**Short Communication** doi: 10.1111/j.1744-313X.2007.00698.x

# HLA analysis of Sri Lankan Sinhalese predicts North Indian origin

**G. N. Malavige,*‡ T. Rostron,* S. L. Seneviratne,† S. Fernando,‡ S. Sivayogan,§ A. Wijewickrama¶ & G. S. Ogg*,**

## Summary

The origin of the Sinhalese population of Sri Lanka is debated. We subtyped HLA-A*02 in 101 Sinhalese and observed a preponderance of the rare allele HLA-A*0211 which was similar to reported frequencies in northern India. Taken with low-resolution typing for the remaining A, B, C, DR and DQ alleles, these data suggest a North Indian origin for the Sri Lankan Sinhalese.

The distribution and the frequency of the human leucocyte antigen (HLA) alleles differ significantly among human populations (Probst *et al*., 2000; Cao *et al*., 2001). Of the different HLA-A alleles, HLA-A*02 is the most heterogeneous allele in the family with over 56 different alleles (Shankarkumar *et al*., 2003). HLA-A2 is frequent in all ethnic groups with a frequency of 50% in Caucasians, 34.6% in African-Americans, 36% in Asians/pacific islanders and 46.9% among the Hispanic population (Ellis *et al*., 2000). However, the HLA-A2 subtypes vary widely among populations, with HLA-A*0201 being thought to be the predominant Western Caucasoid allele, A*0205 is predominantly African while A*0206 and A*0207 occur predominantly in those of South-East Asian origin (Ellis *et al*., 2000). Recently, a relatively rare A2 subtype, HLA-A*0211, was identified among North Indian Asians (Mehra *et al*., 2001).

The study was carried out in the Colombo district, Sri Lanka in which 76.4% are of Sinhalese ethnic origin (Nanayakkara, 2001). One hundred and one EDTA blood samples were collected from unrelated healthy Sinhalese individuals following informed written consent. Ethical clearance for the study was obtained from the Ethical Review Committee of the University of Sri Jayawardanapura, Sri Lanka.

* MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK, † Department of Clinical Immunology, Manchester Royal Infirmary, Manchester, UK, ‡ Department of Microbiology, § Department of Community Medicine, Faculty of Medical Sciences, University of Sri Jayawardanapura, Nugegoda, Sri Lanka, ¶ Infectious Diseases Hospital, Colombo, Sri Lanka and ** Department of Dermatology, Churchill Hospital, Oxford, UK

Received 30 April 2007; revised 30 April 2007; accepted 17 May 2007

Correspondence: Graham Ogg, Cutaneous Immunology Group, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. Tel: +44 186 522 2334; Fax: +44 186 522 2502; E-mail: graham.ogg@clinical-medicine.oxford.ac.uk

HLA class I and class II alleles were typed as previously described (Bunce *et al*., 1995). Briefly, DNA was extracted from EDTA blood using Gentra PureGene Kit (D5000) (Gentra, Minneapolis, MN, USA). Extracted DNA was amplified and DNA typing was undertaken using a 144 sequence-specific primer (SSP) reactions to simultaneously detect all known HLA-A, -B, -C, DRB1, DRB3, DRB4, DRB5 and DQB1 specificities in an allele-specific or group-specific manner using the same method, reagents, polymerase chain reaction (PCR) parameters and protocols for all loci. The PCR products were electrophoresed in 1.0% agarose gels, using xylene cyanol FF and bromophenol blue as marker dyes. The gels were run for 15 min at 15 V/cm in the 0.5 × TBE buffer and visualized using UV illumination. A phototype was considered to be successful when the control amplifications are positive and at least one allele or group was present in each locus. Allele frequencies were estimated from the number of positive typing reactions divided by the total number of haplotypes tested. Differences in gene frequencies, assuming Hardy–Weinberg equilibrium, were calculated using the $\chi^2$ test (GraphPad Prism package, GraphPad Software Inc., San Diego, CA, USA).

We undertook high-resolution typing for the HLA-A*02 subtypes (Table 1). Five A*02 subtypes were identified, with the most common A*02 subtype being A*0211, which was the subtype seen in 40.5% of A*02 individuals. A*0201, which is usually found among Caucasians, was seen in 27% of the A*02 individuals.

Although not part of the original hypothesis, we also undertook low-resolution typing of the remaining HLA-A, -B, -C, DRB and DQB1 loci. Thirteen HLA-A alleles were identified with the predominant alleles being A*24 (20.8%), A*02 (18.3%) and A*33 (17.8%). Twenty-one HLA-B alleles were identified, with B*40 (12.4%), B*35 (11.4%) and B*44 (10.3%) being the most common. Twelve HLA-C alleles were identified among our population, the most common alleles being Cw*07 (25%), Cw*06 (14.4%) and Cw*04 (12.9%). Fourteen DRB1 alleles were identified, with DRB1*15 (21.3%) being the most common DRB1 allele (Table 2). The other common DRB1 alleles were DRB1*07 (20.3%) and DRB1*13 (14.4%). Among the DRB alleles, DRB3 was the most common (39.3%), followed by DRB4 (35.7%) and DRB5 (24%). Only five DQB1 alleles were identified, with DQB1*06 being the most common (30.8%) closely followed by DQB1*05 (29.8%).

 

**Table 1.** Gene frequency of HLA-A, B and C alleles in the Sinhalese

| HLA-A | *N* (F) | HLA-B | *N* (F) | HLA-C | *N* (F) |
|---|---|---|---|---|---|
| 02 | 37 (0.183) | 07 | 17 (0.084) | 01 | 14 (0.069) |
| 0201 | 10 (0.049) | 08 | 3 (0.015) | 02 | 2 (0.009) |
| 0203 | 5 (0.024) | 13 | 5 (0.025) | 03 | 24 (0.119) |
| 0206 | 5 (0.024) | 15 | 18 (0.089) | 04 | 26 (0.129) |
| 0211 | 15 (0.074) | 18 | 3 (0.015) | 05 | 3 (0.015) |
| 0216 | 2 (0.009) | 27 | 2 (0.009) | 06 | 29 (0.144) |
| | | 35 | 23 (0.114) | 07 | 51 (0.25) |
| 0101 | 22 (0.109) | 37 | 5 (0.025) | 08 | 6 (0.029) |
| 03 | 7 (0.035) | 38 | 2 (0.009) | 12 | 15 (0.074) |
| 1101 | 21 (0.104) | 39 | 3 (0.015) | 14 | 7 (0.035) |
| 24 | 42 (0.208) | 40 | 25 (0.124) | 15 | 20 (0.099) |
| 26 | 11 (0.054) | 41 | 1 (0.005) | 16 | 5 (0.025) |
| 29 | 1 (0.005) | 44 | 21 (0.104) | | |
| 30 | 4 (0.019) | 48 | 1 (0.005) | | |
| 31 | 7 (0.035) | 49 | 2 (0.009) | | |
| 33 | 36 (0.178) | 51 | 13 (0.064) | | |
| 68 | 14 (0.069) | 52 | 10 (0.049) | | |
| | | 55 | 13 (0.064) | | |
| | | 57 | 19 (0.094) | | |
| | | 58 | 15 (0.074) | | |
| | | 68 | 1 (0.005) | | |

* Frequency (F %) is expressed as a percentage and is defined as the number of times the allele is present (*N*), divided by the total number of haplotypes in the study panel (2*n*, where *n* is the number of individuals in the panel).

**Table 2.** Gene frequency of HLA-DRB1 and DQB1 alleles in the Sinhalese

| HLA-DRB1 | *N* (F) | HLA-DQB1 | *N* (F) |
|---|---|---|---|
| *01 | 12 (0.059) | *02 | 36 (0.182) |
| *03 | 10 (0.049) | *03 | 38 (0.192) |
| *04 | 14 (0.069) | *04 | 2 (0.010) |
| *07 | 41 (0.203) | *05 | 59 (0.298) |
| *08 | 1 (0.005) | *06 | 61 (0.308) |
| *09 | 1 (0.005) | | |
| *10 | 15 (0.074) | | |
| *11 | 5 (0.025) | | |
| *12 | 6 (0.030) | | |
| *13 | 29 (0.144) | | |
| *14 | 19 (0.094) | | |
| *15 | 43 (0.213) | | |
| *16 | 2 (0.009) | | |

* Frequency (F %) is expressed as a percentage and is defined as the number of times the allele is present (*N*), divided by the total number of haplotypes in the study panel (2*n*, where *n* is the number of individuals in the panel).

We have characterized the HLA-A*02 subtype diversity of 101 unrelated Sinhalese in the Colombo district. We identified a preponderance of HLA-A*0211 allele which is relatively rare in Caucasian populations but has a very high frequency among the North Indian population and indeed is thought to be a novel allele among the North Indians (Mehra *et al*., 2001; Rajalingam *et al*., 2002). Moreover, the overall frequency of the A*0211 subtype (7.4%) in the Sinhalese was not significantly different to that of the North Indian population (6.7%) (Mehra *et al*., 2001). Apart from North India and North-west India (Shankarkumar *et al*., 2002), the A*0211 subtype is also seen in the Bulgarian gypsy population (Naumova, 2002), Ecuador (Trachtenberg *et al*., 1995) and among the Pakistani Sindhi, Brahui and Brusho ethnic groups at relatively lower frequency (Mohyuddin *et al*., 2002). Interestingly, A*0207 and A*0206 that are common A*02 subtypes in South-East Asia (Chandanayingyong *et al*., 1994; Middleton *et al*., 2000) and A*205 and A*202 that are common A2 subtypes seen in the African population were rarely detected in the current study.

The other most common HLA-A allele, A*33 was seen in a similar frequency in the Namboorthi and the Ezhava ethnic groups in Kerala and also not significantly different to that of Maharastra of North-western India (Shankarkumar *et al*., 2002; Thomas *et al*., 2006). However, when comparing the HLA-B alleles, there were some striking differences among the ethnic groups in Kerala and the Sinhalese (Thomas *et al*., 2006). While there were significant differences ($P < 0.05$) in B*07 frequencies between the Sinhalese and South Indian non-tribal and Dravidian populations (Thomas *et al*., 2004, 2006), in contrast, we again found similarities in the frequencies of the predominant HLA-B alleles in the Sinhalese and the North Indian population (Rajalingam *et al*., 2002). Therefore, based on the A*02 subtypes, and HLA-B alleles, the Sinhalese population appears to have a similar genetic background to that of the North Indian population.

As with the differences in the HLA-A-*02 subtypes and B alleles, further differences were seen between the Sinhalese and other Dravidian ethnic groups with regard to other HLA-A alleles (Thomas *et al*., 2004, 2006). For instance, in Tamil Nadu, South India, HLA-A*32 was seen among 29.7% individuals and HLA-A*31 in 16.2% (Vettriselvi *et al*., 2006). However, HLA-A*32 was not seen and A*31 was seen at a very low frequency (7%) in the Sinhalese population.

Of the HLA-DRB1 alleles, DRB1*15 was the most common allele closely followed by DRB1*07 both of which are commonly seen worldwide (http://www.datatechniques.biz/allele/default1.asp). The frequency of DRB1*15 and DRB1*07 alleles was not significantly different to that seen in North Indians (Rajalingam *et al*., 2002; Rani *et al*., 2004). However, DRB1*13 which was present at a frequency of 14.4% in the Sinhalese was seen at a significantly ($P < 0.05$) lower frequency among the North Indians (Rajalingam *et al*., 2002; Rani *et al*., 2004). There are no published data addressing the frequency of HLA-DRB1 and HLA-DQB1 alleles in other Indian populations.

The populations that inhabited early Indian civilizations have been divided into the Aryans and the Dravidians. However, because of repeated invasions, the Dravidians had been driven to South India (Vettriselvi *et al*., 2006). Based on the limited data of HLA types of various ethnic groups in India and also because of the preponderance of

 

the HLA-A*0211 subtype among the Sinhalese population, it appears to be more likely that the Sinhalese have originated from the Aryans rather than the Dravidians. These findings are compatible with some historical data and similarities in the languages (Basa, 1992; Deraniyagala, 1992). The first reference to Sri Lanka was in the great Indian epic the 'Ramayana'. However, most of the history of Sri Lanka has been based on the 'Mahavansa' which is a chronicle composed by the Buddhist monks in Pali. According to this chronicle Vijaya is the central legendary figure and is a son of a North Indian princess. The history of the Sinhalese in Sri Lanka begins with the arrival of a group of 700 people of Indo-Aryan stock led by Prince Vijaya from the Indian mainland in 544 BC. Although, this is a well-enjoyed legendary story, it appears that the genetic data would indeed support such similarities between the Sinhalese and those of North India.

## Acknowledgements

We are grateful to Dr Kelapotha, staff of the Department of Microbiology, University of Sri Jayawardanapura, Sri Lanka for their assistance with the study and to all the donors for their contribution of blood samples. We are grateful to Dr Susan Tonks for critical review of the manuscript.

## References

Basa, K.K. (1992) Early glass beads in India. *South Asian Studies*, **8**, 91.

Bunce, M., O'Neill, C.M., Barnardo, M.C., Krausa, P., Browning, M.J., Morris, P.J. & Welsh, K.I. (1995) Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens*, **46**, 355.

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M. & Fernandez-Vina, M.A. (2001) Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Human Immunology*, **62**, 1009.

Chandanayingyong, D., Stephens, H.A., Fan, L., Sirikong, M., Longta, P., Vangseratthana, R., Lekmak, S., Longta, K., Bejrachandra, S. & Rungruang, E. (1994) HLA-DPB1 polymorphism in the Thais of Southeast Asia. *Human Immunology*, **40**, 20.

Deraniyagala, S.U. (1992) *The Prehistory of Sri Lanka, Pts. I, II*. Archaeological Survey Department, Colombo, Sri Lanka.

Ellis, J.M., Henson, V., Slack, R., Ng, J., Hartzman, R.J. & Katovich Hurley, C. (2000) Frequencies of HLA-A2 alleles in five US population groups. Predominance of A*02011 and identification of HLA-A*0231. *Human Immunology*, **61**, 334.

Mehra, N.K., Jaini, R., Rajalingam, R., Balamurugan, A. & Kaur, G. (2001) Molecular diversity of HLA-A*02 in Asian Indians: predominance of A*0211. *Tissue Antigens*, **57**, 502.

Middleton, D., Williams, F., Meenagh, A., Daar, A.S., Gorodezky, C., Hammond, M., Nascimento, E., Briceno, I. & Perez, M.P. (2000) Analysis of the distribution of HLA-A alleles in populations from five continents. *Human Immunology*, **61**, 1048.

Mohyuddin, A., Ayub, Q., Khaliq, S., Mansoor, A., Mazhar, K., Rehman, S. & Mehdi, S.Q. (2002) HLA polymorphism in six ethnic groups from Pakistan. *Tissue Antigens*, **59**, 492.

Nanayakkara, A.G.W. (2001) Population by ethnicity according to district and sector (provisional). *Census of Population and Housing 2001*. Department of Census and Statistics, Sri Lanka.

Naumova, E. (2002) 13th International Congress of Histocompatibility and Immunogenetics, Seattle, Washington, USA, 18 to 22 May 2002.

Probst, C.M., Bompeixe, E.P., Pereira, N.F., de, O. D. M. M., Visentainer, J. E., Tsuneto, L. T. & Petzl-Erler, M. L. (2000) HLA polymorphism and evaluation of European, African, and Amerindian contribution to the white and mulatto populations from Parana, Brazil. *Human Biology*, **72**, 597.

Rajalingam, R., Krausa, P., Shilling, H.G., Stein, J.B., Balamurugan, A., McGinnis, M.D., Cheng, N.W., Mehra, N.K. & Parham, P. (2002) Distinctive KIR and HLA diversity in a panel of north Indian Hindus. *Immunogenetics*, **53**, 1009.

Rani, R., Sood, A. & Goswami, R. (2004) Molecular basis of predisposition to develop type 1 diabetes mellitus in North Indians. *Tissue Antigens*, **64**, 145.

Shankarkumar, U., Ghosh, K. & Mohanty, D. (2002) Defining the allelic variants of HLA A19 in the western Indian population. *Human Immunology*, **63**, 779.

Shankarkumar, U., Prasanavar, D., Ghosh, K. & Mohanty, D. (2003) HLA A*02 allele frequencies and B haplotype associations in Western Indians. *Human Immunology*, **64**, 562.

Thomas, R., Nair, S.B. & Banerjee, M. (2004) HLA-B and HLA-C alleles and haplotypes in the Dravidian tribal populations of southern India. *Tissue Antigens*, **64**, 58.

Thomas, R., Nair, S.B. & Banerjee, M. (2006) A crypto-Dravidian origin for the nontribal communities of South India based on human leukocyte antigen class I diversity. *Tissue Antigens*, **68**, 225.

Trachtenberg, E.A., Erlich, H.A., Rickards, O., DeStefano, G.F. & Klitz, W. (1995) HLA class II linkage disequilibrium and haplotype evolution in the Cayapa Indians of Ecuador. *American Journal of Human Genetics*, **57**, 415.

Vettriselvi, V., Vijayalakshmi, K., Suganya, S., Krishnan, M., Paul, S.F. & Jayanth, V. (2006) Molecular diversity of HLA-A*19 group of alleles in south Indian population. *International Journal of Immunogenetics*, **33**, 69.

 