunoGeneTics Project/human major histocompatibility complex (IMGT/HLA) database (release 3.23.0). Most of the single nucleotide polymorphisms (SNPs) in the HLA-G coding region are either coding synonymous mutations or intronic variants. Out of 52 alleles which have been officially recognized by IMGT/HLA, only 13 alleles encode 4 different HLA-G full-length molecules and are frequently observed all over the world.<sup>1-3,8,9</sup>

HLA-G shows less diversity but still plays an important role in immune tolerance. The relatively less polymorphism observed in the HLA-G coding region both in humans and apes may represent a strong purifying selective pressure. Studying the diversity among HLA-G may be used to evaluate the impact of expansions in human genome as it is one of the most conserved regions of the non-classical antigens. HLA-G revealed high degree of nucleotide variability in the coding region which causes conformational changes, modifies interaction with cell receptors, modulates immune response and isoform production. All these functions in turn may affect the evolutionary mechanisms. A recent study conducted on African population revealed that natural selection has shaped genetic diversity at 5'-UTR.<sup>9</sup> The present study aims to evaluate how different ethnic groups/populations are genetically linked to each other from HLA-G allelic diversity view point. This in turn will address whether the diversity in the HLA-G alleles play any role in the human migration and selection. The study findings will serve as a reference for future studies.

## MATERIAL AND METHODS

### Study Subjects and Samples

Blood samples of 300 unrelated healthy individuals representing fairly homogeneous ethnic populations residing in North Indian province of Uttar Pradesh, India were randomly selected. The study population comprised of upper caste Hindus. The study was performed after the approval of the Institutional Ethical Committee (IEC) of Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow, UP, India. All individuals were of Indo-Aryan linguistic family and were at least third generation residents of Uttar Pradesh, India. Three generation pedigrees were drawn to ensure no population admixture. Informed consent was obtained from all the individual participants included in the study prior to collection of 5 mL blood in ethylenediaminetetraacetic acid (EDTA) by venipuncture.

The HLA-G frequency data was collected for 30 populations from the allele frequencies.net database<sup>10</sup> and also from published literature. However, only 12 studies were included where the alleles were determined by sequencing. The frequency data so collected represented 2, 3 and 4 allele nomenclatures of HLA-G. Alleles representing all the 3 nomenclatures were considered in-order to satisfy the Hardy Weinberg Equilibrium (HWE). The reason to include different nomenclature was that, when any 1 of the 3 HLA-G nomenclatures was considered the data was not found to be in accordance with the HWE.

### DNA Extraction

Blood samples were collected and processed for deoxyribonucleic acid (DNA) isolation by QiaAmp DNA blood kit (Qiagen; West Sussex, UK).

### HLA-G Sequencing

HLA-G regulatory region alleles were detected using sequencing. Sequencing reaction was performed using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit. Purified products of sequencing reactions were run on electrophoresis on an ABI 3130 DNA Sequencer (Applied Biosystem, Foster City, CA, USA). Allele assignment was accomplished using the DNASTAR v. 11.0. The PCR followed by PAGE was used for genotyping of 14 bp insertion/deletion polymorphism of exon-8 in 3'-UTR).

### Defining HLA-G Allele

Each HLA-G allele name has a unique number like classical HLA nomenclature corresponding up to 4 sets of digits separated by colons. The digits before the first colon describe the type, by serology. The next sets of digits are used to list the subtypes, numbers being assigned in the order in which DNA sequences have been determined. Alleles whose numbers differ in the 2 sets of digits must differ in one or more nucleotide substitutions that

change the amino acid sequence of the encoded protein. Alleles that differ only by synonymous nucleotide substitutions within the coding sequence are distinguished by the use of the 3<sup>rd</sup> set of digits. Alleles that only differ by sequence polymorphisms in the introns, or in the 5' or 3' un-translated regions that flank the exons and introns, are distinguished by the use of the 4<sup>th</sup> set of digits. Alleles that have been shown not to be expressed - 'Null' alleles - have been given the suffix 'N'.<sup>11</sup> The alleles found in the present study were defined on the basis of 5'URR SNPs (-1306 G>A, -1179 G>A, -1155 G>A, -1140 A>T, -1138 A>G, -1121 C>T, -964 G>A, -762 C>T, -725 C>G/T, -716 T>G, -689 A>G, -666G>T, -633 G>A, -486 A>C, -477 G>C, -369 A>C, -201 G>A, and -56 C>T) and 14 bp I/D polymorphism (Table 1).

**Data Analysis and Statistical Methods**

Frequency of each HLA-G allele was determined by direct counting. Phylogenetic tree based on HLA-G allele frequencies was constructed by neighbor joining method<sup>12</sup> using MEGA (version 5.2.2, [www.megasoftware.net](http://www.megasoftware.net)).<sup>13</sup> The principal component analysis (PCA) of HLA-G allele frequencies were carried out using XLSTAT. Further, we have evaluated genetic diversity using SMOGD<sup>14</sup> to observe the evolutionary relationship of the studied populations with North Indians.

**RESULTS**

**HLA-G Allele Frequency**

Eight HLA-G alleles with minor allele frequency greater than one were observed in the present study. Individual allele frequencies and sample size for each population is also mentioned in Table 2. All the 13 different populations selected, representing different ethnicities were in HWE. The populations considered for finding the global landscape of HLA-G alleles were from Brazil (Parana ethnicity),<sup>15</sup> Sao Paulo,<sup>16</sup> Germany,<sup>17</sup> Ghana,<sup>18</sup> Poland,<sup>19</sup> South Korea,<sup>20</sup> Spain,<sup>17</sup> USA African Ameri-

can,<sup>21</sup> Zambia,<sup>21</sup> Iraq,<sup>22</sup> USA Hutterite,<sup>23</sup> and Denmark<sup>24</sup> along with the present study. The alleles found in these populations were G\*01:01:01, G\*01:01:03, G\*01:01:04, G\*01:01:05, G\*01:01:06, G\*01:01:08, G\*01:04:03, G\*01:04:04, G\*01:01, G\*01:03, and G\*01:04.

**GENETIC DISTANCE**

North Indians showed the least genetic distance from all the West Eurasian (otherwise known as Caucasian) populations (Table 3). The West Eurasian populations included in this study were from Germany, Spain, Poland, and a set of heterogeneous West Eurasian population dataset from different parts of the USA. The Nei's genetic distance revealed least genetic distance between North Indian and German (0.15) populations. Meanwhile, population belonging to Mongoloid lineage of (Korean (0.75)), and middle-east lineage (Iraq (0.65)) revealed maximum genetic distance with North Indian data.

**Phylogenetic Assessment of Studied Population**

Phylogenetic analysis showed 3 major ethnicities namely West Eurasians, Africans, and South Americans under 3 clusters (Figure 1). It was evident from the figure that North Indians were in close proximity with West Eurasians (Poland, Germany, Spain, Denmark). USA Hutterite population was found to be placed on a common node with Denmark. African populations from Ghana and Zambia along with USA-African-Americans were placed in a separate cluster. Iraq representing the Middle East and South Koreans belonging to Mongoloid ethnicity was placed on single nodes in the phylogenetic tree.

**Principal Component Analysis**

Principal component analysis (PCA) revealed that HLA-G alleles in different populations were in accordance with their geographical locations (Figure 2). Distinct geographical clusters

SNP → Haplotype ↓	-1306	-1179	-1155	-1140	-1138	-1121	-964	-762	-725	-716	-689	-666	-633	-486	-477	-369	-201	-56	14bp I/D
G*01:01:01:01	G	A	G	A	A	C	G	C	C	T	A	G	G	A	C	C	G	C	Deletion
G*01:01:01:02	G	A	G	A	A	C	G	C	G	T	A	G	G	A	C	C	G	C	Insertion
G*01:01:01:03	G	A	G	A	A	T	G	C	G	T	A	G	G	A	C	C	G	C	Insertion
G*01:01:01:04	G	A	G	A	A	C	G	C	C	T	A	G	G	A	C	C	G	C	-
G*01:01:02	A	G	G	T	A	C	A	T	C	G	G	T	A	C	G	A	A	C	Deletion
G*01:03:01:02	G	G	G	A	G	C	G	C	T	T	A	G	G	A	G	A	G	T	Insertion
G*01:04:01	A	G	A	A	A	C	A	T	C	G	G	T	A	C	G	A	A	C	Deletion
G*01:06	A	G	G	T	A	C	A	T	C	G	G	T	A	C	G	A	A	C	Insertion
G*01:05N	A	G	G	T	A	C	A	T	C	G	G	T	A	C	G	A	A	C	Insertion

**Table 1:** HLA-G allele associated with un-translated region (UTR) haplotype of 5'-URR and 14 bp insertion-deletion polymorphism in exon-8 of 3'-UTR.

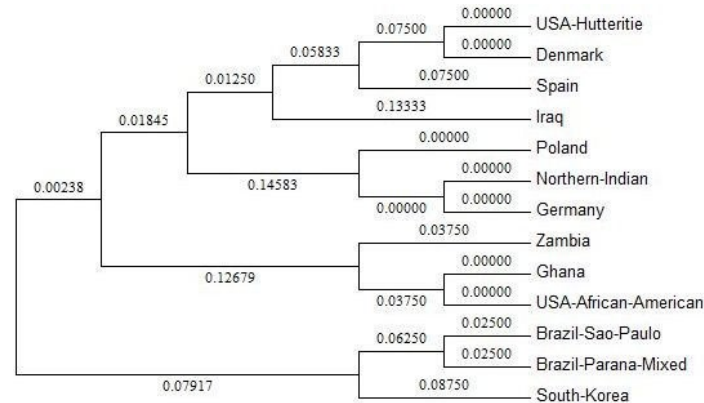
Population → Alleles ↓	Northern Indian (n=300)	Brazil Parana Mixed (n=52)	Brazil Sao Paulo (n=103)	Germany (n=132)	Ghana (n=42)	Poland (n=58)	South Korea (n=486)	Spain (n=114)	Zambia (n=256)	Denmark (n=150)	USA African American (n=42)	USA Hutteritie (n=80)	Iraq (n=50)
G*01:01:01	0.00	0.29	0.40	0.00	0.83	0.53	0.39	0.51	0.41	0.00	0.70	0.00	0.13
G*01:01:02	0.10	0.24	0.20	0.00	0.02	0.21	0.07	0.25	0.21	0.00	0.06	0.00	0.25
G*01:01:03	0.00	0.02	0.05	0.00	0.00	0.07	0.04	0.07	0.00	0.00	0.02	0.00	0.00
G*01:01:04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.00	0.00
G*01:01:05	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
G*01:01:06	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G*01:01:08	0.00	0.18	0.04	0.00	0.00	0.06	0.11	0.00	0.00	0.00	0.00	0.00	0.00
G*01:04:01	0.19	0.12	0.08	0.00	0.00	0.04	0.24	0.00	0.00	0.00	0.00	0.00	0.00
G*01:04:03	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00
G*01:04:04	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G*01:05N	0.07	0.01	0.01	0.02	0.05	0.02	0.02	0.03	0.00	0.00	0.08	0.00	0.09
G*01:01	0.00	0.00	0.00	0.87	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.80	0.20
G*01:03	0.00	0.14	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.43	0.00	0.03	0.05
G*01:04	0.00	0.00	0.00	0.08	0.10	0.00	0.00	0.11	0.00	0.00	0.13	0.13	0.17
G*01:06	0.07	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.11
G*01:01:01:01	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G*01:01:01:03	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G*01:01:01:02	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G*01:01:01:04	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G*01:03:01:02	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

n=Number of individuals against each population.

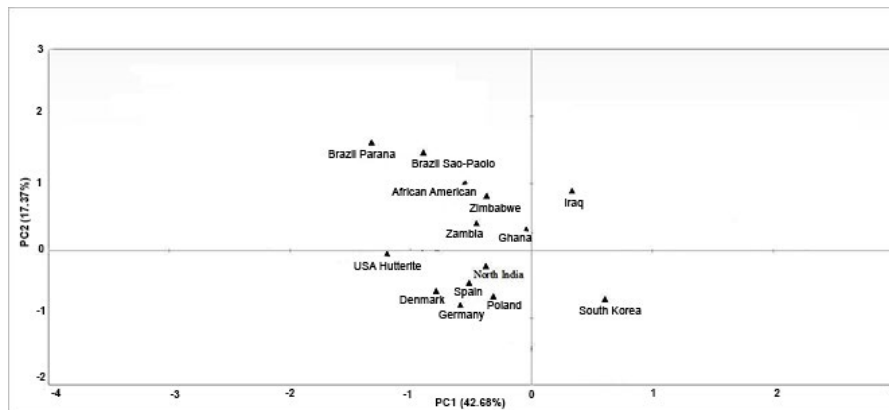
Table 2: Incidence of HLA-G alleles among northern Indians and global populations.

	Northern Indian	Zambia	Spain	Germany	Ghana	South Korea	Poland	USA African American	Brazil Parana Mixed	Iraq	Brazil Sao Paulo	USA Hutteritie	Denmark
Northern Indian		0.5000	0.2500	0.1500	0.4500	0.7500	0.2000	0.5000	0.5000	0.6500	0.5000	0.4000	0.3500
Zambia	0.5000		0.2000	0.3500	0.2500	0.1500	0.1500	0.2000	0.3000	0.3500	0.2000	0.4000	0.3000
Spain	0.2500	0.2000		0.5500	0.4500	0.1500	0.3500	0.4000	0.4000	0.5500	0.4000	0.6000	0.4000
Germany	0.1500	0.3500	0.5500		0.2000	0.5000	0.5000	0.2500	0.3500	0.1000	0.2500	0.0500	0.1500
Ghana	0.4500	0.2500	0.4500	0.2000		0.3000	0.3000	0.0500	0.1500	0.3000	0.0500	0.2500	0.1500
South Korea	0.7500	0.1500	0.1500	0.5000	0.3000		0.2000	0.2500	0.3500	0.5000	0.2500	0.5500	0.3500
Poland	0.2000	0.1500	0.3500	0.5000	0.3000	0.2000		0.2500	0.3500	0.5000	0.2500	0.5500	0.4500
USA African American	0.5000	0.2000	0.4000	0.2500	0.0500	0.2500	0.2500		0.2000	0.3500	0.0000	0.3000	0.2000
Brazil Parana Mixed	0.5000	0.3000	0.4000	0.3500	0.1500	0.3500	0.3500	0.2000		0.2500	0.2000	0.3000	0.3000
Iraq	0.6500	0.3500	0.5500	0.1000	0.3000	0.5000	0.5000	0.3500	0.2500		0.3500	0.0500	0.2500
Brazil Sao Paulo	0.5000	0.2000	0.4000	0.2500	0.0500	0.2500	0.2500	0.0000	0.2000	0.3500		0.3000	0.2000
USA Hutteritie	0.4000	0.4000	0.6000	0.0500	0.2500	0.5500	0.5500	0.3000	0.3000	0.0500	0.3000		0.2000
Denmark	0.3500	0.3000	0.4000	0.1500	0.1500	0.3500	0.4500	0.2000	0.3000	0.2500	0.2000	0.2000	

Table 3: Genetic Distance among studied populations.



**Figure 1:** Evolutionary relationship among thirteen populations on the basis of HLA-G allele diversity  
The evolutionary history was inferred using the Neighbor Joining method. The optimal tree with the sum of branch length = 2.06396995 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary analyses were conducted in MEGA5.



**Figure 2:** Principal component analysis  
Four quartiles show distribution of thirteen populations on the basis of HLA-G allele frequencies.

were seen in the PC plot which includes West Eurasian (Caucasian), Asian, African and Mongoloid populations. This supported the findings of genetic distance and phylogenetic analysis. Middle Eastern population from Iraq was placed in the 1<sup>st</sup> quartile. The 2<sup>nd</sup> quartile represented populations of mainland Africa namely Ghana, Zambia, and African American along with South American populations from Brazil (Sao-Paulo, Parana). The North Indian population huddled with the West Eurasian (Spain, Poland, Germany, Denmark, USA Hutterite) placed in the 3<sup>rd</sup> quartile. Population belonging to the Mongoloid ethnicity (South Korean) fell separately in the fourth quartile.

**Genetic Diversity**

Measurement of genetic diversity was calculated on the basis of HLA-G allele frequency data. Total heterozygosity ( $H_T$ ), absolute differentiation ( $D_{ST}$ ), relative differentiation ( $G_{ST}$ ), heterozygosity among subpopulation ( $H_{ST}$ ), effective number of distinct subpopulations ( $\Delta_{ST}$ ), actual differentiation ( $D$ ), and proportion of total diversity found in the average subpopulation ( $\Delta_S/\Delta_T$ ), and nearly unbiased estimator of total subpopulation heterozygosity ( $H_{T\_est}$ ) (Table 4) were calculated. Higher total hetero-

zygosity ( $H_T$ ) values were observed for for HLA-G\*01:01:02 (0.43), G\*01:01:03 (0.50) and G\*01:01:08 (0.47) alleles. These results suggest no change in relative differentiation ( $G_{ST}$ ). The component of diversity ( $\Delta_{ST}$ ) between subpopulation was observed to be minimum (1.1) for G\*01:01:01:03, G\*01:01:01:02, G\*01:01:01:04, G\*01:03:01:02, and G\*01:04:04. Further, the proportion of total diversity present in the total subpopulation ( $\Delta_S/\Delta_T$ ) was maximum (0.9) for HLA-G alleles like G\*01:01:01:02, G\*01:01:01:04, G\*01:03:01:02, and G\*01:04:04. The actual differentiation ( $D$ ) for different populations was maximum (0.53) for G\*01:01:03, and G\*01:04:01 alleles.

**Multidimensional Scaling (MDS) Analysis**

The primary objective of the multidimensional scaling (MDS) analysis was to obtain the best fit with the smallest number of possible dimensions. For this purpose we have used the scree plot (Figure not shown). Stress was observed at 6<sup>th</sup> dimension with the Iraq population. Subsequently major stress was observed with Brazil Parana (2<sup>nd</sup> dimension) and Brazil Sao Paulo (3<sup>rd</sup> dimension) which revealed higher genetic distances with North Indians. We observed declined stress with increased di-



Locus ID	N	N	H <sub>T</sub>	D <sub>ST</sub>	G <sub>ST</sub>	H <sub>ST</sub>	Δ <sub>ST</sub>	D	Δ <sub>S</sub> /Δ <sub>T</sub>
G*01:01:01	1	13	0.43	0.43	1	0.43	1.76	0.46	0.57
G*01:01:02	2	13	0.43	0.43	1	0.43	1.76	0.46	0.57
G*01:01:03	3	13	0.50	0.50	1	0.50	1.99	0.53	0.50
G*01:01:04	4	13	0.19	0.19	1	0.19	1.23	0.20	0.81
G*01:01:05	5	13	0.19	0.19	1	0.19	1.23	0.20	0.81
G*01:01:06	6	13	0.19	0.19	1	0.19	1.23	0.20	0.81
G*01:01:08	7	13	0.47	0.47	1	0.47	1.87	0.49	0.53
G*01:04:01	8	13	0.50	0.50	1	0.50	1.99	0.53	0.50
G*01:04:03	9	13	0.33	0.33	1	0.33	1.50	0.35	0.67
G*01:04:04	10	13	0.10	0.10	1	0.10	1.11	0.11	0.90
G*01:05N	11	13	0.43	0.43	1	0.43	1.76	0.46	0.57
G*01:01	12	13	0.43	0.43	1	0.43	1.76	0.46	0.57
G*01:03	13	13	0.47	0.47	1	0.47	1.87	0.49	0.81
G*01:04	14	13	0.47	0.47	1	0.47	1.87	0.49	0.53
G*01:06	15	13	0.43	0.43	1	0.43	1.76	0.46	0.81
G*01:01:01:01	16	13	0.27	0.27	1	0.27	1.36	0.28	0.81
G*01:01:01:03	17	13	0.10	0.10	1	0.10	1.11	0.11	0.81
G*01:01:01:02	18	13	0.10	0.10	1	0.10	1.11	0.11	0.90
G*01:01:01:04	19	13	0.10	0.10	1	0.10	1.11	0.11	0.90
G*01:03:01:02	20	13	0.10	0.10	1	0.10	1.11	0.11	0.90

N=number of populations, HT=total heterozygosity, DST=absolute differentiation, GST=relative differentiation, HST=between-subpopulation heterozygosity, ΔST=between-subpopulation component of diversity, or the effective number of distinct subpopulations, D=actual differentiation, ΔS/ΔT=proportion of total diversity that is contained in the average subpopulation.

**Table 4:** Measurement of genetic diversity.

mensionality showing genetic diversity among studied populations for the HLA-G alleles.

## DISCUSSION

The HLA-G locus seen in 6 different HLA-G lineages revealed functional variations prominently in nucleotides of the regulatory regions. Differences were observed at the 5' URR in positions that either coincide with or are close to transcription factor-binding sites<sup>23,25</sup> which may influence the messenger RNA (mRNA) stability. It has been reported earlier that KIR2DL4 acts as a receptor for HLA-G revealing the importance of fetomaternal tolerance.<sup>26</sup> KIR diversity has been well documented for Indian populations.<sup>27-29</sup> In this context there exists a plausibility of high degree of differential selective pressure due to Killer-cell immunoglobulin-like receptor (KIR) and HLA-G interactions leading to population diversity among different populations. The present study has been conducted to correlate the genetic diversity of North Indians with twelve other populations from different continents to provide a global landscape.

Various recombination events in the HLA-G\*01:01 allele have resulted into four variants namely HLA-G\*01:03, \*01:04, \*01:05N, and \*01:06.<sup>16</sup> All these variants have also been

observed among North Indians. The distribution of the HLA-G is not significantly different among populations with respect to HLA-G\*01:01 group. HLA-G\*01:01:01 is the predominant allele with a frequency of 60% in Europe, 50% North America, 40% in South America and Africa, 30% in Asia, and 16% among North Indians. Incidence of HLA-G\*01:01:02 was high, constituting approximately 50% in European populations, 30% in North America, 20% in South America and Africa, 15% in Asia, and nearly 11% in the present study. HLA-G\*01:01:08 is a rare allele in most of the populations with an exception of Brazilian populations (8% and 15%), Poland (5%).<sup>19</sup> We have not observed this allele among North Indians. However, the HLA-G\*01:04:01 allele is observed with a high frequency in Asian (Korea), North Indians, and African populations while it is low in Brazil and Caucasian population. The null allele HLA G\*01:05N is detected with high frequency in Spanish (6%), African-Americans (7%), Iraqi (8%), and North Indians (8%).<sup>17,22,30</sup> This shows concordance of our results with other genetically similar populations.

However, the frequency of the null allele among Caucasian from Europe and South America is low. In Brazil the prevalence of the null allele varies from 1% to 3%, while in Europe from 1% to 6%. In China and in some populations of USA,

the null allele is rare. In Peru, Guatemala, and Japan the null allele is completely absent.<sup>21,31</sup> These reports suggest that the null mutation might have arisen recently in the human population. Thus, it is striking that HLA-G\*01:05N allele is maintained in a substantial frequency in certain human populations. Further studies in worldwide populations show that the highest HLA-G\*01:05 allele frequencies are found in Middle East. This may suggest that either this area was close to the “null” allele origin and some evolutionary pressures maintained the high frequency in the Middle East.

The population comparisons confirmed that most of the HLA-G variability has originated before human dispersion from Africa and that the allele and haplotype frequencies have probably been shaped by strong selective pressures. In addition to alignments between human and other primates coding MHC-G sequences, analysis of HLA-G non-coding regions have proved to be highly informative about the evolutionary history of this gene. Interestingly the primate family Cercopithecinae was assumed to be distantly related to human ancestors and also found to possess the MHC-G “null” allele. This suggests that the null allele may confer some advantage either at maternal/fetal tolerance or at other immune mechanisms involved in the HLA-G functions. Considering the role of HLA-G in the placenta and the selective pressures operating at this stage of development, a new allele could be easily fixed in a population.<sup>21</sup> One previous report suggests that from Africa these might have been introduced into the Spanish population by Arab invaders and into African American population with the slave deportation.<sup>21</sup>

The genetic landscape of Indian populations is highly diverse as shown by previous studies based upon mitochondrial<sup>32</sup> and Y chromosomal DNA.<sup>33</sup> However, looking at the pathological relevance of HLA-G, it is evident that HLA-G plays an important role in the regulation of immune response and immune modulation in spite of low degree of genetic diversity. From evolution view point rare variants are much younger to common variants e.g., alleles found in 1 out of 1,000 individuals have an average age of 22 generations, or roughly 550 years.<sup>34</sup> Interestingly common variants give an insight into gene flow patterns over several millennia, while rare variants aid in investigating patterns on the timescale for centuries. More studies on less polymorphic non classical HLA-G antigens will allow us to study population movements that will enhance the knowledge of human variation, genetic drift and more importantly the selective pressures due to its immense role in the immune surveillance.

#### AUTHOR CONTRIBUTIONS

Dr. Swayam Prakash performed the experiments, and wrote the manuscript. Dr. Maneesh Misra performed experiments. Prof. Suraksha Agrawal conceived the idea, supervised the experiments and gave critical input in preparation of the final manuscript.

#### ACKNOWLEDGEMENTS

The authors are grateful to the Department of Science and Technology, Government of India, New Delhi for the financial support to carry out this work.

#### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

#### REFERENCES

- Hviid TV, Rizzo R, Melchiorri L, Stignani M, Baricordi OR. 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.* 2006; 67: 53-62.
- Yari F, Zavaran Hosseini A, Nemat Gorgani M, Khorramizadeh MR, Mansouri P, Kazemnejad A. Expression of HLA-G in the skin of patients with pemphigus vulgaris. *Iran J Allergy Asthma Immunol.* 2008; 7: 7-12. doi: [07.01/ijaa.712](https://doi.org/10.7171/ijaa.712)
- Hviid TV, Larsen LG, Hoegh AM, Bzorek M. HLA-G expression in placenta in relation to HLA-G genotype and polymorphisms. *Am J Reprod Immunol.* 2004; 52: 212-217. doi: [10.1111/j.1600-0897.2004.00208.x](https://doi.org/10.1111/j.1600-0897.2004.00208.x)
- Aghdaie MH, Azarpira N, Kazemi K, Geramizadeh B, Darai M, Malekhoseini SA. Frequency of HLA-G exon 8 polymorphisms and kidney allograft outcome in Iranian population. *Mol Biol Rep.* 2011; 38: 3593-3597. doi: [10.1007/s11033-010-0470-y](https://doi.org/10.1007/s11033-010-0470-y)
- Mariaselvam CM, Chaaben AB, Salah S, et al. Human leukocyte antigen-G polymorphism influences the age of onset and autoantibody status in rheumatoid arthritis. *Tissue Antigens.* 2015; 85(3): 182-189. doi: [10.1111/tan.12521](https://doi.org/10.1111/tan.12521)
- Silva HP, Ururahy MA, Souza KS, et al. The association between the HLA-G 14-bp insertion/deletion polymorphism and type 1 diabetes. *Genes Immun.* 2016; 17(1): 13-18. doi: [10.1038/gene.2015.45](https://doi.org/10.1038/gene.2015.45)
- Alves BM, Prellwitz IM, Siqueira JD, et al. The effect of human leukocyte antigen G alleles on human papillomavirus infection and persistence in a cohort of HIV-positive pregnant women from Brazil. *Infect Genet Evol.* 2015; 34: 339-343. doi: [10.1016/j.meegid.2015.06.027](https://doi.org/10.1016/j.meegid.2015.06.027)
- Hyllenius S, Andersen AM, Melbye M, Hviid TV. Association between HLA-G genotype and risk of pre-eclampsia: A case-control study using family triads. *Mol Hum Reprod.* 2004; 10: 237-246. doi: [10.1093/molehr/gah035](https://doi.org/10.1093/molehr/gah035)
- Gineau L, Luisi P, Castelli EC, et al. Balancing immunity and tolerance: Genetic footprint of natural selection in the transcrip-

- tional regulatory region of HLA-G. *Genes Immun.* 2015; 16: 57-70. doi: [10.1038/gene.2014.63](https://doi.org/10.1038/gene.2014.63)
10. González-Galarza FF, Takeshita LY, Santos EJ, et al. Allele frequency net 2015 update: New features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res.* 2015; 43(database issue): D784-D788. doi: [10.1093/nar/gku1166](https://doi.org/10.1093/nar/gku1166)
11. Robinson J, Halliwell JA, Hayhurst JH, Flicek P, Parham P, Marsh SGE. The IPD and IPD-IMGT/HLA database: Allele variant databases. *Nucleic Acids Res.* 2015; 43: D423-D431. doi: [10.1093/nar/gku1161](https://doi.org/10.1093/nar/gku1161)
12. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4: 406-425. Web site: <http://mbe.oxfordjournals.org/content/4/4/406.full.pdf+html>. Accessed September 12, 2016.
13. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28: 2731-2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121)
14. Crawford NG. SMOGD: Software for the measurement of genetic diversity. *Mol Ecol Resour.* 2010; 10: 556-557. doi: [10.1111/j.1755-0998.2009.02801.x](https://doi.org/10.1111/j.1755-0998.2009.02801.x)
15. Pirri A, Contieri FC, Benvenuti R, Bicalho Mda G. A study of HLA-G polymorphism and linkage disequilibrium in renal transplant patients and their donors. *Transpl Immunol.* 2009; 20: 143-149. doi: [10.1016/j.trim.2008.09.012](https://doi.org/10.1016/j.trim.2008.09.012)
16. Castelli EC, Mendes-Junior CT, Donadi EA. HLA-G alleles and HLA-G 14 bp polymorphisms in a Brazilian population. *Tissue Antigens.* 2007; 70: 62-68. doi: [10.1111/j.1399-0039.2007.00855.x](https://doi.org/10.1111/j.1399-0039.2007.00855.x)
17. Pfeiffer KA, Fimmers R, Engels G, van der Ven H, van der Ven K. The HLA-G genotype is potentially associated with idiopathic recurrent spontaneous abortion. *Mol Hum Reprod.* 2001; 7: 373-378. doi: [10.1093/molehr/7.4.373](https://doi.org/10.1093/molehr/7.4.373)
18. Ishitani A, Kishida M, Sageshima N, et al. Re-examination of HLA-G polymorphism in African Americans. *Immunogenetics.* 1999; 49(9): 808-811. doi: [10.1007/s002510050555](https://doi.org/10.1007/s002510050555)
19. Sipak-Szmigiel O, Cybulski C, Lubiński J, Ronin-Walknowska E. HLA-G polymorphism in a Polish population and reproductive failure. *Tissue Antigens.* 2008; 71: 67-71. doi: [10.1111/j.1399-0039.2007.00942.x](https://doi.org/10.1111/j.1399-0039.2007.00942.x)
20. Park Y, Park Y, Kim YS, Kwon OJ, Kim HS. Allele frequencies of human leukocyte antigen-G in a Korean population. *Int J Immunogenet.* 2012; 39: 39-45. doi: [10.1111/j.1744-313X.2011.01053.x](https://doi.org/10.1111/j.1744-313X.2011.01053.x)
21. Shao W, Yoo YJ, Tang J, Zhang K, Allen S, Kaslow RA. Tight linkage disequilibrium between HLA-G and HLA-A alleles in native Africans. *Hum Immunol.* 2006; 67(1): S118. doi: [10.1016/j.humimm.2006.08.180](https://doi.org/10.1016/j.humimm.2006.08.180)
22. Jassem RM, Shani WS, Loisel DA, Sharief M, Billstrand C, Ober C. HLA-G polymorphisms and soluble HLA-G protein levels in women with recurrent pregnancy loss from Basrah province in Iraq. *Hum Immunol.* 2012; 73: 811-817. doi: [10.1016/j.humimm.2012.05.009](https://doi.org/10.1016/j.humimm.2012.05.009)
23. Ober C, Rosinsky B, Grimsley C, van der Ven K, Robertson A, Runge A. Population genetic studies of HLA-G: Allele frequencies and linkage disequilibrium with HLA-A1. *J Reprod Immunol.* 1996; 32: 111-123. doi: [10.1016/S0165-0378\(96\)01000-5](https://doi.org/10.1016/S0165-0378(96)01000-5)
24. Steffensen R, Christiansen OB, Bennett EP, Jersild C. HLA-E polymorphism in patients with recurrent spontaneous abortion. *Tissue Antigens.* 1998; 52(6): 569-572. doi: [10.1111/j.1399-0039.1998.tb03088.x](https://doi.org/10.1111/j.1399-0039.1998.tb03088.x)
25. Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, Roger M, Moreau P, Donadi EA. A comprehensive study of polymorphic sites along the HLA-G gene: Implication for gene regulation and evolution. *Mol Biol Evol.* 2011; 28: 3069-3086. doi: [10.1093/molbev/msr138](https://doi.org/10.1093/molbev/msr138)
26. Rajagopalan S, Long EO. KIR2DL4 (CD158d): An activation receptor for HLA-G. *Front Immunol.* 2012; 20(3): 258. doi: [10.3389/fimmu.2012.00258](https://doi.org/10.3389/fimmu.2012.00258)
27. Rajalingam R, Du Z, Meenagh A, et al. Distinct diversity of KIR genes in three southern Indian populations: Comparison with world populations revealed a link between KIR gene content and pre-historic human migrations. *Immunogenetics.* 2008; 60: 207-217. doi: [10.1007/s00251-008-0286-2](https://doi.org/10.1007/s00251-008-0286-2)
28. Rajalingam R, Krausa P, Shilling HG, et al. Distinctive KIR and HLA diversity in a panel of North Indian Hindus. *Immunogenetics.* 2002; 53: 1009-1019. doi: [10.1007/s00251-001-0425-5](https://doi.org/10.1007/s00251-001-0425-5)
29. Prakash S, Alam S, Faridi RM, Sonawane A, Agrawal S. Killer immunoglobulin like receptor gene content diversity among Northern Indian population. *Ann Hum Biol.* 2014; 41: 540-547. doi: [10.3109/03014460.2014.899624](https://doi.org/10.3109/03014460.2014.899624)
30. Matte C, Lacaille J, Zijenah L, Ward B, Roger M. HLA-G and HLA-E polymorphisms in an indigenous African population. The ZVITAMBO Study Group. *Hum Immunol.* 2000; 61: 1150-1156.
31. Arnaiz-Villena A, Enriquez-de-Salamanca M, Areces C, et al. HLA-G(\*01:05N null allele in Mayans (Guatemala) and



Uros (Titikaka Lake, Peru): Evolution and population genetics. *Hum Immunol.* 2013; 74(4): 478-482. doi: [10.1016/j.humimm.2012.12.013](https://doi.org/10.1016/j.humimm.2012.12.013)

32. Rowold DJ, Benedico DP, Garcia-Bertrand R, et al. Ladakh, India: The land of high passes and genetic heterogeneity reveals a confluence of migrations. *Eur J Hum Genet.* 2016; 24: 442-449. doi: [10.1038/ejhg.2015.80](https://doi.org/10.1038/ejhg.2015.80)

33. Negi N, Tamang R, Pande V, et al. The paternal ancestry of Uttarakhand does not imitate the classical caste system of India. *J Hum Genet.* 2016; 61: 167-172. doi: [10.1038/jhg.2015.121](https://doi.org/10.1038/jhg.2015.121)

34. Novembre J, Ramachandran S. Perspectives on human population structure at the cusp of the sequencing era. *Annu Rev Genomics Hum Genet.* 2011; 12: 245-274. doi: [10.1146/annurev-genom-090810-183123](https://doi.org/10.1146/annurev-genom-090810-183123)