

HIGH RESOLUTION HLA-A, -B, -C AND -DRB1 ALLELE AND HAPLOTYPE FREQUENCIES IN THE COSTA RICA CENTRAL VALLEY POPULATION

E Arrieta-Bolaños^{1,2,3}, JA Madrigal^{1,2}, SGE Marsh^{1,2}, BE Shaw^{1,4}, L Salazar-Sánchez³

¹ Anthony Nolan Research Institute, Royal Free Hospital, Pond Street, London, UK, ² Cancer Institute, University College London, Royal Free Campus, London, UK, ³ Centro de Investigaciones en Hematología y Trastornos Afines (CIHATA), Universidad de Costa Rica, San José, Costa Rica, ⁴ Haemato-Oncology Research Unit, Division of Molecular Pathology, The Institute of Cancer Research, London, UK

Introduction

The Costa Rica Central Valley population (CCVP) is the major population in this country, accounting for over 60% of the Costa Rican inhabitants concentrated since colonial times in a 2,500 km² intermontane region. **Figure 1** shows its location and a detail of its major urban centers. Interesting historic, demographic and genetic characteristics of this hybrid population have attracted researchers interested in testing genetic associations for various diseases. However, no study describing Human Leukocyte Antigen (HLA) frequencies by molecular methods had been performed. We have recently described low resolution HLA allele group and haplotype frequencies in a sample of this population. In this report, we extend our study to high resolution by sequence-based typing of exons 2, 3 and 4 for HLA class I, and exon 2 for HLA-DRB1.



Figure 1 Diagram showing the location and detail of the CCVP within Costa Rica. The CCVP is the major population of this country and it concentrates most of the major Costa Rican cities in an intermontane region of approximately 90x25 km.

Materials and Methods

Peripheral blood or saliva samples from healthy unrelated volunteer donors were obtained by venipuncture or by collection of saliva using the ORAGENE-ONE collection kits (DNA Genotek Inc., Ottawa). In the case of blood samples, DNA was extracted by an in-house salting-out method. For the saliva samples, the manufacturer's extraction method was followed. All participants were born in the Costa Rican Central Valley and/or could trace their ancestry to this region, and signed an informed consent. A total of 205 samples from CCVP inhabitants, recruited as part of the University of Costa Rica's Centre for Research in Haematology and Related Disorders (CIHATA) DNA bank, were genotyped.

Generic amplification of the relevant genes was followed by exon-specific sequencing according to Anthony Nolan's in-house SBT protocol. Sequence data were acquired on an ABI 3730xL DNA Analyser. Sequence analysis was performed with Assign SBT software (Conexio Genomics, Freemantle, Australia) using IMGT/HLA database release 3.9.0. Alleles are noted at the highest possible resolution. Class I alleles that have identical sequences in exons 2, 3 and 4 are listed under the the first allele in the group.

Principal Coordinates Analysis (PCO) based on Euclidean distances calculated from 47 HLA-A and HLA-B allele group frequencies from 60 ancestral populations and the CCVP was carried out with the Multi-Variate Statistical Package (MVSP, Kovach Computing Services, Anglesey, Wales).

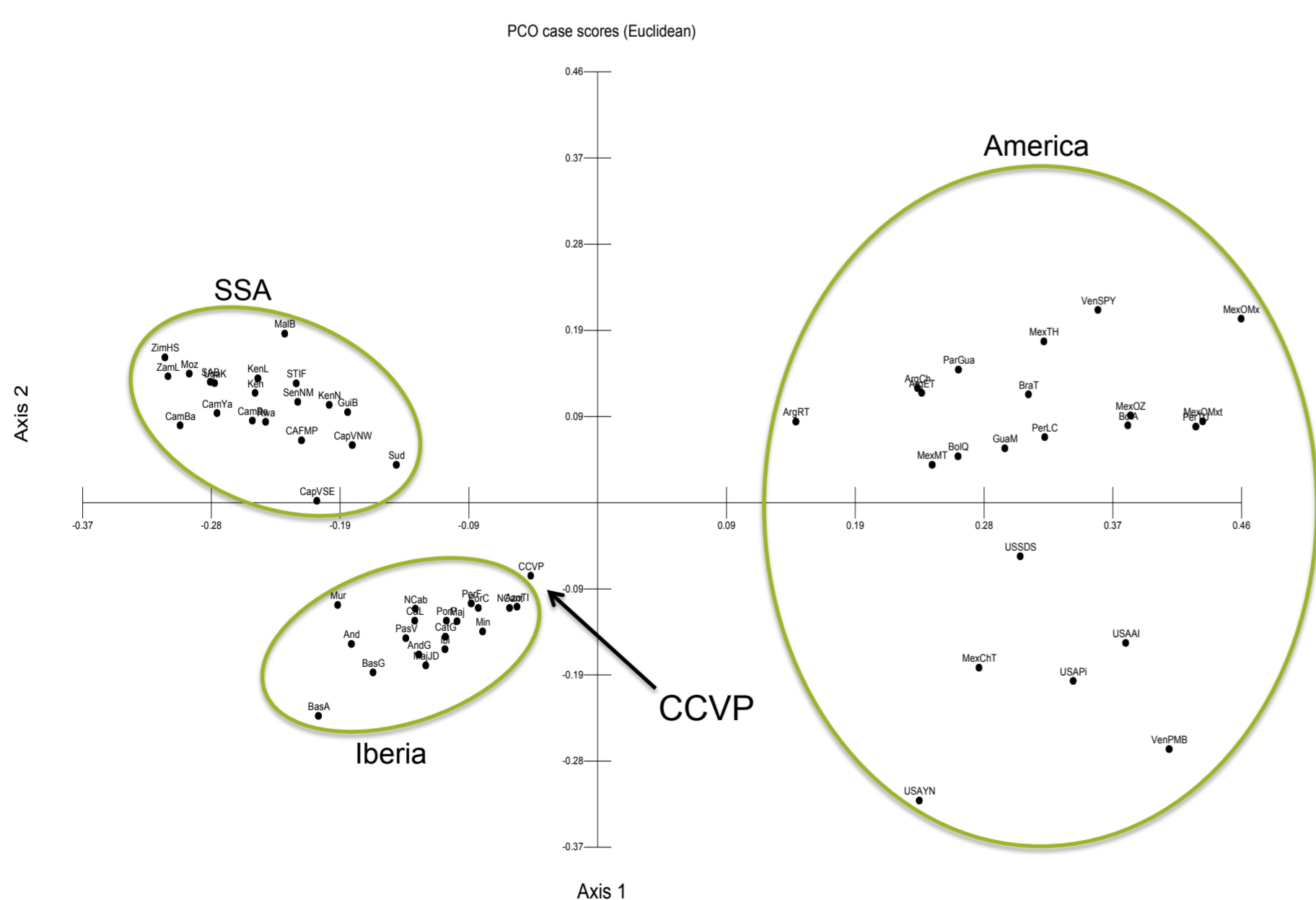


Figure 2 The CCVP has an important Caucasian component. Principal Coordinates Analysis (PCO, 57.4% cumulative variance) of the CCVP and 60 ancestral populations from Sub-Saharan Africa (upper left quadrant), Iberia (lower left quadrant) and America (right quadrants) based on 47 HLA-A and HLA-B allele group frequencies.

Results

A total of 37 HLA-A, 61 HLA-B, 24 HLA-C and 38 HLA-DRB1 alleles were seen in this sample. The five most frequent alleles for these genes are A*02:01:01, A*24:02:01, A*03:01:01, A*01:01:01, A*68:01:02, B*07:02:01, B*35:01:01, B*40:02:01, B*44:02:01, B*14:02:01, C*04:01:01, C*07:02:01, C*03:05, C*06:02:01, C*07:01:01, DRB1*13:01:01G, DRB1*04:07:01G, DRB1*15:01:01G, DRB1*03:01:01G, and DRB1*07:01:01G. Principal coordinates analysis based on HLA-A and -B allele group frequencies reveals that this population clusters closely to Iberian populations (**Figure 2**). Preliminary haplotype estimation results show, as a proxy for admixture proportions, that 68% of the extended haplotypes are Caucasian, while 23% are Amerindian in origin and 9% are clearly Sub-Saharan African (**Figure 3**).

Table 1 High resolution HLA frequencies in the CCVP.

| HLA-A Allele | Count | Frequency | Homozygotes | Phenotype frequency (%) |
|--------------|-------|-----------|-------------|-------------------------|
| 01:01:01 | 32 | 0.0780 | | 15.1 |
| 02:01:01 | 70 | 0.1707 | 5 | 31.7 |
| 02:02 | 1 | 0.0024 | | 0.5 |
| 02:05:01 | 3 | 0.0073 | | 1.5 |
| 02:06:01 | 13 | 0.0317 | | 6.3 |
| 02:11:01 | 1 | 0.0024 | | 0.5 |
| 02:22:01 | 2 | 0.0049 | | 1.0 |
| 03:01:01 | 47 | 0.1146 | 2 | 22.0 |
| 03:02:01 | 1 | 0.0024 | | 0.5 |
| 11:01:01 | 9 | 0.0220 | | 4.4 |
| 23:01:01 | 5 | 0.0122 | | 2.4 |
| 24:02:01 | 57 | 0.1390 | 4 | 25.9 |
| 24:03:01 | 3 | 0.0073 | | 1.5 |
| 25:01:01 | 18 | 0.0439 | | 8.8 |
| 26:01:01 | 18 | 0.0439 | | 8.8 |
| 26:08 | 1 | 0.0024 | | 0.5 |
| 29:01:01 | 2 | 0.0049 | | 1.0 |
| 29:02:01 | 13 | 0.0317 | | 6.3 |
| 30:01:01 | 7 | 0.0171 | 1 | 2.9 |
| 30:02:01 | 9 | 0.0220 | | 4.4 |
| 30:04:01 | 7 | 0.0171 | | 3.4 |
| 31:01:02 | 13 | 0.0317 | | 6.3 |
| 32:01:01 | 13 | 0.0317 | | 6.3 |
| 33:01:01 | 13 | 0.0317 | | 6.3 |
| 33:03:01 | 1 | 0.0024 | | 0.5 |
| 34:02:01 | 2 | 0.0049 | | 1.0 |
| 36:01 | 1 | 0.0024 | | 0.5 |
| 66:01 | 2 | 0.0049 | | 1.0 |
| 66:02 | 1 | 0.0024 | | 0.5 |
| 68:01:01 | 5 | 0.0122 | | 2.4 |
| 68:01:02 | 22 | 0.0537 | 1 | 10.7 |
| 68:02:01 | 6 | 0.0146 | | 2.9 |
| 68:03:01 | 4 | 0.0098 | | 2.0 |
| 68:07 | 4 | 0.0098 | | 2.0 |
| 68:30 | 1 | 0.0024 | | 0.5 |
| 74:01:01 | 1 | 0.0024 | | 0.5 |
| 80:01:01 | 2 | 0.0049 | | 1.0 |
| | | | | Heterozygosity |
| Total | 410 | 1 | 13 | 93.7 |

| HLA-B Allele | Count | Frequency | Homozygotes | Phenotype frequency (%) |
|--------------|-------|-----------|-------------|-------------------------|
| 07:02:01 | 48 | 0.1171 | 1 | 22.9 |
| 07:05:01 | 2 | 0.0049 | | 1.0 |
| 08:01:01 | 15 | 0.0366 | | 7.3 |
| 13:02:01 | 2 | 0.0049 | | 1.0 |
| 14:01:01 | 2 | 0.0049 | | 1.0 |
| 14:02:01 | 20 | 0.0488 | | 9.8 |
| 15:01:01 | 16 | 0.0390 | | 7.8 |
| 15:03:01 | 3 | 0.0073 | | 1.5 |
| 15:09 | 1 | 0.0024 | | 0.5 |
| 15:10:01 | 1 | 0.0024 | | 0.5 |
| 15:15 | 1 | 0.0024 | | 0.5 |
| 15:16:01 | 1 | 0.0024 | | 0.5 |
| 15:17:01 | 4 | 0.0098 | | 2.0 |
| 15:18:01 | 1 | 0.0024 | | 0.5 |
| 18:01:01 | 18 | 0.0439 | | 8.8 |
| 18:05 | 1 | 0.0024 | | 0.5 |
| 27:05:02 | 2 | 0.0049 | | 1.0 |
| 35:01:01 | 35 | 0.0854 | 2 | 16.1 |
| 35:02:01 | 3 | 0.0073 | | 1.5 |
| 35:03:01 | 7 | 0.0171 | | 3.4 |
| 35:08:01 | 8 | 0.0195 | | 3.9 |
| 35:10 | 1 | 0.0024 | | 0.5 |
| 35:12:01 | 6 | 0.0146 | | 2.9 |
| 35:14:01 | 1 | 0.0024 | | 0.5 |
| 35:16 | 1 | 0.0024 | | 0.5 |
| 35:17 | 1 | 0.0024 | | 0.5 |
| 35:43:01 | 8 | 0.0195 | | 3.9 |
| 37:01:01 | 12 | 0.0293 | | 5.9 |
| 38:01:01 | 8 | 0.0195 | | 3.9 |
| 38:03 | 1 | 0.0024 | | 0.5 |
| 39:01:01 | 1 | 0.0024 | | 0.5 |
| 39:05:01 | 3 | 0.0073 | | 1.5 |
| 39:06:01 | 1 | 0.0024 | | 0.5 |
| 39:08 | 11 | 0.0268 | | 5.4 |
| 39:11 | 2 | 0.0049 | | 1.0 |
| 40:01:01 | 6 | 0.0146 | | 2.9 |
| 40:02:01 | 35 | 0.0854 | 4 | 15.1 |
| 40:11:01 | 2 | 0.0049 | | 1.0 |
| 40:12 | 1 | 0.0024 | | 0.5 |
| 40:16 | 1 | 0.0024 | | 0.5 |
| 41:01:01 | 7 | 0.0171 | | 3.4 |
| 41:02:01 | 1 | 0.0024 | | 0.5 |
| 42:01:10 | 1 | 0.0024 | | 0.5 |
| 44:02:01 | 23 | 0.0561 | | 11.2 |
| 44:03:01 | 15 | 0.0366 | | 7.3 |
| 45:01 | 9 | 0.0220 | | 4.4 |
| 48:01:01 | 1 | 0.0024 | | 0.5 |
| 48:02:01 | 1 | 0.0024 | | 0.5 |
| 49:01:01 | 4 | 0.0098 | | 2.0 |
| 50:01:01 | 2 | 0.0049 | | 1.0 |
| 51:01:01 | 11 | 0.0268 | | 5.4 |
| 52:01:01 | 3 | 0.0073 | | 1.5 |
| 53:01:01 | 15 | 0.0366 | 1 | 6.8 |
| 55:01:01 | 3 | 0.0073 | | 1.5 |
| 56:01:01 | 2 | 0.0049 | | 1.0 |
| 57:01:01 | 10 | 0.0244 | | 4.9 |
| 57:02:01 | 2 | 0.0049 | | 1.0 |
| 57:03:01 | 1 | 0.0024 | | 0.5 |
| 58:01:01 | 4 | 0.0098 | | 2.0 |
| 58:02 | 1 | 0.0024 | | 0.5 |
| 73:01 | 1 | 0.0024 | | 0.5 |
| | | | | Heterozygosity |
| Total | 410 | 1.0000 | 8 | 96.1 |

| HLA-C Allele | Count | Frequency | Homozygotes | Phenotype frequency (%) |
|--------------|-------|-----------|-------------|-------------------------|
| 01:02:01 | 15 | 0.0371 | | 7.4 |
| 02:02:02 | 7 | 0.0173 | | 3.5 |
| 02:10 | 2 | 0.0050 | | 1.0 |
| 03:02:01 | 4 | 0.0099 | | 2.0 |
| 03:03:01 | 12 | 0.0297 | | 5.9 |
| 03:04:01 | 18 | 0.0446 | | 8.9 |
| 03:05 | 37 | 0.0916 | 3 | 16.8 |
| 04:01:01 | 70 | 0.1733 | 3 | 33.2 |
| 05:01:01 | 24 | 0.0594 | | 11.9 |
| 05:20 | 1 | 0.0025 | | 0.5 |
| 06:02:01 | 37 | 0.0916 | 1 | 17.8 |
| 07:01:01 | 28 | 0.0693 | 2 | 12.9 |
| 07:02:01 | 69 | 0.1708 | 2 | 33.2 |
| 07:04:01 | 5 | 0.0124 | | 2.5 |
| 08:02:01 | 23 | 0.0569 | | 11.4 |
| 12:02:02 | 3 | 0.0074 | | 1.5 |
| 12:03:01 | 21 | 0.0520 | | 10.4 |
| 14:02:01 | 5 | 0.0124 | | 2.5 |
| 15:02:01 | 2 | 0.0050 | | 1.0 |
| 15:05:01 | 3 | 0.0074 | | 1.5 |
| 16:01:01 | 10 | 0.0248 | | 5.0 |
| 16:02:01 | 2 | 0.0050 | | 1.0 |
| 17:01:01 | 3 | 0.0074 | | 1.5 |
| 18:01:01 | 3 | 0.0074 | | 1.5 |
| | | | | Heterozygosity |
| Total | 404 | 1.0000 | 11 | 94.6 |

| HLA-DRB1 Allele | Count | Frequency | Homozygotes | Phenotype frequency (%) |
|-----------------|-------|-----------|-------------|-------------------------|
| 01:01:01G | 24 | 0.0585 | | 11.7 |
| 01:02:01 | 12 | 0.0293 | | 5.9 |
| 01:03 | 4 | 0.0098 | | 2.0 |
| 03:01:01G | 30 | 0.0732 | 1 | 14.1 |
| 03:02:01 | 1 | 0.0024 | | 0.5 |
| 04:01:01 | 1 | 0.0024 | | 0.5 |
| 04:02:01 | 11 | 0.0268 | | 5.4 |
| 04:03:01 | 8 | 0.0195 | | 3.9 |
| 04:04:01 | 16 | 0.0390 | 1 | 7.3 |
| 04:05:01 | 5 | 0.0122 | | 2.4 |
| 04:07:01G | 38 | 0.0927 | 4 | 16.6 |
| 04:08:01 | 1 | 0.0024 | | 0.5 |
| 04:10:01 | 2 | 0.0049 | | 1.0 |
| 04:11:01 | 13 | 0.0317 | 1 | 5.9 |
| 04:17:02 | 2 | 0.0049 | | 1.0 |
| 07:01:01G | 25 | 0.0610 | | 12.2 |
| 08:01:01G | 1 | 0.0024 | | 0.5 |
| 08:02:01 | 21 | 0.0512 | | 10.2 |
| 08:04:01 | 5 | 0.0122 | | 2.4 |
| 10:01:01 | 5 | 0.0122 | | 2.4 |
| 11:01:01G | 24 | 0.0585 | 1 | 11.2 |
| 11:02:01 | 12 | 0.0293 | | 5.9 |
| 11:03 | 3 | 0.0073 | | 1.5 |
| 11:04:01 | 7 | 0.0171 | | 3.4 |
| 11:45 | 1 | 0.0024 | | 0.5 |
| 12:01:01G | 2 | 0.0049 | | 1.0 |
| 12:02:01 | 1 | 0.0024 | | 0.5 |
| 13:01:01G | 44 | 0.1073 | 2 | 20.5 |
| 13:02:01 | 17 | 0.0415 | | 8.3 |
| 13:03:01 | 4 | 0.0098 | | 2.0 |
| 13:04 | 3 | 0.0073 | | 1.5 |
| 13:05:01 | 4 | 0.0098 | | 2.0 |
| 14:01:01G | 13 | 0.0317 | | 6.3 |
| 14:02 | 9 | 0.0220 | | 4.4 |
| 15:01:01G | 32 | 0.0780 | | 15.6 |
| 15:02:01 | 1 | 0.0024 | | 0.5 |
| 16:01:01 | 4 | 0.0098 | | 2.0 |
| 16:02:01 | 4 | 0.0098 | | 2.0 |
| | | | | Heterozygosity |
| Total | 410 | 1 | 10 | 95.1 |

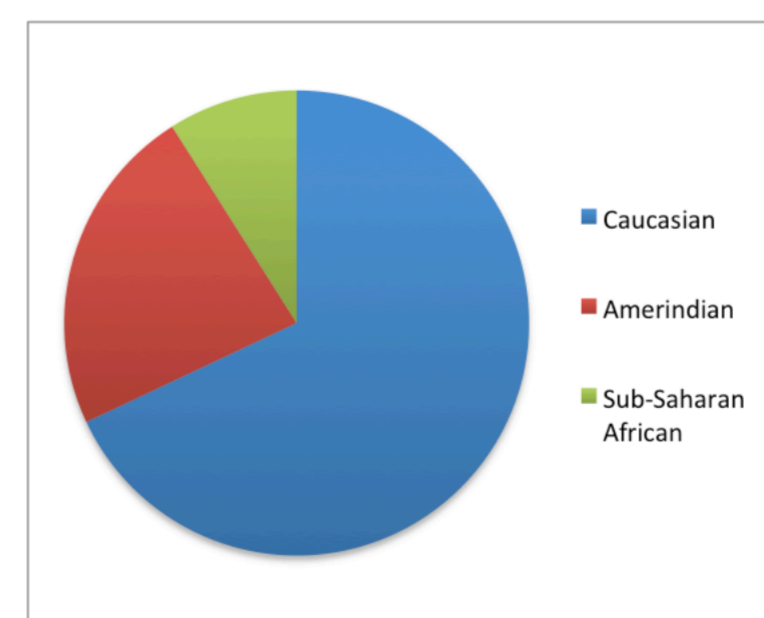


Figure 3 Proposed origin of extended HLA haplotypes found in the CCVP.

Conclusions and future work

The CCVP shows a diverse HLA profile, with an important Caucasian component. We are currently continuing the definition of high-resolution HLA allele and haplotype frequencies for this and other Costa Rican populations.

Acknowledgements

This work was supported by grants from the University of Costa Rica and the Costa Rican National Council for Scientific and Technologic Research (CONICIT). Molecular typing was carried out at Anthony Nolan Research Institute.

