Tissue Antigens ISSN 0001-2815

Population samples and genotyping technology

S. J. Mack[1,2], A. Sanchez-Mazas[3], R. M. Single[4], D. Meyer[5], J. Hill[1,2], H. A. Dron[1,2], A. J. Jani[1,2], G. Thomson[6] & H. A. Erlich[1,2]

1 Children's Hospital Oakland Research Institute, Oakland, CA, USA
2 Roche Molecular Systems, Alameda, CA, USA
3 Laboratory of Anthropology, Genetics and Peopling history, Department of Anthropology and Ecology, University of Geneva, Switzerland
4 Department of Mathematics and Statistics, University of Vermont, Burlington, VT, USA
5 Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, Brazil
6 Department of Integrative Biology, University of California, Berkeley, CA, USA

Key words
anthropology; biostatistics; population genetics; Major Histocompatibility Complex database; International Workshop

Correspondence
Steven J. Mack
Department of Human Genetics
Roche Molecular Systems
Children's Hospital Oakland Research Institute
1145 Atlantic Avenue
Alameda, CA
94501-1155
USA
Tel: 1 510 814 2949
Fax: 1 510 522 1285
e-mail: steve.mack@contractors.roche.com

doi: 10.1111/j.1399-0039.2006.00768.x

Abstract

The 14th International HLA (human leukocyte antigen) Immunogenetics Workshop (14th-IHIWS) Biostatistics and Anthropology/Human Genetic Diversity project continues the population sampling, genotype data generation, and biostatistic analyses of the 13th International Histocompatibility Workshop Anthropology/Human Genetic Diversity Component, with the overall goal of further characterizing global HLA allele and haplotype diversity and better describing the relationships between major histocompatibility complex diversity, geography, linguistics, and population history. Since the 13th Workshop, new investigators have and continue to be recruited to the project and new high-resolution class I and class II genotype data are being generated for 112 population samples from around the world.

Introduction

The Biostatistics and Anthropology/Human Genetic Diversity project of the 14th International HLA (human leukocyte antigen) Immunogenetics Workshop (14th-IHIWS) is a continuation of the research efforts begun by the International Histocompatibility Working Group (IHWG) Anthropology/Human Genetic Diversity Component (AHGDC) as part of the 13th International Histocompatibility Workshop (13th-IHWS).

The primary goals of the 13-IHWS AHGDC were 1) to generate high-resolution genotype data for a global sampling of human populations at the HLA-A, C, B, DRB1, DQA1, DQB1, DPA1, and DPB1 loci, 2) to estimate multilocus haplotype frequencies and linkage disequilibria values from these data, 3) to carry out population genetic analyses of these allele, genotype and haplotype data, with the aim of characterizing the nature and scope of global allele and haplotype diversity, 4) to investigate the relationships between major histocompatibility complex (MHC) diversity, geography, linguistics, and population history, and 5) to make these data and analyses available to the scientific community. Genotype data were generated and analyzed for 95 populations (representing ~12,212 individuals) as described in the 13th-IHWS AHGDC Joint Report (13th-IHWS data set) (1).

The project goals for the 14th-IHIWS Biostatistics and Anthropology/Human Genetic Diversity project were extended to include 6) increased participation from investigators in underrepresented areas of the world, 7) incorporation of new populations (primarily from underrepresented regions of the world) into the study, 8) completion of class I and class II genotyping of 13th-IHWS populations, and 9) development of tools to facilitate amino-acid level population genetic analyses of MHC diversity.

Here, we describe progress made toward goals 6–8 in the 3.5 years that have passed since the 13th-IHWS.

Participation

Forty investigators participated in the 13th-IHWS AHGDC by providing samples and/or performing high-resolution genotyping. The geographic distribution of 13th-IHWS investigators is shown in Table 1, along with the geographic distribution and IHWG labcodes of 14th-IHIWS AHGDC investigators. Australia, South America, Africa, and Southeast Asia were poorly represented among IHWS investigators. Recruitment of participants in these regions, with the twin aims of high-resolution genotyping technology transfer and increased representation of population samples from these regions of the world in the overall data set, is ongoing. During the 14th-IHIWS phase of the project in the past 3.5 years, eight 13th-IHWS investigators contributed new data, and eight new 14th-IHIWS investigators from South Africa, France, Norway, the United Kingdom, China, Taiwan, the United States, and Colombia joined the project.

Table 1 14th-IHIWS Anthropology project participation by nation

Continent	Nation	# of 13th-IHWS investigators	14th-IHIWS investigators
Africa	South Africa	1	ZAFDUT
Europe	Bulgaria	1	**BGRNAU**
	Czech Republic	1	
	Croatia	1	
	Finland	1	
	France		FRADUT
	Italy	**2**	
	Macedonia	1	**MKDSPI**
	Norway		NORTSB
	Slovenia	1	
	Spain	1	
	Switzerland	2	
	United Kingdom	1	UKIBOD
Asia	China		CHNLEE, CHNANX
	India	3	**INDMEH**
	Israel	2	
	Japan	2	
	Korea	1	
	Taiwan	1	**TWNLIN**
	Turkey	1	
	Russia	1	
	Vietnam	1	**VNHNAN**
North America	Canada	2	**CANSHT**
	Mexico	1	
	United States	**9**	**USAERL**, USAHNH
South America	Brazil	2	
	Colombia		COLANA
	Venezuela	1	**VENLAY**
Total investigators		40	16

Investigators that have contributed data specifically to the 14th-IHIWS Anthropology project are shown in the far right column. Of these, those that also contributed to the 13th-IHWS Anthropology project are presented in bold.

13th-IHWS, 13th-International Histocompatibility Workshop; 14th-IHIWS, 14th-International HLA (human leukocyte antigen) Immunogenetics Workshop.

Population sampling

The genotype, demographic, allele frequency, and haplotype (A:B:C, A:B:DRB1, DRB1:DQB1, and DQA1:DQB1) frequency data for the 13th-IHWS data set were made available to the public on the MHC database (dbMHC) in 2003 (2). These data have already proven to be useful for novel studies of MHC selection and diversification (3). In addition, these allele and haplotype frequency data are also available through the AlleleFrequencies.net database (4). Analytical results for this 13th-IHWS data set, describing the results of Hardy–Weinberg testing, tests of selection using the Ewens–Watterson Homozygosity statistic, and linkage disequilibrium values will appear on the dbMHC later in 2006.

For the 14th-IHIWS Anthropology project, new genotype data (for either entirely new population samples, or for individuals that were not included or loci that were not typed in 13th-IHWS data set populations) are being generated for approximately 60 population samples [not including the CEPH-Human Genome Diversity Project (CEPH-HGDP) populations, see below], representing ~11,000 individuals (14th-IHIWS data set). All samples in these populations have been obtained in accordance with applicable laws and regulations at the investigator's institution regarding human subjects research, including any required informed consent for prospective research use. In addition, the identifiers associated with each sample have been coded for anonymity in order to protect the confidentiality of the individual sampled. These populations are described in Table 2.

The 52 populations included in the CEPH-HGDP (5) cell panel have been genotyped at high resolution for class I and II loci as part of the 14th-IHIWS project. While it will not be possible to subject these CEPH-HGDP population samples to every analysis, because the sample size for CEPH-HGDP populations is low (mean $N = 20.4$), these CEPH-HGDP populations have been typed for a large number of non-HLA and non-MHC loci, permitting seven comparisons between HLA and non-HLA loci which have not previously been possible.

The 13th-IHWS data set included a high-density sampling of Aboriginal Taiwanese populations. Analyses of these population samples showed extreme allele and haplotype frequency diversification for these populations (1). The 14th-IHIWS data set includes high-density

Table 2 Population samples genotyped as part of the 14th-IHIWS

Region	Labcode	Population	2*n*	Geographic origin
Sub-Saharan Africa	ZAFDUT	Xhosa	50	South Africa
Europe	BGRNAU	**Roma (Gypsy)**	34	Europe
	MKDSPI	Macedonian	214	Macedonia
Southwest-Asia	INDMEH	Jat Sikhs	100	North India
		Kashmiri Brahmins	100	North India
		Jarwaras	100	North India
	USAERL	Bangladesh	92	Bangladesh
		South Indian	60	Tamil Nadu
		Iran	50	Caucasus
		Kaimykian	50	Caucasus
		Turk	50	Caucasus
		Azerbaidschan	50	Caucasus
		Abchazian	50	Caucasus
		Armenian	50	Caucasus
		Kurd	50	Caucasus
		Ossetian	50	Caucasus
		Darginian	50	Caucasus
		Abazinian	50	Caucasus
		Kubadian	50	Caucasus
		Georgian	50	Caucasus
		Ingushian	50	Caucasus
		CEPH-HGDP	1064	See Text
Southeast Asia	CHNLEE	Zhijiang (Han)	2500	China
	TWNLIN	**Aboriginal Taiwanese**[a]	650	Taiwan
		Chinese (Han)[a]	100	China
	VNHNAN	**Muong**	107	Vietnam
		Kinh	110	Vietnam
	CHNANX	Shui	153	China
		Bouyei	109	China
		Miao	85	China
		Yao	66	China
		Lisu	111	China
		Nu	107	China
		Lahu	72	China
		Daur	100	China
		Sangdung (Han)	98	China
		Maizhan Canton (Han)	80	China
	USAHNH	Chinese	3600	Singapore
Central Asia	UKIBOD	Uzbek	80	Uzbekistan
Oceania	NORTSB	Polynesian	50	Easter Island
North America	MEXGOR	Tarahumaras	100	Mexico
		Mixtecs	100	Mexico
		Lacondones	100	Mexico
		Seris	100	Mexico
Admixed	VENLAY	Curiepe (African admixed)	150	Venezuela
	COLANA	European admixed	100	Madellin-Columbia
	FRABER	African-European admixed	100	Martinique
	USAHNH	African American	564	United States
	ZAFDUT	Cape Colored (African-European admixed)	50	South Africa
Total			12,006	

Populations in boldface are also part of the 13th-IHWS data set; either new individuals have been typed or typing has been extended to include additional loci.
[a] Population name includes multiple discrete population samples.
13th-IHWS, 13th-International Histocompatibility Workshop; 14th-IHIWS, 14th-International HLA (human leukocyte antigen) Immunogenetics Workshop.

samplings of populations from Mainland China, India, and the Caucasus region, in addition to extended genotyping of 13th-IHWS Aboriginal Taiwanese populations. These additional high-density samplings from other regions of the world will provide some context for assessing the results observed in the Aboriginal Taiwanese populations.

As with the 13th-IHWS data set, these 14th-IHIWS genotype and demographic data, allele and haplotype frequency data, and analyses will be made available to the public through online databases such as the dbMHC (2) and AlleleFrequencies.net (4) after participants have had opportunity to publish their findings.

Genotyping technology

In order to facilitate standardized high-resolution class I genotyping for the 13th-IHWS, HLA class I (A, B, and C) reverse-format immobilized sequence-specific oligonucleotide probe arrays (also known as reverse line strip or RLS reagents) were distributed to Anthropology project participants between 2000 and 2002 (6). With the goal of facilitating high-resolution genotyping at class I and class II loci, high-resolution RLS typing reagents for the HLA-A, B, C, DQA1, DQB1, DPA1, and DPB1 loci were distributed to 14th-IHIWS Anthropology project investigators through 2005. The protocols for the class I RLS reagents are the same as described for the RLS reagents used in the 13th-IHWS (6). However, the lists of class I alleles detected using these reagents were updated to include alleles in the March 2003 nomenclature update of the World Health Organization Nomenclature Committee for Factors of the HLA System (7). The March 2003 nomenclature update was used to create the detected allele lists for the class II RLS reagents.

The DQ locus and DP locus RLS consist of probes for both the α and β loci on single strips (i.e. a DQA1/B1 strip and a separate DPA1/B1 strip). The DQA1/B1 strip includes 37 probes, and the DPA1/B1 strip includes 48 probes. For each locus (DQ or DP), the α and β loci are amplified in a single multiplex PCR, so that the use of these co-amp reagents permits four HLA loci to be genotyped in two PCR reactions for each sample.

A two-step amplification/detection system was distributed for DRB1 genotyping. In this system, a 29-probe 'DRB general' strip is used to identify the DRB1 allelic lineages (DR1, DR15, DR16, DR3, DR11, DR13, DR14, DR4, DR7, DR8, DR12, DR9, or DR10) present in a given sample after a DRβ-specific PCR. Allelic lineage-specific primers are used to amplify individual alleles in a second round of PCRs, and a 31-probe DRB1-specific subtyping strip is used to identify the DRB1 alleles in the individually amplified allelic lineages.

14th-IHIWS investigators using the RLS reagents were required to genotype a quality control panel consisting of purified DNA from 15 cell lines (for which the HLA-A, B, C, DRB1, DQA1, DQB1, DPA1, and DPB1 genoptypes had been determined via sequencing-based typing methods) with 93% accuracy for each locus.

While these RLS reagents were made available for all 14th-IHIWS participants, their use was not mandatory, and participants were free to use alternative genotyping methods, provided that high-resolution genotyping results were generated.

Continuing efforts

The work described in this joint report is a continuation of work that was begun in previous workshops, and that will continue in future workshops. Sample collection, genotyping, and data analyses are ongoing, and this report should be considered as an interim progress report; specific 14th-IHIWS data and analyses will be presented and reported as they become available.

Acknowledgments

We thank 14th-IHIWS investigators Anlong Xu, Begum Pillay, Carolyn Hurley, Chen-Chung Chu, Clara Gorodezky, Elissaveda Naumova, Erik Thorsby, Ernette du Toit, John Tsun-Dao Lee, Juan-Manuel Anaya, Marie Lin, Milena Ivanova, Mirko Spiroski, Naridner Mehra, Odille Bera, Sir Walter Bodmer, Susan Tonks, Vu-Trieu-An, and Zulay Layrisse for their collaboration on this project. This work has been supported by National Institutes of Health (USA) grants AI49213 and GM35326, United States Department of Energy grant DE-FG02-00ER45828, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Brazil) grant 03/01583-8, and Swiss National Foundation (Switzerland) grants 3100-49771.96 and 3100A0-112651.

Conflict of Interest Statement

All authors have declared no conflicts of interests.

References

1. Mack SJ, Erlich HA. Anthropology Joint Report. In: Hansen JA, ed. *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference, Volume I. Seattle: IHWG Press, 2007.*
2. http://www.ncbi.nlm.nih.gov/mhc/ihwg.cgi?ID=9&cmd=PRJOV.
3. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol* 2005: **15**: 1022–7.
4. http://www.allelefrequencies.net/.
5. Cann HM. Human genome diversity. *C R Acad Sci III* 1998: **321**: 443–6.
6. Hurley C. IHWG Technology Core Joint Report. In: Hansen JA, ed. *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference, Volume I. Seattle: IHWG Press, 2007.*
7. SGE Marsh. Nomenclature for factors of the HLA system. Update March 2003. *Tissue Antigens* 2003: **62**: 91–2.