MOLECULAR APPROACHES TO TISSUE TRANSPLANTATION AND TRANSFUSION MEDICINE IN MAORI AND POLYNESIANS IN NEW ZEALAND

BY

EDINUR HISHAM ATAN

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

in Cell and Molecular Bioscience

Victoria University of Wellington

2013

Abstract

My thesis is focussed on using new and existing molecular technologies to survey immune system (i.e. HLA, KIR and MICA), blood group and HPA genes in Polynesian and Maori subjects. I have compiled large datasets for these individuals including first ever molecular typing blood group, HPA and MICA. I have also compared my data with those from previous genetic studies of KIR and HLA for the four Polynesian sub-populations and HLA for Maori. Combined with previous information in a small scale meta-analysis, the results presented here were interpreted for the first time in the light of new theories about the hybrid origin of Proto-Polynesians via gender-biased gene flow between Austronesian-speaking Mongoloids and Papuan-speaking Australoids. Phylogenetic and principal component/coordinate analyses show that Polynesian sub-populations are closer to each other and are end populations of the great Austronesian Diaspora. Estimation based on HLA class I and II allele frequencies show a wide range (55:45 to 90:10) of ASM:P-SA ancestral fractions in Maori and other Polynesians. The same phenomenon was observed for blood group, HPA and KIR data. Thus, findings from the present study generally support the account deduced from Y-chromosome, mtDNA and genome wide SNP data, and are in accord with the emergent picture from linguistics and archaeology studies.

The data collected bear more or less directly on questions of health including tissue matching for transfusion and transplant surgery, disease resistance and the relative incidence of autoimmune disease in these ethnic groups. Tissue matching of HLA, MICA, KIR, blood group and HPA is a fundamental element in transfusion and transplant surgery and it is vital that these technologies take accurate account of the ethnic origins of both donors and recipients. This is important due to significant differences between Polynesians and people of European origin, who form the majority segment of the modern day multi-ethnic country of New Zealand. The prevalence of HLA, MICA, KIR, blood group and HPA and their potential influence on the health of modern day Maori and Polynesians were also discussed.

This study was intentionally designed to validate new molecular blood group and HPA typing methods in Polynesian and Maori populations. Scores for the other loci are based on the application of pre-existing technologies. Validation using three different blood group and HPA genotyping methods (i.e. PCR-SSP, SBT and SNP assay) gave 100% concordance and thus reflects the accuracy of the new DNA-based methods. Overall, molecular genotyping does seem to offer the opportunity to improve the efficiency of high quality transfusion services, but should include all known variants that are associated with serological expression.

In conclusion, the present molecular genetics study has successfully compiled new HLA, MICA, blood groups, HPA and KIR datasets for Polynesians and Maori, used the datasets for dual analyses of ancestry and health, and demonstrated the reliability of current genotyping platforms that would suit well the needs of transfusion and transplantation centres. In general, this thesis has strongly endorsed the indigenous biomedicine approach, and particularly will be valuable to Asia-Pacific populations of Austronesian peoples.

Acknowledgements

I wish to take this opportunity to express my deep appreciation to my supervisor Dr Geoffrey Keith Chambers. He has helped me to plan a very good research framework for my PhD, continues to monitor my research progress closely and gives prompt answers to any questions forwarded to him. I am very grateful by the support, consideration, advice and suggestions given by my other supervisors and collaborators, Dr Paul Dunn (New Zealand Blood Service), Dr Rod Lea (Griffith University, Australia), Dr Paul Norman (Stanford University), Dr Peter Parham (Stanford University), Dr Zlatibor Velickovic (University of Sydney, Australia) and Dr Medhat Askar (Cleveland Clinic, Ohio).

My PhD has been supported by a wide range of New Zealand, Malaysian and United States of America agencies. I acknowledge with thanks the support given by the School of Biological Sciences, Victoria University of Wellington, Universiti Sains Malaysia, Ministry of Higher Education, Malaysia, The Tissue Typing Laboratory of the New Zealand Blood Service, Auckland, The Department of Structural Biology, Stanford University of Medicine and Allogen Laboratories and the Transplant Center, Cleveland Clinic, Ohio.

I would like to thank all the participants, local Maori and Polynesian communities including Te Iwi o RakaiPaaka tribal members, New Zealand Central Region Ethics Committee and Victoria University Human Ethics committe for their involvement and approval. I extend my particular thanks for administrative services provided by Paul Marsden (Victoria University), Patricia Stein (Victoria University) and Mary Murray (Victoria University) and technical assistance from Maximiliano Crosetti (Progenika Biopharma S.A and Grifols International), William Perry (Grifols Australia), Martina Prager (BAG Health Care GmbH, Lich, Germany), Gabi Henschel (BAG Health Care GmbH, Lich, Germany), Dr Jordi Bozzo (Progenika Biopharma S.A and Grifols International) and Dr Abeer Madbouly (United States National Marrow Donor Program).

Special thanks are expressed to Dr Josep Azar, Phil Servid, Kosala Sirisena, Amy Hoo, Shalen, Sebastian, Miles, Olivia, Heather Constable, Aashi, Pop, Cong Zong, Maheshini, Leighton Thomas, Zurin, Wan Joe, Mamat, MWPGS and all others that have kept me happy, healthy and rational[©].

To my wife, Tengku Puteri Nadiah Tengku Baharuddin Shah, my parents, Atan Mohamed and Zawiah Mohamad, my beloved kids, Puteri Elena Aisya and Putera Eiman Shah...thank you for your love and all I have is yours.

List of abbreviations

А	adenine
A-SM	Austronesian-speaking Mongoloid
bp	base pair
С	cytosine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic Acid
dNTP	deoxynucleotide triphosphate
EDTA	ethylene diamine tetra acetic
G	guanine
H ₂ O	water
GVHD	graft versus host disease
HLA	human leukocye antigen
HPA	human platelet antigen
HWE	Hardy-Weinberg equilibrium
kb	kilo base
KIR	killer cell immunoglobulin-like receptor
LD	linkage disequilibrium
Μ	molar
MgCl ₂	magnesium chloride
MHC	major histocompatibility complex
MICA	major histocompatibility complex class I chain-related gene A
mM	millimolar
mtDNA	mitochondrial DNA

ng	nanogram
NJ	neigbor-joining
PCA	principle component
РСО	principle coordinate
PCR	polymerase chain reaction
P-SA	Papuan-speaking Australoid
psi	pound force per square inch
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	revolution per minutes
Т	thymine
Taq	Thermus aquatic
SBT	sequence based typing
SNP	single nucleotide polymorphism
SSOP	sequence specific oligonucleotide probe
SSP	sequence specific primer
TBE	tris-borate-ethylene diamine tetra acetic acid
U	unit
μg	microgram
μl	microliter
UV	ultra violet
V	volt
Y-STRs	Y-chromosome short tandem repeats

List of related publications and presentations

A. Journals

- Chambers, G.K. & Edinur, H.A. (2013) Genetic relationships between Malays and Maori. In: *The changing values of Malays, Maori and Pacific Islanders* (eds. Rahman W.R.A. & Higgins T.R.), 6-37. Chair of Malay Studies & Victoria University Foundation, Wellington, New Zealand. ISBN: 9780475124005.
- Edinur, H.A., Dunn, P.P. J., Hammond, L., Selwyn, C., Velickovic, Z.M, Lea, R.A. & Chambers, G.K (2012) Using HLA loci to inform ancestry and health in Polynesian and Maori populations. *Tissue Antigens*, 80, 509-522.
- Riccio, M.E., Buhler, S., Nunes, J.M., Vangenot, C., Cuenod, M., Currat, M., Di, D., Andreani, M., Boldyreva, M., Chambers, G., Chernova, M., Chiaroni, J., Darke, C., Di Cristofaro, J., Dubois V., Dunn, P., Edinur, H.A., Elamin, N., Eliaou, J.-F., Grubic, Z., Jaatinen, T., Kanga, U., Kervaire, B., Kolesar, L., Kunachiwa, W., Lokki, M.L., Mehra, N., Nicoloso, G., Paakkanen, R., Voniatis D.P., Papasteriades, C., Poli, F., Richard, L., Alonso I.R., Slavcev, A., Sulcebe, G., Suslova, T., Testi, M., Tiercy J.-M., Varnavidou, A., Vidan-Jeras, B., Wennerstrom, A. & Sanchez-Mazas, A. (2013) 16th IHIW: Analysis of HLA Population Data, with updated results for 1996 to 2012 workshop data (AHPD project report). *International Journal of Immunogenetics*, 40, 21-30.
- Edinur, H.A., Dunn, P.P.J., Lea, R.A. & Chambers, G.K. (2013) Human platelet antigens frequencies in Maori and Polynesian populations. *Transfusion Medicine*, 23, 330-337.

- Edinur, H.A., Dunn, P.P.J., Hammond, L., Selwyn, C., Askar, M., Velickovic, Z.M., Lea, R.A. & Chambers, G.K. (2013) HLA and MICA polymorphism in Polynesians and New Zealand Maori: Implications for ancestry and health. *Human Immunology*, 74, 1119-1129.
- Edinur, H.A., Dunn, P.P.J., Lea, R.A. & Chambers, G.K. (2013) Molecular approaches to transfusion medicine in Polynesians and Maori in New Zealand. *International Journal of Immunogenetics. In press.*

B. Oral and poster presentations

- Atan, E.H., Lea, R.A., Dunn, P. & Chambers, G.K. (2010) HLA class I antigens in Polynesia, Wellington Health and Biomedical Research Society, 79th Scientific Meeting, Great Hall, Museum Building, Massey University, Wellington, 23 August 2010. *Poster presentation*.
- Edinur, H.A. (2010) HLA class I polymorphism in Maori and Polynesians. Ecology and Evolution Seminar Series 2010, Victoria University of Wellington, New Zealand, 17 September 2010. Oral presentation.
- Edinur, H.A., Lea, R.A., Dunn, P. & Chambers, G.K. (2010) HLA polymorphism in Polynesia. Asia-Pacific Histocompatibility & Immunogenetic Association Conference 2010, Queentown, New Zealand, 16-19 November 2010. *Poster presentation*.
- Chambers, G.K. & Edinur, H.A. (2012) Genetic relationships between Malays and Maori. The 3rd International Seminar: Malay, Maori and Pacific language, art and culture: "The

changing values of Malays, Maori and Pacific Islanders", 23 & 24 March 2012, Te Herenga Waka Marae, Kelburn Parade, Victoria University of Wellington, Wellington, New Zealand. Oral presentation.

- Edinur, H.A., Chambers, G.K., Dunn, P.P. J., Hammond, L., Selwyn, C., Lea, R.A. & Chambers, G.K (2012) Using HLA loci to inform ancestry and health in Polynesian and Maori population. *Tissue Antigens*, **79**, 557.
- 6. Dunn, P., Fisher, R., Darke, C., Cowie, H., Edinur, H.A., Chambers, G.K & Blacklock, H. (2012) HLA alleles and haplotypes & recruitment in five Polynesian populations in the New Zealand Bone Marrow Donor Registry. The Joint 16th International HLA and Immunogenetics Conference/26th European Immunogenetics and Histocompatibility Conference/23rd British Society of Histocompatibility and Immunogenetics Conference, 28th May 3rd June 2012. *Poster Presentation.*
- Dunn, P.P.J., Edinur, H.A., Selwyn, C., Hammond, L. & Chambers, G.K. (2012) HLA alleles and haplotypes in Maori and Polynesians: clues for ancient migration and modern day health. APHIA Conference, November 2012, Adelaide, Australia. *Poster Presentation*.
- Dunn, P.P.J., Edinur, H.A., Hammond, L., Selwyn, C., Brescia, P., Askar, M., Reville, P., Velickovic, Z.M., Lea, R.A. & Chambers, G.K. (2013) High resolution HLA and MICA alleles and haplotypes in Maori and Polynesians. *Tissue Antigens*, 81, 299.

 Aziz, N.A., Edinur, H.A., Dunn, P.P.J., Hammond, L., Selwyn, C., Brescia, P., Askar, M., Reville, P., Velickovic, Z.M., Lea, R.A. & Chambers, G.K. (2013) HLA and MICA alleles and haplotypes: disease association in Maori and Polynesians. *Tissue Antigens*, 81, 368.

Table of Contents

Abstract	i
Acknowledgements	iii
List of abbreviations	v
List of related publications and presentations	vii
Table of Contents	xi
List of Tables	xvi
List of Figures	xviii
Declaration	XX
Chapter 1: General Introduction	1
1.1 Transplantation medicine	1
1.2 Transfusion medicine	4
1.3 Genetic testing in transplantation and transfusion medicine	5
1.4 Donor recruitment and applications of population data	6
1.5 Overview of the thesis layout	8
1.6 Genetic markers	14
1.6.1 Human leukocyte antigen (HLA)	14
1.6.2 Major histocompatibility complex class I chain-related gene A (MICA)	16
1.6.3 Blood group	17
1.6.3.1 ABO	19
1.6.3.2 Rhesus	19
1.6.3.3 MNS	20
1.6.3.4 Kell	20
1.3.3.5 Kidd	21

1.6.3.6 Duffy	22
1.6.3.7 Dombrock	23
1.6.3.8 Colton	23
1.3.3.9 Diego	23
1.6.4 Human platelet antigen (HPA)	24
1.6.5 Killer cell immunoglobulin-like receptor (KIR)	25
1.7 Objectives	29
1.8 Hypotheses	30
References	31
Chapter 2: Extended information on Materials and Methods	49
2.1 Banked DNA sample	49
2.2 General laboratory practice	50
2.3 DNA quantification	50
2.4 Low resolution typing of HLA, MICA and KIR loci	50
2.5 High resolution typing of HLA loci	51
2.6 PCR-SSP typing of ABO, Rhesus, MNS Kell, Kidd, Duffy and HPA	51
2.7 Sequence based typing of Kell, Duffy, JK_{null} and HPA-6	52
2.8 Blood group and HPA genotyping using chip hybridization SNP assay	53
2.9 Statistical analysis	54
2.9.1 Allele, gene and carrier frequencies	54
2.9.2 Haplotype frequency	54
2.9.3 Exact tests for Hardy-Weinberg equilibrium	54
2.9.4 Linkage disequilibrium	55
2.9.5 Testing for natural selection	55
2.9.5.1 Deviation from HWE	55

2.9.5.2 The Ewens-Watterson neutrality test	56
2.9.5.3 T-test	56
2.9.5.4 Ancestral fraction scaling test	56
2.9.6 Exact test of population differentiation	57
2.9.7 Phylogenetic analysis	57
2.9.8 Principal coordinate and component analysis	58
2.9.9 Probability of an exact match	58
2.9.10 Probability of alloimmunization	59
References	60
Chapter 3.1: Using HLA to inform ancestry and health in Polynesian and	
Maori populations	62
Abstract	63
Introduction	64
Materials and methods	65
Results	67
Discussion	84
Acknowledgements	91
References	92
Supplementary information	97
Chapter 3.2: HLA and MICA polymorphism in Polynesians and Maori: implications	
for ancestry and health	110
Abstract	111
Introduction	112
Materials and methods	113
Results	116

xiii

	Discussion	129
	Acknowledgements	138
	References	139
	Supplementary information	148
Chapter 3	.3: Molecular approaches to transfusion medicine in Polynesians and	
	Maori in New Zealand	157
	Summary	158
	Introduction	160
	Materials and methods	161
	Results	165
	Discussion	174
	Conflict of interest	182
	Acknowledgements	182
	References	183
	Supplementary information	190
Chapter 3	.4: Human platelet antigens frequencies in Maori and Polynesian populations	195
	Summary	196
	Introduction	197
	Materials and methods	199
	Results	201
	Discussion	208
	Acknowledgements	214
	Conflict of interest	215
	References	216
	Supplementary information	221

Chapter 3.5: KIR gene content in New Zealand Maori and Polynesians	225
Abstract	226
Introduction	227
Materials and methods	228
Results	230
Discussion	237
Acknowledgements	241
Funding	241
Conflict of interest	242
References	243
Supplementary information	249
Chapter 4: General discussion and conclusion	271
References	278
Appendix 1: Genetic relationships between Malays and Maori	281
Appendix 2: Analysis of the 16th IHIW HLA Population Data, with updated	
results for 1996 to 2012 workshop data (AHPD project report)	321

List of Tables

Chapter 1

Table 1: Blood groups and their molecular bases	18
Table 2: HPA systems and their molecular bases	24
Table 3: KIRs and their ligands	28
Chapter 3.1	
Table 1: HLA allele frequencies for Polynesians and Maori	72
Table 2: HLA haplotypes with respective value of LD	76
Table S1: Exact test of Hardy-Weinberg equilibrium	97
Table S2: Likelihood-ratio test of linkage disequilibrium	97
Table S3: Exact test of population differentiation	98
Table S4: HLA haplotypes in Polynesian and Maori datasets	99
Table S5: Simulations of A-SM:P-SA fractions in Polynesians	101
Chapter 3.2	
Table 1: Frequency of HLA and MICA alleles in present study	121
Table 2: HLA and MICA haplotypes	124
Table S1: Selection test on Maori and Polynesian datasets	148
Table S2: Frequency of HLA and MICA alleles for Polynesian datasets	149
Table S3: Exact tests on Polynesian and Maori datasets	155
Chapter 3.3	
Table 1: ABO and Rhesus phenotype frequencies for Rakaipaaka Iwi	168
Table 2: Blood group phenotype frequencies and probability of match	169
Table S1: Blood group variants captured by molecular technology	190
Table S2: PCR protocols for SBT	192

Table S3: Exact tests on Maori datasets	193
Table S4: Exact tests on Polynesian and European datasets	193
Table S5: Comparison of phenotype prediction	194
Table S6: Summary of the costs associated with molecular methods	194
Chapter 3.4	
Table 1: HPA allele frequencies	203
Table S1: HPA genotype profiles and HWE analysis	221
Table S2: Homogeneity tests	222
Table S3: Probability of transfusion and gestation mismatch	223
Chapter 3.5	
Table 1: Phenotype and genotype frequencies	235
Table 2: Correlation between KIR and HLA carrier frequencies	237
Table S1a: LD analysis of KIR genes in TP group	249
Table S1b: LD analysis of KIR genes in PFA sub-group	250
Table S1c: LD analysis of KIR genes in TM group	251
Table S1d: LD analysis of KIR genes in MFA sub-group	252
Table S2: KIR carrier frequencies for Polynesian datasets	253
Table S3a: Homogeneity tests for KIR2DS4 and 3DS1	254
Table S3b: Homogeneity tests for KIR3DL1 and 2DL1	255
Table S3c: Homogeneity tests for KIR2DL5 and 2DS1	256
Table S3d: Homogeneity tests for KIR2DS2 and 2DS3	257
Table S3e: Homogeneity tests for KIR2DS5 and 3DL3	258
Table S3f: Homogeneity tests for KIR2DP1 and 3DP1	259
Table S3g: Homogeneity tests for KIR2DL3 and KIR2DL2	260
Table S4: KIRs and HLA ligands carrier frequencies	261

List of Figures

Chapter 1

Fig. 1: Genetic markers and their chromosomal locations	11
Fig. 2: Diffusion route of A-SM to Remote Polynesia	13
Fig. 3: Schematic diagram of HLA class I and II antigenic presentation	16
Fig. 4: Schematic diagram of KIR haplotype A and B	27
Chapter 3.1	
Fig. 1: Graphs constructed using HLA class I allele frequencies	80
Fig. 2: Graphs constructed using HLA class II allele frequencies	81
Fig. 3: PCO plot (axis 1 and 2) for HLA class 1 allele frequencies	82
Fig. 4: PCO plot (axis 1 and 2) for HLA class II allele frequencies	83
Fig. S1: PCO plot (axis 1 and 3) for HLA class 1 allele frequencies	102
Fig. S2: PCO plot (axis 1 and 3) for HLA class II allele frequencies	103
Fig. S3: PCO plot - validation and comparison	104
Fig. S4: Ancestral fractions - HLA-A allele frequencies	105
Fig. S5: Ancestral fractions - HLA-B allele frequencies	106
Fig. S6: Ancestral fractions - HLA-C allele frequencies	107
Fig. S7: Ancestral fractions - HLA-DRB1 allele frequencies	108
Fig. S8: Ancestral fractions - HLA class I and II allele frequencies	109
Chapter 3.2	
Fig. 1: LD for HLA-B*-MICA* associations	126
Fig. 2: HLA and MICA allele frequencies - Polynesians, A-SM and P-SA	127
Fig. 3: PCO constructed using HLA class I and II allele frequencies	128
Fig. 4: Structure of <i>HLA-B*40:10</i> and its global distribution	133

xviii

Chapter 3.3

Fig. 1: Schematic diagram of blood group genotyping strategy	164
Fig. 2: Screening of Kell, Kidd and Duffy using PCR-SSP and SBT	171
Fig. 3: N-J tree - ABO, Rhesus and MNS phenotype frequencies	172
Fig. 4: PCO plot - ABO, Rhesus and MNS phenotype frequencies	173
Fig. 5: ABO phenotype frequencies - Polynesians, A-SM and P-SA	180
Fig.6: Rhesus phenotype frequencies - Polynesians, A-SM and P-SA	181
Chapter 3.4	
Fig. 1: Agarose gel electrophoresis of the HPA allele specific products	204
Fig. 2: Recent admixture at polymorphic HPA loci	205
Fig. 3: PCO plots - HPA-1 to -5 allele frequencies	206
Fig. 4: N-J tree - HPA-1 to -5 allele frequencies	207
Chapter 3.5	
Fig. 1: KIR genotype profile	233
Fig. 2: PCA constructed using KIR carrier frequencies	236
Fig. S1a: Correlation between KIR3DL1 and HLA-Bw4	262
Fig. S1b: Correlation between KIR3DL1 and and HLA-Bw4-80I	263
Fig. S1c: Correlation between KIR3DS1 and HLA-Bw4	264
Fig. S1d: Correlation between KIR3DS1 and HLA-Bw4-80I	265
Fig. S1e: Correlation between KIR2DL1 and HLA-C group 2	266
Fig. S1f: Correlation between KIR2DL2 and HLA-C group 1	267
Fig. S1g: Correlation between KIR2DL3 and HLA-C group 1	268
Fig. S1h: Correlation between KIR2DS1 and HLA-C group 2	269
Fig. S1i: Correlation between KIR2DS2 and HLA-C group 1	270

Declaration

"The submitted PhD thesis is an incorporation of articles/manuscripts as results chapters plus General Introduction, Extended information on Materials and Methods and General Discussion and Conclusion. Each chapter reflects the style of the journal to which the articles/manuscripts have been published or submitted. Therefore, the submitted thesis involves contributions from others and all of them are my PhD supervisors and collaborators:

Chapters 3.1:

H.E. designed the research, analyzed the data and wrote the paper, P.D. contributed to the research design, performed the research and writing the paper, L.H., C.S. and Z.M.V. performed the research, R.L. advised on statistics and edited the manuscript and G.K.C. helped to design and fund the research, managed the DNA Bank and obtained ethical approvals and he also contributed to preparation of the paper.

Chapters 3.2:

H.E. designed the research, analyzed the data and wrote the paper, P.D. contributed to the research design, performed the research and writing the paper, L.H., C.S., P.B., M.A., P.R. and Z.M.V. performed the research, R.L. advised on statistics and edited the manuscript and G.K.C. helped to design and fund the research, managed the DNA Bank and obtained ethical approvals and he also contributed to preparation of the paper.

H.E. designed the research, performed the research, analyzed the data and wrote the paper, P.D. contributed to the research design, performed the research and writing the paper, R.L. advised on statistics and edited the manuscript and G.K.C. helped to design and fund the research, managed the DNA Bank and obtained ethical approvals and he also contributed to preparation of the paper.

Chapters 3.5:

H.E. designed the research, performed the research, analyzed the data and wrote the paper, P.J.N. and P.P contributed to the research design, funded the research and advised on statistics, P.D. contributed to the research design and G.K.C. helped to design and fund the research, managed the DNA Bank and obtained ethical approvals and he also contributed to preparation of the paper.

I can confirm that I am the first author for articles/manuscripts included as results chapters and had the greatest input. Other related articles where I am a co-author are included as Appendices."

Chapter 1

General Introduction

Genetics has always been an implicit part of medicine and has now become explicit. Taking family histories guided practitioners for centuries, but genetic factors have only become main stream in the last 100 years basis for inheritance has become better understood. This era start with Archibald Garrod and inborn errors of metabolism and developed through case studies such as the role variants in beta-globin gene variants of sickle cell anaemia. The DNA era began using terminal restriction fragment length polymorphism (T-RFLP) to track markers that are associated with conditions such as Huntington's disease and cystic fibrosis. In more recent times, genome wide association studies (GWAS) are able to map out complex multifactorial diseases and there is now the possibility of universal genome scans for individuals. This thesis devoted to use new DNA-based technology in two areas which have seen particularly significant advances in 20th century; transplantation and transfusion medicine.

1.1 Transplantation medicine

Transplantation is the transfer of a graft (cell, tissue or organ) either from one part of the body to another (autograft) or between a donor and recipient pair of same species (allograft). Successful allograft transplantation depends on having the requisite level of genetic compatibility between donor and recipient, particularly for ABO blood groups and human leukocyte antigens (HLA). For HLA, rejection episodes are significantly reduced with decreasing numbers of allelic mismatched loci particularly for HLA-A, -B and -DRB1, and this requirement is extended to HLA-C and HLA-DQB1 in unrelated bone marrow transplants (Kalwak *et al.*, 2006, Sheldon and Poulton, 2006 and Doxiadis *et al.*, 2006).

Chapter 1

Typically, there are three different phases of rejection; hyperacute, acute and chronic. Hyperacute rejection usually occurs within minutes to hours immediately following transplantation and is due to circulating preformed antibodies (i.e. humoral immunity) against donor ABO and/or HLA. The pre-existing antibodies in the recipient can be produced by a previous transplantation, blood transfusion and pregnancy (Hopkins, 2006). Therefore, serological cross-matching between donor and recipient using panels of reactive antibodies prior to transplatation is of particular importance to avoid hyperacute rejection of the graft (Lydyard *et al.*, 2004).

Acute and chronic rejections may occur later due to HLA incompatibily between donor and recipient. Acute rejection is a form of cell-mediated immunity which takes place within weeks of engraftment (Wise and Carter, 2002 and Lydyard *et al.*, 2004). Two main mechanisms are involved during acute rejection; direct and indirect allorecognition by circulating T cells, respectively. Direct allorecognition involves T cell recognition of HLA molecules on donor antigen processing cells (APC), while indirect allorecognition involves antigenic presentation by HLA class I and II to CD4⁺ and CD8⁺ T cells, respectively (see Elgert, 2009, Pathak and Palan, 2005 and sub-section 1.6.1 of this chapter for more details). The allorereactivity remains in chronic rejection and is further impaired by the development of alloantibodies against the graft. This will be followed by graft loss over months to years after transplantation (Wise and Carter, 2002, Plant and Wood, 2009 and Hopkins, 2006).

The reverse situation occurs in allogenic bone marrow transplantation where the incompatible recipient tissue is attacked by donor immunocompetent T cells (aka graft versus host disease; GVHD). In bone marrow transplantation, patients usually undergo preparative treatments which include chemotheraphy and irradiation. These treatments not only eradicate residual

cancer/diseased cells, but also paralyse patient cell-mediated responses (Elgert, 2009). This condition will lead to direct allorecognition of grafted immunocompetent lymphocytes againts the genetically incompatible host.

In addition to HLA, the polymorphic major histocompatibility complex class I chain-related gene A (MICA) and killer cell immunoglobulin-like receptor (KIR) genes have also been shown to be clinically and practically important in transplantation. The MICA system produces stress-induced antigens expressed on the surface of keratinocytes, endothelial cells and monocytes. These antigens are recognized by NKG2D, an activating receptor for cytotoxic natural killer (NK) and CD8⁺ T cells (see Bauer et al., 1999 and Zwirner et al., 1999 and sub-section 1.6.2). Therefore, MICA is capable of stimulating cellular immunity againts transplanted grafts and/or able to trigger graft versus host reaction against mismatched MICA antigens possessed by donors and recipients (Zou and Stastny, 2010). In contrast, KIRs are not markers of tissue identity, but they do mediate self-recognition by NK cells through their interaction with HLA class I ligands (see sub-section 1.6.5 for details). For example, interaction of inhibitory KIR2DL1 and HLA-C group 1 allotypes provide tolerance to NK cells, while interaction between stimulatory KIR2DS1 and HLA-C group 1 ligands induce NK alloreactivity. The same mechanism occurs in allogenic transplantation, where allografts lacking compatible HLA ligands for inhibitory KIRs are attacked by the recipient's NK cells (Ludajic et al., 2009 and Rajalingam and Gebel, 2011). Nonetheless, missing ligands for KIR were shown to have beneficial in hematopoietic stem cell transplantation, including significantly reduced leukemia relapse and graft rejection (aka antileukemia effect). This was observed in HLA haplotype-mismatched (haplo-identical) transplants where recipient's leukemia cells are recognized (i.e. receptor-ligand mismatched) as foreign by the alloreactive donor's NK cells (Ruggeri et al., 2002).

Chapter 1

Overall, correct matching for blood group and HLA markers and more recently, for MICA and KIR variants will determine success rates in allogenic transplantation. Compared with blood group and HLA, the role of MICA and KIR is considered as a new development in the field of immunogenetics, as well as in transplantation (Bach and Amos, 1967, Moretta *et al.*, 1993, Bahram *et al.*, 1994). Future studies will fill some of the remaining gaps in our present understanding of the biological function of these markers and their synergy in transplantation medicine.

1.2 Transfusion medicine

Transfusion medicine is an area of multidisciplinary expertise; management of blood donation, blood component preparation, blood cell serology and blood transfusion therapy. It is the most common and successful form of tissue transplantation (Davenport and Mintz, 2007 and Elgert, 2009). Unlike other types of transplantation, blood contains little or no HLA class I (Elgert, 2009). However, blood does contain various types of other significant polymorphic antigens. These include the highly immunogenic ABO antigens, which cause hyperacute rejection of vascularised grafts such kidney and heart. Others including Rhesus, Kell, Dombrock, Kidd, MNS and platelet antigens can also be important (Anstee, 2009, Chinen and Bukley, 2011 and Brown and Navarrete, 2011). Routine serological matching is only required for ABO and RhD antigens (Blann, 2006 and Brown and Navarrete, 2011). But, in certain cases (e.g. refractoriness to platelet transfusion and transfused related acute lung injury), patients may require compatibility testing for these other blood group antigens (particularly Kell and Duffy) and may even extend to HLA and HPA matching for transfusion of blood component products (see Brown and Navarrete, 2011 and Chapter 3.4 for further details).

1.3 Genetic testing in transplantation and transfusion medicine

The major barriers to transplantation and transfusion are HLA and blood group antigens, although others, e.g. MICA, HPA and KIR can also contribute towards immunological reaction following transfusion and transplantation (see Chapters 3.2 to 3.5 for further details). The HLA and blood group antigens are targets of the immune system. Transplantation and transfusion success both rely on the level of compatibility between donor and recipient for these markers. Therefore, prior genetic testing plays a critical role for the matching process and this can be done using serological or molecular techniques.

Serological procedures (i.e. combination of HLA/blood phenotyping, antibody screening and crossmatching) still remain as fundamental elements in transplantation and transfusion medicine (see Doxiadis *et al.*, 2006 and Higgins, 2007). Nonetheless, molecular approaches are now being widely used for HLA and blood group genotyping and are already considered superior to serology-based techniques in several circumstances; e.g. genotyping of weakly expressed antigens and determination of blood groups and HLA types in multi-transfused and multi-transplanted patients (see Middleton, 2005, Sheldon and Poulton, 2006, Avent *et al.*, 2009 and Hustinx *et al.*, 2009). In addition, the advent of molecular methods has improved resolution of HLA and blood group polymorphisms to a precision that could not be achieved using serology alone (see Patnaik *et al.*, 2012 and Robinson *et al.*, 2013). Most of the HLA and blood group antigens are now defined at allelic levels, which are difficult to detect serologically. Often there are limited supplies of monoclonal antibody or perhaps none at all (Middleton *et al.*, 1988 and Schonitzer, 2009). The efficiency and accuracy of molecular genotyping is supported by the higher survival rates for unrelated transplantation with

matching at the allelic level, compared with serologically matched pairs (Sheldon and Poulton, 2006).

Overall, there is compelling evidence for the wider application of molecular approaches in transplantation and transfusion. The most accurate and reliable molecular technique is sequence based typing (SBT) as it shows exact variations in particular genes. In fact, SBT is now a 'gold standard' in several transplantation centres (Oliviera *et al.*, 2012). However, certain clinical circumstances require rapid and high throughput genotyping platforms; e.g. for heart and lung transplantation. Currently, hybridization SNP assays (e.g. Luminex and BLOODchip) seem set to become the accepted platforms for transfusion and transplantation services (reviewed by Veldhuisen *et al.*, 2009 and Wu and Csako, 2006).

1.4 Donor recruitment and applications of population data

Successful transplantations and transfusions also rely on the efficiency of pre-operative treatment and donor recruitment, plus post-operative care. Although posttransplantation/transfusion complications can be managed by immunosuppressive therapy (e.g. corticosteroids and cyclosporin), this strategy is usually associated with long term adverse effects, such as infection and drug-specific toxicity (Lawson, 2006 and Higgins, 2007). Therefore, careful donor recruitment seems to be the best way to avoid such complications.

Bone Marrow Donors Worldwide (BMDW) is a good example of an international collaboration that organizes recruitement of HLA-typed donors for unrelated hematopoietic stem cell transplantation. The BMDW database consists of 20,966,656 volunteer stem cell

6

donors and 569,598 units of cord blood are available for transplantation (www.BMDW.org, accessed 14 April 2013). However, majority of the donors recruited are of European origin. This was demonstrated by finding lower hit rate for New Zealand (NZ) Maori and Polynesians than for European patients when the database was searched for HLA-A, -B and -DRB1 matches (Zlatibor and Carter, 1999). Therefore, it has become a policy of the NZ Bone Marrow Registry (NZBMDR) to recruit increased numbers of of Maori and Polynesian donors as they are presently under-represented in BMDW (Dunn et al., 2012). In addition, the feasibility of finding suitable donors for particular patients is enhanced by identification, characterization and compilation of population data for medically important and highly polymorphic genes such as HLA, MICA, blood groups, HPA and KIR. This is because geographically or historically related populations tend to share common and unique alleles/haplotypes. For example, it is difficult to find a donor for a patient with a HLA-A*02:12-C*07:02-B*39:05-DRB1*08:02-DQA1*04:01-DQB1*04:02 haplotype in New Zealand, but it is quite widely distributed among Indigenous Americans. The presence of this haplotype in Polynesian patients is associated with the South American slave trade which operated across the Pacific during the 1800s (see Thorsby, 2012 and Chapter 3.2 for further details). This shows how population survey data can be used to narrow down the scope of searches for alleles/haplotypes that are not common in particular populations, but may be so in others.

Donor and recipient matching will remain as a fundamental element in transplantation and transfusion. Nonetheless, recent advances in molecular biology, immune therapy (e.g. gene transfer and immune suppressive drugs) and bio-engineering may one day provide better solutions (see Lawson, 2006 and Song *et al.*, 2013).

Chapter 1

1.5 Overview of the thesis layout

Many infectious and autoimmune diseases have now been directly linked to human genes (e.g. see Pleasance *et al.*, 2010). Molecular approaches point to genetic alterations due to mutation, failure of recombination processes or cellular repair mechanisms. These changes may influence individual susceptibility or resistance towards particular diseases. In particular, modern molecular methods offer unprecedented insights into human genomic regions. Hence, such investigation should become a priority area for biomedical research.

My PhD study involves molecular screening of HLA, MICA, blood groups, HPA and KIR loci using Victoria University of Wellington (VUW) DNA Bank samples (see Fig. 1 for the chromosal locations of these markers). The study has two major components; a bioinformatic study of HLA and MICA data and laboratory based analyses of blood groups, HPA and KIR (see Chapter 2 for further details). Only one molecular genetic study of HLA variation in New Zealand Maori has previously been carried out by Tracey (2007). Molecular scoring of HLA and KIR for the four Pacific Island sub-populations has been reported by Velickovic (2001) and Velickovic et al. (2006), respectively. The data provided from these three studies are informative but only in a preliminary sense. Further work still needs to be done in order to provide more extensive and reliable data for the Polynesians and Maori (e.g. regarding the effects of European admixture). Nonetheless, earlier data on the Polynesian and Maori subpopulations can be re-interpreted and combined with those from the present study, making larger databases (see Chapter 3.1 and 3.2 for further details). All prior studies on blood groups on Maori and Polynesians have been done via serological typing alone and thus provide an interesting basis for comparisons. This study will also validate and establish the application of high-throughput blood group and HPA typing (i.e. BLOODChip) in NZ

Chapter 1

subjects by comparing those results with similar ones obtained using a polymerase chain reaction-sequence specific primer (PCR-SSP) assay and SBT. Thus, the present study will be a stepping stone for the introduction of molecular approaches for HLA, MICA, blood group, HPA and KIR typing in New Zealand. Several other features also distinguish the present work. My survey uses DNA samples that have been classified according to admixture based on self-declared ancestry and interview. Hence, this new work allows one to examine participants that can be clearly identified as lacking genetic input from European ancestors. Therefore, population data provided by the present survey can not only be used as a base line medical reference standard but also for other advantages (see below). Overall, it is expected to give a better perspective regarding the distribution of HLA, MICA, blood group, HPA and KIR in Maori and Polynesians; e.g. as a guide when searching for donor and health assessments based on diseases known to be associated with specific HLA, MICA, blood group, HPA and KIR.

Population data for highly polymorphic genes such as HLA, MICA, blood groups, HPA and KIR also have significant values for ancestry studies. This complements previous mtDNA, Y-chromosome haplotype scoring and microsatellite DNA analysis of VUW DNA Bank samples. Our presently emergent picture of Polynesian origins is captured by the extended version (Chambers and Edinur, 2013) of the *'Synthetic Total Evidence Model (STEM)*' see Box 1 for other competing models. According to *STEM*, the distal sources of populations that contribute to the settlement of Remote Oceania can be divided into at least 2 groups; first an ancient Papuan-speaking Australoid lineage (P-SA; ancestors of Australian Aborigines, ex-coastal Melanesia, hill tribes of Papua New Guinea, Andaman Islanders and Negritos) moved South from Asia 40,000-70,000 yrs ago and the second group of Austronesian-speaking Mongoloids (A-SM) who first occupied Taiwan around 4-6,000 years ago (see Fig.

9

2). The A-SM people migrated to Island South East Asia (ISEA), coastal Papua New Guinea and Island Melanesia between 5000 to 3,500 years ago (Bellwood *et al.*, 2011 and Chambers and Edinur, 2013). In coastal settlements around northern mainland Papua New Guinea and the neighbouring islands, matrilineal marriage meant that numerous P-SA men were recruited by the A-SM voyagers and contributed towards a gender-biased genetic admixture process; see Chambers and Edinur (2013) for further details. The newly formed hybrid population (Proto Polynesians) was then moved outwards to Remote Oceania (Polynesia) and eventually occupied New Zealand from 650 ybp. Genetic admixture continues today as Maori and Polynesian people intermarry with the descendant of Europeans who arrived in New Zealand over the last 150 years. These historical events have left a 'genetic trail' that stretches from Taiwan to New Zealand, Hawaii and Easter Islands. Many studies have contributed to this account including mitochondrial DNA, Y-STR and other autosomal loci (see Chambers, 2008, Kayser *et al*, 2008 and Belwood *et al.*, 2011). The present study on HLA, MICA, blood group, HPA and KIR loci adds to this story and also informs their medical significance (see the following Chapters).



Fig. 1: Chromosomal locations of genetic markers used in the present study. Figure is modified from Logdberg *et al.* (2005) and Oberti *et al.* (2004).

Box 1: Models of population expansion to Oceania

Model	Descriptions
The <i>Express Train to</i> <i>Polynesia</i> model by Diamond (1988)	Based on the ' <i>Out of Taiwan</i> ' archaeological hypothesis proposed by Bellwood (1985). According to this hypothesis, common Austronesian ancestors are hill tribe aboriginal people of Taiwan who migrated southwards via the Batanes Islands, Philippines to Remote Oceania without measurable pause in Near Oceania.
The <i>Slow Boat Model</i> by Oppenheimer and Richards (2001)	This model suggested Island of Southeast Asia as the origin of Polynesians, which is in marked conflict with archaeological chronology.
The Synthetic Total Evidence Model (STEM) by Chambers (2006)	This model embraces all pattern and process elements including <i>Out of Taiwan</i> , gender-biased reticulated admixture process in Near Oceania and genepool refinement via bottlenecks and founder effects.



Fig. 2: Migration patterns of A-SM in Island South East Asia, Near and Remote Oceania based on archaeological dating. The directions of movements are inferred from sequential chronology and linguistic evidence. Abbreviations: A-SM (Austronesian-speaking Mongoloid), P-SA (Papuan-speaking Australoid) and E (Europeans). This figure is reproduced after Chambers (2008).

1.6 Genetic markers

1.6.1 Human leukocyte antigen (HLA)

The major histocompatibility complex (MHC) lies on the short arm of human chromosome 6 and encodes a number of immune genes, including human leukocyte antigen (HLA) class I and II. The HLA class I antigens consist of non-covalently associated of α -chain glycoprotein and β_2 -microglobulin (Eren and Travers, 2000). The α -chain glycoprotein determines HLA class specificities (i.e. HLA-A, -C and -B) and is coded for by the highly polymorphic *HLA-A*, *-C* and *-B* genes, located within the classical class I sub-region of MHC complex (Milner *et al.*, 2000). In contrast, the β_2 -microglobulin is non-polymorphic and encoded by a single gene on chromosome 15 (Goodfellow *et al.*, 1975). The HLA class II antigens are heterodimers of α - and β -chain glycoproteins which are encoded by three pairs of genes on HLA-D region of chromosome 6 (Trowsdale, 1987). This region codes for the three types of HLA class II antigens, HLA-DP, -DQ and -DR.

The HLA class I antigens are present on all nucleated cells and play role in endogenous antigenic peptide presentation to CD8⁺ T cell (see Fig. 3). The peptide binding clefts of newly synthesized HLA class I molecules bind to endogenously synthesized antigens in the cytosol of the cell. The antigenic peptides associated with HLA class I molecules are then moved out to the cell surface and form a ligand for T-cell receptors (TCRs) of CD8⁺ T cytotoxic cell recognition (Thorsby, 1999). In contrast, HLA class II are usually expressed on the surface of antigen presenting cells and are involved in presentation of exogenous peptides to CD4⁺ T cells (see Fig. 3). The class II ligands are mainly derived from endocytosed plasma membrane proteins and extracellular fluid proteins. The HLA class II-
peptide complexes are then transported to the cell surface for CD4⁺ T helper cell recognition (Crumpton, 1987, Strominger, 1987 and Davies, 1997). Generally, HLA molecules are involved in recognition of foreign (non-self) peptides, which is crucial for disease resistance. However, certain foreign materials also contain components that resemble those expressed by human cells and cause predisposing HLA molecules to bind and present apparent auto-antigens to T-cells. This process is known as molecular mimicry and may trigger autoimmune diseases such as ankylosing spondylitis, rheumatoid arthritis, narcolepsy and coeliac disease (see Sinha *et al.*, 1990 and Thorsby and Lie, 2005).

Serological and molecular studies have revealed marked differences in levels of variation between α - and β -chains of HLA class II antigens. For example, α -chain is constant between individuals for HLA-DR, but are polymorphic for HLA-DQ (Trowsdale, 1987). Mutation and shuffling of closely linked HLA loci by recombinantion mechanism can account for the extensive diversity of HLA loci (Bodmer and Thomson, 1977). There are more than 1800, 2400, 1300, and 1000 of HLA-A, -B, -C and -DRB1 alleles (respectively) have been reported and compiled in public HLA database; www.allelefrequencies.net (Gonzalez-Galarza *et al.*, 2011). The HLA class I and II polymorphism in Polynesian and Maori populations together with their applications in population genetics and modern day health are explained and discussed in more details in Chapters 3.1 and 3.2.



Fig. 3: Schematic diagram of HLA class I and II antigenic peptide presentation to $CD8^+ T$ and $CD4^+ T$ cells, respectively. This figure is modified from Thorsby and Lie (2005).

1.6.2 Major histocompatibility complex class I chain-related gene A (MICA)

The major histocompatibility complex class I chain-related gene A (MICA) molecule is encoded by a highly polymorphic gene located within the classical class I sub-region on the chromosome 6 (Horton *et al.*, 2004). Structurally, the MICA molecule is similar to HLA class I, but not associated with β_2 -microglobulin and is not involved in antigenic peptide presentation (Groh *et al.*, 1996 and Stephens, 2001). Instead, MICA molecules are expressed by keratinocytes, fibroblasts, endothelial cells, monocytes, epithelial cells in response to cellular stress (Bauer *et al.*, 1999). The MICA molecules expressed by these cells stimulate cytotoxic activity of natural killer and CD8⁺ T ($\alpha\beta$ and $\gamma\delta$) cells (Bahram *et al.*, 1994, Bauer *et al.*, 1999, Zwirner *et al.*, 1999).

Currently, there are 83 MICA SNP alleles that have been characterized that give rise to 67 protein variants (Robinson *et al.*, 2010). The high diversity of the MICA gene is thus due to both synonymous and non-synonymous mutation in exons 2, 3 and 4, which code for the three (α 1-3) extracellular domains (Perez-Rodriguez *et al.*, 2002). In addition, transmembrane domain of MICA encoded by exon 5 also shows a unique number of repetitive polyalanine motif, but is shared with a number of MICA alleles (Robinson *et al.*, 2001). Roles and functions of MICA as well as their associations with HLA alleles are discussed in the following section (i.e. Chapter 3.2).

1.6.3 Blood group

The ABO system is the first blood group identified in the early 20^{th} century. Since then, many more systems have been characterized. Currently, there are 30 blood group systems documented in public databases (Patnaik *et al.*, 2012). The following ones are involved in the present study (see Table 1).

System	Chromosome	Protein	Polymorphism	Molecular basis	Amino acid change
ABO	9	Transferase	A→B	526C>G, 703G>A, 796C>A, 803G>C	R176G, G235S, L266M, G268A
		-	А→О	Δ261G	Truncated glycosyltransferase
MNS	4	TM TM	$\begin{array}{c} M \rightarrow N \\ s \rightarrow S \end{array}$	59C>T, 71G>A, 72T>G 143C>T	S20L, G24E T48M
Rhesus	1	TM TM TM	$\begin{array}{c} D \rightarrow \text{variant } D \\ C \rightarrow c \\ e \rightarrow E \end{array}$	Deletion//hybrid/SNPs 48C>G, 178A>C, 203G>A, 307T>C 676G>C	partial, weak and D-negative C16W, I60L, S68N, S103P A226P
Lutheran	19	TM	$Lu^a \rightarrow Lu^b$	230G>A	R77H
Kell	7	TM	k→K	578C>T	T193M
Duffy	1	TM -	$Fy^{a} \rightarrow Fy^{b}$ $Fy^{b} \rightarrow Fy$	125G>A _67T>C (GATA-1)	G42D Not coding
Kidd	18	TM	$Jk^{a} \rightarrow Jk^{b}$	838G>A	D280N
Diego	17	TM	$Di^a \rightarrow Di^b$	2561C>T	P854L
Dombrock	12	GPI	$Do^b \rightarrow Do^a$	793A>G	D265N
Colton	7	TM	$Co^a \rightarrow Co^b$	134C>T	A45V

Table 1: List of blood groups involved in this study and their molecular bases

Abbreviations: TM: transmembrane proteins, GPI: proteins anchored to the membrane by an glycosylphosphatidylinositol tail. Table was adapted from Daniels (2009) and Veldhuisen *et al.* (2009).

1.6.3.1 ABO

The precursor for ABO blood group is the H antigen, encoded by the *H* gene on chromosome 19. The H antigen is then converted to A or B antigens by $\alpha 1, 3$ -*Nacetylgalactosaminyltransferase* and $\alpha 1, 3$ -*D*-*galactosyltransferase*, encoded by the *ABO* gene complex on chromosome 9, respectively (see Harmening and Firestone, 2005 and Table 1 for more details). Therefore, the A and B antigens are the end products of the acting of *A*- and *Bglycosyltransferases* on H antigen. In contrast, the *O* allele contains a deletion ($\Delta 261G$) in exon 6 of *ABO* gene resulting in loss of enzymatic activity and H antigen remains unchanged (Olsson and Chester, 2001 and Daniels, 2009). Together, combinations of *A*, *B* and *O* alleles produce four phenotypes; A, B, AB and O. In addition, inactivating mutations in *H* gene produce the very rare autosomal recessive phenotype (O_h or Bombay phenotype) and carriers develop isoantibodies toward A, B and H antigens (Cooling, 2008).

1.6.3.2 Rhesus

The Rh blood group is clinically important in blood transfusion and gestation (Ridgwell *et al.*, 1992). The Rh blood group system consists of 5 antigens; D, C, c, E, and e. The D antigen is encoded by *RHD* gene while the other antigens are encoded by *RHCE* gene (see Table 1). These two genes are closely linked and located on chromosome 1p34.3-p36.1 (Hemker *et al.*, 1999). Individuals will either be D-positive or D-negative (aka Rhesus-positive and negative, respectively) no matter which RHCE antigens they carry. The D-negative occurred due to mutation in *RHD* gene as well as genetic exchange between the highly homologous *RHD* and *RHCE* genes (Arce *et al.*, 1993 and Daniels, 2005). The same mechanism is also generated other variants of D antigen such as 'weak D' and 'partial D', which are at low incidence in

the world populations (Westhoff, 2008). The most common occurring Rhesus phenotypes are DCcee in White, Dccee in Black and DCCee in Asians (Grunbaum *et al.*, 1980).

1.6.3.3 MNS

MNS antigens are integral membrane proteins encoded by two homologous glycophorin gene family, *GYPA* and *GYPB* located on long arm of chromosome 4 (Huang *et al.*, 1993 and Velliquette *et al.*, 2010). The *GYPA* gene codes for M and N blood group antigens while *GYPB* gene codes for the S and s blood group antigens. These two pairs of polymorphic antithetical and co-dominant antigens are differ by three (59C>T, 71G>A, 72T>G) and one (143C>T) SNPs, respectively (see Table 1). Nonetheless, deletions, misalignment during meiosis and gene conversion between the closely linked and homologous *GYPA* and *GYPB* genes give rise to low incidence MNS variants such as En(a-), S-s-U-, M^k and Miltenberger (Poole *et al.*, 1999 and Leger and Calhoun, 2005).

1.6.3.4 Kell

The Kell blood group antigens are 93-kDa type II glycoproteins that are linked by a single disulfide bond to an integral membrane protein, Kx (Lee *et al.*, 2001). The Kell and Kx antigens are encoded by separate genes, *KELL* and *XK* located on chromosome 7q33 and Xp21.1, respectively (Reid and Lomas-Francis, 2004). There are epistatic interactions between these two gene products as the absence in the KX protein of XK blood group system will reduce expression of Kell antigens (in McLeod syndrome; see Daniels, 2002). The weak expression of Kell antigens can also be seen in the K_{mod} RBC phenotype which is associated

with a missense mutation (1208G>A) in exon 10 of the Kell blood group gene (Lee *et al.*, 2003).

Overall, the Kell blood group system comprises more than 34 serologically defined antigens, including three sets of antigens; K and k, Kp^a, Kp^b and Kp^c and Js^a and Js^b (Patnaik *et al.*, 2012). The K and its antithetical antigen, k allele only differ by 578T>C nucleotide substitution which led to Thr193Met amino acid exchange, respectively (Lee *et al.*, 1995). The Kp^b allele is the most common in all populations and only differs from Kp^a and Kp^c by a single amino acid substitution at codon 281 of exon 8. The Kp^a codon codes for tryptophan (TGG), Kp^b for arginine (CGG) and glutamine (CAG) for Kp^c. Equally, Pro597Leu substitution differentiates between the Js^a and Js^b alleles, respectively (Daniels, 2005 and Lee, 2007).

1.6.3.5 Kidd

The *JK* or *SLC14A1* (solute carrier family 14, member 1) gene located on chromosome 18q11-q12 codes for the Kidd antigens. Variants of this locus, the *JK*A and *JKB* alleles, code for the two co-dominant Kidd antigens, Jka and Jkb, respectively (Daniels, 2002). The *JKA/JKB* polymorphism involves a nucleotide transition (838G>A) and define three phenotype: Jk(a+b-), Jk(a-b+) and Jk(a+b+) (Olives *et al.*, 1997).

The occurrence of Jk_{null} phenotype or Jk(a-b-) has been observed in Polynesian and Finnish populations (Irshaid *et al.*, 2000). The Jk_{null} phenotype occurs due to a mutation in the 3'acceptor splice site of intron 5 (IVS5-1G>A) and 5'- splice site of intron 7 of *JKB* alleles which lead to either skipping of exon 6 (Jk Δ 6 mutation) and exon 7 (Jk Δ 7 mutation),

respectively (Lucien *et al.*, 1998 and Lin and Yu, 2008). In addition, single nucleotide polymorphism (202C>T, 222C>A, 582C>G, 871T>C, 896G>A, 956C>T), 723delA, deletion of intron 3 to intron 5 also contributed towards a Jk_{null} phenotype (Lucien *et al.*, 1998; Irshaid *et al.*, 2000; Wester *et al.*, 2008; Liu *et al.*, 2009).

1.6.3.6 Duffy

The Duffy antigen chemokine receptor (*DARC*) gene located on the long arm of chromosome 1q22-q23 (Reid and Lomas-Francis, 2004). There are two alleles; *FY*A* and *FY*B*, which code for Fya and Fyb antigens, respectively that are expressed on red blood cell (RBC), endothelium, brain, colon and kidney. They are identified as receptors for *Plasmodiium vivax* and *Plasmodium knowlesi* (Tournamille *et al.*, 1995 and Leger and Calhoun, 2005). The *FY*A* and *FY*B* alleles differ by a point mutation at nucleotide 125 (125G>A) that encode glycine and aspartic acid, respectively (Estalote *et al.*, 2005). These alleles determine three of five known phenotypes: Fy (a+b-), Fy (a-b+) and Fy (a+b+).

The fourth phenotype, Fy (a-b-) or Fyb^{Es} (erythroid silent) has resulted from a 67T>C point mutation in the GATA-1 binding motif of the *FY*B* promoter (Tournamille *et al.*, 1995 and Rios *et al.*, 2000). Despite having Fyb^{Es} phenotype, several individuals do express weak Fyb antigens (Fya-b^{weak}) which are associated with *FY*X* allele. The *FY*X* allele is caused by Arg89Cys substitution (286C>T) in the background of FY**B* allele (Tournamille *et al.*, 1998).

1.6.3.7 Dombrock

Dombrock antigens, Do^a and Do^b are glycoproteins encoded by the *DO* gene, located on chromosome 12p13.2-p12.1 (Reid and Lomas-Francis, 2004). These antigens are attached to RBC membrane by glycosylphosphatidylinositol (GPI) and incompatibility can cause a transfusion reaction (Daniels, 2008 and Costa *et al.*, 2010). Multiple single nucleotide polymorphisms in exon 2 of the *DO* gene: 378C>T, 624T>C and 793A>G differentiate the Do^a and Do^b alleles (Storry *et al.*, 2003). While the first two SNPs are silent transitions, the last ones causes an amino acid change at position 265 from asparagine for Do^a to aspartic for the Do^b antigen (Costa *et al.*, 2010).

1.6.3.8 Colton

Colton antigens (Co^a and Co^b) are encoded by a gene on chromosomal 7p14 (Daniels, 2002 and Reid and Lomas-Francis, 2004). The Co^a and Co^b antigens are produced by single nucleotide polymorphism (C>T) at nucleotide position 134 resulting in an alanine to valine change (Arnaud *et al.*, 2010). Alloantibodies against Co^a and Co^b antigens can cause haemolytic transfusion reaction and haemolytic disease of the fetus and newborn (Daniels, 2008).

1.6.3.9 Diego

Diego antigens (Di^a and Di^b) are located on the anion exchanger 1 of red cell membrane glycoprotein, products of *SLC4A1* gene (solute carrier family 4, anion exchanger, member 1) on chromosome 17q12-q21 (Daniels, 2002). A single nucleotide polymorphism (2561C>T)

in exon 9 causes an amino acid substitution (Lys658Glu) and differentiates these two antigens (Reid and Lomas-Francis, 2004). The Di^a antigen is very rare among Africans and Europeans, but slightly higher among Asians and Indigenous Americans (Daniels, 2008).

1.6.4 Human platelet antigens (HPA)

Human platelet antigens (HPA) are glycoproteins on the surface of anucleate platelets and play major role in haemostasis. Here, platelet aggregation takes place through interaction of HPAs with other vascular endothelium and plasma proteins (Norton *et al.*, 2004). The molecular bases of no less than 27 HPA systems have been assigned and approved according to guidelines set up by The Platelet Nomenclature Committee (Metcalfe *et al.*, 2003). According to HPA nomenclature, each HPA system is represented by 'a' and 'b' alleles. The taq 'w' (workshop) is added after the antigen only if antibody against the 'a' antigen has been reported. See Table 2 for the molecular basis of HPA alleles and Chapter 3.4 for further details of HPA work in Polynesian and Maori populations.

Antigen	Glycoprotein	Chromosome	Nucleotide Change	Protein
HPA-1	GPIIIa	17	176T>C	L33P
HPA-2	GPIba	17	482C>T	T145M
HPA-3	GPIIb	17	2621T>G	I843S
HPA-4	GPIIIa	17	506G>A	R143Q
HPA-5	GPIa	5	1600A>G	E505K
HPA-6w	GPIIIa	17	1544G>A	R489Q
HPA-7w	GPIIIa	17	1297C>G	P407A
HPA-8w	GPIIIa	17	1984C>T	R636C
HPA-9w	GPIIb	17	2602G>A	V837M
HPA-10w	GPIIIa	17	263G>A	R62Q
HPA-11w	GPIIIa	17	1976G>A	R633H
HPA-15	CD109	6	2108C>A	S703Y

Table 2: Molecular basis of HPA systems involved in this study

Nucleotide substitutions are shown as changes from a to b alleles. Adapted from IPD database (Metcalfe *et al.*, 2003).

1.6.5 Killer cell immunoglobulin-like receptor (KIR)

Natural killer (NK) cells are large granule lymphocytes involved in cytotoxicity response toward infected and tumour cells. Recognition of healthy and unhealthy cells by NK cells are assisted by various combinations of NK receptors and their ligands (e.g. HLA and MIC). The NK receptors are either immunoglobulin-like or lectin-like structure, encoded by leukocyte receptor complex (LRC) of chromosome 19 and natural killer complex of chromosome 12, respectively (see Yokoyama and Plougastel, 2003 and Martin *et al.*, 2002). Killer cell immunoglobulin-like receptors (KIRs) are among the receptors encoded by the LRC complex. The KIRs are highly diverse and evolve rapidly, compared with the lectin-like receptor genes which are quite well conserved in humans. For this reason, KIRs have become attractive markers in immunogenetic and population studies (see Robinson *et al.*, 2010 and Gonzalez-Galarza *et al.*, 2011).

The name given to KIR antigens are based on their extracellular immunoglobulin domains (2 domains or 3 domains) and the length of cytoplasmic domain (Long: L or Short: S). The KIR proteins with long cytoplasmic domains transmit inhibitory signals via an immune receptor tyrosine-based inhibitory motifs (ITIMs) while the KIR proteins with short cytoplasmic domain lack an ITIM and are activating receptors. The KIR2DL1, KIR2DL3, KIR3DL1, KIR3DL3 and KIR2DL5 genes code for inhibitory receptor whereas KIR2DS4, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS5 and KIR3DS1 genes code for stimulatory receptors. The KIR2DL4 gene has both, stimulatory and inhibitory potential (Robinson *et al.*, 2010). The KIR genes can also be divided into two haplotype groups (see Fig. 3):

a. Haplotype A: *KIR2DL1, KIR2DL3, KIR3DL1, KIR3DL3, KIR2DS4, KIR2DP1, KIR3DP1, KIR3DL2* and *KIR2DL4*

b. Haplotype B: Any combination other KIR genes including those assigned for haplotype A

A given KIR haplotype may have variable number of KIR genes, as there are variations in the presence or absent of KIR genes, except for the framework loci (*KIR3DL3*, *3DP1*, 2DL4 and *3DL2*), which are common to all haplotypes (Middleton *et al.*, 2002). This form of polymorphism is not limited to gene content but also extends to allelic level.

Currently, more than 600 alleles have been reported across all 16 of the KIR loci (see Robinson *et al.*, 2010 and Gonzalez-Galarza *et al.*, 2011). In this respect, the complexity of KIR system reflects to that observed for HLA class I, which is also a ligand for KIRs (see Table 3). The extreme diversity of KIR and HLA class I loci are also thought to be co-evolutionary with distinct KIR-HLA combinations providing varying degrees of protection toward diseases (see Single *et al.*, 2007 and Chapter 3.5 for further details). Nevertheless, those KIRs without HLA ligands might also engage in vital part in phato-physiological functions with ligands other than HLA (Parham *et al.*, 2012 and VandenBussche *et al.*, 2009).



Fig. 4: Schematic diagram of KIR haplotype A and B. Figure is modified from Parham et al. (2012).

KIR	Alleles	Protein variants	HLA ligands
2DL1	43	24	C2
2DL2	28	11	C1,C2
2DL3	34	17	C1,C2
2DL4	46	22	G
2DL5	41	17	unknown
2DS1	15	7	C2
2DS2	22	8	unknown
2DS3	14	5	unknown
2DS4	30	13	A*11, some C
2DS5	16	11	unknown
3DL1	73	58	Bw4
3DS1	16	12	unknown
3DL2	84	61	A*03,-11
3DL3	107	55	unknown
3DP1	22	0	0
2DP1	23	0	0

Table 3: List of KIRs and their ligands

C1 = HLA-C allotypes with serine and asparagines at position 77 and 80 of α 1 domain, respectively C2 = HLA-C allotypes with asparagines and lysine at position 77 and 80 of α 1 domain, respectively and Bw4 = HLA-B allotypes with isoleucine or threonine at position 80 of α 1 domain. Adapted from Robinson *et al.* (2010) and Parham *et al.* (2012).

1.7 Summary of experimental objectives

Objectives of these projects are:

- A. To compile new HLA, MICA, blood group, HPA and KIR data set for Polynesian and Maori subjects
- B. To compare our data from all these systems with those from previous studies
- C. To use data from this study for:
 - i. Tracing Polynesian ancestry and migration pattern
 - ii. Showing old and new genetic admixture events
 - iii. Examining the potential influence of these variants on Maori and Polynesian health
- D. To evaluate new molecular blood group and HPA genotyping methods using:
 - i. Sequence specific primer-based typing (BAGene DNA-SSP; BAG Health Care GmbH, Lich, Germany)
 - ii. Chip hybridization SNP assay (BLOODchip[®] v2.0; Progenika Biopharma S.A., Spain)
 - iii. Sequence based typing as a quality assurance check

The work described here in pursuit of these objectives and allows one to test a range of competing ideas about the significance of HLA, MICA, KIR, blood groups and HPA diversities in Polynesians. These are listed in the next section (sub-section 1.8).

1.8 Hypotheses

In the following chapters, we tested these specific hypotheses:

Chapter 3.1: that HLA data will provide a way to test the claim of the STEM model that Polynesians have unequal genetic contributions from A-SM and P-SA ancestors.

Chapter 3.2: that high resolution HLA and MICA data will help to explain differences in the rates of occurrence of auto-immune diseases and infection diseases between that Polynesians and Europeans.

Chapter 3.3: that molecular techniques provide a reliable alternative technology for determining blood groups in Polynesian subjects.

Chapter 3.4: that Maori and Polynesians will show difference in HPA allelic distribution from Europeans due to population history.

Chapter 3.5: that Polynesians will display a unique repertoire of KIR haplotypes and different from those previously reported in Europeans.

References

Anstee, D.J. (2009) Red cell genotyping and the future of pretransfusion testing. *Blood*, **114**, 248-256.

Arce, M.A., Thompson, S., Wagner, S., Coyne, K.E., Ferdman, B.A. & Lublin, D.M. (1993) Molecular cloning of RhD Cdna derived from a gene present in RhD-positive, but not RhDnegatived individuals. *Blood*, **82**, 651-655.

Arnaud, L., Helias, V., Menanteau, C., Peyrard, T., Lucien, N., Ripoche, P., Lapegue, R., Pham, B-N., Le Pennec, P-Y., Moulds, J.J. & Cartron, J-P. (2010) A funtional AQP1 allele producing a Co(a-b-) phenotype rvises and extends the Colton blood group system. *Transfusion*, **50**, 2106-2116.

Avent, N.D., Martinez, A., Flegel, W.A., Olsson, M.L., Scott, M.L., Nogues, N., Nogues, N.,
Pisacka, M., Daniels, G.L., Muniz-Diaz, E., Madgett, T.E., Storry, J.R., Beibor, S.,
Maaskant-van Wijk, P.M., von Zabern, I., Jimenez, E., Tejedor, D., Lopez, M., Camacho, E.,
Cheroutre, G., Hacker, A., Jinoch, P., Svobodova, I., van der Schoot, E. & de Haas, M.
(2009) The bloodgen project of the European Union, 2003-2009. *Transfusion Medicine and Hemotherapy*, **36**, 162-167.

Bach, F.H. & Amos, D.B. (1967) Hu-1: major histocompatibility locus in man. *Science*, **156**, 1506-1508.

Bahram, S., Bresnahan, M., Geraghty, D.E. & Spies, T. (1994) A second lineage of mammalian major histocompatibility complex class I genes. *Proceedings of the National Academy of Sciences*, **91**, 6259-6263.

Bauer, S., Groh, V., Wu, J., Steinle, A., Phillips, J.H., Lanier, L.L. & Spies, T. (1999)Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science*, 285, 727-729.

Beck, S. (2004) Gene map of the extended human MHC. *Nature Reviews Genetics*, **12**, 889-899.

Bellwood, P. (1985). Prehistory of the Indo-Malay Archipelago. Academic Press, Sydney.

Bellwood, P., Chambers, G.K., Ross, M. & Hung, C-H. (2011) Are "cultures" inherited? Multidisciplinary perspectives on the origins of Austronesian–speaking peoples prior to 1000 BC. In: *Investigating archaeological cultures: material culture, variability and transmission* (eds. Roberts, B. W. & Linden M.V.), 321-354. Springer, Dordrecht.

Blann, A. (2006) Routine blood results explained. M & K Update Ltd, Cumbria.

Bodmer, W. & Thomson, G. (1977) Population genetics and evolution of the HLA system. In: *HLA and disease* (eds. Dausset, J. & Svejgaard, A.), 280-295. Munksgaard, Copenhagen.

Brown, C.J. & Navarrete C.V. (2011) Clinical relevance of the HLA system in blood transfusion. *Vox Sanguinis* **101**: 93-105.

Chambers, G.K. (2006) Polynesian genetics and Austronesian prehistory. In: *Austronesian Diaspora and the Ethnogeneses of People in Indonesian Archipelago* (eds. Simanjuntak T, Pojoh I & Hisyam M.), 299-319. Indonesian Institute of Sciences (LIPI), Jakarta.

Chambers, G.K. (2008) Genetics and the origins of the Polynesians. In: *Handbook of human molecular evolution* (eds. Cooper D.N. & Kehrer-Sawatzki H.), 489-497. John Wiley & Sons Inc, Chichester.

Chambers, G.K. & Edinur, H.A. (2013) Genetic relationships between Malays and Maori. In: *The changing values of Malays, Maori and Pacific Islanders* (eds. Rahman W.R.A. & Higgins T.R.), 6-37. Chair of Malay Studies & Victoria University Foundation, Wellington, New Zealand.

Chinen, J. & Bukley, R.H. (2011) Transplantation immunology: solid organ and bone marrow. *Journal of Allergy and Clinical Immunology*, **125**: S324-S335.

Cooling, L. (2008) ABO, H, and Lewis blood groups and structurally related antigens. In: *Technical manual* (16th edn) (eds. Roback, J.D., Combs, M.R., Grossman, B.J. & Hillyer, C.D.), 361-385. American Association of Blood Banks, Bethesda, Maryland.

Costa, F.P.S., Hue-Roye, K., Sausais, L., Velliquette, R.W., Da Costa Ferreira, E., Lomas-Francis, C. & Reid M.E. (2010) Absence of DOMR, a new antigen in the Dombrock blood group system that weakens expression of Do^b, Gy^a, Hy, Jo^a, and DOYA antigens. *Transfusion*, **50**, 2026-2031. Crumpton, M.J. (1987) Introduction. British Medical Buletin, 43, i-vi.

Daniels, G. (2002) Human blood groups (2nd edn). Blackwell Science, Malden, MA.

Daniels, G. (2005) The molecular genetics of blood group polymorphism. *Transplant Immunology*, **14**,143-153.

Daniels, G. (2008) Other blood groups. In: *Technical manual* (16th edn) (eds. Roback, J.D., Combs, M.R., Grossman, B.J. & Hillyer, C.D.), 411-498. American Association of Blood Banks, Bethesda, Maryland.

Daniels, G. (2009) The molecular genetics of blood group polymorphism. *Human Genetics*, **126**, 729-742.

Davenport, R.D. & Mintz, P.D. (2007) Transfusion Medicine. In: *Henry's clinical diagnosis* and management by laboratory methods (21st edition) (eds. McPherson R.A. & Pincus, M.R.), 670-685. Saunders Elsevier, Philadelphia.

Davies, H. (1997) Introductory immunobiology. Chapman & Hall, London.

Doxiadis, I.I.N., Witvliet, M.D., Persijn, G.G. & Claas, F.H.J. (2006) External proficiency testing in histocompatibility typing, screening and crossmatching for organ transplantation. In: *Standardization of donor-recipient matching in transplantation* (ed. Lange, A.), 105-126. Nova Science Publisher, Inc, New York. Dunn, P.P., Fisher, R., Darke, C. & Blacklock, H. (2012) HLA alleles and haplotypes & recruitment in five Polynesian populations in the New Zealand Bone Marrow Donor Registry. *Tissue Antigens*, **79**, 557.

Elgert, K.D. (2009) *Immunology : understanding the immune system* (2nd edition). Wiley-Blackwell, Hoboken, New Jersey.

Engelfriet, C.P. & Reesink, H.W. (2003) Detection of platelet-reactive antibodies in patients who are refractory to platelet transfusions, and the selection of compatible donors. *Vox Sanguinis*, **84**, 73-88.

Eren, E. & Travers, P. (2000) The structure of the major histocompatibility complex and its molecular interactions. In: *HLA in health and disease* (2nd edn) (eds. Lechler, R. & Warrens, A.), 3-33. Academic Press, San Fracisco, New York.

Estalote, A.C., Proto-Siqueira, R., Da Silva Jr, W.A., Zago, M.A. & Palatnik, M. (2005) The mutation G298A \rightarrow Ala100Tr on the coding sequence of the Duffy antigen/chemokine receptor gene in non-caucasian Brazilians. *Genetics and Molecular Research*, **4**, 166-173.

Gonzalez-Galarza, F.F., Christmas, S., Middleton, D. & Jones, A.R. (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Research*, **39**, D913-D919.

Goodfellow, P.N., Jones, E.A., Heyningen, V.V., Solomon, E. & Borrow M. (1975) The β_2 microglobulin gene is on chromosome 15 and not in the HL-A region. *Nature*, **254**, 267-268.

Groh, V., Bahram, S., Bauer, S., Herman, A., Beauchamp, M. & Spies, T. (1996) Cell stressregulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proceedings of the National Academy of Sciences*, **93**, 12445-12450.

Grunbaum, B.W., Crim, M., Selvin, S., Myhre, B.A. & Pace, N. (1980) Distribution of gene frequencies and discrimination probabilities for 22 human blood genetic systems in four racial groups. *Journal of Forensic Sciences*, **25**, 428-444.

Harmening, D.M. & Firestone, D. (2005) The ABO blood group system. In: *Modern blood banking & transfusion practices* (5th edn) (ed. Harmening, D.M), 108-133. F.A. Davis Company, Philadelphia.

Hemker, M.B., Ligthart, P.C., Berger L., van Rhenen, D.J., van der Schoot, C.E., Maaskantvan Wijk, P.A. (1999) *DAR*, a new RhD variant involving exons 4, 5, and 7, often in linkage with *ceAR*, a new Rhce variant frequently found in African Blacks. *Blood*, **94**, 4337-4342.

Higgins, C. (2007) Understanding laboratory investigations for nurse and health professionals (2nd edn). Blackwell Publishing Ltd, Garsington Road, Oxford.

Hopkins, P.M.A. (2006) Pharmacological manipulation of the rejection response. In: *Transplantation immunology: methods in molecular biology* (Hornick P. & Rose M.), 375-399. Humana Press Inc., Totowa, New Jersey.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C. Jr., Wright, M.W., Wain, H.M., Trowsdale, J., Ziegler, A. &

Beck, S. (2004) Gene map of the extended human MHC. *Nature Review Genetics*, **12**, 889-899.

Huang, C-H., Reid, M., Daniels, G. & Blumenfeld, O.O. (1993) Alteration of splice site selection by an exon mutation in the human Glycophorin A gene. *The Journal of Biological Chemistry*, **268**, 25902-25908.

Hustinx, H., Fontana, S., Gowland, P. & Niederhauser, C. (2009) Will Genotyping Replace Serology in Future Routine Blood Grouping? – Opinion 2. *Transfusion Medicine and Hemotherapy*, **36**, 228-229

Irshaid, N.M., Henry, S.M. & Olsson, M.L. (2000) Genomic characterization of the Kidd blood gene: different molecular basis of the Jk (a-b-) phenotype in Polynesians and Finns. *Transfusion*, **40**, 69-74.

Kalwak, K., Toporski, J., Gorczynska, D., Turkiewicz, D., Slociak, M., Wojcik, D. & Chybicka A. (2006) Do we really need a "perfect match"? Analysis of patient outcome after unrelated donor hematopoietic cell transplantation (URD-HCT) in the department of pediatric hematology/oncology, Wroclaw, Poland. In: *Standardization of donor-recipient matching in transplantation* (ed. Lange, A.), 135-140. Nova Science Publisher, Inc, New York.

Kayser, M., Lao, O., Saar, K., Brauer, S., Wang X., Numberg P., Trent R.J. & Stoneking, M. (2008) Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *American Journal of Human Genetics*, **82**, 194-198.

Lawson, C. (2006) Strategies for gene transfer to solid organs: viral vectors. In: *Transplantation immunology: methods in molecular biology* (Hornick P. & Rose M.), 175-200. Humana Press Inc., Totowa, New Jersey.

Lee, S. (2007) The value of DNA analysis for antigens of the Kell and Kx blood group systems. *Transfusion*, **47**, 32S-39S.

Lee, S, David, C.W., Russo, D.C.W., Alexander, P., Reiner, A.P., Jeffrey, H., Lee, J.H., Sy, M.Y., Telen, M.J., Judd, W.J., Simon, P., Maria, J., Rodrigues, M.J., Chabert, T., Poole, J., Jovanovic-Srzentic, S., Levene, C., Yahalom, V. & Redman, C.M. (2001) Molecular defects underlying the Kell null phenotype. *The Journal of Biological Chemistry*, **276**, 27281–27289.

Lee, S., Russo, D.C.W., Reid, M.E. & Redman, C.M. (2003) Mutations that diminish expression of Kell surface protein and lead to the Kmod RBC phenotype. *Transfusion*, **43**, 1121-1125.

Lee, S., Wu, X., Reid, M., Zelinski, T. & Redman, C. (1995) Molecular basis of the Kell (K1) phenotype. *Blood*, **85**, 912-916.

Leger, R.M. & Calhoun, L. (2005) Other major blood group systems. In: *Modern blood banking & transfusion practices* (5th edn) (ed. Harmening, D.M.), 162-192. F.A. Davis Company, Philadelphia.

Lin, M. & Yu, L-C. (2008) Frequencies of the *JKnull* (IVS5–1g>a) allele in Taiwanese, Fujian, Filipino, and Indonesian populations. *Transfusion*, **48**, 1768-1769.

38

Liu, H-M., Lin, J-S., Chen, P-S., Lyou, J-Y., Chen, Y-J. & Tzeng, C.H. (2009) Two novel Jknull alleles derived from 222C>A in Exon 5 and 896G>A in Exon 9 of the JK gene. *Transfusion*, **49**, 259-264.

Logdberg, L., Reid, M.E., Lamont R.E. & Zelinski, T. (2005) Human blood group genes 2004: chromosomal locations and cloning strategies. *Transfusion Medicine Reviews*, **19**: 45-57.

Lucien, N., Sidoux-Walter, F., Olives, B., Moulds, J., Le Pennec, P-Y., Cartron, J-P. & Bailly, P. (1998) Characterization of the gene encoding the human kidd blood group/urea transporter protein. *The Journal of Biological Chemistry*, **273**, 12973-12980.

Ludajic, K., Balavarca, Y., Bickeboller, H., Rosenmayr, A., Fae, I., Fischer, G.F., Kouba, M., Pohlreich, D., Kalhs, P. & Greinix, H.T. (2009) KIR genes and KIR ligands affect occurrence of acute GVHD after unrelated, 12/12 HLA matched, hematopoietic stem cell transplantation. *Bone Marrow Transplantation*, **44**, 97-103.

Lydyard, P., Whelan A. & Fanger M.W. (2004) *Immunology* (2nd edition). Bios Scientific, London.

Martin, A.M., Kulski, J.K., Witt, C., Pontarotti, P. & Christiansen, F.T. (2002) Leukocyte Iglike receptor complex (LRC) in mice and men. *Trends in Immunology*, **23**, 81-88. Metcalfe, P., Watkins, N.A., Ouwehand, W.H., Kaplan, C., Newman P., Kekomaki, R., de Haas, M., Aster, R., Shibata, Y., Smith, J., Kiefel V. & Santoso, S. (2003) Nomenclature of human platelet antigens. *Vox Sanguinis*, **85**, 240-245.

Middleton D. (2005) HLA typing from serology to sequencing era. *Iranian Journal of Allergy, Asthma and Immunology*, **4**, 53-66.

Middleton, D., Curran, M. & Maxwell, L. (2002) Natural killer cells and their receptors. *Transplant Immunology*, **10**, 147-164.

Middleton, D., Savage, DA., Cullen, C. & Martin, J. (1988) Discrepancies in serological tissue typing revealed by DNA techniques. *Transplant International*, **1**, 161-164.

Milner, C.M., Campbell, R.D. & Trowsdale, J. (2000) Molecular genetics of the human major histocompatibility complex. In: *HLA in health and disease* (2nd edn) (eds. Lechler, R. & Warrens, A.), 35-50. Academic Press, San Fracisco, New York.

Moretta, A., Vitale, M., Bottino, C., Orengo, A.M., Morelli, L., Augugliaro, R., Barbaresi, M., Ciccone, E. & Moretta, L. (1993) P58 molecules as putative receptors for major histocompatibility complex (MHC) class I molecules in human natural killer (NK) cells. Anti-p58 antibodies reconstitute lysis of MHC class I-protected cells in NK clones displaying different specificities. *The Journal of Experimental Medicine*, **178**, 597-604.

Norton, A., Allen, D.L. & Murphy, M.F (2004) Review: platelet alloantigens and antibodies and their clinical significance. *Immunohematology*, **20**, 89-102.

Oberti, C., Wang, L., Li, L., Dong, J., Rao S., Du, W. & Wang Q. (2004) Genome-wide linkage scan identifies a novel genetic locus on chromosome 5p13 for neonatal atrial fibrillation associated with sudden death and variable cardiomyopathy. *Circulation*, **110**, 3753-3759.

Oliveira, S., Tafulo, S., Peixoto, M., Mendes, F., Dias, M. & Alves H. (2012) Sequence based typing lab activity: experience from the north histocompatibility centre, Oporto – Portugal. *Tissue Antigens*, **79**, 466.

Olives, B., Merriman, M., Bailly, P., Bain, S., Barnett, A., Todd, J., Cartron, J-P. & Merriman, T. (1997) The molecular basis of the Kidd blood group polymorphism and its lack of association with type 1 diabetes susceptibility. *Human Molecular Genetics*, **6**, 1017-1020.

Olsson, M.L. & Chester, M.A. (2001) Polymorphism and recombination events at the *ABO* locus: a major challenge for genomic ABO blood grouping strategies. *Transfusion Medicine*, **11**, 295-313.

Olsson, M.L., Irshaid, N.M., Hosseini-Maaf, B., Hellberg, A., Moulds, M.K., Sareneva, H. & Chester, M.A. (2001) Genomic analysis of clinical samples with serologic ABO blood grouping discrepancies: Identification of 15 novel A and B subgroup alleles. *Transfusion medicine*, **98**, 1585-1593.

Oppenheimer, S. & Richards, M. (2001) Fast trains, slow boats, and the ancestry of the Polynesian islanders. *Science Progress*, **84**, 157-181.

Parham, P., Norman, P.J., Abi-Rached, L., Hilton, H.G. & Guethlein LA (2012) Review: Immunogenetics of human placentation. *Placenta*, **33**, S71-S80.

Pathak, S. & Palan, U. (2005) *Immunology : essential and fundamental* (2nd edition). Science Publishers, Enfield, New Hampshire 2005.

Patnaik, S.K., Helmberg, W. & Blumenfeld, O.O. (2012) BGMUT: NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. *Nucleic Acids Research*, **40**, D1023–D1029.

Peh, C.A., Purcell, A.W. & McCluskey, J. (2000) The biology of major histocompatibility complex molecules – II: Antigen processing and presentation. In: *HLA in health and disease* (2nd edn) (eds. Lechler, R. & Warrens, A.), 3-33. Academic Press, San Fracisco, New York.

Perez-Rodriguez, M., Arguello, J.R., Fischer, G., Cox, S.T., Robinson, J., Hossain, E., McWhinnie, A., Travers, P.J., Marsh, S.G.E. & Madrigal J.A. (2002) Further polymorphism of the MICA gene. *European Journal of Immunogenetics*, **29**, 35-46.

Plant, N. & Wood, P. (2009) Transplantation, ABO incompatibility and immunology. *Anaesthesia and Intensive Care Medicine*, **10**, 227-230.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M-L., Ordonez, G.R., Bignell, G.R., Ye, K., Alipaz, J., Bauer, M.J., Beare, D., Butler, A., Carter, R.J., Chen, L., Cox, A.J., Edkins, S., Kokko-Gonzales, P.I., Gormley, N.A., Grocock, R.J., Haudenschild, C.D., Hims, M.M., James, T., Jia, M.,

Kingsbury, Z., Leroy, C., Marshall, J., Menzies, A., Mudie, L.J., Ning, Z., Royce, T., Schulz-Trieglaff, O.B., Spiridou, A., Stebbings, L.A., Szajkowski, L., Teague, J., Williamson, D., Chin, L., Ross, M.T., Campbell, P.J., Bentley, D.R., Futreal, P.A. & Stratton, M.R. (2009) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**: 191-197.

Poole, J., Banks, J., Bruce, L.J., Ring, S.M., Levene, C., Stern, H., Overbeeke, M.A.M. & Tanner, M.J.A. (1999) Glycophorin A mutation Ala65 \rightarrow Pro gives rise to a novel pair of MNS alleles ENEP (MNS39) and HAG (MNS41) and altered Wr^b expression: direct evidence for GPA/band 3 interaction necessary for normal Wrb expression. *Transfusion Medicine*, **9**, 167-174.

Rajalingam, R. & Gebel, H.M. (2011) KIR-HLA mismatching in human renal allograft transplantation: emergence of a new concept. *American Journal of Transplantation*, **11**, 1771-1772.

Reid, M.E. & Lomas-Francis, C. (2004) *The blood group antigen factsbook*. Academic Press, San Diego, California.

Ridgwell, K., Spurr, N.K., Laguda, B., MacGeoch, C., Avent, N.D. & Tanner, M.J.A. (1992) Isolation of cDNA clones for a 50 kDa glycoprotein of the human erythrocyte membrane associated with Rh (Rhesus) blood-group antigen expression. *The Biochemical Journal*, **287**, 223-228. Rios, M., Chaudhuri, A., Mallinson, G., Sausais, L., Gomensoro-Garcia, A.E., Hannon, J.,
Rosenberger, S., Poole, G., Burgess, G., Pogo, O. & Reid, M. (2000) New genotype in Fy (ab-) individual: nonsense mutations (Trp to stop) in the coding sequence of either FY A or FY
B. *British Journal of Haematology*, **108**, 448-454.

Robinson, J., Halliwell, J.A., McWilliam, H., Lopez, R., Parham, P. & Marsh, S.G.E. (2013) The IMGT/HLA database. *Nucleic Acids Research*, **41**, D1222–D1227.

Robinson, J., Mistry, K., McWilliam, H., Lopez, R. & Marsh, S.G.E. (2010) IPD – The immune polymorphism database. *Nucleic Acids Research*, **38**, D863-D869. Robinson, J, Perez-Rodriguez, M, Waller, M.J., Cuiller, B., Bahram, S., Yao, Z., Albert,

E.D., Madrigal, J.A. & Marsh, S.G.E. (2001) MICA sequence 2000. Immunogenetics, 53, 150-169.

Ruggeri, L., Capanni, M., Urbani, E., Perruccio, K., Shlomchik, W.D., Tosti, A., Posati, S., Rogaia, D., Frassoni, F., Aversa, F., Martelli, M.F. & Velardi, A. (2002) Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplantations. *Science*, **295**, 2097-2100.

Schonitzer, D. (2009) Will genotyping replace serology in future routine blood grouping? -Opinion 3. *Transfusion Medicine and Hemotherapy*, **36**, 230-231.

Sheldon, S. & Poulton, K. (2006) HLA typing and its influence on organ transplantation. In: *Transplantation immunology: methods in molecular biology* (Hornick P. & Rose M.), 157-174. Humana Press Inc., Totowa, New Jersey.

Single, R.M., Martin, M.P., Gao, X., Meyer, D., Yeager, M., Kidd, J.R., Kidd, K.K. & Carrington, M. (2007) Global diversity and evidence for coevolution of KIR and HLA. *Nature Genetics*, **39**, 1114-1119.

Sinha, A.A., Lopez, M.T. & McDevitt, H.O. (1990) Autoimmune diseases: the failure of self tolerance. *Science*, **248**, 1380-1388.

Song, J.J., Guyette, J.P., Gilpin, S.E., Gonzalez, G., Vacanti, J.P. & Ott, H.C. (2013) Regeneration and experimental orthotopic transplantation of a bioengineered kidney. *Nature Medicine*, **19**, 646-651.

Stephens, H.A.F. (2001) MICA and MICB genes: can the enigma of their polymorphism be resolved? *Trends in Immunology*, **22**, 378-385.

Storry, J.R. & Olsson, M.L. (2009) Will genotyping replace serology in future routine blood grouping? - Opinion 4. *Transfusion Medicine and Hemotherapy*, **36**, 232.

Storry, J.R., Westhoff, C.M., Charles-Pierre, D., Rios, M., Hue-Roye, K., Vege, S., Nance, S.
& Reid, M.E. (2003) DNA analysis for donor screening of Dombrock blood group antigens. *Immunohematology*, **19**, 73-76.

Strominger, J.L. (1987) Structure of class I and class II HLA antigens. *British Medical Buletin*, **43**, 81-93.

Thorsby, E. (1999) MHC Structure and function. *Transplantation Proceedings*, **31**, 713–716.

Thorsby, E. & Lie, B.A. (2005) HLA aasociated genetic predisposition to autoimmune diseases: genes involved and possible mechanisms. *Transplantation Immunology*, **14**, 175-182.

Thorsby, E. (2012) The Polynesian gene pool: an early contribution by Amerindians to Easter Island. *Philosophical transactions of the Royal Society of London, Series B, Biological sciences*, **367**, 812-819.

Tournamille, C., Colin, Y., Cartron, J.P. & Kim, C.L.V. (1995) Distruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individual. *Nature Genetic*, **10**, 224-228.

Tournamille, C., Le Van Kim, C., Gane, P., Le Pennec, P.Y., Roubinet, F., Babinet, J., Cartron, J.P. & Colin, Y. (1998) Arg89Cys substitution results in very low membrane expression of the Duffy antigen/receptor for chemokines in Fy^x individuals. *Blood*, **92**, 2147-2156.

Tracey, M.C. (2007) *HLA allelic variation in New Zealand Maori*. PhD thesis, University of Otago.

Trowsdale, J. (1987) Genetics and polymorphism: class II antigens. *British Medical Buletin*, **43**, 15-36.

VandenBussche, C.J., Mulrooney, T.J., Frazier, W.R., Dakshanamurthy, S. & Hurley, C. K. (2009) Dramatically reduced surface expression of NK cell receptor KIR2DS3 is attributed to multiple residues throughout the molecule. *Genes Immunity*, **10**, 162-173.

Veldhuisen, B., van der Schoot, C.E. & de Haas, M. (2009) Blood group genotyping: from patient to high-throughput donor screening. *Vox Sanguinis*, **97**, 198–206

Velickovic, Z.M. (2001) *HLA polymorphism in Pacific Islands populations*. PhD thesis, University of Otago, Dunedin, New Zealand.

Velickovic, Z.M. & Carter, J.M. (1999) Feasibility of finding an unrelated bone marrow donor on international registries for New Zealand patients. *Bone Marrow Transplantation*, **23**, 291-294.

Velickovic, M., Velickovic, Z. & Dunckley, H. (2006) Diversity of killer cell immunoglobulin-like receptor genes in Pacific Islands populations. *Immunogenetics* **58**: 523-532.

Velliquette, R.W., Hu, Z., Lomas-Francis, C., Hue-Roye, K., Allen, J.L., Mirabella, D., Reid, M.E. (2010) Novel single-nucleotide change in GYP*A in a person who made an alloantibody to a new high-prevalence MNS antigen called ENEV. *Transfusion*, **50**, 856-860.

Wagner, F.F. (2009) Will genotyping replace serology in future routine blood grouping? - Opinion 1. *Transfusion Medicine and Hemotherapy*, **36**, 226.

Wester, E.S., Johnson, S.T., Copeland, T., Malde, R., Lee, E., Storry, J.R. & Olsson, M.L. (2008) Erythroid urea transporter deficiency due to novel JKnull alleles. *Transfusion*, **48**, 365-372.

Westhoff, C.M. (2008) The ABO blood group system. In: *Mordern blood banking & transfusion practices* (5th edn) (ed. Harmening, D.M.), 387-409. F.A. Davis Company, Philadelphia.

Wise D.J. & Carter G.R. (2002) *Immunology: a comprehensive review*. Iowa State University Press, Ames, Iowa.

Wu, Y.Y. & Csako, G. (2006) Rapid and/or high-throughput genotyping for human red blood cell, platelet and leukocyte antigens, and forensic applications. *Clinica Chimica Acta*, **363**, 165-176.

Yokoyama, W.M. & Plougastel, B.F.M. (2003) Immune functions encoded by the natural killer gene complex. *Nature Reviews Immunology*, **3**, 304-316.

Zou, Y. & Stastny P. (2010) Role of MICA in the immune response to transplant. *Tissue Antigens*, **76**, 171-176.

Zwirner, N.W., Dole, K. & Stastny, P. (1999) Differential surface expression of MICA by endothelial cells, fibroblasts, keratinocytes, and monocytes. *Human Immunology*, **60**, 323-330.

Extended information on Materials and Methods

Individual experimental protocols are given in each result section (Chapter 3.1 - 3.5). The following paragraphs provide further details regarding materials and general laboratory methods.

2.1 Banked DNA sample

Frozen-preserved DNA samples obtained from unrelated Polynesian (TP: Total Polynesians) and Maori (TM: Total Maori) individuals were obtained from the Victoria University of Wellington (VUW) DNA bank. They are all local volunteer participants who provided doubly de-identified blood samples with informed consent and were recruited via the Wellington Blood Service. These volunteers have varying degrees of admixture with Europeans. They are divided based on self-reported pedigree at interview into full ancestry (PFA: Polynesians with Full Ancestry and MFA: Maori with Full Ancestry) for those with four grandparents and no family ancestral knowledge of members from other ethnic groups vs. admixed history (PAH: Polynesians with Admixed History and MAH: Maori with Admixed History) for those with 3 or fewer full grandparents. Ethical approvals were obtained from the New Zealand Central Region Ethic Committee and Victoria University Human Ethics Committee – see Hamilton (1993) and Marshall (1994) for details of collection, interview questions and information sheet given to participants.

2.2 General laboratory practice

All plasticware, glassware, deionized water and pipette tips were sterilized by autoclave at 20 psi (121°C) for 20 minutes. A new set of disposable plastic gloves was used for each reaction set-up. In order to avoid cross contamination, different working areas were designated for sample preparation, PCR setup and handling of post-PCR activities.

2.3 DNA quantification

Quantity of DNA in extraction and PCR products was estimated using NanoPhotometerTM Pearl (IMPLEN). The photometer was calibrated using distilled water (3-5 μ l). An amount of 3-5 μ l of sample was used for DNA quantification and the reading was given in ng/ μ l.

2.4 Low resolution typing of HLA, MICA and KIR loci

Polymerase chain reaction-sequence specific oligonucleotide (PCR-SSO) using Luminex (LABType[®] One Lambda[®], CA, USA) was used to genotype HLA-A, -B, -C, -DRB1, -DQA1, -DQB1, MICA and KIR loci. The target regions were amplified using biotinylated locus-specific primers. Amplicons were then hybridized to complementary bead-bound oligonucleotide probes following denaturation and neutralization procedures. The hybrids were then detected using R-Phycoerythrin-conjugated Streptavidin (SAPE) and fluorescence intensities were read by flow analyzer (LABScanTM 100). Allele/gene scores were assigned based on reaction patterns of reference standards.
Chapter 2

2.5 High resolution typing of HLA loci

Sequencing was performed for high resolution HLA typing. High resolution HLA-DRB1 typing was either performed by DNA sequencing using the method described by Kotsch *et al.* (1999) or by using High Resolution Luminex beads (High Definition beads, test performed by Peter Brescia, One Lambda). Unambiguous class I sequencing (HLA-A, B and C) was performed by DNA sequencing. Group-specific amplification separated the two genes at each locus by PCR and each was sequenced separately (Dunn *et al.*, 2003). The PCR was performed using Expand High Fidelity PCR System (Roche Diagnostics Deutschland GmbH). The amplicons were purified using AMPure magnetic beads (Agencourt Bioscience Corporation) and then subjected to cycle sequencing using BigDye[™] terminators (Life Technologies Corporation). Generally, 6-8 cycle sequencing reactions were performed to obtain DNA sequences (both strands) of exons 2-4 (HLA-A, -B) or exons 2-7 (HLA-C). Unincorporated fluourescent dyes were removed using CleanSEQ beads (Agencourt Bioscience Corporation) and sequencing reaction products were then separated by capillary electrophoresis using a 3130XL Genetic Analyser (Life Technologies). DNA sequences were compiled, analysed and assigned using SBTengine (GenDx, Utrecht, Netherlands).

2.6 PCR-SSP typing of ABO, Rhesus, MNS Kell, Kidd, Duffy and HPA

Commercial DNA-SSP kits (BAG Health Care GmbH, Lich, Germany) were used to genotype ABO (ABO-TYPE variant), Rhesus (RH-TYPE), MNS (MNS-TYPE), Kell, Kidd and Duffy (KKD-TYPE) and human platelet (HPA-TYPE) loci. Each reaction tube contains a prealiquoted and dried reaction mixture of oligonucleotide primer pairs (allele specific and internal control for human growth hormone) and deoxynucleotides. The full reaction mixture

Chapter 2

(10 ul) was then make up by adding with 1 ul of 10x PCR master mix (BAG Health Care GmbH, Lich, Germany), 1ul of DNA (50-100 ng/ul), 8 ul distilled water and 0.08 ul of *Taq* DNA Polymerase (5 U/ul, Roche Diagnostics Deutschland GmbH). Each tube was closed with the provided lid and amplified using GeneAmp[®] PCR System 2700 (Applied Biosystems). Thermal cycling regime is according to the manufacturer's instruction.

Agarose gel electrophoresis was used to to detect the presence of specific PCR products of known size. A 2% agarose gel was prepared by adding 3 g of agarose powder (AppliChem GmbH, Germany) into 150 ml of 0.5X TBE buffer (Roche Diagnostic, GmbH, Germany) and ethidium bromide (0.5 μ g/ml) stain (2 μ l per 100 ml gel solution). A total volume of 10 μ l of the PCR products was loaded into each well. A small volume, 5 μ l of 100 bp DNA size standard ladder (250 μ g/ml) was loaded on each sample row containing PCR amplified products. Electrophoresis was carried out at 80 V for 45 minutes. The product band was visualized and recorded by using image analyzer (UVITEC, Cambridge). The presence of specific PCR products was interpreted by using BAGenotype evaluation software version 1.4.

2.7 Sequence based typing of Kell, Duffy, JK_{null} and HPA-6

Genotyping protocols for Kell, Duffy and JK_{null} (i.e. g>a mutation at the 3' acceptor splice site of intron 5) were described in detail in Chapter 3.3 (i.e. Table S2). The HPA-6 locus was genotyped using forward (5'-CTGGCTGGCTGGGATCCCAGTG-3) and reverse (5'-CCCTGCAGTTCTCCTCACCTGAG-3) primers and amplified using a PCR protocol previously described by Tanaka *et al.* (1996). Briefly, target DNA loci was amplified using a thermo cycler (GeneAmp PCR System 9700, Applied Biosystems, Foster City, CA, USA) in a 50 µl mixture containing 4 ul of 15 ng/ul genomic DNA, 2 ul of each forward and reverse primers (10 uM each), 25 ul of MyTaqTM Red Mix (Bioline) and 17 ul of distilled water. The PCR conditions were as follows: 28 cycles at 95°C for 1 minute, 70°C for 1 minute, 72°C for 1 minute, with an initial denaturation at 95°C for 3 minutes and a final extension at 72°C for 5 minutes. The amplified products were purified using DNA Clean & ConcentratorTM-5 (Zymo Research Corporation) and sequenced using multi-color fluorescence-based technology on ABI3730 DNA Analyzer (Applied Biosystems) at Massey University, New Zealand (Massey Genome Service).

2.8 Blood group and HPA genotyping using chip hybridization SNP assay

High-throughput molecular blood group genotyping was performed using hybridization SNP assay (BLOODchip[®] v2.0) service provided by Progenika Biopharma S.A., Spain (http://www.progenika.com). Briefly, a total of 45 specific regions of genomic DNA ($\geq 2 \mu g$, 50 ng/ μ l) were amplified and labelled using 3 different master mixes. The labelled amplicons were then allowed to hybridize with complementary oligonucleotide probes spotted on glass slides. The oligonucleotide probes are designed to detect 128 SNPs that belong to 9 blood group (ABO, Rhesus, MNS, Kell, Kidd, Duffy, Colton, Dombrock and Diego) and 12 platelet antigen (HPA-1 to -11 and -15) loci. After a washing procedure to remove excess amplicons, the fluorescently labelled amplicon-oligonucleotide complex were measured using a scanner and the presence of specific SNPs are interpreted by the BLOODchip® v2.0 software.

2.9 Statistical analysis

2.9.1 Allele, gene and carrier frequencies

Allele and carrier frequencies were obtained by direct counting. Gene frequency (*GF*) for KIR was calculated using the formula $GF = 1 - \sqrt{1 - F}$ where *F* is observed carrier frequency.

2.9.2 Haplotype frequency

Haplotypes are unknown when individuals are heterozygous for more than one locus. In this study, an expectation-maximization (EM) algorithm integrated in the Arlequin software package was used to estimate haplotypes from genotypic data with unknown gametic phase. Estimation of linkage disequilibrium (D) between alleles at 2 and 3 different loci and their level of significance (p) is described in Chapter 3.1.

2.9.3 Exact tests for Hardy-Weinberg equilibrium

The exact tests were carried out on an extended two-by-two contingency table of arbitrary size (Arlequin) or by chi-squared tests (SPSS Inc., Chicago, IL) and were used to test for Hardy-Weinberg equilibrium (HWE). Deviation from HWE is considered significant at the p-value of < 0.05, with 1 degree of freedom.

2.9.4 Linkage disequilibrium

Determination of linkage disequilibrium (LD) between pairs of loci was done by using exact tests under the null-hypothesis of no association between the two tested loci (Excoffier *et al.*, 2006). Cramer's V statistic (i.e. Pearson's correlation) was also used to estimate LD between pairs of KIR loci (after Single *et al.*, 2008). The correlation coefficient (*r*) is equivalent to Wn and has a range from -1 (negative association) to +1 (positive association). The significant association (p < 0.05) of LD between two loci was then estimated using Fisher's exact test.

2.9.5 Testing for natural selection

2.9.5.1 Deviation from HWE

According to the HWE model, gene variant frequencies in the genepool of an infinitely large and randomly breeding population remains constant throughtout generations provided there is no influence from natural selection, genetic drift, mutation, gene flow and non-random mating. This model is represented by the equations below:

$$\mathbf{p} + \mathbf{q} = \mathbf{1}$$

where p and q is frequencies of a gene with two alleles, f(A) and f(B):

AA
$$(p^2) = f(A) \bullet f(A)$$

AB and BA $(2pq) = 2 f(A) \bullet f(B)$
BB $(q^2) = f(B) \bullet f(B)$

Thus, genotype frequencies are equal to:

$$p^2 + 2pq + q^2 = 1$$

where observed genotypic data can be compared with expectation via a chi-squared test with one degree of freedom.

Deviation from HWE (as heterozygote excess or deficiency) can arise either from violation of one of the assumption of the model or by natural selection. However, deviation for expectation do not indicate either the direction or intensity of selection.

2.9.5.2 The Ewens-Watterson neutrality test

The Ewens-Watterson neutrality test was performed to test hypothesis for neutral evolution against either balancing or directional selection. This test is based on Ewens' sampling theory and the Infinites Alleles Model of neutral theory. The test uses the observed allele frequency spectrum against that expected using the model. The action of natural selection in the population is considered significant if having *p*-value < 0.05 (see Chapter 3.2 for details).

2.9.5.3 T-test

Selection was tested for pairs of KIR and HLA ligands using a T-test with a total of 10,000 permutations. The *p*-value (i.e. of the Pearson correlation co-efficient) was calculated as a percentage of the 10,000 correlation estimation (see Chapter 3.5 for details).

2.9.5.4 Ancestral fraction scaling test

This is a novel method developed in Chapter 3.1 of this thesis. Pricipal coordinate (PCO) analysis was also used to search for evidence of natural selection in Polynesian sub-

populations. Ancestral fractions in these populations were estimated by using set of reference A-SM and P-SA genepools (see Chapter 3.1 for details). The relative ancestral fraction in test population was estimated by projecting the test population onto the line joining the two reference pools. Loci that showed large shifted towards one or other ancestral population were interpreted as having under gene selection favouring alleles that were more common in that ancestor.

2.9.6 Exact test of population differentiation

Homogeneity and heterogeneity between datasets were evaluated using exact or chi-squared (with Yate's continuity correction) tests performed using Arlequin (Excoffiler *et al.*, 2006) and Graphpad (http://www.graphpad.com) software, respectively. Two datasets are considered significantly different if the return *p*-value is < 0.05.

2.9.7 Phylogenetic analysis

Phylogenetic analysis was achieved using programs included in PHYLIP version 3.6 (Felsenstein, 2004). Bootstrap and genetic distance analyses were performed by Seqboot and Gendist programs, respectively. The neighbour-joining routine was used to construct evolutionary trees and a single consensus bootstrapped tree was obtained with Consense program (Felsenstein, 2004). Topology of the consensus tree was visualized with TreeView program (Page, 1996).

Phylogenetic analysis summarizes similarities and differences in a single diagram under the null assumption that they were generated by a strictly tree-like process. Organisms that share

a common ancestor are expected to evolve and show hierarchial clustering. If this natural hierarchical distribution is violated or erased, e.g. due to admixture and hybridization, then phylogenetic analysis may not resolve relationships reliably and yield trees with branches supported by low bootstrap values.

2.9.8 Principal coordinate and component analysis

Principle coordinate (PCO) and component analyses (PCA) were carry out by using Multivariate Statistical Package 3 (MVSP3) (Kovach Computing Services, UK; http://www.kovcomp.com/mvsp). The PCO and PCA transform dissimilarity and similarity matrixes (respectively) into uncorrelated variables called principal coordinates/components (respectively). The PCO/PCA methods collapse genetic variability in several axes and the two most informative were usually represented in a two-dimensional plot. Each population is plotted as a 2-dimentional scatter graph and the distance between populations demonstrates their genetic relationship. Pairs of populations with closer genetic relationship will be plotted together, while those which are distinct will be separated.

2.9.9 Probability of an exact match

Probability of finding a phenotype match between randomly selected donors and recipients was calculated according to the following formula (after Badjie *et al.*, 2011):

where A and B = antigens of particular locus AA = homozygote for A antigen AB = heterozygote for A and B antigensBB = homozygote for B antigen

2.9.10 Probability of alloimmunization

Probabilities of transfusion and pregnancy alloimmunization were calculated according to the following formula (after De La Vega Elena *et al.*, 2008):

i. Probability of transfusion alloimmunization = $(1 - p_p)^2 - (2p_d - p_d^2)$ where p_p = gene frequency of patient's population p_d = gene frequency of donor's population ii. Probability of pregnancy alloimmunization = $(1 - p_m)^2 X 2p_p$ where p_m = gene frequency of mother's population p_p = gene frequency of father's population

References

Badjie, K.S.W., Tauscher, C., Buskirk, C.V., Wong, C., Jenkins, S., Smith, C. & Stubbs, J.R. (2011) Red blood cell phenotype matching for various ethnic groups. *Immunohematology*, **27**, 12-19.

De La Vega Elena, C.D., Nogues, N., Montoya, A.F., Chialina, S., Blanzaco, P.D., Theiller, E., Raillon, M.A., Arancegui, N., Solis, E., Oyonarte, S., Ferrer, V.C., Munoz, A.C. & Muniz-Diaz, E. (2008) Human platelet-specific antigens frequencies in the Argentinean population. *Transfusion Medicine*, **18**, 83-90.

Dunn, P.P.J., Day, S., Williams, S. & Bendukidze, N. (2003) HLA sequencing as a tissue typing tool. In: *Pediatric Hematology* (eds. Goulding, N. & Seward, C.), 233-246. Humana Press, Totowa, New Jersey, USA.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47-50.

Felsenstein, J. (2004) *PHYLIP (Phylogeny Inference Package) version 3.6.* Department of Genome Sciences, University of Washington, Seattle.

Hamilton, J.F. (1993) *Multi-locus and and single-locus DNA profiling in New Zealand*. PhD thesis, Victoria University of Wellington, Wellington, New Zealand.

Kotsch, K., Wehling, J. & Blaszcyk, R. (1999) Sequencing of HLA class II genees based on the conserved diversity of the non-coding regions: sequencing based typing of HLA-DRB genes. *Tissue Antigens*, **53**, 486-497.

Marshall, S.J. (1994) *Multi-locus and and single-locus DNA profiling in New Zealand*. PhD thesis, Victoria University of Wellington, Wellington, New Zealand.

Page, R.D.M. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357-358.

Single, R.M., Martin, M.P., Meyer, D., Gao, X. & Carrington, M. (2008) Methods for assessing gene content diversity of KIR with examples from a global set of populations. *Immunogenetics*, **60**, 711–725.

Tanaka, S., Taniue, A., Nagao, N., Tomita, T., Ohnoki, S., Shibata, H., Okubo, Y., Yamaguchi H. & Shibata Y. (1996) Genotype frequencies of the human platelet antigen, Ca/Tu, in Japanese, determined by PCR-RFLP method. *Vox Sanguinis*, **70**, 40-44.

Using HLA loci to inform ancestry and health in Polynesian and Maori populations

H.A. Edinur¹, P.P.J. Dunn², L. Hammond², C. Selwyn², Z.M Velickovic³, R.A. Lea⁴ and G.K. Chambers¹

¹School of Biological Sciences, Victoria University of Wellington, New Zealand ²New Zealand Blood Service, Auckland

³Molecular Genetics, Tissue Typing, Australian Red Cross Blood Service, Sydney, Australia ⁴School of Medical Sciences, Griffith University, Australia

Status: *Tissue Antigens* 2012, **80**, 509-522.

Key words: HLA, Polynesia, Polynesians, Maori, admixture, ancestry and human health **Correspondence**: Dr Geoffrey K. Chambers, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand. Ph: +64-(0)4-463-6091, Fax: +64-(0)4-463-5331, e-mail: Geoff.Chambers@vuw.ac.nz

Abstract

Human Leukocyte Antigens (HLA) are important genetic markers of tissue identity and accurately reflect ancestral history. The work reported in this paper provides a detailed description of HLA polymorphism in Polynesian and Maori individuals in relation to other populations. Our study concerns HLA class I and II antigens in Polynesian (N = 36) and Maori (N = 114) subjects genotyped at 2 digit resolution by New Zealand Blood Service Laboratory in Auckland using PCR-SSO Luminex and PCR-SSP technologies. We have also compared our data with those from other Austronesian-speaking Mongoloid and Papuan-speaking Australoid populations in order to test previously published account of the origin of Proto-Polynesians via gender-biased gene flow between these two ancestral populations. We use principal coordinate (PCO) analysis for this purpose, arguing this approach to be superior to tree-based methods, due to factors associated with population history and admixture. Our data are in general agreement with earlier work and reflect received wisdom on the dual origin of Proto-Polynesians. They also show the way in which the genetic make up of Polynesian and Maori subjects is changing due to intermarriage with Europeans.

Introduction

Austronesian-speaking Mongoloids (A-SM) from East Asia (Taiwan) first migrated to South East Asia (SEA), coastal New Guinea and Island Melanesia between 5000 to 3,500 years ago, intermixed with Papuan-speaking Australoid (P-SA) residents (1). Their descendants moved outwards to cross the Pacific basin Remote Oceania and occupied New Zealand from 650 ybp. On the route to New Zealand, the genepool ultimately inherited by Polynesians and local Maori people was repeatedly refined by founder effects due to multiple re-settlements; aka the '*Genetic bottleneck in Polynesia*' and see (2) for more details. This means that Polynesian and Maori people are on average more closely related to one another than the members of large outbred populations; say those in Asia or Europe, would be. The recent period has seen the beginning of a further phase of gene flow as Polynesian and Maori people intermarry with Europeans.

Studies on immune system loci have played a leading part in the development of genetic evidence that underpins the present account (3-30). Our present study examines HLA markers in DNA samples that have been classified according to reported admixture. Hence, this new work includes the only individuals ever examined where some participants can be clearly identified as lacking genetic input from European ancestors. This design allows comparisons that were not previously possible. This work represents an independent resampling of these populations, and thus has comparative value too. However, these new data can be combined with previous data on Polynesians and Maori in a small scale meta-analysis.

In the present study, our historical reconstructions compared principal coordinate analysis (PCO) and neighbour-joining (NJ) methodologies. The HLA data generated from this study will be tested under the *Total Synthetic Evidence Model* of Pacific settlement and specifically allowing for recent admixture. They will also be interpreted for the first time in the light of new knowledge about Polynesians genetics including tests for historical geneflow with a 70:30 ratio (A-SM:P-SA) as a null hypothesis based on the SNP data (31) and tested versus values calculated from (32). Wider comparisons have also been made in a new and more appropriate context recognising other Austronesian populations (e.g. Malay and Taiwanese Aboriginals) as being distinct from Polynesians and as representing ancestral type Austronesian-speaking Mongoloid lineages. In making these last comparisons we have also critically scrutinised population labels applied in previous reports for authentic content; e.g. does 'Taiwan' really mean Han Chinese *vs.* Atayal, Bunun and Paiwan etc. Overall, we believe that our compilation and comparison of HLA class I and II in this study and previous works fill some of the gaps left in our present understanding of Polynesian immunogenetics.

Materials and Methods

Samples, HLA genotyping and ethical clearance

DNA samples were obtained from the Victoria University of Wellington DNA Bank and comprise Total Polynesians (TP: N = 36) and Total Maori (TM: N = 114) groups. Each individual was assigned into one of two sub-groups based on their self-reported pedigree: TP = Polynesians with Full Ancestry (PFA: N = 24) + Polynesians with Admixed History (PAH: N = 12) and TM = Maori with Full Ancestry (MFA: N = 49) + Maori with Admixed History (MAH: N = 65). Ethical approvals were obtained from the New Zealand Central Region Ethic Committee and Victoria University Human Ethics Committee (HEC). The HLA class I and II alleles were typed using polymerase chain reaction-sequence specific oligonucleotide

(PCR-SSO) using Luminex (LABType[®] One Lambda[®], CA, USA) and ambiguities were resolved by using Life Technologies AllSetTM Gold HLA low resolution SSP kits.

Statistical analysis

Allele frequencies were obtained by the direct counting method. Direct gene counting could be applied because no typing ambiguities were found. Individuals are considered homozygotes if only one type of allele group detected at particular HLA locus. Arlequin software version 3.1 (33) was used to estimate allele and haplotype frequencies, to test for Hardy-Weinberg equilibrium and linkage disequilibrium between pairs of HLA loci and carry out exact test of population differentiation. Linkage disequilibrium (D) between two alleles at two different loci and level of significance (p-value = 0.05) were calculated using Arlequin software (33). The generated linkage disequilibrium (D) values were then used to calculate linkage disequilibrium (D) between three alleles at three different loci. Formula for calculating 3-locus disequilibrium and their level of significance (p-value = 0.05) are Dkri = Pkri - qiDkr - yrDki - pkDri - pk . yr . qi and $\chi^2 = 2N(Dkri)^2 \div pk(1-pk)$ yr(1-yr) qi(1-qi), respectively (34,35). The Dkri is the linkage desequilibrium (D) of the haplotype kri, Pkri is the frequency of the haplotype kri, qi is frequency of the allele i, yr = the frequency of the allele r, pk = is the frequency of the allele k, Dkr = the linkage designilibrium (D) of thehaplotype kr, Dki = the linkage desiquilibrium (D) of the haplotype ki, <math>Dri = the linkagedesiguilibrium (D) of the haplotype ri and N = the number of samples. Phylogenetic trees were constructed using PHYLIP software package, version 3.6 (36) and the TreeView programme (37). Principal coordinate (PCO) plots were constructed using Multivariate Statistical software package 3 (Kovach Computing Services, UK; http://www.kovcomp.com/mvsp). Proportions of A-SM:PSA genepools in Polynesian and Maori populations were estimated by using set of simulations blending representative of two individuals A-SM and P-SA populations (e.g. Ami vs. Goroko and Ami vs. Rangers) in

various combinations with a range of proportions 50:50 to 90:10 A-SM:P-SA. The A-SM:P-SA combinations were then compared with Polynesian and Mori datasets using X^2 goodness of fit (SPSS Inc., Chicago, IL) and exact test of population differentiation (Arlequin software version 3.1 – ref 33). Admixture coefficient in Polynesian and Maori populations were also estimated by plotting a full range simulated proportions (0:100 to 100:0) of average genetic profiles of two broad groups, A-SM:P-SA genepools on PCO plots.

Results

Name and frequency of HLA alleles (based on two digit allele groups) identified across the class I and II loci in our Polynesian and Maori study groups and sub-groups are shown in Table 1. There is no significant deviation from Hardy-Weinberg equilibrium at any of the six loci (Table S1) and tests for pairwise linkage disequilibrium were significant except for the HLA-A-B and HLA-A-C pairs in PAH (Table S2). The 2- and 3-locus class I and II haplotypes with significant associations between alleles at different loci estimated for Polynesians and Maori in our new survey are listed in Table 2.

Our new HLA class I and II data for PFA closely resemble those from four different Polynesian populations (Cook Islands – CI, Tokelau – TOK, Tonga – TON and Samoa – SA) previously collected (3,4), but which have not previously appeared in the primary literature – see Table 1. This observation holds for both total numbers and types of alleles found. Exact tests of population differentiation on the HLA class I and II data show few significant differences between the four populations except to differentiate Cook Islands and Tonga in HLA class II comparisons (see Table S3). We took these observations as justification for combining these four populations into a larger unified group; VTP (Velickovic Total Polynesians), see Table 1 for comparison with our newly collected data for Polynesians. Exact tests using HLA class I, II or combined of class I and II allele frequencies (see Table S3) shows that the PFA sub-group is most similar to VTP and that PFA is only significantly different from Tokelau (p-value < 0.05), one of the four individual sub-groups that make up VTP. Hence, we judge it legitimate to combine PFA with VTP into PMD (Polynesian Meta Dataset – see Table 1) for comparisons with other (ancestral) populations (see later). The statistically significant differences observed between PFA and PAH may be attributed to recent admixture with Europeans and this is the subject of the next set of analyses with reference European populations – see later.

The 2- and 3-locus haplotypes associations found for Polynesians in our new survey also compare well with those with those previously obtained (3,4) – see Table S4. Although haplotype frequencies do sometimes show quite marked differences among the four Island groups, the combined VTP data set matches PFA fairly closely, particularly with respect to the more common associations discussed above. Hence, we created a haplotype metadata group (PMD) by combining them two.

We also compared the HLA allele frequency scores of our Maori study group with those from a previous study (5,6) on Maori (TTM). In summary, HLA data for MFA and TTM are similar in terms of the most frequent alleles observed. Many alleles which are not detected in MFA were recorded for TTM, e.g. *HLA-A*29*, *-B*35*, *-B*37*, *-B*58* and *-C*16*. All of these are present in MAH. This suggests TTM is more similar to either our MAH sub-group or the wider TM group which contain alleles due to admixture with Europeans. Therefore, we ran exact tests using MFA, MAH and TTM datasets and all comparisons were significantly different from one another (see Table S3). In contrast to Polynesians, where VTP is equivalent to PFA in terms of un-admixed Polynesian gene content, here we have shown TTM most closely resembles MAH and suggests greater degree of admixture in (5,6) datasets than (3,4). To best illustrate these comparisons, we have plotted HLA class I and II allele frequency spectra for the various Maori (Figs. 1 and 2) and Polynesian (not shown) datasets against the two reference European populations, ELAN (class I-ref 7) and ENG (class II-ref 8). In all of the graphs, vertical allele frequency bars plotted for PAH, MAH and TTM are intermediate between values represented by PFA/MFA and ELAN/ENG or shared just with the latter. Hence, we judge it to be optimal to use PMD and MFA as reference groups for studying Polynesian and Maori genealogy. Our original plan to create the largest possible Polynesian and Maori meta datasets by combining TP and VTP and TM and TTM had to be rejected due to extensive European contributions in both, PAH and TTM/MAH, respectively. Therefore, HLA data from the earlier work on TTM is specifically excluded from Tables 1 and S4 due to admixture.

Next, we asked if Maori is typical of other Polynesian populations. The HLA data from MFA are then compared with other Polynesian datasets (see Table 1). In general, the most common alleles in MFA are also typical for the other Polynesians; except for *HLA-B*56*, -*DQA1*03* and -*DRB1*04* which are less frequent in MFA compared to other Polynesian populations. Similar patterns can be seen with the distributions of HLA class I and II haplotypes, but with a few minor exceptions such as; *HLA-DRB1*08-DQB1*05* is only detected in MFA and CI while *HLA-DRB1*08-DQA1*01-DQB1*05* is only observed in MFA. Overall, the new data collected for MFA are consistent with those from other Polynesians, although allele and haplotype frequencies do show some slight differences between sub-populations. It is obvious that variations will arise between individual populations due to their history and the widely accepted way to examine their genetic relationships is by using Neighbour-Joining (NJ) phylogenetic trees and PCO plots. Here we show only those from the latter (Figs. 3 and 4 and Figs. S1 and S2) but note that the NJ trees (not shown) were broadly congruent. We argue in discussion that the PCO method is a superior tool when admixture is a feature of population history. The HLA-A, -B and -C and HLA-DRB1 and -DQB1 PCO diagrams constructed using information obtained from axis 1 and 2 (Figs. 3 and 4, respectively) are perhaps more informative and show clear separation of A-SM and P-SA than those constructed using axis 1 and 3 (Figs. S1 and S2). In Figs. 3 and 4, Polynesians fall as intermediate between cumulative of A-SM and P-SA genepools and Cook Islanders are closer to MFA compared with the other Polynesian populations. Nevertheless, our admixed groups (MAH and TM) as well as TTM are plotted as displaced toward European datasets, ELAN and ENG in the HLA-A, -B and -C and HLA-DRB1 and -DQB1 plots, respectively. We state here that there are no marked differences between PCO plots constructed for combined and original populations from those based only on the original populations (for example see Fig. S3 compared with Fig. 3).

Because comparative genotypic data were not available for wider P-SA and A-SM groups, we were unable to carry out *STRUCTURE* analysis and apportion the genepool of our Polynesian and Maori subjects between their ancestral groups, cf. (31) and (32) studies using SNP data. Therefore, we ran an extensive set of simulations blending representative of two individual P-SA and A-SM (e.g. Ami vs. Goroko) populations in various combinations with a range of proportions 50:50 to 90:10 A-SM:P-SA. We then carried out X^2 goodness of fit analysis and exact tests of population differentiation comparing PFA, VTP and PMD to each of these. All proved to be significantly different (*p*-values = 0, e.g. see Table S5). We

therefore resorted to plotting a full range of simulated proportions (0:100 to 100:0) of A-SM:P-SA genepools on the PCO plots and estimating admixture by projecting the individual points for PMD, VTP etc. on to the line joining the simulated points (see Figs. S4 to S8). For Polynesians, these gave intersections around 60:40 for HLA-A (Fig. S4) and 40:60 for HLA-C (Fig. S6) loci but values fell in the 90:10 to 100:0 intervals for HLA-B (Fig. S5) and may reflect past episodes of natural selection (see Discussion). The only locus for which there are sufficient data to attempt calculation of admixture proportions from HLA class II is HLA-DRB1, since there are no data for Taiwan aborigines for the other class II loci. Here, all Polynesian populations including PFA and the composites VTP and PMD fell close to 100:0 proportion of A-SM:P-SA genepools (Fig. S7). Our estimations show that they are around 30:70 for HLA-A and -C, and 100:0 for HLA-B and -DRB1 (see Figs. S4-S7) proportion of A-SM:P-SA ancestral fractions in Maori. Overall, our simulation using a combined HLA-A, -B, -C and –DRB1 allele frequencies show that there is around a 60:40 genetic proportion of ASM:P-SA ancestral fractions in Maori and other Polynesians (see Fig. S8).

Datasets (N)	TP (36)	PFA (24)	PAH (12)	TM (114)	MFA (49)	MAH (65)	CI (50)	TOK (50)	TON (50)	SA (50)	VTP (200)	PMD (224)	TTM (§)	MMD (§)
HLA-A*														
01	0.06	0.02	0.13	0.07	0.05	0.08	0.01	0.01	-	0.01	0.01	0.01	0.05	0.06
02	0.22	0.21	0.25	0.33	0.36	0.31	0.22	0.17	0.12	0.34	0.21	0.21	0.34	0.34
03	-	-	-	0.04	-	0.07	0.01	-	-	-	0.00	0.00	0.06	0.05
11	0.11	0.15	0.04	0.18	0.16	0.18	0.11	0.16	0.21	0.11	0.14	0.14	0.17	0.17
23	-	-	-	-	-	-	-	-	-	-	0	-	0.00	0.00
24	0.40	0.39	0.42	0.27	0.38	0.19	0.48	0.56	0.50	0.46	0.50	0.49	0.23	0.25
25	-	-	-	0.01	-	0.02	-	-	-	0.01	0.00	0.00	0.00	0.01
26	0.03	0.04	0-	0.01	-	0.02	0.02	0.01	0.06	-	0.03	0.03	0.01	0.01
29	0.01	-	0.04	-	-	-	-	-	-	-	0	-	0.01	0.01
30	-	-	-	0.00	-	0.01	0.01	-	-	-	0.00	0.00	0.00	0.00
31	-	-	-	0.00	-	0.01	-	0.01	-	-	0.00	0.00	0.01	0.01
32	-	-	-	0.00	-	0.01	-	-	-	-	-	-	0.01	0.01
33	-	-	-	0.00	-	0.01	-	-	-	-	-	-	0.00	0.00
34	0.17	0.19	0.13	0.07	0.05	0.08	0.13	0.08	0.11	0.07	0.10	0.11	0.08	0.08
43	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00
66	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00
68	-	-	-	0.01	-	0.02	0.01	-	-	-	0.00	0.00	0.01	0.01
k	7	6	6	13	5	13	9	7	5	6	11	11	17	17
HLA-B*														
07	0.01	-	0.04	0.05	0.02	0.07	-	-	-	0.03	0.01	0.01	0.13	0.10
08	0.01	-	0.04	0.03	0.02	0.04	0.01	-	-	0.01	0.01	0.00	0.02	0.03
13	0.06	0.06	0.04	0.02	0.04	-	0.06	0.04	0.03	0.06	0.05	0.05	0.04	0.03
14	-	-	-	0.02	-	0.04	-	-	-	-	-	-	0.00	0.01
15	0.06	0.02	0.13	0.04	0.01	0.05	0.03	0.08	0.05	0.04	0.05	0.05	0.04	0.04
18	0.01	-	0.04	0.01	-	0.02	-	-	-	-	-	-	-	0.00
27	0.01	0.02	-	0.01	0.01	0.01	0.02	0.01	0.02	-	0.01	0.00	-	0.00
35	-	-	-	0.00	-	0.01	-	-	-	-	-	-	0.02	0.02
37	-	-	-	0.00	-	0.01	-	-	-	-	-	-	0.01	0.01
38	-	-	-	0.00	-	0.01	-	-	-	-	-	-	-	0.00

Table 1: HLA class I and II allele frequencies for Polynesian and Maori datasets.

Table 1 cont.

Datasets (N)	TP (36)	PFA (24)	PAH (12)	TM (114)	MFA (49)	MAH (65)	CI (50)	TOK (50)	TON (50)	SA (50)	VTP (200)	PMD (224)	TTM (§)	MMD (§)
HLA-B*														
39	0.06	0.06	0.04	0.11	0.09	0.12	0.12	0.06	0.03	0.06	0.07	0.07	0.05	0.07
40	0.32	0.42	0.13	0.18	0.20	0.16	0.19	0.51	0.35	0.41	0.37	0.38	0.18	0.18
44	0.04	0.02	0.08	0.05	0.02	0.08	-	-	-	-	-	0.00	0.01	0.03
45	-	-	-	0.00	-	0.01	-	-	-	-	-	-	-	0.00
47	0	-	-	0.00	-	0.01	-	-	-	-	-	-	-	0.00
48	0.07	0.10	-	0.12	0.14	0.10	0.11	0.01	0.11	0.11	0.09	0.09	0.17	0.15
50	-	-	-	0.01	-	0.02	-	-	-	-	-	-	0.01	0.01
51	-	-	-	0.00	-	0.01	0.02	-	-	-	0.01	0.01	0.01	0.01
52	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00
54	-	-	-	0.00	-	0.01	-	-	-	-	-	-	0.02	0.01
55	0.18	0.15	0.25	0.25	0.37	0.16	0.22	0.19	0.07	0.12	0.15	0.15	0.21	0.22
56	0.15	0.13	0.21	0.06	0.07	0.05	0.20	0.09	0.34	0.16	0.20	0.19	0.03	0.04
57	0.01	0.02	-	0.00	-	0.01	-	-	-	-	-	0.00	0.02	0.01
58	-	-	-	0.01	-	0.02	-	-	-	-	-	-	-	0.00
67	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.00
81	-	-	-	-	-	-	-	-	-	-	-	-	0.02	0.01
Blank	-	-	-	0.01	-	0.02	-	-	-	-	-	-	-	0.00
k	13	10	10	23	11	22	10	8	8	9	11	13	20	26
HLA-C*														
01	0.33	0.27	0.46	0.30	0.45	0.18	0.40	0.25	0.35	0.24	0.31	0.31	0.25	0.27
02	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00
03	0.13	0.15	0.08	0.04	0.02	0.06	0.03	0.01	0.12	0.25	0.10	0.11	0.04	0.04
04	0.18	0.21	0.13	0.12	0.17	0.08	0.13	0.46	0.20	0.17	0.24	0.22	0.17	0.15
05	0.03	0.02	0.04	0.05	0.01	0.08	0.01	0.01	-	-	0.01	0.01	0.04	0.05
06	0.01	0.02	-	0.03	-	0.05	0.01	-	-	-	0.00	0.00	0.02	0.02
07	0.11	0.08	0.17	0.20	0.14	0.24	0.16	0.07	0.08	0.12	0.11	0.11	0.19	0.19
08	0.07	0.10	-	0.15	0.13	0.16	0.11	0.01	0.12	0.11	0.09	0.09	0.11	0.13
12	0.06	0.06	0.04	0.03	0.04	0.02	0.04	0.12	0.06	0.03	0.06	0.06	0.03	0.03
14	-	-	-	-	-	-	0.02	-	-	-	0.01	0.01	0.00	0.00

Table 1 cont.

Datasets (N)	TP (36)	PFA (24)	PAH (12)	TM (114)	MFA (49)	MAH (65)	CI (50)	TOK (50)	TON (50)	SA (50)	VTP (200)	PMD (224)	TTM (§)	MMD (§)
HLA-C*														
15	0.07	0.08	0.04	0.06	0.03	0.08	0.09	0.07	0.07	0.08	0.08	0.08	0.08	0.07
16	0.01	-	0.04	0.00	-	0.01	-	-	-	-	-	-	0.04	0.02
18	-	-	-	-	-	-	-	-	-	-	-	-	0.03	0.01
Blank	-	-	-	0.02	-	0.03	-	-	-	-	-	-	-	0.01
k	10	9	8	10	8	10	10	8	7	7	10	10	13	13
HLA-DQA1*														
01	0.36	0.35	0.38	0.29	0.26	0.32	-	-	-	-	-	-	-	-
02	-	-	-	0.04	-	0.07	-	-	-	-	-	-	-	-
03	0.38	0.44	0.25	0.21	0.27	0.17	-	-	-	-	-	-	-	-
04	0.01	0.02	-	0.02	0	0.04	-	-	-	-	-	-	-	-
05	0.22	0.19	0.29	0.41	0.48	0.36	-	-	-	-	-	-	-	-
Blank	0.03	0	0.08	0.03	0	0.05	-	-	-	-	-	-	-	-
k	4	4	3	5	3	5								
HLA-DOB1*														
02	0.01	-	0.04	0.07	0.02	0.10	0.02	-	-	-	0.01	0.00	0.10	0.08
03	0.54	0.56	0.50	0.58	0.68	0.50	0.64	0.66	0.56	0.58	0.61	0.60	0.52	0.54
04	0.06	0.08	-	0.04	0.04	0.04	0.03	0.08	0.05	0.11	0.07	0.07	0.01	0.02
05	0.15	0.17	0.13	0.15	0.12	0.17	0.22	0.10	0.18	0.14	0.16	0.16	0.20	0.18
06	0.21	0.19	0.25	0.14	0.13	0.15	0.09	0.16	0.21	0.17	0.16	0.16	0.18	0.17
Blank	0.03	-	0.08	0.03	-	-	-	-	-	-	-	-	-	0.01
k	5	4	4	5	5	5	5	4	4	4	5	5	5	5
HLA-DRB1*														
01	0.03	0.02	0.04	0.04	0.01	0.05	0.01	-	-	-	0.00	0.00	0.03	0.03
03	0.01	-	0.04	0.04	0.02	0.06	0.01	-	-	0.02	0.01	0.01	0.04	0.04
04	0.22	0.25	0.17	0.15	0.16	0.14	0.28	0.46	0.22	0.25	0.30	0.29	0.18	0.17
07	-	-	-	0.04	-	0.07	0.01	-	-	0.01	0.01	0.00	0.05	0.05
08	0.10	0.10	0.08	0.13	0.18	0.09	0.10	0.12	0.13	0.12	0.12	0.12	0.10	0.11

Table 1 cont.

Datasets (N)	TP (36)	PFA (24)	PAH (12)	TM (114)	MFA (49)	MAH (65)	CI (50)	TOK (50)	TON (50)	SA (50)	VTP (200)	PMD (224)	TTM (§)	MMD (§)
HLA-DRB1*														
09	0.15	0.19	0.08	0.06	0.10	0.03	0.08	0.04	0.16	0.25	0.13	0.14	0.04	0.05
11	0.07	0.06	0.08	0.16	0.18	0.15	0.12	-	0.16	0.08	0.09	0.09	0.15	0.16
12	0.14	0.13	0.17	0.21	0.28	0.16	0.17	0.22	0.08	0.08	0.14	0.14	0.22	0.22
13	0.04	0.02	0.08	0.03	-	0.05	0.01	0.03	-	-	0.01	0.01	0.03	0.03
14	0.14	0.15	0.13	0.09	0.05	0.12	0.14	0.07	0.12	0.09	0.11	0.11	0.10	0.10
15	0.10	0.08	0.13	0.04	0.01	0.05	0.07	0.06	0.13	0.10	0.09	0.09	0.06	0.05
16	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.00
Blank	-	-	-	0.01	-	0.02	-	-	-	-	-	-	-	0.00
k	10	9	10	11	9	11	11	7	7	9	11	11	12	12

N = total number of individuals, k = total number of observed alleles, 0=not observed, 0.00 = <0.01, TP: Total Polynesians, PFA: Polynesians with Full Ancestry, PAH: Polynesians with Admixed History, TM= Total Maori, MFA = Maori with Full Ancestry, MAH: Maori with Admixed History, CI: Cook Islands (3,4), TOK: Tokelau (3,4), TON: Tonga (3,4), SA: Samoa (3,4), VTP: Velickovic Total Polynesians (CI+TOK+TON+SA), PMD: Polynesian Meta Dataset (PFA+VTP), TTM: Tracey Total Maori (5,6), MMD: Maori Meta Dataset (TM+TTM), \S : total number of individuals for TTM and MMD at HLA-A* (176 and 290, respectively), -B* (171 and 285, respectively), - C*(124 and 238, respectively), -DQB1 (187 and 301, respectively) and -DRB1* (200 and 314, respectively) loci.

Table 2: List of the most common (frequencies ≥ 0.06) class I and II haplotypes with											
respective	value	of	linkage	disequilibrium	(D)	and	their	level	of	significance	(<i>p</i>).

HLA-A*-C*	Freq > 0.06	D	<i>n</i> -value
TP (N=36)		D	p value
02-01	0.15	0.06	0.01
11-04	0.07	0.05	0.01
34-15	0.07	0.0579	0
PFA (N=24)	0.07	0.0077	Ū.
02-01	0.19	0.13	0
11-04	0.08	0.05	0.01
34-15	0.08	0.05	0
PAH(N=12)	0.00	0.07	0
01-07	0.08	0.06	0.01
TM(N=114)	0.00	0.00	0.01
02-01	0.20	0.11	0
11-04	0.06	0.04	Ő
11-08	0.08	0.05	Ő
24-07	0.08	0.02	0.04
34-15	0.05	0.04	0
MFA (N=49)	0.05	0.01	0
02-01	0.33	0.16	0
11-04	0.09	0.06	Ő
11-08	0.10	0.00	0.01
24-07	0.10	0.05	0.01
MAH(N=65)	0.10	0.05	0.01
02-01	0.13	0.07	0
11-08	0.09	0.06	Ő
34-15	0.07	0.06	Ő
HLA-C*-B*	Freq > 0.06	D	<i>p</i> -value
TP (N=36)		2	p (allo
01-55	0.18	0.12	0
01-56	0.13	0.07	0
03-40	0.11	0.07	0
04-40	0.14	0.08	0
04-15	0.04	0.03	Ō
07-39	0.06	0.05	0
08-48	0.07	0.06	0
15-40	0.07	0.05	0
PFA (N=24)			
01-55	0.15	0.11	0
01-56	0.10	0.07	0
03-40	0.15	0.09	0
04-40	0.19	0.10	0
07-39	0.06	0.06	0
08-48	0.1	0.09	0
15-40	0.08	0.05	0.01
PAH (N=12)			
01-56	0.21	0.12	0
TM (N=114)			
01-55	0.25	0.18	0
01-56	0.15	0.02	0.01
04-40	0.12	0.10	0
07-39	0.09	0.08	0
08-48	0.12	0.10	0

Table 2 cont.

HLA-C*-B*	Freq ≥ 0.06	D	<i>p</i> -value
MEA(N-40)			F
$\frac{1}{1} \frac{1}{1} \frac{1}$	0.37	0.20	0
01-55	0.06	0.20	0.02
01-50	0.00	0.02	0.02
07-20	0.17	0.14	0
07-59	0.09	0.08	0
08-48	0.15	0.11	0
MAH (N=65)			
01-55	0.16	-0.03	0
04-40	0.08	0.06	0
05-44	0.06	0.06	0
07-07	0.06	0.05	0
07-39	0.10	0.06	0
08-48	0.10	0.07	0
15-40	0.07	0.06	0
HLA-A*-C*-B*	$Freq \ge 0.06$	D	<i>p</i> -value
TP (N=36)	•		•
02-01-55	0.08	-0.15	< 0.01
11-04-40	0.07	0.03	< 0.01
24-01-55	0.09	-0.31	< 0.01
24-01-56	0.12	-0.28	< 0.01
24-03-40	0.07	0.03	< 0.02
24-04-44	0.08	0.04	< 0.01
34-15-40	0.07	0.04	< 0.01
\mathbf{DEA} (N-24)			
02-01-55	0.12	0.05	<0.01
24 08 48	0.06	0.03	<0.01
24-08-46	0.00	0.02	<0.01
	0.00	0.04	<0.01
1 M (N=114)	0.10	0.00	0.01
02-01-55	0.19	0.09	<0.01
11-04-40	0.06	0.03	<0.01
11-08-48	0.07	0.03	<0.01
24-07-39	0.08	0.03	<0.01
MFA (N=49)			
02-01-55	0.29	0.03	< 0.05
11-04-40	0.09	0.04	< 0.01
24-01-56	0.06	0.02	< 0.02
24-07-39	0.09	0.04	< 0.01
24-08-48	0.08	0.02	< 0.01
MAH (N=65)			
02-01-55	0.11	0.05	< 0.01
11-08-48	0.08	0.04	< 0.01
24-07-39	0.07	0.03	< 0.01
34-15-40	0.07	0.05	< 0.01
HLA-DRB1*-DOB1*	Freq > 0.06	D	<i>n</i> -value
TP (N=36)	1 - 0000		r
04-03	0.18	0.06	0.02
08-06	0.08	0.06	0
09-03	0.15	0.07	0
11-03	0.07	0.03	0.04
12-03	0.14	0.06	0
14-05	0.13	0.11	0
15-06	0.10	0.08	0

Table 2 cont.

HLA-DRB1*-DQB1*	$Freq \ge 0.06$	D	<i>p</i> -value
PFA (N=24)			
04-04	0.06	0.04	0.02
08-06	0.08	0.06	0
09-03	0.19	0.08	0
12-03	0.13	0.05	0.02
14-05	0.15	0.12	0
15-06	0.08	0.07	0
PAH(N-12)			
04-03	0.17	0.08	0.04
08-06	0.08	0.00	0.02
12-03	0.17	0.08	0.04
14-05	0.08	0.08	0
15-06	0.13	0.10	0
TM(N-114)			•
1111(11-114) 04.02	0.13	0.04	0
08.06	0.13	0.04	0
00.03	0.07	0.00	0
11 03	0.00	0.05	0
12.03	0.13	0.00	0
12-03	0.21	0.09	0
13-00	0.03	0.03	0
14-05	0.09	0.08	0
MFA (N=49)			
04-04	0.04	0.03	0
08-05	0.06	0.04	0
08-06	0.12	0.91	0
09-03	0.10	1.00	0.02
11-03	0.18	0.06	0
12-03	0.28	0.09	0
MAH (N=65)			
04-03	0.14	0.08	0
11-03	0.13	0.11	0
14-05	0.12	0.13	0
HI A-DRB1*-DOA1*-DOB1*	Freq > 0.06	D	<i>n</i> -value
		D	p vulue
IP (N=30) 08 01 06	0.08	0.02	<0.01
11 05 02	0.08	0.03	< 0.01
12 05 03	0.07	0.02	< 0.02
12-05-05	0.14	0.03	< 0.01
14 01 05	0.13	0.15	< 0.01
15 01 06	0.15	0.03	<0.01
	0.10	0.04	<0.01
PFA (N=24)			0.04
08-01-06	0.08	0.03	<0.01
12-05-03	0.13	0.03	<0.01
13-01-06	0.19	0.18	<0.01
15-01-06	0.08	0.03	< 0.01
PAH (N=12)			
04-03-03	0.17	0.04	< 0.05
13-01-06	0.08	0.04	< 0.05

HLA-DRB1*-DQA1*-DQB1*	$Freq \ge 0.06$	D	<i>p</i> -value
TM (N=114)			
04-03-03	0.13	0.08	< 0.01
08-01-06	0.07	0.03	< 0.01
09-03-03	0.06	0.02	< 0.01
11-05-03	0.15	0.03	< 0.01
12-05-03	0.21	0.02	< 0.02
14-01-05	0.09	0.04	< 0.01
MFA (N=49)			
08-01-06	0.12	0.05	< 0.01
09-03-03	0.10	0.02	< 0.02
11-05-03	0.18	0.04	< 0.01
12-05-03	0.28	2.43	< 0.01
MAH (N=65)			
04-03-03	0.14	0.05	< 0.01
14-01-05	0.12	0.05	< 0.01

Table 2 cont.

N = total number of individuals, TP: Total Polynesians, PFA: Polynesians with Full Ancestry, PAH: Polynesians with Admixed History, TM= Total Maori, MFA = Maori with Full Ancestry, MAH: Maori with Admixed History.



Fig. 1: Distributions of HLA class I alleles in our study group and sub-groups in comparison with TTM from previous study (5) and European population (ELAN-ref 7); (a) HLA-A, (b) HLA-B and (c) HLA-C. Abbreviations: MFA = Maori with Full Ancestry, TM = Total Maori, TTM = Tracey Total Maori, MAH = Maori with Admixed History, ELAN = England Lancaster.



Fig. 2: Graphs constructed using (a) HLA-DQA1, (b) HLA-DQB1 and (c) HLA-DRB1 allele frequencies show European admixtures in our admixed sub-group (MAH) and previous study on Maori (TTM-ref 6), in comparison to our full ancestry Maori sub-group (MFA) and typical European (ENG-ref 8) population. Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TTM = Tracey Total Maori, ENG = England.



Fig. 3: PCO plots (axis 1 and 2) constructed using HLA-A, -B and -C allele frequencies. Reference populations obtained from (3,5,7,11,14,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians Meta Dataset (VTP + PFA), TTM = Tracey Total Maori, PNG = Papua New Guinea.



Fig. 4: Genetic relationships between Polynesians and other Asia Pacific populations (4,6,7,14,16,18,25) were probed using HLA-DRB1 and -DQB1 allele frequencies PCO plots (axis 1 and 2). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians Meta Dataset (VTP + PFA), TTM = Tracey Total Maori, PNG = Papua New Guinea, ENG = England.

Discussion

The data presentation and analysis in the present study were specifically designed to trace the origin of Polynesians and Maori based on multidisciplinary perspectives of Proto-Polynesian diffusion to Remote Oceania, including recent admixture with Europeans (1,38,39). For this reason, our new survey on HLA from full ancestry and admixed history sub-groups was carefully scrutinized and compared with existing data on Polynesians (3,4), Maori (5,6) and the European reference populations (7,8).

The most common HLA genes observed in Polynesians are *HLA-A*02*, -A*24, -B*40, -B*56, -C*01, -DQB1*03 and -DRB1*04 (see Table 1). The predominance of *HLA-A*02*, -A*24, -B*40, -B*55, -DQB1*03 and -DRB1*04 genes in Polynesia can be traced back to Asia as these allele groups combinations are common among Taiwan aborigines and other A-SM populations (7,9,10-13,18,20,22,24-27). The *HLA-A*24* is also commonly seen among P-SA, where it may have also been passed to Proto-Polynesians via this route (see 7,10,11). In contrast, high frequencies of *HLA-B*56* and -C*01 among Polynesians may reflect geneflow from P-SA, as these alleles are also common among Austronesian speaking groups in Melanesia and other Australoids, respectively (7,11).

The *HLA-B*55* and *-DQA1*03* alleles are not common in either putative ancestral population (i.e. P-SA and A-SM). A detailed search of the Allele Frequency Net Database (7) shows Taiwan aborigines and Chinese to have the highest frequency of the *HLA-B*55* allele in worldwide populations. Hence, the presence *HLA-B*55* in Polynesia may be taken as a faithful pointer to Taiwanese ancestors. The HLA-DQA1 allele frequency distributions among Polynesians are generally inconsistent, where both *HLA-DQA1*03* and *-DQA1*05*

have been reported to be the most common alleles in these people (7,18). Because *HLA-B* B*55 and -DQA1*03 allele groups are not the most frequent HLA-B and -DQA1 allele groups in either of A-SM and P-SA peoples, it might be that natural selection or other evolutionary mechanisms have elevated the frequency of this allele in Polynesia (as discovered by others for HLA loci – see 40).

Overall, there is agreement between the new HLA data for PFA and those previously obtained (3,4) – see Table 1. This view is supported by the exact tests of population differentiation (see Table S3). Although individual Polynesian (Cook Islands, Tokelau, Tonga and Samoa) may well be significantly different from each other, their combined genepool as VTP is not significantly different from PFA. The differences between individual Polynesian populations might be due to genetic refinement via founder effects (aka *The Genetic Bottleneck in Polynesia*) during multiple re-settlements and also due to selection during long periods of isolation. The latter will further investigated using high resolution HLA data. In particular, the close similarity of VTP to PFA rather than TP shows that subjects recruited for the previous studies probably have relatively little admixture. Hence, we combined VTP and PFA to gain PMD for wider phylogenetic and PCO comparison (later). In doing so, we consciously rejected the opportunity to increase the overall sample size by combining TP and VTP. We recognized that this sacrificed genuine Polynesian signal present in the TP group, but would increase noise due to inclusion of European genetic component in PAH.

Genetic fractions in Polynesians are arguably best demonstrated by their HLA haplotype distributions. The haplotype analysis increases statistical power by combining multiple loci as a single unit of inheritance. Our estimation of haplotype frequencies might be slightly

affected by the inability of the Arlequin software to account of the presence of very small number of blank in the present study. The most frequent haplotypes detected in our PFA and other Polynesians (see Tables 2 and S4) can all be traced back to Asia and Melanesia. For example, HLA-A*02-C*01, HLA-C*01-B*56, HLA-C*04-B*40 and HLA-A*02-C*01-B*55 associations are recorded from Taiwan, Philippines, Papua New Guinea Great Britain Rabaul, Cook Islands and Tokelau (refer 3,7,10,11,15,22) while HLA-DRB1*09-DQB1*03 is relatively common among Chinese, Japanese, Filipinos, Trobriand Islanders, PNG Roro (22,26,29,30). Thus, the most frequent haplotypes represented by PFA and those previously reported (3,4) can be detected, all along the diffusion route of A-SM to Remote Oceania, as inferred from multidisciplinary perspectives (1). The potential influence of P-SA ancestors on the Polynesian genepool is supported by the presence of HLA-DRB1*15-DQA1*01-DQB1*06 haplotype in the PFA (Table 2). This is commonly reported as the most frequent haplotype among P-SA populations (see 17,19), but does also occur at relatively high frequency among Japanese, Chinese and Koreans (7). Others (19) used high resolution HLA data to show differences between A-SM and P-SA. Their P-SA populations included Australian aborigines have high frequencies of the HLA-DRB1*15:02-DQA1*01:02-DQB1*06:01 haplotype while HLA-DRB1*15:01-DQA1*01:02-DQB1*06:02 is the more common type among their Asian populations: refer also to The Allele Frequency Net website (7) for wider population coverage. Interestingly, the P-SA haplotype is more frequent among Polynesians (19). Therefore, Polynesians seem to have inherited both haplotypes. The HLA data generated from our present study does show the need for high resolution analysis in future in order to better distinguish level of ancestral contributions to the Polynesian genepool, because the current data are insufficient to trace origins and migration patterns in fine detail.
Existing HLA class I (5) and class II (6) data on Maori (TTM) provide the possibility of creating a larger unified group of Maori by combining MFA and TTM. In general, HLA class I and II data for MFA and TTM do have general similarity in terms of the most frequent alleles observed. Nonetheless, there are number of alleles reported in TTM that are absent in MFA, but often detected in MAH and our reference European populations, ELAN (7) and ENG (8). We interpret this as showing European input in both, TTM and MAH. This view is further supported by the absence of nearly all of these alleles among the two ancestors, A-SM and P-SA, of modern Polynesian populations. This recent European admixture effect is demonstrated clearly in Figs. 1 and 2. Although the more common alleles (e.g. *HLA-A*02*, - *C*01* and -*DRB1*04*) in MAH and TTM show quite similar frequencies to MFA, the MAH and TTM frequency bars for other alleles are only shared with ELAN/ENG or are intermediate in frequencies between MFA and ELAN. This situation mirrors those observed in admixed Polynesians (PAH) at both, class I and II loci.

The admixture effects in MAH and TTM are also shown, e.g. by both noting the reduced frequencies of some haplotypes (e.g. *HLA-A*24-C*01* and *HLA-A*24-C*04*) which are quite common in MFA. This might due to the presence of alleles in MAH and TTM which are common to Europeans and lead to the presence of additional haplotypes, which are not otherwise common among Polynesians. For example, the *HLA-A*02-C*05-B*44*, *HLA-DRB1*07-DQB1*02* and *HLA-DRB1*13-DQB1*06* haplotypes observed in MAH and TTM are almost exclusively European types (7).

We conclude that our class I and II HLA data from MFA could be the best reference to use for genealogical comparisons while those belonging to TTM as reported by (5,6) and from our TM group should be treated as admixed groups and best represent the present-day Maori

Chapter 3.1

genepool. So, we take TTM plus TM together as MMD to best represent Maori presently living in New Zealand. In contrast, the previous Polynesian data (3,4) appears to be relatively free of admixture. Thus, the data we collected for PAH and MAH are entirely novel in our experience and show the effect of recent admixture following the arrival of Europeans settlers in New Zealand from 1840 onwards. In our view, recent admixture with Europeans has had significant impact on the genepool of Polynesians and Maori in New Zealand today. This is evident from both, the distributions of HLA alleles and haplotypes. This conclusion is potentially of particular importance for genetic studies with medical applications and is the subject of the future high resolution analysis of these samples via sequence based typing (to be reported later).

Next, we compared HLA data between Polynesians and Maori. The predominant HLA class I alleles (see Table 1) observed in Polynesians such as HLA-A*02, -A*24, -B40, -B55 and -C*01 are also the most common in MFA, except for HLA-B*56 which is slightly reduced in MFA. As discussed earlier, HLA-A*02, -B*40 and -B*55 link Polynesians to A-SM populations, -C*01 and -B*56 to P-SA and -A*24 is in both (3,7,9-15). There are only slight differences in distribution of HLA class II alleles between MFA and other Polynesians (see Table 1). Besides the general inconsistency of HLA-DQA1 loci among Polynesians as previously noted, MFA has HLA-DRB1*12 as the most common allele while HLA-DRB1*04 is predominant among other Polynesians. However, both alleles can be traced back to their A-SM ancestors. The HLA-DRB1*04 allele is the most common form among east coast Taiwanese Ami tribe (0.56) while two central mountain Taiwanese tribes (Taiwan Bunun: 0.27 and and Rukai: 0.33) have HLA-DRB1*12 as their most common form (see 9). Nonetheless, differences at HLA-DQA1 and -DRB1 between Maori and other Polynesians

88

will be the subject of our next analysis using high resolution data (Dunn and Edinur unpublished). Overall, the distribution of HLA class I and II alleles shows the greater representation of A-SM ancestry in the Maori genepool, where HLA-A*24 and -C*01 are the only alleles that can be traced back to P-SA populations.

The same can be seen regarding the distribution of HLA class I haplotypes where those that are typical of Polynesians, link them to A-SM populations and are particularly common in MFA (see Table 1 and Table S1). Nevertheless, there is marked difference for class II haplotypes (see Table 1) where Maori have higher frequencies of *HLA-DRB1*12-* and *HLA-DRB1*12-DQA1*05-* related haplotypes (e.g. *HLA-DRB1*12-DQB1*03* and *HLA-DRB1*12-DQA1*05-* related haplotypes (e.g. *HLA-DRB1*04* and *HLA-DRB1*04-DQA1*03-* related haplotypes are more common among Polynesians (e.g. *HLA-DRB1*04-DQB1*03* and *HLA-DRB1*04-DQB1*03* and *HLA-DRB1*04-DQA1*03-* related haplotype combinations are consistent with the differences between Maori and other Polynesians at HLA-DRB1 and -DQA1 discussed above. However, *HLA-DRB1*12-* and *HLA-DRB1*12-DQA1*05-* related haplotypes associations, which are common in MFA, can also be traced back to Melanesian and Asian populations and indicate genetic affinities between Asia-Pacific populations (4,7,14,25,29).

We used NJ phylogenetic tree (figure not shown) and PCO plots (Figs. 3-4 and Figs. S1-S2) to examine genetic affinities between Polynesian populations and their positions in a wider geographical context, particularly within the Asia-Pacific region. As noted in the results section, geographically related populations tend to cluster in both data mapping/procedures. The NJ analysis does not, however, resolve the wider pattern of relationships as effectively as PCO. We believe the latter to be a superior methodology in that it is empirical cf. theory-based. The NJ algorithm can only return a faithful representation of an evolutionary pattern

Chapter 3.1

if the underlying population level process that generates it, is itself tree-like. When this assumption is violated, e.g. as here due to full or partial reticulation arising from dual origin and gene flow, then the method may not resolve the tree reliably. The PCO method is based only on relative similarity and excludes them from this constraint. Our PCO plots constructed using information from axis 1 and 2 (Figs. 3 and 4) show more clear separation of populations than those constructed using information from axis 1 and 3 (Figs. S1 and S2) and this finding is associated with greater relative amount of genetic information captured by axis 2 compared with axis 3. The PCO plots in Figs. 3 and 4 show that Polynesians are distinct from other Austronesian-speaking groups (Taiwanese and Malays) and their position reflects degree of shared ancestry with the P-SA populations (Australian aboriginals and PNG). The distance between the two Austronesian groups (Taiwan aboriginals and Malays) reflect the diverse genepool of A-SM populations (41) and may be attributed to differences in allele frequency arising from founder effects during migration and/or natural selection associated with disease resistance. This is supported by X^2 goodness of fit analyses and exact tests of population differentiation where all the simulated of A-SM:P-SA populations are significantly different (p-value = 0) from Polynesians, regardless of mixing ratio. We interpret this to mean that the test is too powerful for its intended purpose, i.e. it simply shows that allele frequencies differ between the test groups rather than pointing to the best match.

Therefore, we used PCO plots to estimate genetic fractions of A-SM and P-SA in Polynesian and Maori genepools. Overall, we estimate approximate 60:40 genetic proportions of ASM:P-SA ancestral fractions in Maori and other Polynesians (see Figs. S4 to S8) compared with 70:30 and 85:15 proposed earlier by (31) and (32), respectively. There is a fairly wide range across all loci (55:45 to 90:10). Nonetheless, our estimation based on HLA-A, -B, -C and -DRB1 allele frequencies and those proposed above (31,32) do lead to just one conclusion, the A-SM/P-SA dual origin of Polynesians. Our corresponding estimates for Melanesians are 30:70 (range 24:76 to 50:50) compared with the previously reported value 45:55, but ours are based on far fewer data. Hence, this other value must be rated as more reliable as it is based on many SNP loci. However, the general correspondence (i.e. with respect to both rank order and magnitude) between our calculated values and theirs is encouraging and supports the idea of genetic differentiation between Melanesians and Polynesians due to much higher P-SA content in the genepol of the former.

Analysis of the two-digit resolution HLA class I and II data in this series point out the need for high resolution data to better resolve outstanding ambiguities and also to better determine A-SM/P-SA ancestral fractions. Therefore, this will be now part of the subject matter in our next study using high resolution HLA class I and II data, besides the primary focus to look deeper at evolutionary processes such as bottlenecks, natural selection and, more significantly, the implications for health. Finally, placements of admixed sub-groups (e.g. MAH, TM and PAH) toward European population (ELAN and ENG) in Figs. 3 and 4 are evidence for significant contemporary geneflow from Europeans. In our view, full understanding of these admixed sub-groups should be a primary focus in any new developments in domestic health policies as these emerging lineages corresponds to present day Polynesian and Maori living in New Zealand.

Acknowledgements

We would like to thank all the participants involved in this study. This research was supported by Victoria University of Wellington and New Zealand Blood Service, Auckland.

References

- Bellwood P, Chambers GK, Ross, M. Hung C-H. Are "cultures" inherited? Multidisciplinary perspectives on the origins of Austronesian–speaking peoples prior to 1000 BC. In: Roberts BW, Linden MV, eds. *Investigating archaeological cultures: material culture, variability and transmission. Dordrecht: Springer, 2011, 321-54.*
- Chambers GK. Polynesian genetics and Austronesian prehistory. In: Simanjuntak T, Pojoh I, Hisyam M, eds. Austronesian Diaspora and the Ethnogeneses of People in Indonesian Archipelago. Jakarta: Indonesian Institute of Sciences (LIPI), 2006, 299-319.
- Velickovic ZM. PhD thesis: HLA polymorphism in Pacific Islands populations. Dunedin, New Zealand: University of Otago, 2001.
- Velickovic ZM, Delahunt B, Carter JM. HLA-DRB1 and DQB1 polymorphisms in Pacific Islands populations. *Tissue Antigens* 2002: 59: 397-406.
- Tracey MC. *PhD thesis: HLA allelic variation in New Zealand Maori*. Dunedin, New Zealand: University of Otago, 2007.
- Tracey MC, Carter JM. Class II HLA allele polymorphism: DRB1, DQB1 and DPB1 alleles and haplotypes in the New Zealand Maori population. *Tissue Antigens* 2006: 68: 297-302.
- Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucl Acids Res* 2011: **39**: D913-19.
- 8. Thomson W, Barrett JH, Donn R et al. Juvenile idiopathic arthritis classified by the ILAR criteria: HLA associations in UK patients. *Rheumatology* 2002: **41**: 1183-89.

- 9. Chu CC, Lin M, Nakajima F et al. Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. *Tissue Antigens* 2001: **58**: 9-18.
- Bugawan TL, Mack SJ, Stoneking M, Saha M, Beck HP, Erlich HA. HLA class I allele distribution in six Pacific/Asian populations: evidence of selection at the HLA-A locus. *Tissue Antigen* 1999: 53: 311-19.
- Main P, Attenborough RD, Chelvanayagam G, Gao X. The peopling of New Guinea: Evidence from class I human leukocyte antigen. *Hum Biol* 2001: 73: 365-83.
- 12. Bhatia K, Prasad ML, Barnish G, Koki G. Antigen and haplotype frequencies at three human leukocyte antigen (HLA-A, -B and -C) in the Pawaia of Papua New Guinea. *Am J Phys Anthropol* 1988: **75**: 329-40.
- 13. Serjeantson SW, Ryan DP, Thompson AR. The Colonization of the Pacific: The story according to human leukocyte antigens. *Am J Hum Genet* 1982: **34**: 904-18.
- Edinur HA, Zafarina Z, Spinola H, Nur Haslindawaty AR, Panneerchelvam S, Norazmi MN. HLA polymorphism in six Malay sub-ethnic groups in Malaysia. *Hum Immunol* 2009: **70**: 518-26.
- Lin M, Chu CC, Lee HL et al. Heterogeneity of Taiwan's indigenous population:
 Possible relation to prehistoric Mongoloid dispersals. *Tissue Antigens* 2000: 55: 1-9.
- 16. Gao X, Bhatia K, Trent RJ, Serjeantson SW. HLA-DR, DQ nucleotide sequence polymorphisms in five Melanesian populations. *Tissue Antigens* 1992a: **40**: 31-7.
- 17. Gao X, Veale A, Serjeantson SW. HLA class II diversity in Australian aborigines: unusual HLA-DRB1 alleles. *Immunogenetics* 1992b: **36**: 333-37.
- Gao X, Zimmet P, Serjeantson SW. HLA-DR, DQ sequence polymorphisms in Polynesians, Micronesians, and Javanese. *Hum Immunol* 1992c: 34: 153-61.
- 19. Gao X, Serjeantson SW. Heterogeneity in HLA-DR2-related DR, DR haplotypes in eight populations of Asia-Oceania. *Immunogenetics* 1991: **34**: 401-8.

- 20. Munkhbat B, Sato T, Hagihara M et al. Molecular analysis of HLA polymorphism in Khoton-Mongolians. *Tissue Antigens* 1997: **50**: 124-34.
- 21. Chandanayingyong D, Stephens HAF, Klaythong R et al. HLA-A, -B, -DRB1, DQA1, and DQB1 polymorphism in Thais. *Hum Immunol* 1997: **53**: 174–82.
- 22. Shaw C.-K, Chen L.-L, Lee A, Lee TD. Distribution of HLA gene and haplotype frequencies in Taiwan: a comparative study among Min-nan, Hakka, Aborigines and Mainland Chinese. *Tissue Antigens* 1999: **53**: 51-64.
- 23. Maitland K, Bunce M, Harding RM et al. HLA class-I and class-II allele frequencies and two-locus haplotypes in Melanesians of Vanuatu and New Caledonia. *Tissue Antigens* 2004: **64**: 678-86.
- Ohashi J, Naka I, Kimura R et al. HLA-DRB1 polymorphism on Ha'ano island of the Kingdom of Tonga. *Anthropol Sci* 2006: **114**: 193–98.
- 25. Mack SJ, Bugawan TL, Moonsamy PV et al. Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. *Tissue Antigens* 2000: 55: 383–400.
- 26. Saito S, Ota S, Yamada E, Inoko H, Ota M. Allele frequencies and haplotypic associations defined by allelic DNA typing at HLA class I and class II loci in the Japanese population. *Tissue Antigens* 2000: **56**: 522-29.
- 27. Yang JH, Sohn Y-H, Ko S-Y, Choi S-E, Kim MH, Oh H-B. Anthropological analysis of Koreans using HLA class II diversity among East Asians. *Tissue Antigens* 2010: 76: 282-88.
- 28. Yoshida M, Ohtsuka R, Nakazawa M, Juji T, Tokunaga K. HLA-DRB1 frequencies of non-Austronesian-speaking Gidra in South New Guinea and their genetic affinities with Oceanian populations. *AJPA* 1995: **96**: 177-81.

- 29. Zimdahl H, Schiefenhovel W, Kayser M, Roewer L, Nagy M. Towards understanding the origin and dispersal of Austronesians in the Solomon Sea: HLA class II polymorphism in eight distinct populations of Asia-Oceania. *Eur J Immunogenet* 1999: **26**: 405-16.
- 30. Hagelberg E, Kayser M, Nagy M et al. Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. *Philos Trans R Soc Lond B Biol Sci* 1999: **354**: 141-52.
- 31. Kimura R, Ohashi J, Matsumura Y et al. Gene flow and natural selection in oceanic human populations inferred from genome-wide SNP typing. *Mol Biol Evol* 2008: 25: 1750-61.
- 32. Wollstein A, Lao O, Becker C et al. Demographic history of Oceania inferred from genome-wide data. *Curr Biol* 2010: **20**: 1983-92.
- 33. Excoffier L, Laval G, Schneider S. Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 2005: **1**: 47-50.
- 34. Weir BS. Genetic data analysis II. Sunderland, Massachusetts: Sinauer Associates, 1996.
- 35. Nielson DM, Ehm MG, Zaykin DV, Weir BS. Effect of two- and three-locus linkage disequilibrium on the power to detect marker/phenotype associations. *Genetics* 2004: 168: 1029-40.
- 36. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Seattle: Department of Genome Sciences, University of Washington, 2004.
- Page RDM. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 1996: 12: 357–8.
- 38. Underhill PA, Passarino G, Lin AA et al. Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mutat* 2001: **17**: 271-80.

- 39. Lea RA, Chambers GK. *Pharmacogenomics*. In: Suarez-Kurtz G. ed. *Pharmacogenetics in admixed populations in Polynesia. Austin Texas: Landes Bioscience*, 2007, 164-89.
- 40. Solberg OD, Mack SJ, Lancaster AK et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. *Hum Immunol* 2008: **69**: 443-64.
- 41. Sanchez-Mazas A, Poloni ES, Jacques G, Sagart L. HLA genetic diversity and linguistic variation in East Asia. In: Sagart L, Blench R, Sanchez-Mazas, ed. *The people of East Asia: Putting together archaeology, linguistic and genetics. London and New York: Routledge Curson, 2005, 273-96.*

Supplementary information

Table S1: Exact test of Hardy-Weinberg equilibrium (*p*-values) for the studied population groups and sub-groups.

	HLA-A*	HLA-C*	HLA-B*	HLA-DRB1*	HLA-DQA1*	HLA-DQB1*
TP	0.65	0.43	0.38	0.56	0.75	0.30
PFA	0.32	0.22	0.14	0.46	0.14	0.79
PAH	0.92	0.69	0.23	1	0.56	0.79
ТМ	0.28	0.68	0.62	0.11	0.47	0.37
MFA	0.69	0.77	0.83	0.47	0.24	0.99
MAH	0.16	0.51	0.77	0.43	0.78	0.32

Deviation from HWE is considered significant at p-value < 0.05. Abbreviations: TP: Total Polynesians, PFA: Polynesians with Full Ancestry, PAH: Polynesians with Admixed History, TM= Total Maori, MFA = Maori with Full Ancestry, MAH: Maori with Admixed History.

	<u>HLA-</u> A*-C*	<u>HLA-</u> A*-B*	<u>HLA-C*-</u> B*	<u>DRB1*-</u> DOA1*	HLA-DRB1- DOB1*	HLA-DQA1- DOB1*
TP	0	0	0	0	0	0
PFA	0	0.04	0	0	0	0
PAH	0.47	0.23	0.01	0	0	0
TM	0	0	0	0	0	0
MFA	0	0	0	0	0	0
MAH	0	0	0	0	0	0

Table S2: Likelihood-ratio test of linkage disequilibrium (*p*-values) between pairs of HLA loci in the studied population groups and sub-groups.

Linkage between pairs of loci is considered significant at *p*-value < 0.05. Abbreviations: TP: Total Polynesians, PFA: Polynesians with Full Ancestry, PAH: Polynesians with Admixed History, TM= Total Maori, MFA = Maori with Full Ancestry, MAH: Maori with Admixed History.

					Exact	tests cla	ass I (p	-values)				
	TP	PFA	PAH	CI	TOK	TON	SA	VTP	PMD	TM	MFA	MAH	TTM
ТР	~												
PFA	1	~											
PAH	0.52	0.02	~										
CI	0.25	0.35	0.02	~									
TOK	0.20	0.55	0	0	~								
TON	0.02	0.27	0	0	0								
S A	0.02	0.27	0	0	0	0							
SA VTD	0.13	0.44	0	0	0	0.46	0.04						
	0.01	0.45	0	0.01	0	0.40	0.04	~ 1					
	0.12	0.91		0.01	0	0.42	0.21	1	~				
IM	0.03	0.05	0.08	0	0	0	0	0	0	~			
MFA	0.01	0	0	0.04	0	0	0	0	0	0.48	~		
MAH	0	0	0.04	0	0	0	0	0	0	0.93	0	~	
TTM	0	0	0.01	0	0	0	0	0	0	0	0	0	~
					Exact	tests cla	iss II (p	-values	<u>s)</u>				
	TP	PFA	PAH	CI	TOK	TON	SA	VTP	PMD	TM	MFA	MAH	TTM
TP	~												
PFA	1	~											
PAH	0.96	0.49	~										
CI	0.35	0.62	0.09	~									
TOK	0	0.01	0	0	~								
TON	0.34	0.73	0.02	0.08	0	~							
SA	0	0.03	0	0	Ő	0	~						
VTP	012	0.91	Ő	0 36	õ	0 36	0	~					
PMD	0.12	0.96	Ő	0.30	Õ	0.38	Õ	1	~				
TM	0.03	0.90	0 59	0.04	0	0.50	0	0	0	~			
MEA	0.05	0.04	0.01	0.04	0	0	0	0	0	0.13			
МЛН	0.01	0.01	0.01	0.05	0	0	0	0	0	0.15	0		
TTM	0.01	0.01	0.92	0	0	0	0	0	0	0.09	0	0.01	~
				E	vact test	ts class	I and I	[(<i>n</i> -val	ues)				
	TD		DATI			TON				TM		MATT	
ТD	IP	РГА	РАП	CI	IUK	IUN	SA	VIP	PMD	1 1/1	MГА	МАП	1 1 1/1
	~												
PFA	1	~											
PAH	0.85	0.04	~										
CI	0.33	0.48	0.01	~									
TOK	0	0	0	0	~								
TON	0.05	0.55	0	0	0	~							
SA	0	0.12	0	0	0	0	~						
VTP	0	0.75	0	0.01	0	0.50	0	~					
PMD	0.09	0.99	0	0.01	0	0.42	0	1	~				
TM	0.01	0.01	0.14	0	0	0	0	0	0	~			
MFA	0	0	0	0.01	0	0	0	0	0	0.28	~		
MAH	0	0	0.17	0	0	0	0	0	0	0.97	0	~	
TTM	0	0	0.01	0	0	0	0	0	0	0	0	0	~

Table S3: Exact test of population differentiation (*p*-values) between pairs of Polynesian and Maori datasets

The two populations are considered significantly different if having *p*-value < 0.05. Abbreviations: TP: Total Polynesians, PFA: Polynesians with Full Ancestry, PAH: Polynesians with Admixed History, CI: Cook Islands (3,4), TOK: Tokelau (3,4), TON: Tonga (3,4), SA: Samoa (3,4), VTP: Velickovic Total Polynesians (CI+TOK+TON+SA) and PMD: Polynesian Meta Dataset (PFA+VTP), TM= Total Maori, MFA = Maori with Full Ancestry, MAH: Maori with Admixed History, TTM: Tracey Total Maori (5,6).

$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	0.11 0.04
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$.11 .04
$\begin{array}{cccccccccccccccccccccccccccccccccccc$.04
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	·.0+
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	02
	105
	04
	116
	05
	113
	06
	04
	02
34-01 0.03 0.02 0.04 0 0 0 0.05 0 0 0 0 0	0
34-15 0.07 0.08 0.04 0.05 0.02 0.07 0.06 0.05 0.04 0.04 0.05 0.0	0.05
HLA-C*-B*	
	14
01-55 0.18 0.15 0.19 0.25 0.57 0.10 0.22 0.07 0.18 0.10 0.14 0.1	14
01-50 0.15 0.10 0.21 0.15 0.00 0.02 0.15 0.20 0.07 0.12 0.15 0.1	.14
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.001 0.0	1.08
04-13 0.04 0.02 0 0 0 0.01 0 0.04 0.04 0.02 0.0	102
04-40 0.14 0.19 0.04 0.12 0.17 0.08 0.09 0.14 0.43 0.10 0.19 0.1	0
	0
0.05 - 44 0.05 0.02 0 0.05 0.01 0.06 0 0 0 0 0 0 0 0 0 0	0
	.05
07-59 0.00 0.00 0.04 0.09 0.09 0.10 0.10 0.05 0.04 0.04 0.05 0.0	.03
07-30 0.01 0.02 0 0.02 0.01 0.02 0.04 0.04 0 0 0.02 0.02 0.01 0.02 0.04 0.04 0 0 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.04 0.04 0 0 0.02	02
	.00
12-15 0 0 0 0 0 0 0 0 0 0 0 0.05 0 0.01 0.0 15-40 0.07 0.08 0 0.05 0.02 0.07 0.07 0.07 0.04 0.07 0.06 0.0	01
	.00
HLA-A*-C*-B*	0
01-07-08 0.01 0 0.04 0.03 0.02 0.04 0 0 0 0 0 0	0
02-01-55 0.08 0.12 0.08 0.19 0.29 0.11 0.13 0.04 0.09 0.05 0.08 0.0	.08
02-03-40 0 0 0.04 0 0 0.01 0 0 0.15 0.04 0.0	.04
02-05-44 0 0 0 0.03 0 0.05 0 0 0 0 0 0	0
	0
	0
	.01
	0
24-01-55 0.09 0 0.17 0.05 0.07 0.04 0.05 0 0.08 0 0.04 0.0	1.04
24-01-56 0.12 0.06 0.13 0.03 0.06 0.02 0.11 0.19 0 0.08 0.10 0.1	1.10
	.01
	0
24-04-40 0.04 0.08 0 0.04 0.06 0.02 0.05 0.04 0.27 0.09 0.11 0.1	.11
	0
	0
	0.04
	0
34-15-40 0.07 0.08 0.04 0.05 0.02 0.07 0.06 0.05 0 0.04 0.04 0.04	0.04

Table S4: List of HLA class I and II haplotypes in Maori and Polynesian datasets.

Table S4 cont.

Datasets	TP	PFA	PAH	ТМ	MFA	MAH	CI	TON	TOK	SA	VTP	PMD
(N)	(36)	(24)	(12)	(114)	(49)	(65)	(50)	(50)	(50)	(50)	(200)	(224)
HLA-DRB1	*-DQB	1*										
01-05	0.03	0.02	0.04	0.03	0.01	0.04	-	-	-	-	-	-
03-02	0.01	0	0.04	0.03	0.02	0.05	-	-	-	-	-	-
04-02	0	0	0	0	0	0	-	-	-	-	-	-
04-03	0.18	0.19	0.17	0.13	0.12	0.14	0.24	0.17	0.36	0.15	0.23	0.21
04-04	0.04	0.06	0	0.02	0.04	0	0	0.03	0.06	0.08	0.04	0.05
07-02	0	0	0	0.02	0	0.04	-	-	-	-	-	-
07-03	0	0	0	0.01	0	0.02	-	-	-	-	-	-
08-05	0	0	0	0.02	0.06	0	0.03	0	0	0	0.01	0
08-06	0.08	0.08	0.08	0.07	0.12	0.03	0.03	0.11	0.07	0.07	0.07	0.08
09-03	0.15	0.19	0.08	0.06	0.10	0.03	0.09	0.15	0.04	0.25	0.13	0.16
11-03	0.07	0.06	0.08	0.15	0.18	0.13	0.11	0.16	0	0.07	0.09	0.07
12-02	0	0	0	0	0	0	-	-	-	-	-	-
12-03	0.14	0.13	0.17	0.21	0.28	0.01	0.16	0.06	0.22	0.08	0.13	0.13
13-06	0.03	0.01	0.04	0.03	0	0.05	-	-	-	-	-	-
14-05	0.13	0.15	0.08	0.09	0.05	0.12	0.13	0.11	0.07	0.08	0.10	0.12
15-05	0	0	0	0	0	0	0	0.03	0	0	0.01	0
15-06	0.10	0.08	0.13	0.03	0.01	0.05	0.03	0.07	0.06	0.08	0.06	0.07
HLA-DRB1	*-DQA	1*-DQ	B1*									
01-05-01	0.03	0.02	0.04	0.03	0.01	0.05	-	-	-	-	-	-
03-02-05	0.01	0	0.04	0.03	0.02	0.05	-	-	-	-	-	-
04-03-03	0.18	0.19	0.17	0.13	0.12	0.14	-	-	-	-	-	-
07-02-02	0	0	0	0.03	0	0.05	-	-	-	-	-	-
08-05-01	0	0	0	0.03	0.06	0	-	-	-	-	-	-
08-06-01	0.08	0.08	0.08	0.07	0.12	0.04	-	-	-	-	-	-
09-03-03	0.15	0.19	0.03	0.06	0.10	0.03	-	-	-	-	-	-
11-03-05	0.07	0.06	0.08	0.15	0.18	0.13	-	-	-	-	-	-
12-05-03	0.14	0.13	0.17	0.21	0.28	0.16	-	-	-	-	-	-
13-06-01	0.15	0.19	0.08	0.03	0	0.05	-	-	-	-	-	-
14-01-05	0.13	0.15	0.08	0.09	0.05	0.12	-	-	-	-	-	-
15-01-06	0.10	0.08	0.13	0.03	0.01	0.05	-	-	-	-	-	-

Abbreviations: TP: Total Polynesians, PFA: Polynesians with Full Ancestry, PAH: Polynesians with Admixed History, TM= Total Maori, MFA = Maori with Full Ancestry, MAH: Maori with Admixed History, CI: Cook Islands (3,4), TOK: Tokelau (3,4), TON: Tonga(3,4), SA: Samoa (3,4), VTP: Velickovic Total Polynesians (CI+TOK+TON+SA) and PMD: Polynesian Meta Dataset (PFA+VTP)

Table S5: Exact tests of population differentiation of HLA class I data from Polynesian, Maori and combination genepool of Taiwan Ami (A) and Papua New Guinea Goroko (G).

							Exact te	ests class I	(p-values)						
	PFA	MFA	CI	TOK	TON	SA	VTP	PMD	А	G	50A:50G	60A:40G	70A:30G	80A:20G	90A:10G
PFA	~														
MFA	0	~													
CI	0.29	0.03	~												
TOK	0	0	0	~											
TON	0.29	0	0	0	~										
SA	0.41	0	0	0	0	~									
VTP	0.43	0	0	0	0.46	0.06	~								
PMD	0.92	0	0	0	0.47	0.19	1	~							
А	0	0	0	0	0	0	0	0	~						
G	0	0	0	0	0	0	0	0	0	~					
50A:50G	0	0	0	0	0	0	0	0	0	0	~				
60A:40G	0	0	0	0	0	0	0	0	0	0	1	~			
70A:30G	0	0	0	0	0	0	0	0	0.06	0	0.69	1	~		
80A:20G	0	0	0	0	0	0	0	0	0.74	0	0.01	0.66	1	~	
90A:10G	0	0	0	0	0	0	0	0	1	0	0	0.01	0.59	1	~

The two populations are considered significantly different if having *p*-value < 0.05. Abbreviations: PFA: Polynesians with Full Ancestry, MFA = Maori with Full Ancestry, CI: Cook Islands (3), TOK: Tokelau (3), TON: Tonga (3), SA: Samoa (3), VTP: Velickovic Total Polynesians (CI+TOK+TON+SA) and PMD: Polynesian Meta Dataset (PFA+VTP), A: Taiwan Ami (7,15) and G: Papua New Guinea Goroko (11).



Fig. S1: PCO plots (axis 1 and 3) constructed using HLA-A, -B and -C allele frequencies. Reference populations obtained from (3,5,7,11,14,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians meta dataset (VTP + PFA), TTM = Tracey Total Maori, PNG = Papua New Guinea.



Fig. S2: PCO plots in Fig. 4 were re-construct using genetic percentage represent by axis 1 and 3. References populations obtained from (4,6,7,14,16,18,25). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians meta dataset (VTP + PFA), TTM = Tracey Total Maori, PNG = Papua New Guinea, ENG = England.



Fig. S3: Here, we demonstrate that there are no marked differences between PCO plots constructed using HLA-A, -B and -C allele frequencies (axis 1 and 2) for just the original populations and from those constructed together with combined populations (Fig. 3). Reference populations obtained from (3,5,7,11,14,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, TTM = Tracey Total Mori, PNG = Papua New Guinea.



Fig. S4: HLA-A allele frequencies PCO plots used to estimate genetic fractions of Maori and Polynesian populations based on simulated proportions of A-SM (Taiwan Ami, Atayal, Paiwan, Rukai, Puyuma and Tsou) and P-SA (Papua New Guinea Goroko and West Schrader Ranges) genepools. References populations obtained from (3,5,7,11,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians Meta Dataset (VTP + PFA), TTM = Tracey Total Maori, A-SM = Austronesian-speaking Mongoloids, P-SA = Papuan-speaking Australoids.



Fig. S5: HLA-B allele frequencies PCO plots used to estimate genetic fractions in Maori and Polynesian populations based on simulated proportions of A-SM (Taiwan Ami, Atayal, Paiwan, Rukai, Puyuma and Tsou) and P-SA (Papua New Guinea Goroko and West Schrader Ranges) genepools. Reference populations obtained from (3,5,7,11,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians Meta Dataset (VTP + PFA), TTM = Tracey Total Maori, A-SM = Austronesian-speaking Mongoloids, P-SA = Papuan-speaking Australoids.



Fig. S6: HLA-C allele frequencies PCO plots used to estimate genetic fractions in Maori and Polynesian populations based on simulated proportions of A-SM (Taiwan Ami, Atayal, Paiwan, Rukai, Puyuma and Tsou) and P-SA (Papua New Guinea Goroko and West Schrader Ranges) genepools. Reference populations obtained from (3,5,7,11,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians Meta Dataset (VTP + PFA), TTM = Tracey Total Maori, A-SM = Austronesian-speaking Mongoloids, P-SA = Papuan-speaking Australoids.



Fig. S7: HLA-DRB1 allele frequencies PCO plots used to estimate genetic fractions in Maori and Polynesian populations based on simulated proportions of A-SM (Taiwan Ami, Atayal, Paiwan, Rukai, Puyuma and Tsou) and P-SA (Australian Aborigine Yuendemu, Kimberly, Cape York Peninsula and Papua New Guinea Goroko) genepools. References population obtained from (4,6,7,11,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians Meta Dataset (VTP + PFA), TTM = Tracey Total Maori, A-SM = Austronesian-speaking Mongoloids, P-SA = Papuan-speaking Australoids.



Fig. S8: Combination of HLA-A, -B, -C and –DRB1 allele frequencies PCO plots used to estimate genetic fractions in Maori and Polynesian populations based on simulated proportions of A-SM (Taiwan Ami, Atayal, Paiwan, Rukai, Puyuma and Tsou) and P-SA (Australian Aborigine Yuendemu, Kimberly, Cape York Peninsula and Papua New Guinea Goroko) genepools. Reference populations obtained from (3-7,11,15). Abbreviations: TM = Total Maori, MFA = Maori with Full Ancestry, MAH = Maori with Admixed History, TP = Total Polynesians, PFA = Polynesians with Full Ancestry, PAH = Polynesians with Admixed History, VTP = Velickovic Total Polynesians (Cook Islands + Tokelau + Tonga + Samoa), PMD = Polynesians Meta Dataset (VTP + PFA), TTM = Tracey Total Maori, A-SM = Austronesian-speaking Mongoloids, P-SA = Papuan-speaking Australoids.

HLA and MICA polymorphism in Polynesians and New Zealand Maori: Implications for ancestry and health

H.A. Edinur^a, P.P.J. Dunn^b, L. Hammond^b, C. Selwyn^b, P. Brescia^c, M. Askar^d, P. Reville^d, Z.M Velickovic^{e,f}, R.A. Lea^g and G.K. Chambers^a¶

^aSchool of Biological Sciences, Victoria University of Wellington, New Zealand

^bTissue Typing Laboratory, New Zealand Blood Service, Auckland

^cOne Lambda, Canooga Park, Los Angeles, CA, USA

^dAllogen Laboratories, Transplant Center, Cleveland Clinic, Ohio

^eDepartment of Cell & Molecular Therapies, Royal Prince Alfred Hospital, Sydney, Australia ^fSydney Medical School, University of Sydney, Australia. ^gSchool of Medical Sciences, Griffith University, Australia

Status: Human Immunology, 74, 1119-1129.

Key words: Polynesia, Maori, Austronesian Diaspora, Human Leukocyte Antigen, MICA **¶Corresponding author**:

Dr Geoffrey K. Chambers, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand. Ph: +64-(0)4-463-6091, Fax: +64-(0)4-463-5331, e-mail: Geoff.Chambers@vuw.ac.nz

ABSTRACT

Data from HLA typing studies have made significant contributions to genetic theories about the Austronesian diaspora and the health of descendant populations. To help further unravel pattern and process elements, we have typed HLA and MICA loci at high resolution in DNA samples from well defined groups of Maori and Polynesian individuals. Our results show a restricted set of HLA class I alleles compared with other well characterised populations. In contrast, the class II HLA-DRB1 locus seems to be diverse in Maori and Polynesians and both groups show high frequencies of HLA-DRB1*04:03, -DRB1*08:03, -DRB1*09:01 and -DRB1*12:01. Our survey also provides the first ever MICA datasets for Polynesians and reveal unusual distributions and associations with the HLA-B locus. Overall, our data provide further support for a hybrid origin for Maori and Polynesians. One novel feature of our study is the finding that the gene sequence of the HLA-B*40:10 allele in Polynesians is a recombinant of HLA-B*55:02 and -B*40:01. HLA-B*40:10 is in close association with HLA-C*04:03, an allele identified as a hybrid of HLA-C*04 and -C*02. In this respect, our data resemble those reports on Amerindian tribes where inter-allele recombination has been a common means of generating diversity. However, we emphasize that Amerindian gene content per se is only a very minor element of the overall Polynesian genepool. The wider significance of HLA and MICA allele frequencies across the Pacific for modern day health is also discussed in terms of the frequency relative to reference populations of disease known to be associated with specific HLA and MICA markers. Thus, Polynesians and Maori are largely unaffected by "European autoimmune diseases" such as ankylosing spondylitis, uveitis and coeliac disease, yet there are several Maori- and Polynesian-specific autoimmune diseases where the HLA and MICA associations are still to be determined.

Chapter 3.2

1. Introduction

The structurally homologous classical HLA (human leukocyte antigen) and MICA (major histocompatibility complex class I chain-related gene A) molecules are encoded by the closely-linked genes located on the short arm of human chromosome 6. The HLA and MICA genes are highly polymorphic, reflecting their roles in the adaptive immune system [1-3]. MICA, as HLA, is clinically relevant in organ transplantation [4] and may have a role in haematopoietic stem cell transplantation as it is a ligand for NKG2D which activates NK cells and T cells [5]. Pacific population data from HLA and MICA may also be viewed in the light of an uncontested archaeological chronology [6]. The present survey is the first study of MICA in Polynesian and Maori subjects, while population data generated from HLA studies have been of longstanding value in learning about the origins of Polynesians [7-13]. In general, they show close genetic similarities between Polynesian and Asian populations. Other work illustrates moderate gene flow between Austronesian-speaking Mongoloids (A-SM) and Papuan-speaking Australoids (P-SA) in Melanesia before voyaging began to Remote Oceania; see [6,13] for further details. The emerging picture is one, where both Polynesian and Melanesian people have a common hybrid origin, but that the former have retained relatively greater A-SM content in their genepool. These features mirror conclusions based on SNP data [14,15]. Such findings are entirely consistent with earlier studies from tissue typing laboratories; blood groups and globin gene rearrangements [16-23]. Autosomal markers are thus in accord and show hybrid origin of Polynesians with strongest link to Asia.

Data from maternal and paternal specific markers, mtDNA and Y-chromosome, reveal details of the hybridisation phase – specifically that it was gender-biased with relatively high proportions of P-SA males and A-SM females involved in the diffusion process via Near Oceania and out to wider Polynesia. Hence, mtDNA maps Polynesian origins back to Taiwan [24] while Y chromosome data show a predominantly 'Melanesian' influence in the Polynesian paternal genepool [25,26]. These markers also show variation from West to East as voyaging progressed via repeated founder effects and were repeatedly subject to episodes of natural selection. The same forces may well have shaped HLA and MICA allele frequencies and nature of variation at these loci may indicate the effect of particular types of selection pressure. Cell surface antigens such as HLA and MICA are important in disease resistance and have been associated with a wide range of autoimmune disorders. In this context, variations of HLA and MICA alleles in particular populations may also have consequences for contemporary health, as many studies have shown associations of HLA and MICA alleles with diseases; e.g. multiple sclerosis, rheumatoid arthritis, ankylosing spondylitis, Behcet's disease, type 1 diabetes etc. [2,27,28]. These loci are, therefore, exposed to natural selection and their patterns of allelic diversity can reveal features of significant events in population history. The HLA and MICA data generated from this survey and previous studies will now be used to provide further information about socio-cultural factors prevailing during initial occupancy of Near Oceania by Polynesian ancestors and extended to include the effects of recent admixture with Europeans and their contemporary medical significance.

2. Materials and Methods

2.1. Samples

Human genomic DNA samples from unrelated individuals were obtained from the Victoria University of Wellington DNA Bank and consist of Total Polynesian (TP) and Total Maori (TM) groups. The volunteers have varying degrees of admixture, particularly with Europeans. We have divided them into Full Ancestry (i.e. Polynesians with Full Ancestry: PFA and Maori with Full Ancestry; MFA) sub-groups for those with four grandparents and no family ancestral knowledge of members from other ethnic groups. All genetic materials were obtained with informed consent and ethical approvals were obtained from the New Zealand Central Region Ethic Committee and Victoria University Human Ethic Committee.

2.2. HLA and MICA genotyping

Genetic variation at HLA-A, -B and -C loci were determined by high resolution DNA sequence based typing using a combination of locus- and allele-specific primers [29]. HLA-DRB1 high resolution typing was also by sequence based typing (SBT) [29] or by using Luminex High Definition DRB1 beads ('HD DR' One Lambda, Canooga Park, CA, USA). DNA sequences were compiled and assigned using SBTengine software (Genome Diagnostics, Utrecht Holland). The HLA-DRB3, DRB4, DRB5, DQA1, DPA1, -DPB1 and MICA loci were genotyped using rSSOP Luminex beads (One Lambda, Canoga Park, CA, USA) at the Cleveland Clinic, Ohio. To test for a MICA deletion [30], a multiplex PCR approach was utilized as previously described [31]. Briefly, DNA samples were sent to the Cleveland Clinic and interrogated for a possible MICA deletion using multiplex PCR carried out on a thermo cycler (GeneAmp PCR System 9700, Applied Biosystems, Foster City, CA, USA) in a 20 µl mixture containing 0.1 µg of genomic DNA, three primer pairs (0.5 µM each), 0.2 mM dNTP, 1.5 mM MgCl₂, 5% DMSO, and 1.25 units of AmpliTaq Gold DNA polymerase (Applied Biosystems). The PCR conditions were as follows: 31 cycles at 95 °C for 1 min, 61 °C for 1 min, 72°C for 1 min, with an initial denaturation at 95 °C for 5 min and a final extension at 72°C for 5 min. The MICA locus for Polynesian sub-populations Cook

Chapter 3.2

Islands, Samoa, Tokelau and Tonga was genotyped using SBT as previously described [32,33]. The MICA deletion was not specifically confirmed in Polynesian sub-populations. DNA sequences were compiled and assigned using Match Tools and MT Navigator software (Applied Biosystems, Foster City, CA, USA)

2.3. Statistical analyses

Allele frequencies were obtained by the direct counting method. Direct counting was applied because no typing ambiguities were observed and individuals are considered homozygotes if only one type of allele group detected at particular HLA and MICA locus, unless MICA deletion was detected for the latter. Observed MICA allele frequencies were generated for Polynesian sub-populations (i.e. Cook Islands, Tokelau, Tonga and Samoa) including the blank allele as MICA deletion could not be excluded or confirmed (Velickovic, unpublished). Haplotype frequencies estimation (i.e. Expectation-Maximization), test for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) between pairs of HLA loci, compute LD coefficients between pairs of alleles at different loci, exact test of population differentiation and Ewens-Watterson homozygosity (F) tests were carried out using algorithm included in Arlequin software version 3.1 [34]. The LD between three alleles at three different loci was calculated as described elsewhere in [13]. Multi-Variate Statistical Package version 3 (Kovach Computing Services; http://www.kovcomp.com/mvsp) was used to build PCO plots. The heatmap plot was constructed using ggplots2 and RColorBrewer in R (http://www.r-project.org/).

3. Results

The numbers and frequencies of HLA and MICA alleles observed in our new survey on Polynesian and Maori individuals are listed in Table 1. Our data for MICA are novel and the first reported for Polynesians and Maori. Statistical tests on genotypic HLA and MICA data show that Polynesian and Maori study groups and sub-groups are in HWE except for HLA-C in PFA, HLA-DRB1, -DPB1 in TM, and MICA in both, TM and MFA. There is significant linkage disequilibrium between pairs of HLA-A-C, HLA-C-B, HLA-DQA1-DRB1, HLA-DQB1-DRB1, HLA-DPA1-DPB1 and HLA-B-MICA in both Maori and Polynesian study groups and sub-groups. Ewens-Watterson homozygosity tests based on HLA and MICA allele frequencies (in Table 1) are shown in Table S1. Together with our newly collected data in Table S1 are HLA data from other Polynesian populations [8,9,35-37] - see later for population comparison. Overall, the homozygosity values for Polynesian and Maori datasets are around or below those expected under the assumption of neutrality with only Tonga (HLA-B, -DQB1 and -DRB1), Samoa (HLA-DQB1) and TM (HLA-DPA1) being significantly lower.

We have made every effort to compare our data with all of those available for related populations in other published sources. This is often difficult because different studies examine different sets of loci having greater or lesser overlap with ours. This means that for some loci that we tested, there are more sources with which we can make comparisons than for others. We began with a detailed examination of our data and compared them with [8,9,35] as the closest comparable works on Polynesians (see Table S2). For accuracy, comparisons were only made on HLA allele frequencies as MICA deletion was not specifically confirmed in the previous study of Polynesians (see Table S2 for reason to exclude MICA). The HLA allele frequencies do sometimes show significant differences at some loci among individual Polynesian populations in these earlier studies. Nonetheless, we combined Cook Islands, Tokelau, Tonga and Samoa data into a larger unified group, VTP (Velickovic Total Polynesians). This matches our PFA fairly closely at all of the six tested HLA loci (see Table S3). Our TP group shows good internal agreement with PFA at all loci tested, but is differentiated from VTP at both, HLA-B (p = 0.02) and -DRB1 (p = 0.01). This is probably due to the inclusion of individuals who have admixture with Europeans as a component of TP group [see later and 13]. In contrast, the previous Polynesian datasets [8,9,35] exclusively comprise full ancestry Polynesian individuals. Based on the statistical results, we decided to combine PFA with VTP as PMD (Polynesian meta dataset- see Table S2) for comparisons with other populations (see later). In addition, the most common HLA class I and II haplotype combinations found for PFA match well with those for other Polynesians reported by [8,9,35] - see Table 2. This provides further strong evidence for the value of using the PMD grouping in order to gain more confidence via greater statistical power when testing for Polynesian ancestry. Our present survey also provides novel 2- and 3-locus haplotype combinations for HLA-B*-DRB1*, HLA-A*-B*-DRB1* and HLA-B*-MICA* for Polynesians and these are listed in Table 2.

Next, we compared our Maori data with those from a previous study on Maori (TTM), reported by [36,37]. Here, only 5 out of 7 tested loci are common to both datasets. The most common HLA alleles in MFA are also predominant in TTM, but with slightly reduced frequencies; e.g. *HLA-A*24:02* is 0.36 in MFA compared with 0.21 in TTM, *HLA-B*55:02* is 0.36 in MFA, but only 0.18 in TTM, *HLA-C*01:02* is 0.45 in MFA vs. 0.21 in TTM and *HLA-DRB1*12:01* is 0.24 in MFA, but only slightly lower (0.22) in TTM (see Table S2). The HLA data recorded for TTM have more alleles than MFA and some are missing also in

TM. Both observations are probably due to undisclosed European admixture in TTM as previously recognized by [13]. In addition, statistical tests show some significant differences in allele frequencies between MFA and TTM except at HLA-A and -DPB1 (see Table S3). It is clear that TTM resembles TM more closely than its MFA sub-group. Our original plan to create a larger Maori study group by combining MFA with TTM had to be rejected (after ref 13) due to admixture in the latter. Hence, TTM and TM may be combined as a Maori meta dataset (MMD) representing present day Maori living in New Zealand. In contrast, the HLA data from MFA will be used as a representative group for full ancestry Maori in historical reconstructions. This contrasts with our findings and subsequent analysis in the study of HLA and MICA loci in TP, where the admixture is not too obvious because only a few of the individuals are known to have some European admixture compared with those in TM. The most common haplotypes in our Maori study group and sub-group are also compared with those reported for TTM in Table 2. Visual inspection shows the most frequent haplotypes recorded for MFA have slightly higher values than the corresponding haplotypes in TTM. For example; *HLA-A*02:06-C*01:02* is 0.28 in MFA, but only 0.15 in TTM, *HLA-C*01:02-*B*55:02 is much higher in MFA (0.36) compared with 0.05 in TTM [36]. Therefore, the distributions of haplotype frequencies further support our earlier claim [13] regarding unknown but significant levels of admixture in TTM and our decision to use MFA for the ancestry study of Maori.

We also examined genetic affinities between New Zealand Maori and other Polynesian populations. Generally, the new data collected for MFA are consistent with those from other Polynesians (both, new and previous data) with only minor exceptions, most probably at the HLA-A, -B and -DRB1 loci (see Tables 1 and S2). These differences are due to relatively higher frequencies of *HLA-A*02:06* (0.30 in MFA vs. 0.07-0.17 among Polynesians), -

B*55:02 (0.36 in MFA vs. 0.07-0.22 among Polynesians) and -DRB1*12:01 (0.24 in MFA vs. 0.08-0.22 among Polynesians) in MFA. We also compared the most common 2- and 3-locus haplotype associations between MFA and Polynesians (see Table 2). These datasets share similar allelic combinations where the most common ones observed in MFA, are also abundant in Polynesians, but with different frequencies (see Tables 2). There are a few exceptions to this generality; *HLA-A*24:02-C*03:04* is only detected as a significant association in Maori while *HLA-B*40:01-DRB1*09:01* was only recorded in PFA. Thus, minor differences between MFA and other Polynesians echo the previous observations on individual Polynesian populations [8,9,35] and may be attributed to a combination of historical and local selective forces - see later.

These are the first ever MICA reports for Polynesians. Therefore, we provide a complete analysis of HLA-B-MICA associations for the combined Polynesian and Maori populations (i.e. TP and TM). A heatmap of linkage disequilibrium values (r^2) was constructed for all possible MICA-HLA-B associations possible in the combined Maori and Polynesian populations that were typed at HLA-B and MICA loci (see Fig. 1). For haplotype frequencies $\geq 1\%$, statistically significant ($X^2 \geq 6.64$, p ≤ 0.01 , 1 d.f.) linkage disequilibrium was observed for the following MICA*-HLA-B* haplotypes: *008-0801*, *008-18:01*, *008-18:01*, *008-18:01*, *008-18:01*, *008-48:01*, *010-39:01*, *012-40:10*, *012-55:02*, *045-39:01*, *del-48:01*.

Influences of the presumed A-SM and P-SA ancestors in shaping the descendant Polynesian and Maori genepools were demonstrated by plotting the most common HLA and MICA alleles at each locus on bar diagrams (Fig. 2). Several of the alleles can be linked to both A-SM and P-SA populations [38-40], but no simple correlation can be made to point to a single unique Polynesian origin. We also compared the most common 2-locus haplotype associations (see Table 2) observed in our Polynesians with A-SM and P-SA populations [38,40]. Most of the predominant haplotypes observed in Polynesians and Maori can be traced back to either A-SM or P-SA populations, but a few of them are unique only to Polynesians (e.g. *HLA-A*34:01-C*12:03* and *HLA-C*12:03-B*13:01* - see Table 2).

Finally, a PCO plot based on HLA-A, -B, -C and -DRB1 allele frequencies (Fig. 3) was used to examine genetic relationships between Polynesian populations and to place them in a wider Asia-Pacific context [8,9,35-40]. The PCO plot shows clear separation of A-SM, P-SA and Polynesian populations. In the plot, Maori and other Polynesians appear to be closely related to one another and fall in between clusters representing A-SM and P-SA populations.

HLA-A*	TP (N=29)	PFA (N=22)	TM (N=114)	MFA (N=48)
01:01	0.05	0.05	0.07	0.05
02:01	0.12	0.09	0.08	0.03
02:05	-	-	0.02	0.01
02:06	0.10	0.09	0.22	0.30
02:12	-	-	0.01	0.01
02.12	_	_	0	-
03.01	_	_	0.04	_
11:01	0.14	0.18	0.18	0.17
24.01	-	-	0.10	0.01
24:02	0.36	0.34	0.27	0.37
25:01	-	-	0.01	-
26:01	0.03	0.05	0.01	_
30.02	-	-	0	_
31.01	_	_	0	_
33.01	_	_	Ő	_
34.01	0.19	0.21	0 07	0.05
68:01	-	-	0.01	-
k	7	7	17	9
HLA-B*	TP (N=29)	PFA (N=22)	TM (N=114)	MFA (N=48)
07.02	0.02		0.04	0.01
07:05	-	_	0	0.01
08:01	0.02	0.02	0 03	0.02
13.01	0.02	0.02	0.02	0.02
14.01	-	-	0.01	-
14.02	_	_	0.01	_
15:01	0.02	_	0.03	0.01
15:06	0.05	0.02	0	-
18:01	0.02	-	0.02	-
27:04	-	-	0	0.01
27:05	0.02	0.02	0	-
35:03	-	-	0	-
37:01	-	-	0	-
38:01	-	-	0	-
39:01	0.05	0.07	0.08	0.08
39:05	-	-	0.01	-
39:06	-	-	0.01	_
40:01	0.19	0.23	0.07	0.10
40:02	0.09	0.09	0.05	0.02
40:10	0.07	0.09	0.06	0.08
40:06	-	-	0	-
44:02	0.05	0.02	0	0.02
45:01	-	-	0	-
47:01	-	-	0	-
48:01	0.07	0.09	0.14	0.15
50:01	-	-	0.01	-
51:01	-	-	0	-
54:01	-	-	0	-
55:01	-	-	0	-
55:02	0.12	0.09	0.26	0.37
56:01	0.02	0.02	0.02	0.01
56:02	0.14	0.14	0.04	0.06
57:01	0.02	0.02	0	-
58:01	-	-	0.01	-
k	17	14	34	15

Table 1: Number and frequency of HLA and MICA alleles identified across HLA and MICA loci in our Polynesian and Maori study groups and sub-groups

HLA-C*	TP (N=29)	PFA (N=22)	TM (N=107)	MFA (N=46)
01:02	0.29	0.25	0.31	0.45
03.03	0.02	0.02	0.02	-
03:04	0.10	0.11	0.01	0.01
04.01	0.03	0.05	0.06	0.10
04:03	0.16	0.16	0.07	0.09
05:01	0.03	0.02	0.04	0.01
06.02	0.02	0.02	0.03	_
07:01	0.03	0.02	0.04	0.02
07:02	0.09	0.09	0.16	0.11
07:04	-	-	0.01	0.01
08:01	0.07	0.09	0.13	0.13
08:02	_	_	0.03	-
12:02	-	-	0.01	0.01
12:03	0.05	0.07	0.02	0.03
15:02	0.10	0.09	0.06	0.02
15:05	-	-	0.01	0.01
k	12	12	16	13
HLA-DRB1*	TP (N=37)	PFA (N=26)	TM (N=114)	MFA (N=49)
01.01	0.03	0.02	0.02	0.01
01:02	-	-	0.02	-
01:02	_	-	0.01	_
03:01	0.01	0.02	0.04	0.02
04:01	0.01	-	0.04	0.02
04:01	0.03	0.15	0.05	0.03
04:04	-	-	0	-
04:05	0.05	0.08	0.02	0.04
07:01	-	-	0.04	_
08:01	-	-	0.01	-
08:02	0.01	0.02	0.01	-
08:03	0.10	0.12	0.11	0.18
09:01	0.16	0.19	0.06	0.10
11:01	0.05	0.06	0.17	0.18
12:01	0.15	0.14	0.21	0.25
13:01	0.03	0.02	0.02	-
13:02	0.01	-	0.01	-
13:04	0.01	-	-	-
14:01	-	-	-	0.01
14:04	-	-	0	_
14:05	0.01	-	0	-
14:08	0.05	0.04	0.04	0.02
14:54	0.05	0.08	0.04	0.02
15:01	0.05	0.04	0.04	0.02
15:02	0.04	0.04	0	-
k	18	14	23	13
HLA-DPA1*	TP (N=25)	PFA (N=17)	TM (N=114)	MFA (N=42)
01:03	0.48	0.47	0.40	0.32
01:05	0.02	-	-	-
02:01	0.02	-	0.03	0.02
02:01/02	0.02	0.03	0.01	-
02:02	0.46	0.50	0.50	0.66
k	4	3	3	3
HLA-DPB1*	TP (N=25)	PFA (N=17)	TM (N=91)	MFA (N=42)
--------------	--------------	--------------	----------------------	----------------------
01:01	-	-	0.02	0.01
02:01	0.16	0.18	0.17	0.11
02:02	-	-	0.02	0.01
04:01	0.08	0.06	0.04	0.02
04:01/134:01	0.20	0.21	0.22	0.18
04:02	0.06	0.03	0.01	-
05:01	0.46	0.47	0.48	0.66
09:01	-	-	0.01	0.01
10:01	0.02	0.03	-	-
13:01/107:01	-	-	0.01	-
14:01	0.02	0.03	-	-
15:01	-	-	0.01	-
15:01/74:01	-	-	0.01	-
17:01	-	-	0.01	-
20:01	-	-	0.01	-
35:01	-	-	0.01	-
k	7	7	14	7
MICA*	TP (N=25)	PFA (N=17)	TM (N=91)	MFA (N=42)
001	-	-	0.02	-
002	-	-	0.04	-
007	0.02	0.03	0.01	-
008	0.44	0.56	0.34	0.27
009	-	-	0.02	0.02
010	0.02	-	0.04	0.02
011	-	-	0.02	-
012	0.36	0.21	0.43	0.56
017	-	-	0.01	-
019	0.02	0.03	0.01	-
033			0.01	0.02
	-	-	0.01	0.02
045	0.08	0.09	0.01	0.02
045 del	0.08 0.06	0.09 0.09	0.01 0.03 0.02	0.02 0.06 0.04

Table 1 cont.

Abbreviations: N= total number of individuals, k = total number of alleles observed, 0 = < 0.01, - = not observed, TM: Total Maori, MFA: Maori with Full Ancestry, TP: Total Polynesians, PFA: Polynesians with Full Ancestry

HLA-A*-C*	TP (N=49)	PFA (N=22)	TM (N=107)	MFA (N=41)	Polynesians	Possible origins
02:06-01:02	0.07	$0.02^{\$}$	0.19	0.28	0.03-0.15 ^[35,36]	TA (0-0.07) ^[38]
11:01-04:03	0.09	0.11	0.06	0.09	$0.03 - 0.12^{[35,36]}$	Filipinos (0.04) ^[38] , TA (0.01-0.34) ^[38]
11:01-08:01	0.03	$0.05^{\$}$	0.08	$0.04^{\$}$	$0.03 - 0.07^{[35,36]}$	$TA^{(0.01-0.06)^{[38]}}$
24:02-01:02	0.15 [§]	0.20	0.10	$0.16^{\$}$	$0.02 - 0.21^{[35,36]}$	Filipinos (0.01) ^[38] ,TA (0.01-0.11) ^[38]
24:02-03:04	$0.08^{\$}$	$0.07^{\$}$	-	-	$0.03 - 0.09^{[35]}$	Filipinos (0.03) ^[38] , TA (0.05-0.47) ^[38]
24:02-04:01	$0.02^{\$}$	$0.02^{\$}$	0.05	$0.06^{\$}$	$0.03 - 0.28^{[35,36]}$	Filipinos (0.04) ^[38] , TA (0.01-0.16) ^[38]
24:02-04:03	$0.07^{\$}$	-	-	-	$0.04 - 0.08^{[35,36]}$	$TA(0.01-0.06)^{[38]}$
24:02-07:02	-	-	0.08	0.09	$0.05 - 0.11^{[35,36]}$	Filipinos (0.05) ^[38] , TA (0.02-0.49) ^[38]
34:01-12:03	0.05	0.07	0.01	0.03	0.05 ^[35]	(recombinant of A-SM-P-SA)
34:01-15:02	0.08	0.09	0.05	0.02	$0.04 - 0.06^{[35]}$	Filipinos $(0.07)^{[38]}$
HLA-C*-B*	TP (N=49)	PFA (N=22)	TM (N=107)	MFA (N=41)	Polynesians	Possible origins
01:02-55:02	0.12	0.09	0.25	0.36	0.07-0.22 ^[35,36]	TA (0.02-0.13) ^[38]
01:02-56:02	0.14	0.14	0.04	0.07	$0.04 - 0.12^{[35]}$	PNGH (>0.02) ^[40]
03:04-40:01	0.10	0.11	0.01	0.01	$0.08-0.21^{[35]}$	Filipinos (0.08) ^[38] ,TA (0.02-0.23) ^[38]
04:01-40:01	0.03	0.05	0.06	0.10	$0.02 - 0.29^{[35,36]}$	Filipinos (0.07) ^[38] , TA(0.01-0.21) ^[38]
04:03-40:10	0.07	0.09	0.07	0.09	$0.04 - 0.06^{[35]}$	USA Asian (0.06) ^[38]
07:02-39:01	0.05	0.07	0.08	0.09	$0.03 - 0.10^{[35]}$	Filipinos (0.03) ^[38] , TA (0.04-0.54) ^[38]
08:01-48:01	0.07	0.09	0.13	0.13	$0.10-0.11^{[35]}$	Filipinos (0.04) ^[38] , TA (0.04-0.26) ^[38]
12:03-13:01	0.05	0.07	0.01	0.03	0.03 ^[35]	(recombinant of A-SM-PSA)
15:02-40:02	0.09	0.09	0.05	0.02	$0.04 - 0.07^{[35]}$	PNGH (>0.02) ^[40]
HLA-A*-C*-B*	TP (N=29)	PFA (N=22)	TM (N=107)	MFA (N=46)	Polynesians	Possible origins
02:06-01:02-55:02	0.05	$0.02^{\$}$	0.19	0.26	0.04-0.13 ^[35]	•
11:01-04:03-40:10	0.05	0.07	0.06	0.09	$0.05^{[35]}$	•
11:01-08:01-48:01	0.03	0.05	0.08	$0.04^{\$}$	0.04 ^[35]	•
24:02-01:02-55:02	$0.07^{\$}$	$0.07^{\$}$	0.05	$0.09^{\$}$	$0.03 - 0.08^{[35]}$	•
24:02-01:02-56:02	0.10	0.14	0.04	0.07	$0.03 - 0.11^{[35]}$	•
24:02-04:01-40:01	$0.02^{\$}$	0.02	0.06	0.06	$0.05 - 0.27^{[35]}$	•
24:02-03:04-40:01	0.07	$0.07^{\$}$	-	-	$0.05^{[35]}_{[35]}$	•
24:02-04:03-15:06	0.05	-	-	-	$0.04^{[35]}$	•
24:02-07:02-39:01	-	-	0.07	0.09	0.03-0.09 ^[35]	•
24:02-08:01-48:01	$0.03^{\$}$	0.05	0.04^{8}	0.08	$0.04 - 0.06^{[35]}$	•
34:01-12:03-13:01	0.05	0.07	0.01	0.03	-	•
34:01-15:02-40:02	0.09	0.09	0.05	0.02	$0.04 - 0.06^{135}$	•

Table 2: List of 2-locus haplotypes in Polynesians and Maori in comparison with other Asia Pacific populations

Table 2 cont.

HLA-B*-DRB1*	TP (N=27)	PFA (N=21)	TM (N=111)	MFA (N=48)	Polynesians	Possible origins
48.01 11.01	0.02	0.02	0.00	0.06		Eilipipos $(0.03)^{[38]}$ TA $(0.02.0.07)^{[38]}$
40.01-11.01	0.02	0.02	0.09	0.00	•	$TA (0.03, 0.04)^{[38]}$
55.02 00.01	0.07	0.05	0.18	0.24	•	TA (0.03 - 0.04) TA (0.02 0.03) ^[38]
40.01 11.01	0.00	0.05	0.05	0.09	•	$Filipipos (0.01)^{[38]} TA (0.01.0.13)^{[38]}$
40.01-11.01	-	-	0.03	0.08	•	Thiplinos $(0.01)^{[38]}$, TA $(0.01-0.13)$
40.01-00.03	0.04	0.03	0.04	0.08	•	Filipinos $(0.01)^{[38]}$ TA $(0.01-0.08)^{[38]}$
HLA-A*-B*-DRB1*	TP (N=27)	PFA (N=21)	TM (N=111)	MFA (N=48)	Polynesians	Possible origins
02.06 55.02 12.01			0.19	0.24		
02:06-55:02-12:01	0.06	-	0.18	0.24	•	•
11:01-48:01-11:01	0.02	-	0.06	0.04	•	• • • [38]
24:02-40:01-09:01	0.06°	0.10	-	-	•	Asian populations ¹⁰⁰
24:02-40:01-11:01	- 0.00 ⁸	- 0.00 [§]	0.05	0.08	•	Taiwanese
24:02-48:01-08:03	0.02°	0.02°	0.02	0.06	•	•
24:02-55:02-09:01	0.06	0.05	0.04	0.06	•	•
24:02-56:02-04:03	0.04*	0.05*	0.02	0.04^{s}	•	•
34:01-40:02-14:54	0.04	0.05	-	-	•	•
HLA-DPA1*-DPB1*	TP (25)	PFA (17)	TM (91)	MFA (42)	Polynesians	Possible origins
01:03-02:01	0.14	0.15	0.16	0.11	0.19-0.32 ^[8]	PNGH (0.36) ^[38]
01:03-04:01/134:01	0.16	0.18	0.21	0.18	-	•
01:03-04:01	$0.06^{\$}$	$0.06^{\$}$	0.04	0.02	0.19-0.39 ^[8]	PNGH (0.37) ^[38]
02:02-05:01	0.46	0.47	0.48	0.66	0.26-0.56 ^[8]	Japan (0.38) ^[38] , PNGH (0.15) ^[38]
HLA-B*-MICA*	TP (N=20)	PFA (N=15)	TM (N=90)	MFA (N=42)	Polynesians	Possible origins
13:01-045	0.08	0.10	0.02	0.05	•	China $(0.08)^{[38]}$, South Korea $(0.03)^{[38]}$
39:01-008	0.03 [§]	0.03 [§]	0.08	0.07	•	•
40:01-008	0.25	0.30	0.07	0.11	•	China (0.12) ^[38]
	··=-	0.20				
40:02-008	0.08	0.10	0.06	0.02	•	Thailand (0.03) ^[38]
40:02-008 40:10-012	0.08 0.05	$0.10 \\ 0.07^{\$}$	0.06 0.06	0.02 0.08	•	Thailand (0.03) ^[38]
40:02-008 40:10-012 48:01-del	0.08 0.05 0.08	0.10 0.07 [§] 0.10	0.06 0.06 0.02	0.02 0.08 0.02	• •	Thailand (0.03) ^[38]
40:02-008 40:10-012 48:01-del 55:02-012	0.08 0.05 0.08 0.10	0.10 0.07 [§] 0.10 0.03 [§]	0.06 0.06 0.02 0.28	0.02 0.08 0.02 0.38	• • •	Thailand (0.03) ^[38] • •

Abbreviations: N= total number of individuals, - = not observed, \bullet = not reported, TM: Total Maori, MFA: Maori with Full Ancestry, TP: Total Polynesians, PFA: Polynesians with Full Ancestry, § = alleles are not significantly associated (*p*-value>0.05), TA = Taiwan Aborigines, PNGH = Papua New Guinea Highlanders



Fig. 1: Linkage disequilibrium p-values for HLA-B*-MICA* associations in combined Polynesian and Maori populations (N=110).



Fig. 2: Comparison of the most common HLA and MICA alleles in Polynesians (purple fill) and Maori (bright purple fill) to those in Austronesian-speaking Mongoloid (A-SM)/Asia (blue fill) and Papuan-speaking Australoid (P-SA - red fill) populations. Data for reference populations were obtained from [38-40]. Abbreviations; PMD: Polynesian meta dataset, MFA: Maori with Full Ancestry, AYA: Australian Yuendemu Aborigine, ACYPA: Australia Cape York Peninsula Aborigine, AKA: Australian Kimberly Aborigine, PNGH: Papua New Guinea Highland, *: the highest allele frequencies recorded in Allele Frequencies.net or observed in the present survey.



Fig. 3: Relationships of Polynesian and Maori datasets shown in a wider Asia Pacific context. Reference populations were obtained from [8,9,35-40]. Abbreviations; PMD: Polynesians meta dataset, VTP: Velickovic Total Polynesians, TP: Total Polynesians, PFA: Polynesians with Full Ancestry, TM: Total Maori, MFA: Maori with Full Ancestry, PNG: Papua New Guinea, A-SM: Austronesian-speaking Mongoloids, P-SA: Papuan-speaking Australoids, TTM: Total Tracey Maori, MMD: Maori Meta dataset.

Chapter 3.2

4. Discussion

Ethnic diversity in the New Zealand population provides valuable opportunities for genetic studies of ancestry and health. The 4 millions living there today include of Europeans: 68%, Maori: 15%, Pacific people: 7%, Asians: 9% and others: 1%. The Pacific peoples principally include Samoans, Cook Islanders, Tongans, Fijian and Tokelauans (Statistics New Zealand; http://www.stats.govt.nz). Previously [13], we tested the expectation that both, Polynesians and Maori will be intermediate between A-SM and P-SA but closer to the former (minimally 70% according to the SNP data – see Introduction). We found that the HLA genepool may be up to 100% A-SM at some loci - e.g. HLA-B and -C. Our previous findings thus generally support the account [6,41] deduced from Y-chromosome, mtDNA and genome wide SNP data [10,14,15,25], but may point to past locus-specific selection events.

Here, we have further characterized HLA and MICA loci using a combination of high resolution SBT and Luminex typing in the same subjects reported earlier [13]. In particular, statistically significant differences observed between our admixed groups (TP and TM) and full ancestry sub-groups (PFA and MFA) are associated with European admixture. The same effects were observed in previous studies on Maori [36,37]. In contrast, slight differences between our full ancestry study sub-groups (PFA and MFA) and Polynesian sub-populations (Cook Islands, Tokelau, Tonga and Samoa) might be due to genetic refinement via selective pressure and longer-term bottlenecks as the genepool of Polynesian sub-populations were repeatedly refined by founder effects due to multiple re-settlements [8,9,35]. Overall, these have led us to the same conclusion regarding A-SM/P-SA hybrid ancestry of Polynesians and extensive European contributions in genepools of modern day Maori and Polynesians living in New Zealand (see Results and Fig. 3). Presently, we lack adequate comparative data from

129

A-SM and P-SA populations for the MICA locus. However, allele frequencies for MICA in Polynesians and Maori are most similar to those observed in Asian populations, rather than being similar to Europeans, Africans or Amerindians [38,42-47]. Two subjects were homozygous for HLA-B*48:01 (one MFA and one PFA) and no MICA alleles were detected. In addition, 19 of the 20 individuals who carried HLA-B*48:01 and another HLA-B allele, had only one MICA allele detected. Interestingly, only two of the 19 subjects carrying the possible MICA deletion haplotypes were PFA and the remaining were Maori (8 MFA and 9 admixed). When we tested specifically using pairs of primers for detecting the MICA gene, the presence of MICA*del in Maori and Polynesians which is associated with two individuals who are homozygous for HLA-B*48:01 and is shown to be in significant LD with HLA-B*48:01 (see Fig. 1). There is one caveat regarding this conclusion because there are six samples which have not been able to re-analyse. These include one MICA*008 homozygote and one MICA*012 homozygote. Thus, it is still possible (albeit unlikely) that these conceal MICA-del. In addition, we also provide the first ever HLA-B and MICA associations for Polynesian and Maori populations (see Table 2). We would like to point out that our haplotype analysis (Table 2) for HLA and MICA should be regarded as suggestive and provisional rather that definitive. This caveat is associated with less the use of extensive and sophisticated analysis and is due to several recombination hotsports present in between HLA and MICA loci that might harm haplotype predictions. There was a high frequency of the HLA-B*55:02-MICA*012 and HLA-B*40:01-MICA*008 haplotypes. We also observed potentially novel associations between HLA-B*39:01 and MICA*045 and HLA-B*39:01 and MICA*010. However, these associations may only highlight the general lack of MICA data for Asia Pacific populations, particularly from A-SM and P-SA. Frequencies of MICA alleles and B-MICA haplotypes are of clinical interest in the light of the literature suggesting that MICA antigens could be target for immune allorecognition with consequent antibody mediated rejection in kidney, pancreas and heart transplantation and deleterious transplant outcomes [48-50].

Our findings from this present survey conclusively exclude Amerindians as the first settlers in Polynesia [51]. However, we did detect common Amerindian alleles/haplotypes in two Maori individuals with admixed history and one Cook Island individual. Their genotypes are HLA-A*02:12/11:01-C*07:02/15:02-B*39:05/40:06-DRB1*08:02/14:04-DQA1*01GN/04MS-DQB1*04:02/05:03, HLA-A*02:12/03:01-C*06:02/07:02-B*39:05/47:01-DRB1*01:03/08:02-DQA1*04MS/05EF-DQB1*04:02/03CJB and HLA-A*02:12/34:01-C*07:02/12:03-B*13:01/39:05-DRB1*04:05/08:02-DQA1*03:01/04:01.

The highlighted alleles are common in Amerindians; see [52-54] for details. The *HLA-*A*02:12 and -DRB1*08:02 alleles were also detected in our MFA and PFA groups, respectively, but these samples were observed to have allele/haplotype combinations which are not common in Amerindians. However, we would not entirely exclude the possibility of a small amount of Amerindian admixture even among our self-proclaimed full ancestry Maori or Polynesian individuals. An *ad hoc* query to the Tissue Typing Laboratory Histotrac (SystemLink, Chantilly, VA, USA) database (New Zealand Blood Service) for either *HLA-*A*02:12 or -B*39:05 or -DRB1*08:02 shows at least 10 patients or donors with the full Amerindian haplotype *HLA-A*02:12-C*07:02-B*39:05-DRB1*08:02*. There may be more, because not all routine typing resolves to the allelic level (Paul Dunn, unpublished observations). The presence of an Amerindian haplotype in our data is most likely to be a historical signal due to a South American slave trade which operated across the Pacific before Europeans colonised New Zealand (Erik Thorsby, personal communication).

Overall, the present study on Maori and Polynesians shows a restricted set of alleles for HLA class I and MICA loci, compared with the more diverse nature of HLA class II (i.e. HLA-DRB1). Nonetheless, we propose that gene conversion involving a small number of HLA genes may be a mechanism of maintaining diversity at HLA class I protein level in Maori and Polynesians. For example, HLA-B*40:10 allele really does look like a genuine hybrid between HLA-B*55 or -B*56 and -B*40:01. The DNA sequence of the gene starts in the 5'UTR/promoter region exactly matching HLA-B*55 and -B*56 sequences and continues like this through exon 1 to the beginning of exon 2 (first 100 nucleotides), but then changes to match HLA-B*40:01 for all remaining exons and introns into 3'UTR (see Fig. 4). Therefore, amino acid differences between the protein encoded by HLA-B*40:10 compared with that encoded by HLA-B*40:01 are in the peptide binding groove, an immensely important region for generating diversity. The HLA-B*40:10 allele is almost absent from all other populations (see Fig. 4) and not reported anywhere for South American populations [38,55], but quite common among Polynesians (0.01-0.09). This suggests that HLA-B*40:10 was formed by gene conversion of ancestral (A-SM or P-SA) alleles. This is a reasonable proposal because, HLA-B*55 is quite frequent among the A-SM populations [38,39] while HLA-B*56 is more common among P-SA [38,40]. HLA-B*40 is common in both, A-SM and P-SA populations [38-40] and observed here (see Table S2) as B60 protein (HLA-B*40:01 and -B*40:02 alleles) and B61 (HLA-B*40:02 allele). This same mechanism has been inferred in Papuan populations in generating HLA-C*04:03, which is a frequent HLA-C allele among Polynesians and thought to be a hybrid of HLA-C*04 and -C*02 alleles [56]. The HLA-C*04:03 is also in associations with *HLA-B*40:10* in Polynesian populations (see Table 2).



Fig. 4: Recombinant structure of *HLA-B*40:10* allele and its distribution in various populations. The allele frequency maps were adapted from [55].

We should now examine whether natural selection has played an important role in shaping of HLA and MICA allele distributions in Polynesians. We would like to note that natural selection might be a reason for observed HWE deviation in PFA (HLA-C), TM (HLA-DRB1 and -DPB1) and TM and MFA (MICA), although linkage between pairs of HLA loci, founder effects and nature of our samples that make up our study groups and sub-groups might also be contributing factors. Homozygosity tests on HLA and MICA allele frequency spectra suggested that balancing selection is acting on these loci in Polynesian and Maori populations (see Table S1). These findings reflect those apparent from other studies [11,12,55]. Thus natural selection can be considered as a factor for having elevated and/or maintain high frequencies of several HLA and MICA alleles (e.g. HLA-B*55:02 and MICA*012 in MFA and HLA-DRB1*04:03 in Tokelauans) among individual Polynesian population, compared with those in their A-SM and P-SA ancestors [38-40,43]. However, there is also possible evidence for directional selection acting on HLA-A in MFA, HLA-C in Tokelau, HLA-DPA1 in Samoa, Tokelau and MFA and MICA in PFA (see Table S1). These explanation arise from finding relatively high frequencies of several ancestral (A-SM and P-SA) alleles (e.g. HLA-A*02:06, -A*11:01 and -A*24:02 in MFA and HLA-C*01:02 and -C*04:01 in Tokelauans). In our opinion, this apparent evidence of directional change on HLA loci in Polynesians might be better interpreted as due to founder effects, rather than directional selection. In Tokelauns, this is most likely due to a massive depletion of genepool in the 1800s, as a result of slave trade when ~95% of able bodied men were taken from the island in one year [57]. In addition, selective forces acting on particular Polynesian populations may well reflect the influence of more localised geographic factors on particular diseases, as these people are well scattered over the large Pacific basin in Remote Oceania.

Thus, it seems clear that maintenance of HLA and MICA diversity within Pacific populations is influenced by several factors which include both, local selective forces and founder effects. It is widely accepted that susceptibility and resistance to disease are genetically linked, especially for HLA and MICA loci. Ankylosing spondylitis (AS), uveitis (UV) and coeliac disease (CD) are autoimmune diseases strongly associated with tissue types

of European (CD) or European and Asian (AS and UV) origin. AS and UV are usually associated with B*27:05 in Europeans and B*27:04 with Asians [58,59]. The HLA-B*27 allele is at low frequency in Maori and Polynesians (0.01-0.02) populations characterized here and are probably only present at all due to European admixture. Coeliac disease (CD), or gluten intolerance as it is also known, is another common (~0.01) autoimmune disease in populations from Europe, South America, Australasia and the USA [60,61]. The risk of coeliac disease is strongly associated with the DQ2.5 (HLA-DQB1*02-DQA1*05:01/05) heterodimer and to a lesser extent with the DQ8.1 (HLA-DQB1*0302-DQA1*0301) heterodimer [62]. In the Maori and Polynesian samples, there were a small number (0.06-0.09) of DQ2.5 types, but many more (0.16-0.38) DQ8.1 haplotypes – see [13]. The low numbers of DQ2.5 haplotype may explain the low frequency of CD in Maori and Polynesians, but this is certainly not true for the DQ8.1 haplotype. Therefore, we now question the importance of the DQ8.1 haplotype in diagnosis of CD particularly in these subjects. The three HLA-associated diseases, namely CD, AS and UV, are not usually observed in Maori and Polynesians (Paul Dunn, unpublished observations). In addition, increasing admixture with Europeans may mean that CD, AS and UV will perhaps become known as emerging "European autoimmune diseases" in the Maori and Polynesian populations of New Zealand.

The pathogenesis of rheumatoid arthritis is also associated with HLA; *HLA-DRB1*04:01* in Europeans and *-DRB1*04:05* in Polynesians [63]. *HLA-DRB1*04:05* was only seen in a small number of the Maori and Polynesian individuals studied here (0-0.09) and thus might reflect the low prevalence of rheumatoid arthritis in these populations. However, *HLA-DRB1*04:03* is relatively common in our Maori and Polynesian populations, with frequencies from 0.08 in TM and up to 0.38 in the Tokelauans (see Table S2). *HLA-*

*DRB1*04:03* is considered 'not associated with', and possibly protective of Type 1 Diabetes (however, see further) and rheumatoid arthritis. Thus, selection in favour of *HLA-DRB1*04:03* might be thought of as a "health advantage" for Maori and Polynesians. Conversely, subacute cutaneous lupus erythematosus (SCLE) is much higher in Polynesians, compared with New Zealand Europeans [64]. The SCLE condition is thought to be genetically linked to the *HLA-A1-B8-DR3* haplotype [65], which is relatively uncommon in Maori and Polynesians (see Tables 1 and S1). Therefore, susceptibility factors to SCLE in Polynesians might involve other HLA antigens/candidate genes within MHC complex, as the effects of disease susceptibility genes may differ among ethnic groups [see 66].

Subcutaneous T cell paniculitis is a rare skin lymphoma found in various different ethnic populations. In Maori and Polynesians, this condition is treated well with cyclosporine indicating T cell involvement (Hilary Blacklock, personal communication). We have analysed the HLA types of 6 patients with this condition but have not observed an obvious association (Paul Dunn, unpublished observation). There is also no clear HLA association with type II diabetes in Maori and Polynesians, although *HLA-B22* and *-B40* have been implicated with this form of diabetes [67,68]. Recently, we have identified six Polynesian adult patients with 'late onset type I diabetes' - a hitherto unknown entity in Polynesians. Four of the five patients tested had HLA-DRB1-DQ haplotypes previously associated in European subjects with the latent autoimmune diabetes in adults (LADA) variant of type 1 diabetes, [68]. Three of our subjects were *HLA-DRB1*09:01-DQA1*03:01-DQB1*03:03* and one was *HLA-DRB1*04:03-DQA1*03:01-DQB1*03:02*. LADA may be more common in adult Polynesians than previously thought and this HLA information, in conjunction with anti-GAD antibody testing and C-peptide assay, could prove valuable in future diagnoses (Tangi Purea, Tim Cundy and Paul Dunn personal communication).

The MICA locus is also associated with a number of autoimmune diseases and most of them are also HLA associated [1]. This might be due to composition of the MHC region itself which contains several categories of genes related to innate and adaptive immunity; antigen processing (e.g. HLA class I and II), stress response (e.g. MICA and MICB), the complement cascade (e.g. BF and C2), inflammation (e.g. ABCF1 and AIF1), immunoglobulin superfamily (e.g. AGER and BTN1A1) and immune regulation (NFKBIL1 and RXRB) loci – see [70] for details. It is noteworthy that two individuals were positive for MICA*007:01 (one MFA and one Polynesian with admixed history subjects), an allele that has been reported in association with T1D in Japanese [71], Kawasaki disease in Taiwanese [72], and seronegative spondylarthropathies in Japanese [73] and Europeans [74]. Both MICA*007:01 carrying individuals appear to carry the following probable extended haplotype HLA-A*02:01-C*01:02-B*27:05-MICA*007:01-DRB1*01:01-DQA1*01:01-DQB1*05:01. However, many of the previous MICA-HLA and disease association studies involved a repetitive GCT motif in exon 5, encoding the MICA transmembrane domain. This appeared to be less informative as the repetitive polyalanine encoding motif is shared with a number of other well-characterized MICA alleles [3,75]. In our combined opinion, future studies should focus on association of HLA and MICA polymorphism with diseases, e.g. by looking at HLA and MICA alleles which are closely linked and crucial in immune responses In addition, HLA and MICA polymorphisms are also important in [see 1.2,70]. transplantation as the success rate of transplantation is dependent on the degree of HLA and MICA matching between donor and recipient [4,76,77]. The HLA and MICA data presented here suggest that the feasibility of finding suitable donors for Maori and Polynesian patients is within their own ethnic group(s), as they share a number of common and unique HLA and MICA alleles.

The complex genepool of Polynesians as previously described (41) can now be better understood. Re-interpretations of HLA and MICA data in the light of new scholarship [6] have brought a clearer picture of A-SM/P-SA ancestral fractions in these people, although HLA and MICA alleles do sometimes show slight differences between them in tracing their putative ancestors. Our survey also points out significant European gene flow in our TP and TM groups, which are more typical of modern day Polynesians and Maori living in New Zealand, respectively. This later phase of gene flow is medically significant as interaction between European and Polynesian genes might influence new patterns of disease resistance and vulnerability. Therefore, we believe that the HLA and MICA data from this study could be used as a preliminary medical reference standard for Maori and Polynesians living in New Zealand, also helping to redefine the immunogenetics of autoimmune diseases such as coelics disease and Type 1 diabetes.

Acknowledgements

The authors acknowledge Dr. Abeer Madbouly from the United States National Marrow Donor Program (NMDP) for bioinformatics consultation. This work is a part of collaborations between New Zealand Blood Service, Auckland, Victoria University of Wellington and Cleveland Clinic, Ohio.

References

- 1. Stephens HAF. MICA and MICB genes: can the enigma of their polymorphism be resolved? Trends Immunol 2001;22:378-85.
- Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. Clin Microbiol Rev 2009;22:370-85.
- Collins RWM. Human MHC class I chain related (MIC) genes: their biological function and relevance to disease and transplantation. Eur J Immunogenet 2004;31:105-14.
- 4. Zou Y, Stastny P. Role of MICA in the immune response to transplants. Tissue Antigens 2010;76:171-76.
- 5. Deng L, Mariuzza RA. Structural basis for recognition of MHC and MHC-like ligands by natural killer cell receptors. Seminars Immunol 2006;18:159-66.
- Bellwood P, Chambers GK, Ross M, Hung C-H. Are "cultures" inherited? Multidisciplinary perspectives on the origins of Austronesian–speaking peoples prior to 1000 BC. In: Roberts BW, Linden MV, editors. Investigating archaeological cultures: material culture, variability and transmission, Dordrecht: Springer; 2011, p. 321-54.
- Gao X, Zimmet P, Serjeantson SW. HLA-DR,DQ sequence polymorphisms in Polynesians, Micronesians, and Javanese. Hum Immunol 1992;34:153-61.
- Velickovic ZM, Carter JM. HLA-DPA1 and DPB1 polymorphism in four Pacific Islands populations determined by sequencing based typing. Tissue Antigens 2001;57:493-501.
- Velickovic ZM, Delahunt B, Carter JM. HLA-DRB1 and HLA-DQB1 polymorphisms in Pacific Islands populations. Tissue Antigens 2002;59:397-406.

- 10. Hagelberg E, Kayser M, Nagy, M, Roewer L, Zimdahl H, Krawczak M, et al. Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. Phil Trans R Soc Lond B 1999;354:141-52.
- 11. Bugawan TL, Mack SJ, Stoneking M, Saha M, Beck HP, Erlich HA. HLA class I allele distribution in six Pacific/Asian populations: evidence of selection at the HLA-A locus. Tissue Antigens 1999;53:311-9.
- Mack SJ, Bugawan TL, Moonsamy PV, Erlich JA, Trachtenberg EA, Paik YK, et al. Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. Tissue Antigens 2000;55:383-400.
- 13. Edinur HA, Dunn PPJ, Hammond L, Selwyn C, Velickovic ZM, Lea RA, et al. Using HLA loci to inform ancestry and health in Polynesian and Maori populations. Tissue Antigens 2012;80:509-22.
- 14. Kimura R, Ohashi J, Matsumura Y, Nakazawa M, Inaoka T, Ohtsuka R, et al. Gene flow and natural selection in Oceanic human populations inferred from genome-wide SNP typing. Mol Biol Evol 2008;25:1750-61.
- 15. Wollstein A, Lao O, Becker C, Brauer S, Trent RJ, Nurnberg P, et al. Demographic history of Oceania inferred from genome-wide data. Curr Biol 2010;20:1983-92.
- Simmons RT. Blood group genes in Polynesians and comparisons with other Pacific peoples. Oceania 1962;32:198-210.
- 17. Hill AVS, Gentile B, Bonnardot JM, Roux J, Weatherall DJ, Clegg JB. Polynesian origins and affinities: globin gene variants in Eastern Polynesia. Am J Hum Genet 1987;40:453-63.

- 18. Hertzbergh MS, Mickleson KNP, Trent RJ. α-Globin gene haplotype in Polynesians: their relationship to population groups and gene rearrangements. Am J Hum Genet 1988;43:971-7.
- 19. Trent RJ, Buchanan JG, Webb A, Goundar RPS, Seruvatu LM, Mickleson KNP. Globin genes are useful markers to identify genetic similarities between Fijians and Pacific Islanders from Polynesia and Melanesia. Am J Hum Genet 1988;42:601-7.
- 20. O'Shaughnessy DF, Hill AVS, Bowden DK, Weatherall DJ, Clegg JB. Globin genes in Micronesia: origins and affinities of Pacific Island peoples. Am J Hum Genet 1990;46:144-55.
- 21. Tsintsof AS, Hertzberg MS, Prior JF, Mickleson KNP, Trent RJ. α-Globin gene markers identity genetic differences between Australian Aborigines and Melanesians. Am J Hum Genet 1990;46:138-43.
- 22. Ohashi J, Naka I, Ohtsuka R, Inaoka T, Ataka Y, Nakazawa M, et al. Molecular polymorphism of *ABO* blood group in Austronesian and non-Austronesian populations in Oceania. Tissue Antigens 2004;63:355-61.
- 23. Ohashi J, Naka I, Kimura R, Tokunaga K, Nakazawa M, Ataka Y, et al. HLA-DRB1 polymorphism on Ha'ano Island of the Kingdom of Tonga. Anthropological Science 2006;114:193-8.
- 24. Benton M, Macartney-Coxson D, Eccles D, Griffiths L, Chambers G, Lea R. Complete mitochondrial genome sequencing reveals novel haplotypes in a Polynesian population. PLoS ONE 2012;7:e35026.
- 25. Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky LA, et al. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradient across the Pacific. Mol Biol Evol 2006;23:2234-44.

- 26. Underhill PA, Passarino G, Lin AA, Marzuki S, Oefner P, Cavalli-Sforza LL, et al. Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. Hum Mutat 2001;17:271-80.
- 27. Martinez A, Fernandez-Arquero M, Balsa A, Rubio A, Alves H, Pascual-Salcedo D, et al. Primary association of a MICA allele with protection against rheumatoid arthritis. Arthritis Rheum 2001;44:1261-5.
- 28. Mizuki N, Meguro A, Tohnai I, Gul A, Ohno S, Mizuki N. Association of major histocompatibility complex class I chain-related gene A and HLA-B alleles with Behcet's disease in Turkey. Jpn J Ophthalmol 2007;51:431-6.
- 29. Dunn P, Day S, Williams S, Bendukidze N. HLA sequencing as a tissue typing tool. In: Goulden NJ, Steward CG, editors. Pediatric hematology. Methods and protocols, New Jersey: Humana Press; 2003, p.233-46.
- 30. Ota M, Bahram S, Katsuyama Y, Saito S, Nose Y, Sada M, et al. On the MICA deleted-MICB null, HLA-B*4801 haplotype. Tissue Antigens 2000;56:268-71.
- 31. Komatsu-Wakui M, Tokunaga K, Ishikawa Y, Leelayuwat C, Kashiwase K, Tanaka H, et al. Wide distribution of the MICA-MICB null haplotype in East Asians. Tissue Antigens 2001;57:1-8.
- 32. Katsuyama Y, Ota M, Ando H, Saito S, Mizuki N, Kera J, Bahram S, Nose Y, Inoko H. Sequencing based typing for genetic polymorphisms in exons 2, 3 and 4 of the MICA gene Tissue Antigens 1999; 54:178-84.
- 33. Mizuki N, Ota M, Kimura M, Ohno S, Ando H, Katsuyama Y, Yamazaki M, Watanabe K, Goto K, Nakamura S, Bahram S, Inoko H. Triplet repeat polymorphism in the transmembrane region of the MICA gene: a strong association of six GCT repetitions with Behçet disease. Proc Natl Acad Sci 1997;94:1298-303.

- 34. Excoffier L, Laval G, Schneider S. Arlequin version 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 2005;1:47-50.
- 35. Velickovic ZM. PhD thesis: HLA polymorphism in Pacific Islands populations. Dunedin, New Zealand: University of Otago; 2001.
- 36. Tracey MC. PhD thesis: HLA allelic variation in New Zealand Maori. Dunedin, New Zealand: University of Otago; 2007.
- 37. Tracey MC, Carter JM. Class II HLA allele polymorphism: DRB1, DQB1 and DPB1 alleles and haplotypes in the New Zealand Maori population. Tissue Antigens 2006;68:297-302.
- 38. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucl Acids Res 2011;39:D913-9.
- 39. Chu CC, Lin M, Nakajima F, Lee H-L, Chang S-L, Juji T, Tokunaga K. Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. Tissue Antigens 2001;58:9-18.
- 40. Main P, Attenborough RD, Chelvanayagam G, Gao X. The peopling of New Guinea: evidence from class I human leukocyte antigen. Hum Biol 2001;73:365-83.
- 41. Chambers GK. Genetics and the origins of the Polynesians. In: Cooper DN, Kehrer-Sawatzki H, editors. Handbook of human molecular evolution, Chichester: John Wiley & Sons Inc; 2008, p. 489-97.
- 42. Petersdorf E, Shuler KB, Longton GM, Spies T, Hansen JA. Population study of allelic diversity in the human MHC class I-related MIC-A gene. Immunogenetics 1999;49:605-12.
- 43. Obuchi N, Takahashi M, Nouchi T, Satoh M, Arimura T, Ueda K, et al. Identification of MICA alleles with a long Leu-repeat in the transmembarane region and no

cytoplasmic tail due to a frameshift-deletion in exon 4. Tissue Antigens 2001;57:520-35.

- 44. Zhang Y, Lazaro AM, Zou Y, Lavingia B, Moraes ME, Moraes JR, et al. MICA polymorphism in South American Indians. Immunogenetics 2002;53:900-6.
- 45. Pyo C-W, Hur S-S, Kim Y-K, Choi H-B, Kim T-Y, Kim T-G. Distribution of MICA alleles and haplotypes associated with HLA in the Korean population. Hum Immunol 2003;64:378-84.
- 46. Tian W, Boggs DA, Uko G, Essiet A, Inyama M, Banjoko B, et al. MICA, HLA-B haplotypic variation in five population groups of sub-Sahara African ancestry. Genes Immun 2003;4:500-5.
- 47. Zhang Y, Han M, Vorhaben R, Giang C, Lavingia B, Stastny P. Study of MICA alleles in 201 African Americans by multiplex single nucleotide extension (MSNE) typing. Hum Immunol 2003;64:130-6.
- 48. Hankey KG, Drachenberg CB, Papadimitriou JC, Klassen DK, Philosophe B, Bartlett ST, et al. MIC expression in renal and pancreatic allografts. Transplantation 2002;73:304-6.
- 49. Suarez-Alvarez B, Lopez-Vazquez A, Diaz-Pena R, Diaz-Molina B, Blanco-Garcia RM, Alvarez-Lopez MR, et al. Post-transplant soluble MICA and MICA antibodies predict subsequent heart graft outcome. Transpl Immunol 2006;17:43-6.
- 50. Zou Y, Stastny P, Susal C, Dohler B, Opelz G. Antibodies against MICA antigens and kidney-transplant rejection. N Engl J Med 2007;357:1293-300.
- 51. Heyerdahl T. American Indians in the Pacific. The theory behind the Kon-Tiki expedition. London, UK: Allen and Unwin; 1952.

- 52. Lie BA, Dupuy BM, Spurkland A, Fernandez-Vina MA, Hagelbergh E, Thorsby E. Molecular genetic studies of natives on Easter Island: evidence of early European and Amerindian contribution to the Polynesian gene pool. Tissue Antigens 2006;69:10-18.
- Thorsby E. The Polynesian gene pool: an early contribution by Amerindians to Easter Island. Phil Trans R Soc B 2012;367:812-9.
- 54. Thorsby E, Flam ST, Woldseth B, Dupuy BM, Sanchez-Mazas A, Fernandez-Vina MA. Further evidence of an Amerindian contribution to the Polynesian gene pool on Easter Island. Tissue Antigens 2009;73:582-5.
- 55. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum Immunol 2008:69;443-64.
- 56. Little AM, Mason A, Marsh SG, Parham P. HLA-C typing of eleven Papua New Guineans: identification of an HLA-Cw4/Cw2 hybrid allele. Tissue Antigens 1996;48:113-7.
- 57. Maude HE. Slavers in paradise: the Peruvian slave trade in Polynesia, 1862-1864.Australia: Australian National University Press; 1981.
- 58. Lopez-Larrea C, Sujirachato K, Mehra NK, Chiewsilp O, Isarangkura D, Kanga U, et al. HLA-B27 subtypes in Asian patients with ankylosing spondylitis. Evidence for new associations. Tissue Antigens 1995;45:169-76.
- 59. Brown MA, Pile KD, Kennedy LG, Calin A, Darke C, Bell J, et al. HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. Ann Rheum Dis 1996;55:268-70.
- 60. Catassi C. The world map of coeliac disease. Acta Gastroenterol Latinoam 2005;35: 37-55.

- 61. Van Heel DA, West J. Recent advances in coeliac disease. Gut 2006;55:1037-46.
- 62. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, et al. HLA types in coeliac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Coeliac Disease. Hum Immunol 2003;64:469-77.
- 63. Tan OL, Farmiloe S, Roberts M, Geursen A, Skinner MA. HLA-DR4 subtypes in New Zealand Polynesians. Predominance of Dw13 in the healthy population and association of Dw15 with rheumatoid arthritis. Arthritis Rheum 1993;36:15-9.
- 64. Jarrett P, Scragg R, Thornley S. The prevalence of cutaneous lupus in Counties Manukau District Health Board. Australas J Dermatol 2011;52:A1-10.
- 65. Sontheimer RD, Stastny P, Gilliam JN. Human histocompatibility antigen associations in subacute cutaneous lupus erythematosus. J Clin Invest 1981;67:312-6.
- 66. Pickering MC, Perraudeau M, Walport MJ. HLA and systemic vasculitides, systemic lupus erythematosus and Sjögren's syndrome. In: Lechler R, Warrens A, editors. HLA in health and disease, London: Academic Press; 2000, p. 327-64.
- 67. Dittmer I, Woodfield G, Simpson I. Non-insulin-dependent diabetes mellitus in New Zealand Mori: a relationship with class I but not class II histocompatibility locus antigens. New Zeal Med J 1998;111:294-6.
- 68. Serjeantson SW, Ryan DP, Zimmet P, Taylor R, Cross R, Charpin M, et al. HLA antigens in four Pacific populations with non-insulin-dependent diabetes mellitus. Ann Hum Biol 1982;9:69-84.
- Pettersen E, Skorpen F, Kvaloy K, Midthjell K, Grill V. Genetic heterogeneity in latent autoimmune diabetes is linked to various degrees of autoimmune activity. Diabetes 2010; 59:302-10.

- 70. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet 2004;12:889-99.
- 71. Kawabata Y, Ikegami H, Kawaguchi Y, Fujisawa T, Hotta M, Ueda H, et al. Agerelated association of MHC class I chain-related gene A (MICA) with type 1 (insulindependent) diabetes mellitus. Hum Immunol 2000;61:624-9.
- 72. Huang Y, Lee YJ, Chen MR, Hsu CH, Lin SP, Sung TC, et al. Polymorphism of transmembrane region of MICA gene and Kawasaki disease. Exp Clin Immunogenet 2000;17:130-7.
- 73. Tsuchiya N, Shiota M, Moriyama S, Ogawa A, Komatsu-Wakui M, Mitsui H, et al. MICA allele typing of HLA-B27 positive Japanese patients with seronegative spondylarthropathies and healthy individuals: differential linkage disequilibrium with HLA-B27 subtypes. Arthritis Rheum 1998;41:68-73.
- 74. Yabuki K, Ota M, Goto K, Kimura T, Nomura E, Ohno S, et al. Triplet repeat polymorphism in the MICA gene in HLA-B27 positive and negative caucasian patients with ankylosing spondylitis. Hum Immunol 1999;60:83-6.
- 75. Robinson J, Perez-Rodriguez M, Waller MJ, Cuiller B, Bahram S, Yao Z, et al. MICA sequence 2000. Immunogenetics 2001;53:150-69.
- 76. Zwirner NW, Marcos CY, Mirbaha F, Zou Y, Stastny P. Identification of MICA as a new polymorphic alloantigen recognized by antibodies in sera of organ transplant recipients. Hum Immunol 2000;61:917-24.
- 77. Erlich HA, Opelz G, Hansen J. HLA DNA typing review and transplantation. Immunity 2001;14:347-56.

Supplementary information

		TP	PFA	CI	SA	TOK	TON	VTP	PMD	MFA	TM	TTM
HLA-A	obs. F	0.35	0.35	0.30	0.31	0.21	0.19	0.24	0.36	0.49	0.12	0.14
	exp. F	0.34	0.32	0.22	0.34	0.30	0.34	0.28	0.27	0.30	0.20	0.13
	p	0.08	0.11	0.86	0.42	0.78	0.49	0.70	0.70	0.40	0.48	0.74
HLA-B	obs. F	0.10	0.10	0.21	0.12	0.23	0.21	0.11	0.12	0.18	0.07	0.05
	exp. F	0.12	0.14	0.13	0.17	0.22	0.22	0.17	0.15	0.17	0.09	0.05
	p	0.19	0.23	0.25	0.46	0.66	0.01	0.20	0.31	0.69	0.87	0.96
HLA-C	obs. F	0.11	0.17	0.21	0.27	0.51	0.20	0.18	0.17	0.18	0.16	0.16
	exp. F	0.19	0.17	0.19	0.25	0.27	0.27	0.23	0.23	0.20	0.20	0.13
	p	0.24	0.21	0.73	0.08	0.22	0.13	0.21	0.15	0.82	0.27	0.15
HLA-DQB1	obs. F exp. F p	• •	• •	0.27 0.30 0.23	0.21 0.30 0.02	0.35 0.38 0.08	0.31 0.33 0.02	0.27 0.31 0.05	• •	• •	• • •	0.16 0.24 0.08
HLA-DRB1	obs. F	0.09	0.14	0.10	0.22	0.18	0.18	0.45	0.19	0.13	0.13	0.14
	exp. F	0.13	0.15	0.17	0.21	0.30	0.22	0.18	0.17	0.21	0.14	0.15
	p	0.12	0.13	0.29	0.06	0.24	0.01	0.16	0.17	0.22	0.25	0.07
HLA-DPA1	obs. F exp. F p	0.53 0.63 0.23	0.79 0.76 0.04	0.65 0.67 0.18	0.98 0.82 0.05	0.96 0.81 0.20	0.40 0.66 0.22	0.76 0.74 0.12	0.77 0.75 0.11	0.95 0.80 0.15	0.80 0.83 0.03	• •
HLA-DPB1	obs. F	0.31	0.49	0.37	0.45	0.49	0.43	0.51	0.18	0.35	0.22	0.17
	exp. F	0.43	0.41	0.37	0.49	0.49	0.43	0.33	0.33	0.48	0.31	0.22
	p	0.33	0.37	0.44	0.35	0.13	0.43	0.63	0.63	0.59	0.70	0.88
MICA	obs. <i>F</i> exp. <i>F</i> <i>p</i>	0.21 0.28 0.46	0.45 0.30 0.68	• •	• •	• •	• •	• •	•	0.29 0.32 0.35	0.17 0.21 0.72	• •

Table S1: Homozygosity (F) values for Maori and Polynesian datasets

Abbreviations: TP: Total Polynesians, PFA: Polynesians with Full Ancestry, TM= Total Maori, MFA = Maori with Full Ancestry, CI: Cook Islands [8,9,31], SA: Samoa [8,9,31], TOK: Tokelau [8,9,31], TON: Tonga [8,9,31], VTP: Velickovic Total Polynesians (CI+SA+TOK+TON), PMD: Polynesian Meta Dataset (PFA+VTP), TTM: Tracey Total Maori [32,33], MMD: Maori Meta Dataset (TM+TTM), obs: observed, exp: expected, F = homozygosity, p: significant level (<0.05), • = not reported

$\frac{A^{*}}{(29)} (22) 30) (30) (30) (400) (444) (48) (114) (176)$) (290)
01:01 0.05 0.05 0.01 0.01 0 0.01 0.01 0.01	0.06
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0.09
	0
	0.01
02:06 0.10 0.09 0.17 0.15 0.07 0.12 0.13 0.12 0.30 0.22 0.22	0.22
	0
02:12 - 0.01 - 0 0 0.01 0.01 0.01 0.01 0.01 0	0.01
	-
	0
02:119 0 - 0.04 - 0.01	0 05
0.01 0.01 0.01 0.04 0.00	0.05
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0.17
23.01	0
	0
24.02 0.36 0.34 0.48 0.56 0.50 0.46 0.50 0.48 0.36 0.27 0.2	0.24
24.03 0	0.21
24.05 0	0
24:14 0.0	0.01
25:01 0.01 0	0.01
26:01 0.03 0.05 0.02 0.01 0.05 0.01 0.02 0.02 - 0.01 0.0	0.01
26:05 0.01 - 0 0	-
29:02 0.0	0.01
30:01 0.01 0 0 0	0
30:02 0 -	0
31:01 0.01 0 0 - 0 0.0	0.01
32:01 0.0	0.01
33:01 0 -	0
33:03 0	0
34:01 0.19 0.20 0.13 0.08 0.11 0.07 0.10 0.11 0.05 0.07 0.09	0.08
43:01 0	0
66:03 0	0
68:01 0.01 0.0	0.01
68:02 0.01 0 0 0	-
K / / 13 9 8 8 15 15 9 1/ 2/	51
HLA- TP PFA CI TOK TON SA VTP PMD MFA TM TTM	I MMD
$B^{*} (29) (22) (50) (50) (50) (50) (400) (444) (48) (114) (171)$) (285)
07:02 0.02 0.03 0.01 0.01 0.01 0.04 0.03	0.06
07:03 0	0
07:04 0	0
07:05 0.01 0 0.02	0.02
07:06 0	0
07:07 0	0
07:10 0	0
07:45 0.0	0
08:01 0.02 0.02 0.01 0.01 0.01 0.01 0.02 0.03 0.02	0.02
	0
13:01 0.05 0.07 0.04 0.04 0.03 0.06 0.04 0.05 0.04 0.02 -	0.01
13:02 0.02 0.01 0 0.01	0.02
	0
14:02 0.01 0 0 - 0.01 0	0.01
15.01 0.02 - - - - - - 0.01 0.03 0.0.	0.02
	0
15:06 0.05 0.02 0.03 0.07 0.05 - 0.04 0.04 - 0 0.0	0.01

Table S2: HLA and MICA allele frequencies for Polynesian and Maori datasets

Table S2 cont.

HLA-	TP	PFA	CI	TOK	TON	SA	VTP	PMD	MFA	ТМ	TTM	MMD
B*	(29)	(22)	(50)	(50)	(50)	(50)	(400)	(444)	(48)	(114)	(171)	(285)
15:21	-	-	-	0.01	-	0.04	0.01	0.01	-	-	-	-
15:28	-	-	-	-	-	-	-	-	-	-	0	0
15:31	-	-	-	-	-	-	-	-	-	-	0	0
15:33	-	-	-	-	-	-	-	-	-	-	0	0
15:34	-	-	-	-	-	-	-	-	-	-	0	0
18:01	0.02	-	-	-	-	-	-	-	-	0.02	-	0.01
27:04	-	-	0.02	0.01	0.02	-	0.01	0.01	0.01	0	-	0
27:05	0.02	0.02	-	-	-	-	0	0	-	0	-	0
35:01	-	-	-	-	-	-	-	-	-	-	0.01	0.01
35:03	-	-	-	-	-	-	-	-	-	0	0.01	0.01
35:10	-	-	-	-	-	-	-	-	-	-	0	0
35:60	-	-	-	-	-	-	-	-	-	-	0	0
37:01	-	-	-	-	-	-	-	-	-	0	-	0
37:02	-	-	-	-	-	-	-	-	-	-	0.01	0
38:01	-	-	-	-	-	-	-	-	-	0	-	0
39:01	0.05	0.07	0.11	0.06	0.03	0.05	0.06	0.06	0.08	0.08	0.01	0.04
39:02	-	-	-	-	-	-	-	-	-	-	0.02	0.01
39:05	-	-	0.01	-	-	0.01	0.01	0	-	0.01	-	0
39:06	-	-	-	-	-	-	-	-	-	0.01	0.02	0.01
39:10	-	-	-	-	-	-	-	-	-	-	0	0
40:01	0.19	0.23	0.07	0.42	0.17	0.32	0.25	0.24	0.10	0.07	0.12	0.10
40:02	0.09	0.09	0.07	0.05	0.09	0.08	0.07	0.07	0.02	0.05	0.01	0.02
40:03	-	-	-	-	-	-	-	-	-	-	0	0
40:00	-	-	-	-	-	-	-	-	-	0	-	0.01
40:07	-	-	0.01	-	-	-	0.05	0 05	-	-	0.02	0.01
40:10	0.07	0.09	0.04	0.05	0.09	0.01	0.05	0.05	0.08	0.00	0.02	0.04
40:12	-	-	-	-	-	-	-	-	-	-	0	0
40.33	-	-	-	-	-	-	-	-	-	-	0	0
40.49	0.05	0.02	0.01	-	-	-	0	0	0.02	0.04	0.02	0.03
44.02 11.03	0.05	0.02	0.01	_	_	_	0	0	0.02	0.04	0.02	0.03
44.05 44.23N	-	-	-	-	-	-	-	-	-	-	0.01	0.01
44.231 15.01	_	_	-	_	_	_	_	_	_	0	0	0
43.01 47.01	_	_	_	_	_	_	_	_	_	0	_	0
48.01	0.07	0.09	0.11	0.01	0.11	0.10	0.08	0.08	0.15	0 14	0 14	0 14
48.03	-	-	-	-	-	0.10	0.00	0.00	-	-	0.02	0.01
48.05	-	_	_	_	_	-	-	-	_	_	0.02	0.01
48.14	-	_	_	_	_	_	_	-	_	_	Ő	0
50:01	-	-	-	-	-	-	-	-	-	0.01	0.01	0.01
51:01	-	-	0.01	-	-	-	0.01	0	-	0	0.01	0.01
51:02	-	-	0.01	-	-	-	0	Ō	-	-	-	-
51:09	-	-	-	-	-	-	-	-	-	-	0	0
52:01	-	-	-	-	-	-	-	-	-	-	Õ	0
54:01	-	-	-	-	-	-	-	-	-	0	0.02	0.01
55:01	-	-	-	-	-	-	-	-	-	0	0.02	0.01
55:02	0.12	0.09	0.22	0.19	0.07	0.12	0.15	0.14	0.36	0.26	0.18	0.21
55:13	-	-	-	-	-	-	-	-	-	_	0	0
56:01	0.02	0.02	0.08	0.04	0.15	0.04	0.08	0.07	0.01	0.02	0.01	0.01
56:02	0.14	0.14	0.11	0.05	0.18	0.11	0.11	0.11	0.06	0.04	0.02	0.03
56:04	-	-	0.01	-	0.01	0.01	0.01	0.01	-	-	-	-
57:01	0.02	0.02	-	-	-	-	0	0	-	0	0.02	0.01
58:01	-	-	-	-	-	-	-	-	-	0.01	-	0
67:01	-	-	-	-	-	-	-	-	-	-	0.01	0.01
81:01	-	-	-	-	-	-	-	-	-	-	0.02	0.01
k	17	14	20	12	12	15	23	25	15	34	57	69

Table S2 cont.

HLA-	TP	PFA	CI	TOK	TON	SA	VTP	PMD	MFA	TM	TTM	MMD
C*	(29)	(22)	(50)	(50)	(50)	(50)	(400)	(444)	(46)	(107)	(124)	(231)
01:02	0.29	0.25	0.40	0.25	0.35	0.24	0.31	0.30	0.45	0.31	0.21	0.26
01:03	-	-	-	-	-	-	-	-	-	-	0.03	0.02
02:02	-	-	-	-	-	-	-	-	-	-	0	0
03:02	-	-	-	-	-	-	-	-	-	-	0.02	0.01
03:03	0.02	0.02	-	-	0.02	0.01	0.01	0.01	-	0.02	0.01	0.02
03:04	0.10	0.11	0.03	0.01	0.10	0.24	0.10	0.10	0.01	0.01	0.01	0.01
04:01	0.03	0.05	0.06	0.30	0.10	0.11	0.14	0.13	0.10	0.06	0.07	0.07
04:03	0.16	0.16	0.07	0.16	0.10	0.06	0.10	0.10	0.09	0.07	0.11	0.09
05:01	0.03	0.02	0.01	0.01	0	0	0.01	0.01	0.01	0.04	0.04	0.04
06:02	0.02	0.02	0.01	0	0	0	0	0	-	0.03	0.02	0.02
07:01	0.03	0.02	0.01	0	0	0.02	0.01	0.01	0.02	0.04	0.07	0.05
07:02	0.09	0.09	0.15	0.07	0.08	0.10	0.10	0.10	0.11	0.16	0.11	0.14
07:04	-	-	-	-	-	-	-	-	0.01	0.01	0	0.01
07:05	-	-	-	-	-	-	-	-	-	-	0.01	0.01
08:01	0.07	0.09	0.10	0.01	0.12	0.11	0.09	0.09	0.13	0.13	0.09	0.11
08:02	-	-	0.01	-	-	-	0	0	-	0.03	0.02	0.03
12:02	-	-	0.02	0.02	0.02	-	0.02	0.01	0.01	0	0.02	0.01
12:03	0.05	0.07	0.02	0.10	0.04	0.03	0.05	0.05	0.03	0.02	0.01	0.02
14:02	-	-	0.02	-	-	-	0.01	0	-	-	0	0
15:02	0.10	0.09	0.09	0.07	0.07	0.07	0.08	0.08	0.02	0.06	0.06	0.06
15:03	-	-	-	-	-	0.01	0	0	-	-	-	-
15:04	-	-	-	-	-	-	-	-	-	-	0.02	0.01
15:05	-	-	-	-	-	-	-	-	0.01	0	-	0
16:01	-	-	-	-	-	-	-	-	-	-	0.03	0.02
16:02	-	-	-	-	-	-	-	-	-	-	0	0
18:01	-	-	-	-	-	-	-	-	-	-	0.03	0.02
k	12	12	14	10	10	11	16	16	13	16	24	25
HLA-			CI	TOK	TON	SA	VTP				TTM	
DQB1*	TP	PFA	(50)	(50)	(50)	(50)	(400)	PMD	MFA	TM	(187)	MMD
02:01	•	•	0.02	-	-	-	0.01	•	•	•	0.08	٠
02:02	•	•	-	-	-	-	-	•	•	•	0.02	•
02:03	•	•	-	-	-	-	-	٠	٠	•	0	٠
03:01	•	•	0.30	0.22	0.23	0.16	0.23	٠	٠	•	0.30	٠
03:02	•	•	0.25	0.38	0.17	0.16	0.24	٠	٠	•	0.12	٠
03:03	•	•	0.09	0.06	0.16	0.26	0.14	٠	٠	•	0.06	٠
03:04	٠	•	-	-	-	-	-	٠	٠	•	0.03	٠
04:01	٠	•	-	-	-	0.01	0	٠	٠	•	-	٠
04:02	٠	•	0.03	0.08	0.05	0.10	0.07	٠	٠	•	0.01	٠
05:01	٠	•	0.01	-	-	-	-	٠	٠	•	0.04	٠
05:02	٠	•	0.08	-	0.01	0.03	0.03	٠	٠	•	0.06	٠
05:03	٠	•	0.13	0.10	0.17	0.11	0.13	٠	٠	•	0.11	٠
06:01	٠	•	0.08	0.14	0.19	0.14	0.14	٠	٠	•	0.08	٠
06:02	٠	•	0.01	-	0.02	0.03	0.02	٠	٠	•	0.06	٠
06:03	٠	•	-	0.02	-	-	0.01	٠	٠	•	0.01	٠
06:04	•	•	-	-	-	-	-	٠	٠	•	0.02	•
06:05	•	•	-	-	-	-	-	•	•	•	0	•
k	•	٠	10	7	8	9	12	•	•	•	16	•

Table S2 cont.

HLA-	TP	PFA	CI	TOK	TON	SA	VTP	PMD	MFA	TM	TTM	MMD
DRB1*	(37)	(26)	(50)	(50)	(50)	(50)	(400)	(452)	(49)	(114)	(200)	(314)
01:01	0.03	0.02	0.01	-	-	-	0	0	0.01	0.02	0.03	0.03
01:02	-	-	-	-	-	-	-	-	-	0.01	-	0
01:03	-	-	-	-	-	-	-	-	-	0	-	0
03:01	0.01	0.02	0.01	-	-	0.02	0.01	0.01	0.02	0.04	0.04	0.04
03:07	-	-	-	-	-	-	-	-	-	-	0.01	0
04:01	0.03	-	-	-	-	-	-	-	0.03	0.05	0.05	0.05
04:03	0.14	0.15	0.26	0.38	0.17	0.16	0.24	0.23	0.11	0.08	0.11	0.10
04:04	-	-	-	-	-	-	-	-	-	0	0.01	0
04:05	0.05	0.08	0.03	0.07	0.04	0.09	0.06	0.06	0.04	0.02	0	0.01
04:07	-	-	-	0.01	-	-	0	0	-	-	-	-
04:10	-	-	-	-	0.01	-	0	0	-	-	0.01	0
07:01	-	-	0.01	-	-	0.01	0.01	0	-	0.04	0.05	0.05
08:01	-	-	-	-	0.01	-	0	0	-	0.01	0	0.01
08:02	0.01	0.02	0.02	-	-	0.02	0.01	0.01	-	0.01	-	0
08:03	0.09	0.12	0.08	0.12	0.12	0.10	0.11	0.11	0.18	0.11	0.09	0.10
09:01	0.16	0.19	0.08	0.04	0.16	0.25	0.13	0.14	0.10	0.06	0.04	0.05
11:01	0.05	0.06	0.12	-	0.16	0.08	0.09	0.09	0.18	0.17	0.14	0.15
11:02	-	-	-	-	-	-	-	-	-	-	0	0
11:03	-	-	-	-	-	-	-	-	-	-	0	0
11:04	-	-	-	-	-	-	-	-	-	-	0	0
11:11	-	-	-	-	-	-	-	-	-	-	0	0
12:01	0.15	0.13	0.16	0.22	0.08	0.08	0.14	0.13	0.24	0.21	0.22	0.21
12:02	-	-	0.01	-	-	-	0	0	-	-	0.01	0
13:01	0.03	0.02	-	0.03	-	-	0.01	0.01	-	0.02	0.01	0.01
13:02	0.01	-	-	-	-	-	-	-	-	0.01	0.01	0.01
13:04	0.01	-	-	-	-	-	-	-	-	-	-	-
13:05	-	-	-	-	-	-	-	-	-	-	0.01	0
13:10	-	-	-	-	-	-	-	-	-	-	0	0
13:12	-	-	0.01	-	-	-	0	0	-	-	-	-
14:01	-	-	0.07	-	0.05	0.01	0.03	0.03	0.01	0	0.03	0.02
14:04	-	-	-	-	-	-	-	-	-	0	-	0
14:05	0.01	-	-	-	-	-	-	-	-	0	-	0
14:07	-	-	-	-	-	-	-	-	-	-	0	0
14:08	0.05	0.04	0.07	0.07	0.07	0.08	0.07	0.07	0.02	0.04	0.06	0.05
14:23	-	-	-	-	-	-	-	-	-	-	0	0
14:54	0.05	0.08	-	-	-	-	0	0.01	0.02	0.04	-	0.01
15:01	0.05	0.04	0.01	-	0.03	0.03	0.02	0.02	0.02	0.04	0.05	0.05
15:02	0.04	0.04	0.06	0.06	0.10	0.07	0.07	0.07	-	-	0.01	0.01
16:01	-	-	-	-	-	-	-	-	-	-	0	0
16:02	-	-	-	-	-	-	-	-	-	-	0	0
k	18	14	16	9	12	13	20	21	13	23	31	37

Table S2 cont.

HLA-	TP	PFA	CI	TOK	TON	SA	VTP	PMD	MFA	ТМ		
DPA1*	(25)	(17)	(N=50)	(50)	(50)	(50)	(400)	(434)	(42)	(114)	TTM	MMD
01:03	0.48	0.47	0.54	0.73	0.46	0.42	0.54	0.53	0.32	0.47	•	•
01:05	0.02	-	-	-	-	-	-	-	-	-	•	•
02:01	0.02	-	-	0.01	-	0.01	0.01	0	0.02	0.03	•	•
02:01/												
02	0.02	0.03	-	-	-	-	-	0	-	0.01	•	•
02:02	0.46	0.50	0.44	0.26	0.53	0.57	0.45	0.45	0.65	0.49	•	•
04:01	-	-	0.02	-	0.01	0	0.01	0.01	-	-	•	•
k	4	3	3	3	3	3	4	4	3	3	•	•
HLA-	TP	PFA		TOK	TON	SA	VTP	PMD	MFA	TM	TTM	MMD
DPB1*	(25)	(17)	CI (50)	(50)	(50)	(50)	(400)	(434)	(42)	(91)	(199)	(290)
01:01	-	-	-	-	-	-	-	-	0.01	0.02	0.01	0.01
02:01	0.16	0.18	0.26	0.32	0.24	0.19	0.25	0.25	0.11	0.17	0.16	0.16
02:02	-	-	-	-	-	-	-	-	0.01	0.02	0.01	0.01
03:01	-	_	0.02	0.01	0.01	_	0.01	0.01	_	_	0.02	0.01
04:01	0.08	0.06	0.23	0.39	0.19	0.21	0.26	0.24	0.02	0.04	0.29	0.21
04:01/	0.20	0.21	-	-	-	_	-	0.02	0.18	0.22	-	0.07
134:01		•										
04:02	0.06	0.03	0.02	0.01	0.01	0.02	0.02	0.02	_	0.01	0.06	0.04
05.01	0.00	0.05	0.44	0.01	0.52	0.56	0.45	0.02	0.65	0.48	0.00	0.43
06.01	-	-	-	-	-	-	-	-	-	-	0.10	0
08.01	_	_	_	_	_	_	-	-	_	-	0	Ő
09.01	_	_	_	_	_	_	-	-	0.01	0.01	0	Ő
10.01	0.02	0.03	_	0.01	_	_	0	0	-	-	0	Õ
11.01	-	-	_	-	_	_	-	-	_	-	0.01	0.01
13.01	_	_	0.01	_	_	_	0	0	_	_	0.01	0.01
13.01/	_	_	-	_	_	_	-	-	_	0.01	-	Ő
107.01										0.01		Ū
14:01	0.02	0.03	-	-	0.01	0.01	0.01	0.01	-	-	-	_
15:01	-	_	-	-	_	-	-	_	-	0.01	-	0
15:01/	-	_	-	-	-	_	-	-	-	0.01	_	0
74:01												,
16:01	-	-	-	-	-	-	-	-	-	-	0.01	0.01
17:01	-	-	-	-	-	-	-	-	-	0.01	0.01	0.01
20:01	-	_	0.01	-	-	_	0	0	-	0.01	_	0
21:01	-	_	-	-	-	_	-	-	-	-	0	Ő
23:01	-	_	_	_	_	_	-	-	_	-	0.02	0.01
25:01	-	_	-	-	-	-	-	-	-	_	0	0
28:01	-	_	0.01	-	-	-	0	0	-	_	-	-
29:01	-	_	-	-	-	0.01	Ő	Õ	-	-	-	-
31:01	-	-	-	-	0.02	-	0.01	Õ	-	-	-	-
35:01	-	-	-	-	-	-	-	-	-	0.01	-	0
65:01	-	_	-	-	-	_	-	-	-	-	0	Ő
k	7	7	8	6	7	6	12	13	7	14	19	25

Table S	S2 cont.
---------	----------

	TP	PFA	Cl§	TOK [§]	TON [§]	SA [§]	VTP [§]		MFA	TM		
MICA*	(25)	(17)	(43)	(36)	(38)	(41)	(158)	PMD	(42)	(91)	TTM	MMD
001	-	-	-	-	-	-	-	•	-	0.02	•	٠
002	-	-	-	-	-	-	-	•	-	0.04	•	•
004	-	-	-	-	-	0.01	0.01	•	-	-	•	•
006	-	-	0.01	-	-	-	0.01	•	-	-	•	•
007	0.02	0.03	-	-	0.01	0.01	0.01	٠	-	0.01	•	٠
008	0.44	0.56	0.20	0.38	0.18	0.34	0.27	٠	0.27	0.34	•	٠
009	-	-	-	-	-	-	-	٠	0.02	0.02	•	٠
010	0.02	-	-	-	-	-	-	٠	0.02	0.04	•	٠
011	-	-	-	-	-	-	-	•	-	0.02	•	•
012	0.36	0.21	0.38	0.23	0.34	0.22	0.30	٠	0.56	0.43	•	٠
017	-	-	-	-	-	-	-	•	-	0.01	•	•
019	0.02	0.03	0.01	0.08	0.07	0.04	0.04	•	-	0.01	•	•
027	-	-	0.07	0.10	0.07	0.06	0.07	•	-	-	•	•
033	-	-	0.02	0.01	0.01	-	0.01	٠	0.02	0.01	•	٠
045	0.08	0.09	0.03	0.03	0.03	0.05	0.03	٠	0.06	0.03	•	٠
del	0.06	0.09	-	-	-	-	-	•	0.04	0.02	•	•
blank	-	-	0.29	0.17	0.30	0.29	0.25	•	-	-	•	٠
k	6	5	7	6	6	7	9	•	6	12	•	٠

HLA data for CI, TOK, TON and SA were obtained from [8,9,35 and Velickovic, unpublished] and from [36,37] for TTM. We considered creating PMD by adjusting allele frequency values in VTP assuming *MICA*del* = 0.09, as in PFA. We decided not to do so due to high blank values in various Polynesian sub-populations that make up VTP. Some of these might be authentic *MICA-del* homozygotes. Abbreviations: number in bracket = total number of individuals, k = total number of observed alleles, 0 = < 0.01, - = not observed), TP: Total Polynesians, PFA: Polynesians with Full Ancestry, TM= Total Maori, MFA = Maori with Full Ancestry, CI: Cook Islands, TOK: Tokelau, TON: Tonga, SA: Samoa, VTP: Velickovic Total Polynesians (CI+TOK+TON+SA), PMD: Polynesian Meta Dataset (PFA+VTP), TTM:Tracey Total Maori, MMD: Maori Meta Dataset (TM+TTM), • = not reported and § = MICA allele frequencies were determined by direct counting (counting once for homozygotes) and blank might include MICA-del.

Table S3:	Exact	tests	of	population	differentiation	(<i>p</i> -values)	on	Polynesian	and	Maori
datasets										

HLA-A										
	TP	PFA	ТМ	MFA	CI	SA	TOK	TON	VTP	PMD
TP	~	~	~	~	~	~	~	~	~	~
PFA	1.00	~	~	~	~	~	~	~	~	~
TM	0.19	0.30	~	~	~	~	~	~	~	~
MFA	0.00	0.01	0.58	~	~	~	~	~	~	~
CI	0.16	0.23	0.00	0.02	~	~	~	~	~	~
SA	0.07	0.03	0.00	0.00	0.00	~	~	~	~	~
TOK	0.01	0.02	0.00	0.02	0.84	0.00	~	~	~	~
TON	0.08	0.25	0.00	0.00	0.12	0.00	0.17	~	~	~
VTP	0.15	0.17	0.00	0.00	0.21	0.13	0.54	0.48	~	~
PMD	0.34	0.38	0.00	0.00	0.32	0.13	0.47	0.56	1.00	~
TTM	0.32	0.38	0.36	0.14	0.00	0.00	0.00	0.00	0.00	0.00
]	HLA-B					
	TP	PFA	TM	MFA	CI	SA	TOK	TON	VTP	PMD
TP	~	~	~	~	~	~	~	~	~	~
PFA	1.00	~	~	~	~	~	~	~	~	~
	0.07	0.06	~	~	~	~	~	~	~	~
MFA	0.07	0.01	0.94	~	~	~	~	~	~	~
CI SA	0.10	0.22	0.01	0.04	~	~	~	~	~	~
SA TOV	0.00	0.29	0.00	0.00	0.00	~	~	~	~	~
TON	0.01	0.01	0.00	0.00	0.00	0.00	~	~	~	~
VTD	0.04	0.10	0.00	0.00	0.01	0.00	0.00	$\tilde{0.25}$	~	~
	0.02	0.62	0.00	0.00	0.02	0.20	0.15	0.23	~	~
TTM	0.09	0.02	0.00	0.00	0.02	0.24	0.12	0.02	0.00	0.00
	0.00	0.00	0.00	0.00	HLA-C	0.00	0.00	0.00	0.00	0.00
	TP	PFA	ТМ	MFA	CI	SA	TOK	TON	VTP	PMD
TP	~	~	~	~	~	~	~	~	~	~
PFA	1.00	~	~	~	~	~	~	~	~	~
TM	0.08	0.13	~	~	~	~	~	~	~	~
MFA	0.01	0.03	0.20	~	~	~	~	~	~	~
CI	0.18	0.20	0.31	0.54	~	~	~	~	~	~
SA	0.07	0.26	0.00	0.00	0.00	~	~	~	~	~
TOK	0.00	0.00	0.00	0.00	0.00	0.00	~	~	~	~
TON	0.30	0.49	0.00	0.11	0.26	0.19	0.00	~	~	~
VTP	0.09	0.26	0.00	0.01	0.06	0.07	0.00	0.97	~	~
PMD	0.20	0.43	0.00	0.01	0.06	0.09	0.00	0.98	1.00	~
TTM	0.09	0.16	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
HLA-DQB1										
	TP	PFA	ТМ	MFA	CI	SA	TOK	TON	VTP	~
TP	~	~	~	~	~	~	~	~	~	~
PFA	~	~	~	~	~	~	~	~	~	~
TM	~	~	~	~	~	~	~	~	~	~
MFA	~	~	~	~	~	~	~	~	~	~
CI	~	~	~	~	~	~	~	~	~	~
SA	~	~	~	~	0.00	~	~	~	~	~
TOK	~	~	~	~	0.00	0.00	~	~	~	~
TOM			~	~	0.00	0.28	0.01	~	~	~
TON	~				0.00					
VTP	~	~	~	~	0.00	0.11	0.04	0.75	~	~

Table	S 3	cont.
-------	------------	-------

HLA-DRB1										
	TP	PFA	TM	MFA	CI	SA	TOK	TON	VTP	PMD
TP	~	~	~	~	~	~	~	~	~	~
PFA	1.00	~	~	~	~	~	~	~	~	~
ТМ	0.01	0.01	~	~	~	~	~	~	~	~
MFA	0.02	0.03	0.52	~	~	~	~	~	~	~
CI	0.16	0.44	0.00	0.00	~	~	~	~	~	~
SA	0.22	0.59	0.00	0.00	0.01	~	~	~	~	~
TOK	0.00	0.00	0.00	0.00	0.00	0.00	~	~	~	~
TON	0.10	0.33	0.00	0.00	0.24	0.26	0.00	~	~	~
VTP	0.01	0.43	0.00	0.00	0.63	0.28	0.00	0.52	~	~
PMD	0.02	0.68	0.00	0.00	0.62	0.49	0.00	0.59	1.00	~
TTM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HLA-DPA1										
	ТР	PFA	TM	MFA	CI	SA	TOK	TON	VTP	PMD
TP	~	~	~	~	~	~	~	~	~	~
PFA	1.00	~	~	~	~	~	~	~	~	~
TM	0.25	1.00	~	~	~	~	~	~	~	~
MFA	0.05	0.14	0.03	~	~	~	~	~	~	~
CI	0.38	0.69	0.05	0.00	~	~	~	~	~	~
SA	0.28	0.69	0.54	0.22	0.06	~	~	~	~	~
TOK	0.00	0.01	0.00	0.00	0.01	0.00	~	~	~	~
TON	0.61	1.00	0.52	0.06	0.45	0.57	0.00	~	~	~
VTP	0.15	0.59	0.10	0.00	0.45	0.07	0.00	0.32	~	~
PMD	0.14	0.59	0.13	0.00	0.36	0.07	0.00	0.34	0.97	~
				Η	LA-DPB	1				
	TP	PFA	TM	MFA	CI	SA	TOK	TON	VTP	PMD
TP	~	~	~	~	~	~	~	~	~	~
PFA	0.98	~	~	~	~	~	~	~	~	~
TM	0.56	0.51	~	~	~	~	~	~	~	~
MFA	0.08	0.16	0.38	~	~	~	~	~	~	~
CI	0.34	0.46	0.38	0.01	~	~	~	~	~	~
SA	0.35	0.42	0.63	0.33	0.20	~	~	~	~	~
TOK	0.02	0.05	0.00	0.00	0.02	0.00	~	~	~	~
TON	0.22	0.41	0.23	0.06	0.54	0.66	0.00	~	~	~
VTP	0.43	0.37	0.01	0.01	0.66	0.52	0.03	0.60	~	~
PMD	0.52	0.47	0.01	0.01	0.60	0.56	0.04	0.64	1.00	~
TTM	0.70	0.58	0.05	0.08	0.23	0.19	0.09	0.05	0.00	0.00
MICA										
	ТР	PFA	TM	~	~	~	~	~	~	~
TP	~	~	~	~	~	~	~	~	~	~
PFA	0.84	~	~	~	~	~	~	~	~	~
TM	0.61	0.11	~	~	~	~	~	~	~	~
MFA	0.33	0.01	0.62	~	~	~	~	~	~	~

Abbreviations: TP: Total Polynesians, PFA: Polynesians with Full Ancestry, TM= Total Maori, MFA = Maori with Full Ancestry, CI: Cook Islands [8,9,35], SA: Samoa [8,9,35], TOK: Tokelau [8,9,35], TON: Tonga [8,9,35], VTP: Velickovic Total Polynesians (CI+SA+TOK+TON), PMD: Polynesian Meta Dataset (PFA+VTP), TTM: Tracey Total Maori [36,37], MMD: Maori Meta Dataset (TM+TTM), p: significant level (<0.05).

Molecular approaches to transfusion medicine in Polynesians and Maori in

New Zealand

H.A. Edinur^{*}, P.P.J. Dunn[§], R.A. Lea^{\dagger} and G.K. Chambers^{*}

*School of Biological Sciences, Victoria University of Wellington, New Zealand *New Zealand Blood Service, Auckland

[†]School of Medical Sciences, Griffith University, Australia

Status: In press

Journal: International Journal of Immunogenetics

Correspondence: Dr Geoffrey K. Chambers, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand. Ph: +64-(0)4-463-6091, Fax: +64-(0)4-463-5331, e-mail: Geoff.Chambers@vuw.ac.nz

Chapter 3.3

Summary

In recent years with the application of genotyping technology, there has been a substantial increase in the number of reported blood group alleles. The present survey was designed to evaluate new molecular blood group genotyping methods and compile reference blood group datasets for Polynesian and Maori subjects. Subsequent analyses of these results were used to calculate probability of random match, to trace Polynesian ancestry and migration patterns and to reveal past and present episodes of genetic admixture. Genomic DNA samples from Maori and Polynesian subjects were drawn from the Victoria University of Wellington DNA Bank and genotyped using combination of commercial PCR-SSP kits, hybridization SNP assay services or sequence based typing. The present survey also involves compilation of serological ABO and Rhesus blood group data from RakaiPaaka Iwi Tribal members for comparison with those generated during our molecular blood group study. We observed perfect consistency between results obtained from all molecular methods for blood group genotyping. The A, O, DCcEe, DCCee, MNs, K-k+, Jk(a+b-), Jk(a+b+), Fy(a+b-), Fy(a+b+), Di(a+b-), Co(a+b-) and Do(a-b+) were predominant blood group phenotypes in both Polynesians and Maori. Overall, our survey data show only small differences in distributions of blood group phenotypes between Polynesian and Maori groups and their sub-groups. These differences might be associated with selection, population history and gene flow from Europeans. In each case, we estimate that patients with certain blood groups have a very low probability of an exact phenotypic match, even if the patients were randomly transfused with blood from donors of their own ethnicity. The best way to avoid haemolytic transfusion reaction in such cases is to perform a pre-transfusion cross-match and recruit increased numbers of donors with rare phenotype profiles. The conclusion of this study is that application of molecular method covering as many known variants as possible may help to
improve the accuracy blood group genotyping and potentially conserve the routine requirements of transfusion centres.

Introduction

Blood group antigens are polymorphic genetic markers located on the surface of red blood cells. Currently, substantial genetic variation, amounting to 1251 alleles belonging to 30 blood group systems, has been documented (see Daniels and Reid, 2010; Patnaik *et al.*, 2012). Following the elucidation of the molecular basis underpinning the structure of blood group antigens, a wide range of molecular techniques have been developed. These range from simple polymerase chain reaction (PCR) techniques (e.g. PCR-fragment length polymorphism and sequence specific primer) to the rapid high-throughput SNP arrays (BeadChip and BLOODchip) genotyping platforms; see also Wu and Csako, (2006); Avent (2007); Hashmi *et al.* (2007); Veldhuisen *et al.* (2009) for reviews on blood group typing.

It is already well established that ABO, Rhesus and Kell blood group antigens are clinically important in transfusion compatibility and in certain cases are valuable as prenatal diagnostics in the management of hemolytic disease of the fetus and newborn (HDFN). As well as ABO, Rhesus and Kell, there are other blood group antigens which can stimulate clinically significant antibodies and these include Dombrock, Kidd and MNS (Anstee, 2009). New Zealand comprises a unique population mixture and there are no recent blood group data available for any of its sub-ethnic groups. This presents challenges and opportunities for molecular blood group screening. The present survey of blood groups is a part of wider screening exercise (i.e. blood group and HPA loci; see Edinur *et al.*, 2013 for details) and was designed to assess application of new molecular blood group typing methods in Polynesian and Maori populations. This is because most of the molecular blood group platforms which have been tested have been used with subjects (e.g. Europeans and Africans), who might have differing 'molecular' blood groups profiles compared with Polynesians. Blood group

data generated from this survey can also be used for dual analysis of ancestry and health. The data presented here will prove useful in the supply of 'safe blood' to multi- or recently-transfused patients, or to patients with pre-existing autoantibodies. Overall, the blood group data presented here complement our recent studies of HLA and HPA genes, which are medically important genetic markers and reflect ancestral history (Edinur *et al.*, 2012; Edinur *et al.*, 2013).

Materials and methods

Serological data

The present survey involves compilation of serological ABO and Rhesus blood group data for comparison with our new DNA-based blood group study in Maori and Polynesian subjects. The serological data comprise information for 269 Te o Iwi RakaiPaaka members (Total RakaiPaaka: TR), one of Maori tribe living in New Zealand (NZ). These data were generated during the Te Iwi o Rakaipaaka Health and Ancestry Study (RHAS), under permit from RHAS Management Team, NZ Multi-Regional Ethics Committee and Victoria University of Wellington (VUW) Human Ethics Committee (refer Eccles, 2011 for details). Data from those un-admixed individuals (i.e. those with four grandparents and no family ancestral knowledge of members from other ethnic groups) in this collection were then grouped into a smaller sub-group; RakaiPaaka with Full Ancestry (RFA).

Genomic DNA

The molecular blood group study involved genomic DNA typing of Maori (Total Maori: TM) and Polynesian with Full Ancestry (PFA) subjects, drawn from the VUW DNA Bank. We have selected a Full Ancestry (MFA) sub-group from the TM group containing for those with four grandparents and no familial knowledge ancestors from other ethnic groups by means of family history interview (see Box 1A). All genetic materials used in the present survey were

161

obtained with informed consent and full ethical approval (New Zealand Central Region Ethics Committee and VUW Human Ethics Committee).

Blood group genotyping

The blood group genotyping strategy for the present survey is shown in Fig. 1. All subjects were genotyped using either commercial PCR-SSP kits (BAGene DNA-SSP ABO-TYPE variant, RH-TYPE, MNS-TYPE and KKD-TYPE kits; BAG Health Care GmbH, Lich, Germany) or by hybridization SNP assay (BLOODchip[®] v2.0; Progenika Biopharma S.A., Spain) services (see Table S1 for details of variants covered by each genotyping platform). Nonetheless, we had to perform sequence based typing (SBT) for JK_{null} (i.e. g>a mutation at the 3' acceptor splice site of intron 5; Wester *et al.*, 2008) which is not covered by the commercial PCR-SSP kits (see Table S2 for details of PCR and SBT protocols). A total of 10 samples were genotyped using both, PCR-SSP and BLOODchip, as cross-validation of the molecular methods used in this study. As a further quality assurance measure, an additional 30 samples were randomly selected for Kell (578T>C for *Kell*A* and *Kell*B* alleles, respectively) and Duffy (125G>A for *FY*A* and *FY*B* alleles, respectively) genotyping (see Table S2 for practical details and references) using SBT.

Statistical analysis

Phenotype frequencies were estimated by direct counting while the probability of finding a phenotype match between randomly selected donors and recipients was calculated according to Badjie *et al.* (2011). Exact tests were performed using Arlequin, version 3.1 (Excoffier *et al.*, 2005). Neighbor-joining trees based on Nei's genetic distance were constructed using the PHYLIP software package, version 3.6 (Felsenstein, 2004) and visualized with the TreeView program (Page, 1996). Principal coordinate (PCO) analysis was performed using Multi-Variate Statistical Package, version 3 (MVSP3).

Box 1A: Note on study groups and sub-groups.

We provide additional information about the identity and ancestry of our study groups and sub-groups (see Edinur *et al.*, 2012 for further details).

Population	Description
TR	Member of Maori tribal group Te Iwi O RakaiPaaka. Majority are resident of the village of Nuhaka, New Zealand.
TM	Maori volunteers of mixed tribes from the Wellington region in New Zealand and whose genetic material form part of the VUW DNA Bank.
PFA	Pacific Islanders from different geographical sources (e.g. Tonga, Samoa and Cook Island) living in New Zealand. Only those with four grandparents and no family knowledge of members from other ethnic groups were included.
RFA	RFA is a sub-set of TR. The RFA sub-group consists of those with four grandparents and no family ancestral knowledge of members from other ethnic groups.
MFA	MFA is a sub-set of TM. The MFA sub-group consists of those with four grandparents and no family ancestral knowledge of members from other ethnic groups.



Fig. 1: Schematic diagram of blood group genotyping strategy. Abbreviations are given in the Materials and methods section. Pooled data are shown in Table 2, except for HPA where have been reported elsewhere (Edinur *et al.*, 2013).

Results

Serologically determined blood group antigen frequencies collected from the RakaiPaaka Iwi are given in Table 1 and DNA-based blood group phenotype frequencies for VUW DNA Bank volunteers are listed in Table 2. Included in Table 2 are reference European (Euro) populations obtained from Grunbaum et al. (1980) and Hashmi et al. (2007). The present survey used two different molecular methods for predicting blood group phenotype frequency. To ensure that we have covered all commonly reported variants that affect blood group antigen expression (see Table S1) and may be present in Polynesians, we also performed SBT for the Kidd g>a mutation at the 3' acceptor splice site of intron 5 (IVS5-1g>a) (Irshaid et al., 2000), which is not covered by the commercial PCR-SSP kits. In addition, we note that several less well known blood group systems such as Diego, Dombrock, Colton, Kp^a, Kp^b and Kp^c and Js^a and Js^b, K_{mod}-1 and Gp.Mur can only be tested using BLOODchip (see Table S1). These data are not included in Table 2 because almost all loci tested across 50 samples were found to be fixed; i.e. 100% are Di(a-b+), Co(a+b-), Kp^b and Js^b. No samples were found with K_{mod}-1 and Gp.Mur phenotypes. Only the Dombrock antigens were polymorphic and phenotype frequencies are given in Table 2. Comparison of PCR-SSP and BLOODchip methods using 10 samples gave 100% agreement. Our further validation using SBT on the 30 randomly selected individuals for Kell and Duffy blood groups showed results fully consistent with those obtained from PCR-SSP and BLOODchip (see Fig. 2 for examples). In addition, Polynesian and Maori subjects were observed to have most common type variants at other SNPs in the region flanked by the sequencing primers (e.g 577A for Kell and 293G for Duffy blood groups; see Patnaik et al., 2012).

The newly observed ABO and Rhesus blood group data for TM and MFA were compared to those collected from the RakaiPaaka Iwi study. Exact tests show no significant differences between Maori sub-groups except for TM and RFA, for both, ABO and Rhesus (see Table S3). We judge this might be due to admixture in TM (see later). Next, we compared blood group data from MFA with Polynesians (i.e. PFA). Visual examination shows surprisingly high frequencies of ABO B and Rhesus DCCee phenotype frequencies in PFA, compared with MFA (see Table 2). Overall, statistical tests show that the MFA data are not generally different from PFA, except for the ABO, Rhesus and Kidd blood groups (refer Table S4). Blood group frequencies from Maori and Polynesian datasets were compared with reference European (Euro) populations. Exact tests (Table S4) show significant differences between Maori/Polynesian and European datasets particularly for Rhesus, MNS and Duffy blood groups.

The blood group data listed in Table 2 were also used to estimate the probability of match between randomly selected donors and recipients. The Diego and Colton blood groups are monomorphic and therefore have 100% matches between all datasets. For the polymorphic loci, only the Kell blood type was observed to have a high probability of match between a randomly selected recipient and donor of similar ethnicity. In every case, the probability values also suggested a lower chance of an exact match for Maori or Polynesians with blood sourced from European donors, compared with donor and recipient pairs of similar ethnicity.

Finally, a phylogenetic tree (Fig. 3) and PCO plot (Fig. 4) were constructed using ABO, Rhesus and MNS phenotype frequencies and used to examine genetic relationships between Polynesian populations and other populations including their putative A-SM and P-SA populations (see Box 1B for detailed description of these populations). In the phylogenetic tree, Maori datasets form a distinct branch while PFA and Tongan appear to be closely related and grouped together with A-SM population, including Filipinos - similar to that captured by the PCO plot. Our admixed TM group and previous study on Maori (i.e. Maori NZ) are plotted closer to Mexico and European populations in the top hand corner of Figure 4.

Blood group	TR (N=269)	RFA (N=126)
ABO		
А	0.54	0.54
В	0.05	0.06
AB	0.04	0.02
0	0.38	0.38
Rhesus		
Rh+	0.97	0.99
Rh-	0.03	0.01

Table 1: ABO and Rhesus phenotype frequencies from the RakaiPaaka Iwi Study

Abbreviations: TR: Total RakaiPaaka, RFA: RakaiPaaka with Full Ancestry, N = sample size

Table 2: Blood group phenotype frequencies and probability of match for European, Maori and Polynesian datasets

ABO	Euro (N=914)	TM (N=65)	MFA (N=45)	PFA (N=25)
А	0.41	0.48	0.53	0.36
В	0.09	0.00	0.00	0.24
AB	0.05	0.05	0.02	0.00
0	0.45	0.48	0.44	0.40
PM [⊙]	0.38	0.46	0.48	0.35
PM^{\P}		0.41	0.42	0.35
Rhesus*	Euro (N=914)	TM (N=65)	MFA (N=45)	PFA (N=24)
DCCEe	0.00	0.00	0.00	0.00
DCCee	0.17	0.18	0.16	0.58
DCcEE	0.00	0.00	0.00	0.00
DCcEe	0.13	0.42	0.47	0.21
DccEE	0.01	0.22	0.27	0.00
Dccee	0.04	0.00	0.00	0.00
DCcee	0.32	0.12	0.11	0.13
DccEe	0.12	0.00	0.00	0.04
dccee	0.18	0.05	0.00	0.00
dCcee	0.01	0.02	0.00	0.00
dccEE	0.00	0.00	0.00	0.00
dccEe	0.01	0.00	0.00	0.04
PM [☉]	0.20	0.27	0.33	0.40
PM^{\P}		0.14	0.13	0.17
MNS	Euro (N=751)	TM (N=65)	MFA (N=45)	PFA (N=25)
MNSs	0.21	0.09	0.09	0.08
MNS	0.04	0.00	0.00	0.00
MNs	0.24	0.45	0.47	0.44
MSs	0.15	0.05	0.04	0.04
MS	0.07	0.00	0.00	0.00
Ms	0.11	0.18	0.18	0.32
NSs	0.03	0.06	0.04	0.00
NS	0.01	0.03	0.04	0.00
Ns	0.16	0.14	0.13	0.12
PM [⊙]	0.16	0.27	0.28	0.32
PM^{\P}		0.18	0.18	0.18
Kell	Euro (N=911)	TM (N=74)	MFA (N=45)	PFA (N=25)
K+k-	0.00	0.00	0.00	0.00
K+k+	0.09	0.04	0.04	0.04
K-k+	0.90	0.96	0.96	0.96
PM [☉]	0.83	0.95	0.96	0.92
PM^{\P}		0.88	0.89	0.87
Kidd	Euro (N=514)	TM (N=74)	MFA (N=45)	PFA (N=25)
Jk (a+b-)	0.28	0.62	0.71	0.44
Jk (a-b+)	0.25	0.09	0.02	0.12
Jk (a+b+)	0.47	0.28	0.27	0.44
PM [☉]	0.36	0.47	0.58	0.40
PM [¶]		0.33	0.33	0.36
				-

Duffy	Euro (N=913)	TM (N=74)	MFA (N=45)	PFA (N=25)
Fy(a+b-)	0.22	0.41	0.40	0.64
Fy(a-b+)	0.32	0.16	0.16	0.00
Fy(a+b+)	0.45	0.43	0.44	0.36
Fy(a-b-)	0.01	0.00	0.00	0.00
PM [☉]	0.35	0.38	0.38	0.54
PM^{\P}		0.33	0.34	0.30
Dombrock	Euro (N=1243)	TM (N=40)	MFA (N=20)	PFA (N=10)
Do(a+b-)	0.13	0.08	0.00	0.10
Do(a-b+)	0.51	0.75	0.85	0.60
Do(a+b+)	0.36	0.18	0.15	0.30
PM [⊙]	0.41	0.60	0.75	0.46
PM [¶]		0.46	0.49	0.43

Table 2, cont.

Abbreviations: Euro = Europeans, TM =Total Maori, MFA= Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, N = sample size, PM° = probability of match between donor and recipient of similar ethnic background, PM^{\parallel} = probability of math between European donor and Maori/Polynesian datasets, * = individual with weak D is considered as donor and recipient D positive while individual with partial D is considered as donor D positive and recipient negative. Reference data for Europeans obtained from Grunbaum *et al.*, 1980 (ABO, Rhesus, MNS, Kell, Kidd and Duffy) and Hashmi *et al.*, 2007 (Dombrock). Abbreviations: TM =Total Maori, MFA= Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, > = donor to receipient.



Fig. 2: Comparative screening of Kell, Kidd and Duffy blood group polymorphisms using combination of PCR-SSP and SBT.



Fig. 3: NJ tree constructed using ABO, Rhesus and MNS phenotype frequencies. Reference populations for A-SM (blue), P-SA (red), Polynesians (green) and others (black) were obtained from Gajdusek *et al.*, 1978; Sanger *et al.*, 1951; Lai and Bloom, 1982; Grunbaum *et al.*, 1980; Stavely and Douglas, 1959; Woodfield *et al.*, 1987; Windhof and Walter, 1983; Nakajima *et al.*, 1971; Chen *et al.*, 1985.



Fig. 4: Position of Maori and Polynesian datasets in the PCO plots compared to those probed in Fig. 3.

Box 1B: Note on nomenclature of Asian and Pacific peoples.

We have adopted the system below to clearly distinguish modern and ancestral populations. This is necessary to give a clear description of admixed descendants such as Melanesians whose genetic make-up may not correspond to observed cultural traits.

Population	Description
Taiwan aboriginal and other Asian including Filipino, Malay and Indonesian	Austronesian-speaking Mongoloid (A-SM)
Australian Aborigine and Papuan interior	Papuan-speaking Australoid (P-SA)
Coastal PNG plus islands and other Melanesians	~50:50 (A-SM: P-SA) ^{ab}
Polynesian including Maori	70:30 to 87:13 (A-SM: P-SA) ^{ac}

^aWollstein *et al.* (2010), ^bFriedlaender *et al.* (2008), ^cKimura *et al.* (2008)

Discussion

From the discovery of the ABO blood groups in the early 20th century, phenotyping remains as a fundamental element in transfusion medicine. Currently, molecular methods have begun to emerge as a new technique for blood typing. The molecular bases of many blood groups have been characterized, including those associated with null variants – see Patnaik *et al.* (2012) for details. As a consequence, many molecular methods have since been designed for blood group typing, with most of them aiming at multiple blood group loci (see Montpetit *et al.*, 2006; Hashmi *et al.*, 2007; Karpasitou *et al.*, 2008; Avent *et al.*, 2009).

Previously, we have demonstrated the consistency of molecular diagnostics (i.e. SNP array and PCR-SSP) for human platelet loci (Edinur *et al.*, 2013). Here, we further characterized the same sets of samples in the larger study of blood group polymorphism using a combination of commercial PCR-SSP kits, BLOODchip and SBT (see Figs. 1 and 2), plus comparison with serological references. Throughout this study, we have shown the reliability

of these three molecular methods (i.e. PCR-SSP, SBT and SNP arrays) for blood group genotyping and thus our work reflects findings from other studies (Beiboer et al., 2005; Hashmi et al., 2007; Kappler-Gratias et al., 2011; Jungbauer et al., 2012). Generally, antigen expression can be predicted by genotyping and any discrepancy between serology and molecular typing is usually associated with new mutations, which may be cryptic to antibody probes. Therefore, the capability of molecular methods to show actual serological expression is only dependent on the number of known variants included in the genotyping platform. For example, the JK*IVS5-1g>a SNP is not available in the commercial PCR-SSP kit used in the present survey, but is a common JK_{null} [Jk(a-b-)] variant in A-SM and Polynesian populations (see Irshaid et al., 2000; Lin and Yu, 2008 and below for values). Subsequently, the genotype profile obtained from this kit alone will lead to false positive prediction of Jk(a-b+) or Jk(a+b+) phenotypes (see Table S5). Comparison of phenotype predictions based on PCR-SSP alone and in combination with SBT (JK*IVS5-1g>a) shows 19% discordant results. This concern is made evident by the high frequency of JK*IVS5-1g>a observed in this study (0.09, 0.11 and 0.14 in TM, MFA and PFA, respectively). Several other molecular bases of the JK_{null} variant have also been reported: single nucleotide polymorphism (202C>T, 222C>A, 582C>G, 871T>C, 896G>A, 956C>T), 723delA, deletion of intron 3 to intron 5 and 5'-donor splice site of intron 7 (Lucien et al., 1998; Irshaid et al., 2000; Wester et al., 2008; Liu et al., 2009). These variants were not tested in the present survey and are uncommon in A-SM and Polynesians (see Irshaid et al., 2000; Lin and Yu, 2008). However, we cannot conclusively exclude their occurrence in Polynesians. Other blood group variants such as Di(a+b-) and Gp.Mur are not included in PCR-SSP kits, but are available via BLOODchip service (see Table S1). These variants were not observed in the present survey, but were reported for their putative ancestral populations (see Daniels, 2008; Lin and Broadberry, 1998). We cannot exclude the possibility of their existence in other Polynesians,

and future studies using larger sample sizes could yet reveal these phenotypes. Therefore, we feel that inclusion of as many known SNPs as possible related to serological expression will help to improve accuracy of molecular blood group genotyping. Nonetheless, molecular method is already superior to serology in many respects and circumstances; e.g. fetal blood group diagnosis, determination of blood group types in multi-transfused patients and resolution of ABO and RHD serological discrepancies (Hustinx *et al.*, 2009).

The obvious barriers for application of blood group genotyping are operational and related to cost and turnaround time (Hustinx et al., 2009; Storry and Olsson, 2009; Schonitzer, 2009; Wagner, 2009 and Table S6 for parameters and comments). Routine serological matching is only required for ABO and RhD and may in certain cases be extended to RhCE and Kell antigens (Wagner, 2009; Hustinx et al., 2009; Storry and Olsson, 2009). This may be one reason why the traditional serological approach in transfusion medicine appears to be cheaper than molecular genotyping. The question of costs remains an issue that merits further exploration. The overall costs of tests are given in Table S6 for the panels detailed in Table S1. In short, test for six blood groups plus six HPA loci by PCR-SSP kits is much cheaper (c. €128 than testing nine blood groups and twelve HPA loci via BLOODchip (c. €350). However, these figures do not take proper account of cost per test, technician time and clinical priority. There are several clinical applications where RBC genotyping has advantages over agglutination tests for the provision of safe blood. These include: (a) patients who are recently transfused, (b) patients who have a positive direct antiglobulin test because of antibodies bound to RBCs: such as in haemolytic anemia, (c) presence of autoantibodies hinders serological RBC typing but not genotyping, or (d) patient with a pre-existing clinically significant RBC antibody. Genotyping is also extremely useful to clarify equivocal results or where there is evidence of weakly expressed antigens. In addition, there are

worldwide shortages of 'traditional' grouping antisera that is leading to escalating test costs. Recently, genotyping has been used to assess the Rh type of the fetus in cases where the fetus is at risk of haemolytic disease. A real-time polymerase chain reaction method has been developed using the mother's circulating blood rather than amniotic cells as source of fetal DNA, and this method is now in widespread use (Finning *et al.*, 2008).

At present, the output from high-throughput molecular methods such as BLOODchip more than matches the routine requirements of transfusion centres and can be used to screen not only for ABO, Rhesus, MNS, Kell, Kidd and Duffy, but extended to Colton, Dombrock, Diego and 12 HPA loci (see Table S1). Recently, the Lutheran blood group system has also been included in the new generation of BLOODchip analysis (BLOODchip[®] v4.1; http://www.progenika.com). Overall, molecular genotyping does seem to offer the opportunity to improve the efficiency of high quality transfusion services and in our opinion high throughput blood genotyping might become an ideal platform in the near future.

Blood group data generated from the present survey suggest that successful blood donor recruitment in New Zealand is dependent on public support through blood donations from both, major and minor ethnic groups living in New Zealand. The genepool of Polynesian and Maori people is significantly different from those of Europeans, who are the majority in New Zealand (see Table S4 plus other references). The match probabilities shown in Table 2 demonstrated that blood groups are highly variable, not only between populations of geographically unrelated region (e.g. Europeans and Polynesians), but also within the populations themselves. For instance, even if patient and donor are compatible for ABO, Rhesus and Kell blood groups, the probability of mismatch is remains high for other blood groups (e.g. MNS, Kidd and Duffy), regardless of ethnic group. Therefore, it is important to

emphasize that getting a match for as many blood group antigens as possible between donor and recipient would significantly reduced risk of alloimmunization.

Patterns and processes involved during migration of A-SM speakers to Polynesia particularly gene flow from P-SA populations are demonstrated using NJ tree (Fig. 3) and PCO plot (Fig. 4). The tree and plot were constructed using only ABO, Rhesus and MNS phenotype frequencies, since only limited data are available for other blood groups. These blood group loci are the most polymorphic, the ones of greatest medical significance and, therefore, the ones for which most data exist. The tree and plot were unable to fully distinguish populations (e.g. Polynesians and A-SM) and show varying degrees of A-SM and P-SA genetic contribution to Polynesians, perhaps due to less discriminative genetic information being used than in earlier studies. Our previous findings on HLA loci showed a fairly wide range (55:45 to 90:10) of A:SM and P:SA proportions in Polynesian and Maori populations (see Edinur et al., 2012). Therefore, the inability of blood group data to show A:SM:P-SA ancestral fractions might also be due to founder effects and natural selection, as previously observed for HLA loci (see Mourant et al., 1978; Chambers and Edinur, 2013; Edinur et al., 2012). Differences between PFA and MFA on the plots might be due to high frequencies of ABO B and Rhesus DCCee phenotype frequencies in the former, compared with MFA (see Table 2). To better illustrate this point, we have plotted ABO and Rhesus phenotype frequency spectra for Polynesian, A-SM and P-SA populations in Figs. 5 and 6, respectively. We observed reduced genetic variation and frequencies of B, AB and DCCee phenotypes from the original base of settlement movements (i.e. Taiwan) to the most remote corners of Polynesia (i.e. Easter Island and New Zealand). In contrast, there are increased of DCcEe and DccEE phenotype frequencies with distance from Island of South East Asia and Melanesia. These observations support the presence of selective pressure and/or longer-term

bottleneck as the genepool of Polynesian sub-populations were repeatedly refined by founder effects due to multiple re-settlements. In addition, inferences made based on ABO and Rhesus frequencies are unable to link Maori (i.e. MFA and RFA) to any populations with absolute certainty, since only limited data available for other Polynesian sub-populations. Nonetheless, location of Polynesian populations at the lower right axis of the plot (Fig. 4) is evidence of their close genetic relationship to Asian populations, rather than Europeans and African. In contrast, differences observed between MFA and TM or Maori NZ is due to contemporary gene flow from Europeans as TM and Maori NZ (Woodfield *et al.*, 1987) were recruited from samples that are known to have European admixture. Therefore, the molecular blood group data from MFA provides good reference standard for the wider unadmixed Maori group, as this Maori sub-group is also matched well with the serological data for RFA (see Table S3).

As a conclusion, the present survey provides the first molecular blood group genotyping data for the Maori and Polynesian population living in New Zealand. Our validation using three different blood group genotyping methods gave 100% concordance and thus reflects the possibility of applying a range of molecular platforms in transfusion medicine. In addition, the collected blood group data are in general account accord with the of Polynesian origins (Chambers, 2008; Bellwood *et al.*, 2011; Chambers and Edinur, 2013; Chambers, 2013), although not as conclusive as those provided by genome-wide SNP, HLA and autosomal microsatellites markers (Hagelberg *et al.*, 1999; Zimdahl *et al.*, 1999; Friedlaender *et al.*, 2008; Kimura *et al.*, 2008; Wollstein *et al.*, 2010; Edinur *et al.*, 2012).



Fig. 5: ABO blood group phenotype frequency spectra in various Polynesian sub-populations. Reference populations were obtained from Nigg, 1930; Shapiro, 1940; Sanger *et al.*, 1951; Simmons *et al.*, 1955; Simmons and Graydon, 1957; Stavely and Douglas, 1958; Stavely and Douglas, 1959; Nakajima *et al.*, 1971; Lai and Bloom, 1982; Windhof and Walter, 1983.



Fig. 6: Rhesus blood group phenotype frequency spectra in various Polynesian sub-populations. Reference populations were obtained from Sanger *et al.*, 1951; Simmons *et al.*, 1955; Stavely and Douglas, 1958; Stavely and Douglas, 1959; Nakajima *et al.*, 1971; Lai and Bloom, 1982; Windhof and Walter, 1983.

Conflict of interest

The authors have no conflicts of interest

Acknowledgements

This study received financial support from Victoria University of Wellington and Ministry of Higher Education, Malaysia. We wish to acknowledge Maximiliano Crosetti (Progenika Biopharma S.A and Grifols International), William Perry (Grifols Australia), Martina Prager and Gabi Henschel (BAG Health Care GmbH, Lich, Germany) for their technical advice and assistance.

References

- Anstee, D.J. (2009) Red cell genotyping and the future of pretransfusion testing. *Blood*, **114**, 248.
- Avent, N.D. (2007) Large scale blood group genotyping clinical implications. *British* Journal of Haematology, **144**, 3.
- Avent, N.D., Martinez, A., Flegel, W.A., Olsson, M.L., Scott, M.L., Nogues, N. et al. (2009)
 The bloodgen project of the European Union, 2003-2009. *Transfusion Medicine and Hemotherapy*, 36, 162.
- Beiboer, S.H., Wieringa-Jelsma, T., Maaskant-Van Wijk, P.A., van der Schoot, C.E., van Zwieten, R., Roos, D. *et al.* (2005) Rapid genotyping of blood group antigens by multiplex polymerase chain reaction and DNA microarray hybridization. *Transfusion*, 45, 667.
- Badjie, K.S.W., Tauscher, C., Buskirk, C.V., Wong, C., Jenkins, S., Smith, C. *et al.* (2011)
 Red blood cell phenotype matching for various ethnic groups. *Immunohematology*, 27, 12.
- Bellwood, P., Chambers, G.K., Ross, M. & Hung, C-H. (2011) Are "cultures" inherited?
 Multidisciplinary perspectives on the origins of Austronesian–speaking peoples prior to
 1000 BC. In: *Investigating archaeological cultures: material culture, variability and transmission* (eds. B.W. Roberts, M.V. Linden), pp. 321. Springer, Dordrecht.
- Chambers, G.K. (2008) Genetics and the origins of the Polynesians. In: Handbook of human molecular evolution (eds. D.N. Cooper & H. Kehrer-Sawatzki), pp. 489. John Wiley & Sons Inc, Chichester.
- Chambers, G.K. (2013) Genetics and the Origins of the Polynesians. In: *Encyclopedia of Life Sciences*. John Wiley & Sons Ltd, Chichester.

- Chambers, G.K. & Edinur, H.A. (2013) Genetic relationships between Malays and Maori. In: *The changing values of Malays, Maori and Pacific Islanders* (eds. W.R.A. Rahman & T.R. Higgins), pp. 6. Chair of Malay Studies & Victoria University Foundation, Wellington, New Zealand.
- Chen, K.H., Cann, H., Chen, T.C., Van West B. & Cavalli-Sforza, L. (1985) Genetic markers of an aboriginal Taiwanese population. *American Journal of Physical Anthropology*, 66, 327.
- Daniels, G. (2008) Other blood groups. In: *Technical manual* (eds. J.D. Roback., M.R. Combs., B.J. Grossman & C.D. Hillyer), pp. 411. American Association of Blood Banks, Bethesda, Maryland.
- Daniels, G. & Reid M.E. (2010) Blood groups: the past 50 years. Transfusion, 50: 281.
- Eccles D.A. (2011) Genomic analysis of human population structure. Victoria University of Wellington, Wellington, New Zealand.
- Edinur, H.A., Dunn, P.J.J., Hammond, L., Selwyn, C., Velickovic, Z.M., Lea, R.A. *et al.* (2012) Using HLA loci to inform ancestry and health in Polynesian and Maori populations. *Tissue Antigens*, **80**, 509.
- Edinur, H.A., Dunn, P.J.J., Lea, R.A. & Chambers, G.K. (2013) Human platelet antigens frequencies in Maori and Polynesian populations. *Transfusion Medicine*, re-submitted.
- Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1: 47.
- Felsenstein, J. (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle.
- Finning, K., Martin, P., Summers, J., Massey, E., Poole, G. & Daniles, G. (2008) Effect of high throughput RhD typing of foetal DNA in maternal plasma on use of anti-RhD

immunoglobulin in RhD negative pregnant women: prospective feasibility study. *British Medical Journal*, **336**, 816.

- Friedlaender, J.S., Friedlaender, F.R., Reed, F.A., Kidd, K.K., Kidd, J.R., Chambers, G.K. et al. (2008) The genetic structure of Pacific Islanders. PLoS Genetics, 4, 174.
- Gajdusek, D.C., Leyshon, W.C., Kirk, R.L., Blake, N.M., Keats, B. & McDermid E.M. (1978) Genetic differentiation among populations in Western New Guinea. *American Journal of Physical Anthropology*, 48, 47.
- Grunbaum, B.W., Crim, M., Selvin, S., Myhre, B.A. & Pace, N. (1980) Distribution of gene frequencies and discrimination probabilities for 22 human blood genetic systems in four racial groups. *Journal of Forensic Sciences*, 25, 428.
- Hagelberg, E., Kayser, M., Nagy, M., Roewer, L., Zimdahl, H., Krawczak, M. *et al.* (1999)
 Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. *Philosophical Transactions of the Royal Society B*, **354**, 141.
- Hashmi, G., Shariff, T., Zhang, Y., Cristobal, J., Chau, C., Seul, M., *et al.* (2007) Determination of 24 minor red blood cell antigens for more than 2000 blood donors by high-throughput DNA analysis. *Transfusion*, **47**, 736.
- Hustinx, H., Fontana, S., Gowland, P. & Niederhauser, C. (2009) Will genotyping replace serology in future routine blood grouping? - Opinion 2. *Transfusion Medicine and Hemotherapy*, 36, 228.
- Irshaid, N.M., Henry, S.M. & Olsson, M.L. (2000) Genomic characterization of the Kidd blood group gene: different molecular basis of the Jk(a-b-) phenotype in Polynesians and Finns. *Transfusion*, **40**, 69.

- Jungbauer, C., Hobel, CM., Schwartz, D.W.M. & Mayr W.R. (2012) High-throughput multiplex PCR genotyping for 35 red blood cell antigens in blood donors. *Vox Sanguinis*, **102**, 234.
- Kappler-Gratias, S., Peyrard, T., Beolet, M., Amiranoff, D., Menanteau, C., Dubeaux, I. *et al.*(2011) Blood group genotyping by high-throughput DNA analysis applied to 356 reagent red blood cell samples. *Transfusion*, **51**: 36
- Karpasitou, K., Drago, F., Crespiatico, L., Paccapelo, C., Truglio, F., Frison, S. *et al.* (2008) Blood group genotyping for Jk^a/Jk^b, Fy^a/Fy^b, S/s, K/k, Kp^a/Kp^b, Js^a/Js^b, Co^a/Co^b, and Lu^a/Lu^b with microarray beads. *Transfusion*, **48**, 505.
- Kimura, R., Ohashi, J., Matsumura, Y., Nakazawa, M., Inaoka, T., Ohtsuka, R. *et al.* (2008) Gene flow and natural selection in oceanic human populations inferred from genomewide SNP typing. *Molecular Biology and Evolution*, **25**, 1750.
- Lai, L.Y.C. & Bloom, J. (1982) Genetic variation in Bougainville and Solomon Islands populations. American Journal of Physical Anthropology, 58, 369.
- Lee, S., Wu, X., Reid, M., Zelinski, T. & Redman, C. (1995) Molecular basis of the Kell (K1) phenotype. *Blood*, **4**, 912.
- Lin, M. & Broadberry, R.E. (1998). Immunohematology in Taiwan. Transfusion Medicine Reviews, 12, 56.
- Lin, M. & Yu, L-C. (2008) Frequencies of the *JK*^{null} (IVS5-1g>a) allele in Taiwanese, Fujian, Filipino, and Indonesian populations. *Transfusion*, **48**, 1768.
- Liu, H-M., Lin, J-S., Chen, P-S., Lyou, J-Y., Chen, Y-J. & Tzeng, C.H. (2009) Two novel Jknull alleles derived from 222C>A in Exon 5 and 896G>A in Exon 9 of the JK gene. *Transfusion*, **49**, 259.

- Lucien, N., Sidoux-Walter, F., Olives, B., Moulds, J., Le Pennec, P-Y., Cartron, J-P. *et al.* (1998) Characterization of the gene encoding the human kidd blood group/urea transporter protein. *The Journal of Biological Chemistry*, **273**, 12973.
- Montpetit, A., Phillips, MS., Mongrain, I., Lemieux, R. & St-Louis, M. (2006) Highthroughput molecular profiling of blood donors for minor red cell and platelet antigens. *Transfusion*, 46, 841.
- Mourant, A.E., Kopec, A.C. & Domaniewska-Sobczak, K. (1978) Blood groups and diseases: a study of associations of diseases with blood groups and other polymorphisms. Oxford University Press, London.
- Nakajima, H., Ohkura, K., Huang, M-C., Saito, R. & Seto, T. (1971) The distribution of several serological and biochemical traits in East Asia. IV. The distribution of the blood groups in the Taiwanese mountain aborigines. *Jinrui Idengaku Zasshi*, **16**, 57.
- Nigg, C. (1930) A study of the blood group distribution among Polynesians. *The Journal of Immunology*, **19**, 93.
- Page, R.D.M. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357.
- Patnaik, S.K., Helmberg, W. & Blumenfeld, O.O. (2012) BGMUT: NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. *Nucleic Acids Research*, 40, D1023.
- Rios, M., Cash, K., Strupp, A., Uehlinger, J. & Reid, M. (1999) DNA from urine sediment or buccal cells can be used for molecular genotyping. *Immunohematology*, 15, 61.
- Sanger, R., Walsh, R.J. & Kay, M.P. (1951) Blood types of natives of Australia and New Guinea. *American Journal of Physical Anthropology*, **9**, 71.
- Schonitzer, D. (2009) Will genotyping replace serology in future routine blood grouping? -Opinion 3. *Transfusion Medicine and Hemotherapy*, **36**, 230.

- Shapiro, H. (1940) The distribution of blood groups in Polynesia. American Journal of *Physical Anthropology*, **26**, 409.
- Simmons, R.T. & Graydon J.J. (1957) A blood group genetical survey in Eastern and Central Polynesians. American Journal of Physical Anthropology, 15, 357.
- Simmons, R.T., Graydon, J.J., Semple N.M. & Fry E.I. (1955) A blood group genetical survey in Cook Islanders, Polynesia, and comparisons with American Indians. *American Journal of Physical Anthropology*, **13**, 667.
- Staveley, J.M. & Douglas, R. (1958) Blood groups in Maoris. The Journal of the Polynesian Society, 67, 239.
- Staveley, J.M. & Douglas, R. (1959) Blood groups in Tongans (Polynesia). The Journal of the Polynesian Society, 68, 348.
- Storry, J.R. & Olsson, M.L. (2009) Will genotyping replace serology in future routine blood grouping? - Opinion 4. *Transfusion Medicine and Hemotherapy*, **36**, 232.
- Veldhuisen, B., van der Schoot, C.E. & de Haas, M. (2009) Blood group genotyping: from patient to high-throughput donor screening. *Vox Sanguinis*, **97**, 198.
- Wagner, F.F. (2009) Will genotyping replace serology in future routine blood grouping? -Opinion 1. *Transfusion Medicine and Hemotherapy*, 36, 226.
- Wester, E.S., Johnson, S.T., Copeland, T., Malde, R., Lee E., Storry, J.R. *et al.* (2008) Erythroid urea transporter deficiency due to novel JKnull alleles. *Transfusion*, **48**, 365.
- Windhof, O. & Walter, H. (1983) Blood group, serum protein and red cell enzyme polymorphisms in Filipinos. *Human Heredity*, **33**, 357.
- Woodfield, D.G, Simpson, L.A., Seber, G.A.F., & McInerney P.J. (1987) Blood groups and other genetic markers in New Zealand Europeans and Maoris. Annals of Human Biology, 14, 29.

- Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R.J., Nurnberg, P. et al. (2010) Demographic history of Oceania inferred from genome-wide data. *Current Biology*, 20, 1983.
- Wu, Y.Y. & Csako, G. (2006) Rapid and/or high-throughput genotyping for human red blood cell, platelet and leukocyte antigens, and forensic applications. *Clinica Chimica Acta*, 363, 165.
- Zimdahl, H., Schiefenhovel, W., Kayser, M., Roewer, L. & Nagy, M. (1999) Towards understanding the origin and dispersal of Austronesians in the Solomon Sea: HLA class II polymorphism in eight distinct populations of Asia-Oceania. *European Journal of Immunogenetics*, 26, 405.

Supplementary information

Blood group variants	BAGene DNA-SSP	Progenika SNP
ABO	ABO-TYPE variant	
А	+	+
В	+	+
0	+	+
Rhesus	RH-TYPE	
RhD+	+	+
RhD-	+	+
Weak D	+	+
Partial D	+	+
C	+	+
c	+	+
E	+	+
E	+	+
MNS	MNS-TYPE	
М		+
Ν	+	+
S	+	+
S	+	+
U	+	+
Mia	-	+
Kell	KKD-TYPE	
Κ	+	+
k	+	+
Kpa	-	+
Kpb	-	+
Kpc	-	+
Jsa	-	+
Jsb	-	+
Kmod	-	+
Kidd		
Jka	+	+
Jkb	+	+
IVS5-1g>a	-	+
Duffy		
Fya	+	+
Fyb	+	+
FyGATA	+	+
Fyx	+	+
Diego	-	
Dia	-	+
Dib	-	+
Colton	-	
Coa	-	+
Cob	_	+

Table S1: Blood group variants captured by DNA-SSP and BLOODchip technology.

Tab	le S1,	cont.
-----	--------	-------

Blood group variants	BAGene DNA-SSP	Progenika SNP
Dombrock	-	
Doa	-	+
Dob	-	+
200		·
HPA	HPA-TYPE	
HPA-1a	+	+
HPA-1b	+	+
HPA-2a	+	+
HPA-2b	+	+
HPA-3a	+	+
HPA-3b	+	+
HPA-4a	+	+
HPA-4b	+	+
HPA-5a	+	+
HPA-5b	+	+
НРА-ба	-	+
HPA-6b	-	+
HPA-7a	-	+
HPA-7b	-	+
HPA-8a	-	+
HPA-8b	-	+
HPA-9a	-	+
HPA-9b	-	+
HPA-10a	-	+
HPA-10b	-	+
HPA-11a	-	+
HPA-11b	-	+
HPA-15a	+	+
HPA-15b	+	+

Locus	Locus Primers PCR reaction set-up PCR cycling c					
			Step	Temp	Time	Cycles
IVS5-1A>G	F: CAAGTGCAACCAAAGCTCAC	Template = 2ul (15ng/ul)	P.denaturation	94	7min	1
	R: CTGCCATATAACAACTCCCATTC (Wester et al., 2008)	Forward primer = 1ul (10uM)	Denaturation	96	30s	
		Reverse primer = 1ul (10uM)	Annealing	58	30s	
		MyTaqTM Red Mix (Bioline) =12.5ul	Extension	72	60s	35
		Distilled water = 9.5ul	Final extension	72	7min	1
Locus	Primers	PCR reaction set-up	PCR cycling condition			
			<u>Step</u>	<u>Temp</u>	Time	Cycles
Kell	F: TTTAGTCCTCACTCCCATGCTTCC	Template = $2ul (15ng/ul)$	P.denaturation	94	5min	1
	R: TATCACACAGGTGTCCTCTCTCC (Lee et al., 1995)	Forward primer = 1ul (10uM)	Denaturation	94	20s	
		Reverse primer = 1ul (10uM)	Annealing	62	15s	35
		MyTaqTM Red Mix (Bioline) =12.5ul	Extension	72	20s	
		Distilled water = 9.5ul	Final extension	72	10min	1
Locus	Primers	PCR reaction set-up	PCR cycling condition			
			Step	<u>Temp</u>	Time	<u>Cycles</u>
Duffy	F: TCCCCCTCAACTGAGAACTC (Rios et al., 1999)	Template = 2ul (15ng/ul)	P.denaturation	94	5min	1
	R: AAGGCTGAGCCATAYYAGAC (this study)	Forward primer = 1ul (10uM)	Denaturation	94	20s	
		Reverse primer = 1 ul (10 uM)	Annealing	62	15s	35
		MyTaqTM Red Mix (Bioline) =12.5ul	Extension	72	20s	
		Distilled water = 9.5ul	Final extension	72	10min	1

Table S2: PCR protocols for SBT of Kell, IVS5-1 and Duffy variants

All amplified products were sequenced using ABI3730 DNA Analyzer (Applied Biosystems) by Massey Genome Service (Massey University, New Zealand)

	A	BO	
	TR	TM	RFA
TR	~	~	~
TM	0.14	~	~
RFA	0.67	0.05	~
MFA	0.50	0.74	0.30
	Rh	esus	
	TR	ТМ	RFA
TR	~	~	~
TM	0.16	~	~
RFA	0.18	0.01	~
MFA	0.37	0.08	1

Table S3: Exact tests on Maori datasets

Abbreviations: TR: Total RakaiPaaka, TM: Total Maori, RFA: RakaiPaaka with Full Ancestry, MFA: Maori with Full Ancestry

	ABO)			Rhesus	5	
	ТМ	MFA	PFA		TM	MFA	PFA
ТМ	~	~	~	TM	~	~	~
MFA	0.73	~	~	MFA	0.83	~	~
PFA	0.00	0.01	~	PFA	0.00	0.00	~
Euro	0.02	0.06	0.11	Euro	0.00	0.00	0.00
	MNS	5			Kell		
	TM	MFA	PFA		ТМ	MFA	PFA
TM	~	~	~	ТМ	~	~	~
MFA	1.00	~	~	MFA	1.00	~	~
PFA	0.82	0.84	~	PFA	1.00	1.00	~
Euro	0.00	0.00	0.02	Euro	0.11	0.30	0.74
	Kidd	l			Duffy		
	ТМ	MFA	PFA		ТМ	MFA	PFA
TM	~	~	~	TM	~	~	~
MFA	0.30	~	~	MFA	1.00	~	~
PFA	0.27	0.04	~	PFA	0.03	0.05	~
Euro	0.00	0.00	0.17	Euro	0.00	0.02	0.00
	Dombro	ock					
	TM	MFA	PFA				
TM	~	~	~				
MFA	0.67	~	~				
PFA	0.41	0.21	~				
Euro	0.01	0.01	0.91				

Table S4: Exact tests on Polynesian and European datasets

Abbreviations: Euro = Europeans, TM =Total Maori, MFA= Maori with Full Ancestry, PFA = Polynesians with Full Ancestry

Table S5: Comparison of phenotype predictions based on PCR-SSP and combination of PCR-SSP and SBT (*JK*IVS5-1g>a*).

Kidd	PCR-SSP (N)	PCR-SSP plus JK*IVS5-1g>a (N)	Discordant results (N)
Jk(a+b-)	21	32	
Jk(a-b+)	5	5	11 (19%)
Jk(a+b+)	33	22	

Abbreviations: N = sample size

Table S6: Summary of the costs associated with the three molecular methods used in the present survey.

Methods	Blood groups	HPA	DNA	Cost	Time	Comments
PCR-SSP	ABO, Rhesus, Kell, Kidd, Duffy and MNS	1 to -5 and - 15	65ul (50ng/ul)	c. €128 per sample	3h/samples/ blood group	Labour extensive, time consuming and require lots of DNA template
BLOODchip	ABO, Rhesus, Kell, Kidd, Duffy, MNS, Dombrock, Diego and Colton	1 to -11 and - 15	40 ul (50ng/µl)	c. €350 per sample	20 samples, 10h	Expensive but cover multiple number of antigens and short turnaround time
SBT	Kell and Duffy	HPA-6	6ul (15ng/ul)	c. €10 per sample	2-3 days	Accurate but laborious, costly and time-consuming

* The overall cost is given for the blood group tests covered by each technology, excluding technician time and other related cost.
Human platelet antigens frequencies in Maori and Polynesian populations

H.A. Edinur^{*}, P.P.J. Dunn[§], R.A. Lea^{\dagger} and G.K. Chambers^{*}

*School of Biological Sciences, Victoria University of Wellington, New Zealand

[§]Tissue Typing Laboratory, New Zealand Blood Service, Auckland

[†]School of Medical Sciences, Griffith University, Australia

Status: Transfusion Medicine, 23, 330-337.

Key words: Platelet, alloimmunization, Polynesia, Polynesians, Maori

Correspondence: Dr Geoffrey K. Chambers, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand. Ph: +64-(0)4-463-6091, Fax: +64-(0)4-463-5331, e-mail: Geoff.Chambers@vuw.ac.nz

Chapter 3.4

SUMMARY

Background: Allele frequencies of human platelet antigens (HPA) reflect population history and possibility of platelet-specific alloimmunization. Here, we report on screening of variants at HPA loci for Polynesian and Maori subjects.

Objectives: Our aims are to evaluate new HPA genotyping methods, compile and analyse new HPA datasets for these subjects, use HPA data for tracing ancestry, migration patterns, genetic admixture and its potential influence on health.

Materials and Methods: A total of 75 Maori and 25 Polynesian DNA samples were genotyped using commercial BAGene HPA-TYPE DNA-SSP kits, BLOODchip hybridization SNP assays and DNA sequence based typing.

Results: Genotyping was successful and cross validation of PCR-SSP and BLOODchip gave 100% agreement. Among the HPA loci tested, only 6 are dimorphic (HPA-1 to -3, -5, -6 and -15) and all others are monomorphic. The Polynesians and Maori have the 'a' allele form as the most common for all loci except HPA-15.

Conclusions: The newly observed HPA data as well as PCO analysis clearly indicate genetic contributions from both, Asia and Australasia in Maori and Polynesian populations together with recent admixture with Europeans. In addition, different prevalences of HPA alleles among Polynesian, Maori and European populations contribute towards different risk profiles for platelet specific alloimmunization. This is the first report for these populations and our findings are of direct practical relevance for blood transfusion centres, the management of pregnancies, assessment of NAIT and management of multi-transfused patients.

Chapter 3.4

INTRODUCTION

Human platelet antigens (HPA) are immunogenic transmembrane glycoprotein antigens expressed on the surface of anucleate platelets (Robinson *et al.*, 2010 and Rozman, 2002). Neonatal alloimmune thrombocytopenia (NAIT), post transfusion purpura (PTP) and platelet transfusion refractoriness are among clinical consequences of alloantibodies against HPA that develop during pregnancy and after blood or platelet transfusion (see Lubenow *et al.*, 2000; Ohto *et al.*, 2000; Spencer and Burrows, 2001; Engelfriet and Reesink, 2003; Metcalfe *et al.*, 2003), although there is evidence that PTP is autoantibody rather than alloantibody induced (Watkins *et al.*, 2012). The molecular basis of 22 serologically defined HPA (see Robinson *et al.*, 2010) have been assigned and approved according to guidelines set up by The Platelet Nomenclature Committee (Metcalfe *et al.*, 2003). Alloantibodies against both, high and low frequency alleles (designated as a and b, respectively) are only recorded for six HPA systems (i.e. HPA-1 to -5, and -15).

Population genetic studies have shown that distributions of HPA alleles reflect ancestry and the potential prevalence of alloimunization (e.g. see Feng *et al.*, 2006; De La Vega Elena *et al.*, 2008). We are now investigating this proposition by undertaking a new study on previously uncharacterized populations of New Zealand Maori and Polynesians. Previous studies on genes, language and culture have thrown up contrasting theories that each account for parts of the patterns and some of the processes involved during diffusion of Austronesian-speaking Mongoloid (A-SM) peoples to Remote Oceania. These ideas are known by various exotic names; *Slow Boats, Fast Trains, Out of Taiwan* etc. This competition has been further reinforced by apparently conflicting data from genetic studies of mtDNA vs. Y-chromosome

vs. autosomal markers, only recently resolved by adding in gender biased gene flow in matrilineal societies (Chambers, 2006).

The emergent consensus is captured by the following updated version of our '*Synthetic Total Evidence Model (STEM)*' as an extension to the one first proposed Chambers (2006). This comprehensive model combines all previous ideas and data relating to the origins of Polynesian peoples (see also Chambers and Edinur, 2012 and Chambers, 2013 for the updated version of *STEM*). The model involves greater or lesser blending of the genepools of two distinct peoples; A-SM (ex-Taiwanese Aboriginals) and Papuan-speaking Australoid (P-SA: ex-coastal Melanesia and islands of the Bismarck Archipelago). It is thus a reticulated history of human evolution in the Pacific, rather than an exclusively branching one. Under this model, A-SM people migrated south from Taiwan (Bellwood *et al.*, 2011) via Island South East Asia (ISEA) and engaged in gender biased genetic exchange with P-SA people in the Bismarck Archipelago creating Lapita culture and displacing local Papuan languages. Low levels of genetic introgression from P-SA to A-SM created Proto-Polynesians and later gave rise to those now known as Melanesians (Friedlaender *et al.*, 2008; Kayser, 2010) via continued directional admixture.

Movement into the oceanic islands of Near Oceania by Proto-Polynesians was the next step carrying Lapita pottery skills with them. They were later followed as far as Fiji by Melanesians who spoke Austonesian languages and displaced most, but not all of the Proto-Polynesians from Vanuatu, New Caledonia etc. (Bellwood, 1978). The Proto-Polynesians moved on to occupy Remote Oceania reaching a core area in Eastern Polynesia and then rapidly expanding to fill up the Polynesian Triangle (Hawaii, Easter Island and New Zealand). In New Zealand, Maori remained as an isolated people from around 650ybp until the arrival of Europeans over the last 150 years.

This recently extended version of the basic STEM model is supported by multiple lines of evidence and subsumes all previous partial pattern and process models (see above and Bellwood et al., 2011), except for those that exclude Taiwan as the origin of the A-SM component. It makes testable predictions with respect to populations of origin, the genderbiased reticulated admixture process leading to differentiation of male and female lineages and genepool refinement via bottlenecks and founder effects. The aims of present study are to evaluate new molecular HPA genotyping (i.e. sequence specific primers vs. single nucleotide polymorphism array), compile and analyse new HPA datasets for Polynesian and Maori subjects, use HPA data from this study for tracing ancestry, migration patterns and genetic admixture and to evaluate their potential influence on health. To the best of our knowledge, this is the first molecular genetics study of HPA for Maori and the only available HPA data for Polynesians are those from Ma'ohis individuals, the native population of the Tuamotu Archipelago reported by Halle et al. (2004). These data complement our recent study of HLA in Maori and Polynesians and its relevance to transplantation and autoimmune diseases (Edinur et al., 2012) and provide valuable information on frequencies of HPA antigens, important in transfusion sciences.

MATERIALS AND METHODS

Samples

A total of 75 Maori (Total Maori = TM) and 25 Polynesian with Full Ancestry (PFA) genomic DNA samples obtained with informed consent and ethical approval were drawn from the Victoria University of Wellington DNA Bank. The larger Maori group was then

divided into a pair of smaller sub-groups based on self-defined pedigree: Maori with Full Ancestry (MFA: N = 47) + Maori with Admixed History (MAH: N = 28).

HPA genotyping

The genomic DNA were screened for HPA variants using commercial (BAGene DNA-SSP HPA-TYPE kit; BAG Health Care GmbH, Lich, Germany) polymerase chain reaction (PCR)sequence specific primer (SSP) kit and hybridization single nucleotide polymorphism (SNP) assay service (BLOODchip; Progenika Biopharma S.A., Spain) for genotyping of HPA-1 to -5 and-15, and HPA-1 to -11 and -15, respectively. A total of 60 and 50 samples were genotyped using these methods, respectively. In order to cross validate molecular typing methods used in this study, ten of these samples were typed using both, PCR-SSP kits and BLOODchip analysis service. We also performed sequence based typing (SBT) for HPA-6 on the samples which were genotyped using the commercial SSP kits. Our pair of forward and reverse primers and PCR protocol for HPA-6 genotyping is based on Tanaka *et al.* (1996). The PCR products were sequenced using ABI3730 DNA Analyzer (Applied Biosystems) by Massey Genome Service (Massey University, Palmerston North, New Zealand).

Statistical analysis

Allele and genotype frequencies were estimated by direct counting. Chi-squared tests (SPSS Inc., Chicago, IL) were used to test for Hardy-Weinberg equilibrium (HWE) and to compare HPA genotype frequencies between pairs of our Maori and Polynesian study groups and sub-groups. Principal coordinate (PCO) analysis was performed using Multivariate Statistical Software Package 3 (Kovach Computing Services, UK; http://www.kovcomp.com/mvsp). Neighbor-joining (NJ) phylogenetic tree was constructed from Nei's genetic distances using PHYLIP software package, version 3.6 (Felsenstein, 2004) and the topology of the tree was

visualized using TreeView program (Page, 1996). Probabilities of transfusion and pregnancy alloimmunization were calculated according to De La Vega Elena *et al.* (2008).

RESULTS

The HPA alleles and frequencies at each locus observed in Maori and Polynesian study groups and sub-groups are listed in Table 1 and see Fig. 1 in Discussion. Together with our newly collected data in Table 1 are HPA data (HPA-1 to 11 from Jones *et al.*, 2003 and HPA-15 from Ertel *et al.*, 2005) from reference (EUR) European populations (see later for population comparison and estimation of transfusion and gestation incompatibility). Most of the HPA loci tested are polymorphic, except for HPA-4, and -7 to -11. In addition, only homozygous HPA-5a and -6a individuals were observed in the MFA sub-group. All HPA genotype profiles for the polymorphic HPA systems in Maori and Polynesian study groups and sub-groups are in HWE (see Table S1; non-polymorphic HPA systems are removed for simplicity). Chi-squared tests on genotype frequencies (Table S2) show no significant differences between pairs of study population groups and sub-groups except for PFA vs. MFA at HPA-3, -5 and -6 and TM vs. PFA at HPA-6. Admixture effects in our TM and MAH were then shown by plotting them on a bar diagram (Fig. 2) where polymorphic HPA alleles in TM and MAH are usually intermediate between MFA and our reference Europeans populations (Jones *et al.*, 2003 and Ertel *et al.*, 2005).

Genetic relationships between Polynesians and Melanesians and their putative Mongoloid and Australoid ancestors from elsewhere in Asia and Australasia were probed by building HPA-1 to -5 allele frequency PCO plots (Fig. 3) and NJ phylogenetic tree (Fig. 4). In these figures, A-SM and P-SA are the ancestral types who have contributed to the genepool of Polynesians and Maori. The PCO plots show more clear separation of populations; A-SM, P-SA and Europeans. In the plots, the PFA group is well separated and fall between A-SM populations and Australian aborigines compared with MFA which tends to cluster more closely with their A-SM relatives. As might be expected, our admixed sub-group, MAH, is plotted as displaced toward European populations in the lower right corner of Fig. 3, reflecting recent European admixture.

We also calculated probabilities of transfusion and foetomaternal incompatibility for Maori and Polynesian datasets and used reference European (EUR) populations (Jones *et al.*, 2003 for HPA-1, -2, -3, -5, -6 and Ertel *et al.*, 2005 for HPA-15) to represent the majority European ethnic group in New Zealand (see Table S3 for details and how the data can be interpreted). For all datasets, HPA-4 and -7 to -11 are monomorphic. Thus, it is unlikely for alloimunization to develop for these HPA loci and, therefore, they have been removed from Table S3.

Table 1: HPA allele frequencies in Polynesian and Maori study groups and sub-groups and
reference European [§] (EUR) populations

HPA-	<u>TM (N=75)</u>	<u>MAH (N=28)</u>	<u>MFA (N=47)</u>	<u>PFA (N=25)</u>	EUR*
1a	0.960	0.946	0.968	0.980	0.840
1b	0.040	0.054	0.032	0.020	0.161
2a	0.933	0.911	0.947	0.920	0.925
2b	0.067	0.089	0.053	0.080	0.0075
3a	0.580	0.607	0.564	0.720	0.627
3b	0.420	0.393	0.436	0.280	0.373
4a	1.000	1.000	1.000	1.000	1.000
4b	0.000	0.000	0.000	0.000	0.000
5a	0.987	0.964	1.000	0.940	0.914
5b	0.013	0.036	0.000	0.060	0.086
6a	0.987	0.964	1.000	0.900	1.000
6b	0.013	0.036	0.000	0.100	0.000
7a	1.000	1.000	1.000	1.000	1.000
7b	0.000	0.000	0.000	0.000	0.000
8a	1.000	1.000	1.000	1.000	1.000
8b	0.000	0.000	0.000	0.000	0.000
9a	1.000	1.000	1.000	1.000	1.000
9b	0.000	0.000	0.000	0.000	0.000
10a	1.000	1.000	1.000	1.000	1.000
10b	0.000	0.000	0.000	0.000	0.000
11a	1.000	1.000	1.000	1.000	1.000
11b	0.000	0.000	0.000	0.000	0.000
15a	0.340	0.304	0.362	0.300	0.512
15b	0.660	0.696	0.638	0.700	0.488

HPA: human platelet antigen, N = sample size, TM: Total Maori, MAH: Maori with Admixed History, MFA: Maori with Full Ancestry and PFA: Polynesians with Full Ancestry, § = HPA data for Europeans (EUR) obtained from Jones et al., 2003 (HPA-1 to -6) and is Ertel et al., 2005 (HPA-15) and * = sample size for HPA-1 to -6 is 134 and 407 for HPA-15.



Fig. 1: Agarose gel electrophoresis of the HPA allele specific products amplified using BAGene DNA-SSP HPA-TYPE kit.



Fig. 2: Bar graph for polymorphic HPA loci show that our MAH is intermediate between MFA, TM and Europeans (EUR). HPA data for EUR were obtained from Jones *et al.*, 2003 (HPA-1 to -3, -5 and -6) and is Ertel *et al.*, 2005 (HPA-15).



Fig. 3: PCO plots constructed using HPA-1 to -5 allele frequencies. Reference populations for Europeans (grey), Austronesian-speaking Mongoloid (blue font), Papuanspeaking Australoid (red) and Polynesians (green) obtained from Seller *et al.*, 1999; Chu *et al.*, 2001; Bennet *et al.*, 2002; Jones *et al.*, 2003, Shih *et al.*, 2003; Halle *et al.*, 2004; Robinson *et al.*, 2010; Tan *et al.*, 2012. Abbreviations: TM: Total Maori, MAH: Maori with Admixed History, MFA: Maori with Full Ancestry and PFA: Polynesians with Full Ancestry.



Fig. 4: Position of Polynesians and Maori in N-J phylogenetic tree constructed using HPA-1 to -5 allele frequencies compared to those probed in Fig. 3. Abbreviations and references are as assigned for Fig. 3.

Chapter 3.4

DISCUSSION

Serological typing has been integral to platelet phenotyping until the early 1990s (Norton *et al.*, 2004). However, the serological assays have certain well known limitations, especially with limited supply of antisera. Introduction of molecular techniques has contributed toward a better understanding of the genetic and morphological variation of HPA alleles. Differences between the HPA alleles are mostly due to single nucleotide polymorphism leading to a single amino acid substitution (see Metcalfe *et al.*, 2003 for molecular basis of HPA alleles). With the elucidation of the molecular basis of HPA, many DNA-based methods are now available for HPA characterization, including regular PCR-based (PCR-SSP and PCR-RFLP; restriction fragment length polymorphism) and high-throughput (SNP array and Luminex) genotyping platform; see also Veldhuisen *et al.*, 2009; Wu and Csako, 2006; Avent, 2007 for reviews on molecular techniques available for HPA typing.

Genotyping of all samples was successful and cross validation using 10 samples gave 100% agreement. The Progenika Biopharma service is simple and convenient. DNA samples (with minimum concentration and amount of 50 ng/µl and 2 µg, respectively) are sent via courier to facilities in Derio, Spain. It takes about 2 weeks and data reports are returned as electronic spreadsheets. We conducted our tests as part of a wider blood group screening exercise (128 SNPs for 9 blood groups and 12 HPA loci; Edinur, unpublished) and cost c. €350 per sample. The Progenika Biopharma does offer platelet only assay using a combination of multiplex PCR and Luminex technology (i.e. IDHPA). Genotyping of HPA using BAGene DNA-SSP HPA-TYPE kit was done in-house. The SSP method requires PCR followed by determination of product size by using agarose gel electrophoresis – see Fig. 1. Gel is easily interpreted 'by eye' or with manufacturer's software. The PCR-SSP kits use a simple

procedure, work well, require $14\mu l$ (50-100ng/ μl) of DNA, relatively inexpensive but restricted to limited number of HPA specificities (i.e. without HPA-6 to -11 variants).

There were no inconsistencies among the ten samples genotyped using both, PCR-SSP kits and BLOODchip and this indicates the accuracy of molecular diagnostics. However, we needed to perform SBT for HPA-6 locus on those samples genotyped using the commercial PCR-SSP kits, because this locus is polymorphic in Maori and Polynesians and is not included in their kit (see later). The service provided by Progenika Biopharma is more expensive compared with BAGene PCR-SSP kits or any PCR-based typing, but some may judge the price to be worthwhile when one considers the total number of SNPs (i.e 128 SNPs for full detailed coverage of 9 blood groups and 12 HPA systems) included on the BLOODchip. Several PCR-SSP kits or PCR protocols are needed to genotype larger number of antigens which in turn are costly, laborious, time consuming and require lots of DNA template. Overall, DNA-based methods provide the best option for accurate and specific characterization of HPA and are superior to serogical phenotyping for prenatal diagnostic and management of multi-transfusion patients (see Wu and Csako, 2006; Avent, 2007). Nonetheless, a high-throughput molecular method such as BLOODchip would be preferable for mass screening of donors with rare genotypes at multiple HPA loci. There is a single caveat. All DNA markers show potential for expression of a particular marker type, but serology shows the actual expression. Thus, molecular analysis of DNA SNP may not reveal a variant mutation that affects expression or modification of this allele. Serological tests are unlikely to miss novel null variant in individual except in the sense that they will appear as homozygote. Novel variant present different problem as they may cross react weakly or not at all with antibody-based reagent.

Visual comparisons on the distributions of HPA allele frequencies (Table 1) and statistical analysis on genotype frequencies (see Table S2) show no large contrast between MFA and MAH/TM, except for evidence of reduced variability in MFA. Among the 12 HPA loci tested, only 4 are dimorphic in MFA, compared with 6 in both, TM and MAH. This finding of reduced diversity in MFA is consistent with factors associated with founder effects and longer-term bottlenecks which are known to reduce allelic diversity, specifically via the preferential loss of rare alleles (i.e. HPA-5b and -6b in MFA). Nonetheless, we are able to show the effect of recent European admixture with confidence as can be readily seen in Fig. 2 where in almost every case the length of each bar representing the MAH sub-group is intermediate between those for TM, MFA and EUR. Therefore, we use HPA data from MFA for tracing Maori lineage and history. In contrast, HPA data from TM is very important for platelet alloimunization estimation, because they represent the genepool of present day Maori living in New Zealand due to its combination of MFA and MAH components.

Differences between PFA and MFA (see Table 1 and S2) are really only obvious for HPA-3, -5 and -6. These are due to the particularly low occurrence of HPA-3b in PFA (0.28 vs. 0.44 in MFA) and monomorphic nature of the HPA-5 and -6 systems in MFA. Small differences between PFA and MFA might be due to factors associated with sample collection and population histories. Higher diversity observed in PFA compared with MFA is related to the nature of our PFA subjects, which are combination of Pacific Islanders from different geographical sources living in New Zealand.

Allele frequencies for the HPA-1, -2, -5 and -6 loci in PFA and MFA (see Table 1) are similar to those observed in Asian populations, particularly those for Taiwan aboriginals (Seo *et al.*, 1998; Chu *et al.*, 2001; Shih *et al.*, 2003; Halle *et al.*, 2004; Feng *et al.*, 2006; Tan *et*

al., 2012), rather than Europeans (Seller *et al.*, 1999; Bennet *et al.*, 2002; Jones *et al.*, 2003; Robinson *et al.*, 2010). In addition, the presence of the HPA-6b allele in PFA reflects close genetic relationship with populations in East Asia as this allele is only observed among Asian populations (see Chu *et al.*, 2001; Shih *et al.*, 2003; Feng *et al.*, 2006) and in the one previous study on Polynesians (Halle *et al.*, 2004).

However, allele frequencies for HPA-3 locus in PFA are not only different to MFA but also to those previously reported for Ma'ohis Polynesians (Halle *et al.*, 2004). The PFA study group has a low frequency of HPA-3b (0.28), which is similar to indigenous Taiwanese Tsou (0.25; see Chu *et al.*, 2001), but not anywhere near as low as to those recorded for aboriginal communities (0.07-0.06) from Australia (Chen *et al.*, 1997; Bennet *et al.*, 2002). Both, PFA and MFA were observed to have relatively low frequencies of HPA-15a (0.36 and 0.30, respectively) compared with other populations from Asia, Africa and European populations, who often have almost equal values for HPA-15a and -b or higher frequencies of HPA-15a then HPA-15b (see Robinson *et al.*, 2005 for wider comparisons). Allele frequencies at the HPA-15 locus have not been studied for many populations. Hence, we presently lack important reference data for ancestral groups; Taiwan aborigines and P-SA populations. The HPA-15a and -b allele frequencies reported for our PFA group are respectively the lowest and highest values for any characterized population.

We believe that genetic relationships between Polynesians and their putative Mongoloid and Australoid ancestors from elsewhere in Asia and Australasia as well as recent admixture with Europeans are best demonstrated by using PCO plots (Fig. 3), rather than by using methods such as NJ (see Fig. 4), that assume a tree-like evolutionary process. This is shown by occasional inability of NJ procedure to group particular populations into their geographic localities and known population labels, besides showing low bootstrap values. Our PCO analysis shows that Polynesians (i.e. PFA) are distinct from other A-SM populations (Taiwan aborigines, Malays and Filipinos) and their position reflects their degree of shared A-SM/P-SA ancestry. Slight differences between PFA and other Polynesian populations (i.e. MFA and Ma'ohis) on the plots might appear to reflect different proportions of genetic inputs from their ancestors, but more likely arise from genepool refinement via founder effects after multiple re-settlements and long term isolation. In addition, the PCO plots also show evidence for gene flow in MAH as it lies closer to European populations. Overall, the new data collected for PFA and MFA are consistent with the hybrid origin of Polynesians (Kimura *et al.*, 2008; Wollstein *et al.*, 2010; Bellwood *et al.*, 2011) and European admixture in present day Maori is shown by our TM group.

The findings from the present survey show how important it is to recognize that New Zealand is a heterogeneous multi-ethnic country. This is of particular importance because it may have medical consequences. Different prevalence of HPA alleles among Polynesian, Maori and European populations will contribute towards different risk profiles for platelet specific alloimmunization. For instance, mismatch probability of random transfusion is quite high for HPA-1b and -15a among Europeans and HPA-3b and -6b among pairs of PFA individuals (see Table S3). These calculated population specific mismatch probabilities will influence medical thinking regarding the risk of random platelet transfusion. In addition, if one considers Europeans as major blood donors, HPA alloimmunization will only increase for HPA-1b, -5b and -15a in TM, MAH, PFA, MFA and only for HPA-3b in PFA. In contrast, mismatch probability for Europeans will be slightly reduced for any of these HPA systems when transfused with platelets from Maori and Polynesian donors, regardless of group. In addition, alloimmunization against HPA-6b is also potentially developed if platelets from

PFA are transfused to Europeans and Maori (TM/MAH/MFA). Datasets in Table S3 also suggest high probability of developing anti- HPA-3b and -15a antibodies during pregnancy in We note that the probability of developing antibodies against all groups and sub-groups. HPA gene products depends not only on the relative frequencies of the alleles, but also the immunogenicity of the antigens. The possibility of alloimunization during pregnancy is also particularly high for HPA-1b, -5b and -15b among Europeans, for HPA-2b among MAH, for HPA-3a among MFA and for HPA-6b among PFA. For inter-group marriages, the theoretical occurrence of NAIT increases for mixed couples of Europeans with others (TM/MAH/MFA/PFA) for HPA-1b and -15a, Maori (TM/MAH/MFA) and Europeans with PFA for HPA-3b and mixed couples of PFA with others (TM/MAH/MFA/Europeans) for HPA-6b. The probability of transfusion and fetomaternal incompatibility is also supported by the presence of rare HPA genotypes; HPA-1bb, HPA-3bb, and HPA-5bb among Europeans (reviewed in Steffensen et al., 1996; Bennet et al., 2002; Ertel et al., 2005; Hauck-Dlimi et al., 2012) and HPA-6bb in PFA. At first sight, the high incidence of HPA-1a observed in this study means that NAIT caused by maternal antibodies against a mismatched HPA-1a on foetal platelets is likely to be a rare occurrence in Maori and Polynesians, but noting at the same time that blood group O frequency in the range (0.40-0.48) and HLA-DRB3*01 allele is 0.40-0.45, which may reduce one's optimism in this respect (present author unpublished data). This is because HPA-1a alloimmunization is strongly associated with HLA-DRB3*01:01, but not with blood group 0 phenotype (see Williamson et al., 1998; Ahlen et al., 2012).

Overall, data generated from this study have shown the value and ability of DNA-based methods for HPA typing, the prevalence of HPA alleles in Maori and Polynesians and allowed us to use these data for studying ancestry, admixture and to evaluate their clinical consequences. Our findings are of practical relevance for blood transfusion centre protocols and management of pregnancies. Transfusion centres may now wish to recruit HPA- and HLA- typed donors and the use high-throughput molecular methods such as BLOODchip, would be a good way to conduct mass donor screening and typing. Knowledge of the incidence of HPA in different populations helps in the management of possible NAIT.

ACKNOWLEDGEMENT

This research was supported in part by Victoria University of Wellington and Ministry of Higher Education, Malaysia. We would like to thank all the participants for their involvement in this study and Dr Dalice Lim (Victoria University of Wellington) for statistical consultation. We are very much appreciated the technical assistance given by Maximiliano Crosetti (Progenika Biopharma S.A and Grifols International), William Perry (Grifols Australia), Martina Prager (BAG Health Care GmbH, Lich, Germany) and Gabi Henschel (BAG Health Care GmbH, Lich, Germany). We extend our particular thanks to Dr Jordi Bozzo (Progenika Biopharma S.A and Grifols International) for extensive and detailed commentary on the earlier version of our manuscript.

H.E. designed and performed the research, analyzed the data and wrote the paper, P.D. contributed to the research design and writing the paper, R.L. advised on statistics and edited the manuscript and G.C. helped to design and fund the research, managed the DNA Bank and obtained ethical approvals and he also contributed to preparation of the paper.

CONFLICT OF INTEREST

The authors have no competing interests.

REFERENCES

- Ahlen, M.T., Husebekk, A., Killie, M.K., Kjeldsen-Kragh, J., Olsson, M.L. & Skogen, B.
 (2012) The development of severe neonatal alloimmune thrombocytopenia due to anti-HPA-1a antibodies is correlated to maternal ABO genotypes. *Clinical and Developmental Immunology*, Article 156867.
- Avent, N.D. (2007) Large scale blood group genotyping. *Transfusion Clinique et Biologique*, 14, 10-15.
- Bellwood, P. (1978) Man's Conquest of the Pacific: The Prehistory of Southeast Asia and Oceania. Collins, Auckland.
- Bellwood, P., Chambers, G.K., Ross, M. & Hung, C-H. (2011) Are "cultures" inherited?
 Multidisciplinary perspectives on the origins of Austronesian-speaking peoples prior to
 1000 BC. In: *Investigating archaeological cultures: material culture, variability and transmission* (eds. Roberts, B.W & Linden, M.V.), 321-354. Springer, Dordrecht.
- Bennet, J.A., Palmer, L.J., Musk, A.W. & Erber W.N. (2002) Gene frequencies of human platelet antigens 1-5 in indigenous Australians in Western Australia. *Transfudion Medicine*, **12**, 199-203.
- Chambers, G.K. (2006) Polynesian genetics and Austronesian prehistory. In: Austronesian Diaspora and the Ethnogeneses of People in Indonesian Archipelago (eds. Simanjuntak, T., Pojoh, I. & Hisyam, M.), 299-319. Indonesian Institute of Sciences (LIPI), Jakarta.
- Chambers, G.K. (2013) Genetics and the Origins of the Polynesians. In: *Encyclopedia of Life Sciences*. John Wiley & Sons Ltd, Chichester.
- Chambers, G.K. & Edinur, H.A. (2013) Genetic relationships between Malays and Maori. In: *The changing value of Malays, Maori and Pacific Islanders* (eds. Rahman W.R.A. &

Higgins T.R.), 6-37. Chair of Malay Studies & Victoria University Foundation, Wellington, New Zealand.

- Chen, Z., Lester, S., Boettcher, B. & McCluskey, J. (1997) Platelet antigen allele frequencies in Australian aboriginal and Caucasian populations. *Pathology*, **29**, 392-398.
- Chu, C-C., Lee, H-L., Chu, T.W. & Lin, M. (2001) The use of genotyping to predict the phenotypes of human platelet antigens 1 through 5 and of neutrophil antigens in Taiwan. *Transfusion*, **41**, 1553-1558.
- De La Vega Elena, C.D., Nogues, N., Montoya, A.F., Chialina, S., Blanzaco, P.D., Theiller,
 E., Raillon, M.A., Arancegui, N., Solis, E., Oyonarte, S., Ferrer, V.C., Munoz, A.C. &
 Muniz-Diaz, E. (2008) Human platelet-specific antigens frequencies in the Argentinean population. *Transfusion Medicine*, 18, 83-90.
- Edinur, H.A., Dunn, P.P.J., Hammond, L., Selwyn, C., Velickovic Z.M., Lea, R.A. & Chambers, G.K. (2012) Using HLA loci to inform ancestry and health in Polynesian and Maori populations. *Tissue Antigens*, **79**, 557.
- Engelfriet, C.P. & Reesink, H.W. (2003) Detection of platelet-reactive antibodies in patients who are refractory to platelet transfusions, and the selection of compatible donors. *Vox Sanguinis*, **84**, 73-88.
- Ertel, K., Al-Tawil, M., Santoso, S. & Kroll, H. (2005) Relevance of the HPA-15 (Gov) polymorphism on CD109 in alloimmune thrombocytopenic syndromes. *Transfusion*, 45, 366-373.
- Felsenstein, J. (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle.
- Feng, M.L., Liu, D.Z., Shen, W., Wang, J.L., Guo, Z.H., Zhang, X., Du, K.M., Qian, K.C. & Zhao, T.M. (2006) Establishment of an HPA-1-to-16-typed platelet donor registry in China. *Transfusion Medicine*, **16**, 369-374.

- Friedlaender, J.S., Friedlaender, F.R., Reed, F.A., Kidd, K.K., Kidd, J.R., Chambers, G.K., Lea, R.A., Loo, J-H., Koki, G., Hodgson, JA., Merriwether, D.A. & Weber JL. (2008) The genetic structure of Pacific Islanders. *PloS Genetics*, 4, 173-190.
- Halle, L, Bach, K.H., Martageix, C., Bianchi, F., Le T Kim., T., Morel-Kopp, M.C., Soulier,J., Tetaria, C. & Kaplan, C. (2004) Eleven human platelet systems studied in theVietnamese and Ma'ohis Polynesian populations. *Tissue Antigens*, 63, 34–40.
- Hauck-Dlimi, B., Hammon, K., Eckstein, R., Ott, S., Zimmermann, R., Dengler, T. & Ringwald, J. (2012) Human platelet antigen genotypes in Turkish and Caucasian blood donors in Germany. *Tissue Antigen*, **80**, 214-218.
- Jones, D.C, Bunce, M., Fuggle, S.V., Young, N.T. & Marshall, S.E. (2003) Human platelet alloantigens (HPAs): PCR-SSP genotyping of a UK population for 15 HPA alleles. *European Journal of Immunogenetics*, **30**, 415–419.
- Kayser, M. (2010) The human genetic history of Oceania: Near and remote views of dispersal. *Current Biology*, 20, R194-R201.
- Kimura, R., Ohashi, J., Matsumura, Y., Nakazawa, M., Inaoka, T., Ohtsuka, R., Osawa, M. & Takunaga, K. (2008) Gene flow and natural selection in oceanic human populations inferred from genome-wide SNP typing. *Molecular Biology and Evolution*, 25, 1750-61.
- Lubenow, N., Eichler, P., Albrecht, D., Carlsson, L.E., Kothmann, J., Rossocha, W-R., Hahn, M., Quitmann, H. & Greinacher A. (2000) Very low platelet counts in post-transfusion purpura falsely diagnosed as heparin-induced thrombocytopenia: report of four cases and review of literature. *Thrombosis Research*, **100**, 115-125.
- Metcalfe, P., Watkins, N.A., Ouwehand, W.H., Kaplan, C., Newman P., Kekomaki, R., de Haas, M., Aster, R., Shibata, Y., Smith, J., Kiefel V. & Santoso, S. (2003) Nomenclature of human platelet antigens. *Vox Sanguinis*, 85, 240-245.

- Norton, A., Allen, D.L. & Murphy, M.F (2004) Review: platelet alloantigens and antibodies and their clinical significance. *Immunohematology*, **20**, 89-102.
- Ohto, H., Yamaguchi, T., Takeuchi, C., Tohyama, Y., Sato, A., Morita, S. & Collaborative study group (2000) Anti-HPA-5b-induced neonatal alloimmune thrombocytopenia: antibody titre as a predictor. *British Journal of Haematology*, **110**, 223-227.
- Page, R.D.M. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357-358.
- Robinson, J., Mistry, K., McWilliam, H., Lopez, R. & Marsh, S.G.E. (2010) IPD The immune polymorphism database. *Nucleic Acids Research*, **38**, D863-D869.
- Rozman, P. (2002) Platelet antigens. The role of human platelet alloantigens (HPA) in blood transfusion and transplantation. *Transplant Immunology*, **10**, 165-181.
- Seller, J., Thompson, J., Guttridge, M.G. & Darke C. (1999) Human platelet antigens: typing by PCR using sequence-specific primers and their distribution in blood donors resident in Wales. *European Journal of Immunogenetics*, 26, 393-397.
- Seo, D.H., Park, S.S., Kim, D.W., Furihata, K., Ueno, I. & Han, K.S. (1998) Gene frequencies of eight human platelet-specific antigens in Koreans. *Transfusion Medicine*, 8, 129-132.
- Shih, M.C., Liu, T.C., Lin, S.F., Chen, C.M. & Chang J.G. (2003) Gene frequencies of the HPA-1 to HPA-13, Oe and Gov platelet antigen alleles in Taiwanese, Indonesian, Filipinos and Thai populations. *International Journal of Molecular Medicine*, **12**, 609-614.
- Spencer, J.A. & Burrows, R.F. (2001) Feto-maternal alloimmune thrombocytopenia: a literature review and statistical analysis. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, **41**, 45-55.
- Steffensen, R., Kaczan, E., Varming, K. & Jersild, C. (1996) Frequency of platelet-specific alloantigens in a Danish population. *Tissue Antigens*, 48, 93-96.

- Tan, J-Y., Lian, L-H. & Nadarajan, V.S. (2012) Genetic polymorphisms of human platelet antigens-1 to -6, and -15 in the Malaysian population. *Blood Transfusion*, **10**, 368-376.
- Tanaka, S., Taniue, A., Nagao, N., Tomita, T., Ohnoki, S., Shibata, H., Okubo, Y., Yamaguchi H. & Shibata Y. (1996) Genotype frequencies of the human platelet antigen, Ca/Tu, in Japanese, determined by PCR-RFLP method. *Vox Sanguinis*, **70**, 40-44.
- Veldhuisen, B., van der Schoot, C.E. & de Haas, M. (2009) Blood group genotyping: from patient to high-throughput donor screening. *Vox Sanguinis*, **97**, 198-206.
- Watkins, N.A., Smethurst P.A., Allen, D., Smith G.A. & Ouwehand W.H. (2002) Platelet αIIbβ3 recombinant autoantibodies from the B-cell repertoire of a post-transfusion purpura patient. *British Journal of Haematology*, **116**, 677-685.
- Williamson, L.M., Hackett, G., Rennie, J., Palmer, CR., Maciver, C., Hadfield, R., Hughes, D., Jobson, S. & Ouwehand, W.H. (1998) The natural history of fetomaternal alloimmunization to the platelet-specific antigen HPA-1a (PI^{A1}, Zw^a) as determined by antenatal screening. *Blood*, **92**, 2280-2287.
- Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R.J., Nurnberg, P., Stoneking, M. & Kayser, M. (2010) Demographic history of Oceania inferred from genome-wide data. *Current Biology*, **20**, 1983-1992.
- Wu, Y.Y. & Csako, G. (2006) Rapid and/or high-throughput genotyping for human red blood cell, platelet and leukocyte antigens, and forensic applications. *Clinica Chimica Acta*, 363, 165-176.

Supplementary information

Table S1: List of polymorphic HPA genotype profiles and Hardy-Weinberg analysis

		TM (N	=75)		MAH (N=	28)		MFA (N=	47)		PFA (N=	=25)
HPA genotype	0	Ε	HWE analysis	0	Ε	HWE analysis	0	Ε	HWE analysis	0	Ε	HWE analysis
1a/1a	0 920	0.922		0 893	0 896		0 936	0 937		0.960	0 960	
1a/1a	0.080	0.077	$X^2 = 0.130$	0.107	0.101	$X^2 = 0.090$	0.064	0.062	$X^2 = 0.050$	0.940	0.039	$X^2 = 0.010$
1b/1b	0.000	0.002	P = 0.718	0.000	0.003	P = 0.764	0.000	0.001	P = 0.823	0.000	0.000	P = 0.920
29/29	0.880	0.871		0.857	0.820		0.894	0.806		0.760	0.774	
2a/2a 2a/2h	0.880	0.071	$X^2 - 1.530$	0.107	0.829	$X^2 - 3.260$	0.894	0.890	$X^2 - 0.150$	0.700	0.774	$X^2 - 0.460$
2a/20 2b/2b	0.107	0.124	P = 0.216	0.036	0.105	R = 0.071	0.100	0.101	P = 0.150 P = 0.699	0.240	0.211	P = 0.400
20/20	0.015	0.004	I = 0.210	0.030	0.008	I = 0.071	0	0.005	I = 0.099	0.000	0.014	I = 0.498
3a/3a	0.320	0.336		0.393	0.369		0.277	0.318		0.560	0.518	
3a/3b	0.520	0.487	$X^2 = 0.340$	0.429	0.477	$X^2 = 0.290$	0.574	0.492	$X^2 = 1.330$	0.320	0.403	$X^2 = 1.064$
3b/3b	0.160	0.176	P = 0.560	0.179	0.154	P = 0.590	0.149	0.190	P = 0.249	0.120	0.078	P = 0.302
5a/5a	0 973	0 973		0 929	0.930		1 000	1 000	na	0 880	0 884	
5a/5a 5a/5h	0.027	0.026	$X^2 - 0.010$	0.071	0.069	$X^2 - 0.040$	0.000	0.000	11.a	0.000	0.004	$X^2 - 0.100$
5b/5b	0.027	0.020	P = 0.920	0.000	0.001	P = 0.841	0.000	0.000		0.020	0.004	P = 0.752
50/50	0.000	0.000	1 = 0.920	0.000	0.001	1 = 0.011	0.000	0.000		0.000	0.001	1 = 0.752
6a/6a	0.973	0.973		0.929	0.930		1.000	1.000	n.a	0.840	0.810	
6a/6b	0.027	0.026	$X^2 = 0.010$	0.071	0.069	$X^2 = 0.040$	0.000	0.000		0.120	0.180	$X^2 = 2.780$
6b/6b	0.000	0.000	P = 0.920	0.000	0.001	P = 0.841	0.000	0.000		0.040	0.010	P = 0.095
15a/15a	0.093	0.116		0.071	0.092		0.106	0.131		0.120	0.090	
15a/15b	0.493	0.449	$X^2 = 0.740$	0.464	0.423	$X^2 = 0.270$	0.511	0.462	$X^2 = 0.530$	0.360	0.420	$X^2 = 0.510$
15b/15b	0.413	0.436	P = 0.390	0.464	0.485	P = 0.603	0.383	0.407	P = 0.467	0.520	0.490	P = 0.475

HPA: human platelet antigen, N = sample size, HWE: Hardy-Weinberg equilibrium, O: Observed frequency, E: Expected frequencies, X^2 : Chi-squared test, n.a. not applicable, p = significant value (p<0.05), TM: Total Maori, MAH: Maori with Admixed History, MFA: Maori with Full Ancestry and PFA: Polynesians with Full Ancestry.

		HPA-1				HPA-2	
	ТМ	MAH	MFA		TM	MAH	MFA
TM	*			ТМ	*		
MAH	$X^2 = 0.188,$ P = 0.701	*		MAH	$X^2 = 0.538,$ P = 1.000	*	
MFA	$X^2 = 0.111,$ P = 1.000	$X^2 = 0.447,$ P = 0.665	*	MFA	$X^2 = 0.633,$ P = 1.000	$X^2 = 1.705,$ P = 0.527	*
PFA	$X^2 = 0.461,$ P = 0.677	$X^2 = 0.853,$ P = 0.613	$X^2 = 0.177,$ P = 1.000	PFA	X2 = 3.032, P = 0.385	$X^2 = 2.419,$ P = 0.278	$X^2 = 2.251,$ P = 0.173
		HPA-3				HPA-5	
	TM	MAH	MFA		TM	MAH	MFA
TM	*			ТМ	*		
MAH	$X^2 = 0.705,$ P = 0.728	*		MAH	$X^2 = 1.094,$ P = 0.572	*	
MFA	$X^2 = 0.361,$ P = 0.846	$X^2 = 1.556,$ P = 0.471	*	MFA	$X^2 = 1.274,$ P = 0.522	$X^2 = 3.449,$ P = 0.136	*
PFA	$X^2 = 6.222,$ P = 0.045	$X^2 = 2.539,$ P = 0.283	$X^2 = 7.183,$ P = 0.026	PFA	$X^2 = 3.439,$ P = 0.098	$X^2 = 0.365,$ P = 0.658	$X^2 = 5.885,$ P = 0.039
		HPA-6				HPA-15	
	TM	MAH	MFA		TM	MAH	MFA
TM	*			ТМ	*		
MAH	$X^2 = 1.094,$ P = 0.572	*		MAH	$X^2 = 0.271,$ P = 0.898	*	
MFA	$X^2 = 1.274,$ P = 0.522	$X^2 = 3.449,$ P = 0.136	*	MFA	$X^2 = 0.134$, P = 0.964	$X^2 = 0.587,$ P = 0.776	*
PFA	$X^2 = 6.621,$ P = 0.033	$X^2 = 1.567,$ P = 0.493	$X^2 = 7.962,$ P = 0.012	PFA	$X^2 = 0.675,$ P = 0.745	X2 = 0.463, P = 0.789	$X^2 = 0.819,$ P = 0.682

Table S2: Homogeneity tests on genotype frequencies of polymorphic HPA loci observed in Maori and Polynesian study group and sub-groups

HPA: human platelet antigen, TM: Total Maori, MAH: Maori with Admixed History, MFA: Maori with Full Ancestry and PFA: Polynesians with Full Ancestry, p = significant value (p < 0.05).

Table S3: Probability of transfusion and gestation mismatch for Maori, Polynesian and reference European[§] (EUR) populations.

	Transfusion	Gestation
HPA-	TM>TM/MAH/MFA/PFA/EUR	TM>TM/MAH/MFA/PFA/EUR
1a	0.002/0.003/0.001/0.000/0.026	0.002/0.003/0.001/0.000/0.025
1b	0.072/0.070/0.073/0.075/0.055	0.037/0.036/0.037/0.038/0.028
2a	0.004/0.008/0.003/0.006/0.006	0.004/0.007/0.003/0.006/0.005
2b	0.112/0.107/0.116/0.109/0.110	0.058/0.055/0.060/0.056/0.057
3a	0.145/0.127/0.157/0.065/0.115	0.102/0.090/0.110/0.045/0.081
3b	0.223/0.245/0.211/0.344/0.261	0.141/0.155/0.134/0.218/0.165
5a	0.000/0.001/0.000/0.004/0.007	0.000/0.001/0.000/0.004/0.007
5b	0.026/0.025/0.026/0.023/0.022	0.013/0.012/0.013/0.012/0.011
ба	0.000/0.001/0.000/0.010/0.000	0.000/0.001/0.000/0.010/0.000
6b	0.026/0.025/0.026/0.021/0.026	0.013/0.012/0.013/0.011/0.013
15a	0.246/0.274/0.230/0.277/0.134	0.148/0.165/0.139/0.167/0.081
15b	0.102/0.082/0.116/0.080/0.232	0.076/0.061/0.086/0.059/0.173
HPA-	MAH>TM/MAH/MFA/PFA/EUR	MAH>TM/MAH/MFA/PFA/EUR
1a	0.002/0.003/0.001/0.000/0.026	0.002/0.003/0.001/0.000/0.024
1b	0.096/0.093/0.098/0.100/0.073	0.049/0.048/0.050/0.051/0.038
2a	0.004/0.008/0.003/0.006/0.006	0.004/0.007/0.003/0.006/0.005
2b	0.149/0.141/0.153/0.144/0.146	0.078/0.074/0.080/0.076/0.076
3a	0.149/0.131/0.161/0.066/0.118	0.107/0.094/0.116/0.048/0.084
3b	0.212/0.233/0.201/0.327/0.248	0.132/0.145/0.125/0.204/0.154
5a	0.000/0.001/0.000/0.004/0.007	0.000/0.001/0.000/0.003/0.007
5b	0.068/0.065/0.070/0.062/0.059	0.035/0.033/0.036/0.032/0.030
6a	0.000/0.001/0.000/0.010/0.000	0.000/0.001/0.000/0.010/0.000
6b	0.068/0.065/0.070/0.057/0.070	0.035/0.033/0.036/0.029/0.036
15a	0.224/0.250/0.210/0.252/0.123	0.132/0.147/0.124/0.149/0.072
15b	0.105/0.084/0.119/0.082/0.238	0.081/0.064/0.091/0.063/0.183
HPA-	MFA>TM/MAH/MFA/PFA/EUR	MFA>TM/MAH/MFA/PFA/EUR
1a	0.002/0.003/0.001/0.000/0.026	0.002/0.003/0.001/0.000/0.025
1b	0.058/0.056/0.059/0.060/0.044	0.029/0.029/0.030/0.031/0.022
2a	0.004/0.008/0.003/0.006/0.006	0.004/0.008/0.003/0.006/0.005
2b	0.090/0.086/0.093/0.088/0.089	0.046/0.044/0.048/0.045/0.046
3a	0.143/0.125/0.154/0.063/0.113	0.099/0.087/0.107/0.044/0.078
3b	0.229/0.251/0.217/0.354/0.268	0.147/0.161/0.139/0.266/0.171
5a	0.000/0.001/0.000/0.004/0.007	0.000/0.001/0.000/0.004/0.007
5b	0.000/0.000/0.000/0.000/0.000	0.000/0.000/0.000/0.000/0.000
6a	0.000/0.001/0.000/0.010/0.000	0.000/0.001/0.000/0.010/0.000
6b	0.000/0.000/0.000/0.000/0.000	0.000/0.000/0.000/0.000/0.000
15a	0.258/0.287/0.241/0.290/0.141	0.158/0.175/0.147/0.177/0.086
15b	0.100/0.080/0.114/0.078/0.228	0.074/0.059/0.084/0.057/0.167

Table S3, cont.

	Transfusion	Gestation
HPA-	PFA>TM/MAH/MFA/PFA/EUR	PFA>TM/MAH/MFA/PFA/EUR
1a	0.002/0.003/0.001/0.000/0.026	0.002/0.003/0.001/0.000/0.025
1b	0.036/0.035/0.037/0.038/0.028	0.018/0.018/0.019/0.019/0.014
2a	0.004/0.008/0.003/0.006/0.006	0.004/0.007/0.003/0.006/0.005
2b	0.134/0.127/0.138/0.130/0.131	0.070/0.066/0.072/0.068/0.068
3a	0.163/0.142/0.175/0.072/0.189	0.127/0.111/0.137/0.056/0.100
3b	0.162/0.178/0.153/0.250/0.189	0.094/0.103/0.089/0.145/0.110
5a	0.000/0.001/0.000/0.004/0.007	0.000/0.001/0.000/0.003/0.007
5b	0.113/0.108/0.116/0.103/0.097	0.058/0.056/0.060/0.053/0.050
ба	0.000/0.001/0.000/0.010/0.000	0.000/0.001/0.000/0.009/0.000
6b	0.185/0.177/0.190/0.154/0.190	0.097/0.093/0.100/0.081/0.100
15a	0.222/0.247/0.208/0.250/0.121	0.131/0.146/0.122/0.147/0.071
15b	0.105/0.084/0.119/0.082/0.239	0.081/0.065/0.092/0.063/0.184
HPA-	EUR>TM/MAH/MFA/PFA/EUR	EUR>TM/MAH/MFA/PFA/EUR
1a	0.002/0.003/0.001/0.000/0.025	0.001/0.002/0.001/0.000/0.022
1b	0.273/0.265/0.277/0.284/0.208	0.148/0.144/0.151/0.155/0.113
2a	0.004/0.008/0.003/0.006/0.006	0.004/0.007/0.003/0.006/0.005
2b	0.126/0.120/0.129/0.122/0.124	0.065/0.062/0.067/0.063/0.064
3a	0.152/0.133/0.164/0.067/0.120	0.111/0.097/0.119/0.049/0.087
3b	0.204/0.224/0.193/0.315/0.239	0.125/0.137/0.119/0.193/0.147
5a	0.000/0.001/0.000/0.004/0.007	0.000/0.001/0.000/0.003/0.007
5b	0.160/0.153/0.165/0.145/0.138	0.084/0.080/0.086/0.076/0.072
6a	0.000/0.001/0.000/0.010/0.000	0.000/0.001/0.000/0.010/0.000
6b	0.000/0.000/0.000/0.000/0.000	0.000/0.000/0.000/0.000/0.000
15a	0.332/0.370/0.310/0.373/0.181	0.223/0.248/0.209/0.251/0.122
15b	0.085/0.068/0.097/0.066/0.193	0.056/0.045/0.064/0.044/0.128

HPA: human platelet antigen, TM: Total Maori, MAH: Maori with Admixed History, MFA: Maori with Full Ancestry and PFA: Polynesians with Full Ancestry, \$ = HPA data for Europeans (EUR) obtained from Jones *et al.*, 2003 (HPA-1 to -6) and Ertel *et al.*, 2005 (HPA-15), >: donor to recipient for transfusion and father to mother populations for gestation.

KIR and HLA diversity in Polynesians and Maori of New Zealand

H.A. Edinur^a, P. J. Norman^b, , P.P.J. Dunn^c, and G.K. Chambers^a, P. Parham^{b,¶}

^aSchool of Biological Sciences, Victoria University of Wellington, New Zealand

^bDepartment of Structural Biology, Stanford University School of Medicine, Stanford, CA,

USA

^cTissue Typing Laboratory, New Zealand Blood Service, Auckland

Status: Submission ready for thesis

Journal: *Immunogenetics* (format)

Key words: Polynesians, Maori, KIR, HLA, A and B haplotypes

Corresponding author: Peter Parham, Department of Structural Biology, Stanford University, 299 Campus Drive West, Fairchild D-157, Stanford, CA 94305, USA.

e-mail: peropa@stanford.edu

Abstract

The KIR gene region is highly polymorphic and variability is shown in both haplotype-gene content and at the allelic level. In this respect, the KIR system variations reflect that observed for the highly polymorphic HLA class I loci, where products are ligands for KIR. In the present study, we examine KIR gene content in a set of Polynesian and Maori individuals previously genotyped for HLA loci. The haplotype A homozygote was the most common KIR genotype profile observed which is consistent with high frequencies (0.87-1.00) of individual KIR genes that make up the A haplotype. Other profiles observed were AB heterozygote haplotypes involving mostly combinations of centromeric (KIR2DL2-2DS2) and telomeric (2DL5-2DS1, 2DS1-3DS1 and 2DS5-3DS1) B pairs of loci, which were in strong positive association. A principal component plot constructed using KIR carrier frequency shows that geographically proximal populations tend to cluster together fairly well reflecting the shared history among populations. However, KIR variation at the population level may also reflect adaptation to local selective forces. Therefore, we compared correlations of KIR and HLA ligand frequencies in order to avoid potential issues of nonindependence due to the shared genetic history among the populations. There were significant positive correlations between KIR3DLI and HLA-Bw4/Bw4-80I and significant negative correlations for pairs of 3DSI-Bw4/Bw4-80I, 2DL2-C1 and 2DS1-C2. Overall, our research findings together with subsequent analyses provide a valuable perspective on Polynesian immunogenetics and support the view that global patterns of KIR and HLA diversity are shaped by both population history and adaptation by natural selection.

Chapter 3.5

Introduction

Killer cell immunoglobulin-like receptors (KIR) are transmembrane glycoproteins expressed by natural killer cells. These receptor proteins are classified based on the number of extracellular immunoglobulin domains (either 2D; 2 domains or 3D; 3 domains) and the characteristics of their cytoplasmic domain (S; short or L; long). KIR with the long cytoplasmic domains contain an immunoreceptor tyrosine-based inhibition motif and transduce inhibitory signals, while those with the short cytoplasmic domain transmit an activating signal through association with DAP-12 molecule (Wilson et al. 1997; Carrington and Norman 2003).

The KIR glycoproteins are encoded by genes on human chromosome 19q13.4 (Suto et al. 1998). The KIR genes are highly polymorphic and variability is shown in both, haplotype gene content and at the allelic level (reviewed by Middleton and Gonzelez 2010). A total of 16 KIR genes have been identified and can be divided into two haplotype groups, known as A and B. The four framework KIR genes (KIR3DL3, 3DP1, 2DL4 and 3DL2) are common to both haplotypes. The A haplotype consists of KIR2DL3, 2DP1, 2DL1, 3DL1 and 2DS4 genes whereas haplotype B is assigned based on the present of KIR3DS1, 2DL5, 2DS1, 2DS5, 2DL2, 2DS2 and 2DS3 genes (Rajalingam et al. 2008). To date, there are more than 600 different alleles that have been reported across all of the 16 KIR loci (see Robinson et al. 2010; Gonzalez-Galarza et al. 2011). In this respect, the observed diversity in the KIR system is similar to that observed for the highly polymorphic human leukocyte antigen (HLA) class I, ligand for KIR. The high polymorphism of KIR and HLA class I loci are thought to have co-evolved with distinct KIR-HLA ligand combinations being associated in pathogenesis of autoimmune diseases, virus infections, tumours, fertility and the success of

227

transplantation (see Ruggeri *et al.*, 2002; Yawata et al. 2006; Single et al. 2007; Ahlenstiel et al. 2008; Kurkani et al. 2008; Moesta and Parham, 2012; Parham et al. 2012a,b; Parham and Moffett 2013). Therefore, evolutionary forces acting effectively at the population level may simultaneously influence allele and haplotype frequency for HLA class I and KIR loci.

In recent years, several large scale studies have been conducted to explore worldwide patterns of KIR variation. These provide exceptional value for authenticating population history (e.g. see Hollenbach et al. 2010; Hollenbach et al. 2012). This study is the first KIR gene content screening in Maori. To date, the only KIR data available for Polynesian sub-populations (Cook Islands, Tonga, Samoa and Tokelau) are those reported by Velickovic et al. (2006). The present study is intended to examine gene content polymorphism in previously HLA genotyped Polynesian and Maori individuals (see Edinur et al. 2012). The newly collected KIR data will be used for ancestry study and can also be used as reference medical standard for disease association studies in Maori and Polynesians.

Materials and Methods

Study population group and sub-groups

Genomic DNA samples from 108 unrelated Maori (Total Maori = TM) and 48 Polynesian (TP = Total Polynesians) unrelated individuals were drawn from the Victoria University of Wellington DNA Bank. The larger Maori and Polynesian groups were then divided into a pair of smaller sub-groups based on a self-defined pedigree: Maori with Full Ancestry (MFA: N = 50) and Polynesians with Full Ancestry (PFA: N = 34), after Edinur et al. (2012). Ethical approvals were obtained from New Zealand Central Region Ethics Committee and VUW Human Ethics Committee.

KIR and HLA genotyping

KIR genotyping was carried out using polymerase chain reaction-sequence specific oligonucleotides (PCR-SSO) using Luminex (LABType[®] SSO KIR Typing Test, One Lambda[®], CA, USA). HLA genotyping was performed using Luminex and sequence based typing as previously described and data was presented in Edinur et al. (2012) and Edinur et al. (2013).

Statistical analysis

Observed carrier frequency (F) was determined by direct counting. Gene frequency (GF) was calculated using the formula $GF = 1 - \sqrt{1 - F}$. Group A and B haplotypes were predicted according to Rajalingam et al. (2008). Chi-squared tests (SPSS Inc., Chicago, IL) were used to test for Hardy-Weinberg equilibrium (HWE) while linkage disequilibrium between pairs of KIR genes were estimated using Wn or Cramer's V statistic, as previously described by Single et al. (2008). The correlation coefficient is equivalent to Wn, the summary of relationship between two loci. The Wn has a range from -1 to +1, which depends on the direction of association. The statistical significance of LD between a pair of loci was then estimated using Fisher's exact test. Differences in carrier KIR frequencies between Polynesian datasets were evaluated using Chi-squared tests with Yate's continuity correction available Graphpad in the software (http://www.graphpad.com/quickcalcs/contingency1.cfm). A principal component analysis (PCA) was performed using Multivariate Statistical Software Package 3 (Kovach Computing Services, UK; http://www.kovcomp.com/mvsp). Correlations between KIR and HLA ligand carrier frequencies were examined using T-tests on transformed correlation coefficients. We performed the labels for the KIR genes and calculated the correlation between the HLA and KIR genes for each random permutation. A total of 10,000 permutations were applied. The *p*-value for each correlation was calculated as a percentage of the 10,000 correlation runs.

229

Thus calculated values that were as large as or larger than the one observed. This *p*-value is a two-tailed permutation *p*-value.

Results

Genotype profiles and haplotype assignments for study population groups and sub-groups are illustrated in Fig. 1. A total of 38 different genotype profiles were observed in the present study; 33 in TM, 21 in MFA, 19 in TP and 14 profiles for PFA and no significant departure from HWE was observed for these study population groups and sub-groups. Thirty of these profiles have previously been reported in allele frequency net database (www.allelefrequencies.net). For each genotype, there are more inhibitory than stimulatory genes. The homozygous haplotype A genotype is the most common form and the highest frequency was observed in PFA (0.50). All the 16 previously known types of KIR genes were detected in the present study and their observed carrier and estimated gene frequencies are listed in Table 1. Overall, frequencies of inhibitory (KIR2DL1, 2DL3, 3DL1, 2DL4, 3DL3) and pseudo-genes (KIR2DP1 and 3DP1) are higher than stimulatory genes (KIR2DS1, 2DS2, 2DS3, DS5 and 3DS1), except for KIR2DS4 which was present in more than 97% of individuals. As expected, the framework genes (KIR3DL3, 2DL4, 3DP1 and 3DL2) were present in almost all of the individuals tested. Linkage disequilibrium (LD) analysis shows strong positive associations for the gene pairs KIR2DS2-2DL2, 2DL5-2DS1, 2DS5-3DS1 and 2DS1-3DS1 (see Table S1a-d). The framework genes were excluded from the LD analysis because they are present in all samples. In addition, KIR2DP1 and 2DL1 were present in all TP group and PFA sub-group individuals and also excluded from LD analysis.
The KIR carrier frequencies and genotype profiles obtained in the present study were compared with those of previously studied Polynesian sub-populations; see Table S2 and Velickovic et al. (2006) for details. The four Polynesian sub-populations are generally not significantly different between each other except for the KIR2DS4 gene. Nonetheless, their pool gene frequency as Velickovic Total Polynesians (VTP) is perfectly matched with our new collected data for PFA (see Table S3a-g). Therefore, we combined PFA with VTP, forming a larger dataset, PMD (Polynesian meta dataset) for wider population comparisons. Statistical tests showed our MFA is similar to PFA but is different with VTP only for frequency of KIR2DS4. Although our TM group is only different with MFA for KIR3DS1, the former is significantly different with VTP and PFA (see Table S3a-g). In addition, comparisons of the KIR profiles showed quite a marked differences between our partially admixed groups (i.e. TP and TM) and the full ancestry sub-groups (i.e. PFA and MFA). Only 21 and 14 KIR gene profiles observed in MFA and PFA (respectively), compared with 33 and 19 in TM and TP (respectively). We believe this reflects European admixture effects in TP and TM. Admixture effects are not so obvious in TP compared with TM. This might be due to the smaller sample size of the former group. Therefore, we used PMD and MFA to represent the genepools of Polynesians and Maori (respectively) for tracing history, while those from TP and TM represent modern day Polynesians and Maori living in New Zealand. We used PCA for a wider comparison of KIR distribution in Polynesians and other populations. The PCA plot constructed using KIR carrier frequencies (see Fig.2) demonstrated that most populations cluster together according to their geographic locations.

Finally, we examined the correlation between each combination of KIR-HLA ligands using reference populations (see Table S4) reported by Single et al. (2007). We observed significant positive correlation between KIR3DLI and HLA-Bw4/ Bw4-80I and significant

negative correlation for pairs of KIR3DSI and HLA-Bw4/Bw4-80I, KIR2DL2 and HLA-C group 1 ligand and KIR2DS1 and HLA-C group 2 (see Table 2, Fig. S1a-i).

ID	HG	Fra	meworl	k and p	seudog	gens	Н	aplotyp	A gene	s			Haplo	type B	genes			тр	DEA	тм	МЕА
ID	щ	3D	3D	2D	3D	2D	2D	2D	3D	2D	2D	2D	2D	2D	2D	2D	3D	11	IIA	1 101	WII A
		L2	L3	L4	P1	P1	L1	L3	L1	S4	S1	S2	S 3	S5	L2	L5	S1				
1	AA																	22	17	37	18
N6	AA																	0	0	1	0
200	AB																	0	0	1	1
433	AB																	1	1	0	0
267	AB																	1	0	0	0
55	AB																	0	0	1	0
4	AB																	3	0	20	6
31	AB																	2	2	1	1
5	AB																	2	1	3	2
22	AB																	0	0	2	1
33	AB																	1	0	2	0
2	AB																	2	2	3	1
8	AB																	2	2	4	4
28	AB																	0	0	1	1
381	AB																	0	0	2	0
9	AB																	0	0	1	0
3	AB																	2	2	7	3
11	AB																	1	1	0	0
7	AB																	1	1	1	1
6	AB																	2	1	2	1
242	AB																	0	0	1	1

Fig. 1: Genotype profile observed in the study population groups and sub-groups. Haplotype ID was assigned according to allele frequency net database (Gonzalez-Galarza et al. 2011). Abbreviations: HG: haplogroup, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry.

Ш	HC	Fra	amewor	k and p	seudog	ens	Н	aplotyp	A gene	s			Haplo	type B	genes			тр	ΡΕΛ	тм	ΜΕΔ
Ш	по	3D	3D	2D	3D	2D	2D	2D	3D	2D	2D	2D	2D	2D	2D	2D	3D	11	ITA	1 101	WIT'A
		L2	L3	L4	P1	P1	L1	L3	L1	S4	S1	S2	S3	S5	L2	L5	S1				
N4	AB																	1	0	0	0
N3	AB																	1	1	0	0
N2	AB																	0	0	1	1
N1	AB																	0	0	1	1
N5	AB																	0	0	1	0
81	BB																	0	0	2	1
176	BB																	0	0	1	0
71	BB																	0	0	2	1
90	BB																	0	0	1	0
73	BB																	2	1	1	1
75	BB																	1	1	0	0
87	BB																	1	1	1	0
74	BB																	0	0	1	1
N9	BB																	0	0	1	0
N7	BB																	0	0	2	1
76	BB																	0	0	2	2
172	BB																	0	0	1	0
							Total	number	r of ind	ividuals								48	34	108	50
							Frequ	uencies	of haple	otype A								0.69	0.71	0.61	0.61
							Frequ	uencies	of haple	otype B								0.31	0.29	0.39	0.39

Fig. 1 cont.

KIR	TP (N-48)	PFA (N-34)	TM (N-108)	MFA (N=50)
	(11-40)	(11-34)	(14=108)	(11-30)
	Phe	enotype freque	ency	
3DL3	1	1	1	1
2DS2	0.38	0.32	0.54	0.50
2DL2	0.38	0.32	0.54	0.50
2DL3	0.96	0.97	0.87	0.86
2DL5	0.48	0.50	0.44	0.52
2DS3	0.29	0.29	0.20	0.28
2DS5	0.23	0.24	0.25	0.28
2DP1	1	1	0.91	0.88
2DL1	1	1	0.94	0.92
3DP1	0.96	0.97	0.97	0.98
2DL4	1	1	1	1
3DL1	0.96	0.94	0.95	0.96
3DS1	0.31	0.29	0.64	0.42
2DS1	0.31	0.35	0.34	0.40
2DS4	0.98	0.97	0.99	0.98
3DL2	1	1	1	1
	Ge	notype freque	ncy	
3DL3	1	1	1	1
2DS2	0.21	0.18	0.32	0.29
2DL2	0.21	0.18	0.32	0.29
2DL3	0.80	0.83	0.64	0.63
2DL5	0.28	0.29	0.26	0.31
2DS3	0.16	0.16	0.11	0.15
2DS5	0.12	0.13	0.13	0.15
2DP1	1	1	0.70	0.65
2DL1	1	1	0.76	0.72
3DP1	0.80	0.83	0.83	0.86
2DL4	1	1	1	1
3DL1	0.80	0.76	0.79	0.80
3DS1	0.17	0.16	0.40	0.24
2DS1	0.17	0.20	0.19	0.23
2DS4	0.86	0.83	0.90	0.86
3DL2	1	1	1	1

Table 1: Phenotype and genotype frequencies in study population groups and sub-groups

Abbreviations: TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, N = sample number.



Fig. 2: PCA constructed using KIR2DL1-5, 3DL1-3, 2DS1-5 and 3DS1 carrier frequencies. Reference populations were obtained from Hollenbach et al. (2010), Hollenbach et al. (2012), Gonzalez-Galarza et al. (2011), Taniguchi and Kawabata (2009), Velickovic et al. (2006), Velickovic et al. (2008) and Velickovic et al. (2010).

KIR and HLA pair	Correlation (<i>r</i>)	Permutation <i>p</i> -value	
3DL1-Bw4	0.36	0.02	
3DL1-Bw480I	0.40	0.01	
3DS1-Bw4	-0.53	0	
3DS1-Bw480I	-0.60	0	
2DL1-C group 02	0.05	0.77	
2DL2-C group C1	-0.35	0.04	
2DL3-C group C1	0.23	0.18	
2DS1-C group 02	-0.45	0.01	
2DS2-C group C1	-0.33	0.06	

Table 2: Correlation between KIR and HLA carrier frequencies

Discussion

A detailed account of Polynesian origins and a chronology of their associated archaeology is presented by Bellwood et al. (2011). In short, the Austronesian-speaking Mongoloids (A-SM) from Taiwan voyaged to Melanesia via the Philippines and Borneo. Gender-biased genetic admixture occurred between A-SM and residents of coastal Papua New Guinea and the islands of the Bismarck Archipelago (i.e. Papuan-speaking Australoid: PSA), before the first moved outwards to Remote Oceania (Polynesia) and occupied New Zealand from 650 ybp. The general account above has received independent confirmation from a Bayesian phylogenetic analysis of Pacific and East Asian languages (Gray *et al.*, 2009) and large-scale independent molecular phylogenetic studies (Kimura et al. 2008 and Woolstein et al. 2010).

We have previously shown evidence for the hybrid origin of Polynesian populations by using HLA data (Edinur et al. 2012 and Edinur et al. 2013) and thus support the account (Chambers 2006; Bellwood et al. 2011). Here, we extend this research and characterize KIR loci in the same sets of samples (see Table 1 and Fig. 1). Statistical analyses (see Table S3a-g) of the KIR data showed a close genetic relationship between our Polynesians full ancestry sub-

group (i.e. PFA) and the previous study on the four (Cook Island, Samoa, Tonga and Tokelau) Polynesian sub-populations (Zlatibor et al. 2006). Therefore, we combined them to create a larger dataset (i.e. PMD) for Polynesians. In contrast, there are significant differences between our Maori full ancestry sub-group (i.e. MFA) and the larger admixed group (i.e. TM). This might be due to European admixture in TM as previously discovered (Edinur et al. 2012). The admixture is not so obvious in TP and this might reflect the effects of smaller sample size represented by this study group. All things considered, we used MFA and PMD for genealogical studies while those represented by TP and TM represented the genepools of modern day Polynesians and Maori living in New Zealand.

In wider context, distribution of KIR genotypes among Maori and Polynesians is most similar to Asian populations. These have been often reported to have slightly higher frequency of haplotype A and genotype AA homozygote profiles, compared with Europeans, African and P-SA populations (see Yawata et al. 2002; Whang et al. 2005; Rajalingam et al. 2008; Ashouri et al. 2009; Hollenbach et al. 2010; Hollenbach et al. 2012; John et al. 2012). Positions of Polynesian and Maori KIR datasets in a wider geographical context are shown in Fig. 2. In the PCA plot, geographically proximal populations tend to cluster together fairly well. As expected, Oceanian populations form two well separated groups which represent at least two different lineages of people; A-SM and P-SA plus others (Polynesians and Melanesians) who have greater or lesser influence from these two putative ancestral populations. Here, Maori and Polynesians are close to Asian populations and reflect more A-SM influence than P-SA/Melanesians. Some of the HLA loci also showed a greater A-SM ancestral fraction (i.e. HLA-B and -DRB1) in Polynesians (~100:0 genetic proportions of ASM:P-SA ancestral fractions) compared with 60:40 for HLA-A and -C loci (Edinur et al. 2012). Overall, KIR gene content frequency may appear to be rather less directly informative

Chapter 3.5

compared with HLA or SNP data for historical reconstructions, but it may also reflect natural selection (see later). In addition, it is hard to make an effective comparison because most KIR population data are not reported for all the 16 KIR genes and few studies are reported at the allelic level.

Population level variation of KIR may reflect local adaptation especially for these medically important genes. Studies have shown selective forces on both KIR gene content and allelic variants (Single et al. 2007; Norman et al. 2007). However, testing for natural selection may be confounded by issues such as non-independence and shared history among populations, as demonstrated above (see Fig. 2). This difficulty can be resolved by application of two unlinked loci (e.g. KIR and HLA) and using a permutation-based approach (refer Single et al. 2008 for details). Based on this statistical method, the two unlinked loci are assumed to have experienced the same demographic history and the permutation-based method is used to generate a null hypothesis of no functional relationship between the two loci. We therefore examined correlation between frequencies of KIR and their HLA ligand (see Table 2, Fig. S1a-i). Generally, there is evidence of a positive correlation between inhibitory KIR-HLA ligand combinations and negative correlation between activating KIR-HLA ligands in worldwide populations. This might be due to selection pressures that favour more inhibitory KIR-HLA complex to avoid auto-reactivity. The significant negative correlation observed between the KIR2DL2 and HLA-C group 1 ligand is conflicting with a positive correlation of other inhibitory KIR-HLA complexes. The HLA-C group 1 is also a ligand for the more common inhibitory KIR2DL3 and negative correlation between KIR2DL2 and HLA-C group 1 ligand might be compromised by the positive correlation of KIR2DL3 and HLA-C group 1 (see also Single et al. 2007 and Hollenbach et al. 2012). Functional study has shown that the

presence of one inhibitory KIR-HLA complex is sufficient for self-tolerance (Valiante et al. 1997).

In Polynesians and Maori, each individual has at least one inhibitory KIR that recognize self HLA class I molecule. This is evident from the predominance of haplotype A (0.61-0.71) among Polynesians and Maori (see Fig. 1). Thus, selection in favour of inhibitory KIR genes and their ligands may provide protection from autoimmune diseases (e.g. psoriasis, and rheumatoid vasculitis) in Polynesians and Maori. This observation reflects our inference based on the HLA data (see Edinur et al. 2013). Nonetheless, high frequencies of haplotype A homozygote individuals (0.36 - 0.50) among Maori and Polynesians may mean that they are vulnerable to infectious diseases, cancer and have decreased rate of reproduction success. These individuals possibly possess no stimulatory KIR (due to the predominance of null 2DS4 allele), which are implicated with reduced risk of infectious (e.g. human immunodeficiency virus, tuberculosis and human cytomegalovirus) and reproductive (e.g. preeclampsia) diseases (reviewed in Kulkarni et al. 2008). Hiby et al. (2004) have shown that a combination of maternal homozygote for haplotype A and fetal with HLA-group C2 is associated with an occurrence of preeclampsia. They hypothesized that an inverse relationship between the KIR AA genotype and HLA-group C2 frequencies in various populations as a mechanism of selective pressures for maintaining reproductive success. However, haplotype A homozygote and HLA-C2 frequencies are in balance among Polynesians (0.35-0.50 vs 0.39-0.52), as compared with Australian aborigines (0.02 vs 0.60) and Asian populations (0.35-0.60 vs 0.10-0.20). Evolutionary processes due to natural selection might be violated in Polynesians due their A-SM:PSA dual origin and the clinical consequences of this admixture could be an important subject of disease association study (see Edinur et al. 2012). In addition, KIR has been shown to be of relevance in transplantation medicine as the success rate is also dependent on the degree of KIR matching between donor and recipient. The Maori and Polynesians seem to share a number of common and unique KIR gene content profiles and this could facilitate the search of donors for allograft transplantation, but see Ruggeri et al. (2002).

Overall, our research findings together with subsequent analyses support the view that global patterns of KIR and HLA diversity are shaped by migration patterns, demographic history as well as human adaptation toward episodes of natural selection (Sanchez-Mazas et al. 2011). We believe that the KIR data reported in this study provides increased understanding of Polynesian immunogenetics and could be used as a medical reference standard for Polynesians and Maori living in New Zealand.

Acknowledgments

The authors acknowledge Dr Dalice Sim (Victoria University of Wellington), Dr Zlatibor M Velickovic (Department of Cell & Molecular Therapies, Royal Prince Alfred Hospital and Medical School, University of Sydney, Australia) and Dr Eduardo Jose Melo dos Santos (Human and Medical Genetics, Federal University of Para) for bioinformatics consultation.

Funding

This study was supported by the Stanford University School of Medicine, Stanford, New Zealand Blood Service, Auckland, Victoria University of Wellington and Ministry of Higher Education, Malaysia.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Ahlenstiel G, Martin MP, Gao X, Carrington M, Rehermann B (2008) Distinct KIR/HLA compound genotypes affect the kinetics of human antiviral natural killer cell responses. J Clin Invest 118(3): 1017-1026. doi: 10.1172/JCI32400
- Ashouri E, Farjadian S, Reed EF, Ghaderi A, Rajalingam R (2009) KIR gene content diversity in four Iranian populations. Immunogenetics 61(7):483-492. doi: 10.1007/s00251-009-0378-7
- 3. Bellwood P, Chambers GK, Ross M, Hung C-H (2011) Are "cultures" inherited? Multidisciplinary perspectives on the origins of Austronesian–speaking peoples prior to 1000 BC. In: Roberts BW, Linden MV (eds) Investigating archaeological cultures: material culture, variability and transmission, Springer, Dordrecht, pp 321-354
- Carrington M, Norman P (2003) The KIR Gene Cluster. National Center for Biotechnology Information (US), Bethesda (MD). http://www.ncbi.nlm.nih.gov/books/NBK10134. Accessed 19 February 2011
- Chambers GK (2006) Polynesian genetics and Austronesian prehistory. In: Simanjuntak T, Pojoh I, Hisyam M (eds) Austronesian diaspora and the ethnogeneses of people in Indonesian Archipelago, Indonesian Institute of Sciences (LIPI), Jakarta, pp 299-319
- Edinur HA, Dunn PJ, Hammond L, Selwyn C, Velickovic ZM, Lea RA, Chambers GK (2012) Using HLA loci to inform ancestry and health in Polynesian and Maori populations. Tissue Antigens 80(6):509-522. doi: 10.1111/tan.12026
- 7. Edinur HA, Dunn PPJ, Hammond L, Selwyn C, Askar M, Velickovic ZM, Lea RA & Chambers GK (2013) HLA and MICA polymorphism in Polynesians and New

Zealand Maori: Implications for ancestry and health. *Human Immunology. Re-submitted.*

- Gray RD, Drummond AJ, Greenhill SJ (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science: 323: 479-483. doi: 10.1126/science.1166858
- Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 39(Database issue): D913-919. doi: 10.1093/nar/gkq1128
- 10. Hiby SE, Walker JJ, O'Shaughnessy KM, Redman CWG, Carrington M, Trowsdale J, Moffett A (2004) Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. J Exp Med 200(8):957-965. doi: 10.1084/jem.20041214
- 11. Hollenbach JA, Meenagh A, Sleator C, Alaez C, Bengoche M, Canossi A, Contreras G, Creary L, Evseeva I, Gorodezky C, Hardie R.-A, Karlsen T.H, Lie B, Luo M, Martinetti M, Navarette C, de Oliveira DC, Ozzella G, Pasi A, Pavlova E, Pinto S, Porto LC, Santos P, Slavcev A, Srinak D, Tavoularis S, Tonks S, Trachtenberg E, Vejbaesya S, Middleton D (2010) Report from the killer immunoglobulin-like receptor (KIR) anthropology component of the 15th International Histocompatibility Workshop: worldwide variation in the KIR loci and further evidence for the co-evolution of KIR and HLA. Tissue Antigens 76(1):9-17. doi: 10.1111/j.1399-0039.2010.01459
- 12. Hollenbach JA, Nocedal I, Ladner MB, Single RM, Trachtenberg EA (2012) Killer cell immunoglobulin-like receptor (KIR) gene content variation in the HGDP-

CEPH populations. Immunogenetics 64(10):719-737. doi: 10.1007/s00251-012-0629-x

- 13. John E, Christiansen FT, Mueller I, Schofield L, Senitzer D, Siba P, Witt CS (2012)
 Distinct distribution of killer-cell immunoglobulin-like receptor genes in the
 Mugil and Ilaita areas of Papua New Guinea. Tissue Antigens 79(4), 263-271.
 doi: 10.1111/j.1399-0039.2012.01848.x
- 14. Kimura R, Ohashi J, Matsumura Y, Nakazawa M, Inaoka T, Ohtsuka R, Osawa M, Takunaga K (2008) Gene flow and natural selection in Oceanic human populations inferred from genome-wide SNP typing. Mol Biol Evol 25(8):1750-61. doi: 10.1093/molbev/msn128
- 15. Kurkani S, Martin MP, Carrington M (2008) The Yin and Yang of HLA and KIR in human disease. Semin Immunol 20(6): 343-352. doi: 10.1016/j.smim.2008.06.003
- 16. Martin AS, Freitas EM, Witt CS, Christiansen FT (2000) The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. Immunogenetics 51(4): 268-280.
- 17. Middleton D, Gonzelez F (2010) The extensive polymorphism of KIR genes. Immunology 129(1):8-19. doi: 10.1111/j.1365-2567.2009.03208.x
- Moesta AK, Parham P (2012) Diverse functionality among human NK cell receptors for the C1 epitope of HLA-C: KIR2DS2, KIR2DL2, and KIR2DL3. Front Immunol 3: Article 336. doi: 10.3389/fimmu.2012.00336
- 19. Norman PJ, Abi-Rached L, Gendzekhadze K, Korbel D, Gleimer M, Rowley D, Bruno D, Carrington CVF, Chandanayingyong D, Chang Y-H, Crespi C, Saruhan-Direskeneli G, Fraser PA, Hameed K, Kamkamidze G, Koram KA, Layrisse Z, Matamoros N, Mila J, Park MH, Pitchappan RM, Ramdath DD, Shiau

M-Y, Stephens HAF, Struik S, Verity DH, Vaughan RW, Tyan D, Davis RW, Riley EM, Ronaghi M, Parham P (2007) Unusual selection on the KIR3DL1/S1 natural killer cell receptor in Africans. Nat Genet 39(9): 1092-1099. doi:10.1038/ng2111

- 20. Parham P, Moffett A (2013) Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. Nat Rev Immunol 13(2): 133-144. doi: 10.1038/nri3370
- 21. Parham P., Norman P.J., Abi-Rached, L., Hilton, H.G. & Guethlein LA (2012a) Review: Immunogenetics of human placentation. Placenta 33, S71-S80. doi: 10.1016/j.placenta.2011.11.020
- 22. Parham P, Norman PJ, Abi-Rached L, Guethlein LA (2012b) Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. Philos Trans R Soc Lond B Biol Sci 367(1590): 800-811. doi: 10.1098/rstb.2011.0266
- 23. Rajalingam R, Du Z, Meenagh A, Luo L, Kavitha VJ, Pavithra-Arulvani R, Vidhyalakshmi A, Sharma SK, Balazs I, Reed EF, Pitchappan RM, Middleton D (2008) Distinct diversity of KIR genes in three southern Indian populations: comparison with world populations revealed a link between KIR gene content and pre-historic human migrations. Immunogenetics 60(5): 207-217. doi: 10.1007/s00251-008-0286-2
- 24. Robinson J, Mistry K, McWilliam H, Lopez R, Marsh SGE (2010) IPD The immune polymorphism database. *Nucleic Acids Res* 38 (Database issue): D863-D869. doi: 10.1093/nar/gkp879
- 25. Ruggeri L, Capanni M, Urbani E, Perruccio K, Shlomchik WD, Tosti A, Posati S, Rogaia D, Frassoni F, Aversa F, Martelli MF, Velardi A (2002) Effectiveness of

donor natural killer cell alloreactivity in mismatched hematopoietic transplantations. Science 295(5562): 2097-2100. doi: 10.1126/science.1068440

- 26. Sanchez-Mazas A, Fernandez-Vina F, Middleton D, Hollenbach JA, Buhler S, Di D, Rajalingam R, Dugoujon J-M, Mack SJ, Thorsby E (2011) Immunogenetics as a tool in anthropological studies. Immunology 133(2): 143-164. doi: 10.1111/j.1365-2567.2011.03438.x
- 27. Single RM, Martin MP, Gao X, Meyer D, Yeager M, Kidd JR, Kidd KK, Carrington M (2007) Global diversity and evidence for coevolution of *KIR* and *HLA*. Nature Genetics 39(9):1114-1119. DOI: 10.1038/ng2077
- 28. Single RM, Martin MP, Meyer D, Gao X, Carrington M (2008) Methods for assessing gene content diversity of KIR with examples from a global set of populations. Immunogenetics 60(12):711-725. doi: 10.1007/s00251-008-0331-1
- 29. Suto Y, Ishikawa Y, Kasahara M, Kasai F, Yabe T, Akaza T, Juji T (1998) Gene arrangement of the killer cell inhibitory receptor family on human chromosome 19q13.4 detected by fiber-FISH. Immunogenetics 48(4): 235-241. PMID: 9716642
- 30. Taniguchi M, Kawabata M (2009) KIR3DL1/S1 genotype and 2DS4 allelic variants in the AB KIR genotypes are associated with Plasmodium-positive individuals in malaria infection. Immunogenetics 61(11-12), 717-730. doi: 10.1007/s00251-009-0401-z
- 31. Valiante NM, Uhrberg M, Shilling HG, Lienert-Weidenbach K, Arnett KL, D'Andrea A, Phillips JH, Lanier LL, Parham P (1997) Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. Immunity 7(6): 739-751. doi:10.1016/S1074-7613(00)80393-3.

- 32. Velickovic M, Velickovic Z, Dunckley H (2006) Diversity of killer cell immunoglobulin-like receptor genes in Pacific Islands populations. Immunogenetics 58(7): 523–532. doi: 10.1007/s00251-006-0124-3
- 33. Velickovic M, Velickovic Z, Panigoro R, Dunckley H (2010) Diversity of killer cell immunoglobulin-like receptor genes in Indonesian populations of Sumatra, Sulawesi and Moluccas Islands. Tissue Antigens 76(4), 325-330. doi: 10.1111/j.1399-0039.2010.01525.x
- 34. Whang DH, Park H, Yoon JA, Park MH (2005) Haplotype analysis of killer cell immunoglobulin-like receptors genes in 77 Koreans families. Human Immunology 66(2): 146-154. doi:10.1016/j.humimm.2004.10.013
- 35. Wilson MJ, Torkar M, Trowsdale J (1997) Genomic organization of a human killer cell inhibitory receptor gene. Tissue Antigens 49(6): 574-579. PMID: 9234478
- 36. Wollstein A, Lao O, Becker C, Brauer S, Trent RJ, Nurnberg P, Stoneking M, Kayser M (2010) Demographic history of Oceania inferred from genome-wide data. Curr Biol 20(22):1983-1992. doi: 10.1016/j.cub.2010.10.040
- 37. Yawata M, Yawata N, Draghi M, Little A-M, Partheniou F, Parham P (2006) Roles for HLA and KIR polymprhisms in natural killer cell repertoire selection and modulation of effector function. J Exp Med 203(3): 633-645. doi: 10.1084/jem.20051884

Supplementary information

						Wn					
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	3DP1	3DL1	3DS1	2DS1	2DS4
2DS2	~	1	-0.27	0.55	0.45	0.29	-0.27	-0.05	0.22	0.31	0.11
2DL2	0	~	-0.27	0.55	0.45	0.29	-0.27	-0.05	0.22	0.31	0.11
2DL3	0.14	0.14	~	-0.22	-0.32	-0.38	-0.04	-0.04	-0.31	-0.31	-0.03
2DL5	0	0	0.22	~	0.67	0.57	-0.22	-0.22	0.70	0.70	-0.15
2DS3	0	0	0.08	0	~	0.41	-0.10	-0.32	0.46	0.56	-0.23
2DS5	0.07	0.07	0.05	0	0.01	~	0.11	-0.38	0.70	0.70	-0.27
3DP1	0.14	0.14	1	0.22	0.50	1	~	-0.04	0.14	0.14	-0.03
3DL1	1	1	1	0	0.08	0.05	1	~	-0.31	-0.31	0.70
3DS1	0.20	0.20	0.09	0	0	0	1	0.09	~	0.90	-0.22
2DS1	0.05	0.05	0.09	0	0	0	1	0.09	0	~	-0.22
2DS4	1	1	1	0.48	0.29	0.23	1	0.04	0.31	0.31	~
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	3DP1	3DL1	3DS1	2DS1	2DS4
						<i>p</i> -value					

Table S1a: Linkage disequilibrium analysis of KIR genes in TP group

						Wn					
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	3DP1	3DL1	3DS1	2DS1	2DS4
2DS2	~	1	-0.25	0.69	0.52	0.36	-0.25	-0.09	0.28	0.41	0.12
2DL2	0	~	-0.25	0.69	0.52	0.36	-0.25	-0.09	0.28	0.41	0.12
2DL3	0.32	0.32	~	-0.17	-0.27	-0.31	-0.03	-0.04	-0.24	-0.24	-0.03
2DL5	0	0	1	~	0.65	0.55	-0.17	-0.25	0.74	0.74	-0.17
2DS3	0	0	0.29	0	~	0.25	-0.27	-0.39	0.47	0.60	-0.27
2DS5	0.08	0.08	0.24	0	0.19	~	0.10	-0.45	0.75	0.75	-0.31
3DP1	0.32	0.32	1	1	0.29	1	~	-0.04	0.13	0.13	-0.03
3DL1	1	1	1	0.48	0.08	0.05	1	~	-0.34	-0.34	0.70
3DS1	0.14	0.14	0.35	0	0.02	0	1	0.12	~	0.87	-0.24
2DS1	0.03	0.03	0.35	0	0	0	1	0.12	0	~	-0.24
2DS4	1	1	1	1	0.29	0.24	1	0.06	0.35	0.35	~
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	3DP1	3DL1	3DS1	2DS1	2DS4
						<i>p</i> -value					

Table S1b: Linkage disequilibrium analysis of KIR genes in PFA sub-group

							Wn						
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	2DP1	2DL1	3DP1	3DL1	3DS1	2DS1	2DS4
2DS2	~	1	-0.36	0.38	0.19	0.32	-0.30	-0.23	-0.04	-0.20	0.27	0.28	-0.09
2DL2	0	~	-0.36	0.38	0.19	0.32	-0.30	-0.23	-0.04	-0.20	0.27	0.28	-0.09
2DL3	0	0	~	-0.38	-0.22	-0.22	0.54	0.63	-0.07	0.44	-0.28	-0.30	0.25
2DL5	0	0	0	~	0.57	0.65	-0.23	-0.27	0.04	-0.25	0.84	0.81	-0.11
2DS3	0.06	0.06	0.04	0	~	0.24	0.08	0.12	0.09	-0.22	0.43	0.31	0.05
2DS5	0	0	0.04	0	0.02	~	-0.18	-0.23	-0.03	-0.38	0.72	0.66	-0.17
2DP1	0	0	0	0.02	0.68	0.12	~	0.76	0.14	0.23	-0.29	-0.31	0.30
2DL1	0.03	0.03	0	0.01	0.34	0.03	0.00	~	-0.04	0.33	-0.32	-0.34	0.40
3DP1	1	1	1	1	1	1	0.26	1	~	-0.04	0.01	0.00	-0.02
3DL1	0.06	0.06	0	0.02	0.06	0	0.07	0.02	1	~	-0.29	-0.31	0.44
3DS1	0.01	0.01	0.01	0	0	0	0	0	1	0.01	~	0.92	-0.13
2DS1	0	0	0	0	0	0	0	1	1	0	0	~	-0.13
2DS4	1	1	0.13	0.44	1	0.25	0.09	0.06	1	0.05	0.36	0.34	~
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	2DP1	2DL1	3DP1	3DL1	3DS1	2DS1	2DS4
							<i>p</i> -value						

Table S1c: Linkage disequilibrium analysis of KIR genes in TM group

							Wn						
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	2DP1	2DL1	3DP1	3DL1	3DS1	2DS1	2DS4
2DS2	~	1	-0.40	0.48	0.18	0.45	-0.37	-0.25	-0.14	-0.20	0.36	0.33	-0.14
2DL2	0	~	-0.40	0.48	0.18	0.45	-0.37	-0.25	-0.14	-0.20	0.36	0.33	-0.14
2DL3	0.01	0.01	~	-0.39	-0.13	-0.39	0.56	0.63	-0.06	0.51	-0.36	-0.38	0.35
2DL5	0	0	0.01	~	0.60	0.60	-0.35	-0.24	-0.14	-0.20	0.82	0.78	-0.14
2DS3	0	0.35	0.38	0	~	0.11	0.09	0.16	0.09	-0.10	0.46	0.40	0.09
2DS5	0	0	0.01	0	0.50	~	-0.32	-0.41	-0.23	-0.33	0.73	0.67	-0.23
2DP1	0.02	0.02	0	0.02	0.66	0.04	~	0.68	-0.05	0.24	-0.43	-0.45	0.39
2DL1	0.23	0.23	0	0.24	0.55	0.02	0	~	-0.04	0.38	-0.30	-0.31	0.57
3DP1	1	1	1	1	1	0.28	1	1	~	-0.03	-0.17	-0.17	-0.02
3DL1	0.49	0.49	0.02	0.49	0.49	0.07	0.23	0.12	1	~	-0.24	-0.25	0.70
3DS1	0.02	0.02	0.03	0	0	0	0	0.07	0.42	0.17	~	0.96	-0.17
2DS1	0.04	0.04	0.01	0	0.01	0.00	0	0.06	0.40	0.16	0	~	-0.17
2DS4	1	1	0.14	1.00	1	0.28	0.12	0.06	1	0.04	0.42	1	~
	2DS2	2DL2	2DL3	2DL5	2DS3	2DS5	2DP1	2DL1	3DP1	3DL1	3DS1	2DS1	2DS4
							<i>p</i> -value						

Table S1d: Linkage disequilibrium analysis of KIR genes in MFA sub-group

Population	Ν	3DL1	2DL1	2DL3	2DS4	2DL2	2DL5	3DS1	2DS1	2DS2	2DS3	2DS5	2DL4	3DL2	3DL3	2DP1	3DP1
ТМ	108	0.95	0.94	0.87	0.99	0.54	0.44	0.64	0.34	0.54	0.20	0.25	1	1	1	0.91	0.97
TP	48	0.96	1	0.96	0.98	0.38	0.48	0.31	0.31	0.38	0.29	0.23	1	1	1	1	0.96
MFA	50	0.96	0.92	0.86	0.98	0.50	0.52	0.42	0.40	0.50	0.28	0.28	1	1	1	0.88	0.98
PFA	34	0.94	1	0.97	0.97	0.32	0.50	0.29	0.35	0.32	0.29	0.24	1	1	1	1	0.97
TOK	47	0.85	0.89	0.96	0.96	0.40	0.57	0.32	0.36	0.43	0.34	0.26	1	1	0.98	0.96	0.96
TON	49	0.84	1	0.94	0.73	0.51	0.53	0.35	0.29	0.49	0.35	0.39	1	1	1	1	0.96
SA	50	0.92	0.98	0.98	0.94	0.38	0.48	0.38	0.34	0.40	0.34	0.22	1	1	1	0.98	0.98
CI	48	0.90	0.92	0.92	0.77	0.58	0.44	0.29	0.40	0.56	0.15	0.35	1	1	1	0.94	0.94
VTP	194	0.88	0.95	0.95	0.85	0.47	0.51	0.34	0.35	0.47	0.29	0.30	1	1	1	0.97	0.96
PMD	228	0.89	0.96	0.95	0.87	0.45	0.51	0.33	0.35	0.45	0.29	0.29	1	1	1	0.97	0.96

Table S2: Comparison of KIR carrier frequencies between various Polynesian datasets

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

					21	NC 4				
					21	054				
	TP	PFA	TM	MFA	TON	TOK	CI	SA	VTP	PMD
TP	~	0.80	0.55	0.98	0	0.99	0.01	0.64	0.03	0.05
PFA	0.86	~	0.97	0.78	0.01	0.76	0.03	0.90	0.10	0.15
TM	0.00	0.00	~	0.57	0	0.45	0	0.18	0	0
MFA	0.37	0.35	0.02	~	0	0.96	0	0.61	0.02	0.04
TON	0.88	0.79	0	0.59	~	0.01	0.86	0.01	0.09	0.03
TOK	0.94	0.81	0	0.41	0.94	~	0.02	0.70	0.09	0.14
CI	0.82	0.98	0	0.27	0.71	0.95	~	0.04	0.27	0.13
SA	0.62	0.56	0	0.84	0.89	0.68	0.48	~	0.15	0.24
VTP	0.85	0.74	0	0.38	0.93	0.92	0.64	0.72	~	0.70
PMD	0.91	0.80	0	0.32	0.99	0.99	0.70	0.64	0.96	~
	TP	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD
					30	DS1				

Table S3a: Homogeneity tests for Polynesian datasets (KIR2DS4 and 3DS1)

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

					3D	DL1				
	TP	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD
TP	~	0.72	0.90	0.97	0.10	0.15	0.43	0.71	0.19	0.24
PFA	NA	~	0.77	0.69	0.27	0.36	0.75	0.71	0.46	0.54
ТМ	0.22	0.36	~	0.86	0.03	0.06	0.31	0.63	0.06	0.09
MFA	0.14	0.24	0.81	~	0.09	0.13	0.40	0.67	0.17	0.21
TON	NA	NA	0.22	0.13	~	0.85	0.58	0.34	0.55	0.42
TOK	0.06	0.13	0.43	0.92	0.06	~	0.73	0.45	0.75	0.61
CI	0.13	0.23	0.76	0.95	0.12	0.97	~	0.95	0.98	0.91
SA	0.32	0.41	0.55	0.36	0.32	0.18	0.33	~	0.60	0.71
VTP	0.23	0.37	0.88	0.67	0.22	0.29	0.62	0.56	~	0.89
PMD	0.29	0.44	0.84	0.48	0.28	0.17	0.44	0.70	0.89	~
	TP	PFA	TM	MFA	TON	TOK	CI	SA	VTP	PMD
					20	DL1				

Table S3b: Homogeneity tests for Polynesian datasets (KIR3DL1 and 2DL1)

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, NA = not applicable, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

					2D	DL5				
—	ТР	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD
TP	~	0.77	0.17	0.18	0.66	0.98	0.67	0.97	0.78	0.85
PFA	0.81	~	0.18	0.19	0.89	0.76	0.78	0.78	0.90	0.96
ТМ	0.09	0.05	~	0.86	0.32	0.18	0.57	0.06	0.03	0.02
MFA	0.30	0.17	0.79	~	0.33	0.19	0.57	0.07	0.06	0.04
TON	0.26	0.14	0.89	0.92	~	0.68	0.98	0.60	0.79	0.99
ТОК	0.93	0.61	0.18	0.46	0.40	~	0.69	0.96	0.80	0.87
CI	0.07	0.04	0.72	0.53	0.60	0.12	~	0.33	0.62	0.53
SA	0.96	0.77	0.10	0.31	0.27	0.97	0.07	~	0.56	0.61
VTP	0.31	0.17	0.31	0.82	0.72	0.52	0.21		~	0.88
PMD	0.45	0.24	0.16	0.60	0.52	0.70	0.12	0.48	0.73	~
	TP	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD

Table S3c: Homogeneity tests for Polynesian datasets (KIR2DL5 and 2DS1)

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

	2DS2											
	ТР	PFA	TM	MFA	TON	TOK	CI	SA	VTP	PMD		
TP	~	0.96	0.05	0.21	0.25	0.62	0.07	0.80	0.20	0.31		
PFA	0.98	~	0.05	0.17	0.20	0.48	0.06	0.63	0.17	0.24		
ТМ	0.32	0.39	~	0.79	0.71	0.27	0.90	0.15	0.31	0.16		
MFA	0.90	0.89	0.39	~	0.92	0.60	0.68	0.42	0.82	0.60		
TON	0.71	0.79	0.05	0.62	~	0.67	0.61	0.49	0.92	0.70		
TOK	0.77	0.84	0.34	0.67	0.95	~	0.26	0.96	0.71	0.91		
CI	0.14	0.18	0.53	0.17	0.04	0.05	~	0.16	0.32	0.20		
SA	0.77	0.84	0.10	0.67	0.94	1.00	0.05	~	0.47	0.65		
VTP	0.97	0.95	0.14	0.90	0.53	0.60	0.07	0.59	~	0.20		
PMD	0.98	0.96	0.12	0.89	0.53	0.60	0.06	0.59	0.99	~		
	TP	PFA	TM	MFA	TON	TOK	CI	SA	VTP	PMD		
					20	S3						

Table S3d: Homogeneity tests for Polynesian datasets (KIR2DS2 and 2DS3)

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

	2DS5												
	ТР	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD			
TP	~	0.95	0.94	0.73	0.14	0.95	0.26	0.91	0.44	0.50			
PFA	NA	~	0.86	0.84	0.23	0.84	0.36	0.87	0.58	0.65			
TM	NA	NA	~	0.84	0.12	0.94	0.25	0.83	0.44	0.53			
MFA	NA	NA	NA	~	0.36	0.96	0.57	0.74	0.93	0.89			
TON	NA	NA	NA	NA	~	0.24	0.89	0.11	0.31	0.24			
TOK	0.99	0.39	0.67	0.98	0.98	~	0.41	0.87	0.68	0.77			
CI	NA	NA	NA	NA	NA	0.99	~	0.21	0.57	0.47			
SA	NA	NA	NA	NA	NA	0.98	NA	~	0.35	0.41			
VTP	0.62	0.67	0.45	0.61	0.61	0.84	0.62	0.61	~	0.92			
PMD	0.65		0.49	0.64	0.64	0.77	0.65	0.64	0.91	~			
	TP	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD			
					3D	DL3							

Table S3e: Homogeneity tests for Polynesian datasets (KIR2DS5 and 3DL3)

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, NA = not applicable, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

	2DP1											
-	TP	PFA	TM	MFA	TON	ТОК	CI	SA	VTP	PMD		
TP	~	NA	0.07	0.04	NA	0.47	0.24	0.32	0.47	0.55		
PFA	0.77	~	0.15	0.10	NA	0.62	0.37	0.41	0.65	0.73		
ТМ	0.65	0.96	~	0.80	0.06	0.46	0.75	0.18	0.04	0.02		
MFA	0.97	0.78	0.77	~	0.04	0.31	0.53	0.12	0.03	0.01		
TON	0.98	0.78	0.67	0.99	~	0.46	0.23	0.32	0.46	0.54		
TOK	0.98	0.76	0.63	0.96	0.97	~	0.66	0.96	0.69	0.90		
CI	0.65	0.87	0.56	0.58	0.98	0.66	~	0.58	0.54	0.40		
SA	0.97	0.78	0.77	1.00	0.99	0.96	0.58	~	0.68	0.80		
VTP	0.99	0.74	0.78	0.77	0.99	0.97	0.81	0.77	~	0.78		
PMD	0.94	0.78	0.82	0.80	0.97	0.92	0.75	0.80	0.93	~		
	TP	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD		

Table S3f: Homogeneity tests for Polynesian datasets (KIR2DP1 and 3DP1)

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, NA= not applicable, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

	2DL3											
	ТР	PFA	ТМ	MFA	TON	TOK	CI	SA	VTP	PMD		
TP	~	0.77	0.17	0.18	0.66	0.98	0.67	0.97	0.78	0.85		
PFA	0.81	~	0.18	0.19	0.89	0.76	0.59	0.78	0.90	0.96		
TM	0.09	0.05	~	0.86	0.32	0.18	0.57	0.06	0.03	0.02		
MFA	0.30	0.17	0.79	~	0.33	0.19	0.57	0.07	0.06	0.04		
TON	0.26	0.14	0.89	0.92	~	0.68	0.98	0.60	0.79	0.99		
TOK	0.93	0.61	0.18	0.46	0.40	~	0.69	0.96	0.80	0.87		
CI	0.07	0.04	0.72	0.53	0.60	0.12	~	0.33	0.62	0.53		
SA	0.96	0.77	0.10	0.31	0.27	0.97	0.07	~	0.56	0.61		
VTP	0.31	0.17	0.31	0.82	0.72	0.52	0.21	0.33	~	0.88		
PMD	0.45	0.24	0.16	0.60	0.52	0.70	0.12	0.48	0.73	~		
	TP	PFA	TM	MFA	TON	TOK	CI	SA	VTP	PMD		
					2D	DL2						

Table S3g: Homogeneity tests for Polynesian datasets (KIR2DL3 and KIR2DL2)

KIR data for the four Polynesian sub-populations (TOK, TON, SA and CI) were obtained from Zlatibor et al. (2006). Abbreviations: N = sample number, TM = Total Maori, TP = Total Polynesians, MFA = Maori with Full Ancestry, PFA = Polynesians with Full Ancestry, TOK = Tokelau, TON = Tonga, SA = Samoa, CI = Cook Islands, VTP = combined carrier frequencies for the four Polynesian sub-populations (TOK, TON, SA and CI) and PMD = combined carrier frequencies of VTP and PFA.

Table S4: KIR2DL1, 2DL2, 2DL3, 3DL1, 2DS1, 2DS2, 3DS1 and HLA-C1, -C2 and -Bw4 carrier frequencies

Population	ID	C1	C2	Bw4	Bw4- 80I	2D L1	2D L2	2D L3	3D L1	2D S1	2D S2	3D S1
Biaka	1	0.69	0.73	0.79	0.49	0.97	0.58	0.80	1	0.01	0.57	0.03
Ethiopian	2	0.72	0.76	0.80	0.77	0.90	0.81	0.77	0.97	0.26	0.77	0.23
Hausa	3	0.64	0.85	0.79	0.67	1	0.37	0.97	1	0.14	0.32	0.03
Ibo	4	0.63	0.83	0.78	0.75	0.96	0.61	0.76	1	0.08	0.58	0.06
Mbuti	5	0.68	0.70	0.73	0.46	0.94	0.66	0.62	0.97	0.28	0.66	0.11
Yoruba	6	0.67	0.84	0.65	0.56	1	0.38	0.92	0.99	0.15	0.35	0.12
Adygei	7	0.72	0.79	0.78	0.57	1	0.50	0.93	0.98	0.35	0.50	0.35
CEPH_UT	8	0.84	0.54	0.62	0.26	0.97	0.57	0.89	0.94	0.40	0.59	0.39
Danish	9	0.83	0.61	0.50	0.20	0.98	0.52	0.96	0.98	0.37	0.52	0.38
European	10	0.82	0.66	0.56	0.26	0.98	0.46	0.92	0.93	0.38	0.47	0.40
Finns	11	0.77	0.63	0.51	0.26	1	0.39	0.97	0.91	0.55	0.39	0.53
Irish	12	0.88	0.59	0.66	0.20	0.99	0.52	0.92	0.97	0.28	0.52	0.32
Russian	13	0.85	0.61	0.67	0.26	0.94	0.44	0.86	0.98	0.23	0.44	0.21
Druze	14	0.74	0.80	0.69	0.57	0.97	0.72	0.78	0.97	0.24	0.72	0.31
Yemenites	15	0.84	0.81	0.69	0.59	0.98	0.47	0.95	0.98	0.31	0.50	0.36
Ami	16	0.98	0.33	0.05	0.00	0.92	0.63	0.85	1	0.38	0.60	0.33
Atayal	17	1	0.10	0	0.00	1	0	1	1	0.33	0.00	0.34
Cambodian	18	0.89	0.61	0.69	0.13	0.91	0.45	0.86	0.95	0.43	0.43	0.48
Hakka	19	1	0.30	0.55	0.32	1	0.33	0.97	0.98	0.50	0.33	0.53
Han_SF	20	1	0.38	0.68	0.40	1	0.14	1	0.98	0.37	0.14	0.36
Han_Taiwan	21	1	0.38	0.79	0.51	0.98	0.19	0.94	0.98	0.35	0.19	0.37
Japan	22	0.96	0.18	0.60	0.38	1	0.09	1	0.96	0.48	0.09	0.49
Micronesian	23	0.53	0.89	0.09	0.03	0.97	0.43	0.97	0.97	0.50	0.43	0.51
Nasioi	24	0.95	0.24	0.25	0.00	0.91	0.96	0.59	0.59	0.91	0.91	0.73
Yakut	25	0.92	0.75	0.88	0.46	0.98	0.28	0.92	0.96	0.46	0.37	0.42
Maya	26	0.89	0.38	0.09	0.07	0.94	0.38	0.94	0.94	0.62	0.38	0.56
Pima	27	0.93	0.60	0.23	0.14	0.88	0.64	0.88	0.77	0.77	0.64	0.72
Katritiana	28	0.80	0.52	0	0	0.86	0.67	0.86	0.82	0.80	0.67	0.80
Surui	29	0.91	0.55	0.07	0.07	0.98	0.26	0.98	0.91	0.48	0.26	0.30
Ticuna	30	0.92	0.54	0.31	0.19	0.94	0.39	0.94	0.94	0.62	0.39	0.46
TP	31	0.87	0.49	0.23	0.02	1	0.39	0.95	0.96	0.39	0.36	0.32
PFA	32	0.89	0.52	0.18	0.03	1	0.33	0.96	0.94	0.44	0.30	0.36
TM	33	0.94	0.47	0.22	0.06	0.95	0.54	0.88	0.95	0.34	0.55	0.36
MFA	34	0.96	0.39	0.14	0	0.91	0.50	0.85	0.96	0.39	0.50	0.43

Reference populations obtained from Single et al. (2007).



Fig. S1a: Correlation between KIR3DL1 and HLA-Bw4 carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.



Fig. S1b: Correlation between KIR3DL1 and and HLA-Bw4-80I carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.



Fig. S1c: Correlation between KIR3DS1 and HLA-Bw4 carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding population ID in Table S4.



Fig. S1d: Correlation between KIR3DS1 and HLA-Bw4-80I carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.



Fig. S1e: Correlation between KIR2DL1 and HLA-C group 2 carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.


Fig. S1f: Correlation between KIR2DL2 and HLA-C group 1 carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.



Fig. S1g: Correlation between KIR2DL3 and HLA-C group 1 carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.



Fig. S1h: Correlation between KIR2DS1 and HLA-C group 2 carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.



Fig. S1i: Correlation between KIR2DS2 and HLA-C group 1 carrier frequencies. Reference populations were obtained from Single et al. (2008). Numbers are corresponding to population ID in Table S4.

General Discussion and Conclusion

My PhD work has been focussed on using recently developed molecular technologies to survey immune system (i.e. HLA, KIR and MICA), blood group and HPA genes in Polynesian and Maori subjects. It began with a bioinformatics study of low (Chapter 3.1) and high (Chapter 3.2) resolution HLA and MICA data and followed by my laboratory based analyses of blood groups (Chapter 3.3) and HPA (Chapter 3.4). The systematic study on KIR (Chapter 3.5) adds extra information on the analysis of high resolution HLA data because both genes are shaped by their shared histories of pathogen encounters.

Overall, the present study has successfully compiled new HLA, MICA, blood group, HPA and KIR datasets for Polynesian and Maori subjects [Objective: A - refer to General These genetic markers are are highly diverse, reflecting their roles in Introduction]. immunity and/or as markers of tissue identity. I would like to note that the present survey provides the first ever blood group, HPA and MICA molecular data for Polynesians and Maori. I have also compared my data with those from previous genetic studies of KIR and HLA for four Polynesian sub-populations (Velickovic, 2001 and Velickovic et al., 2006) and HLA for Maori (Tracey, 2001) [Objective: B]. Statistical tests on the HLA and KIR datases generally show no significant differences between our Polynesians full ancestry sub-group (i.e. PFA: Polynesians with Full Ancestry) and the four individual Polynesian subpopulations (Cook Island, Tokelau, Tonga and Samoa) reported earlier. Therefore, they were combined to make larger dataset, PMD (Polynesians meta dataset) for geneaological study. This was not the case for our Maori group where the HLA data collected for present study on our Maori full ancestry sub-group (MFA: Maori with Full Ancestry) is significantly different from those reported by Tracey (2001). The difference can be attributed to European

Chapter 4

admixture in the latter. Therefore, HLA data from MFA was used for ancestry study while those reported by Tracey (2011) were combined with Maori admixed study group (TM: Total Maori) to represent genepool of Modern day Maori.

The new HLA, KIR, blood group, MICA and HPA datasets/meta datasets were also compared with other those from Asia Pacific populations **[Objectives: C.i. and C.ii]**. Phylogenetic and principal component/coordinate analyses show that Polynesian sub-populations are closely related to another, as compared with those from other Asia Pacific populations. Estimations made based on HLA-A, -B, -C and -DRB1 allele frequencies show wide range (55:45 to 90:10) of ASM:P-SA ancestral fractions in Maori and other Polynesians. The same phenomenon was also observed for blood group, HPA and KIR data (see Chapters 3.1 - 3.5). Inability of these data to show A:SM:P-SA ancestral fractions clearly might appear to reflect different proportions of genetic inputs from their ancestors, but more likely associated with population histories, which include founder effects and local selection forces.

Overall, findings from the present study are entirely consistent with the expectation that both, Polynesians and Maori are intermediate between A-SM and P-SA populations (Kimura *et al.*, 2008 and Wollstein *et al.*, 2010). In particular, Polynesian sub-populations share a common Taiwanese ancestor and have experienced gender-biased admixture with P-SA populations. These observations thus support the account that emerges from other disciplines, particularly from anthropology, archaeology and linguistics (see Chambers and Edinur, 2013 and Chamber, 2013). Genetic data collected for Polynesians over the years contribute significant support for key elements of the *Synthetic Total Evidence Model* of Pacific Settlement. This thesis brings them together for the first time since Chambers (2006) and extends the

Chapter 4

Bellwood *et al.* (2011) and Chambers (2013) accounts which are focussed on support for the Out of Taiwan component. Today, the genepools of Polynesians and Maori are expanding through opportunities for intermarriage with members of other ethnic group (i.e. Europeans) that form significant part of domestic population. Genetic data collected from admixed Maori and Polynesian study groups are significantly different from that full ancestry study groups and show the significant effect of recent admixture following the arrival of Europeans settlers in New Zealand **[Objective: C.ii]**.

Nature of loci tested also bear more or less directly on questions of health [Objective C. iii]. This is particularly important because my thesis is largely concerned with finding efficient new ways to score these markers for transplantation and transfusion medicine. Tissue matching of HLA, MICA, KIR, blood group and HPA is a fundamental element in transfusion and transplant surgery and it is vital that these technologies take accurate account of the ethnic origins of both donor and recipient (see Chapters 3.1 - 3.5). This is due to significant differences between Polynesians and those of European origin, who form the majority segment of the modern day multi-ethnic country of New Zealand. For example, different prevalence of blood groups (e.g. ABO and Rhesus) and HPA alleles (e.g. HPA-1b, -5b and -15a) between Polynesian and European populations were shown to increase risk of alloimunization (see Chapters 3.3 and 3.4). In general, feasibility of finding suitable donors for Polynesian and Maori patients is higher within their own ethnic groups as they share common and unique genepools. Nonetheless, Polynesian and Maori genepools are today becoming more diverse due to inter-group marriage with those from the European lineages. As a consequence, the numbers of available donors for un-admixed Polynesian and Maori patients become smaller, but increase the number of potential donor and recipient pairs that belong to the admixed group. Therefore, I have argued in Chapters 3.2 to 3.4 that it should

become a policy for transfusion and transplant centres to perform a pretransfusion/transplantation cross-match and recruit increased numbers of donors with rare phenotype profiles.

The prevalence of HLA, MICA and KIR data deposited in this thesis were also discussed for their wider potential influence on the health of modern day Maori and Polynesians. In general, Polynesians and Maori are unaffected by autoimmune diseases such as type I diabetes, psoriasis, rheumatoid vasculitis and coeliac disease, which are associated with particular types of HLA, MICA and KIR that are common to Europeans. However, increasing admixture with Europeans may mean that "European autoimmune diseases" such as ankylosing spondylitis and coeliac disease may become common in Maori and Polynesians. In addition, there are several autoimmune diseases (e.g. subacute cutaneous lupus erythematosus and diabetes) that are common to Polynesians and their genetic associations are still to be determined.

My laboratory work on blood group and HPA (see Chapters 3.3 and 3.4) was also designed to assess the new molecular blood group typing methods in Polynesian and Maori populations **[Objective D.i.-iii]**. Scores for the other loci are based on the application of the pre-existing and established screening technologies (e.g. Luminex vs. SBT for HLA and KIR typing). Validation using three different blood group and HPA genotyping methods (i.e. PCR-SSP, SBT and SNP assay) gave 100% concordance and thus reflects the accuracy of DNA-based methods. However, each genotyping method should try to include all known variants that are associated with serological expression. For example, commercial PCR-SSP kits used in the present survey does not include Jk_{null} and HPA-6 which are common/polymorphic in Polynesians and Maori. Thus, routine uncritical application of these kits would lead to false

Chapter 4

predictions and will subsequently increase risk of alloimunization. Overall, there is compelling evidence for the wider application of molecular approaches for HLA, KIR, MICA, blood group and HPA typing. Molecular methods show exact variations in particular genes and have shown to improve resolution of polymorphic loci to a precision that could not previously be achieved using serological approaches (see also Patnaik et al., 2012 and Robinson et al., 2013). In addition, molecular methods are already superior in many ways to the traditional phenotyping and have solved problem of unavaibility of reliable antisera, poor cell expression and cross reactivity (see Chapter 3.3 and 3.4). Currently, molecular methods as simple as PCR-SSP to rapid and high-throughput PCR-SSOP based platforms such as Luminex and SNP genotyping assay have been tested and used by others for HLA, MICA, KIR, blood group and HPA typing (reviewed by Middleton, 2005, and Wu and Csako, 2006, Veldhuisen et al., 2009 and Dunn, 2011). At present, several different commercial kits/SNP genotyping are used to genotype HLA, MICA, KIR, blood group and HPA loci, individually. Nonetheless, it is expected that these loci could be genotyped simultaneously in near future, probably by designing mega DNA chip. All looks promising for the application of molecular approaches in clinical settings and to the extend that never been achieved using serological approaches.

In conclusion, the present molecular genetics study has successfully compiled new HLA, MICA, blood groups, HPA and KIR datasets for Polynesians and Maori, used the datasets for dual analyses of ancestry and health, demonstrated the reliability of current genotyping platforms that would suit well the needs of transfusion and transplantation centers and provide test of the hypotheses presented in the Introduction on p. 30:

Chapter 3.1: that HLA data will provide a way to test the claim of the STEM model that Polynesians have unequal genetic contributions from A-SM and P-SA ancestors.

Accepted: The HLA allele frequencies in Polynesians and Maori are intermediate between those for reference A-SM and P-SA populations and thus our data are entirely consistent with *STEM* model. Ancestral fractions are generally around 60:40 (Fig.S4-S8, pp. 105-109). But HLA-B and -DRB1 loci are exceptional being around 100% A-SM.

Chapter 3.2: that high resolution HLA and MICA data will help to explain differences in the rates of occurrence of auto-immune diseases and infection diseases between that Polynesians and Europeans.

Accepted: Differences in *HLA* and *MICA* distributions between Europeans and Polynesians help to explain why the latter are unaffected by autoimmune diseases and infection diseases that are common among Europeans; e.g. ankylosing spondylitis and uveitis are associated with *HLA-B*27* which is a low frequency allele group in Polynesians: see Table 1, p. 121 and discussion on pp.134-137.

Chapter 3.3: that molecular techniques provide a reliable alternative technology for determining blood groups in Polynesian subjects.

Accepted: There is compelling evidence for the application of molecular blood grouping in Polynesians, but should include all known variants that are associated with serological expression. Both PCR-SSP and BLOODchip technologies proved accurate and effective. *Chapter 3.4: that Maori and Polynesians will show difference in HPA allelic distribution from Europeans due to population history.*

Rejected: Distribution of HPA in Polynesians is only significantly different from Europeans for some loci, e.g. HPA-3 and HPA-15 – see Table 1 on p. 203.

Chapter 3.5: that Polynesians will display a unique repertoire of KIR haplotypes and different from those previously reported in Europeans.

Accepted: Polynesians do have many unique KIR haplotypes compared with Europeans – see Fig. 1 on pp. 233-234 and discussion on pp. 237-239 which explains that Polynesians are most closely related to Asians in this respect.

In addition, novel information gained and deposited in this thesis has strongly endorsed the indigenous biomedicine approach, and will be particularly also valuable to Asia-Pacific populations of Austronesian peoples. In the view of present study, future study and analysis of high resolution HLA and KIR data (to be reported later) will further help to inform ancestry and health in Polynesian and Maori people and will be wide interest to the biomedical community.

References

Bellwood, P., Chambers, G.K., Ross, M. & Hung, C-H. (2011) Are "cultures" inherited? Multidisciplinary perspectives on the origins of Austronesian–speaking peoples prior to 1000 BC. In: *Investigating archaeological cultures: material culture, variability and transmission* (eds. Roberts, B.W & Linden, M.V.), 321-354. Springer, Dordrecht.

Chambers, G.K. (2006) Polynesian genetics and Austronesian prehistory. In: *Austronesian Diaspora and the Ethnogeneses of People in Indonesian Archipelago* (eds. Simanjuntak, T., Pojoh, I. & Hisyam, M.), 299-319. Indonesian Institute of Sciences (LIPI), Jakarta.

Chambers, G.K. (2013) Genetics and the Origins of the Polynesians. In: *Encyclopedia of Life Sciences*. John Wiley & Sons Ltd, Chichester.

Chambers, G.K. & Edinur, H.A. (2013) Genetic relationships between Malays and Maori. In: *The changing values of Malays, Maori and Pacific Islanders* (eds. W.R.A. Rahman & T.R. Higgins), 6-37. Chair of Malay Studies & Victoria University Foundation, Wellington, New Zealand.

Dunn, P.P.J. (2011) Human leucocyte antigen typing: techniques and technology, a critical appraisal. *International Journal of Immunogenetics* **38**: 463-467.

Edinur, H.A., Dunn, P.P.J., Hammond, L., Selwyn, C., Askar, M., Velickovic, Z.M., Lea, R.A. & Chambers, G.K. (2013) HLA and MICA polymorphism in Polynesians and New Zealand Maori: Implications for ancestry and health. *Human Immunology* **74**, 1119-1129.

Edinur, H.A., Dunn, P.P. J., Hammond, L., Selwyn, C., Velickovic, Z.M, Lea, R.A. & Chambers, G.K (2012) Using HLA loci to inform ancestry and health in Polynesian and Maori populations. *Tissue Antigens* **80**, 509-522.

Edinur, H.A., Dunn, P.P.J., Lea, R.A. & Chambers, G.K. (2013) Human platelet antigens frequencies in Maori and Polynesian populations. *Transfusion Medicine* **23**, 330-337.

Edinur, H.A., Dunn, P.P.J., Lea, R.A. & Chambers, G.K. (2013) Molecular approaches to transfusion medicine in Polynesians and Maori in New Zealand. *International Journal of Immunogenetics*. *In-press*.

Kimura, R., Ohashi, J., Matsumura, Y., Nakazawa, M., Inaoka, T., Ohtsuka, R., Osawa, M. & Takunaga, K. (2008) Gene flow and natural selection in Oceanic human populations inferred from genome-wide SNP typing. *Molecular Biology and Evolution* **25**:1750-1761.

Middleton, D. (2005) HLA typing from serology to sequencing era. *Iranian Journal of Allergy, Asthma and Immunology*, **4**, 53-66.

Tracey, M.C. (2007) *HLA allelic variation in New Zealand Maori*. PhD thesis, University of Otago.

Veldhuisen, B., van der Schoot, C.E. & de Haas, M. (2009) Blood group genotyping: from patient to high-throughput donor screening. *Vox Sanguinis*, **97**, 198-206.

Velickovic, Z.M. (2001) *HLA polymorphism in Pacific Islands populations*. PhD thesis, University of Otago, Dunedin, New Zealand.

Velickovic, M., Velickovic, Z. & Dunckley, H. (2006) Diversity of killer cell immunoglobulin-like receptor genes in Pacific Islands populations. *Immunogenetics* **58**: 523-532.

Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R.J., Nurnberg, P., Stoneking, M. & Kayser, M. (2010) Demographic history of Oceania inferred from genome-wide data. *Current Biology* **20**:1983-1992.

Wu, Y.Y. & Csako, G. (2006) Rapid and/or high-throughput genotyping for human red blood cell, platelet and leukocyte antigens, and forensic applications. *Clinica Chimica Acta*, **363**, 165-176.

Genetic relationships between Malays and Maori

Geoffrey K. Chambers and Hisham A. Edinur

School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand

Status: In: The changing values of Malays, Maori and Pacific Islanders (eds. W.R.A. Rahman & T.R. Higgins), pp. 6. Chair of Malay Studies & Victoria University Foundation, Wellington, New Zealand. ISBN: 9780475124005.

Key words: Austronesian Diaspora, Malay, Maori, Polynesia, genetic admixture

Correspondence: Geoffrey K. Chambers, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand. Ph: +64-(0)4-463-6091, Fax: +64-(0)4-463-5331, e-mail: Geoff.Chambers@vuw.ac.nz

Abstract

Malay and Maori people are both end populations of the great Austronesian Diaspora. Under multidisciplinary received wisdom of this model they represent closely related but divergent descendants of Taiwanese farmers who became transoceanic voyagers from around 5000 ybp. The greater part of today's Malay sub-ethnic groups represent this ancestral stock. They arrived relatively recently via Borneo and Indonesia and are now a major component of the modern multicultural population of today's Malaysia.

In contrast, Maori are a geographically and culturally distinct element of the widespread Polynesian people. This larger group arose around 2500 ybp via limited gender-biased geneflow between Austronesian voyagers and Papuan residents of coastal Papua New Guinea and the islands of the Bismarck Archipelago. Maori themselves are now a substantial minority component of the largely European population of modern New Zealand.

In this presentation we will survey genetic evidence collected from Maori and Malay volunteers which inform the above account. Thus it can be said with confidence that Malays and Maori spring from the same original source and remain closely related today. The data to hand include maternal (mitochondrial DNA haplotypes), paternal (Y chromosome nucleotide polymorphisms and microsatellites), and biparental (Human Leucocyte Antigen) systems. Our data are of significance for better understanding of health issues in these populations and endorse the emergent picture from linguistics and archaeology. They also underline our assertion that studies on either one of these two populations have value for the other and indeed for people all across the Pacific Regions of Near and Remote Oceania.

Introduction: The emergent model of Malay, Maori and Pacific genetic history

The developing deep time story of Polynesian genetic history has been presented by one of the present authors in Chambers (2006, 2008) and later phases described in detail by Kayser (2010). In short, the now more or less standard reconstruction features three ancient lineages; (1) island South-East Asians (Austronesian-speaking Mongaloids, A-SM), (2) ancient peoples including the various Negrito groups, Australian Aborigines and New Guineans (Papuan-speaking Australoids, P-SA) and (3) Europeans (Caucasoids) - and see Box 1 for more on terminology. It has previously been the prevailing view that these lineages arose by multiple dispersals out of Africa (Lahr and Foley, 1994). Thus, it has been held that Lineages (1) and (2) separated shortly after anatomically modern humans first left Africa (Figure 1). Lineage (1) people were believed to have travelled to eastern Asia via the central Asian Steppe vs. Lineage (2) ancestors who went by a southern coastal route via India. The ancient Europeans, Lineage (3) simply turned to the North. While much of the available genetic data are consistent with this established view, it has recently been challenged (see Stoneking and Krause, 2011) by extensive single nucleotide polymorphism (SNP) surveys (e.g. by Wollstein et al, 2010). The data are argued to provide a best fit to a model with a single dispersal which gave rise to all the three lineages. Similar doubts have been expressed based on a study of the Aboriginal genome (Rasmussen et al, 2011). These authors suggest later separation of Lineage (1) and (2) and/or extensive geneflow between them - see Gibbons (2011) for commentary. Oppenheimer (2009) seems to endorse the single out of Africa origin for Lineages (1) and (2), but often seems to be in two minds regarding just when the split between them occurred. Recent microbiological data concerning the stomach bacterium,

Helicobacter pylori, from Moodley et al. (2009) and Bayesian dating of Austronesian language trees by Gray et al. (2007) support the idea (Bellwood et al., 2011) that A-SM from Taiwan moved south via the Philippines (Bellwood and Dizon, 2005) and Borneo with a western branch going on to Malaysia and Indonesia and an eastern branch heading for the Pacific (Figure 2). Proto-Polynesian populations were formed when descendants of Lineage (1) voyagers above were re-united with Lineage (2) somewhere in the Bismarck Archipelago off the northern coast of what is today's Papua New Guinea (PNG). There is one caveat which now renders this older orthodox view provisional; i.e. that the HUGO Pan-Asian SNP Consortium (2009) has recently presented an extensive data set and claimed that the Asian Lineage (1) may be of much greater significance than those who followed the putative southern coastal route leading to Lineage (2); see above and discussions in Stoneking and Krause (2011). We note that this view has received some recent support from Reich et al. (2011) who show that the Melanesian genepool in Bourgainville contains relict signals from admixture with the ancient and enigmatic Denisovian people who are presently only known from their remains in Siberian caves. The Bouganville Islanders tested are dark skinned Papuan speakers and thus it seems relatively unlikely that they obtained their Denisovian DNA via contact with Austronesians. However, one should probably suspend judgement on this question as there is no present extended evidence for or against Denisovian genes in any Asian population. Finally, the original Polynesian settlers of New Zealand (Maori) and the various Malay sub-ethnic groups in Peninsula Malaysia now find themselves living amongst more recent settlers. In New Zealand the majority of these people belong to the European Lineage (3). The potential pharmacogenomic consequences of this contemporary admixture have been explored by Lea and Chambers (2007). While in Malaysia historic immigrant populations have come from India, China and Europe. Hence, the contemporary population is a sophisticated multicultural melange consisting of ancient indigenous groups, such as the

Orang Asli, Malay sub-ethnic groups who are predominantly, but perhaps not exclusively, Austronesian arrivals old and new (see Edinur *et al.*, 2009 for an account) plus immigrants old and new.



Figure 1: Diagram showing the conventional Out of Africa Model with multiple dispersals (lilac arrows) giving rise to Lineage (1) shown in pink, Lineage (2) in blue and European Lineage (3) in green. This figure is reproduced from Chambers, 2008 (with permission requested).



Figure 2: Settlement of Island South East Asia, Near and Remote Oceania based on archaeological dating. The directions of movements are inferred from sequential chronology and lingusitic evidence. This figure is reproduced from Bellwood *et al.*, 2008 (with permission requested).

The supporting evidence

The most recent chapter in this history dates starting from around 4,500 ybp (see Chambers, 2006) and forms part of the greater 'Austronesian Diaspora' (Simanjuntak et al., 2006). The common Austronesian ancestors of Malays and all Polynesians (i.e. including Maori) are hill tribe aboriginal people of Taiwan (Amis, Atayal, Paiwan etc.) who migrated southwards via the Batanes Islands and Philippines etc. according to the 'Out of Taiwan' archaeological hypothesis (Bellwood and Dizon, 2005). Recent molecular studies (Delfin et al., 2011, Gunnarsdóttir et al., 2011 and Tabbada et al. 2010) document the passage of maternal and paternal genetic markers across this 'populated viaduct' and gives confidence to those who hold this view. A fully detailed current account of the evidence supporting this pathway and its associated chronology is presented by Bellwood et al. (2011). The onward migration process was dubbed 'The Express Train to Polynesia' by Diamond (1988), but is perhaps better captured by the 'Slow Boat Model' (see Kayser et al., 2000) and later elaborations such

as the 'Slow Boat from Asia Model' (Kayser et al., 2006 and 2008). It is important that this latter model should not be confused with a less orthodox version (Richards et al., 1988) which confusingly still goes under the same title (e.g. as recently presented in Soares et al., 2008). Nonetheless, one must acknowledge that to the minds of some investigators this competing idea may have gained some support with the publication of the recent report from the HUGO Pan-Asian SNP Consortium (2009). However, the present authors caution that the new data point to a significantly earlier chronology for these hypothetical population movements and that there is archaeological support for neither (see later).

Further southward movement brought the voyagers to Borneo and Sulawesi. From here they split with one branch going west and eventually reaching Malaysia and even Madagascar. The other branch went east towards the Pacific (see later). The origins of Proto- and Deutero-Malays have been fully described by Edinur *et al.* (2009). In brief this episode began between 4500 and 3500 ybp as Austronesian speakers moved on to Central Java and Eastern Indonesia. The Proto-Malay population formed in Peninsula Malaysia via migration across the Straits of Malacca 3500 to 2500 ybp. They were joined by a wide range of local immigrant groups, Thai, Chinese etc. to form a Deutero-Malay population that displaced the ancient indigenous peoples into mountain forests, apparently without significant admixture. More recently new Austronesian people have arrived in Malaysia from neighbouring countries and have formed more or less easily recognisable sub-ethnic groups; e.g. Minangkabau from Sumatra moved to Negeri Sembilan (c14th) and Bugis to Johor and Selangor from Makasar (c17th) – see Edinur *et al.* (2009) and references therein for a more complete account.

Continued south and westward movement eventually (by c. 3300 ybp) brought the other branch of the Austronesian voyagers into contact with the Papuan language-speaking descendants of Lineage (2). These were people who had been resident in the general PNG area for more than 20,000 years. In contrast with events in Malaysia, the result of this interaction was widespread genetic and linguistic replacement in the settled indigenous populations in situ. This gives rise to the long-recognised differences between coastal and highland peoples in PNG (Stoneking et al., 1990) and has left clear signals in their descendant gene pools (Friedlaender et al., 2008 and Kayser et al., 2008). For instance, Soares et al. (2011) show that the famous Polynesian CCG mtDNA motif first arose in the Bismarck Archipelago. These new findings are entirely consistent with the account we have presented here and can also be found in much the same form and described at length by Wollstein et al. (2010) and by Kayser (2010). The study by Soares et al. (2011) presents molecular clock dates that are at odds with our chronology derived using extensive evidence from other disciplines. Their new calculations are based on more extensive whole mitochondrial genome data and do allow for the effects of natural selection to accommodate earlier shortcomings as recognised by the same authors (Soares et al., 2009). This helps to bring their dates forward and reduce the confidence intervals around them, but not sufficiently to bring them into line with received wisdom. Once again this leads the present authors to urge a prudent approach when interpreting molecular clock estimates which depend on in-built assumptions, both implicit and explicit. There is a fascinating asymmetry between maternal (mtDNA haplotypes) and paternal (Y-chromosome SNPs and microsatellite loci) markers. This is clearly a result of gender-biased geneflow (see Chambers, 2006 and 2008 and Kayser et al., 2008). Hence, these differences are attributed to matrilocal marriage practice (men move to join their wife's family on their ancestral land) coupled with matrilineal inheritance of property and social status among Austronesian

groups. This behavioural culture is in marked contrast to the patrilocal practices of Papuan groups in the same area (Hage and Marck, 2003). A detailed picture of the evidence upon which this reconstruction stands has been presented recently by Kasyer (2010) and informatively placed in a wider regional and cultural context by Donohue and Denham (2010).

Spreading out from this original base, new settlement movements proceeded to fill up dispersed habitable space and extending to Central Polynesia (Marquesas Is) by 1300 ybp and reaching the most remote corners of the Polynesian Triangle; Hawaii (by 1100 ybp), Easter Island (1100 ybp) and New Zealand (700 ybp). A recent review of available 14-C dating evidence by Wilmshurst et al. (2010) suggests that this phase took place over a much more condensed time period. Caution is presently advised regarding their interpretations as the work has yet to receive wider scrutiny from the academic community. Evidence supporting the general account above has been presented by Bellwood et al. (2011) and receives independent confirmation from the large-scale independent molecular study of Friedlaender et al. (2008). Further, and as noted earlier, a molecular phylogenetic study (Moodley et al., 2009) of H. pylori (an obligate human commensal responsible for the development of stomach ulcers) together with Bayesian analysis of Pacific and East Asian languages (Gray et al., 2009) makes a compelling case for the pathway linking Taiwan with the far distant reaches of Remote Oceania and the directionality of migration. The latter study also reveals sharp temporal discontinuities which are co-incident with major geographical boundaries (Figures 3A and 3B). In contrast, and as also noted above, the SNP study by the HUGO Pan-Asian Consortium (2009) might be interpreted as support for the alternate 'Out of Island SE Asia' aka 'Slow Boat from Melanesia' (Kayser, 2010) hypothesis, which is in such marked conflict with the well established position taken by Bellwood *et al.* (2011). It will clearly take

quite some time before this new information can be accommodated into a single unified mainstream theory. The empirical virtues of the HUGO study (large scale with many loci) are balanced by its failure to account for the manifest differences between Melanesian and Austronesian gene pools (see Kayser *et al.*, 2008 and references) and lack of congruence with a now extensive archaeological chronology (Bellwood *et al.*, 2010). One possibility is that the data analysis methods used by the HUGO team seem to necessarily constrain the data to fit a divaricating (tree-like) process. So, if the actual historical pattern includes reverse migration and/or reticulation, then such methods are prone to converge on a misleading, but otherwise convincing, result.

Recently it has become possible (and affordable) to carry out really large scale genetic screening exercises which are capable of partitioning modern genepools into ancestral fractions (see Box 1). Here we note Kimura *et al.'s* (2008) initial estimate based on extensive nuclear SNP data that the genepool of Polynesians in Tonga is comprised 70% Asian Austronesian and just 30% Papuan. Their reference group for Papuan-speaking Australoids is drawn from the Gidra people who come from the Western Province lowlands of mainland Papua New Guinea and would not be expected to have admixture from Austronesian – speaking Mongoloids coming down the migration route from Taiwan etc. The admixture estimates of Kimura *et al.* (2008) are rather higher than those of Wollstein *et al.* (2010). But nonetheless taken together these studies do serve to show that a significant historical exchange of genes did indeed take place, but never to the extent of complete admixture. They call their model the '*Slow Train*'. The sex biased nature of this process as explained here and by Wollstein *et al.* (2010) is not apparent in the earlier study due to the nature of their data. This later report by Wollstein *et al.* (2010) covers similar ground with two important differences. First, their analysis of their SNP survey data includes a

comparison between rates of polymorphism on X-chromosome loci and those on autosomes confirming directional gender biased admixture as discussed above. Second, their estimate for admixture proportions is 87:13 for East Asian:Near Oceanian (i.e. A-SM:P-SA). Such proportions suggest rather more limited admixture than has conventionally been accepted and is more in keeping with the original Fast Train Model rather than any version of the Slow Boat Model. But perhaps of even greater fascination is Wollstein et al. 's (2010) finding that a Melanesian population in Fiji has much greater Near Oceanian gene content than typical Polynesian populations in the central Pacific islands (Remote Oceania). They quote figures of 65:35 Polynesian:Near Oceanian for their Fijian subjects. This ratio extrapolates to 57:43 East Asian:Near Oceanian using the split above. In our view, this is of key significance in explaining the persistent differences between typical present day Melanesian and Polynesian populations. It also goes a long way towards reconciling the difference in admixture estimates between their study and that of Kimura et al. (2008) i.e. that the latter study included authentic Melanesian type subjects. Hence, it is now apparent that a second wave of migrants left Near Oceania following the Proto-Polynesians and subsequently largely replaced them in the Vanuatu archipelago, New Caledonia and Fiji, but did not extend further. We know nothing about the geographical source(s) of these second wave migrants, or about the reasons why they set out. We can only speculate about the timing of these events, but it seems likely that they correspond to the demise of Lapita culture in these sites. It has long been clear (Marshall, et al., 2005 and Chambers, 2006) that all of the above mentioned models only capture part(s) of the pattern or process. For instance Trejaut et al. (2011) now agree and recognise that:

"Express train", "Slow boat" and "Out of Taiwan" models are all components of a single phenomenon.

However, our own 'Synthetic Total Evidence Model' as first advanced in Chambers (2006) does encompass most features including: Out of Taiwan, admixture in Near Oceania with gender biased geneflow plus Proto-Polynesian Diaspora with Bottlenecks (actually founder effects) – see Figure 4. It must now be extended to include the later expansion of Melanesians across Near Oceania as originally described by Bellwood (1978), Chapter 9 pp. 233-280, but adding in a modern view of the genetic make up of these voyagers according to Wollstein *et al.* (2010).



Figure 3A: Map showing locations where Austronesian languages are spoken and showing the distribution of various language groups colour coded to match the tree in Figure 3B. This figure is reproduced from Gray *et al.*, 2009 (with permission requested).



Figure 3B Evolutionary tree of Austronesian languages. Tree showing ancestor-descendant relationships between Austronesian languages. The inferred locations of the pauses in linguistic evolution are shown in Figure 3A and match discontinuities in the archaeological record. This figure is reproduced from Gray *et al.*, 2009 (with permission requested).



Figure 4: The Synthetic Total Evidence Model shown in diagrammatic form. Movements of the Lineage (1) A-SM people is shown in green and Lineage (2) P-SA populations in orange. The dispersal of Proto-Polynesians is shown by purple arrows. This figure is reproduced from The Dominion Post newspaper (with permission requested).

Box 1 Note on nomenclature of Asian and Pacific peoples.

We have adopted the system below to clearly distinguish modern and ancestral populations. This is necessary to give a clear description of admixed descendants such as Melanesians whose genetic make-up may not correspond to observed cultural traits.

<u>Population</u>	Description
Taiwan aboriginal and other Asian including Filipino, Malay and Indonesian	Austronesian-speaking Mongoloid (A-SM)
Australian Aborigine and Papuan interior	Papuan-speaking Australoid (P-SA)
Coastal PNG plus islands and other Melanesians	~50:50 (A-SM: P-SA) ^{ab}
Polynesian including Maori	70:30 to 87:13 (A-SM: P-SA) ^{ac}

^aWollstein et al. (2010), ^bFriedlaender et al. (2008), ^cKimura et al. (2008)

A short history of research contributions from USM and VUW

Some of information generated in these two centres has already been referenced among the evidence which today supports the emergent model and the work of scientists at VUW has been presented in greater detail and in temporal context by Marshall *et al.* (2005) and by Chambers (2006). Nonetheless, we have judged it appropriate to give a further brief account of this rather diverse set of studies stressing their natures, lessons learned along the way and showing how they are integrated to motivate the present programme which is being carried out by a USM scholar at VUW. This new initiative forms the subject of the final section of this article.

We begin around 1988 with an era of forensic science projects at VUW intended to develop DNA profiling technology for case work. Technology progressed through multi- and single-locus minisatellite probes (Hamilton, 1994) to multiplex PCR amplification of microsatellite loci using fluorescent labels (Watson, 2000). Analysis of these data (Clark *et al.*, 1995) confirmed that the Maori (and Polynesian) genepool was both different from that of Europeans and less variable – a manifestation of the '*population bottleneck in Polynesia*' as according to Flint *et al.* (1989). As might well be expected, these effects were most marked in Maori, being the population right at the end of the series of island hopping migrations. Most importantly this early work secured our set of banked DNA samples together with reliable data on the ethnic groups of the volunteer contributors.

Our next project involved a survey of SNPs in genes for alcohol metabolism. These had been implicated as protective factors in a series of studies on Japanese, Korean and Chinese populations. These markers were known as ADH2*2 and ALDH 2*2 (todays revised nomenclature is ADH1B*2 for the former). Our study confirmed the presence (allele frequency = 0.42) and protective effect of ADH2*2 in young male Maori (Chambers *et al.*, 2002). This gene variant is very rare (frequency generally below 0.02) in Caucasians. Unlike many Asian populations the even more protective ALDH2*2 variant is not found in Maori of Polynesians. This is something that they have in common with Taiwanese aboriginals and lends support to the 'Out of Taiwan' idea.

Our next two studies looked at one matrilineal marker (mtDNA: mitochondrial DNA) and one set of paternal markers (Y-chromosome SNPs) as potential contributors to a missing persons database. Low levels of variability turned out to make both of these of limited utility for Maori and Polynesian subjects in New Zealand. However, the ancestry of these systems proved to be very interesting. The mtDNA markers tracked back somewhat unconvincingly to Taiwan (Whyte, Marshall and Chambers, 2005). But later work by Benton (2010) produced full mtDNA sequences for the first time in our laboratory and his phylogenetic analysis shows how they can be traced back to Taiwanese ancestors (Figure 5). The Y-chromosomes point to PNG (Underhill *et al.*, 2001). This may seem to offer a contradiction with respect to the mtDNA data, but is entirely in line with previous observations, as can be appreciated by examining the wider picture obtained by others (Figure 6). This is a striking illustration of the consequences of previous episodes of gender-biased geneflow. Because there were pre-existing databases of Maori and Polynesian mtDNA haplotypes, Whyte, Marshall and Chambers (2005) were also able to estimate the number of founding females who first settled Aotearoa (later to become known as New Zealand) returning values in the range 40-380, depending on what assumptions are built into the calculations. The fact that all of these are relatively large numbers suggested to us that the settlement process that took place around 700-650 ybp was well planned and deliberate.

More recently we had a chance to collaborate with the Marshfield Clinic programme in the USA (Friedlaender *et al.*, 2008). This allowed our collaborators to score a large number of loci including microsatellites and insertion/deletion polymorphisms (indels). Extensive analysis of this very large dataset including external reference populations helped to build a strong case that Polynesians arose via admixture between A-SM (ex-Taiwan) and P-SM (northern PNG residents); see Figure 7 for an illustration. Hence, taken together these new data plus those from the proceeding work at VUW may fairly be said to illuminate many features of the emergent model presented earlier. These include and A-SM origin in Taiwan, formation of Proto-Polynesians via gender biased admixture with P-SA followed by refinement of the gene pool during colonisation of the Pacific to the most distant islands of

Remote Oceania. Together they go to build the platform upon which the Synthetic Total Evidence Model stands.

Before moving on to describe our present programme at VUW which is focussed on developing an indigenous approach to transplant and transfusion medicine, we need to consider foundation work done at USM which motivates this new venture. Edinur et al. (2009) used molecular methods to examine Class I and II HLA (Human Leucocyte Antigen) systems (see Figure 8 for an explanatory diagram). These markers, HLA-A, -B, -C (formerly Cw), class I and HLA-DRB1 and -DQB1, Class II are important antigens in tissue matching for transplant surgery. They are also extremely diverse and so it is important to survey them widely across big samples drawn from potential donor and recipient pools. The same properties make them excellent tools for detecting population subdivision and ancestral relationships. The study looked at six sub-ethnic groups from various geographic locations within Peninsula Malaysia (Figure 9). As mentioned previously, many of these had only recently (within the past 2-300 years) arrived in Malaysia from neighbouring territories. These include Banjar, Bugis and Jawa who proved to be particularly closely related and who all come from islands around the Java Sea (Java, Kalimantan, and Sulawesi). Equally, Minangkabau and Rawa (both originally from Sumatra) were found to group together.

The outlier population is Kelantan Malays. This is not surprising as they do show cultural and historical links to northern populations Thai and perhaps even further to Vietnam and/or China. However, this does not show clearly in the phylogenetic trees or principal coordinate plots. There may be historic or technical reasons for these observations.

However, the experience gained in the above project does provide an incentive to collect the same information for the VUW DNA Bank samples and will more or less complete delivery of our analytical contract with the original participants. There are precedents (see below for full references) for this type of study in New Zealand, but these lacked clearly defined ancestry records. Hence, one cannot say if they represent modern admixed populations (the health focus) or original genepool representatives (as required for accurate ancestry determination). Thus our new programme not only offers to expand the datasets, but also to better determine how they may best be understood. Groups with full ancestry can then be used to reconstruct patterns of relationship across the Pacific and out to Asia and Australia following the methodology established by Edinur *et al.* (2009).



Figure 5A: Genetic relationships between Maori and Taiwanese Aboriginals as revealed through studies of complete mtDNA sequences. The phylogenetic tree of Maori and Taiwanese mtDNA sequences. Maori sequences (marked as RHAS = Rakaipaaka Health and Ancestry Study) are nested inside those from Amis, Atayal, Paiwan and Tao. Note that various Polynesian mtDNA sequences are also included inside the Maori group. This figure is reproduced from Benton (2010).



Figure 5B: Map of the homelands of Taiwanese Aboriginal Hill Tribes. B: Map of Taiwan showing where the various tribes live. This figure is reproduced from Trejaut *et al.*, 2005 (with permission requested).



Figure 6: Map showing distributions of Y-chromosome haplotypes in A-SM and P-SA populations. This figure is reproduced from Kayser *et al.*, 2003 (with permission requested).



Figure 7: Phylogenetic and Structure analysis of data from many nuclear loci showing the relationship of Polynesians to other population groups worldwide. The branch leading to Polynesians and Maori lies at the intersection of the regions covered by Taiwanese (light blue) and Melanesians (grey). This figure is reproduced from Friedlaender *et al.*, 2008 (with permission requested).



Figure 8: Genetic map of human chromosome 6 showing the locations of class I and II HLA loci together with experimental procedures for low- and high-resolution typing.


Figure 9: Diagram of locations of six Malay sub-ethnic groups who participated in the HLA study by Edinur *et al.* (2009). The figure is reproduced from this source with permission (requested).

The new HLA screening programme at VUW

Throughout the development of the Synthetic Total Evidence Model and its predecessors, data from HLA loci have played a significant part (Sarjeantson *et al.*, 1982) especially in dealing with inconsistent findings between mtDNA and Y-chromosome markers (see Hagelberg *et al.*, 1999, Lie *et al.*, 2006, Mack *et al.*, 2000, and Ohashi *et al.*, 2006). However, interpretation of these HLA data has not had the opportunity to take into consideration the

hybrid origin of Proto-Polynesians via gender-biased gene flow between Austronesianspeaking Mongoloids and Papuan-speaking Australoids. Therefore, we are now returning to examine HLA loci in light of new understanding. As explained above we are using and developing the approach employed by Edinur *et al.* (2009) and will also have the added advantage of being able to incorporate data from their study of extra Austronesian people, the six Malay sub-ethnic groups.

Several screening studies have already been done on Polynesians (see Zimdahl *et al.*, 1999, Mack *et al.*, 2000, Velickovic and Carter 2001, Velickovic *et al.*, 2002 and Gonzalez-Galarza *et al.*, 2011), but they focus mostly on HLA Class II markers. There are also unpublished data for HLA Class I (Velickovic, 2001) from the same samples first used for HLA Class II typing in Velickovic *et al.* (2001). However, these samples too contain individuals who are nominally Polynesians, but who are known to have some Caucasian and/or Asian admixture.

Our present study examines DNA samples that have been classified according to reported admixture (e.g. MFA = Maori with Full Ancestry *vs.* MAH = Maori with Admixed Heritage etc.). Hence, this new work includes the only individuals ever examined where some participants can be clearly identified as lacking genetic input from Caucasian ancestors. This design allows comparisons that were not previously possible. The new work represents an independent re-sampling of these populations, and thus has comparative value too. In brief, we find that those Polynesian populations (VTP = Velickovic Total Polynesians) examined by Velickovic (2001) best match our PFA (Polynesian with Full Ancestry subgroup) and thus represent the ancestral genepool. In contrast, the data (Tracey and Carter, 2006) obtained from Maori samples (TTM = Tracey Total Maori) collected by Tracey (2007) are a much

better match to our pooled TM group (Total Maori = MFA + MAH) or to MAH than they are to MFA. This indicates to us that the earlier study was indeed compromised to some extent by undisclosed admixture. However, our new data can now be combined effectively with previous data for small scale meta-analyses. Thus PMD (combines VTP + PFA) provides a large (2n = 448) sample for ancestry analysis and MMD (combines TTM + TM) provides an increased group size representing the current Maori population in New Zealand. Importantly, our MFA subgroup becomes the best representative of the original Maori genepool for comparison with PMD and data from external source P-SA and A-SM populations, the latter including Malays. The effect of present day admixture with Europeans on the Maori genepool is illustrated in Figure 10. This shows the relative frequencies of HLA-C alleles. Consider, for example HLA-C*01. Here we see that this allele is present in around 37% of MFA genes examined, but only in about 3% of the European reference group. The corresponding value for the admixed MAH subgroup is intermediate at 19%. There are many other examples in the diagram and we have corresponding pictures for all the other loci too.

We can now proceed to historical reconstructions of relationships between populations. We will also compare principal coordinate analysis (PCO) and neighbour-joining (NJ) methodologies along the way. The HLA data are examined in the light of the Total Synthetic Evidence Model of Pacific settlement and include tests for historical geneflow with a 70:30 ratio (A-SM-P-SA) as a null hypothesis based on the SNP data of Kimura *et al.* (2008) versus the 87:13 estimate of Wollstein *et al.* (2010). Wider comparisons have also been made in a new and more appropriate context recognising other Austronesian populations (e.g. Malay and Taiwanese Aboriginals) as being distinct from Polynesians and as representing "the pure ancestral type Austronesian-speaking Mongoloid" lineages. These plots are for the class I

HLA data alone and shown for illustration in Figures 11 (NJ) and 12 (PCO). They are taken from our new report on this work (Edinur *et al.*, 2012).

The NJ analysis (Figure 11) has a fairly conventional outcome and groups Maori with Polynesians closest to Cook Islanders. The PCO plot (Figure 12) is more informative and includes both A-SM and P-SA populations (shown as ellipses that encompass the metadata collections gleaned from the literature and online databases). Here one can clearly see that that the Polynesian populations all lie intermediate between A-SM and P-SA reflecting this hybrid origin. We have done mathematical simulations to estimate A-SM: P-SA fractions in these populations (not shown). This type of exercise typically returns values around 60:40 for most HLA loci – rather lower than those reported for the big SNP surveys (see earlier). However, some individual loci such as HLA-B show close to 100% A-SM ancestry content. We think this is a very exciting finding and interpret it to reflect past episodes of natural This seems to be a reasonable conclusion given that HLA loci have been selection. implicated in defensive responses against a wide range of pathogens (see references in figure legends for more details). The admixed groups like MAH are displaced towards the European reference as might be expected. The MFA group lies closer to the main body of Polynesian data points, again notably closest to Cook Islanders. Melanesian populations represented by Madang and Rabaul show greater affinity with the P-SA group as might be expected. Interestingly the Island South East Asia group (ISEA) which includes Malays lies away from the body of the source A-SM populations from Taiwan and in our view reflects historic changes in HLA allele frequency rather than bulk P-SA input. This view is supported by the fact that the Malay sub-ethnic groups reported on by Edinur et al. (2009) do not seem to have HLA alleles or overall allelic profiles that are in any way similar to those of P-SA populations (data not shown).



Figure 10: Allele frequency spectrum for MFA and MAH compared with a standard European reference population (ELAN). The histogram bars are colour coded according to the key on the diagram. This figure is reproduced from Edinur *et al.*, 2012 (with permission requested).



Figure 11: A neighbour joining tree for combined HLA class I data. Those branches with the greatest statistical support are shown as darker lines according to the key in the diagram. Names of populations are also colour coded as to type (A-SM etc.) for easy reference.



Figure 12: A principal coordinates plot for combined HLA class I data. The colour coding scheme is as for Figure 11. The data for this diagram come from many published sources, but principally those referenced in the text plus others (includes Bhatia *et al.*, 1998, Bugawan *et al.*, 1999, Chen *et al.*, 2007, Chu *et al.*, 2001, Edinur *et al.*, 2009, Gonzalez-Galarza *et al.*, 2010, Lie *et al.*, 2006, Shaw *et al.*, 1999, Main *et al.*, 2001, Lin *et al.*, 2001, Velickovic, 2001).

Concluding observations

The genetic data we have collected over the years contribute significant support for key elements of the Synthetic Total Evidence Model of Pacific Settlement. This article brings them together for the first time since Chambers (2006) and extends the Bellwood *et al.* (2011) account which is focussed on support for the Out of Taiwan component. As previously we hold that our data support accounts that emerge from other disciplines, particularly from anthropology, archaeology, linguistics and indigenous oral histories. Together they provide an enriched worldview of the Austronesian Diaspora which gave rise to both Malays and Maori. Here we show how Austronesian-speaking voyagers have spread round the globe to

reach an estimated population size of more than 250 million in dispersed populations stretching from Madagascar to Easter Island, Hawaii and New Zealand. We have shown that Malays and Maori are closely related via their Taiwanese ancestors and differ in that the Polynesian genetic trail includes restricted and gender-biased admixture with Papuan-speaking Australoids. Today the genepools of both Malays and Maori are expanding through opportunities for intermarriage with members of other ethnic groups that form significant parts of their respective domestic populations. Thus it can be said with confidence that Malays and Maori spring from the same original source and remain closely related.

From a philosophical perspective it is instructive for us to note that the VUW laboratory never consciously engaged in any attempt to investigate this area of scholarship through the design of our experiments. Each new investigation had a strongly applied focus such as designing forensic databases or improving medical outcomes. However, the genetic basis of all of our work means that the data collected do also bear more or less directly on questions of ancestry. Hence, we have taken every opportunity to present our research findings together with analyses and interpretation in this context both via the scientific literature, public talks and the popular media. In doing so we are conscious of an ethical obligation to share this information and its potential significance with those who supported the work and contributed samples.

Our more recent work on pharmacogenomics and HLA screening are anticipated to have medical benefits. Further the recognition of the close relationships between Maori and Malays teaches us that studies carried out on one population can have relevance for the other and indeed for the wider Asia-Pacific community. We seek to further this symbiosis via the present USM and VUW collaboration as described here.

310

Acknowledgements

The experimental work in the VUW laboratory has been supported by a wide range of New Zealand agencies over the years. Edinur's PhD has scholarship support from USM and the Malaysian Government. Our colleague Dr Paul Dunn at the New Zealand Blood Service in Auckland has been instrumental in collecting the HLA data which we report in Edinur *et al.* (2012) and which we highlight here. We wish to express our gratitude for his extensive and high quality contribution to the programme.

References

Bellwood P (1978) Man's Conquest of the Pacific: The Prehistory of Southeast Asia and Oceania. Collins, Auckland.

Bellwood P, Chambers GK, Ross, M. *et al.* (2011) Are "cultures" inherited? Multidisciplinary perspectives on the origins of Austronesian–speaking peoples prior to 1000 BC. pp. 321-234 in: Roberts B. Van der Linden M, eds. *Investigating archaeological cultures: material culture, variability and transmission.* Springer. Dordrecht

Bellwood P, Dizon E. (2005). The Batanes Archaeological Project and the Out of Taiwan hypothesis for Austronesian dispersal. *J Austronesian Stud*: **1**: 1-33.

Benton M (2010). Mitochondrial genome variation and metabolic traits in a Maori community. MSc(Hons) Thesis, Victoria University of Wellington

Bhatia K, Prasad ML, Barnish G *et al.* (1988). Antigen and haplotype frequencies at three human leukocyte antigen (HLA-A, -B and -C) in the Pawaia of Papua New Guinea. *Am J Phys Anthropol* **75**: 329-40.

Bugawan TL, Mack SJ, Stoneking M *et al.* (1999). HLA class I allele distribution in six Pacific/Asian populations: Evidence of selection at the HLA-A locus. *Tissue Antigens* **53**: 311-19.

Chambers GK (2006). Polynesian genetics and Austronesian prehistory. pp. 299-319 in: Simanjuntak T, Pojoh I, Hisyam M, ed. *Austronesian Diaspora and the Ethnogeneses of People in Indonesian Archipelago*. Indonesian Institute of Sciences (LIPI), Jakarta.

Chambers GK (2008). Genetics and the origins of the Polynesians. pp. 1-9 In: *Encylopedia of Life Sciences*. John Wiley & Sons, Chichester UK.

Chambers GK, Marshall SJ, Robinson GM *et al.* (2002). The genetics of alcohol metabolism in Polynesians: Alcohol and aldehyde dehydrogenase genotypes in young men. *Alcohol Clin Exp Res* **26**: 949-955.

Chen S, Hu Q, Xie Y *et al.* (2007). Origin of Tibeto-Burman speakers: evidence from HLA allele distribution in Lisu and Nu inhabiting Yunnan of China. *Hum Immunol* **68**: 550–9.

Chu CC, Lin M, Nakajima F. *et al.* (2001). Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. *Tissue Antigens* **58**: 9-18.

Clark AG, Hamilton JF Chambers GK (1995). Inference of population subdivision from the VNTR distributions of New Zealanders. *Genetica* **96**: 37-49

Delfin F, Salvador JM, Calacal GC *et al.* (2011). The Y-chromosome landscape of the Philippines: Extensive heterogeneity and vaying genetic affinities of Negrito and non-Negrito groups. *Eur J Hum Genet* **19**: 224-230.

Diamond J. (1988). Express train to Polynesia. Nature 336: 307–308.

Donohue M, Denham T (2010). Farming and language in Southeast Asia: Reframing Austronesian history. *Curr Anthropol* **51**: 223-256.

Edinur HA, Dunn PPJ, Hammond C (2012). Using HLA loci to inform ancestry and health in Polynesian and Maori populations. *Tissue Antigens* (submitted)

Edinur HA, Zafarina Z, Spinola H *et al.* (2009). HLA polymorphism in six Malay sub-ethnic groups in Malaysia. *Hum Immunol:* **70**: 518-526.

Flint J, Boyce AJ, Martinson J *et al.* (1989) Population bottlenecks in Polynesia revealed by minisatellites. *Hum Genet* **83**: 257-263.

Friedlaender JS, Friedlaender F, Reed FA, *et al.* (2008) The genetic structure of Pacific Islanders. *PloS Genetics*: **1.** e19.

Gibbons A (2011). Aboriginal genome shows two-wave settlement of Asia. *Science* **333**: 1689-1691.

Gonzalez-Galarza FF, Christmas S, Middleton D *et al.* (2001). Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucl Acids Res*: **39**: D913-19.

Gray RD, Drummond AJ, Greenhill SJ (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*: **323**: 479-483.

314

Gunnarsdóttir ED, Mingkun L Bauchet M *et al.*, (2011). High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res* **21**: 1-11.

Hage P, Marck J (2003). Matrilineality and the Melanesian origin of Polynesian Ychromosomes. *Curr Anthropol* **44**: 121-127.

Hagelberg E, Kayser M, Nagy M *et al.* (1999). Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. *Phil Trans Roy Soc Lond B*: **354**: 141–152.

Hamilton JF (1994) Multi-locus and single-locus DNA profiling in New Zealand. PhD Thesis, Victoria University of Wellington.

HUGO Pan-Asian SNP Consortium (2009). Mapping human genetic diversity in Asia. *Science* **326**: 1541-1545.

Kayser M (2010). The human genetic history of Oceania: Near and remote views of dispersal. *Curr Biol* **20**: R194-R201.

Kayser M, Brauer S, Cordaux R *et al.* (2006). Melanesian origin of Polynesians: mtDNA and Y-chromosome gradients across the Pacific. *Mol Biol Evol*: **23**: 2234–2244.

Kayser M, Brauer S, Weiss G *et al.* (2000). Melanesian origin of Polynesian Y chromosomes. *Curr Biol*: **10**: 1237-1246.

Kayser M, Brauer S, Weiss G *et al.* (2003). Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from New Guinea. *Am J Hum Genet* **72**: 281-302.

Kayser M, Choi Y, van Oven M *et al.* (2008). The impact of the Austronesian expansion: Evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol Biol Evol* **25**: 1362-74.

Kayser M, Lao O, Saar K *et al.* (2008). Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am J Hum Genet*: **82**: 194-198.

Kimura R, Ohashi J, Matsumura Y *et al.* (2008). Gene flow and natural selection in oceanic human populations inferred from genome-wide SNP typing. *Mol Biol Evol* **25**: 1750-61.

Lahr M, Foley R (1994). Multiple dispersals and modern human origins. *Evol Anthropol* **3**: 48-60.

Lea RA, Chambers GK (2007). Pharmacogenetics in admixed populations in Polynesia. pp. 164-189 in: Suarez-Kurtz G. ed. *Pharmacogenomics*. Landes Bioscience, Austin Texas. Lie BA, Dupuy BM, Spurkland A *et al.* (2006). Molecular genetic studies of natives on Easter Island: evidence of an early European and Amerindian contribution to the Polynesian gene pool. *Tissue Antigens* **69**: 10-18.

Lin M, Chu CC, Lee HL *et al.* (2001). Heterogeneity of Taiwan's indigenous population: Possible relation to prehistoric Mongoloid dispersals. *Tissue Antigens* **55**: 1-9. Mack SJ, Bugawan TL, Moonsamy PV *et al.* (2000) Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. *Tissue Antigens* **55**: 383–400.

Main P, Attenborough RD, Chelvanayagam G. (2001) The peopling of New Guinea: Evidence from class I human leukocyte antigen. *Hum Biol* **73**: 365-83.

Marshall SJ, Whyte ALH, Hamilton JF *et al.* (2005). Austronesian prehistory and Polynesian genetics: A molecular view of migration across the Pacific. *New Zealand Science Review* **63**: 75-80.

Moodley Y, Linz BO, Yamaoka Y *et al.* (2009). The peopling of the Pacific from a bacterial perspective. *Science* **323**: 527-530.

Ohashi J, Naka I, Kimura R *et al.* (2006). HLA-DRB1 polymorphism on Ha'ano island of the Kingdom of Tonga. *Anthropol Sc* **114**: 193–198.

Oppenheimer S (2009). The great arc of dispersal of modern humans: Africa to Australia. *Quat Internat* **202**: 2-13.

Rasmussen M, Guo X, Wang, Y *et al.* (2011). An Aboriginal Australian genome reveals separate dispersals into Asia. *Science* **334**: 94-98.

Reich D, Patterson N, Kircher M *et al.* (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* **89**: 516-528.

Richards M, Oppenheimer S, Sykes B. (1998). mtDNA suggests Polynesian origins in eastern Indonesia. *Am J Hum Genet*: **57**: 1463-1475.

Serjeantson SW, Ryan DP, Thompson AR (1982). The Colonization of the Pacific: The story according to human leukocyte antigens. *Am J Hum Genet*: **34**: 904-918.

Shaw C.-K, Chen L.-L, Lee A *et al.* (1999). Distribution of HLA gene and haplotype frequencies in Taiwan: a comparative study among Min-nan, Hakka, Aborigines and Mainland Chinese. *Tissue Antigens* **53**: 51-64.

Simanjuntak T, Pojoh I, Hisyam M eds, (2006). *Austronesian Diaspora and the Ethnogeneses* of People in Indonesian Archipelago. LIPI, Jakarta.

Soares P, Ermini L, Thomson N *et al.* (2009). Correcting for Purifying Selection: An improved human mitochondrial molecular clock. *Am J Hum Genet* **84**: 740-59.

Soares P, Rito T, Trejaut JA *et al.* (2011). Ancient voyaging and Polynesian origins. *Am J Hum Genet* **88**: 1-9.

Soares P, Trejaut JA, Loo J.-H *et al.* (2008). Climate change and postglacial human dispersals in Southeast Asia. *Mol Biol Evol* **25**: 1209–1218.

Stoneking M, Jorde LB, Bhatia K, *et al.* (1990). Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124**: 717-733.

318

Stoneking M and Krause J (2011) Learning about human population history from ancient and modern genomes. *Nat Rev Genet* **12**: 603-614.

Tabbada KA, Trejaut J, Loo J.-H, *et al.* (2010). Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia. *Mol Biol Evol* **27**: 21-31.

Tracey MC (2007). HLA allelic variation in New Zealand Maori. PhD thesis, University of Otago.

Tracey MC, Carter JM (2006). Class II HLA allele polymorphism: DRB1, DQB1 and DPB1 alleles and haplotypes in the New Zealand Maori population. *Tissue Antigens* **68**: 297-302.

Trejaut JA, Kivisild T, Loo JH (2005). Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PloS Biol* **3**: e247.

Trejaut JA Lee C-L, Yen J-C, *et al.* (2011). Ancient migration routes in Austronesianspeaking populations in Southeast Asia and Melanesia might mimic the spread of nasopharangyal carcinoma. *Chin J Cancer* **30**: 96-105.

Underhill PA, Passarino AA, Lin S *et al.* (2001). Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mutat* **17**: 271-280.

Velickovic ZM. (2001). HLA polymorphism in Pacific Islands populations. PhD thesis University of Otago, Dunedin, New Zealand,

Velickovic ZM, Carter JM (2001). HLA-DPA1 and DPB1 polymorphism in four Pacific Islands populations determined by sequencing based typing. *Tissue Antigens* **57**: 493–501.

Velickovic ZM, Delahunt B, Carter JM (2002). HLA-DRB1 and DQB1 polymorphisms in Pacific Islands populations. *Tissue Antigens* **59**: 397-406.

Watson E (2000). Investigation of autosomal tetranucleotide STR loci and male lineages among UK Leicestershire and Polynesian populations. PhD thesis, University of Loughborough, UK.

Whyte, ALH, Marshall SJ, Chambers GK (2005). Human evolution in Polynesia. *Hum Biol* **77**: 157-177.

Wilmshurst JM, Hunt T, Lipo CP, *et al.* (2010). High precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proc Natl Acad Sci USA* **108** : 1815-1820.

Wollstein A, Lao O, Becker C *et al.* (2010). Demographic history of Oceania inferred from genome-wide data. *Curr Biol*: **20**: 1983-1992.

Zimdahl H, Schiefenhovel W, Kayser M, *et al.* (1999). Towards understanding the origin and dispersal of Austronesians in the Solomon Sea: HLA class II polymorphism in eight distinct populations of Asia-Oceania. *Eur J Immunogenet* **26**: 405–16.

16th IHIW: Analysis of HLA Population Data, with updated results for 1996 to 2012 workshop data (AHPD project report)

M. E. Riccio*, S. Buhler*, J. M. Nunes*, C. Vangenot*, M. Cuenod*, 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. Papaioannou Voniatis¶¶¶¶, C. Papasteriades****, F. Poli†††††, L. Richard‡‡‡‡‡, I. Romon Alonso§§§§§, A. Slavcev****, G. Sulcebe¶¶¶¶¶, T. Suslova¶, M. Testi†, J.-M. Tiercy¶¶¶, A. Varnavidou¶¶¶, B. Vidan-Jeras*****, A. Wennerstrom‡‡‡‡ & A. Sanchez-Mazas*

* Laboratory of Anthropology, Genetics and Peopling history (AGP lab), University of Geneva, Geneva, Switzerland, † Laboratory of Immunogenetics and Transplant Biology, IME Foundation, Rome, Italy, ‡ Institute of Immunology FMBA of Russia, Moscow, Russia, § School of Biological Sciences, Victoria University, Wellington, New Zealand, ¶ Chelyabinsk State University, Chelyabinsk, Russia, ** Universite de la Mediterranee, CNRS, EFS, Marseilles, France, †† Welsh Blood Service, Wales, UK, ‡‡ Etablissement Francais du Sang (EFS) Rhone-Alpes, Lyon, France, §§ Tissue typing Laboratory, New Zealand Blood Service, Auckland, New Zealand, ¶¶ Institute for Transfusion Medicine, Hannover Medical School, Hannover, Germany, *** Department of Immunology, University Hospital of Montpellier, Montpellier, France, †† University Hospital Zagreb, Zagreb, Croatia, ‡‡‡ Finnish Red Cross Blood Service, Helsinki, Finland, §§§ All India Institute of Medical Sciences, New Delhi, India, ¶¶¶ National Reference Laboratory for Histocompatibility, University Hospital Geneva, Geneva, Switzerland, **** Institute for Clinical and Experimental Medicine, Prague, Czech Republic, †††† Chiang Mai University, Chiang Mai, Thailand, ‡‡‡‡ University of Helsinki, Helsinki, Finland, §§§§ Foundation Swiss Blood Stem Cells, Bern, Switzerland, ¶¶¶¶ Makarios III Hospital, Nicosia, Cyprus, ***** Department of Immunology and Histocompatibility, Evangelismos Hospital, Athens, Greece, ††††† Fondazione IRCCS Ca'Granda - Ospedale Maggiore Policlinico, Milan, Italy, ‡‡‡‡ Hema-Quebec, Montreal, Canada, §§§§§ Laboratorio de Histocompatibilidad, Hospital Universitario Marques de Valdecilla, Santander, Spain, ¶¶¶¶ Laboratory of Immunology and Tissue Typing, University Hospital Center, Tirana, Albania and ***** Blood Transfusion Centre of Slovenia, Ljubljana, Slovenia.

Status: International Journal of Immunogenetics, 2013, 40, 21-30.

Correspondence: Alicia Sanchez-Mazas, Laboratory of Anthropology, Genetics and Peopling history (AGP lab), Department of Genetics and Evolution – Anthropology Unit, University of Geneva, 12 rue Gustave-Revilliod, CH-1211 Geneva 4, Geneva, Switzerland. Tel: +4122379 6984; Fax: +4122 379 3194; E-mail: alicia.sanchez-mazas@unige.ch

Summary

We present here the results of the Analysis of HLA Population Data (AHPD) project of the 16th International HLA and Immunogenetics Workshop (16IHIW) held in Liverpool in May-June 2012. Thanks to the collaboration of 25 laboratories from 18 different countries, HLA genotypic data for 59 new population samples (either well-defined populations or donor registry samples) were gathered and 55 were analysed statistically following HLA-NET recommendations. The new data included, among others, large sets of well defined populations from north-east Europe and West Asia, as well as many donor registry data from European countries. The GENE[RATE] computer tools were combined to create a GENE[RATE] computer pipeline to automatically (i) estimate allele frequencies by an expectation-maximization algorithm accommodating ambiguities, (ii) estimate heterozygosity, (iii) test for Hardy-Weinberg equilibrium (HWE), (iv) test for selective neutrality, (v) generate frequency graphs and summary statistics for each sample at each locus and (vi) plot multidimensional scaling (MDS) analyses comparing the new samples with previous IHIW data. Intrapopulation analyses show that HWE is rarely rejected, while neutrality tests often indicate a significant excess of heterozygotes compared with neutral expectations. The comparison of the 16IHIW AHPD data with data collected during previous workshops (12th-15th) shows that geography is an excellent predictor of HLA genetic differentiations for HLA-A, -B and -DRB1 loci but not for HLA-DQ, whose patterns are probably more influenced by natural selection. In Europe, HLA genetic variation clearly follows a north to south-east axis despite a low level of differentiation between European, North African and West Asian populations. Pacific populations are genetically close to Austronesian-speaking South-East Asian and Taiwanese populations, in agreement with current theories on the peopling of Oceania. Thanks to this project, HLA genetic variation is more clearly defined

worldwide and better interpreted in relation to human peopling history and HLA molecular evolution.

Introduction

The Analysis of HLA Population Data (AHPD) component of the 16IHIW held in Liverpool (UK) in May-June 2012 was set up as a continuation of the AHPD component of the 15th IHIW held in Buzios (Brazil) in September 2008 (Nunes et al., 2010). The main objectives of the project were to analyse statistically new HLA molecular data for different collections of individuals (mostly population samples and bone marrow donor registries) provided by participating laboratories, to update the results of previous IHIW population studies and to discuss them in relation to the history of human genetic differentiations. Another aim of the project was to improve the statistical methods and computer programs, which were developed within the frame of both the European network HLA-NET (http://www.cost.eu/domains_actions/bmbs/Actions/BM0803) and different research projects supported by the Swiss FNS (Nunes et al., 2011; Buhler et al., 2012) to handle the huge sets of complex HLA data submitted to the AHPD team during this workshop.

Materials and methods

Collection of new population data

Thanks to several announcements made at the European Federation for Immunogenetics (EFI) and the American Society of Histocompatibility and Immunogenetics meetings, through the EFI mailing list and Newsletter and on the 16IHIW website, tissue typing laboratories were invited to participate first by filling the *Population Questionnaire* recommended by *HLA-NET* (http://geneva.unige.ch/ahpd/ahpd_form. php) (Sanchez-Mazas *et al.*, 2012b) and then by sending electronically the HLA data of population or donor samples to the AGP (*Anthropology, Genetics and Peopling history*) team at the University of Geneva

(Switzerland). By May 2012, a total of 55 samples totalizing up to 67 623 individuals were submitted by 23 teams from 17 different countries (Table 1). These data included 32 samples from *well-defined populations* (i.e. populations sampled on purpose for an anthropological study and well characterized geographically and linguistically) and 23 samples from donor registries (typed for either one or several loci, in majority HLA-DRB1). Four additional samples were submitted by 2 teams (from 2 countries, including 1 new) joining the AHPD project during the workshop, bringing the total number of submitted samples to 59 (36 from well-defined populations and 23 from donor registries, see Table 1). These last data could not be included in time for the population comparisons of the present report, but will be used in further analyses.

One to 7 HLA loci were typed by sequence-based typing (SBT), sequence-specific oligoprobes (SSO) and/or sequence-specific primer (SSP) molecular techniques in the submitted samples. Overall, the 16IHIW AHPD data thus represent 166 series of HLA single-locus genotypic data. A detailed description of all data sets including population names, geographic locations, linguistic information, loci typed, sample sizes per locus, molecular techniques used to type the samples and resolution levels, as well as the rules adopted to classify populations into geographic regions (Sanchez-Mazas, 2012) can be found on the AHPD website (http://geneva.unige.ch/ahpd/).

For interpopulation analyses, the new data collected within the scope of the 16IHIW AHPD project were analysed jointly with the population data collected during the last workshops (12th–15th IHIWs). Overall, these data include a total of 116, 113, 75, 56, 57, 97 and 153 samples tested for HLA-A, -B, -C, -DPB1, -DQA1, -DQB1 and -DRB1, respectively

(after removing samples not in Hardy–Weinberg equilibrium (HWE), see below). Their geographic distribution is shown in Fig. 1.

Table 1: List of participants (alphabetical order) who submitted data to the AHPD project of the 16th IHIW.

			Number of submitted	
			samples	
No.	Participant(s) name	Country	Population	Registry
1	Marco Andreani & Manuela Testi	Italy	3 ^a	
2	Margarita Boldyreva	Russia	18	
3	Geoffrey Chambers, Paul Dunn & Hisham	New Zealand	1	
	Atan Edinur			
4	Jacques Chiaroni, Julie Di Cristofaro,	France &	5	
	Jean-Marie Tiercy, Barbara Kervaire	Switzerland		
	& Alicia Sanchez-Mazas' AGP team			
5	Chris Darke	UK		1
6	Valerie Dubois	France		1
7	Paul Dunn	New Zealand		5
8	Nadra Elamin	Germany		1
9	Jean-Francois Eliaou	France		1
10	Zorana Grubic	Croatia	1	
11	Taina Jaatinen	Finland		2
12	Marja Liisa Lokki, Annika Wennerstrom	Finland	1 ^b	
	& Riitta Paakkanen			
13	Narinder Mehra	India		1
14	Dia Papaioannou Voniatis	Cyprus		1
	& Agathi Varnavidou			
15	Chryssa Papasteriades	Greece	1	
16	Francesca Poli	Italy		1
17	Lucie Richard	Quebec		1
18	Inigo Romon Alonso	Spain		1
19	Antonij Slavcev & Libor Kolesar	Czech Republic		1
20	Genc Sulcebe	Albania	1	
21	Tatiana Suslova & Maria Chernova	Russia		3
22	Jean-Marie Tiercy & Grazia Nicoloso	Switzerland		1
23	Blanka Vidan-Jeras	Slovenia	1	2
Total 1			32	23
24	Uma Kanga	India	3	
25	Warunee Kunachiwa	Thailand	1	
Total 2	36		36	23

^aSamples of patient families (unrelated father and mother of the patient); ^bA sample of controls Total 1: data submitted by May 2012 and included in the present analyses

Total 2: total data including additional samples submitted during the 16th AHPD (not included in the analyses).

Statistical analyses

During the preparation of the 16IHIW AHPD project, recommendations were given to all participants on how to report typing data, ambiguities and standard formats before sending the data. The data received by the AHPD team were then screened carefully to check allele names according to the new nomenclature adopted since the last IHIW (Marsh *et al.*, 2010). When present, NMDP codes were expanded to the corresponding lists of alleles. The data were also converted from Text, Excel or LUMINEX output formats to UNIFORMAT (extension .unif), the standard input format for using the GENE[RATE] computer tools (http://geneva.unige.ch/generate/) recommended by HLANET (Sanchez-Mazas *et al.*, 2012a,b). After that, the following analyses were performed for each population thanks to the GENE[RATE] programs:

• Estimation of allele frequencies at each HLA locus at different levels of resolution: first field (in all cases, except HLA-DPB1 see hereafter), second field (when possible) and mixed resolution levels (to keep the maximum information available). To estimate allele frequencies, GENE[RATE] uses an expectation-maximization algorithm under a model of HWE and specifically adapted to take into account all genotyping ambiguities (Nunes *et al.*, 2010, 2011).

• Test for HWE at each HLA locus. To perform this test, GENE[RATE] compares the log likelihood for frequency estimates obtained under a HWE model with that btained under an inbreeding model (Nunes *et al.*, 2010). Populations tested at several HLA loci were considered to deviate significantly from HWE when a significant deviation (5% level) was found for at least one locus after Bonferroni's correction for the number of loci typed in each population. In such cases, the estimated frequencies were not used for interpopulation analyses.

• Estimation of population heterozygosity and test of selective neutrality. The GENE[RATE] tool used here is an adaptation of the classical Ewens–Watterson (Ewens, 1972; Watterson, 1978) and Slatkin (Slatkin, 1994) tests for analysing data with ambiguities. This procedure was first described by Nunes *et al.* (2010) and improved by Nunes *et al.* (2011).

The 'summary statistics' described above were generated together with a colour-coded graph of allele frequencies at each locus and for each population, as was made for the 15IHIW AHPD report (Nunes et al., 2010).

Following intrapopulation analyses, interpopulation analyses comparing the data submitted to the present and previous (i.e. 12th–16th) IHIW were carried out. These analyses included the following:

• Estimation of pairwise Reynolds' genetic distances (Reynolds *et al.*, 1983) between populations.

• Multidimensional scaling (MDS) analyses (Kruskal, 1964) using Reynolds' genetic distances. To better evaluate the genetic relationships between populations, which are sometimes difficult to assess in MDS when numerous populations are included, four data sets were considered successively: (i) the whole set of populations, (ii) the whole set of minus Amerindians, Australian Aborigines, Melanesians, populations from the Pacific area and Taiwanese (all these populations except Amerindians and Taiwanese are called Oceanians hereafter), (iii) Oceanians and Austronesian-speaking populations from South-East Asia (including Taiwanese) and (iv) Europeans and West Asians for data considered at the second field level of resolution.

• Mantel tests (Mantel, 1967) to assess the correlation between genetic and great circle geographic distances (Ray, 2002).

330

The GENE[RATE] computer tools were improved to perform automatically genetic distance and two- and three-dimensional MDS analyses on a set of data (*GENE[RATE] pipeline*). This was performed by integrating R facilities to the ARLEQUIN (Excoffier *et al.*, 2005) program. GNU/Linux scripts and the R software were used to perform Mantel tests. The analyses described above were performed at each HLA locus by considering the first field level of resolution to include a maximum number of populations. Additional analyses were further performed on a lower number of populations from Europe and West Asia by considering the second field level of resolution.



Figure 1: Geographic location of the population samples submitted to the 12th to 16th International HLA and Immunogenetics Workshops and analyzed in this study. The 16IHIW samples are shown in red.

Results

For all population or donor registry samples submitted to the 16IHIW AHPD project, colourcoded graphs of allele frequencies (at the maximum resolution available) and summary statistics estimated at all loci tested were generated and can be found on the AHPD website (http://geneva.unige.ch/ahpd/). Synthesis of the results obtained by considering the data at the first field of resolution (and second field for HLA-DPB1, as alleles at this locus are not subdivided into allele families) is presented in Table 2. Note that some samples, which were initially tested at a medium level of resolution, were recorded at the first field level because of an excessive degree of complexity in the initial data. This applies to the six samples from the Pacific islands and the samples from the Czech Republic, Quebec and France Montpellier (only HLA-B data in the latter case).

As shown in Table 2, the most numerous data set is that for the HLA-DRB1 locus (55 samples representing almost 70 000 individuals), with more well-defined populations than donor registry samples. Many samples have also been tested for HLA-A and -B (32 populations representing more than 40 000 individuals), but this time the proportion of donor registry samples is higher. As expected, mean sample sizes are much larger for donor registry data (up to about 2700 individuals for HLA-DRB1). However, well-defined populations are very satisfactorily represented by sample sizes of about 100-200 individuals. Mean heterozygosities at the first field level of resolution are lower for HLA-DQA1 and -DQB1 (around 70%) compared with HLA-A, -B, -C and -DRB1 (around 85–95%). HLA-DPB1 is close to the latter loci with mean heterozygosities between 80% and 85% at the second field level of resolution. Well-defined population samples do not differ remarkably from donor registry samples with this respect. Very few rejections from HWE are observed (9.6% on all

series of data, and 4.8% after Bonferroni's correction for the number of loci typed in each population, see Table 2). Interestingly, these rejections principally affect donor registry samples: apart from Belorussians from Vitebsk and Kalmyks-El-Mari tested exclusively at HLA-DRB1, HWE is rejected for French, French from Montpellier, Maori, Tongans, Sudanese, Quebecers, Swiss and Slovenians donor, all of them being represented by donor registries. Moreover, after Bonferroni's correction on the number of loci, significant rejections are only observed for donor registries: French (at the second field but not first field level of resolution), Sudanese (at first and second fields), Quebecers (at both fields, with second field only available for HLA-DRB1) and Swiss (at first and mixed first-second fields). These data were thus not included in the interpopulation analyses. Note that for the Swiss, the data for most regional recruitment centres (11 of 13 regional centres) were in HWE, although the whole national registry was not (Buhler *et al.*, 2012), as discussed below.

In contrast with HWE tests, for a high number of samples tested for HLA-B, -C and -DRB1 (much less for HLA-A and -DPB1, and none for HLA-DQA1 and -DQB1), the frequency distributions do not match neutrality expectations, showing an excess of heterozygotes both in well-defined population and donor registry samples (Table 2). For this analysis, some discrepancies are observed between the results obtained at the first and second field levels of resolution, but not systematically in the same way (i.e. rejected at first field and not at second field or the reverse, data not shown).

Correlation coefficients (r) between genetic and geographic distances and their significance are given in Table 3. On the left part of the table, the coefficients shown were computed on data taken at the first field level of resolution, first for all samples and then for all samples minus Amerindians, Taiwanese and Oceanians to clarify the relationships among

the other populations. All coefficients are highly significant (P < 0.001). However, at all loci except DQB1, r is much greater when Amerindian, Taiwanese and Oceanian populations are removed from the analysis. On the right part of the table, the coefficients were computed on data taken at the second field level of resolution, first for all populations from Europe and West Asia taken together, then for Europeans and West Asians taken separately (the analyses at second field level of resolution were limited to geographic regions for which new 16IHIW data were available). Most coefficients are significant (P < 0.001 or P < 0.01) except for Europeans at loci HLA-A and HLA-DPB1, West Asians at locus HLA-C, and Europeans plus West Asians at locus HLA-DQB1.

The MDS analyses (for the different data sets described above) are given in Supporting Information S1–S19 and S20–S29, except (to show an example) the plots for HLA-A that are also presented in Fig. 2 for the total data set and in Fig. 3 for the total data set minus Amerindians, Taiwanese and Oceanians (first field in both cases). Good *Stress* values, around 0.1, were obtained for both analyses. Amerindian, Taiwanese and Oceanian populations are widely scattered on the plot shown in Fig. 2, which suggests a high level of differentiation (high F_{ST}), probably due to rapid genetic drift. By contrast, Fig. 3 reveals a very close relationship between genetic and geographic distances when the latter populations are removed, in agreement with the high and highly significant correlation coefficient found (r = 0.539). On the whole, similar patterns are observed for the HLA-B and -DRB1 loci. On the other hand, the results found for HLA-DQA1 and HLA-DQB1 indicate a greater deviation between genetic and geographic distances (r = 0.418 and 0.390, respectively, for the data set excluding Amerindians, Taiwanese and Oceanians), a discrepancy which is also visible on the MDS. The results found for HLA-C are intermediate (r = 0.445).

Considering the new populations tested for HLADRB1, which are the most numerous (55), the main observations are the following: the inclusion of more new European samples to the analyses allows confirming previous results indicating that European populations are genetically differentiated along a North-West/Central/South-East axis (Nunes et al., 2010; Buhler et al., 2012). Indeed, the Saami and Finns are plotted at one end of the MDS (Supporting Information S18) and are rather close to North-East Asians (among which the new Nenet, Buriat, Tuvinian and Bashkir populations tested in this study); the Greeks, Kosovo Albanians and Italians are plotted at the other end of the MDS and are closer to West Asians (among which the new Iraqi Kurdistan and Syrian populations); and the other Europeans (among which the new populations from the Russian area) are plotted in between. Similar results are found for HLA-A and -B, showing, for example, that Cypriots make a genetic link between South-EastEuropeans and West Asians. The MDS for HLA-DRB1 also shows that, while the Nenets, Buriats, Tuvinians and Bashkirs cluster with other Altaicspeaking populations from north-east Asia (as mentioned earlier), the Udmurts and the Tatars are much more distant. Among the new populations from Afghanistan, the Pashtuns and to a lesser extent the Hazara are genetically closer to the Indians than the Tajiks, Uzbeks and Turkmens. Besides these main trends, the MDS plots show some overlap between European, West Asian, North African and, to a lesser extent, North-East Asian populations, while the populations from the other geographic areas (sub-Saharan Africa and South-East Asia) are more discriminated, in agreement with their geographic location.

As said above, Oceanian and Taiwanese (plus Amerindian) populations are widely scattered on the MDS plots (first series with all data). This is the reason why additional MDS analyses including the new populations from the Pacific together with other Oceanian and Austronesian-speaking populations from South-East Asia have been performed (Supporting Information S4, S7, S10, S13, S16 and S19). A remarkable result is that for the 3 loci HLA-A, -B and -DRB1 (because among the 16IHIW data only one Pacific population was tested for HLA-C, -DQA1 and -DQB1 and none for HLA-DPB1), the new Pacific populations tested (i.e. Cook Islanders, Maori, Maori-Polynesians, Niueans, Samoans and Tongas) are genetically close to each other and to Austronesian-speaking populations from South-East Asia and Taiwan.

The analyses performed with data considered at the second field level of resolution are given in Supporting Information S20–29.

Locus	Collection	Number of	Number of	Mean sample	Mean	HWE rejection	HWE rejection (5%)	Neutrality	
		populations	individuals	size (± s.d)	Heterozygosity ^a	Heterozygosity ^a (%) at 5% level		rejection (%) at	
		tested	tested		(± s.d)		correction	5% level	
А	Total set	32	46 797	1462(±2813)	0.838(±0.049)	3 (9.4)	1	3 (10)	
	wd	10	1768	177(±90)	$0.842(\pm 0.052)$	0	0	2 (25)	
	dr	22	45029	2047(±3246)	0.837(±0.049)	3 (13.6)	1	1 (4.5)	
В	Total set	32	44375	$1387(2618\pm$)	0.907(±0.031)	6 (18.8)	4	21 (67.7)	
	wd	10	1761	176(± 90)	0.913(±0.022)	0	0	6 (75)	
	dr	22	42614	1937(±3017)	0.904(±0.034)	6 (27.3)	4	15 (65.2)	
С	Total set	16	13962	873(±1220)	0.858(±0.022)	3 (18.8)	1	10 (62.5)	
	wd	7	911	130(±56)	0.865(±0.015)	0	0	3 (60)	
	dr	9	13051	1450(±1390)	0.0852(±0.027)	3 (33.3)	1	7 (77.8)	
DPB1	Total set	8	2041	255(±341)	0.838(±0.047)	0	0	1 (12.5)	
	wd	6	726	121(±60)	0.851(±0.042)	0	0	1 (16.7)	
	dr	2	1315	$658(\pm 605)$	0.801(±0.051)	0	0	0	
DQA1	Total set	5	1306	261(±231)	0.701(±0.024)	0	0	0	
	wd	1	146	-	-	0	0	0	
	dr	4	1160	290(±257)	0.703(±0.027)	0	0	0	
DQB1	Total set	18	8745	486(±552)	0.721(±0.046)	1 (5.6)	1	0	
	wd	8	1502	188(±98)	0.703(±0.055)	0	0	0	
	dr	10	7243	724(±652)	0.735(±0.035)	1 (9.1)	1	0	
DRB1	Total set	55	67623	1230(±3553)	0.874(±0.018)	3 (5.5)	1	29 (52.7)	
	wd	32	4348	136(±65)	0.875(±0.018)	2 (6.5)	0	16 (51.6)	
	dr	23	63275	2751(±5180)	0.872(±0.018)	1 (4.2)	1	13 (54.2)	
All loci	Total set	166				16 (9.6)	8 (4.8)	64 (38)	

Table 2: Summary statistics at each HLA locus (first field level of resolution, except HLA-DPB1 at second field)

wd; well-defined populations; dr, donor registry samples.^aSamples rejecting HWE are not considered.

	First field resolution	Second field resolution						
Locus	Collection	k	r	<i>P</i> -value ^a	Collection	k	r	<i>P</i> -value ^a
А	All samples	116	0.288	0.0001	EUR & WASI	28	0.436	0.0001
	All samples minus AME, Taiwanese and OCE	76	0.539	0.0001	EUR	17	0.124	0.2148
					WASI	11	0.163	0.0292
В	All samples	113	0.340	0.0001	EUR & WASI	30	0.424	0.0001
	All samples minus AME, Taiwanese and OCE	75	0.540	0.0001	EUR	15	0.549	0.0001
					WASI	15	0.321	0.0002
С	All samples	75	0.284	0.0001	EUR & WASI	23	0.327	0.0001
	All samples minus AME, Taiwanese and OCE	50	0.445	0.0001	EUR	14	0.441	0.0006
					WASI	9	0.211	0.0578
DPB1					All samples	56	0.395	0.0001
					All samples minus AME, Taiwanese and OCE	39	0.551	0.0001
					EUR & WASI	24	0.287	0.0007
					EUR	17	0.122	0.1240
					WASI	7	0.405	0.0190
	All samples	57	0 364	0.0001				
DQAI	All samples	35	0.304	0.0001				
	All samples	33 07	0.410	0.0001		25	0 154	0.0705
DQDI	All samples	97	0.455	0.0001		55 24	0.134	0.0703
	All samples minus AME, Talwanese and OCE	12	0.390	0.0001		54 1	0.205	0.0131
		152	0.422	0.0001		1	-	-
DKRI	All samples	153	0.422	0.0001		40	0.331	0.0001
	All samples minus AME, Talwanese and OCE	104	0.504	0.0001		28	0.3/4	0.0001
					WASI	12	0.421	0.0001

Table 3: Correlation analyses between genetic and geographic distances

k, number of populations; *r*, Mantel correlation coefficient; EUR, Europeans; WASI, West Asians; AME, Amerindians; OCE, Oceanians (see text). ^aBased on 9999 permutations.


Figure 2: Two-dimensional multidimensional scaling analysis for the total data set at HLA-A (first field of resolution). The three-dimensional plot and legend with population numbers are provided in Supporting Information S2.



Figure 3: Two-dimensional multidimensional scaling analysis for the total data set minus Amerindians, Taiwanese and Oceanians at HLA-A (first field of resolution). The three-dimensional plot and legend with population numbers are provided in Supporting Information S3.

Discussion and conclusion

The new data submitted to the 16IHIW AHPD component considerably improve the population data set brought together during previous HLA workshops (Imanishi *et al.*, 1992; Clayton *et al.*, 1997; Mack *et al.*, 2007; Nunes *et al.*, 2010). This time, Europe is particularly well represented, including an important set of populations from the Russian area, a region which was almost completely unknown for HLA before. A significant improvement was also

Appendix 2

the inclusion of several samples from West Asia. Finally, a novelty of the 16IHIW AHPD project was to gather a large set of donor registry data (mostly from European countries) in addition to well-defined population samples.

The first remarkable result of the analyses performed in this study is that almost all populations tested are in HWE, despite many rejections of selective neutrality. This means that, although evidence is shown of an excess of HLA heterozygotes compared with neutral expectations, probably due to their selective advantage against pathogens (Meyer & Thomson, 2001; Prugnolle et al., 2005; Solberg et al., 2008; Buhler & Sanchez-Mazas, 2011; Sanchez-Mazas et al., 2012a,b), this likely natural selection did not create deviations from HWE. In addition to the fact that the Hardy-Weinberg law is very robust (i.e. almost always verified), any significant deviation from HWE should then be considered seriously as an indicator of either inaccurate population sampling (and typing errors, but this should be considered as less and less likely with current typing methods) or strong deviation from equilibrium due to demographic or selective factors. Interestingly, HWE was rejected for some donor registry samples analysed in this study (French, Sudanese, Quebecers and Swiss), suggesting that such samples are too heterogeneous (in regard to the origin of the individuals tested) to be used for population comparisons. However, as recently found for the Swiss stem cell registry (Buhler et al., 2012), samples collected within regional recruitment centres may be at equilibrium, although the national registry is not. This may reflect genetic variation among regions within the corresponding country and indicate that regional instead of national data can sometimes be used for anthropological studies. The HLA data of bone marrow donor (BMD) registries ought to be analysed more thoroughly in this perspective.

The significant improvement in the HLA-typed population data set resulting from the 16IHIW AHPD project also allows preparing more detailed HLA genetic maps of human populations, which are very useful to investigate human migrations' history. Continental or regional maps are particularly relevant as the huge amount of existing data now make global maps difficult to read. East Asia has recently been described in detail for HLA genetic variation, and a challenging model of human migrations in this part of the world has been proposed (Di & Sanchez-Mazas, 2011). Africa still needs to be explored, as, unfortunately, very few HLA-typed population samples are available for such a large continent (see Fig. 1) whose history seems to be particularly complex (Excoffier et al., 1987; Poloni et al., 2009; Tishkoff et al., 2009). By contrast, Europe is now very well represented by HLA population data, and a clear genetic variation (yet with still unexplained discrepancies between loci, probably due to selection) is observed, mainly along a north to south-east geographic axis (Buhler et al., 2006, 2012). Although neighbouring populations such as North Africans and West Asians overlap to some extent to Europeans, possibly as a signature of past immune adaptations to pathogen-rich environments (Currat et al., 2010; Sanchez-Mazas et al., 2012a,b), the observed pattern may easily be related to the demic diffusion of Neolithic farmers from Anatolia advocated by many scholars based on other genetic results (Ammerman & Cavalli-Sforza, 1984; Chikhi et al., 1998, 2002; Barbujani & Bertorelle, 2001; Barbujani, 2006; Balaresque et al., 2010). Finally, as several samples from the Pacific area ('Polynesians' taken in a broad sense) have been submitted to the present project (Cook Islanders, Maori, Maori-Polynesians, Niueans, Samoans and Tongas), our new results can also be interpreted in relation to population genetic relationships in Oceania. Being in HWE, these donor registry samples may be considered as representative of their corresponding local populations. However, the results show that they are much less differentiated among them and from South-East Asian populations than would be predicted from the large geographic

Appendix 2

distances separating the different Pacific islands from where they originate. Of course, we cannot rule out some admixture effect as except for the Maori-Polynesian (well-defined population sampled around Wellington in New Zealand), the individuals are taken from donor registries, which recruit people from all over Polynesia. On the other hand, the position of the Polynesian samples on the MDS close to Austronesian-speaking South-East Asian and Taiwanese is congruent with the widely supported 'out of- Taiwan' hypothesis of the peopling history of Oceania (Diamond & Bellwood, 2003; Bellwood & Dizon, 2008). Additional analyses on the new HLA data submitted to the AHPD 16IHIW component are currently in progress.

Last but not least, we would like to emphasize that the 16IHIW AHPD project also promoted the development of very useful computer tools now combined into a pipeline for automatic population genetics' analyses. Thanks to this workshop and the impulse of the *HLA-NET* project, the GENE[RATE] pipeline will soon be available to the scientific community to automatically perform basic HLA analyses starting from laboratory typing outputs to ending with population genetic comparisons. We hope that this fundamental contribution will catalyse HLA population genetic studies, most useful for immunogenetic, epidemiological andanthropological studies.

Acknowledgements

This work was supported by grants from the Swiss National Science Foundation (FNS #3100-49771.96 and 3100A0-112651, the ESF (Europe) COST Action BM0803 HLA-NET, and the Swiss State Secretariat for Education and Research (SER #C08.0131) to A.S.M. We

also thank Stephan O. Weber for his precious technical assistance, as well as Derek Middleton and Steven G.E. Marsh for their successful organization of the 16IHIW in Liverpool.

Supporting Information

Additional supporting information may be found in the online version of this article.

References

- Ammerman, A. & Cavalli-Sforza, L. (1984) The Neolithic Transition and the Genetics of Populations in Europe, Princeton University Press, Princeton, New Jersey.
- Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T. E., Rosser, Z.H. et al. (2010) A predominantly neolithic origin for European paternal lineages. *PLoS Biology*, 8, e1000285.
- Barbujani, G. (2006). Population History of Europe: Genetics. Encyclopedia of Life Science, doi: 10.1002/9780470015902. a0005962.
- Barbujani, G. & Bertorelle, G. (2001) Genetics and the population history of Europe.
 Proceedings of the National Academy of Sciences of the United States of America, 98, 22.
- Bellwood, P. & Dizon, E. (2008). 1. Austronesian cultural origins: Out of Taiwan, via the Batanes Islands, and onwards to Western Polynesia. In: Past Human Migrations in East Asia (ed. A. Sanchez-Mazas, R. Blench, M.D. Ross, I. Peiros & M. Lin), pp. 23–39.
 Routledge, London and New York.
- Buhler, S. & Sanchez-Mazas, A. (2011) HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events. *PLoS One*, 6, e14643.
- Buhler, S., Megarbane, A., Lefranc, G. & Tiercy, J.-M. (2006) HLA-C molecular characterization of a Lebanese population and genetic structure of 39 populations from Europe to India-Pakistan. *Tissue Antigens*, 68, 44.
- Buhler, S., Nunes, J.M., Nicoloso, G., Tiercy, J.M. & Sanchez-Mazas, A. (2012) The Heterogeneous HLA Genetic Makeup of the Swiss Population. *PLoS One*, 7, e41400.

- Chikhi, L., Destro-Bisol, G., Bertorelle, G., Pascali, V. & Barbujani, G. (1998) Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 9053.
- Chikhi, L., Nichols, R.A., Barbujani, G. & Beaumont, M.A. (2002) Y genetic data support the Neolithic demic diffusion model. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11008.
- Clayton, J., Lonjou, C. & Whittle, D. (1997) Allele and haplotype frequencies for HLA loci in various ethnic groups. In: *Genetic Diversity of HLA: Functional and Medical Implication. 12th Int. Histocompatibility Workshop and Conference (Paris, June 1996)* (ed. D. Charron), pp. 665–820. EDK, Paris. 1: workshop.
- Currat, M., Poloni, E.S. & Sanchez-Mazas, A. (2010) Human genetic differentiation across the Strait of Gibraltar. *BMC Evolutionary Biology*, **10**, 237.
- Di, D. & Sanchez-Mazas, A. (2011) Challenging views on the peopling history of East Asia: the story according to HLA markers. *American Journal of Physical Anthropology*, 145, 81.
- Diamond, J. & Bellwood, P. (2003) Farmers and their languages: the first expansions. *Science*, **300**, 597.
- Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87.
- Excoffier, L., Pellegrini, P., Sanchez-Mazas, A., Simon, C. & Langaney, A. (1987) Genetics and history of Sub-Saharan Africa. *Yearbook of Physical Anthropology*, **30**, 151.
- Excoffier, L., Laval, G.& Schneider, S. (2005) ARLEQUIN (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47.

- Imanishi, T., Akaza, T., Kimura, A., Tokunaga, K. & Gojobori, T. (1992) Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: HLA 1991. Vol.1 (ed. K. Tsuji, M. Aizawa & T. Sasazuki), pp. 1065–1220. Oxford University Press, Oxford.
- Kruskal, J. (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29, 115.
- Mack, S.J., Tsai, Y., Sanchez-Mazas, A. & Erlich, H.A. (2007) 13th International Histocompatibility Workshop Anthropology/Human Genetic Diversity Joint Report. Chapter 3: Anthropology/human genetic diversity population reports. In: *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference*, Vol. 1 (ed. J.A. Hansen), pp. 580–652. IHWG Press, Seattle.
- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209.
- Marsh, S.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A. *et al.* (2010) Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, **75**, 291.
- Meyer, D. & Thomson, G. (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*, **65**, 1.
- Nunes, J.M., Riccio, M.E., Buhler, S., Di, D., Currat, M., Ries, F. et al. (2010) Analysis of the HLA population data (AHPD) submitted to the 15th International Histocompatibility/Immunogenetics Workshop by using the Gene[rate] computer tools accommodating ambiguous data (AHPD project report). *Tissue Antigens*, **76**, 18.
- Nunes, J.M., Riccio, M.E., Tiercy, J.M. & Sanchez-Mazas, A. (2011) Allele frequency estimation from ambiguous data: using resampling schema in validating frequency estimates and in selective neutrality testing. *Human Biology*, 83, 437.

- Poloni, E.S., Naciri, Y., Bucho, R., Niba, R., Kervaire, B., Excoffier, L., Langaney, A. & Sanchez-Mazas, A. (2009) Genetic Evidence for Complexity in Ethnic Differentiation and History in East Africa. *Annals of Human Genetics*, **73**, 582.
- Prugnolle, F., Manica, A., Charpentier, M., Guegan, J.F., Guernier, V. & Balloux, F. (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology*, 15, 1022.
- Ray, N. (2002). *Geodist*. Computer program, Laboratory of Genetics and Biometry, University of Geneva.
- Reynolds, J., Weir, B.S. & Cockerham, C.C. (1983) Estimation of the Coancestry Coefficient:Basis for a Short-Term Genetic Distance. *Genetics*, **105**, 767.
- Sanchez-Mazas, A. (2012) Adjusting Database Contents to Set Up a HLA map of Europe. STSM report. Action COST BM0803 HLA-NET STSM, Liverpool, UK.
- Sanchez-Mazas, A., Lemaitre, J.F. & Currat, M. (2012a) Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. *Philosophical Transactions of the Royal Society, London, B Biological Sciences*, 367, 830.
- Sanchez-Mazas, A., Vidan-Jeras, B., Nunes, J.M., Fischer, G., Little, A.M., Bekmane, U. *et al.* (2012b) Strategies to work with HLA data in human populations for histocompatibility, clinical transplantation, epidemiology and population genetics: HLA-NET methodological recommendations. *International Journal of Immunogenetics*, 39, 459.
- Slatkin, M. (1994) An exact test for neutrality based on the Ewens sampling distribution. Genetic Research, 64, 71.
- Solberg, O.D., Mack, S.J., Lancaster, A.K., Single, R.M., Tsai, Y., Sanchez-Mazas, A. & Thomson, G. (2008) Balancing selection and heterogeneity across the classical human

leukocyte antigen loci: A meta-analytic review of 497 population studies. *Human Immunology*, **69**, 443.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A. *et al.* (2009) The genetic structure and history of Africans and African Americans. *Science*, 324, 1035.

Watterson, G.A. (1978) The homozygosity test of neutrality. Genetics, 88, 405.