

HLADQA1 ALLELE-FREQUENCIES IN THE WORLD USING A BIPLOT TO VISUALISE ALLELES AND POPULATIONS SIMULTANEOUSLY

A.D. Kloosterman*, M. Sjerps*, D. Eerhart* and N. Dimo**

Gerechtelijk Laboratorium, Ministerie van Justitie, Rijswijk, NL*
Institute Med Légale, Université de Lausanne, Switzerland**

INTRODUCTION

Population genetic studies show differences within and between racial groups for the HLADQA1 locus^{1 2}. The chi-square test is a relatively simple test which is often used when comparing population samples. However, when a large number of samples are compared, it is difficult to get a clearly structured overall picture. A quick overview of the distances between populations and deviant allele frequencies can be obtained from a so-called biplot³. This type of plot is related to principal components analysis and has many variants, which are increasingly used in anthropometric and genetic analyses. Using the biplot graphic display we analysed the HLADQA1 allele-frequency data of 31 population samples^{4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19}.

STATISTICAL METHODS

Table 1 gives the overview of the 31 populations that were compared. A biplot¹ is used to visualise the allele frequency data table which consists of rows (populations) and columns (alleles). The type of plot we use is obtained from relative allele frequencies by subtracting column means and scaling of row factors with the singular values. The data analysis was performed using the software Spectramap²⁰. The biplots are interpreted in the following way. Alleles are represented by squares, and populations by circles. Populations that are close together in the plot have similar allelic compositions, in contrast to populations that are far apart. The population frequencies of any allele *a* can be compared by perpendicular projection of the populations on a line drawn through allele “*a*” and the origin (Fig.1). The horizontal and vertical direction in the biplot each account for a certain fraction of the total information (e.g., the total variance of the allele frequencies). These fractions are reported in the figure legend.

RESULTS

In the biplot of all examined populations (Fig.1), the Caucasians appear a relatively homogeneous group, and so do the Black populations. The Asian populations are scattered over the plot, indicating heterogeneity. By perpendicular projection of the population points on the axis drawn through the centre of the plot (+) and allele 3, it can be seen that the Japanese sample shows an exceptionally high frequency for this allele, whereas the Indonesian sample scores very low. Other differences can be detected in a similar way.

SPECTRAMAP

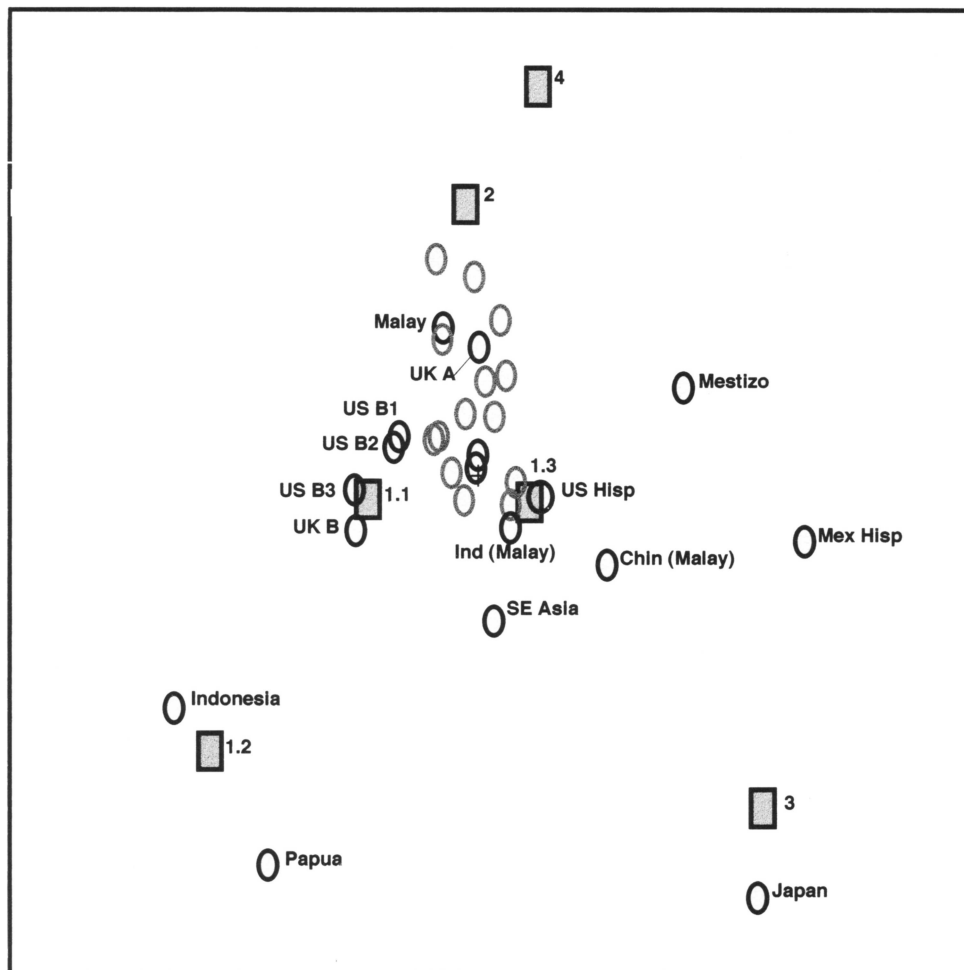


Fig 1 Biplot of the HLADQA1-locus for all populations investigated (abbreviations as in Table 1). The biplot graphic display accounts for 73% of the total variance (horizontal and vertical direction 46% and 27% resp).

The Caucasian-population samples which appear a relatively homogeneous group are given by grey circles and not further specified.

CONCLUSIONS

The biplot allows a quick overview of similarities and differences between the allele frequency distributions when many samples are analysed. A nice feature of the plot is that it visualises both alleles and populations in the same plot. It should be noted, however, that some of the information is lost because the biplot reduces a high dimensional space to only two dimensions. Nevertheless, the biplots in this paper reproduce 73% of the total variation, hence most of the information is retained.

Table 1 Overview of the analysed population-samples

Country	Racial group	N ind	Ref	Biplot code	Country	Racial group	N ind	Ref
US	Black	224	1	US B1	US	Cauc	413	1
US	Black	172	1	US B2	Italy	Cauc	103	18
Marion US	Black	203	8	US B3	Lisboa Port	Cauc	106	5
UK	Asian	191	11	UK A	Porto Port	Cauc	325	6
Japan	Asian	92	1	Japan	Andalusia Sp.	Cauc	174	7
Mexico	Mestizo	100	1	Mestizo	Marion US	Cauc	185	8
Mexico	Hispanic	169	1	Mex. Hisp	US	Cauc	174	1
US	Hispanic	146	1	US Hisp	Bavaria Ger.	Cauc	213	9
SE Asia	Asian	87	1	SE Asia	S-W Swit	Cauc	237	19
Indonesia	Asian	144	1	Indonesia	NL	Cauc	157	15
NW Guinea	Papua	134	1	Papua	Germany	Cauc	212	10
Malaysia	Asian Malay	130	17	Malay	Finland	Cauc	112	12
Malaysia	Asian Chinese	125	17	Chin (Malay)	Australia	Cauc	280	16
Malaysia	Asian Indian	137	17	Ind (Malay)	Switzerland	Cauc	200	13
UK	Black	202	11	UK B	Spain	Cauc	120	14
					UK	Cauc	201	11

REFERENCES

- 1 R. Helmuth et al (1990); HLADQA1 allele and genotype frequencies in various human populations; *Am J Hum Genet* **47**:515-523
- 2 A Piazza A et al (1980); The HLA-A,B gene frequencies in the world: migration or selection? *Hum Immunol* **1**:297-304
- 3 Gabriel KR (1971) The biplot graphic display of matrices with applications to principal components analysis. *Biometrika* **58**: 453-467
- 4 Cavalli-Sforza LL, Menozzi P, Piazza A (1994). The history and geography of human genes. Princeton University Press, New Jersey.
- 5 R Espinheira et al; *Adv For Haemogen* **5**:499-501
- 6 MF Pinheiro et al; *Adv For Haemogen* **5**:559-561
- 7 P Sanz et al. *Adv For Haemogen* **5**:572-574
- 8 MA Tahir et al; *Adv For Haemogen* **5**:596-598
- 9 GM Weichold et al; *Adv For Haemogen* **5**:599-602
- 10 PM Schneider et al (1991); *Arch Krimino* **188**:167-174
- 11 KM Sullivan et al (1992); *Int J Leg Med* **105**:17-20
- 12 A Sajantila et al (1991); *Int J Leg Med* **104**:181-184
- 13 MN Hochmeister et al The application of PCR to forensic science. Birkhauser Verlag, Basel
- 14 JA Lorente et al (1992); *Proceedings Promega Int Symp Human Identification*
- 15 AD Kloosterman et al (1993); *Int J Leg Med* **105**:233-238
- 16 CS Harrington et al (1991); *Forensic Sc Int* **51**:147-157
- 17 Chong-Lek Koh et al (1994); *Hum Hered* **44**:150-155
- 18 A. Tagliabracci et al (1992); *Int J Leg Med* **105**:161-164
- 19 N. Dimo (1995), personal communication
- 20 Smit Consult, Drunen, The Netherlands; Spectramap for biplot analysis