Data Article

Human Leukocyte Antigen (HLA) class I and II datasets for sudanese

Ameer Mohamed Dafalla [a, *, 1], Hisham Atan Edinur [b, 1], Mohammed Abdelwahed [a], Almutaz Abbas Elemam [a], Amel Abdeen Ibrahim [a], Ahmed Mohamadani [d], Mubarak Magzoub [a], Gehad ElGhazali [c, 1]

[a] HLA Laboratory, National Cancer Institute, University of Gezira, Wad Medani, P.O. Box 20, Sudan
[b] Forensic Science Programme, School of Health Sciences, Universiti Sains Malaysia, 16150, Kubang Kerian, Kelantan, Malaysia
[c] Department of Immunology, Pathology and Laboratory Medicine Services, Sheikh Khalifa Medical City, Abu Dhabi, P.O Box 51900, United Arab Emirates
[d] Department of Pathology, Faculty of Medicine, University of Gezira, P.O. Box 20, Sudan

ARTICLE INFO

ABSTRACT

Sudan is located in the heart of Africa and surrounded by eight countries with people of different ethnic origins. Historical records show that the population of the Sudan is a mixture of Arabic, West Asian Arabic and sub-Saharan African elements. The present survey provides data on allele lineages, and haplotype frequencies of Human Leukocyte Antigen (HLA) class I (HLA-A and HLA-B) and class II (HLA-DR and −HLA-DQ) loci in 11 Sudanese populations. The sampled individuals are all local transplant donors who provided informed consent for HLA analyses on their blood samples and were registered at National Cancer Institute, University of Gezira, Wad Medani. The HLA class I and II data reported here can be subjected for future analyses of genetic structure and health in Sudan. These include as reference datasets for identifying the association between HLA and diseases and for designing donor recruitment strategies.

* Corresponding author.
  E-mail address: ameerdafalla@gmail.com (A.M. Dafalla).
[1] Author who equally contributed to this work.

Specification table

Subject area	Immunology
More specific subject area	Immunogenetics
Type of data	Tables and figure
How data was acquired	Genomic DNA was extracted from blood samples using QIAGEN QIAamp DNA Mini Kit® (Qiagen, Hilden, Germany) and HLA class I and II loci were typed using sequence specific typing primer (SSP) typing kit (R.O·S.E. GenTec Ltd., Europe GmbH, Germany and Olerup SSP, Stockholm, Sweden). Allele and haplotype frequencies were estimated using algorithm implemented in Python for Population Genomics (PyPop vs. 0.7.0) while principal component plot was constructed using Multivariate Statistical Package 3 (MVSP3) software
Data format	Raw and analysed
Experimental factors	Blood samples collection, DNA extraction and purification, SSP amplification, HLA allelic scoring using provided software and HLA data analyses
Experimental features	DNA SSP amplification of HLA loci followed by population genetic analyses
Data accessibility	Data is available with this article
Data source location	Sudan
Related research article	F.F. González-Galarza, L.Y. Takeshita, E.J. Santos, F. Kempson, M.H. Maia, A.L. da Silva, A.L. Teles e Silva, G.S. Ghattaoraya, A. Alfirevic, A.R. Jones, D. Middleton, Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations, Nucleic Acids Res. 43 (2015) D784-788 [1].

Value of the data

- Most Sudanese populations included in this survey (except Nuba) have not been previously characterized for HLA class I and II loci [1].
- HLA data have been widely used for population studies [2] and the HLA class I and II data reported here can be relevant for future analyses of population structure in Sudan
- Many diseases found to be associated to HLA markers and the data collected from Sudanese populations can be used as reference for future studies on diseases associated with HLA [3].
- HLA antigens are essential markers for transplantation compatibility. Thus, HLA data for Sudanese populations provide unprecedented insights into allelic spectra in Sudan and can be used for designing donor recruitment strategies.

1. Data

In this survey, we typed HLA loci in 4701 individuals who were registered as transplant donors at the National Cancer Institute, University of Gezira, Wad Medani. These individuals were socio-culturally classified into eleven Sudanese populations of distinct ethnic origins (Table 1). Their HLA

Table 1
List of Sudanese populations tested in this study and their group ID.

Group ID	Number of individuals sampled	Populations	Ethnicities included in each population
T1	1973	Gaalia	Gaalia, Shagia, Umarab, Bederia, Salamab Manaseer, Rekabia, Awamra, Merfab, and Rubatab
T2	994	Juhayna	Abdellab, Arakia, Gawasma, Halaween, Hassania, Kawahla, Khawalda, Kinana, Masalamia, Fadnia, Rufaa, Shokria, Dghumia, Dubaseen, Ageelia, Bataheen and Hadaria
T3	537	Baggara	Maalia, Messeria, Rezigat, Taaisha, Baggara, Hapania, Banihalpa, Hamar, Gawamia, Gamuia, Darhamed, Darhussain, Garrara, and Kababeesh
T4	531	Nile Nubians	Danagla, Mahas, Halfawi and Scoot
T5	91	Nuba	Nuba, and Taggalia
T6	160	Darfurians	Tama, Tunjor, Zagawa, Perti, Four, Gammari, Miema, Burno, Bazaa, Dajo, Taragma, Burgo, Masaleet and Selahab
T7	181	Moroccans and Egyptian Origions	Gaafra, Ashraf, Basawla, Mawaleed Knoze, Magrba and Nagada
T8	76	Falata	Falata, Folani and Hawsa
T9	128	Beja	Hadandwa, Baniamir, Busharia and Halanga
T10	9	Angasana	Fong
T11	21	Rashayda	Rashayda

Table 2
HLA-A*, HLA-B, HLA-DRB1* and HLA-DQB1* allele frequencies in eleven Sudanese populations. See Table 1 for group IDs.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
A*01	0.09	0.09	0.08	0.08	0.08	0.06	0.10	0.13	0.07	0.22	0.02
A*02	0.26	0.25	0.25	0.29	0.31	0.28	0.26	0.24	0.35	0.22	0.31
A*03	0.08	0.08	0.08	0.08	0.07	0.03	0.08	0.07	0.08	0.00	0.05
A*11	0.03	0.02	0.02	0.01	0.01	0.01	0.03	0.00	0.04	0.00	0.00
A*23	0.03	0.04	0.04	0.03	0.05	0.05	0.04	0.08	0.04	0.11	0.07
A*24	0.07	0.06	0.04	0.07	0.03	0.03	0.10	0.05	0.07	0.00	0.02
A*25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A*26	0.02	0.03	0.01	0.01	0.03	0.03	0.03	0.01	0.02	0.00	0.00
A*29	0.02	0.02	0.03	0.02	0.00	0.03	0.01	0.04	0.03	0.11	0.02
A*30	0.16	0.12	0.17	0.18	0.15	0.18	0.15	0.11	0.09	0.00	0.17
A*31	0.03	0.04	0.06	0.02	0.03	0.02	0.02	0.03	0.02	0.00	0.05
A*32	0.06	0.05	0.04	0.05	0.04	0.07	0.03	0.04	0.04	0.00	0.02
A*33	0.04	0.04	0.05	0.04	0.03	0.06	0.02	0.09	0.03	0.00	0.05
A*34	0.02	0.02	0.03	0.02	0.03	0.02	0.02	0.01	0.00	0.00	0.02
A*36	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.02
A*43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A*66	0.02	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.01	0.00	0.02
A*68	0.07	0.08	0.07	0.06	0.09	0.10	0.08	0.07	0.09	0.17	0.12
A*69	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.11	0.00
A*74	0.01	0.03	0.01	0.01	0.02	0.00	0.01	0.00	0.02	0.06	0.02
A*80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
B*07	0.05	0.06	0.06	0.04	0.05	0.05	0.07	0.04	0.07	0.00	0.07
B*08	0.03	0.02	0.03	0.03	0.06	0.03	0.03	0.01	0.03	0.00	0.10
B*13	0.03	0.03	0.04	0.03	0.04	0.03	0.04	0.06	0.07	0.00	0.00
B*14	0.04	0.03	0.02	0.03	0.04	0.02	0.02	0.00	0.00	0.00	0.02
B*15	0.08	0.06	0.11	0.10	0.15	0.10	0.08	0.07	0.09	0.11	0.14
B*18	0.02	0.02	0.02	0.02	0.03	0.01	0.01	0.03	0.02	0.11	0.02
B*27	0.02	0.01	0.02	0.03	0.02	0.06	0.02	0.07	0.03	0.06	0.00
B*35	0.06	0.07	0.06	0.07	0.04	0.05	0.08	0.05	0.03	0.11	0.07
B*37	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.03	0.01	0.00	0.00
B*38	0.05	0.04	0.03	0.04	0.00	0.02	0.04	0.03	0.03	0.00	0.00
B*39	0.06	0.07	0.07	0.08	0.09	0.06	0.05	0.04	0.06	0.06	0.05
B*40	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.03	0.00	0.00
B*41	0.08	0.05	0.05	0.10	0.04	0.03	0.07	0.03	0.05	0.00	0.07
B*42	0.02	0.02	0.04	0.02	0.02	0.03	0.04	0.07	0.00	0.00	0.05
B*44	0.02	0.02	0.02	0.02	0.01	0.03	0.01	0.01	0.02	0.00	0.00
B*45	0.03	0.03	0.02	0.03	0.05	0.06	0.02	0.04	0.00	0.06	0.00
B*46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B*47	0.01	0.02	0.01	0.02	0.04	0.02	0.02	0.01	0.01	0.06	0.00
B*48	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
B*49	0.05	0.04	0.03	0.05	0.03	0.04	0.04	0.04	0.03	0.00	0.02
B*50	0.06	0.09	0.04	0.05	0.03	0.04	0.09	0.03	0.09	0.00	0.14
B*51	0.11	0.13	0.09	0.07	0.04	0.06	0.08	0.05	0.13	0.06	0.05
B*52	0.06	0.06	0.04	0.04	0.00	0.00	0.07	0.03	0.04	0.11	0.00
B*53	0.02	0.02	0.05	0.01	0.04	0.07	0.03	0.07	0.03	0.06	0.02
B*54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B*55	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.02
B*56	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
B*57	0.04	0.03	0.02	0.05	0.01	0.03	0.03	0.04	0.05	0.06	0.02
B*58	0.03	0.05	0.07	0.03	0.09	0.10	0.03	0.08	0.05	0.17	0.12
B*59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B*67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B*73	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.00
B*78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
B*81	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00
B*82	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
B*83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DQB1*02	0.23	0.27	0.21	0.19	0.17	0.13	0.27	0.30	0.30	0.17	0.26
DQB1*03	0.23	0.22	0.24	0.29	0.29	0.32	0.22	0.26	0.17	0.22	0.17
DQB1*04	0.02	0.02	0.02	0.03	0.01	0.04	0.03	0.04	0.04	0.11	0.05

Table 2 (*continued*)

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
DQB1*05	0.16	0.16	0.23	0.15	0.29	0.27	0.13	0.20	0.18	0.06	0.31
DQB1*06	0.36	0.33	0.30	0.34	0.24	0.24	0.35	0.20	0.30	0.44	0.21
DRB1*01	0.05	0.04	0.05	0.05	0.10	0.04	0.04	0.05	0.06	0.00	0.10
DRB1*03	0.11	0.13	0.11	0.08	0.09	0.09	0.16	0.14	0.11	0.00	0.17
DRB1*04	0.06	0.10	0.06	0.07	0.04	0.04	0.10	0.10	0.12	0.11	0.10
DRB1*07	0.09	0.11	0.07	0.09	0.04	0.04	0.10	0.07	0.12	0.11	0.14
DRB1*08	0.09	0.06	0.12	0.10	0.09	0.22	0.06	0.16	0.05	0.17	0.07
DRB1*09	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
DRB1*10	0.05	0.05	0.07	0.05	0.04	0.09	0.04	0.06	0.04	0.06	0.05
DRB1*11	0.12	0.10	0.15	0.14	0.23	0.16	0.11	0.11	0.09	0.06	0.12
DRB1*12	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.01	0.00	0.00	0.00
DRB1*13	0.22	0.21	0.20	0.24	0.22	0.21	0.18	0.18	0.26	0.28	0.24
DRB1*14	0.02	0.01	0.03	0.01	0.02	0.03	0.02	0.04	0.02	0.00	0.00
DRB1*15	0.17	0.16	0.13	0.16	0.11	0.07	0.19	0.06	0.13	0.22	0.02
DRB1*16	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00

class I and II genotypes and allele lineage frequencies are reported in Table S1 and Table 2, respectively. HLA-A*~B*, HLA-B*~DRB1*, HLA-DRB1*~DQB1*, HLA-A*~B*~DRB1* and HLA-A*~B*~DRB1*~DQB1* haplotype frequencies are reported in Tables S2—S6, respectively. No significant deviation from Hardy-Weinberg equilibrium was observed for HLA-A*, -B*, -DQB1* and −DRB1* in Angasana, HLA-A*, -B* and −DRB1* for Gaalia and Rashayda, HLA-A* and -DQB1* in Nuba and Darfurians, HLA-A* in Moroccans and Egyptian Origions, HLA-A* and -DRB1* in Falata and HLA-DRB1* in Beja. A principal component analysis (PCA) for the eleven Sudanese populations and other populations [1,4] from North Africa, Sub-Saharan Africa, Europe and Asia is shown in Fig. 1. The Sudanese populations form a cluster with other North African and Sub-Saharan African populations on the upper and lower left of the PCA while other populations from Asia and Europe are projected towards the right corner of the plot.
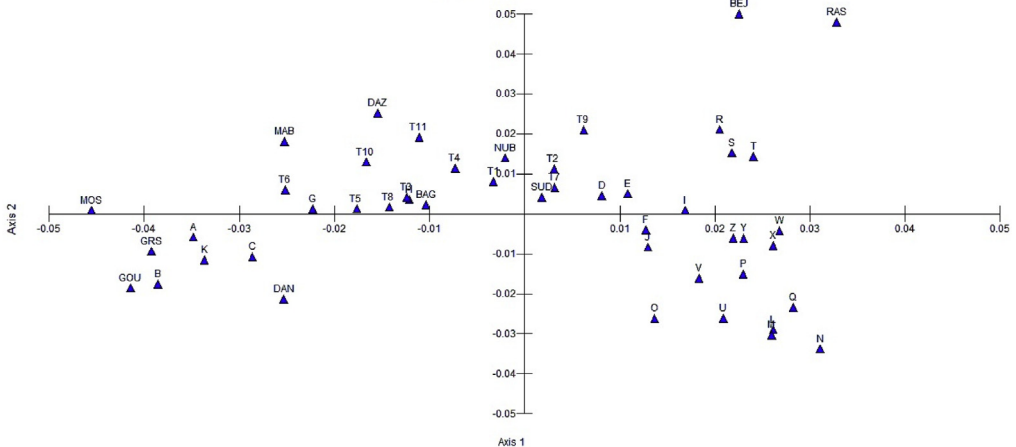

Fig. 1. Principal component analysis of eleven Sudanese populations tested in this study (T1 to T11 and see Table 1 for details) and other North African, Sub-Saharan African, European and Asian populations from ref. (1—4). A: Burkina Faso Mossi; B: Burkina Faso Rimaibe; C: Kenya Nyanza; D: Morocco; E: Morocco Atlantic; F: Morocco pop 2; G: Rwanda; H: Sudan mixed; I: Tunisia Gabes; J: Tunisia pop 3; K: Zimbabwe Harare; L: Tunisia pop 3; M: Iraq Kurdistan; N: Turkey pop 2; O: Iran Baloch; P: Iran Gorgan; Q: Iran Kurds; R: Saudi Arabia Guraian and Hail; S: Saudi Arabia pop 2; T: Saudi Arabia pop 4; U: Greece Crete; V: Italy Rome; W: France Rennes; X: Spain Andalusia; Y: Portugal Center; Z: Germany Essen; SUD: Sudanese Arab; RAS: Rashaayda Bedouins; BEJ: Beja Hadendowa; NUB: Nubians; DAZ: Daza; BAG: Baggara Arabs; MAB: Maba; DAN: Dangaleat; MOS: Mossi; GRS: Gurunsi Kassena; GOU: Gourmantche.

2. Experimental design, materials and methods

2.1. Sample collection and genomic DNA extraction

Blood samples were obtained with written informed consent from 4701 healthy Sudanese volunteers and this study was approved by ethics committee of University of Gezira, Wad Medani, Sudan. DNA was extracted using QIAamp DNA Mini Kit® (Qiagen, Hilden, Germany), according to the manufacture instructions.

2.2. PCR SSP HLA typing

The isolated genomic DNA was typed using sequence specific primer (SSP) typing kits (R.O·S.E. GenTec Ltd., Europe GmbH, Germany and Olerup SSP Stockholm, Sweden). The HLA specific PCR amplicons were separated using 1% ethidium bromide stained agarose gel electrophoresis and visualized by exposure to ultraviolet light. The separated amplicons were analysed and HLA scored using the provided software.

2.3. Statistical analysis

Allele and haplotype frequencies were estimated using an expectation-maximization algorithm implemented in Python for Population Genomics (PyPop vs. 0.7.0) [5]. The PyPop software was also used to carry out Hardy-Weinberg equilibrium (HWE) and selective neutrality tests. Principal coordinate (PCA) plots were conducted for Sudanese and other populations obtained from other sources [1,4] using Multivariate Statistical Package 3 (MVSP3) software (Multivariate Statistical Software Package 3, Kovach Computing Services, Pentraeth, Isle of Anglesey, UK).

Transparency document

Transparency document associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2019.104027.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.104027.

References

[1] F.F. González-Galarza, L.Y. Takeshita, E.J. Santos, F. Kempson, M.H. Maia, A.L. da Silva, A.L. Teles e Silva, G.S. Ghattaoraya, A. Alfirevic, A.R. Jones, D. Middleton, Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations, Nucleic Acids Res. 43 (2015) D784–D788.

[2] M.E. Riccio, S. Buhler, J.M. Nunes, C. Vangenot, M. Cuénod, M. Currat, D. Di, M. Andreani, M. Boldyreva, G. Chambers, M. Chernova, J. Chiaroni, C. Darke, J. Di Cristofaro, V. Dubois, P. Dunn, H.A. Edinur, N. Elamin, J.F. Eliaou, Z. Grubic, T. Jaatinen, U. Kanga, B. Kervaire, L. Kolesar, W. Kunachiwa, M.L. Lokki, N. Mehra, G. Nicoloso, R. Paakkanen, D.P. Voniatis, C. Papasteriades, F. Poli, L. Richard, I. Romón Alonso, A. Slavčev, G. Sulcebe, T. Suslova, M. Testi, J.M. Tiercy, A. Varnavidou, B. Vidan-Jeras, A. Wennerström, A. Sanchez-Mazas, 16(th) IHIW: analysis of HLA population data, with updated results for 1996 to 2012 workshop data (AHPD project report), Int. J. Immunogenet. 40 (2013) 21–30.

[3] H.A. Edinur, P.P. Dunn, L. Hammond, C. Selwyn, P. Brescia, M. Askar, P. Reville, Z.M. Velickovic, R.A. Lea, G.K. Chambers, HLA and MICA polymorphism in Polynesians and New Zealand Maori: implications for ancestry and health, Hum. Immunol. 74 (2013) 1119–1129.

[4] A. Sanchez-Mazas, V. Cerny, D. Di, S. Buhler, E. Podgorna, E. Chevallier, L. Brunet, S. Weber, B. Kervaire, M. Testi, M. Andreani, J.-M. Tiercy, J. Villard, J.M. Nunes, The HLA-B landscape of Africa: signature of pathogen-driven selection and molecular identification of candidate alleles to malaria protection, Mol. Ecol. 26 (2017) 6238—6252.

[5] A.K. Lancaster, R.M. Single, O.D. Solberg, M.P. Nelson, G. Thomson, PyPop update–a software pipeline for large-scale multilocus population genomics, Tissue Antigens 69 (2007) 192—197.