MOLECULAR GENETIC ANALYSIS OF GENES ASSOCIATED WITH INHERITED RESISTANCE TO MALARIAL PARASITAEMIAS

A thesis submitted for the degree of Doctor of Philosophy

by

Gillian E. Wagner

John Curtin School of Medical Research
The Australian National University
Canberra

February 1995
This thesis describes research carried out in the Human and Molecular Genetics Groups, under the supervision of Professor S. Serjeantson and Dr P. Board, at the John Curtin School of Medical Research, Australian National University, from March 1991 until September 1994, during which time I received an Australian National University PhD scholarship. The results presented in this thesis are my own original work, unless otherwise stated in the thesis text or acknowledgements.

Gillian E. Wagner
ACKNOWLEDGMENTS

I would like to acknowledge Prof. Sue Serjeantson for her initial supervision and introduction to the research world of malaria. I have greatly benefited from the supervision and guidance of Dr Phil Board, and particularly appreciate his help in the physical submission of this thesis in my absence. I have valued the discussions with advisors Dr Simon Easteal and Dr Taka Suzuki and am especially indebted to Dr Kuldeep Bhatia, whose encouragement, support and good humour helped me through the last crucial months of writing.

I would like to thank the members of the Human and Molecular Genetics Groups at the John Curtin School of Medical Research for their technical assistance, with special thanks to Ms Marjorie Coggan for her enduring patience and never ending help. I thank Mr Stephan Pluskche for providing the α thalassaemia and Gerbich negative results analysed in this thesis. I also acknowledge his assistance in the joint typing of ovalocytosis in the Wosera. Thank you to the staff of the John Curtin photography unit for their help in producing the photographic material presented in this thesis. I also appreciate the proofreading assistance of Dr Juleen Cavanaugh and Dr Rohan Baker. I am indebted to the friendships of Ms Angela Whittington, who was always there to lend a hand and Ms Alusha Mamchak who helped me to "get over it".

I would like to thank Dr Eleanor Riley for giving me the opportunity to pursue my interests in malaria research before finalising the submission of my thesis.

Finally I would like to thank the three most important people in my life; my parents, for the emotional and financial support in getting me to Canberra, and my fiancé, Ray, who has stood by and supported me through the difficult months of writing my thesis and has been my strength when I have had none. To him I owe so much.
ABSTRACT

Several genetic variants described from the malaria endemic regions of Papua New Guinea (PNG) are considered to be present in high frequencies due to the protection they provide against malarial infection. Prominent among these are α and β thalassaemia, Gerbich negative blood group, glucose-6-phosphate dehydrogenase (G6PD) deficiency, ovalocytosis and HLA antigens. Although the frequency of these genes have been studied in a number of populations and related to their role in protection from malarial infection, no study has investigated the combined presence of protective genes in a single population. The aim of this thesis was to determine the presence and frequency of these malaria-protective genes in Wosera, East Sepik Province, using current molecular genetic techniques.

The frequency of the α thalassaemia deletion, -α4.2/, was the highest yet recorded for a PNG population and appears to be approaching fixation in Wosera. The gene underlying the Gerbich blood group deficiency (Ge-) was also detected at a high frequency in this population. On the other hand, G6PD deficiency was found to be at low frequency. Band 3 deletion (ovalocytosis) was altogether absent from Wosera.

Biochemical analysis of G6PD deficiency was carried out to compare the Wosera variant with previously described biochemical variants from other PNG populations. This analysis determined that the Wosera variant was most like G6PD variants of neighbouring populations in the East Sepik Province. The molecular basis of the Wosera deficiency along with the molecular characterisation of another deficiency variant, from the southern coast population of Kalo, confirmed genetic heterogeneity underlying G6PD deficiency in PNG. A molecular screening study of seven additional PNG populations detected the Kalo variant in 3 further populations but confirmed that the G6PD heterogeneity was not limited to the two variants characterized at the molecular level in this study.
Molecular typing of HLA class II alleles identified limited allelic diversity in Wosera. The high frequency of particular class II alleles in the population from endemic malarial regions of Wosera and Madang compared with their frequencies in the non-endemic highland population of Goroka suggests that these alleles may play a role in protection from malaria.

The prevalence of infection by any plasmodium species was found to be decreased in Gerbich negative homozygotes (Ge -/-) compared with heterozygotes or Gerbich positive individuals. Protection of homozygotes was only marginally significant for *P. malariae* infection and Gerbich heterozygotes were marginally protected from *P. falciparum* infection compared to Gerbich negative or positive homozygotes.

The influences of gene migration and high mutation rates on the presence of protective genotypes in Wosera are also discussed. The recent arrival of G6PD deficiency to Wosera would explain its low frequency. The presence of variant genes is also dependent upon their introduction by different colonizing populations of PNG. For example, the variant band 3 gene characterising the ovalocytosis defect is associated with AN-speaking populations and is not found in the NAN-speaking population of Wosera. The presence of Gerbich negative is common to the NAN-speaking populations found in this region of PNG. Haplotype analysis of the α globin gene cluster of the Wosera -α⁴.² gene variant suggests that it has the same origin as other variants described in other northern coast populations of PNG.

Finally, the presence and effect of combinations of protective genotypes in the Wosera is discussed. The near fixation of the -α⁴.² gene in this population does not appear to affect the selection of the Gerbich gene and together they may contribute to increased genetic protection from malarial infection.


# TABLE OF CONTENTS

Title page......................................................................................................i  
Statement......................................................................................................ii  
Acknowledgments..........................................................................................iii  
Abstract.............................................................................................................iv  
Publications..................................................................................................vi  
Table of Contents......................................................................................vii  
List of Tables.............................................................................................xii  
List of Figures ..........................................................................................xiv  
Abbreviations............................................................................................xvi  

## Chapter 1 Introduction......................................................................1  
1.1 Introduction.......................................................................................1  
1.1.1 Malarial life cycle........................................................................1  
1.1.2 Host acquired immunity..........................................................3  
1.1.3 Control strategies.......................................................................4  
1.1.4 Vaccine development...............................................................5  
1.2 The Malaria Hypothesis......................................................................7  
1.2.1 Malaria as a selector....................................................................7  
1.3 Malaria in Papua New Guinea..........................................................11  
1.3.1 Past control strategies...............................................................14  
1.3.2 Protective effect of protein deficiency.................................15  
1.3.3 Distribution of malaria resistant genes...................................16  
1.3.4 Factors influencing the gene distributions...........................17  
1.4 A Microepidemiological Study of Malaria.........................................20  
1.4.1 Study population.......................................................................20  
1.5 Detection of Malaria Resistant Genetic Variants..............................21  
1.5.1 Serology....................................................................................23  
1.5.2 Variant proteins.......................................................................23  
1.5.3 Microscopy.................................................................................24  
1.5.4 DNA analysis.............................................................................24  
1.6 Aims of the Project.......................................................................25
Chapter 2 Materials and Methods ........................................................................27

2.1 Introduction ......................................................................................................27
  2.1.1 Wosera samples .......................................................................................27
  2.1.2 DNA extraction .........................................................................................27

2.2 Biochemical Analysis of G6PD ........................................................................28
  2.2.1 Starch gel electrophoresis ........................................................................28
  2.2.2 G6PD Enzyme purification .......................................................................28
  2.2.3 Biochemical characterisation ....................................................................30
    2.2.3.1 pH optimum .........................................................................................30
    2.2.3.2 Substrate affinities .............................................................................30
    2.2.3.3 Affinity for alternative substrates ....................................................30
    2.2.3.4 Thermal stability ...............................................................................31
    2.2.3.5 Starch gel electrophoresis ..................................................................31

2.3 Polymerase Chain Reaction (PCR) Amplification .........................................31
  2.3.1 G6PD .........................................................................................................32
  2.3.2 HLA class II .............................................................................................32
  2.3.3 Band 3 protein..........................................................................................32
  2.3.4 PCR product purification .........................................................................35

2.4 Heteroduplex analysis ....................................................................................35

2.5 Sequence Specific Oligonucleotide (SSO) Typing ..........................................36
  2.5.1 Dot blotting ..............................................................................................36
  2.5.2 Prehybridization and hybridization .........................................................36
  2.5.3 G6PD .........................................................................................................37
  2.5.4 HLA class II .............................................................................................38

2.6 Restriction Analysis of G6PD PCR Products ................................................40

2.7 G6PD Sequencing ..........................................................................................40
  2.7.1 M13 Phage cloning ...................................................................................40
    2.7.1.1 M13 ligation ......................................................................................40
    2.7.1.2 M13 transformation ..........................................................................40
  2.7.2 Plasmid vector cloning .............................................................................41
    2.7.2.1 Plasmid ligation ................................................................................42
    2.7.2.2 Plasmid transformation ......................................................................42
    2.7.2.3 Boiling miniprep plasmid preparation ..............................................42
    2.7.2.4 Religation into M13 ..........................................................................43
  2.7.3 M13 ssDNA preparation ..........................................................................43
  2.7.4 Dideoxy sequencing ..................................................................................43
    2.7.4.1 Gel preparation ..................................................................................44
    2.7.4.2 Sequencing reaction ..........................................................................44
Chapter 3 Biochemical Characterisation of G6PD Deficiency in Wosera

3.1 Introduction............................................................................................47
3.2 Materials and Methods...................................................................49
  3.2.1 Starch gel electrophoresis.......................................................49
  3.2.1 Family study...........................................................................50
  3.2.2.1 Sample collection..............................................................50
  3.2.1.2 Biochemical characterisation.............................................50
3.3 Results....................................................................................................52
  3.3.1 G6PD activity.................................................................................52
  3.3.2 Biochemical characteristics of the Wosera G6PD-deficient variant..........................52
3.4 Discussion...........................................................................................62

Chapter 4 Molecular Characterisation of G6PD Deficiency in PNG

4.1 Introduction.............................................................................................64
4.2 Materials and Methods.....................................................................64
  4.2.1 Populations studied...........................................................................64
  4.2.2 Strategies for mutation detection.........................................65
  4.2.3 Screening strategy used in this study....................................67
4.3 Results....................................................................................................69
  4.3.1 Heteroduplex analysis.................................................................69
  4.3.2 Sequencing G6PD genes.................................................................69
    4.3.2.1 Wosera G6PD normal gene....................................................69
    4.3.2.2 Wosera G6PD-deficient gene.....................................................69
    4.3.2.3 Kalo G6PD-deficient gene..........................................................73
    4.3.2.4 Taq Polymerase errors................................................................73
  4.3.3 Verification and Screening by restriction analysis.................77
  4.3.4 Screening using SSO hybridization.................................................77
Chapter 7 \(\alpha\) Thalassaemia and Ovalocytosis in Wosera

7.1 \(\alpha\) thalassaemia

7.1.1 Molecular organisation of \(\alpha\) globin gene

7.1.2 \(\alpha\) thalassaemia and protection from malaria

7.1.3 \(\alpha\) thalassaemia in PNG

7.2 Ovalocytosis

7.2.1 Ovalocytosis in PNG

7.2.2 Ovalocytosis and protection from malaria

7.2.3 Molecular organisation of band 3 gene

7.3 Materials and Methods

7.3.1 \(\alpha\) thalassaemia typing by RFLP analysis

7.3.2 PCR detection of band 3 deletion

7.3.3 Collection of parasitaemia data

7.4 Results

7.4.1 \(\alpha\) thalassaemia in Wosera

7.4.2 Ovalocytosis in Wosera

7.4.3 Decreased parasitaemia vs genotype

7.5 Discussion

Chapter 8 Discussion

8.1 Project aim

8.2 Experimental results

8.2.1 Frequencies of protective genes in Wosera

8.2.2 Effect of host genotype on parasitaemia rates

8.2.3 \(\text{G6PD} \text{ deficiency in PNG - a genetic trail}

8.3 Origins and Affinities of Protective Genotypes in Wosera

8.4 Combined affect of protective genotypes

8.5 Mechanisms of intervention by protective genotypes

8.5.1 Plasmodium species-specific selection

8.6 Advantages and disadvantages of DNA-based analysis

8.7 Limitations of present study

8.8 Conclusions

8.9 Future directions of research

References
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>World distribution of malaria-protective genotypes</td>
<td>9</td>
</tr>
<tr>
<td>2.1</td>
<td>Oligonucleotides and thermal conditions for G6PD amplification</td>
<td>33</td>
</tr>
<tr>
<td>2.2</td>
<td>HLA PCR primers and conditions for the amplification of the hypervariable region of DRB1 (including DR2 subtyping), DQA1 and DQB1</td>
<td>34</td>
</tr>
<tr>
<td>2.3</td>
<td>Sequence specific oligonucleotides (SSO) for the detection of PNG G6PD mutants</td>
<td>37</td>
</tr>
<tr>
<td>2.4</td>
<td>Sequence specific oligonucleotides hybridization patterns of HLA DR2 (DRB1 and DRB5) specific alleles</td>
<td>38</td>
</tr>
<tr>
<td>2.5</td>
<td>Sequence specific oligonucleotide (SSO) hybridization patterns of DQA1 and DQB1 alleles</td>
<td>39</td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of the biochemical characteristics of the Wosera G6PD-deficient variant with similar PNG variants</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Biochemical variants of G6PD in Papua New Guinea</td>
<td>59</td>
</tr>
<tr>
<td>3.3</td>
<td>Biochemical variants described outside of PNG that have biochemical similarity to the Wosera G6PD-deficient variant</td>
<td>61</td>
</tr>
<tr>
<td>4.1</td>
<td>Occurrence of polymorphism in G6PD-deficient genes from nine PNG populations detected by SSO hybridization</td>
<td>80</td>
</tr>
<tr>
<td>4.2</td>
<td>G6PD biochemical variants with nt1388 mutation</td>
<td>84</td>
</tr>
<tr>
<td>4.3</td>
<td>Occurrence of polymorphism in G6PD genes from 12 populations detected by SSO hybridization</td>
<td>85</td>
</tr>
<tr>
<td>5.1</td>
<td>HLA class II allele frequencies in Wosera</td>
<td>98</td>
</tr>
<tr>
<td>5.2</td>
<td>Hardy-Weinberg equilibrium at DQA1, DQB1 and DRB1 loci</td>
<td>100</td>
</tr>
<tr>
<td>5.3</td>
<td>Comparisons of HLA-DQA and -DQB gene frequencies in 3 PNG populations</td>
<td>101</td>
</tr>
<tr>
<td>5.4</td>
<td>HLA-DR, DQ two-locus haplotypes with significant linkage disequilibrium</td>
<td>104</td>
</tr>
<tr>
<td>6.1</td>
<td>Hardy-Weinberg equilibrium of the Gerbich genotypes in the Wosera population</td>
<td>115</td>
</tr>
</tbody>
</table>
Table 6.2  % individuals infected by malarial parasitaemias with respect to Gerbich phenotypes and genotypes in a Wosera population sample (July 1990).................................116
Table 7.1  Gene frequencies of $\alpha$ globin and band 3 deletion genotypes in Wosera..............................................................133
Table 7.2  Percent individuals infected by malarial parasites with respect to $\alpha$ thalassaemia genotypes in Wosera (July 1990).................................................................135
Table 8.1  Summary of the frequency of malaria protective genes in Wosera.................................................................143
# LIST OF FIGURES

| Figure 1.1 | The life cycle of *Plasmodium* .................................................. 2 |
| Figure 1.2 | Malaria Endemicity in Papua New Guinea ............................................ 13 |
| Figure 1.3 | Map of Papua New Guinea ........................................................................ 22 |
| Figure 3.1 | Role of Glucose-6-Phosphate Dehydrogenase in pentose shunt pathway ................. 48 |
| Figure 3.2 | Wosera G6PD deficiency Family Pedigree ................................................ 51 |
| Figure 3.3 | Starch Gel Electrophoresis in three buffer systems ................................... 53 |
| Figure 3.4 | Substrate Affinity for G6P (Km) .............................................................. 54 |
| Figure 3.5 | Substrate Affinity for NADP (Km) ........................................................... 54 |
| Figure 3.6 | pH Optimum .................................................................................................. 55 |
| Figure 3.7 | Thermal Stability (46°C) ........................................................................... 55 |
| Figure 3.8 | Affinity for Alternative Substrates ............................................................ 56 |
| Figure 4.1 | PCR amplification of the coding regions of the G6PD gene .................................. 68 |
| Figure 4.2 | Heteroduplex analysis of G6PD exon 10 ...................................................... 70 |
| Figure 4.3 | Sequence analysis of Wosera nt1388A variant .......................................... 71 |
| Figure 4.4 | Sequence analysis of Kalo nt1360T variant .............................................. 72 |
| Figure 4.5 | Restriction analysis of intron 11 NlaIII polymorphism and nt1360T variant ............ 74 |
| Figure 4.6 | Sequence analysis of G6PD intron 11 polymorphism ...................................... 75 |
| Figure 4.7 | Sequence analysis of Taq Polymerase error .............................................. 76 |
| Figure 4.8 | SSO hybridization of PNG G6PD variants and controls ..................................... 78 |
| Figure 4.9 | Populations screened for G6PD molecular variants, nt1360T and nt1388A .............. 81 |
| Figure 5.1 | Genomic organisation of the HLA genes ..................................................... 90 |
| Figure 6.1 | Schematic representation of normal and Gerbich negative products of proteins synthesis of the Gerbich gene ...................................................... 110 |
| Figure 6.2 | Schematic representation of misalignment and homologous recombination at the GYPC locus .......................................................... 112 |
| Figure 6.3 | Restriction analysis (BgIII) of normal (+/+) and Gerbich negative heterozygote (+/-) and homozygote (-/-) samples ........................................................... 117 |
| Figure 7.1 | Schematic representation of α globin gene cluster organisation and recombination ................................................................. 124 |
| Figure 7.2 | Ovalocytosis in Papua New Guinea ............................................................ 127 |
Figure 7.3  Comparison of band 3 cDNA sequence in normal and ovalocyte band 3..............................................................129
Figure 7.4  Bg/II (Bg) restriction site analysis of normal (αα/), rightward (-α3.7) and leftward (-α4.2) deletion variants of the α globin gene cluster.................................................131
Figure 7.5  Autoradiograph of Bg/II digested genomic DNA probed with a ζ globin gene probe...................................................134
Figure 7.6  PCR amplification of normal (N) and ovalocyte affected (O) individuals. .................................................................136
ABBREVIATIONS

aa    amino acid(s)
AMP   ampicillin
APS   ammonium persulphate
bp    base pair(s)
BPB   bromophenol blue
cDNA  DNA complementary to RNA
dATP  2'-deoxyadenosine 5'-triphosphate
dCTP  2'-deoxycytidine 5'-triphosphate
ddH2O double distilled water
DNA   deoxyribonucleic acid
dNTPs dideoxyribonucleoside triphosphates (dATP, dCTP, dGTP and dTTP)
dGTP  2'-deoxyguanosine 5'-triphosphate
dMSO  dimethylsulphoxide
DTT   dithiothreitol
dTTP  2'-deoxythymidine 5'-triphosphate
EDTA  ethylenediaminetetra acid
EtBr  Ethidium Bromide
EtOH  ethanol
H2O   water
IPTG  isopropyl-β-D-thiogalactopyranoside
kb    kilo base pair(s)
M     Molar
MTT   methyl thiazolyl tetrazolium
min   minute(s)
ml    millilitre
NADP  nicotinamide-adenine dinucleotide phosphate
ng    nanograms
nt    nucleotide
°C    degrees Celcius
O/N   overnight
PCR   polymerase chain reaction
pmoles picomoles (x10-9 moles)
rpm   revolutions per minute
RT    room temperature (~25°C)
s     second(s)
SDS   sodium dodecyl sulphate
ssDNA single stranded DNA
SSO   sequence specific oligonucleotide
TEMED N,N,N',N'-tetra methylethyl-enediamine
U     units (enzyme)
μM    micromolar                ul = microlitre
V     Voltage
vol   volume(s)
X-gal  X-galactosidase
Chapter 1

INTRODUCTION
1.1 Introduction

Malaria occurs in approximately 100 countries, affecting more than 270 million people and leading to premature deaths of 1 to 2 million people (mainly children) each year (WHO 1990; 1991). Malarial transmission has occurred as far north as latitude 64° north and as far south as latitude 31°, which includes Africa, the Americas, South-East Asia, Europe, Eastern Mediterranean and Western Pacific (WHO 1990). As transmission requires an average temperature exceeding 15°C for at least one month a year, it has not been found to occur at altitudes above 3000 metres (Trustees of the Wellcome Trust 1991).

1.1.1 Life cycle

The life cycle of the human malarial parasites (*Plasmodium falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*) requires two hosts: human and the female *Anopheles* mosquito. Figure 1.1 shows the various stages of plasmodium development within these hosts. In humans, a lethal *P. falciparum* infection can occur if merozoites circulating in the blood sequester and, in an as yet identified manner, cause cerebral malaria (Lockyer and Holder 1989).

The malarial parasites undergo four phases of development; three asexual and one sexual. The sexual phase occurs in the mosquito stomach after the ingestion of a plasmodium-infected blood meal. The male and female gametes fuse to form a zygote, which migrates to the epithelium of the stomach wall and differentiates into an oocyst. Here the first asexual stage of growth (sporogony) occurs, leading to the formation of thousands of sporozoites that eventually lyse from the oocyst and migrate into the mosquito’s salivary glands.

The sporozoites are transmitted to the human host during the next blood meal and within minutes invade liver parenchymal cells. The second stage of sporogony occurs within the hepatocytes, where the parasites develop into trophozoites, which grow and divide to produce thousands of invasive merozoites. In *P. vivax* some of the hepatic-invading sporozoites develop into dormant hypnozoites, which can later become active and lead to recrudescence of the
The sexual and first asexual phases take place in an *Anopheles* mosquito, while the second and third asexual phases occur in a vertebrate host, in this case man. *Plasmodium* alternates between the hosts, and in its asexual phases alternates between intracellular growth forms and extracellular invasive forms. The third asexual phase is repeated many times. Phases in the mosquito are shown above the horizontal line (reproduced with permission from the Trustees of the Wellcome Trust, 1991)
infection (see Figure 1.1). The merozoites invade circulating erythrocytes through a series of stages; recognition and attachment (Dvorak et al. 1975), junction formation, invagination of the erythrocyte membrane, sealing of the invaginated membrane into a vacuole and resealing of the penetrated erythrocyte membrane (Aikawa et al. 1978). The third stage of sporogony occurs once the merozoites have invaded, leading to the development of trophozoites and the subsequent growth and development of eight to sixteen new merozoites. When the erythrocytes burst, the merozoites are released for the next cycle of erythrocyte invasion. This phase repeats itself many times and results in the accumulation of a large number of invasive merozoites in the blood. A few merozoites develop into male or female gametocytes which are taken up by the next mosquito’s blood meal to begin the parasite life cycle again.

1.1.2 Host immune response

In populations where periodic (or unstable) malaria transmission occurs, individuals do not acquire adequate immunity memory and are vulnerable to each new wave of infection. However, in populations under constant exposure to endemic malaria (stable transmission) immunity is acquired progressively, with severe disease, reducing to infrequent episodes of illness. The protection provided by progressive adaptation of the immune system to the antigenic diversity of the local parasite population usually requires multiple exposures (Luzzatto 1979; Weidanz 1982; Greenwood 1990; Miller and Scott 1990). This immunity may be mediated by IgG antibodies, which confer protective immunity when transferred from immune to non-immune individuals (McGregor 1982), although the exact mechanism has not been fully established.

The human leukocyte antigens (HLA) play an important role in the host’s acquisition of immunity. Whereas class I antigens are responsible for presentation of endogenous peptides to cytotoxic T cells (CD8+), class II antigens present exogenous peptides on the surface of B cells and macrophages for recognition by helper T (CD4+) cells. Thus, class I antigens present components of malarial antigens from the invasive stage of the life cycle (e.g. in
hepatocytes), while class II antigens present malaria antigens of extracellular origin (Roitt et al. 1991). This ability of the HLA molecules to recognise foreign malarial peptides is fundamental to the development of immunity. HLA antigenic diversity of populations living in malarious areas allows the recognition of a wide variety of plasmodium epitopes and may be partly maintained in response to the evolution of plasmodium antigenic diversity.

Neonates acquire passive immunity, perhaps by the transplacental transfer of maternal IgG antibodies (McGregor 1986). There is also the evidence that the initial persistence of foetal haemoglobin \((\alpha_2\gamma_2)\) in neonates may reduce parasite growth (Pasvol et al. 1977). In populations where stable malarial transmission occurs there is a non-immune “window” in children between six months and five years of age during which most severe malarial infection, and the majority of all malaria-related fatalities, are known to occur. This is the time when maternally-acquired immunity finally wanes and, after a period of some years during which malaria tolerance protects them, when acquired immunity begins to take hold. It is during this period that other protective mechanisms, such as innate resistance to malarial infection, become most effective.

Not all children succumb to malarial infections during this non-immune phase and a proportion of those who survive malarial infection may do so as they possess one, or a number, of protective genetic variants. These variants may help decrease mortality through decreased parasite density or even allow total evasion of the parasite life cycle (Pasvol and Wilson 1982; Serjeantson 1989e; Luzzi and Pasvol 1990).

1.1.3 Malaria control strategy

The use of insecticides in controlling the *Anopheles* mosquito has been a significant measure in the control of malaria since the 1940s (Pant 1988). These insecticides included the chlorinated hydrocarbons (e.g. DDT), organophosphorous compounds, carbamates, pyrethrins and pyrethroids. The initial use of DDT (2,2-bis-[p-chlorophenyl]-1,1,1-trichloroethane) was particularly favoured because of its contact toxicity, low cost, and persistence several months after spraying.
The antimalarial drugs used to treat *P. falciparum* were cinchona alkaloids (e.g. quinine), 4-aminoquinolines (e.g. chloroquine), 4-quinoline methanol, 9-phenanthrenemethanol, and acridine derivatives (for review see Wellems 1991). Mass administration of antimalarial drugs was employed by the WHO in an attempt to control and eliminate malaria. Chloroquine was used primarily because of its low production cost and minimal side effects.

Despite initial successes in malarial control using insecticides and anti-malarial drugs (Perlmann and Wigzell 1988; Najera 1989), the incidence of malaria has begun to rise again, due mainly to the increased occurrence of drug resistant parasites and insecticide resistant mosquitoes (WHO 1990). In the 1950s the proportion of DDT-resistant mosquitoes increased significantly (see Pant 1988). By the 1960s, chloroquine-resistant *P. falciparum* was discovered in Southeast Asia and South America. This resistance spread so rapidly that by 1985 most countries in malaria endemic zones had reported chloroquine-resistant *P. falciparum* (WHO 1987). The increasing spread of chloroquine resistance, and the mounting evidence of the toxicity and lack of efficacy of alternative antimalarial drugs, has complicated the development of effective prophylactic drugs for different regions (Keystone 1990).

1.1.4 Vaccine development

There are many stages of plasmodium development in the human host and mosquito vector toward which a malaria vaccine could be directed. These have been outlined by Alpers *et al.* (1992) and include the gamete, zygote, ookinete, oocyst and sporozoite stages in the mosquito vector, and the sporozoite, preerythrocytic liver and blood stages (both asexual and sexual) within the human host.

The first attempt to induce an effective immune response against malaria with irradiated sporozoites was made in 1936 (Boyd and Kitchen 1936), but experimental vaccination against *P. falciparum* and *P. vivax* infections was not successful with this technique until much later (Clyde *et al.* 1975). The theoretical goal of sporozoite vaccination is to provide complete protection from transmission of malaria and its clinical symptoms. This is a
difficult task that requires the elimination of all sporozoites and the maintenance of long-lasting immunity (Alonso et al. 1994). Unfortunately, large scale human vaccination with irradiated sporozoites is impractical due to the expense and unfeasibility of large-scale mosquito breeding.

Blood-stage vaccines, those targeting circulating merozoites, are a more practical alternative, as they control (rather than eliminate) malaria by decreasing merozoite numbers and, as a result, reducing or preventing malaria-related morbidity and mortality (Playfair et al. 1990). A Colombian group, headed by M.E. Patarroyo, has developed a blood-stage subunit malaria vaccine, SPf66, which consists of three peptides derived from merozoite stage-specific peptides (SPf 55.1, SPf 83.1 and SPf 35.1) linked by a 4 amino acid circumsporozoite (CS) repeat sequence of P. falciparum (Patarroyo et al. 1988). It is the first subunit vaccine to undergo phase III field trials, first in Venezuela, Colombia (Amador et al. 1992), and now in Tanzania (Teuscher et al. 1994), The Gambia, and Thailand (cf Dye and Targett 1994). Although initial monitoring of immune effectiveness was hampered by the exclusion of placebo controls (Marshall 1992), the vaccine does appear to provide some protection (Amador et al. 1992; Valero et al. 1993); current vaccine trials have incorporated appropriate controls (Alonso et al. 1994).

Vaccines producing antibodies toward gametocytes (sexual-stage) prevent transmission of malaria, but do not provide protection for the vaccine recipient. Therefore, sexual-stage peptides are appropriate components of a vaccine “cocktail”, offering protection to both the individual and the community (Kaslow et al. 1992).

There are several reasons why the path to a malaria vaccine has been a difficult one. Firstly, the malarial parasite appears to have developed effective mechanisms for evading the host immune system. Some malarial peptides exhibit antigenic diversity (i.e. are polymorphic within the local plasmodium population) and therefore do not stimulate immunity developed against another strain (McCutchan et al. 1988). Plasmodium sp. has also developed the capacity to constantly evolve its antigenic determinants within the host and thus evade immune recognition (Mendis et al. 1991). It
has also been proposed that during invasion of the human host, the parasite presents a number of non-essential antigens that probably act as a "smoke screen" to swamp the immune system with responses delaying the development of an effective immune response (Marsh and Howard 1986; Miller et al. 1994).

Secondly, the effectiveness of a vaccine can be limited by the extent of host immune restriction and innate resistance. Low HLA restriction, whereby certain HLA antigens are able to present some epitopes less efficiently than others, may affect the ability of a vaccine to elicit an appropriate T cell response. However, HLA restriction of a potential vaccine epitope was not evident in a recent study in Madang, PNG (Graves et al. 1989).

Thirdly, the genetic background is important in the evaluation of vaccine efficacy, as some protective variants may have become well established in a population. Two examples of these are the presence of sickle cell haemoglobin (providing protection against P. falciparum) (Pasvol et al. 1978) and the absence of the Duffy antigens (preventing P. vivax invasion) (Miller et al. 1976). As the effectiveness of a vaccine is dependent upon its ability to boost the natural immune response, individuals possessing protective gene variants may fail to develop an effective vaccine-based immunity. A comprehensive understanding of the role of these protective genotypes is important in the evaluation of vaccine efficacy.

1.2 The Malaria Hypothesis

It is now well accepted that some genetic variants have become established in human populations because of the protection they provide from malarial infection. J.B.S. Haldane was the first to suggest that the co-distribution of sickle cell anaemia and endemic malaria might advocate a selective advantage for individuals carrying a sickle cell haemoglobin gene (HbS) (Haldane 1949). The clinical advantage for heterozygotes was demonstrated by Brain (1952) and was later shown to correspond with a decreased incidence of plasmodium parasitaemia (Allison 1954). However, it was not until the success of P. falciparum in vitro culturing by Trager and Jensen (1976) that this protection could be
experimentally demonstrated and the protective mechanism elucidated (Friedman 1978).

HbS is not the only haemoglobin variant believed to play a protective role in resistance to plasmodium infection (see Table 1.1). HbC, found predominantly in West Africa (Allison 1956), and HbE, in South-East Asia (Myint et al. 1991), have also been linked to protection from malaria (Livingstone 1971). The geographical correlation has been further supported by in vitro demonstration of decreased \textit{P. falciparum} invasion in HbCC erythrocytes (Friedman et al. 1979). Similar in vitro experiments have shown that erythrocyte invasion by \textit{P. falciparum} is inhibited in HbEE erythrocytes (Nagel et al. 1981). Unlike the sickle cell haemoglobin, there is no apparent in vitro protection for HbAC or HbAE heterozygotes (Friedman et al. 1979; Nagel et al. 1981).

The intraerythrocytic variants \( \alpha \) and \( \beta \) thalassaemia and glucose-6-phosphate dehydrogenase (G6PD) deficiency are also considered to provide protection from malaria (Allison 1954; 1960; Motulsky 1960). There is now considerable epidemiological evidence suggesting a relationship between the joint distribution of \( \alpha \) thalassaemia and malaria (Higgs and Weatherall 1983; Hill et al. 1987). However, in vitro evidence has so far failed to show any decrease in parasitic invasion of \( \alpha \) thalassaemic erythrocytes, or inhibition of their intraerythrocytic growth, unless under oxidative stress (Pasvol and Wilson 1982).

Siniscalco et al. (1966) have demonstrated a correlation between \( \beta \) thalassaemia and malaria in Sardinia. The persistence of foetal haemoglobin (HbF) in \( \beta \) thalassaemic children has been suggested as the mechanism by which protection is conferred, as increased levels of HbF in erythrocytes have been demonstrated to decrease \textit{P. falciparum} growth in vitro (Pasvol et al. 1976, 1977). However, as with \( \alpha \) thalassaemia, in vitro studies have failed to show any decrease in parasitic invasion or growth in \( \beta \) thalassaemic erythrocytes, unless under conditions of physiological stress (Flint et al. 1986).

The protective advantage of G6PD deficiency is complicated as G6PD is encoded by an X-linked gene. As a result of X-inactivation,
## Table 1.1 World distribution of malaria-protective genetic variants

<table>
<thead>
<tr>
<th>Region</th>
<th>Reported cases of malaria in 1987 (x1000)</th>
<th>α thal&lt;sup&gt;1&lt;/sup&gt;</th>
<th>β thal&lt;sup&gt;2&lt;/sup&gt;</th>
<th>HbC&lt;sup&gt;3&lt;/sup&gt;</th>
<th>HbE&lt;sup&gt;4&lt;/sup&gt;</th>
<th>HbS&lt;sup&gt;5&lt;/sup&gt;</th>
<th>Duffy -ve&lt;sup&gt;6&lt;/sup&gt;</th>
<th>G6PD def.&lt;sup&gt;7&lt;/sup&gt;</th>
<th>Gerb -ve&lt;sup&gt;8&lt;/sup&gt;</th>
<th>Oval&lt;sup&gt;8&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>3309</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Americas</td>
<td>1019</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SE Asia</td>
<td>2823</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Europe</td>
<td>27</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>E Mediterranean</td>
<td>564</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Western Pacific</td>
<td>758</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

NB: Regions are classified according to the WHO<sup>1</sup> Abbreviations: α thalassemia (α thal), β thalassemia (β thal), haemoglobin C (HbC), haemoglobin E (HbE), haemoglobin S (HbS), Duffy negative (Duffy -ve), Glucose-6-Phosphate Dehydrogenase (G6PD) deficiency, Gerbich negative (Gerb -ve) and Ovalocytosis (Oval). References: <sup>1</sup>WHO 1990 <sup>2</sup>Weatherall 1987 <sup>3</sup>Motulsky 1960 & Allison 1961 <sup>4</sup>Hill 1992 <sup>5</sup>Allison 1961 <sup>6</sup>Miller <i>et al.</i> 1976 <sup>7</sup>WHO 1989 <sup>8</sup>Serjeantson <i>et al.</i> 1992
approximately half of the erythrocytes of female G6PD deficiency heterozygotes are G6PD deficient, while the remaining half are G6PD normal (Beutler et al. 1962). This erythrocyte mosaic is considered to be unsuitable for optimal plasmodium development, thus increasing survival of female G6PD deficient heterozygotes (Bienzle et al. 1972; 1979).

Populations of African origin, who commonly lack the Duffy antigens (Fy a-b-) (Sanger et al. 1955), also appear to be largely resistant to *P. vivax* infection (Young et al. 1955; Bray 1958). Unfortunately there is, as yet, no method of *in vitro* culturing of *P. vivax* to directly demonstrate this phenomenon. The simian parasite, *P. knowlesi*, which shares many similarities with *P. vivax*, can be cultured and, as a consequence, many of the *in vitro* predictions of how the Duffy negative erythrocytes inhibit *P. vivax* invasion are based on studies of *P. knowlesi* (Miller et al. 1975; 1979). Nevertheless, the total absence of *P. vivax* infection in Duffy negative erythrocytes in individuals of all ages is undoubtedly the most convincing evidence of *in vivo* protection (Young et al. 1955; Bray 1958).

HbS and Duffy negative phenotypes are not present in the malaria affected countries of the Western Pacific (Serjeantson et al. 1992). However, there are several local variants which are considered to be involved in resistance to malarial infection (the Gerbich-negative blood group and ovalocytosis). *In vivo* studies showing decreased parasitaemia or parasite density in carriers of these variant genes have been used to demonstrate their role in protection from malarial infection (Serjeantson et al. 1977; Cattani et al. 1987; Schuurkamp et al. 1989; Serjeantson 1989e). These variants, and their distributions in the Western Pacific, will be discussed in detail later in this chapter.

1.2.1 Malaria as a selector

There are several suggested mechanisms by which malaria, unlike any other infectious disease, is able to successfully maintain itself in so many populations. Its success is largely attributable to its high infectivity rate and its ability to maintain a constant
parasite reservoir in the human host, even under circumstances of low population density. It does this in several ways.

A single exposure to malaria infection is not sufficient to induce life-long immunity. This is because of Plasmodium sp. ability to generate enormous antigenic diversity. Life-long immunity is only possible once an individual has experienced infection with all antigenic variants in the local parasite population (Gupta et al. 1994). As the parasite numbers are relatively high in non-immune and infected individuals, there will always be a reservoir of infected individuals in the host population to help maintain the plasmodium transmission.

Acquired immunity can be lost during pregnancy (Riley 1983) or in the absence of regular exposure. In populations experiencing stable malaria, infection (and therefore transmission) can be maintained in non-immune children, pregnant women and immigrants. Even in low density human populations, malaria transmission is maintained because it alternates between two hosts, bypassing the requirement of other infectious diseases (e.g. measles and influenza) for person to person contact. Regular transmission is also aided by the lengthy duration of infection; the gametocytes can persist in the blood stream for up to 3 months.

Intermediate pathogenicity is another important aspect of malaria's success. During infection, merozoite numbers need to be large enough so that gametocytes will be picked up in the next mosquito's blood meal. However, parasite density is not often so high as to lead to the elimination of the host as this would prevent subsequent transmission. Ironically, P. falciparum, responsible for the majority of malaria-related fatalities, is also the most potent selector of host protective genotypes, as the host has developed protective genetic mechanisms to avoid P. falciparum infection.

1.3 Malaria in Papua New Guinea

The island of New Guinea is located in the Western Pacific region of Melanesia, north of Australia, between the equator and latitude 11° south and between longitudes 130° and 156° east. Politically,
the island is divided into two halves; the western half is a state of Indonesia, the independent nation of Papua New Guinea (PNG) occupies the eastern half. Consisting of the mainland and associated island groups, PNG is characterised by extreme diversity in both its people and its geography. Several mountain ranges, rising to over 4000m above sea level in some parts, divide the northern and southern regions of the mainland. A number of highland valleys are located within the mountainous ranges. The lowlands and coastal areas are covered by alluvial soils and extensive swamps. It lies entirely within the tropics, although climatic conditions vary greatly from region to region (Allen 1983, 1992).

Malaria is a major health concern in PNG with an estimated two-thirds of the population living in potentially malarious areas. Conditions for malaria transmission are ideal in the lowland and coastal regions where temperature, humidity, rainfall and population density are high and there is an abundance of the highly efficient *Anopheles punctulatus* mosquito vector (van Dijk and Parkinson 1974).

A characteristic symptom of malaria is an enlarged spleen. The spleen rate (proportion of children aged between 2 and 10 years with enlarged spleens) is regularly used as a measure of the intensity, or endemicity, of malaria transmission (Christophers 1924). Holoendemicity is defined when the spleen rate (splenomegaly) of children is constantly over 75%. Hyperendemicity is defined by 50-75% splenomegaly, mesoendemicity 10-50% and hypoendemicity as <10% (see Figure 1.2).

Surveys in the highland regions indicate that malaria is largely absent above 2000 metres, but transmission can vary between regions. For example, *Anopheles punctulatus* has been recorded as high as 2300m near Mt Hagen, but was not reported in the Goroka Valley, where the altitude is between 1600 and 1700m (Peters et al. 1958). The altitude dependency of malarial transmission is attributed to the 15°C minimum temperature requirement for plasmodium development (Trustees of the Wellcome Trust 1991). Malaria transmission is uncommon in the highlands; however,
Figure 1.2 Malaria Endemicity in Papua New Guinea
(from Parkinson et al. 1974, reprinted with permission of Papua New Guinea Medical Journal)
transmission has increased since European contact due to the improvement of roads, building of airstrips and creation of temporary collections of water ideal for *Anopheles* breeding (Radford et al. 1976).

1.3.1 Past control strategies

The past strategy for malaria control in PNG has been comprehensively reviewed by Spencer (1992) and Montanari et al. (1992). This control strategy involved three stages. The first, from 1946-1956, concentrated on the control of mosquito larvae and administration of antimalarial drugs. The second stage involved residual indoor spraying of DDT, starting in 1957 in Maprik (East Sepik Province) and Goodenough Island (Milne Bay Province) and gradually spreading to most parts of the country (see Parkinson 1974). Spraying was supplemented by mass drug administration (MDA) (eg. nivaquine and paludrine) in 1965. As the third stage of the strategy, residual spraying efforts continued from 1970 to 1983. Although the initial hope was for malaria eradication it was accepted by 1969 that this was too difficult a task and the aim instead became the control of malaria (Spencer 1992). Both spraying and MDA were discontinued after 1983, due to substantial cuts in financial and human resources, but epidemiological surveillance has been maintained.

All four human *Plasmodium* species (*P. vivax, P. falciparum, P. malariae* and *P. ovale*) are present in PNG (van Dijk and Parkinson 1974). The predominant species in the 1950s was *P. vivax* (Peters et al. 1958), but by 1990, 76% of over 100,000 malaria positive slides were *P. falciparum* with only 20% *P. vivax* (Montanari et al. 1992). The change in species dominance can largely be attributed to the interruption of malaria control, changing attitudes to drug usage (Cattani 1992) and rising incidence of *P. falciparum* resistance to chloroquine (see Spencer 1992). If the latter is a major factor, the more recent occurrence of *P. vivax* resistance to chloroquine (Rieckmann et al. 1989; Schuurkamp et al. 1992) might readjust the *P. vivax / P. falciparum* balance.

The primary *Anopheles* vectors in PNG are of the *Anopheles punctulatus* complex: *A. punctulatus* Donitz, *A farauti* Laveran and
A. koliensis Owen (Spencer et al. 1974), although it has been suggested that the Anopheles punctulatus complex may consist of as many as nine species (Bryan et al. 1990).

The populations of PNG have survived in the presence of malaria, and several factors are believed to have contributed to their ability to do so. One of the factors is the population acquisition and amplified frequency of protective genotypes. However, before considering the factors involved in innate resistance to malaria, the contribution of other protective factors should be considered.

1.3.2 Protective effect of protein deficiency

Several environmental influences have been suggested as possible mechanisms for protection from malarial infection. There is extensive debate as to whether protein deficiency promotes or hinders malarial infection. One epidemiological study has shown that deaths due to malaria rose 200% a few months after famine relief efforts had commenced in The Bengal (Ramakrishnan 1954). Animal model studies have also indicated that nutritional deficiencies can limit the multiplication of malarial parasites. These include resistance to P. lophurae in vitamin A-deficient chickens (Ross et al. 1946), P. berghei depression in p-aminobenzoic acid-deficient mice (Hawking 1954) and inhibition of P. knowlesi in starved and ascorbic acid deficient monkeys (McKee and Geiman 1946). However, as malnutrition also impairs T cell function, important in modulating protective immunity, it has been suggested that protein deficiency might in fact enhance malarial infection in neonates (Chandra 1977; Watson and McMurray 1979).

The role of iron deficiency (mostly as a result of an inadequate dietary protein) in immunity to malarial infection is also controversial. Several studies have recorded an increase in malarial infection following iron therapy (see Oppenheimer et al. 1986). In a controlled trial of iron therapy in neonates from the Madang Province, PNG, it was demonstrated that the rate of infectious diseases, and incidence of malarial infection, increased in neonates administered with iron dextran (Oppenheimer et al. 1986a, b). However, a study of semi-immune school children from the same region showed no such increase in plasmodium infections.
in children receiving iron supplements (Harvey et al. 1989). The authors suggest that any protection provided by iron deficiency from malarial infection probably occurs prior to the development of malaria immunity (i.e. in the non-immune window of neonates).

1.3.3 Distribution of protective genotypes in the Pacific

In Oceania the distribution of malaria protective genotypes is restricted to α+ thalassaemia, β thalassaemia, G6PD deficiency, Gerbich negative phenotype and ovalocytosis. As mentioned previously, HbS and the Duffy negative phenotype have not been detected in Oceanic populations. As malaria transmission only occurs in a subset of Western Pacific countries, this region provides a convenient testing ground for the assessing whether the presence of protective genotypes in a given population is indicative of a selective response to malarial infection, or merely reflects a random genetic event.

Professor Sue Serjeantson and her colleagues (Serjeantson et al. 1977; Serjeantson 1989e) recognised that the distribution of the Gerbich-negative blood group phenotype was limited to northern coastal populations of PNG where malaria is endemic, and proposed that the gene responsible for this phenotype might be selected by malaria. The limited distribution of ovalocytosis, a condition in which erythrocytes are oval in shape, to Southeast Asian and Papua New Guinean populations has also been similarly attributed to its selection by malaria (Lie-Injo 1965; Baer et al. 1976; Amato and Booth 1977; Serjeantson et al. 1977; Holt et al. 1981).

Numerous studies have demonstrated that the high frequency and distribution of the α thalassaemia resulting from a single α globin gene deletion (-α/) in the Western Pacific is due to malaria selection (Oppenheimer et al. 1984; Hill et al. 1985, 1987; Hill 1986, 1992; Yenchitsomanus et al. 1985, 1986a, b; Yenchitsomanus 1986; Flint et al. 1986).

The correlation between the distribution of β thalassaemia and malaria in the Pacific is less clear. An early study of the distribution of β thalassaemia in PNG found β thalassaemia to be present in high frequencies in coastal populations but largely
absent from the highlands (Curtain et al. 1962). However, a subsequent study in the Markham Valley failed to confirm this altitude-dependent association (Giles et al. 1967).

Apart from minor variation in gene frequencies attributable to genetic drift, founder effect or migration, the high incidence of G6PD deficiency in the lowland and coastal regions and its absence from the highlands has also highlighted its possible role in malaria protection in PNG (Yenchitsomanus et al. 1986b).

1.3.4 Factors influencing the gene distributions

There are several factors that influence the presence and frequency of a gene within a given population. These include the original gene pool, gene migration, founder effect, random genetic drift and selection. Before selection by malaria can be ascertained to be responsible for the increased frequency of a protective genotype in a given population, these factors must be considered. In PNG, for example, the difference in gene frequencies between the highland and coastal populations might reflect their different original colonising populations or a founder effect, rather than selection by malaria. To better understand the effect of colonising populations on gene distribution in PNG, a brief summary of the current theories of the colonisation of the Pacific region, based on linguistic, archaeological and genetic evidence, is presented in the following paragraphs.

The distributions of the three major language groups (Australoids, Papuans or Non-Austronesians, and Austronesians) in Oceania provide a rough guide to migrational movements within the region. The consensus amongst linguists is that the first human populations to colonise the Pacific were Austrqloid speakers, who migrated into Australia, New Guinea, the Bismarck Archipelago and Northern Solomons from the west between 30,000 and 50,000 years ago while the present day New Guinea and Australia were still joined together as Sahul-land (White and O'Connell 1982). Languages descending from the original Austroloid speakers are still spoken by Australian Aborigines. However, the original Australoid speakers in PNG have apparently been replaced or absorbed by incoming populations as no present day Papua New
Guinean languages are proposed to have descended directly from the Austridoid speakers.

The next wave of colonisers, Papuan or Non-Austronesian (NAN) speakers, are believed to have first arrived in PNG some 15,000 years ago (Wurm 1983), with subsequent waves of NAN speakers arriving after the physical separation of New Guinea and Australia over 8,000 years ago (Serjeantson 1985). According to Wurm, a major NAN expansion took place about 5,000 years ago, spreading the populations, and their languages, eastwards towards the Bismarck Archipelago and the Solomon Islands (Wurm 1983).

The Austronesian (AN) speakers, originating from the southern China and Taiwan region, migrated to Oceania around 3,500 to 4,000 years ago (Bellwood 1989), intermarrying with the coastal NAN speakers of New Guinea, Vanuatu and New Caledonia (Serjeantson 1985). A second wave of Austronesian speakers, characterised by their Lapita-style pottery, arrived from Indonesia about 3,000 years ago and by-passed the northern coast of PNG, settling in Island Melanesia and western Polynesia (Green 1979). However, language group similarity and archaeological artifacts may not always predict the actual relationships and origins of present day populations, as cultural skills can be passed from one population to the next without the need for genetic mixing.

Polymorphisms of red cell antigens, serum proteins and enzymes have been exploited to postulate the settlement of Oceania. In addition, these variants have been used to predict variation within localised island and mainland communities of PNG (see R Kirk, 1989 and 1992 for reviews). A study by Boyce et al. (1978), investigating the genetic relationships between 14 village clusters of the two linguistic groups of Karkar Island, the NAN Waskia-speakers and the AN Takia-speakers, demonstrated genetic similarity between populations of similar linguistic background. Serjeantson et al. (1983) extended this study to include 17 populations from the north coast of PNG and demonstrated a significant correlation between kinship and geographical distance, suggesting that distance, rather than linguistic similarity, might determine genetic affinities. The Waskia and Takia villages are also
geographically separated from one another, which might explain their apparent genetic differences.

The extreme diversity of the HLA system has also been used to predict the level of genetic association between Oceanian populations (Serjeantson et al. 1982, 1987; Serjeantson 1985, 1989b, c; Serjeantson and Easteal 1991). Broadly, the linguistic relationships of these populations are supported by this genetic evidence, although exceptions exist (Serjeantson et al. 1992).

DNA-based analysis has allowed the tracing of population-specific variant genes through Oceanic populations. Hertzberg et al. (1988) used α globin haplotypes to propose that Polynesians have mixed with Melanesian populations, as both populations share common haplotypes (Hertzberg et al. 1988). The presence of similar phenylalanine hydroxylase (PAH) haplotypes in Polynesian and South East Asian populations further supports this hypothesis (Hertzberg et al. 1989a).

Polymorphism within mitochondrial DNA (mtDNA) have also been used to anticipate population associations. An Asian-specific 9bp deletion in mtDNA was found in 93% of Polynesian populations, suggesting an Asian component to pre-Polynesian history (Hertzberg et al. 1989b). Mitochondrial DNA analysis has also demonstrated that even though the islanders of Fiji have an equal mix of Melanesian and Polynesian nuclear DNA variants (Serjeantson et al. 1982; Trent et al. 1988), the original female colonisers of Fiji were probably of Polynesian stock (Hertzberg et al. 1989b).

However, the predicted association of populations based purely on DNA must also be interpreted with caution, as the DNA polymorphisms may be subject to the evolutionary pressures of random drift, fixation, founder effects and selection (Feldman and Christiansen 1975; Serjeantson et al. 1982). The most reliable picture of past colonisation of the Pacific is obtained when each of these lines of evidence are considered in conjunction with one another (Serjeantson 1989a).
Most genetic studies have shown a clear division between the more recent arrival of the coastal (AN-speaking) Melanesians to PNG and the comparatively ancient genes found in the PNG highlanders that are believed to have been introduced by progressive waves of Papuans (NAN-speaking) (Serjeantson et al. 1982). This population division also reflects the separation of those populations affected by endemic malaria into the AN-speakers of the low lying and coastal regions and the highland NAN speakers not affected by endemic malaria. Therefore genes found at high frequency in malaria endemic areas may be a function a common genetic heritage rather than selection by malaria. The NAN-speaking populations in low lying regions of PNG allow assessment of the selective pressure of malaria on gene frequencies independent of their original genetic roots.

1.4 A Microepidemiological Study of Malaria Resistant Markers

The majority of studies investigating genetic resistance to malaria in Papua New Guinea have concentrated on one or two genotypes, their frequencies and their geographical distribution in relation to the distribution of endemic malaria (Kidson 1961; Giles et al. 1967; Serjeantson et al. 1977, 1989e; Oppenheimer et al. 1984; Yenchitsomanus et al. 1986b; Cattani et al. 1987; Hill et al. 1987, 1988; Schuurkamp et al. 1989; Brabin and Brabin 1990). Although the correlations in general are high, the relationship is less than perfect as a number of exceptions exist. Besides, most of these studies have failed to account for the combined impact of protective genotypes at an individual or a population level. The microepidemiological approach of this study offers a more comprehensive picture of the effect of malaria as a selector of protective genotypes at this level.

1.4.1 Study population

Since the populations most likely to experience the selective forces of malaria on their gene frequency distribution are those under constant exposure to malaria infection (i.e. stable transmission), the Wosera population, located in the north of the
mainland PNG and where a number of protective genotypes have been previously described, has been selected for this study.

The Wosera area of the East Sepik Province (see Figure 1.3) is located south of the Maprik township and is approximately 100km inland, directly north of the Sepik river. The Woserans are Abelam speakers, members of the Ndu family and Sepik-Ramu phylum of non-Austronesian languages (Laycock 1973). The Abelam speakers are assumed to have migrated from the Sepik River region, where cultural and linguistic links still remain, moving north across the Sepik plains into the forested area. Their agricultural practices turned forests into grasslands, and they continued to move in a north westerly direction to the base of the Torricelli foothills. European intervention in the 1930s halted their encroachment on the people of the Torricelli hills, and with nowhere further to expand, the Abelam people have experienced an increase in population density (Allen 1983). They experience stable malarial transmission, high mortality and morbidity, have a relatively poor nutritional status (Gibson et al. 1991) and a low level of socioeconomic development (Alpers et al. 1992).

1.5 Detection of Malaria Resistant Phenotypes

To determine whether malaria protective genotypes and/or protein variants have been influenced by the selective pressures of malaria, the frequency of these variants are often compared between populations that are, and those that are not, exposed to malarial infection. As the protective influence of a variant genotype may be subtle, the detection of such a protective relationship is heavily reliant upon the accurate typing of variant genotypes. Traditionally serology, protein electrophoresis and microscopy have been used to detect these variants. Over the years new techniques, including RFLP analysis and PCR based DNA typing, have been developed to enable a better assessment of genotypes. This study has used a mixture of these techniques to detect malaria-associated polymorphisms.
1.5.1 Serology

Initially a number of genetic variants were identified by their ability (or inability) to be recognised by common antibodies. For example, the absence of Gerbich blood group antigens in PNG was discovered when a New Guinea woman produced a strong immune IgG type antibody (anti-Ge\(^a\)) during blood transfusion (Booth et al. 1970). These anti-Ge\(^a\) antibodies were then used to screen populations for the absence of the blood group (Booth and McLoughlin 1972). The depression of the \(I^T\) blood group antigens similarly was reported in coastal Papua New Guineans by Booth and colleagues (Booth 1972; Booth and Simmons 1972), which was later demonstrated to correspond to individuals with oval erythrocytes (Booth et al. 1977). Lately HLA typing has been carried out in PNG using serological approaches (Crane et al. 1985).

Serological techniques are limited by their requirement of intact erythrocytes or viable lymphocytes which are often difficult to maintain in a field situation. Non-availability of population-specific antisera compounds the problem further. For example, the complement-dependent lymphocyte microcytotoxicity test (Terasaki and McClelland 1964) is the standard method used for the determination of class I and class II HLA antigens. However, the antisera panels for HLA typing in PNG were largely derived from European populations which often resulted in a high incidence of serological blanks, especially at the HLA-C locus (Crane et al. 1985; Bhatia et al. 1989).

1.5.2 Variant proteins

Many population studies of G6PD deficiency have used enzyme assay techniques to detect the decreased G6PD activity of the sample (e.g. "fluorescent spot test" (Beutler and Mitchell 1968), and "MTT spot test" (Harris and Hopkinson 1978). While these techniques are a quick and convenient method of detecting severely deficient hemizygous males, the evaluation of female heterozygotes is unreliable as X-inactivation generates a mixed pool of G6PD-deficient and G6PD-normal erythrocytes.
Haemoglobin variants (e.g. HbS and HbJ\textsuperscript{Tongariki}) were initially identified because of their altered electrophoretic mobility (Pauling \textit{et al.} 1949; Gajdusek \textit{et al.} 1967). Similarly, prior to DNA analysis, $\alpha^+$ heterozygotes (-$\alpha$/+$\alpha$) with very mild phenotypes were difficult to detect and could only be identified by the raised levels of Hb Barts ($\gamma_4$) in $\alpha^+$ thalassaemia neonates. $\beta$ thalassaemia carriers, who have elevated levels (>3.5 per cent) of HbA$_2$ ($\alpha_2\delta_2$), were also detected using a microchromatographic quantitation method (Huisman \textit{et al.} 1975).

1.5.3 Microscopy

Microscopy is used routinely to determine the presence of plasmodium infections in thick and thin blood slides (Shute 1988). The technique is inexpensive and, as there is no requirement for electricity, provides a convenient method for field based research. It is still one of the most accurate and reliable methods for determining parasite readings (Lalloo and Naraqi 1992). Microscopy has also been used for the diagnosis of ovalocytosis, characterised by the presence of normochromic and oval erythrocytes, the presence of stomatocytes, knizocytes and the lack of erythrocyte rouleaux formation (Amato and Booth 1977). As erythrocytes may also become oval in shape in response to stress (e.g. hypochromic anaemia), the secondary factors that characterise this condition must be identified before an accurate diagnosis can be made (Nurse 1981).

1.5.4 DNA Analysis

The advent of molecular genetic analysis has allowed a more accurate determination of the nature of variant genes (see Steel 1984; Hayes \textit{et al.} 1989; Miles and Wolf 1989 for reviews). This variation may be due to a small change in the coding sequence, leading to missense mutations, amino acid substitutions, or changes in reading frame leading to a change of termination signals, or due to large deletions, insertions or reorganisations. For example, restriction fragment length polymorphism (RFLP) analysis now allows detection of significant aberrations in the genome that lead to protein variants (e.g. Gerbich negative and $\alpha$ thalassaemia gene deletions) and linked haplotype information.
from normal genes (e.g. $\alpha$ globin), which can shed light on the evolutionary distribution of genes (Hertzberg et al. 1988). The mode of inheritance of variant genotypes (e.g. recessive or dominant) can also be determined as DNA analysis allows the discrimination of homozygotes and heterozygotes. Genetic variants with asymptomatic phenotypes (e.g. $\alpha^+$ thalassaemic heterozygotes) can similarly be accurately determined.

Major hurdles in the application of DNA based analysis to this type of research are expense and the time required for analysis. The determination of the molecular basis of an unknown variant often requires DNA sequencing and can be a lengthy process if the gene in question is large. Even though the DNA analysis is valuable in understanding both the genetic basis of selection and the protein variants, the importance of the phenotype can not be ignored, as phenotypic expressions often provide valuable information about the protective mechanisms (Vulliamy et al. 1993). For example, the absence of antigens in Duffy negative ($P. vivax$ invasion-resistant) erythrocytes indicate their possible role in the invasion process. It has since been demonstrated that the Duffy antigens are responsible for merozoite apical junction formation prior to erythrocyte invasion (Miller et al. 1979).

1.6 Aims of the project

This study aims to assess the presence and frequency of protective genotypes within a single Papua New Guinea population using highly specific molecular genetic techniques.

Previous biochemical characterisation of G6PD deficiencies in PNG has suggested a high degree of heterogeneity (Kirkman et al. 1968; Yoshida et al. 1973; Chockkalingam and Board 1980; Chockkalingam 1981). Detailed biochemical characterisation of a Wosera G6PD deficient sample was undertaken (Chapter 3) to allow comparison of the molecular basis of this deficiency (Chapter 4) with previously characterised biochemical variants in PNG. The molecular analysis of the Wosera G6PD-deficient sample, and a G6PD sample from the southern coast of PNG (Kalo), confirms the underlying heterogeneity of G6PD deficiency in PNG previously suggested by biochemical analysis. Screening of G6PD-deficient
samples from nine PNG populations reveals the presence of further undetected molecular variants.

Immune diversity within this population was assessed by determining the frequency of HLA class II alleles in the population (Chapter 5). The HLA allele frequencies were determined to assess whether the presence of malaria may have lead to an increase in HLA allele diversity and heterozygosity. Analysis has identified further haplotypes between HLA-DRB1, -DQA1 and HLA-DRB1, -DQB1 loci. A comparison of HLA allele frequency in this population with HLA alleles in coastal (endemic) and highland (non-endemic) populations was also made.

The gene frequencies of Gerbich negative (Chapter 6), α thalassaemia and ovalocytosis (Chapter 7) were determined. A comparison of parasite infection and density with the α thalassaemia, Gerbich negative, and ovalocytosis genotypes has been made to determine whether the presence or absence of these genotypes influenced the presence and/or density of *Plasmodium sp.* in the Wosera population.

The Discussion draws together this information, assessing the contribution that selection by malaria has made to the frequency distribution of each protective genotype in this population. The influence of combined genotypes on protection from malarial infection is also considered. Suggested future directions that work in this field might take conclude this thesis.
Chapter 2

MATERIALS AND METHODS
2.1 Introduction

Evaluation of mechanisms by which certain genes offer protection from malarial infection has often been limited by the ability of available techniques to clearly distinguish the precise genotype of individuals. This study has attempted to overcome this limitation by using molecular genetic techniques in typing some of these genotypes.

A screening protocol using starch gel electrophoresis was initially used to determine the presence of G6PD deficiency. Polymerase chain reaction (PCR) amplification, heteroduplex analysis, dideoxy sequencing, restriction analysis and sequence specific oligonucleotide (SSO) typing techniques were used to further characterise the molecular basis of this deficiency. A comparative study of the biochemical characteristics of G6PD deficiency within an affected family was also carried out. A PCR based protocol was similarly utilised to identify the 27bp deletion in the band 3 gene involved in ovalocytosis. Restriction fragment length polymorphism (RFLP) analysis helped determine the presence of α globin gene deletions of α thalassaemia and the 3.5kb deletion in the glycophorin C gene characterising the Gerbich negative blood group phenotype. PCR-SSO typing was used to characterise the HLA class II alleles present in the study population.

2.1.1 Wosera samples

Blood samples were collected from 237 individuals from the Wosera region of Papua New Guinea (PNG) and airfreighted to Canberra, Australia for further analysis. The DNA samples were extracted from the buffy coats of 85 Wosera adults. Further blood samples were collected from a Wosera family to allow biochemical analysis of the G6PD enzyme. Blood and subsequent DNA samples of Caucasian controls were obtained from members of the Molecular Genetics Group, JCSMR.

2.1.2 DNA extraction

Genomic DNA was prepared using a DNA extraction protocol developed by Miller et al. (1988), with some modifications. Briefly
2 ml of buffy coats were thawed at RT, collected into a 15 ml polypropylene centrifuge tube, and spun at 1500g for 5 min. The pelleted cells were washed in 2 ml TE (10 mM Tris, 1 mM EDTA, pH 7.8) several times before resuspending in 800 ul TE. Eighty microlitres of 10% SDS were added and the suspension mixed thoroughly. To this, 60 ul of a 10 mg/ml solution of Proteinase K was gently added. The cells were incubated at 52°C for 2 hr and vortexed after 1 hr.

A 1/10th volume of saturated sodium acetate was added and mixed by vortexing for 15 s, then centrifuged for 15 min at 1500 g. The supernatant was transferred to a new tube and an equal volume of isopropyl alcohol added to precipitate the DNA. The resultant DNA was transferred to a 1.5 ml Eppendorf tube and resuspended in 100-400 ul ddH2O, depending upon the quantity of DNA obtained. The OD readings were measured at 260 nm and 280 nm to determine DNA concentration and DNA concentrations adjusted to 0.2 ug/ul.

2.2 Biochemical Analysis of G6PD

2.2.1 Starch gel electrophoresis

Red blood cells from 237 individuals were screened for G6PD deficiency using TEB starch gel electrophoresis (Harris and Hopkinson 1978). Only one male was identified as hemizygous for the deficiency. Subsequently, blood samples were collected from the proband and five members of his immediate family to enable biochemical characterisation of enzyme deficiency. The latter samples were airfreighted to Canberra and were processed for partial G6PD purification immediately. The Caucasian samples were collected at the same time and stored at 4°C until the Wosera samples arrived in Canberra. These controls were processed concurrently.

2.2.2 G6PD enzyme purification

G6PD enzyme activity from six Wosera samples and two Caucasian controls was assessed using the World Health Organization (WHO) approved techniques described by Beutler et al. (1968). The activity was characterised for the following parameters: pH
optimum (5-11.5), Km (NADP), Km (G6P), utilisation of alternative substrates (2-deoxy G6P, deamino NADP and NAD), thermostability at 46°C and starch gel electrophoretic mobility. The starch gel electrophoresis was carried out using three different buffer systems: TEB, phosphate and Tris-chloride.

Briefly, packed red blood cells were obtained from 20-40ml of whole blood and washed three times in 150mM cold NaCl solution. On ice, the cells were haemolysed in a 9x volume of 0.05% β-mercaptoethanol, 0.1% EDTA and 0.01mM NADP and centrifuged at 16,000g in a Sorval centrifuge (SS34 rotor) for 20min. The supernatant was removed to a fresh tube to which 3 volumes of dilute phosphate buffer (4.64mM K₂HPO₄, 15.36mM KH₂PO₄, 0.05% β-mercaptopoethanol) was added. A 0.5-1.0ml aliquot was removed from each sample for G6PD activity assay. DEAE-cellulose was prepared in a buchner funnel and red blood cell haemolysate was filtered through over a period of 10-20min. Free haemoglobin was eluted by passing large volumes of dilute phosphate buffer through the column under a strong vacuum. Once the effluent was free of haemoglobin, the funnel was transferred to a new flask and 50ml of cold eluting buffer (116mM K₂HPO₄, 384mM KH₂PO₄) was slowly passed through. The cellulose was then gently broken up and another 50ml of the cold eluting buffer passed through.

To the pooled eluant, 0.1% neutralized EDTA, 0.01mM NADP and 0.001% β-mercaptoethanol was added, followed by a very gradual addition of 43.6g of solid ammonium sulphate. Once the salt had dissolved, the solution was centrifuged at 16,000g for 30min and the precipitate was resuspended in 1ml holding solution (3.3M NH₄SO₄, 0.05M Tris, 0.1% EDTA, 0.01mM NADP, and 0.05% β-mercaptopoethanol), 2.5ml of dialysis solution (0.05M Tris, 0.1% EDTA, 0.01mM NADP, and 0.05% β-mercaptopoethanol) and 0.5ml saturated NH₄SO₄. The G6PD activity was determined from 10-50ul of the solution (depending upon the activity of the enzyme) and the remainder recentrifuged at 16,000g for 20min. The precipitate was saved and the supernatant assayed for G6PD activity. If the G6PD activity of the supernatant was equivalent to that of the uncentrifuged sample, then the pellet (with no G6PD activity) was discarded; 1ml of 0.5ml saturated NH₄SO₄ was added to the supernatant and centrifuged as before. Once the G6PD
activity in the supernatants began to fall, the pellets were retained. When there was no detectable G6PD activity in the supernatant, the pellets containing the bulk of the G6PD activity were pooled and resuspended in 1ml holding solution. Aliquots of enzyme suspension were dialyzed in 1cm (flat width) Selbys dialysis tubing in circulating cold dialysis solution immediately prior to enzyme characterisation.

G6PD enzyme activity was assessed at 340nm at 25°C using a Cary 1 UV-Visible Spectrophotometer. A standard reaction mix contained 0.1M Tris, 10mM MgCl$_2$, 0.2mM NADP, 0.6mM G6P and 50ul dialyzed enzyme to a final volume of 1ml. A blank control assay was carried out in parallel replacing the enzyme solution with ddH$_2$O.

2.2.3 Biochemical characterisation

2.2.3.1 **pH optimum**

pH optimum for each G6PD enzyme was determined by the use of a range of buffers varying between pH 5.5 and pH 10.9. The assay consisted of pH buffer (50mM Tris, 50mM glycine, 50mM NaH$_2$PO$_4$), 0.6mM G6P, 0.2mM NADP, 20ul dialyzed enzyme and ddH$_2$O to a final volume of 1ml. All assays were duplicated and whenever possible, triplicated. The change in optical density at 340nm was recorded as a percentage of the activity observed at pH 7.

2.2.3.2 **Substrate affinities**

To determine the G6P substrate affinity, Km(G6P) of each sample was assayed in a reaction mix of 0.1M Tris, 10mM MgCl$_2$, and 0.2mM NADP with varying concentrations of G6P (20-240uM) in a 1ml final volume. Km(NADP) was determined by assaying in a reaction mix of 0.1M Tris, 10mM MgCl$_2$, 0.6mM G6P with varying concentrations of NADP (from 5-100uM) in a 1ml final volume.

2.2.3.3 **Affinity for alternative substrates**

A standard enzyme assay was carried out in 0.1M Tris, 10mM MgCl$_2$ and 0.2mM NADP and substituting 0.6mM deoxy G6P for 0.6mM G6P.
Enzyme assays of NADP analogues (NAD and deamino NADP) were carried out in the same reaction mix, but NADP was substituted instead of G6P. The enzyme activity was recorded as a percentage of activity of the normal substrate analogue.

2.2.3.4 Thermal stability

Mutant and control enzymes were assayed and diluted with dialysis solution (0.05M Tris, 0.1% neutralized EDTA, 0.01mM NADP and 0.05% β-mercaptoethanol) so that the mutant enzyme activity was 1/10th that of the control. Tube 1 contained 300ul mutant enzyme, diluted to 330ul with dialysis solution, Tube 2 30ul control enzyme, again diluted with dialysis solution and Tube 3 300ul mutant enzyme and 30ul control enzyme. Tubes 2 and 3 were included as reaction controls (i.e. Tube1 + Tube2 = Tube 3). All tubes contained a final concentration of 0.3mg bovine albumin. The tubes were incubated at 46°C and 50ul of each reaction assayed after 0, 20, 40, and 60mins of incubation. The activity was assessed in a reaction mix of 0.1M Tris, 10mM MgCl₂, 0.6mM G6P and 0.2mM NADP. The enzyme activities assayed from Tube 1 were plotted as a percentage of activity at time 0min.

2.2.3.5 Starch gel electrophoresis.

Starch gel electrophoresis was carried out following Harris et al. (1978) on a horizontal starch gel system using three buffer systems: phosphate (pH 7.0), Tris-chloride (pH 8.8), and TEB (pH 8.6). Electrophoresed bands containing G6PD activity were detected by overlaying a sliced starch gel with a light-sensitive solution of 2mg (3[4,5 dimethylthiazoly1 1-2] 2,5 diphenyl-tetrazolium bromide) (MTT), 2mg phenozine methosulphate, 3mM G6P, 0.2mM NADP, 0.1M Tris pH 8.0 and 10mM MgCl₂ in a final volume of 10ml.

2.3 Polymerase Chain Reaction (PCR) Amplification

The technique of the Polymerase Chain Reaction (PCR) became the basis of the majority of work presented in this thesis. Careful measures were taken to ensure that no contamination occurred. This included regular cleaning of Gilson pipetteman used for
aliquoting PCR reagents, a sterile work environment, sterile 1.5ml Eppendorf tubes, and sterilized mineral oil overlay. DNA-negative PCR controls were run with every PCR batch. Any reagents or products of amplification suspected of contamination were discarded. PCR product quantity and purity was verified by running a 1/10th volume on a 1.5% agarose gel in TAE (40mM Tris-acetate, 1mM EDTA) alongside a molecular weight marker (MspI digested pBR322). Non-specific PCR amplification was eliminated by increasing the annealing temperature.

2.3.1 G6PD

The PCR strategy used for amplifying the coding regions of the G6PD gene (and introns 3, 6, 10, 11 and 12) was originally designed by Poggi et al. (1990) and is described in some detail in Chapter 4 (Section 4.2.3, p67). For this study, the conditions for PCR amplification were modified; these new conditions are summarised in Table 2.1. Each amplification reaction consisted of 200ng genomic DNA, 7-25pmoles each primer of a pair, 10% Taq polymerase buffer (Promega), 200uM dNTPs (Sigma), +/- 10% dimethyl sulphoxide (DMSO), and 2U Taq polymerase (Promega) in a 50ul volume with ddH₂O.

2.3.2 HLA class II

The second (hypervariable) exons of DRB1, DQA1 and DQB1 of Woseran genomic DNAs were amplified. The standard reaction mixture contained 200ng genomic DNA, 50pmole of each primer, 1x Taq polymerase buffer, 200uM dNTPs and 2.5U Taq polymerase, to a 50ul volume. PCR primer sequences, specificities, and amplification conditions are as described by Gao and Serjeantson (1991b) and are summarised in Table 2.2.

2.3.3 Band 3 protein

PCR primers p198 and p199 (Jarolim et al. 1991) were used to amplify a 174bp region of the band 3 gene (nt1098-1253) that spans the 27bp deleted region, using 200ng of genomic DNA, 10 pmoles of each primer, 1x Taq polymerase buffer, 30uM dNTPs and 1 U Taq polymerase in a final volume of 50ul was amplified by 30
Table 2.1. Oligonucleotides and thermal conditions for G6PD amplification.

<table>
<thead>
<tr>
<th>Primer</th>
<th>5'-3' sequence</th>
<th>Exon</th>
<th>DMSO (10%)</th>
<th>Annealing Temp/Time (°C/sec)</th>
<th>Ext. time (sec)</th>
<th>Fragment size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6P2F</td>
<td>CTCTAGAAGGGGCTAATTCTCAA</td>
<td>2</td>
<td>+</td>
<td>56 60</td>
<td>60</td>
<td>241</td>
</tr>
<tr>
<td>G6P2B</td>
<td>GGAATTCCTGCGTTTTAAGATGGG</td>
<td>3,4</td>
<td>+</td>
<td>52 60</td>
<td>60</td>
<td>308</td>
</tr>
<tr>
<td>G6P3F</td>
<td>CAGCCACTTCTAACCAC</td>
<td>5</td>
<td>+</td>
<td>52 60</td>
<td>60</td>
<td>301</td>
</tr>
<tr>
<td>G6P5F</td>
<td>AAGAGAGGGGCTGACAT</td>
<td>6,7</td>
<td>-</td>
<td>54 45</td>
<td>120</td>
<td>542</td>
</tr>
<tr>
<td>G6P6B</td>
<td>ACTGCCGAGAGGAGG</td>
<td>8</td>
<td>-</td>
<td>50 60</td>
<td>60</td>
<td>164</td>
</tr>
<tr>
<td>G6P8B</td>
<td>CCAGCTCCGAGGAGGA</td>
<td>9</td>
<td>-</td>
<td>50 60</td>
<td>60</td>
<td>253</td>
</tr>
<tr>
<td>G6P9F</td>
<td>CAGGAGGGGCTTCTCT</td>
<td>10,11</td>
<td>-</td>
<td>54 60</td>
<td>120</td>
<td>497</td>
</tr>
<tr>
<td>G6P10F</td>
<td>CAGGAGAGCTGCTGCT</td>
<td>11,12</td>
<td>-</td>
<td>52 45</td>
<td>120</td>
<td>665</td>
</tr>
</tbody>
</table>

NB: All PCR cycles included an initial 1min 96°C denaturation step. Extension (Ext.) cycles were at 72°C.
DMSO: dimethyl sulfoxide. aFollowing Poggi et al. (1990).
Table 2.2  HLA PCR primers and conditions for the amplification of the hypervariable region of DRB1 (including DR2 subtyping), DQA1 and DQB1.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Annealing Temp/Time (°C/min)</th>
<th>Codon</th>
<th>Equivalent name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQA15P</td>
<td>52 2</td>
<td>12-17</td>
<td>GH26</td>
<td>Scharf et al. 1986</td>
</tr>
<tr>
<td>DQA13P</td>
<td>86-80</td>
<td></td>
<td>GH27</td>
<td>Scharf et al. 1986</td>
</tr>
<tr>
<td>DQB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE17</td>
<td>54 2</td>
<td>6-13</td>
<td>P2</td>
<td>Morel et al. 1988</td>
</tr>
<tr>
<td>SE8a</td>
<td>86-79</td>
<td></td>
<td>P4</td>
<td>Morel et al. 1988</td>
</tr>
<tr>
<td>DRB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAMP</td>
<td>56 1</td>
<td>94-87</td>
<td>GAMPDRB1</td>
<td>Todd et al. 1987</td>
</tr>
<tr>
<td>GLP</td>
<td>17-23</td>
<td></td>
<td>GLPDRB1</td>
<td>Todd et al. 1987</td>
</tr>
<tr>
<td>DR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRP5</td>
<td>58 1</td>
<td>7-13</td>
<td>PSP-67</td>
<td>Moraes et al. 1991</td>
</tr>
<tr>
<td>XG5</td>
<td>5-11</td>
<td></td>
<td>PSP-68</td>
<td>Moraes et al. 1991</td>
</tr>
<tr>
<td>SSDRM13R</td>
<td>94-87</td>
<td></td>
<td>GAMP with EcoRI/M13R</td>
<td>Todd et al. 1987</td>
</tr>
</tbody>
</table>
cycles of 96°C (70s), 58°C (50s) and 72°C (50s). The size of resultant PCR products was verified on 2.5% agarose/TAE gels.

2.3.4 PCR product purification

PCR products were cleaned and purified using a Geneclean II BIO 101 Kit from Bresatec. Fifty microlitres of each PCR product was run on a 1.5% agarose TAE gel, visualized under longwave UV light, then cut out from the gel using a scalpel blade. The agarose was sliced into small pieces and incubated at 50-55°C with a 2.5 volume of sodium iodide solution for 5 min (or until the agarose dissolved). The tube was transferred to ice and 10 ul glassmilk suspension added by mixing. After 5 min incubation on ice the glassmilk was pelleted and washed twice with 400 ul of "new wash" (Bresatec). The DNA was removed by resuspending the pellet in a half final volume of ddH2O followed by incubating at 50-55°C for 2 min. The suspension was then spun at 14,000 rpm for 30 s and the supernatant transferred to a fresh tube. This was repeated with a second volume of ddH2O. A small volume of the cleaned PCR product (typically 3/20 ul) was run on a 1.5% agarose TAE gel to determine the quantity and quality of extracted DNA.

2.4 Heteroduplex Analysis

Heteroduplex analysis (White et al. 1992), was utilised to quickly screen the G6PD PCR products for sequence mismatches. Five microlitres of PCR products from affected and normal individuals were mixed with 3 ul standard loading dye (15% Ficol, 0.05% bromophenol blue and 0.05% xylene cyanol), denatured at 95°C, then allowed to reanneal slowly. Any mismatches between the control and variant fragments should form heteroduplexes that migrate more slowly on a polyacrylamide gel than do the PCR samples loaded individually.

Heteroduplex reaction products were loaded onto a 6% or 7.5% polyacrylamide gel (with stacking) depending upon the size of the PCR products (products ~300bp were loaded onto a 7.5% gel and of products ~500bp were loaded onto a 6% gel). The gels were 200x400x1.5mm in dimension. Heteroduplex gel plates were cleaned and siliconized with dimethyldichlororosilane (BDH). The
running gel (1xTBE, 6.0 or 7.5% polyacrylamide bis (49:1), 10% glycerol) was degassed and ammonium persulphate (APS) and tetramethylenediamine (TEMED) were added to allow polymerization and the gel was added up to 100mm from the top of the gel cassette. Once polymerization had occurred, the top interface was decanted and the stacking gel (1xTBE, 9 or 7.5% polyacrylamide respectively, 10% glycerol, 50% urea) poured. A 15 well comb was added and the gel left to set for a minimum of 1 hr. Samples were loaded into dry heteroduplex gel wells (previously washed with 1xTBE (90mM Tris, 90mM boric acid, 2mM EDTA) buffer) which were then sealed with 1% TBE agarose. The heteroduplex samples were electrophoresed in 1xTBE at 20mA for 6% gels and 30mA for 7.5% gels. The samples were left to run overnight.

2.5 Sequence Specific Oligonucleotide (SSO) Typing

2.5.1 Dot blotting

This procedure was performed in accordance with the conditions set down by Gao et al. (1992). Briefly, PCR products were electrophoretically separated on six identical sections of a 150x200mm 0.8% TBE agarose gel and transferred onto a nylon membrane (Hybond N+, Amersham) by sandwich blotting. A Polaroid photograph was taken of the gel before blotting for later comparison of PCR amplification efficiency with SSO hybridization intensity.

2.5.2 Prehybridization and hybridization

The two blotted membranes were each cut into six identical filters and marked for individual sequence specific oligonucleotide (SSO) probing. Each gel allowed a maximum of 12 specific probes to be used. Each filter was rolled into a 10ml plastic tube and prewet with 3xSSC (3M NaCl, 0.3M sodium citrate, pH 7.0). Two ml of hybridization solution (15% formamide, 0.1% Denharts solutions (Sambrook et al. 1989), 5x SSPE, 1%SDS, 5% dextran sulphate, 0.2 mg/ml boiled salmon sperm DNA) was added to each tube and the tubes rotated in a hybridization oven for 3-4hr at 42°C.
Three microlitres of each end-labelled SSO probe (see 2.9.2, p46) was added to the appropriate prehybridizing filter and hybridized for 1hr. The filters were collectively rinsed in 3xSSC for 2x 5min at RT and then washed in a TMAC wash solution (3M tetramethyl ammonium chloride, 50mM Tris pH 8.0, 2mM EDTA, and 0.1% SDS) at 59°C for 30min. The blotted dry filters were wrapped in gladwrap and exposed to photographic x-ray (Fuji) film.

2.5.3 G6PD

Genomic DNA samples were amplified using G6P11F and G6P11B primer pairs. PCR products were transferred to nitrocellulose and probed with each of the six G6PD oligonucleotides summarised in Table 2.3.

Table 2.3 Sequence specific oligonucleotides (SSO) for the detection of PNG G6PD mutants.

<table>
<thead>
<tr>
<th>SSO</th>
<th>Codons</th>
<th>5'-&gt; 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EX11K</td>
<td>451-IVS11/4</td>
<td>CAC-TTC-GTG-IGC-AGG-tga-g</td>
</tr>
<tr>
<td>EX11N</td>
<td>451-IVS11/4</td>
<td>CAC-TTC-GTG-CGC-AGG-tga-g</td>
</tr>
<tr>
<td>INT11T</td>
<td>IVS11/82-100†</td>
<td>aagccatacctgtccccct</td>
</tr>
<tr>
<td>INT11C</td>
<td>IVS11/82-100†</td>
<td>aagccataccgtgccct</td>
</tr>
<tr>
<td>EX12W</td>
<td>466-460*</td>
<td>G-TGA-AAA-TAT-GCC-AGG-CCT</td>
</tr>
<tr>
<td>EX12N</td>
<td>460-466</td>
<td>AGG-CCT-GGC-GTA-TTT-TCA-C</td>
</tr>
</tbody>
</table>

NB: EX11K hybridizes to the Kalo nt1360T G6PD variant and EX12W to the Wosera nt1388A G6PD variant. INT11T and INT11C hybridize to the intron 11 T and C-alleles respectively. The sites of sequence variation are underlined. Oligonucleotides with the suffix "N" are from the normal G6PD B sequence (Persico et al. 1986). * Antisense strand sequence; † nucleotide position in intron 11; intronic sequence is denoted in lowercase.
2.5.4 HLA class II

HLA class II region specific PCR products were blotted and screened with sequence specific oligonucleotides (Gao and Serjeantson 1991b). DRB1 SSO probes (SR1-5,7-10,12 and 14) were specific for each major DR allele. Genomic DNA from all DR2 positive individuals was amplified with DR2 specific primers and probed with DR2 subtype-specific SSO probes (Gao and Serjeantson 1991b). HLA DR2, DQA1 and DQB1 alleles were scored from the hybridization patterns obtained (see Table 2.4 and 2.5).

**Table 2.4** Sequence specific oligonucleotides hybridization patterns of HLA DR2 (DRB1 and DRB5) specific alleles.

<table>
<thead>
<tr>
<th></th>
<th>SR26</th>
<th>2</th>
<th>27</th>
<th>28</th>
<th>85</th>
<th>86</th>
<th>87</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DRB1</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1501</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>1502</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>15.3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>1601</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>1602</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><strong>DRB5</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0101</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>0102</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

NB: + positive hybridization
Table 2.5 Sequence specific oligonucleotide (SSO) hybridization patterns of DQA1 and DQB1 alleles

<table>
<thead>
<tr>
<th>QA</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQA1*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0101</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0102</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0103</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0201</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0301</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0401</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0501</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0601</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SQ</th>
<th>B2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQB1*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0201</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0301</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0302</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0303</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0402</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0501</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0502</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0503</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0601</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0602</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0603</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0604</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0605</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

NB: + positive hybridization
2.6 Restriction Analysis of G6PD PCR products

G6PD amplified fragments obtained using the G6P11F and G6P11B primer pair were digested with NlaIII (New England Biolabs) and HhaI restriction enzymes following conditions specified by the enzyme suppliers. Digested PCR products were run on a 6% polyacrylamide gel (1xTBE, 6% polyacrylamide:bis (19:1)) and stained with EtBr for band visualisation.

2.7 G6PD Sequencing

2.7.1 M13 phage cloning

To provide blunt ends for cloning the G6PD PCR products, M13 vector mp18 was digested with SmaI (Pharmacia) using standard conditions (Sambrook et al. 1989). After digestion, the vector was dephosphorylated by the addition of 1U alkaline phosphatase (Boehringer Mannheim) and 1uM Tris/Cl (pH 8.0) in a 50ul final volume. The mixture was incubated at 37°C for 20-30min. Subsequent phenol, phenol/chloroform and chloroform extractions were carried out and the final supernatant precipitated in 1/2 volume of 7M NH4acetate and 3 volumes of absolute ethanol. After 30min at -20°C, the precipitated M13 DNA was centrifuged at 14,000rpm for 15min, washed in 70% ethanol and resuspended to a concentration of 50ng/ul.

2.7.1.1 M13 ligation

Varying concentrations of PCR product were mixed with 100ng SmaI digested M13 vector to allow for an optimum ligation ratio of insert to vector. The ligation mixture contained 100ng vector, genecleaned PCR product insert, 1xBRL ligating buffer (250mM Tris/Cl pH7.6, 50mM MgCl2, 25% w/v PEG, 5mM ATP, 5mM DTT), and 1U T4 DNA ligase (Pharmacia) in a final volume of 10ul ddH2O. The ligations were incubated overnight at R/T.

2.7.1.2 M13 transformation

To enable transformation of M13 (containing PCR inserts) into E. coli (TG1) cells, the ligation reaction was mixed with 100ul E. coli
TG1 competent cells (Dagert and Erhlich 1977), placed on ice for a minimum of 1/2hr and heatshocked at 42°C for 2min. Three mls of H-Top agar (1% bactotryptone, 0.8% NaCl, 0.7% agar, pH 7.0) was added to glass centrifuge tubes and preheated to 48°C. During heatshock, isopropyl-B-D-thiogalactosidase (IPTG) (Progen) and X-galactosidase (X-gal) (Progen) were added to the H-Top to a final concentration of 1uM and 0.02% respectively.

After heatshock, the cells were returned to ice, 200ul of overnight TG1 culture cells (5ml LB inoculated by a single TG1 colony, cultured overnight at 37°C) were added and together these were mixed with the H-Top and plated out on minimal plates (3.75% agar, 1mM thiamine, 1mM MgSO₄, 0.1mM CaCl₂, 1x M9 salts (42mM Na₂HPO₄, 22mM KH₂PO₄, 19mM NH₄Cl and 9mM NaCl) and 0.2% glucose). After overnight incubation at 37°C, the clear plaques, containing M13 with DNA inserts, were grown in separate tubes containing 2ml Luria broth (LB) with 10ul overnight TG1 cells for 5-7hrs.

At least one blue plaque (containing M13 vector with intact β-galactosidase gene and no DNA insert) was grown as a control. The presence of a DNA insert in the M13 vector was confirmed by mixing 20ul supernatant from centrifuged TG1 cells with 5ul DIGE dye (3% SDS in bromophenol blue) and heating at 65°C for 10min. After cooling, supernatant preparations from each plaque culture were electrophoresed alongside similarly treated supernatant from blue plaques without DNA inserts on a 0.7% agarose TAE gel for 3hrs at 70V.

2.7.2 Plasmid vector cloning

Some PCR products did not ligate into M13 vectors. To overcome this problem, the DNAs were ligated into the pGEM-T (Promega) vector. As Taq polymerase sometimes adds a 3' dATP, the pGEM-T vector enables sticky ended, and more reliable, ligations. The pGEM-T multiple cloning sites allow several restriction enzymes (RE) to be used to cleave the DNA insert from the pGEM-T vector and sticky-end clone it more effectively into an M13 vector digested with the same RE combination. pGEM-T ligations and
transformations were carried out in accordance with the manufacturers instructions (Promega Technical Bulletin 1992).

2.7.2.1 Plasmid ligation

Fifty nanograms of pGEM-T vector, 1U T4 ligase and 1x ligation buffer (30mM Tris-HCl, pH 7.5, 10mM MgCl₂, 10mM DTT, 1mM ATP) were added to varying volumes of gene cleaned PCR products in a final volume of 10μl. Once mixed, the ligations were incubated at 15°C for 3hrs. The reaction was stopped by heating the tube to 70°C for 10min, then allowed to cool to RT before pGEM-T vectors (containing inserts) were transformed into TG1 competent cells.

2.7.2.2 Plasmid transformation

The ligation mix was added to 50μl of TG1 competent cells and incubated on ice for 30min. The cells were then heatshocked at 42°C for 50s and placed on ice for 5 min. One ml of Luria broth (LB) was added and the cells incubated at 37°C for 1hr. LB plates (1.5% agar in LB) containing 1mg/ml ampicillin (Amp), and coated with 100μl of 100mg/ml IPTG and 50μl of 2% X-Gal were used to plate out 150-400μl of transformed cells. Once dried, the plates were incubated overnight at 37°C. White colonies, containing inserts, were picked and grown in 5ml LB containing 100ug/ml Amp for 7-16hrs. Plasmid DNA was isolated using a rapid boiling method (Holmes and Quigley 1981).

2.7.2.3 Miniprep plasmid preparation

To isolate the pGEM plasmid DNA from E. coli TG1 cells, 1.5ml of a pGEM-T culture was spun at 14,000rpm for 5min. The resultant pellet was resuspended in 150μl STET (8% sucrose, 5% triton-X100, 50mM EDTA, 50mM Tris pH 8.0) and 1mg/ml lysozyme. After 15min at RT, the cell suspension was boiled for 1min in a water bath. The cells were pelleted by centrifuging for 10min, the globular pellet removed using a sterile toothpick and discarded and 100ul isopropanol was added to the remaining supernatant. After 2-3min at RT the precipitated plasmids were pelleted by centrifuging for 15-20min, washed in 70% ethanol, dried and resuspended in 50ul TE. DNA concentration was determined by running a small volume
(2/50μl) of miniprep mixed with an equal volume of 10x loading buffer containing 0.1mg/ml DNase-free RNase on a 1.5% agarose TAE gel.

2.7.2.4 Religation into M13

Plasmid DNA was digested at common restriction enzyme sites present in the polylinker regions of both the pGEM-T plasmid and M13 phage (typically SphI/PstI and SphI/SalI). The plasmid digest was run on a 1.5% agarose TAE gel and the appropriate band was cut out and genecleaned. The cleaned product was ligated into the equivalently digested M13 vector in a reaction mix of 1x BRL buffer and 1U T4 DNA ligase, incubated at 16°C for 3hrs and transformed as described in Section 2.7.1 above.

2.7.3 M13 ssDNA preparation

Single-stranded DNA was isolated from the insert-containing M13 phage. Eight hundred microlitres of phage supernatant was mixed with 200μl of 20% PEG6000/2.5M NaCl and incubated on wet ice for 15min. The mixture was then centrifuged at 14,000rpm for 5min. The supernatant was completely removed before resuspending the white pellet in 100μl TES buffer (20mM Tris/Cl pH7.5, 0.1mM EDTA, 10mM NaCl). Sixty microlitres of phenol was added, the mixture vortexed, left at RT for 5min, vortexed again, then spun at 14,000rpm for 5min. The aqueous supernatant was transferred into another Eppendorf tube and extracted in 140μl chloroform.

The chloroform extraction was repeated before the supernatant was transferred to a final tube and reprecipitated in a 1/10vol 3M NaOAc and 2.5vol absolute ethanol. The precipitating DNA was left at -20°C for >30min before being spun again for 10min, washed in 70% ethanol and finally resuspended in 25μl ddH2O.

2.7.4 Dideoxy sequencing

DNA sequencing was carried out using Sanger's dideoxy sequencing method (Sanger et al. 1977). The sequencing reactions were manually electrophoresed in 6% polyacrylamide gels, the gels mounted on blotting paper and exposed to x-ray film. All sequences
were read and verified manually and the data entered into a PCGene DNA sequence-processing computer package.

2.7.4.1 Gel preparation

Sequencing gels were made using the Sequagel (TM) System (National Diagnostics). Glass sequencing plates (200mm x 380mm) were separated by 0.3mm thick spacers and 28-well shark tooth combs. The 6% polyacrylamide was prepared by mixing 9.6ml concentrate (25% acrylamide:bis (19:1) & 8.3M urea), 26.4ml diluent (8.3M urea) and 4ml buffer (10xTBE & 8.3M urea). The acrylamide was polymerlized by adding 320ul of 10% ammonium persulphate (APS) and 20ul TEMED immediately prior to gel pouring.

2.7.4.2 Sequencing reaction

The sequencing reaction was performed using the Amersham Multiwell microtitre plate DNA sequencing system - T7 DNA polymerase kit. This kit enables ssDNA, primed by a universal M13 primer, to incorporate radioactively-labelled $\alpha^{35}$S-dATP into the synthesized DNA strands. The kit utilises the Sanger method of dideoxy chain-termination to produce labelled fragments of DNA sequence that can be visualized by running on a polyacrylamide gel and exposing to a photographic film. The film detects the final ddNTP termination of each fragment at the appropriate position within the sequence.

2.8 Restriction Fragment Length Polymorphism (RFLP) Analysis

2.8.1 DNA digestion and Southern blotting

Genomic DNA was digested with BglII, electrophoresed on 0.8% agarose TBE alongside HindIII/EcoRI- and HindIII-digested $\lambda$ DNA molecular weight markers (Bresatec), and transferred onto Biotrace HP membranes (Gelman Sciences). Nick-translated radioactively labelled cDNA probes were hybridized to the filters. The Gerbich negative genotype was defined by hybridization to the pGPC cDNA probe (pFCF22) provided by Dr. J-P Cartron (Le Van Kim et al. 1987). The $\alpha$ thalassaemia genotypes were defined by hybridization to a zeta globin (pBR$\zeta$) probe (Lauer et al. 1980)
2.8.2 Prehybridization and hybridization

Southern blotted (Southern 1975) filters were wet with 3xSSC, then prehybridized overnight at 65°C in 13.5ml standard Nasmyths hybridization solution (0.55M NaCl, 166mM Na₂HPO₄, 5.5mM EDTA, 18.5% dextran sulphate, 1.85% sarcosine) (Nasmyth 1982). Immediately prior to hybridization, 115ul salmon sperm DNA and 11ml ddH₂O were boiled together for 10min, cooled on ice and added to the emptied hybridization bag along with new hybridization solution, labelled pGPC probe, and 3-5ul λ HindIII/EcoRI-digested marker DNA. The filters were then incubated again at 65°C overnight.

2.8.3 Washing

Once the filters were removed from the hybridization bag, they were rinsed in 2X SSC/0.1% SDS, washed in 2X SSC/0.1% SDS for 5min, then in 1X SSC/0.1% SDS at 65°C for 30min. The filters were wrapped in Gladwrap and exposed to Fuji film.

2.9 Labelling of probes with ³²P

2.9.1 Nick translation

Nick translation of the pGPC probe was carried out in accordance with the instructions in the Amersham Nick Translation Kit. Briefly, 250ng of probe was used per 150x200mm filter. The probe DNA and 30ul ddH₂O was boiled for 10min and then placed on ice for 2min. The dNTPs, α³²P-ATP and enzyme was added, the contents mixed and briefly spun down, and then incubated at 15°C for 1hr.

The reaction was stopped by adding a final concentration of 50mM EDTA. A 1/5th volume of salmon sperm DNA (ssDNA), an equal volume of 4M NH₄Ac and a 3x vol ethanol (EtOH) was added and incubated on ice for 1hr to allow precipitation of the radioactively incorporated DNA probe. After the probe was centrifuged at 14,000rpm for 15min, the radioactive counts of the supernatant
versus the pellet was assessed. This was either done manually, or by DE81 radioactive incorporation (Sambrook et al. 1989). If pellet count/total counts was less than 30%, then the probe was discarded and fresh probe prepared (ideally radioactive incorporation is >60%). Marker DNA (λ HindIII/EcoRI and HindIII) was also radioactively labelled through nick translation.

2.9.2 End-labelling

The SSO probes were radioactively labelled by the 5’ addition of γ\(^{32}\)P-ATP. In the labelling reaction, 21pmoles oligonucleotide, 1x one-phor-all buffer, 3.5ul γ\(^{32}\)P-ATP (4,000 Ci/m mole; Bresatec) and T\(_4\) polynucleotide kinase (Pharmacia) were mixed in a final volume of 30ul with ddH\(_2\)O and incubated at 37°C for 30min. Once the reaction was terminated (by adding EDTA to 0.5mM) the probes were spun down and 1/15th of the reaction (3ul) added to each of the prehybridizing filters.
Chapter 3

BIOCHEMICAL CHARACTERISATION OF G6PD DEFICIENCY IN WOSERA
3.1 Introduction

Glucose-6-Phosphate Dehydrogenase (G6PD) catalyses the first step in the pentose phosphate shunt and reduces NADP to NADPH (Figure 3.1). NADPH is required for the further reduction of oxidized glutathione (GSSG) to GSH. As GSH is responsible for many detoxification and reduction reactions, cells without G6PD have difficulty metabolising products of oxygen metabolism (WHO 1989). Erythrocytes, unlike other cell types, do not possess a nuclear genome and are unable to synthesize their own proteins. Subsequently, G6PD deficiency has a more pronounced effect on erythrocytes than in nucleated cells. Ineffective detoxification of products of oxygen metabolism during oxidative stress can result in cell death, which is why individuals with G6PD deficiency often suffer from haemolytic anaemia (Beutler 1990). The G6PD gene is located on the X chromosome (band Xq28) (Martini et al. 1986), thus deficiencies occur most frequently in males.

The world distribution of glucose-6-phosphate dehydrogenase (G6PD) deficiency largely matches the world-wide distribution of endemic malaria (Allison 1960). In Papua New Guinea (PNG), endemic malaria is largely confined to the lower lying and coastal regions of the country (Motulsky 1960; Allison 1961). Similarly, the distribution of G6PD deficiency is also confined to these regions (Kidson 1961; Yenchitomanus et al. 1986a). This shared pattern of distribution led Motulsky to suggest that G6PD deficiency might provide protection from malarial infection (Motulsky 1960).

Most individuals with G6PD deficiency are not affected by illness on a daily basis. However, a life threatening haemolytic crisis can occur upon infection, ingestion of certain foods, or administration of certain drugs that generate oxidant stress (e.g. primaquine). Chloroquine, a commonly used antimalarial drug, does not cause haemolysis in G6PD deficient individuals. However, due to the rising incidence of antimalarial drug resistance in PNG (Sapak et al. 1991), other drugs that cause hemolysis in G6PD deficient individuals are often required and it is becoming increasingly important to establish the G6PD status of individuals to enable appropriate antimalarial treatment.
Figure 3.1 Role of Glucose-6-Phosphate Dehydrogenase in pentose shunt pathway.
Numbers in bold refer to the following enzymes:
1 glucose-6-phosphate dehydrogenase;
2 6-phosphoglyconolactonase (or lactonase);
3 6-phosphogluconate dehydrogenase;
4 xylulose-5-phosphate epimerase;
5 ribose-5-phosphate isomerase;
6 transketolase;
7 transaldolase;
8 transketolase [same as 6];
9 glutathione reductase.
Abbreviations: G-6-P, glucose 6-phosphate; 6-PGL, 6-phosphogluconolactone; 6-PG, 6-phosphogluconate; Ru-5-P, ribulose 5-phosphate; R-5-P, ribose 5-phosphate; X-5-P, xylulose 5-phosphate; Ga-3-P, glyceraldehyde 3-phosphate; S-7-P, sedoheptulose 7-phosphate; F-6-P, fructose 6-phosphate; E-4-P, erythrose 4-phosphate; PRPP, 5-phosphoribosyl-1-pyrophosphate; DHAP, dihydroxy acetone phosphate; GSH, reduced glutathione; GSSG, oxidized glutathione. (from Eaton and Brewer in The Red Blood Cell Ed. D.M. Surgenor, Academic Press, 2nd Edition, 1974, p. 437)
Glucose

ATP

ADP

G-6-P

6-PGL

6-PG

GSH GSSG

NADP NADPH

H₂O H⁺

Ru-5-P

R-5-P

X-5-P

Ga-3-P

S-7-P

E-4-P

AMP+PRPP ← ATP+

Adenine Nucleotide Synthesis

DHAP

Ga-3-P

Lactate

F-6-P

F-1,6-DP
A large number of studies have investigated the frequency of G6PD deficiencies in PNG populations (Kidson 1961; Kidson and Gorman 1962; Parsons and Ryan 1962; Porter et al. 1964; Buchbinder and Clark 1972; Malcolm et al. 1972; Yoshida et al. 1973; Young et al. 1974; Booth et al. 1982; Mourant et al. 1982; Yenchitsomanus et al. 1986b; Schuurkamp et al. 1989; Brabin and Brabin 1990). Some early studies identified G6PD variants based on their electrophoretic mobility in TEB starch gel electrophoresis (Porter et al. 1964; Malcolm et al. 1972). More detailed biochemical characterisation was started by Kirkman et al. (1968) in the Bukawa group of the Markham valley, identifying the Mediterranean-like and Markham G6PD deficient variants. Later studies in the Markham Valley by Yoshida et al. (1973) identified 5 variants (including G6PD Mediterranean-like) and demonstrated inter- and intra-village G6PD heterogeneity. The work of Chockkalingham and colleagues has contributed the bulk of knowledge of the G6PD deficient variants in PNG with the characterisation of 18 variants (Chockkalingam and Board 1980; Chockkalingham 1981; Chockkalingam et al. 1982). These included samples from people of the PNG islands of New Ireland, West New Britain, Manus, Karkar, and Goodenough and the PNG provinces of East Sepik, Madang, Morobe, Oro and Central. Although some of the variants from these studies were common (G6PD Salata, Goodenough, Kalaun and Markham) the majority were rare variants (Chockkalingham 1981).

The presence of G6PD deficiency has been determined in a sample of the Wosera population, from the East Sepik Province, PNG. A family study of the biochemical characteristics of a Wosera G6PD deficient enzyme variant has been carried out to allow comparison of this variant with those previously described in PNG.

3.2 Materials and Methods

3.2.1 Starch gel electrophoresis

Red blood cell samples from 237 Woserans were electrophoresed on TEB starch gels (pH 8.6). G6PD activity was identified by a standard electron transfer dye staining method (Harris and Hopkinson 1978). G6PD deficiency was scored when the stained
G6PD band was barely detectable relative to the intensity of the accompanying haemoglobin band.

3.2.1 Family study

The Wosera G6PD-deficient family pedigree is summarised in Figure 3.2 and consists of a mother, father, G6PD-deficient proband, G6PD-deficient brother and two G6PD normal brothers. As G6PD is an X-linked gene, and as not all sons are G6PD-deficient, it is anticipated that the mother of the proband is heterozygous for the deficiency.

3.2.1.1 Sample collection

Blood was collected from the proband (40ml) and other family members (30ml) and transported to Canberra, Australia, within 48hrs. The blood samples were wrapped in newspaper and packed on dry ice to maintain a chilled but not frozen environment during transportation. On receipt in Canberra, the G6PD enzyme from all PNG blood samples, along with the blood from two Caucasian G6PD normal controls, was partially purified using WHO standard procedures described in detail by Beutler et al. (1968).

3.2.1.2 Biochemical characterisation

The purified G6PD enzymes from the G6PD-deficient proband, G6PD-deficient brother, G6PD heterozygous mother, and one Caucasian G6PD normal control, were biochemically characterised within 48 hours of receipt. The other purified and G6PD normal enzymes (from the father, G6PD normal brothers, and a second Caucasian control) were stabilised in ammonium sulphate solution and characterised four days later. Biochemical characterisation followed the WHO recommendations for biochemical characterisation of G6PD described by Beutler et al. (1968). Enzyme assays were carried out in duplicate (or triplicate whenever possible) and the data is presented as averaged results.
Figure 3.2  Wosera G6PD deficiency Family Pedigree

- hemizygous G6PD deficient male
- hemizygous G6PD normal male
- heterozygous female
3.3 Results

3.3.1 G6PD Activity

The activity of the G6PD enzyme in the G6PD normal father and sons ranged from 91 to 138% of the Caucasian controls. A G6PD enzyme activity of 47% from the mother of the proband is as expected for a G6PD heterozygote, although non-random X-inactivation can contribute to higher or lower values (Beutler et al. 1962). The brother of the proband, who is also G6PD-deficient, showed 6% G6PD activity of the control. No activity was detected in initial readings of the Wosera proband's haemolysed erythrocytes. However, once the proband's enzyme had been purified and concentrated, the activity of the enzyme was sufficiently restored to enable characterisation. Unfortunately the enzyme activity of the G6PD-deficient brother was too low after purification of the G6PD enzyme to allow accurate characterisation.

3.3.2 Biochemical characterisation of the Wosera G6PD-deficient variant

Upon partial purification of the G6PD enzymes, the following parameters were determined: Km for G6P and NADP substrates, affinity for deamino NADP, 2 deoxy G6P and NAD, pH optimum, starch gel electrophoresis in phosphate (pH 7.0), TEB (pH 8.6) and Tris/Cl (pH 8.8) gels, and thermal stability. The biochemical characteristics of the Wosera proband are presented in Figures 3.3-3.8 along with the control range of pooled data from the G6PD normal father and brothers and a Caucasian G6PD normal control (Study controls). A tabulation of these results, and comparison with the G6PD B (the most common form of normal G6PD enzyme) control range (Yoshida and Beutler 1978) is shown in Table 3.1.

When the Wosera G6PD-deficient results were compared with the G6PD B control range, the Wosera G6PD variant showed decreased G6PD activity and increased electrophoretic mobility in one of the buffering systems (Tris/Cl pH 8.8). Electrophoretic mobility was unchanged in the TEB (pH 8.6) and phosphate (pH 7.0) buffering systems (Figure 3.3). The lower Km for NADP indicated that the
Figure 3.3 Starch Gel Electrophoresis in three buffer systems.
Wosera G6PD deficient proband (W), probands G6PD heterozygous mother (M), probands G6PD deficient brother (B) and G6PD normal Caucasian control (C)
THCl (pH 8.8)

TEB (pH 8.6)

Phos. (pH 7.0)

W  M  B  C
Figure 3.4 Substrate Affinity for G6P (Km)

Figure 3.5 Substrate Affinity for NADP (Km)
Figure 3.6 pH Optimum

Figure 3.7 Thermal Stability (46°C)
Figure 3.8  Affinity for Alternative Substrates

- □ Wosera variant
- I Normal range

<table>
<thead>
<tr>
<th>Substrate</th>
<th>% Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 deoxy G6P</td>
<td>200</td>
</tr>
<tr>
<td>Deamino NADP</td>
<td>150</td>
</tr>
<tr>
<td>NAD</td>
<td>100</td>
</tr>
<tr>
<td>G6PD Variant</td>
<td>Activity (% of normal)</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Amboin</td>
<td>4.7</td>
</tr>
<tr>
<td>Bukitu</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Wosera</td>
<td>&lt;2</td>
</tr>
<tr>
<td>Study</td>
<td>91</td>
</tr>
<tr>
<td>controls</td>
<td>138</td>
</tr>
<tr>
<td>Normal</td>
<td>100.0</td>
</tr>
</tbody>
</table>

1 Chockkalingam et al. 1982, 2 Beutler and Yoshida 1988, 3 Chockkalingam and Board 1980, 4 Yoshida and Beutler 1978
Wosera variant binds the NADP substrate with a stronger affinity than the normal G6PD enzyme, however, its binding affinity for the G6P substrate fell within the control range. The Wosera variant utilises the NADP analogue, deamino NADP, more efficiently as a co-enzyme than the control but the utilisation of the two other alternative substrates tested (2 deoxy G6P and NAD) fell within the control range. The thermal stability and pH optimum did not show any variation from that expected for G6PD B.

Table 3.1 includes the biochemical range of the study control G6PD enzymes. The G6PD activity range indicates that these samples are not deficient, but do vary from both the control data published by Yoshida and from one another. The utilisation of deamino NADP, ranging from 28% to 183% of the activity with normal substrates, is the most extreme example of this inter-sample variation. Relative to these control values, the Wosera variant has increased Km for G6P and the normal utilisation of deamino NADP.

Table 3.2 summarises the biochemical characteristics of all G6PD variants previously described in PNG. Interestingly, when the Wosera G6PD variant is compared to these, the G6PD variants from neighbouring populations in the East Sepik Province show the greatest degree of similarity (Chockkalingam and Board 1980; Chockkalingam et al. 1982). The biochemical characteristics of two of these variants (Amboin and Bukitu) are presented with the Wosera variant in Table 3.1. The Amboin variant has a lower Km for G6P and both the Amboin and Bukitu variants have an increased Km for NADP. They also show increased electrophoretic mobility in the TEB gel, decreased thermal stability at 46°C and can utilise NAD as a co-enzyme.

When the biochemical properties of the Wosera G6PD-deficient variant were compared with other G6PD deficient variants described worldwide, those showing closest similarity were G6PD Chainat, Gifu, Hofu, Sapporo and Yokohama (Table 3.3). These variants are all from Asian populations, the latter four from Japan and G6PD Chainat from Thailand. Although similar, each of these variants differs from the Wosera variant in either their affinity for G6P, electrophoretic mobility in Tris/Cl, thermal stability or affinity for deamino NADP.
### Table 3.2 Biochemical variants of G6PD in Papua New Guinea

<table>
<thead>
<tr>
<th>G6PD Variant</th>
<th>Activity (% of normal)</th>
<th>Electrophoretic Mobility (% of normal)</th>
<th>Km G6P (μM)</th>
<th>Km NADP (μM)</th>
<th>Utilization of substrate analogues</th>
<th>Heat Stability (46°C)</th>
<th>pH optimum</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2dG6P (% of G6P)</td>
<td>NADP (% of NADP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TEB pH 8.6</td>
<td>THC pH 8.6</td>
<td>Phos pH 7.0</td>
<td></td>
</tr>
<tr>
<td>Amboin</td>
<td>4.7</td>
<td>102.8</td>
<td>104.2</td>
<td>-</td>
<td>29.9</td>
<td>10.6</td>
<td>9.9</td>
<td>73.7</td>
</tr>
<tr>
<td>Angoram</td>
<td>7.6</td>
<td>96.3</td>
<td>87.3</td>
<td>-</td>
<td>15.3</td>
<td>3.2</td>
<td>9.7</td>
<td>71.6</td>
</tr>
<tr>
<td>Bogia</td>
<td>&lt;10</td>
<td>96.0</td>
<td>95.0</td>
<td>-</td>
<td>55.5</td>
<td>7.0</td>
<td>8.6</td>
<td>63.3</td>
</tr>
<tr>
<td>Bukitu</td>
<td>&lt;10</td>
<td>103.0</td>
<td>100.0</td>
<td>-</td>
<td>74.7</td>
<td>11.0</td>
<td>12.0</td>
<td>64.6</td>
</tr>
<tr>
<td>Castilla-like</td>
<td>8.8</td>
<td>105.3</td>
<td>105.3</td>
<td>-</td>
<td>31.4</td>
<td>4.7</td>
<td>10.6</td>
<td>87.9</td>
</tr>
<tr>
<td>Canton-like</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.4</td>
<td>-</td>
<td>162.2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.3</td>
<td>-</td>
<td>222</td>
<td>-</td>
</tr>
<tr>
<td>Corinth-like</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>20-30</td>
<td>55-60</td>
<td>&lt;3.0</td>
</tr>
<tr>
<td>Goodenough</td>
<td>&lt;2</td>
<td>101.1</td>
<td>100.0</td>
<td>7.8 ±</td>
<td>7.0 ±</td>
<td>68.5 ±</td>
<td>365.0</td>
<td>16.0 ±</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0 ±</td>
<td>8.0 ±</td>
<td>100.4 ±</td>
<td>102.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Kaluan</td>
<td>&lt;10</td>
<td>98.0</td>
<td>97.0</td>
<td>161.4</td>
<td>11.4</td>
<td>4.4</td>
<td>20.0</td>
<td>10.2</td>
</tr>
<tr>
<td>Kar Kar</td>
<td>&lt;2</td>
<td>93.5</td>
<td>87.2</td>
<td>92.0</td>
<td>10.4</td>
<td>10.0</td>
<td>83.0</td>
<td>60.0</td>
</tr>
<tr>
<td>Madang</td>
<td>&lt;2</td>
<td>95.3</td>
<td>95.0</td>
<td>96.0</td>
<td>4.8</td>
<td>2.0</td>
<td>10.4</td>
<td>102.0</td>
</tr>
<tr>
<td>Mainoki</td>
<td>&lt;2</td>
<td>95.6</td>
<td>88.8</td>
<td>93.1</td>
<td>3.5</td>
<td>4.3</td>
<td>85.3</td>
<td>44.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.0</td>
<td>300-</td>
<td>60-</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Manum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>Manus</td>
<td>4.3</td>
<td>96.8</td>
<td>96.0</td>
<td>94.6</td>
<td>16.7</td>
<td>1.3</td>
<td>12.1</td>
<td>79.0</td>
</tr>
<tr>
<td>Markham</td>
<td>&lt;10</td>
<td>103-</td>
<td></td>
<td>106.0</td>
<td>6.7-</td>
<td>1.9-</td>
<td>144-</td>
<td>341-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>104</td>
<td>14.7</td>
<td>2.5</td>
<td>159</td>
</tr>
<tr>
<td>MarkhamK</td>
<td>1.5-</td>
<td></td>
<td>105-</td>
<td></td>
<td>4.4-</td>
<td>162-</td>
<td>280-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td>108</td>
<td></td>
<td>6.3</td>
<td>222</td>
<td>350</td>
<td></td>
</tr>
<tr>
<td>Mediterranean-like</td>
<td>1.4-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23-27</td>
<td>300-</td>
</tr>
<tr>
<td>an-like</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palakau</td>
<td>3.5</td>
<td>100.0</td>
<td>87.6</td>
<td>95.4</td>
<td>51.4</td>
<td>8.0</td>
<td>18.9</td>
<td>46.7</td>
</tr>
<tr>
<td>Polondetta</td>
<td>&lt;2</td>
<td>95.6</td>
<td>88.8</td>
<td>93.1</td>
<td>5.0</td>
<td>1.0</td>
<td>66.7</td>
<td>295.3</td>
</tr>
<tr>
<td>Salata</td>
<td>&lt;5</td>
<td>97.0</td>
<td></td>
<td>83.0</td>
<td>73-78</td>
<td>7-10</td>
<td>6-11</td>
<td>57-59</td>
</tr>
<tr>
<td>Swit</td>
<td>&lt;2</td>
<td>101.0</td>
<td>86.0</td>
<td>98.0</td>
<td>5.5</td>
<td>0.8</td>
<td>106.0</td>
<td>344.0</td>
</tr>
<tr>
<td>Union-like</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>180.0</td>
<td>400.0</td>
<td>&lt;3</td>
</tr>
<tr>
<td>Wewak</td>
<td>&lt;2</td>
<td>97.3</td>
<td>90.0</td>
<td>94.6</td>
<td>35.0</td>
<td>1.0</td>
<td>23.9</td>
<td>83.5</td>
</tr>
<tr>
<td>Wosera</td>
<td>&lt;2</td>
<td>100.0</td>
<td>104.6</td>
<td>100.0</td>
<td>67.0</td>
<td>1.0</td>
<td>&lt;10.0</td>
<td>110.0</td>
</tr>
<tr>
<td>Yangoru</td>
<td>11.4</td>
<td>93.9</td>
<td>95.6</td>
<td>90.0</td>
<td>19.6</td>
<td>5.9</td>
<td>15.2</td>
<td>73.1</td>
</tr>
<tr>
<td>Normal</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>50-70</td>
<td>2.9-</td>
<td>&lt;4.0</td>
<td>55-60</td>
<td></td>
</tr>
</tbody>
</table>

4.4

Table 3.3 Biochemical variants described outside of PNG that have biochemical similarity to the Wosera G6PD-deficient variant

<table>
<thead>
<tr>
<th>G6PD Variant</th>
<th>Activity (% of normal)</th>
<th>Electrophoretic Mobility (% of normal)</th>
<th>Km (uM)</th>
<th>Km G6P (uM)</th>
<th>Km NADP (uM)</th>
<th>Thermal Stability (46°C)</th>
<th>pH optimum 8.6</th>
<th>pH optimum 8.8</th>
<th>pH optimum 7.0</th>
<th>TEB</th>
<th>THCl</th>
<th>Phos.</th>
<th>2dG6P Phos. analogues (% of NADP)</th>
<th>De-aminase NADP (% of G6P)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chainat</td>
<td>0.4</td>
<td>98.0</td>
<td>-</td>
<td>67</td>
<td>5</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>Panich &amp; Na-Nakorn 1980</td>
</tr>
<tr>
<td>Gifu</td>
<td>1.2</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>48</td>
<td>3</td>
<td>4.8</td>
<td>8</td>
<td>1.2</td>
<td>Increased</td>
<td>Normal</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>Fuji et al. 1984</td>
</tr>
<tr>
<td>Hofu</td>
<td>&lt;10</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>25</td>
<td>5</td>
<td>5.2</td>
<td>105</td>
<td>7.5</td>
<td>Normal</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>Fuji et al. 1977</td>
</tr>
<tr>
<td>Sapporo</td>
<td>3.6</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>51</td>
<td>3</td>
<td>6.6</td>
<td>7.5</td>
<td>7.5</td>
<td>Unstable</td>
<td>Normal</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>Miwa et al. 1981</td>
</tr>
<tr>
<td>Yokohama</td>
<td>1.9</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>70</td>
<td>6</td>
<td>6.0</td>
<td>6.3</td>
<td>6.3</td>
<td>Unstable</td>
<td>Normal</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>Miwa et al. 1978</td>
</tr>
<tr>
<td>Wosera</td>
<td>&lt;2</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>67</td>
<td>1</td>
<td>&lt;4.0</td>
<td>110</td>
<td>55-60</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>100.0</td>
<td>100.0</td>
<td>Yoshida &amp; Beutler 1978</td>
</tr>
</tbody>
</table>
3.4 Discussion

The results from this study suggest that there is variability amongst G6PD normal controls (Table 3.1). To account for the biochemical range of G6PD B (normal) enzymes, Yoshida et al (1971) pooled results from 3 studies. In an attempt to minimise laboratory to laboratory variation, a standard methodology for the biochemical characterisation of G6PD was agreed upon (Betke et al. 1967). Despite these measures, it is still recognized that the most reliable comparisons can only be made between enzymes characterised at the same time in the same laboratory (Beutler 1992). As the molecular basis of many biochemical variants are described, there are increasing numbers of biochemically-different G6PD variants with the same molecular basis (Beutler and Yoshida 1988; Hirono et al. 1989; Beutler et al. 1991a).

The distinctive characteristics of the Wosera variant show a decreased Km for NADP and an increased utilisation of deamino NADP when compared with the published normal G6PD enzyme. The comparison of this variant with others described in PNG shows that the Wosera variant is most like the other East Sepik G6PD-deficient variants (G6PD Salata, Amboin, Angoram, Castilla-like and Bukitu).

Of the 400 biochemical variants that have been described worldwide, it seems more than coincidental that four of the five G6PD variants that most closely match the biochemistry of the Wosera variant (G6PD Hofu, Gifu, Sapporo and Yokohama) are from Japanese populations, and the other (G6PD Chainat) is from an Asian population. The similarity between Asian and PNG G6PD-deficient variants has also been noted in a study of G6PD-deficient variants from China (Du et al. 1988). As Wurm (1983) has suggested that the NAN speaking people of the East Sepik originated from Asia between 10-15,000 years ago, it is possible that the G6PD-deficient genes found in the Wosera (and PNG) may predate the arrival of the Woseran ancestors to PNG.

Biochemical characterisation of G6PD deficiencies have demonstrated that this enzyme is very heterogeneous. As mentioned earlier, over 400 apparently unique variants have been
described at the biochemical level throughout the world (Beutler and Yoshida 1988). This heterogeneity is particularly evident in PNG where at least 23 unique variants have previously been described (see Table 3.2).

It is interesting to speculate why such a high level of G6PD deficiency heterogeneity occurs in PNG. As the distribution of G6PD-deficient variants are largely confined to populations affected by endemic malaria, it has been suggested that malaria has acted as a selective force on the preservation of G6PD deficient variants (Chockkalingam et al. 1982). The division of many of the PNG populations into distinct tribal groups (with over 700 languages being spoken across the country) might also contribute to G6PD heterogeneity, although there is evidence of G6PD variation within, as well as between, villages (Kirkman et al. 1968; Yoshida et al. 1973).

Molecular characterisation of G6PD-deficient variants in PNG is the next step in determining whether the G6PD-deficient variation seen at the biochemical level is a true reflection of G6PD genetic heterogeneity. Once the molecular basis of G6PD deficiencies is known, more informative conclusions can be drawn about the distribution and origin of these variants. Together these answers may shed further light on the mechanism maintaining the high frequency of G6PD deficiency in PNG.
Chapter 4

MOLECULAR CHARACTERISATION OF G6PD DEFICIENCY IN PNG
4.1 Introduction

The vast degree of G6PD deficiency heterogeneity detected by biochemical analysis infers that a large number of separate mutational events have occurred in the G6PD gene and, in populations exposed to endemic malaria, these G6PD-deficient variants have been preferentially selected and amplified in frequency. The determination of the cDNA (Persico et al. 1986; Takizawa et al. 1986), and later genomic sequence (Chen et al. 1991) of G6PD, has allowed the precise determination of the molecular basis of many of these G6PD deficiencies. It is now realised that some apparently distinct biochemical variants have the same molecular basis (Chui et al. 1991; Beutler 1992). The level of heterogeneity of this deficiency therefore does not appear to be as high as is suggested by the biochemical data.

As yet, no other studies have investigated the molecular basis of G6PD deficient variants in PNG. By doing so, this study aims to provide a clearer view of the true degree of G6PD deficiency heterogeneity in the country. It is also anticipated that this information will indicate whether these G6PD deficiency variants are unique to PNG, supporting the conclusions of previous biochemical studies (Chockkalingam and Board 1980; Chockkalingam 1981; Chockkalingam et al. 1982), or have been derived from other populations.

4.2 Materials and Methods

4.2.1 Populations studied

Thebuffy coat samples from of 233 individuals from the Wosera district were collected and tested for the presence of G6PD deficiency using starch gel electrophoresis. To increase the probability of detecting G6PD deficiency molecular heterogeneity in PNG, samples from geographically distant populations were chosen for molecular analysis. An additional 279 obtained from the Kupiano (Wanigela and Kalo), and National Capital (Koki) districts of PNG, were tested. As the G6PD gene is located on the X chromosome, the search for variation in G6PD is simplified in males who possess only one copy of the gene. Two G6PD deficient
males, one from the Wosera and one from the Kalo district along with a Wosera G6PD normal (+) control were chosen for molecular characterisation.

4.2.2 Strategy for mutation detection

There are many methods of detecting genetic variation (for reviews see Rossiter and Caskey 1990 and Cotton 1993). These range from the detection of large chromosomal alterations such as deletions, insertions and rearrangements, to methods that identify single base changes. In the case of molecular lesions leading to G6PD deficiency, the majority of variants have been identified as single base pair substitutions (Beutler et al. 1991b), thus an experimental strategy sensitive enough to identify such a small change was required.

The methods developed for the detection of single base changes can be roughly split into two categories; those that detect mutations of known sequence, and those that detect new mutations. The former category includes restriction analysis, PCR-sequence specific oligonucleotide (SSO) hybridization (Saiki et al. 1986), allele-specific oligonucleotide primer amplification (Dozy et al. 1990), artificial-restriction fragment length polymorphisms (A-RFLP) (Ward et al. 1991). The latter includes chemical cleavage of mismatch (Cotton et al. 1988), RNase A cleavage (Myers et al. 1985a), denaturing gradient gel electrophoresis (DGGE) (Myers et al. 1985b), single-strand conformation polymorphism (SSCP) (Orita et al. 1989), heteroduplex analysis (White et al. 1992), dideoxy fingerprinting (ddF) (Sarkar et al. 1992), and sequencing (Sanger et al. 1977).

Each method has potential advantages and disadvantages and selection of a mutation detection protocol is largely based on the features of the gene of interest and the sample size being tested. For example, the β globin gene is relatively small (3 exons) and β thalassaemia mutations can be realistically screened by automated sequencing (Tamary et al. 1994). Other diseases have definitive mutations that can be directly detected by restriction fragment length polymorphism (RFLP) analysis (as is the case for sickle cell
anaemia) or by detection of deletions by differing PCR product sizes (e.g. ovalocytosis, see Jarolim et al. 1991).

Initial studies of mutation detection in G6PD deficient samples were carried out by the laborious task of cloning G6PD specific fragments from cDNA libraries, sequencing them, then looking for sequence variation (Martini et al. 1986; Persico et al. 1986; Takizawa et al. 1986). The introduction of the Polymerase Chain Reaction (PCR) has revolutionised molecular biological techniques. The PCR allows the amplification of highly specific regions of DNA from very small quantities of initial template DNA. The resultant large quantity of highly specific DNA can be easily manipulated and consequently the detection of mutations in variant genes has been greatly simplified.

Following the determination of G6PD intronic sequence, and the design of PCR primers amplifying the coding exons of the G6PD gene (Persico et al. 1986; Viglietto et al. 1990; Chen et al. 1991), the identification of variation in the G6PD gene has been rapid. The most widely used methods of determining known base substitutions all require PCR amplification of the coding exons of the G6PD gene (Poggi et al. 1990) followed by either; restriction endonuclease (RE) digestion (when the substitution present alters a RE site) (Beutler et al. 1989), sequence specific oligonucleotide (SSO) hybridization (Beutler et al. 1991b), or artificial introduction of restriction sites into variant sequences (ARFLP) (Kurdi et al. 1990). For detection of new substitutions the most commonly used strategy has been PCR amplification followed by DNA sequencing (Hirono et al. 1989). Some studies (Corcoran et al. 1992; Baronciani et al. 1993; Calabro et al. 1993; Hsia et al. 1993; Rovira et al. 1994) have introduced an initial screening step to predict which PCR products contain variation from the normal sequence. From 1992 the most popular method of screening has been single-stranded conformation polymorphism (SSCP). However, this method of screening has not always detected all variants present (Calabro et al. 1993).
4.2.3 Screening strategy used in this study

In an attempt to speed the discovery of mutations in the PNG G6PD-genes, amplified exons were screened by heteroduplex analysis (White et al. 1992). This relatively simple technique relies on the "loopout" of mismatched bases formed by mutant/normal heteroduplexes that migrate more slowly than their homoduplex counterparts. It is believed to be more sensitive than SSCP, as it does not rely on mismatches in sequence leading to conformational changes, and safer, as it does not require the use of radioactive detection. However, this screening technique and all other screening protocols will not detect 100% of nucleotide variation (Rossiter and Caskey 1990), and in this study, heteroduplex analysis failed to show an altered migration of the variant sequences compared with controls (see below).

It was decided that the most time effective and reliable strategy to determine the molecular basis of the G6PD deficiencies in PNG was to sequence the coding regions of the G6PD gene. This strategy was chosen, over those mentioned above, for several reasons.

1. Only two G6PD deficient samples were initially available for analysis. Even with a third control sample it was decided that full sequencing of the coding exons would ultimately prove more time effective and thorough.

2. As mentioned previously (4.2.2) the sensitivity of screening protocols are variable, depending on the PCR product size (Rossiter and Caskey 1990; Calabro et al. 1993) and the resultant substitution. If screening was undertaken and no variation determined it could not necessarily be concluded that no variation was present, only that it hadn't been detected. If this were the case then the coding sequence would still require sequencing.

3. As G6PD is an X-linked gene and male G6PD-deficient hemizygotes possess only one X chromosome, the usual difficulty of sequencing both chromosomes in a variant gene would not be encountered.

Because of the impracticality of analysing the entire 17.5kb G6PD gene in each case, most studies have sequenced the coding region of
Figure 4.1  PCR amplification of the coding regions of the G6PD gene.

Schematic representation of the G6PD gene. Coding exons are shown as filled boxes, introns as open boxes, and untranslated exons as hatched boxes. Arrows denote amplification by PCR primer pairs; F = forward primer, B = backward primer. Exon numbers are shown beneath each exon (modified from Poggi et al. 1990).
the gene, as changes in this region are the most likely to cause G6PD deficiencies. As PCR amplification strategies for determination of the G6PD coding sequence were already available, the sequencing strategy used in this study followed that developed by Poggi et al. (1990) using 8 PCR primer pairs (See Figure 4.1). This strategy included the amplification of coding exons 2 to 13 and introns 3, 6, 10, 11 and 12. The PCR amplification conditions were adapted for the thermal cycler (Bartelt Instruments, Victoria, Australia) and these are summarized in Table 2.1 p33.

4.3 Results

4.3.1 Heteroduplex Analysis

Heteroduplex analysis was carried out on all the amplified exons. No variation was detected in the Kalo or the Wosera variants. As an example of the results obtained in these experiments heteroduplex analysis of exon 11 from the Kalo variant is shown in Figure 4.2. It is now evident that this analysis did not detect the C to T mutation that was subsequently detected in that exon by sequence analysis (see below).

4.3.2 Sequencing G6PD genes

4.3.2.1 Wosera G6PD normal gene

The G6PD A variant found in African populations is one example of population specific polymorphism not associated with enzyme deficiency (Takizawa et al. 1987). As it was possible that differences found in Papua New Guinean G6PD genes were due to silent population polymorphisms, the G6PD gene from a Wosera G6PD+ individual was characterised along with the G6PD- Wosera and Kalo individuals. This sequence was found to be identical to the previously described G6PD B gene (Persico et al. 1986).

4.3.2.2 Wosera G6PD-deficient gene

Within the Wosera G6PD-deficient variant gene a G->A transition was found at nt1388 within exon 12 resulting in an Arg463->Cys substitution. This sequence change is shown in Figure 4.3. This
Figure 4.2 Heteroduplex analysis of G6PD exon 10.

G6PD PCR samples electrophoresed on a 6% heteroduplex gel. C = control DNA. K = Kalo DNA. M = marker DNA (pBR322 plasmid digested with MspI). Marker DNA band sizes (in bp) indicated on either side of the gel.
Figure 4.3 Sequence analysis of Wosera nt1388A variant
G A T C G A T C

Trp [TGCA]
His [CAT]
Ileu [ATT]

5' Trp
Arg
Ileu
3'

Wosera (nt1388A)
Control (nt1388G)
Figure 4.4 Sequence analysis of Kalo nt1360T variant
Val [GAG] → [GTG] Val
Arg [GCG] → [GTC] Cys
Ser [AGG] → [GCG] Ser

Kalo (nt1360T)  Control (nt1360C)
nucleotide change did not lead to the loss or creation of a restriction site, so it's authenticity was verified by the resequencing of a new PCR product, as well as by subsequent SSO hybridization.

4.3.2.3 Koki G6PD-deficient gene

The Kalo G6PD-deficient variant gene contained a C->T nucleotide transition at nucleotide position nt1360 within exon 11 which results in an Arg^{454}->Cys substitution (Figure 4.4). This mutation leads to the loss of an HhaI restriction site and could be confirmed by HhaI digestion (Figure 4.5A, lane7). An additional T to C change in intronic sequence was seen at nt 92 of intron 11 in the Kalo G6PD deficiency gene, 14bp upstream of exon 12. (Figure 4.6). This change also created a restriction site and could be detected by NlaIII digestion (Figure 4.5B).

4.3.2.4 Taq Polymerase errors

Any protocol based on PCR amplification runs the risk of the Thermus aquaticus (Taq) Polymerase enzyme incorporating errors into the target sequence (Gelfand 1989). This problem is particularly pertinent when variations of interest, as in the case of G6PD deficiency variants, are the result of single base substitutions. Furthermore, if PCR products are sequenced from a single clone, the presence of contaminating sequence may not be obvious. There are several means by which this problem can be minimised.

1. A thermostable DNA polymerase with proof reading capabilities (e.g. Pfu) (Lundberg 1991) can be used in place of Taq polymerase.
2. All sequences obtained can be compared to the published sequence. Any sequence discrepancies can then be verified by restriction analysis of the original PCR product if the substitution alters an RE site. Alternatively, a new PCR product can be resequenced, substantially reducing the probability of redetecting the same Taq polymerase error.
Figure 4.5 Restriction analysis of intron 11 NlaIII polymorphism and nt1360T variant.

PCR amplified region of the G6PD gene (see Figure 4.2) fractionated on a 6% polyacrylamide gel and stained with ethidium bromide.

A. Lanes 1, 5, and 8 are pBR322 digested with MspI. Sizes (bp) of specific marker bands are shown alongside; Lane 2, NlaIII digestion of Kalo G6PD-deficient DNA containing intron 11 nt93C; Lane 3, NlaIII digestion of DNA amplified from the mother of the Wosera G6PD-deficient proband containing intron 11 nt93T; Lane 4, NlaIII digestion of DNA derived from the Wosera proband containing intron 11 nt93T; Lane 6, HhaI digestion of normal control DNA containing nt1360C and Lane 7, HhaI digestion of Kalo G6PD-deficient containing nt1360T.

B. A schematic diagram showing (I) variation in the NlaIII fragments associated with the intron 11 NlaIII+ and - haplotypes (nt93C and nt93T alleles, respectively) and (II) variation in the HhaI fragments associated with the presence of Kalo G6PD-deficient nt1360T allele.
A

![Image of gel electrophoresis with bands labeled NlaIII and HhaI.]

B

<table>
<thead>
<tr>
<th>Sample</th>
<th>Lanes</th>
<th>Sizes (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalo (nt93C)</td>
<td>1</td>
<td>114, 215, 173, 163</td>
</tr>
<tr>
<td>Wosera (nt93T)</td>
<td>2</td>
<td>114, 388, 163</td>
</tr>
<tr>
<td>Control (nt1360C)</td>
<td>3</td>
<td>185, 45, 435</td>
</tr>
<tr>
<td>Kalo (nt1360T)</td>
<td>4</td>
<td>185, 480</td>
</tr>
</tbody>
</table>

Legend:

NlaIII
HhaI
Figure 4.6 Sequence analysis of G6PD intron 11 polymorphism
GATCGATC

5' ATACATGT

3' ATACTATGT

Kalo (nt93C) Wosera (nt93T)
Figure 4.7 Sequence analysis of Taq Polymerase error
normal (nt42\textsuperscript{T}) Taq error (nt42\textsuperscript{A})
Taq Polymerase errors were seen on a number of occasions (see Figure 4.7) and any sequence variation was validated by restriction digestion or resequencing of new PCR products. Pfu polymerase enzyme was not used as the time required to modify the PCR amplification conditions would have been inhibitory.

4.3.3 Verification and screening by restriction analysis

The nt1360T allele abolishes an HhaI site allowing the detection of this mutation by HhaI digestion of the 655 bp, exon 10-13, PCR product (see Figure 4.5A). A PCR product from a normal G6PD gene is cut twice by HhaI, yielding fragments of 185, 45 and 435bp. The second HhaI site is absent in products from individuals carrying the nt1360T mutation and the 480bp product remains undigested. A second HhaI site found upstream of the nt1360 site acts as a digestion control (Figure 4.5B).

The NlaIII polymorphism in intron 11 was identified in a number of samples by digestion of the same 655bp PCR product; the C-allele yielding 4 fragments of 114, 215, 175 and 163bp and the T-allele yielding only 3 fragments of 114, 388 and 163bp due to the loss of an NlaIII site (Figure 4.5B).

Restriction analysis could be used to screen large numbers of samples for these two variations of the G6PD gene. However, as the Wosera G6PD deficiency variant (nt1388A) did not alter a restriction site, it was decided that additional samples would be screened for the nt1360T, nt1388A and intron 11 variants by SSO hybridization.

4.3.4 Screening using SSO hybridization

Six 19mer oligonucleotides were chemically synthesized, two to match either the nt1360T or nt1388A sequences, two to match their G6PD B sequence counterparts and two to match either the intron 11 nt93 C or T alleles. The sequences of these oligonucleotides are given in Table 2.3, p37.

The probands family members were screened by SSO hybridization for the nt1388A variant to confirm that transmission of G6PD
Figure 4.8 SSO hybridization of PNG G6PD variants and controls.

SSO hybridization probes are indicated on the right hand side. PCR products tested are from Kalo (K), mother of Wosera proband (M), Wosera proband (W), G6PD-deficient brother of Wosera proband (B), and Caucasian female control (C) samples. The PCR amplification products bound to nitrocellulose filters are from PCR amplification using primers 11F & 11B (see Figure 4.1).
deficiency to the proband was genetic and not due to a sporadic mutation, as is the case of G6PD Santiago de Cuba (Vulliamy et al. 1988). These results are shown in Figure 4.8. As expected, the brother of the proband, who also had low G6PD activity, was hemizygous for the nt1388 G->A change. The proband's mother was heterozygous for the nt1388A variant (nt1388G/A). All Wosera G6PD genes associated with the T allele at nt93 of intron 11. A Caucasian female G6PD normal control was heterozygous (nt93C/T) for the intron 11 polymorphism.

4.3.5 Population studies

DNA samples of twenty eight G6PD-deficient males from seven additional PNG populations were screened for the presence of the nt1360T and nt1388A G6PD variants. Two of the samples from Aroana and Kaparoko (Central Province) and one from Madang carry the nt1360T substitution seen in the Kalo variant (Figure 4.9, Table 4.1). These were all associated with the NsallI+ (C) allele of the intron 11 polymorphism and the remaining 25 samples with the NsallII- (T) allele. None of the 28 samples screened contained the nt1388A (Wosera) variant.

4.4 Discussion

The Kalo G6PD variant is the result of a C to T transition at nt1360. This substitution is located closer to the putative NADP binding site at aa386 (Hirono et al. 1989) than the putative G6P binding site at aa205 (Camardella et al. 1988). The Arg to Cys amino acid change results in the loss of a positive charge, and as the G6PD protein carries an overall negative charge, this change is anticipated to migrate more quickly than the normal G6PD protein upon starch gel electrophoresis.

The Wosera variant's G to A transition at nt1388 is also closer to the putative NADP binding site than the putative G6P binding site. The resultant aa463Arg->His change, like that of the Kalo variant, leads to the substitution of a basic amino acid for a neutral one, and therefore is also predicted to migrate faster than the normal G6PD protein upon starch gel electrophoresis.
Table 4.1 Occurrence of polymorphism in G6PD-deficient genes from nine PNG populations detected by SSO hybridization.

<table>
<thead>
<tr>
<th>Population</th>
<th>N</th>
<th>EX11N</th>
<th>EX11K</th>
<th>INT11C</th>
<th>INT11T</th>
<th>EX12N</th>
<th>EX12W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wosera (ES)</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Madang (M)</td>
<td>17</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>16</td>
<td>17</td>
<td>-</td>
</tr>
<tr>
<td>Karkar Is (M)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Karikari (N)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Aroana (C)</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Kaparoko (C)</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Kalo (C)</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Karimui (S)</td>
<td>6</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Jimi (WH)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>32</td>
<td>28</td>
<td>4</td>
<td>4</td>
<td>28</td>
<td>29</td>
<td>3</td>
</tr>
</tbody>
</table>

NB: N = no. of G6PD deficient genes studied. Provinces of PNG are indicated in brackets; East Sepik (ES), Madang (M), Northern (N), Milne Bay (MB), Central (C), Northern (N), Simbu (S), and Western Highlands (WH).
Figure 4.9 Populations screened for G6PD molecular variants, nt\textsubscript{1360}T and nt\textsubscript{1388}A.
The biochemical characterisation of the Wosera variant has been determined and is summarised in Chapter 3. As predicted by the molecular change, an increase in electrophoretic mobility is noted in T/HCl (pH8.8) gels (see Figure 3.3, p53), and is emphasized by the thick band of the heterozygous mother, who produces both the normal and deficient G6PD enzyme. The increased binding affinity for NADP and the increased utilization of deamino NADP suggests that this variant may affect normal NADP substrate binding. However, the structural effect on NADP binding can only be accurately predicted once the crystal structure of G6PD is solved.

In Chapter 3 the Wosera G6PD variant was compared to other biochemical G6PD deficiency variants from PNG to determine whether this variant, and therefore the nt1388A mutation, may have been studied previously. Comparisons of this nature may not be reliable as, in some cases, G6PD variants that appeared to be very different when compared by standard biochemical methods have proven to be the same upon sequence analysis. Biochemical comparison between the Wosera G6PD variant and 5 Chinese biochemical variants; G6PD Anant, Kaiping, Sapporo-like, Dhon, and Petrich-like (Du et al. 1988), who all share the single G->A transition at nt1388 (Chui et al. 1991), reveals considerable biochemical variation (Table 4.2).

Conversely, some G6PD deficient enzymes with the same molecular basis might in fact have genuinely different biochemical characteristics, possibly as a result of differing post-transcriptional or post-translational modifications (De Vita et al. 1989; Cappellini et al. 1994). A recent study of two related males with G6PD deficiency characterised by the same molecular variant showed that their clinical presentations were drastically different (Zuo et al. 1992). Other family studies have also shown quite marked differences in the biochemical properties of G6PD deficient enzymes assumed to be identical (Beutler 1993). Additional factors may contribute to the severity, and subsequent biochemical expression, of G6PD deficiency.

This variable biochemistry may also be due to the extremely labile nature of the G6PD protein leading to varying levels of biochemical degradation during storage and transport of samples before
characterisation (Beutler et al. 1991a). Subsequently, although the Wosera G6PD appears biochemically unique in PNG, other PNG variants that share some biochemical similarity with it, for example, Amboin and Bukitu may also possess the nt1388A substitution (see Table 3.1, p57). Molecular analysis of these previously described G6PD biochemical variants is required to confirm whether they do, in fact, carry the same molecular lesion.

The G6PD nt1388A mutation found in the Wosera G6PD deficiency gene has the same molecular basis as G6PD Kaiping, found commonly in Chinese populations (Chui et al. 1991, 1993; Chang et al. 1992; Tang et al. 1992) (Table 4.2). In one Chinese study the nt1388A mutation accounts for up to 21% of G6PD deficiencies (Chang et al. 1992). As the Wosera nt1388A has also been described in Asian populations the possibility that this variant may have been introduced from Asia must be considered. Three lines of evidence support an Asian origin. First, the nt1388A is a very common G6PD variant in Chinese populations and Papua New Guinea has been influenced by migration from South East Asia, from where the non-Austronesian (NAN) speakers of PNG are believed to have originated (Serjeantson et al. 1983; Bellwood 1989). Second, the G6PD biochemical variants that share most similarity with G6PD Wosera also came from Asian populations (see Table 3.3, p61). Third, a linked polymorphism (haplotype) found in intron 11 (Nlalll-) of the Wosera G6PD deficiency gene is also present in all nt1388A variants described in the Chinese (D.Y.T. Chui, personal communication). However, as the Nlalll- polymorphism is found in 100% of all G6PD genes screened from the Wosera population and 92% of the 72 PNG G6PD genes tested (Table 4.3), the possibility that the Wosera nt1388A mutation may have originally occurred sporadically on a Wosera G6PD gene containing the common Nlalll-polymorphism cannot be discounted.

The distribution of the nt1360T variant is nowhere near as localized as the nt1388A variant is to Asian populations. The G6PD Union nt1360T variant found in the Kalo has previously been noted in individuals of Italian (Calabro et al. 1993), Spanish (Rovira et al. 1994), Chinese (Perng et al. 1992), Laotian, Filipino (Beutler et al. 1992a), Japanese, Afro-American (Hsia et al. 1993) and Vanuatuan descents (M. Ganczakowski & L. Luzzatto, personal communication).
### Table 4.2 G6PD biochemical variants with nt1388 mutation

<table>
<thead>
<tr>
<th>G6PD Variants*</th>
<th>Activity (% of normal)</th>
<th>Electrolyte mobility (% of normal)</th>
<th>Km of NADP (uM)</th>
<th>Km of G6P (uM)</th>
<th>Thermal stability (%)</th>
<th>pH optimum</th>
<th>Other Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anant</td>
<td>100.0</td>
<td>100.0 ±2.1</td>
<td>0.0 ±0.5</td>
<td>4.0 ±0.0</td>
<td>20.5 ±9.3</td>
<td>80.0</td>
<td>Biphasic</td>
</tr>
<tr>
<td>Kalping</td>
<td>100.0</td>
<td>100.0 ±2.1</td>
<td>0.0 ±0.5</td>
<td>4.0 ±0.0</td>
<td>12.5 ±2.9</td>
<td>83.9 ±6.0</td>
<td>Biphasic</td>
</tr>
<tr>
<td>Sapporo-like</td>
<td>100.0</td>
<td>100.0 ±2.1</td>
<td>0.0 ±0.5</td>
<td>4.0 ±0.0</td>
<td>10.0 ±1.5</td>
<td>71.3 ±3.5</td>
<td>Biphasic</td>
</tr>
<tr>
<td>Dhon</td>
<td>100.0</td>
<td>100.0 ±2.1</td>
<td>0.0 ±0.5</td>
<td>4.0 ±0.0</td>
<td>9.5 ±1.8</td>
<td>68.9 ±4.3</td>
<td>Peak at pH 7.5</td>
</tr>
<tr>
<td>Petrich-like</td>
<td>100.0</td>
<td>100.0 ±2.1</td>
<td>0.0 ±0.5</td>
<td>4.0 ±0.0</td>
<td>11.0 ±1.0</td>
<td>55-60</td>
<td>Normal</td>
</tr>
<tr>
<td>Wosera</td>
<td>100.0</td>
<td>100.0 ±2.1</td>
<td>0.0 ±0.5</td>
<td>4.0 ±0.0</td>
<td>2.9-4.4</td>
<td>40.4 ±13.0</td>
<td>Normal</td>
</tr>
</tbody>
</table>

* from Du et al. 1983, Yoshida and Beutler 1978 and the present study
Table 4.3 Occurrence of polymorphism in G6PD genes of males from 12 populations detected by SSO hybridization

<table>
<thead>
<tr>
<th>Population</th>
<th>N</th>
<th>EX11N</th>
<th>EX11K</th>
<th>INT11C</th>
<th>INT11T</th>
<th>EX12N</th>
<th>EX12W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wosera (ES)</td>
<td>6</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Kalo (C)</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Karkar Is (M)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Karimui (S)</td>
<td>6</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Aroana (C)</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Kaporoko (C)</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Karikari (N)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Goodenough Is (MB)</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Jimi (WH)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Madang (M)</td>
<td>50</td>
<td>49</td>
<td>1</td>
<td>3</td>
<td>47</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>Asian</td>
<td>6</td>
<td>6</td>
<td>-</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Caucasian</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-</td>
</tr>
</tbody>
</table>

Total 82 78 4 11 71 79 3

NB: N = no. G6PD genes studied. Provinces of PNG are denoted in brackets (see Table 4.1).
This far reaching distribution is most likely indicative of many separate mutational events that have occurred at the same position of the G6PD gene, suggesting that nt1360 may be "hot spot" for mutation. These hot spots appear to occur in CpG islands when the C is methylated and is repaired as a T (Schorderet and Gartler 1992) and a summary of G6PD variants shows a predominance of C to T translocations in the G6PD gene (Vulliamy et al. 1988).

However, the nt1360T variant may not have originated in PNG. A concurrent study of G6PD deficiency in Melanesians indicates that the nt1360T variant in Vanuatu is associated with the same NlaIII+ haplotype as the Kalo nt1360T variant (M. Ganczakowski & L. Luzzatto, personal communication). This haplotype has a lower frequency in most populations (Beutler et al. 1992b), including PNG (0.08 in Madang and only seen associated with the nt1360T mutation in 3 other populations), and it is possible that the Kalo and Vanutuan nt1360T variant have the same origin. In non-Oceanic populations the nt1360T mutation is not always seen associated with the NlaIII+ haplotype (E. Beutler, personal communication), indicating that this mutation must have occurred as at least two separate events.

Screening of G6PD deficient individuals from seven other PNG populations was carried out to provide an idea of the distribution of the nt1360T and nt1388A G6PD deficiency variants in PNG. The Wosera nt1388A variant was not found in any of the additional PNG populations screened in this study, although none of those populations were from the same East Sepik district, where the biochemical variants sharing the closest similarity were seen. As the sample size of this preliminary screening study is small, the actual distribution of this variant can not be resolved. However, from the available molecular evidence, this variant does not appear to have a widespread distribution across PNG.

The Kalo nt1360T variant is present in 3 other population samples; Aroana, Kaparoko and Madang (see Table 4.1 & Figure 4.9). It appears particularly common in the south coast of PNG where 3/3 G6PD deficiency genes described carry this mutation. The presence of the nt1360T variant in Madang, on the northern coast of PNG, also suggests that its distribution may be more widespread. The
remaining 16 G6PD deficiency genes from the Madang district do not hybridized to either the Kalo or Wosera variant SSOs, suggesting that there is at least one more molecular variant that is responsible for G6PD deficiency in this region. This finding is not surprising, as previous biochemical studies have suggested extensive G6PD heterogeneity in this region (Chockkalingam and Board 1980; Chockkalingam et al. 1982).

The widespread distribution of the nt1360T variant may be the result of population migration, intermarriage or genetic drift. Giles et al. 1967 have previously proposed that genes flow between populations of related language groups (Giles 1967) and in PNG genes are anticipated to flow between either Austronesian (AN) speakers or Non-Austronesian (NAN) speakers. However a study by Serjeantson et al. (1983) of 17 populations from the northern coast of PNG suggests that populations cluster together by geographical distance rather than by linguistic types. This is supported by Schwidetzky (1971) who also found that anthropological similarity was more closely related to geographical closeness than language similarity. The nt1360T variant is distributed between speakers of both AN and NAN languages. As the populations involved (excluding Madang) are geographically close (see Figure 4.9), the evidence presented here supports Serjeantsons conclusion that linguistic similarity cannot wholly determine genetic closeness of PNG populations.

The difference in the molecular basis of the Wosera and Kalo G6PD deficient variants, and the high proportion of G6PD-deficient samples screened (25/28 or 89%) who do not contain either of these variants, establishes the molecular heterogeneity of G6PD deficiency in PNG.

It has been suggested that G6PD deficiency genes have evolved and spread due to the protection that they provide from malarial infection. This was first proposed by Allison who noted that the worldwide distribution of G6PD deficiency correlated well with present or past malarial endemicity (Allison 1960). How does a potentially detrimental genotype provide protection from malarial infection? It was first thought that a G6PD deficient erythrocyte was an unsuitable environment for P. falciparum
trophozoite development. However, this was challenged by a study demonstrating that *P. falciparum* density was equivalent in G6PD A- and G6PD B+ erythrocytes (Porter *et al.* 1964). Further studies by Bienzle *et al.* (1972) showed that it was in fact G6PD A-/B+ female heterozygotes who had lower parasite counts of, and theoretically greater resistance to, *P. falciparum*. The increased protection of female heterozygotes was confirmed in a later study comparing the morbidity rates in Nigerian children with and without G6PD deficiency. Interestingly, this study concluded that G6PD A-/B+ (but not B-/B+) heterozygotes have a protective advantage against malarial infection (Bienzle *et al.* 1979) whereas an *in vitro* study in Sardinia found that both hemizygous males and heterozygous females showed reduced numbers of parasitized cells, compared with G6PD B+ controls (Roth *et al.* 1983).

As G6PD is coded on the X chromosome, X-inactivation produces an approximately equal ratio of G6PD normal (+) to G6PD deficient (-) erythrocytes (Beutler *et al.* 1962). Usanga demonstrated that invading merozoites could adapt to a G6PD deficient erythrocyte environment *in vitro* by producing their own G6PD enzyme. However, parasite growth was reduced each time that a merozoite originating from a G6PD+ erythrocyte invaded a G6PD- one. This adaptive process is suggested to result in the decreased parasite densities seen in G6PD+/- heterozygotes (Usanga and Luzzatto 1985).

The distribution of G6PD deficiency in PNG correlates well with malarial endemicity (Yenchitsomanus *et al.* 1986b; Schuurkamp *et al.* 1989). However, this correlation is not absolute, as although the Wosera population lives in a hyperendemic region of PNG, the frequency of G6PD deficiency genes in this population is surprisingly low. In the Woseran population it is likely that other genes play a more important role in protection from malarial infection.
Chapter 5

CLASS II HLA VARIATION IN WOSERA
5.1 Introduction

It is now well recognised that human leukocyte antigens (HLA) play an important role in susceptibility to disease (Tiwari and Terasaki 1985) and organ transplant survival (van Rood et al. 1985). HLA antigens help bind foreign peptides and present them to the immune system. If particular HLA variants are more effective at eliciting protective immune response to infectious diseases such as malaria, they would presumably be found in higher frequencies in malaria endemic areas. Similarly, if allele heterozygosity increased the probability of HLA antigens adequately presenting pathogens to the immune system, the expected number of individuals heterozygous at their HLA loci should increase. To investigate whether malaria has a selective effect on HLA allelic diversity and/or heterogeneity, the HLA class II alleles and their frequencies have been investigated in Wosera. Other selective mechanisms that might affect HLA diversity are also discussed.

5.1.1 MHC structure and organisation

The human major histocompatibility complex (MHC) is located on the short arm of chromosome 6 (6p21.3). The genes located in the complex are organised into 3 regions. The class I region contains the genes that code for HLA-A, HLA-B and HLA-C antigens. The class II region contains genes coding for HLA-DP, -DR and -DQ antigens. The class III genes, located between the class I and II genes, encode the complement components C2, C4A and C4B, and factor B (Bf), as well as genes for 21 hydroxylase (21OH) and tumour necrosis factors α and β (TNFα and β). A schematic representation of this gene family is shown in Figure 5.1.

The HLA class I antigens are 45kDa membrane spanning polypeptides. The N-terminal extracellular domain of the antigens consists of 3 globular domains (α1, α2, α3). A 12kDa β2-microglobulin molecule non-covalently associates with the α3 domain (Cresswell et al. 1973; Grey et al. 1973). Alloantigenic sites within these molecules are contained within the α1 and α2 domains.
The DRB5 gene is only present associated with the DR2 allele at the DRB1 locus. Likewise, DRB2 and DRB3 gene loci are only present associated with the DRB1 3,11,12,13 and 14 alleles. NB: not all genes shown. Modified from Campbell and Trowsdale (1993)
In contrast, the class II antigens consist of a non-covalently bound α and β polypeptide chain of molecular weights 34kD and 28kD, respectively. Each chain has two domains (α1, α2 and β1, β2) with the majority of antigenic polymorphism located in the β chain (Roitt et al. 1991).

Both class I and class II antigens are involved in the presentation of foreign peptides to the immune system. Class I antigens, expressed on the surface of all nucleated cells, associate with intracellular peptides and present these on the cell surface for recognition by CD8+ cytotoxic T lymphocytes (CTLs). Protein products of foreign DNA (incorporated into the host genome) will be processed and presented on the surface of the cell by the class I antigen, but unlike endogenous peptides, will stimulate the CTLs to target cells presenting the same antigen. Class II antigens, on the other hand, are expressed on the surface of B cells, antigen presenting cells (APC) and macrophages. These cells ingest external proteins, process them internally and present processed peptides on their cell surface in association with the class II antigens. Foreign peptide recognition by the CD4+ T helper (T\textsubscript{H}) cells leads to further humoral or cellular immune responses (Roitt et al. 1991). The class III region includes genes which encode components of the complement system, involved in the degradation and clearance of immune complexes (Male et al. 1987).

5.1.2 HLA polymorphism

Both class I and class II HLA genes are highly polymorphic. The World Health Organization (WHO) nomenclature committee has officially designated 41 HLA-A, 71 HLA-B, 18 HLA-C, 2 HLA-DRA, 60 HLA-DRB1, 4 HLA-DRB3, 4 HLA-DRB5, 14 HLA-DQA1, 19 HLA-DQB1, 8 HLA-DPA1 and 38 HLA-DPB1 alleles (Bodmer et al. 1992). In addition to this allelic polymorphism, variation is also generated by allelic combinations of linked genes (known as haplotypes). If the alleles associate randomly, the number of possible allelic combinations is 41×71×18×2×60...(etc.) = 8×10\textsuperscript{12}. As each individual co-dominantly expresses two HLA haplotypes, the possible variation increases to 6.6×10\textsuperscript{25}. However, this level of heterogeneity is not seen. As the HLA genes are in close proximity to one another, there is a tendency for alleles to be
inherited as a cluster in linkage disequilibrium with one another (Serjeantson 1985), which reduces the possible number of haplotypes in a population considerably.

5.1.3 Maintenance of HLA polymorphism

There are several hypotheses which propose how HLA polymorphism arose and is maintained. These include selective neutrality, negative assortive mating, maternal-foetal incompatibility, high mutation rate, frequency-dependent selection and overdominant selection (selection favouring heterozygotes) (for reviews see Bodmer 1972, Serjeantson 1989d, Hill et al. 1991b).

Selective neutrality assumes that the fitness of all alleles is equal, that allelic variation is a function of mutation rates and that the maintenance of new alleles is dependent upon population size. Potts et al. 1991 conducted a study of half wild mice of known MHC-type in an enclosed, but naturalistic, environment and showed that the number of MHC-heterozygous offspring was greater than expected. This heterozygosity was demonstrated to be the result of preferential selection by females of MHC-incompatible mates (negative assortive mating). The association of distinct urinary odours with different MHC alleles was thought to be the basis for mating preference (Potts et al. 1991). HLA-restricted mating may also occur in human populations, as most populations forbid consanguineous, and therefore possible HLA-compatible, matings (Glass and Meggitt 1969). Supporting the hypothesis of maternal-foetal incompatibility, mouse experimental models have shown that HLA incompatibility between the foetus and mother results in increased placental size, favouring foetal survival and thus promoting heterozygote selection (Clarke and Kirby 1966).

5.1.3.1 Selection by malaria

Assuming that HLA polymorphism is pathogen-driven there are two different mechanisms by which this polymorphism could have occurred and been maintained; frequency-dependent selection or overdominant selection (selection favouring heterozygotes). Frequency-dependent selection maintains that individuals with
newly arisen MHC alleles have a selective advantage, as the invading pathogen has not had time to adapt to infecting cells expressing the new MHC antigen (Bodmer and Bodmer 1978). Thus, there is a constant turnover of new alleles as the old alleles lose their resistance to pathogens.

Hughes and Nei (1988) do not support this model as they point out that some MHC alleles persist in populations. Instead, they favour the model of overdominant selection whereby individuals heterozygous at their MHC loci have a selective advantage, as the ability of HLA heterozygotes to recognise foreign epitopes is almost double that of homozygotes (Doherty and Zinkernagel 1975). To establish whether MHC polymorphism is influenced by overdominant selection, the rate of nonsynonymous (amino acid altering) substitutions was compared with the rate of synonymous substitutions in the antigen recognition sites (ARS) of MHC class I antigens in both humans and mice (Hughes and Nei 1988). It was reasoned that, in the presence of overdominant selection, the rate of codon substitutions would be higher in the ARS compared to non ARS regions of the MHC DNA, as individuals carrying new MHC alleles would have a selective advantage. The results indicated that this was in fact the case, whereas in non ARS coding regions of the MHC the reverse was true, thus supporting the overdominant selection hypothesis. If selection was neutral, or a high mutation rate or interlocus genetic exchange was responsible for MHC diversity, then the rates of synonymous and nonsynonymous substitutions in the ARS should be more or less the same.

Overdominant selection differs from frequency-dependent selection in that it assumes that all alleles have equal fitness. This pathogen-driven selection is expected to raise the frequency of a number of alleles to similar levels, whereas frequency dependent selection favours the accumulation of many low frequency alleles.

5.1.4 HLA in Papua New Guinea

The distribution and frequency of HLA alleles is varied between different populations in Papua New Guinea. Interpopulation comparisons of HLA class I alleles show a clear separation
between the PNG highland and coastal populations. This is illustrated by the lack of the HLA class I allele, A2, in highlanders which is present in appreciable frequency in coastal populations (0.037) and other Pacific populations (Serjeantson et al. 1982; Crane et al. 1985). Another study investigating the correlation between HLA-based genetic distances and geographical, altitude and language differences in PNG suggested that there was a marked altitude cline in the frequencies of some HLA-A antigens (Smith et al. 1994). They attributed this altitudinal cline to natural selection by malaria, present in the lowlands but absent in the highlands.

The common class II alleles in the highland populations are DR2 (DR15), DR4, and DRw6. The coastal populations share these alleles but also have a significant frequency of DR5 (DR11 and DR12) alleles (Crane et al. 1985). Further separation of DR allele subtypes between the highland and coastal populations was achieved using mixed lymphocyte typing (MLC) and is dependent upon their Dw associations (e.g. DR2 Dw12, or DRB1*1502 in coastal Melanesians and DR2 Dw2, or DRB1*1501, in highland populations and Australian Aborigines) (Honeyman et al. 1986).

The most striking feature of past HLA typing studies in PNG was the apparently low HLA allelic diversity in PNG highland populations (Crane et al. 1985, Honeyman et al. 1986). For example, A11, A24 and Aw34 were almost exclusive representatives of HLA-A alleles (Serjeantson et al. 1992). Studies of class II antigens in PNG using serological techniques confirmed this genetic homogeneity (Bhatia et al. 1984). Furthermore, when the linked alleles (haplotypes) were compared between highlanders and those found in other Oceanic populations, only haplotypes between the very closely linked HLA B-C genes were still shared. This reflects an apparent isolation of the highland populations from external gene flow. In contrast, HLA alleles in Melanesian populations from coastal regions of PNG have a greater degree of HLA allelic diversity and share alleles and haplotypes with other populations (Crane et al. 1985).

The HLA homogeneity seen in the PNG highlanders has been suggested to be the result of a population bottleneck followed by a
population explosion. The increase in population size would have limited population migration into this region, thus limiting mixing of genes from peripheral populations. (Serjeantson and Hill, 1989).

More recent studies using Restriction Fragment Length Polymorphism (RFLP) analysis of HLA class II genes have revealed greater heterogeneity and unique haplotype associations in PNG highlanders at DR2, DR5, DRw6 and DRw8 (Serjeantson et al. 1987, 1992; Kohonen-Corish et al. 1988) and this heterogeneity was confirmed by DNA-based haplotype analysis (Bhatia et al. 1991; Gao et al. 1992).

5.1.5 Typing techniques

To assess HLA allelic frequency, diversity and heterozygosity, alleles must be accurately assigned. Many HLA studies have been based on serology utilising panels of defined antisera and have unfortunately been somewhat limited in PNG due to the restricted set of locally produced antibodies (Honeyman et al. 1986) and by difficulties in obtaining viable lymphocytes in the field (Gao et al. 1992; Serjeantson et al. 1992). MLC typing exploits the phenomenon that T lymphocytes proliferate in the presence of a HLA class II mismatch and has similar limitations as it is labour intensive, has a high percentage of blanks and requires a comprehensive and population specific set of stimulating cells (Gao 1992).

DNA based HLA typing has the advantage that there is no requirement for viable lymphocytes or well defined antisera as analysis is carried out from genomic DNA samples (Serjeantson 1989a). DNA can be obtained from any source of nucleated cells and is most conveniently collected from buffycoats. Unusual banding hybridization patterns can be used to identify previously undefined alleles (Gao 1992). However, the DNA typing is wholly based on determination of sequence changes between differing alleles and does have the disadvantage of not indicating the functional significance of new alleles.

Restriction fragment length polymorphism (RFLP) analysis has been a popular method of DNA based HLA typing since the early 1980s.
This method is based on the association of different HLA alleles with different restriction fragment patterns when digested gDNA samples are hybridized to HLA locus-specific cDNA probes. DRB, DQB and DQA TaqI RFLPs were successfully used for typing HLA-DR and -DQ antigens in Caucasians, Pacific and Oriental specimens (see Kohonen-Corish et al. 1988). However, RFLPs rely on each HLA allele associating with unique restriction patterns which, when small nucleotide changes are involved, are not always detectable. Some RFLP DR typing also assumes linkage disequilibria between DQA1 and DQB1 alleles (Cohen et al. 1986).

The arrival of PCR-based HLA typing techniques (Saiki et al. 1986) has allowed closer examination of the variation between HLA allele DNA sequences and has led to the elucidation of many hitherto unknown population specific alleles (Gao et al. 1992). PCR-SSO (PCR-sequence specific oligonucleotide) typing permits sequence comparisons between different alleles and the identification of new alleles through novel probe hybridization patterns (Gao and Serjeantson 1991b). It is possible that the HLA homogeneity seen in the PNG highlanders in the past was, at least in part, due to the insensitivity of serological and MLC typing protocols to detect novel alleles. Further DNA-based typing, particularly at the class I loci, is required to confirm this.

5.1.5.1 Class II DNA typing in the East Sepik Province

To date a small number of studies have defined class II allele frequencies in PNG using PCR-SSO typing techniques (Bhatia et al. 1991; Gao and Serjeantson 1991a,b; Gao et al. 1992). Of these studies, and those using serological, MLC or RFLP based typing methods, none have investigated HLA class II allelic diversity and frequencies in East Sepik populations of PNG.

5.2 Materials and Methods

5.2.1 HLA SSO typing

Fifty-eight individuals from the Wosera were PCR-SSO typed for DRB1, DQA1 and DQB1. Two PCR primers, GLPS' and GAMP3', were used to amplify codons 17-94 of the DRB1 gene (Todd et al. 1987).
Sequence specific oligonucleotides (SSO) were hybridized to the resultant DRB1 PCR products to identify the presence of the DR 1, 2, 4, 5, 7-12 and DR 14 alleles as described by Gao and Serjeantson (1991b). Those individuals positive for DR2 were further subtyped for DRB1*1501, *1502, *1601, *1602 and *15.3 alleles using the same DRB1 specific 3' primer with a DR2B1 specific 5' PCR primer (SRP-5) and were also subtyped at the DRB5 locus for DRB5*0101, *0102, and *02 alleles using a DR2B5 specific 5' PCR primer (SR5P). To do this, genomic DNA was once again PCR amplified, blotted and typed by SSO hybridization. The DQA1 and DQB1 alleles were typed using a similar strategy to DRB1 without the requirement for subtyping (see Table 2.2 p34 for primer details). Not all 58 individuals were typed at all loci. Sequence specific oligonucleotides (SSOs) hybridized to locus-specific PCR products are described by Gao and Serjeantson (1991b). The resultant allele-specific hybridization patterns are summarised in Table 2.4 and 2.5 (p38 and 39 respectively).

5.2.2 Statistical analysis

The gene frequencies were determined by simple gene counting. Deviation of observed numbers of genotypes from those expected under Hardy-Weinberg equilibrium was determined using a chi-squared goodness-of-fit. Two-locus coefficient of linkage disequilibrium (Δ) and its deviation from zero contingency were determined using a computer programme written by Gao (1992). In this a 2x2 chi-square test was used to evaluate the significance of whether observed numbers of genotypes fell below 0.05. Differences in gene frequency were considered marginally significant when p values fell between 0.05 and 0.1.

5.3 Results

5.3.1 HLA class II allele frequencies in Wosera

Only three (DQA1*0101, *0102 and *0501) of the 8 DQA1 alleles screened were detected in Wosera. DQA1*0102 and *0501 were the most frequent alleles (42% and 57% respectively) with only two examples of DQA1*0102 (Table 5.1). Five individuals were not typed at DQA1. Four of the 14 detectable HLA-DQB1 alleles were
Table 5.1  HLA class II allele frequencies in Wosera

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>No.</th>
<th>Gene frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DQA1</td>
<td>0101</td>
<td>2</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>0102</td>
<td>44</td>
<td>0.415</td>
</tr>
<tr>
<td></td>
<td>0501</td>
<td>60</td>
<td>0.566</td>
</tr>
<tr>
<td>(2n=106*)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DQB1</td>
<td>0301</td>
<td>49</td>
<td>0.480</td>
</tr>
<tr>
<td>(2n=102)</td>
<td>0502</td>
<td>33</td>
<td>0.324</td>
</tr>
<tr>
<td></td>
<td>0503</td>
<td>1</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>0601</td>
<td>19</td>
<td>0.186</td>
</tr>
<tr>
<td>HLA-DRB1</td>
<td>11</td>
<td>49</td>
<td>0.471</td>
</tr>
<tr>
<td>(2n=104)</td>
<td>12</td>
<td>10</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>3</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>1502</td>
<td>21</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>1602</td>
<td>21</td>
<td>0.202</td>
</tr>
</tbody>
</table>

NB: * Number of alleles typed varied between loci.
found in Wosera. Of these, DQB1*0301 had the highest frequency (48%) followed by *0502, *0601 and *0503 alleles. Seven individuals were not typed for DQB1.

The DRB1 generic alleles 2 (DR15 and 16) and 11 were found at high frequencies (40.4% and 47.1%, respectively) in the Wosera population. Other alleles detected were DR12, 14 and 4 in order of decreasing frequency. Six Woserans were not typed at the DRB1 locus. Subtyping was carried out on the DR2 positive samples. DRB1 alleles *1502 and *1601 appear common in this population and were found at equal frequencies. At the DRB5 locus, the alleles *0101 and *02 were common. Three of the 39 DR2 samples were not typed for DRB5.

Table 5.1 gives an overview of the DRB1, DQA1 and DQB1 allele frequencies in the Wosera population. There is very little apparent allelic diversity in this population with the majority of DR alleles being DR2 subtypes DRB1*1502 and DRB1*1602, DR11 and DR12, with very low frequency of DR14 alleles. Diversity is also limited at the DQA1 and DQB1 loci, with only 3 and 4 alleles respectively. To determine whether the alleles associated as expected under Hardy Weinberg equilibrium, a chi-squared analysis was calculated (Table 5.2). At DQA1 $p<0.1$, at DQB1 $p<0.001$ and at DRB1 $p<0.02$. The DQA1 allelic associations do not reach significance and therefore fit Hardy Weinberg expectation. The allelic associations at the DQB1 and DRB1 loci do not fit Hardy Weinburg expectations as the result of a significantly higher than expected number of DQB1*0502, DQB1*0301, DRB1*11 and DRB1*12 homozygotes.

5.3.2 Comparisons with other PNG populations

The HLA class II allelic diversity in Wosera was compared with the allele distribution of two other PNG populations, one from the Eastern Highlands Province (Goroka), where malaria transmission is low, and the other from the hyperendemic coastal population of Madang, where malarial transmission is high (Gao et al. 1992).

The similarity between the Madang and Wosera populations is evident at the DQA1 and DQB1 loci, with notable high frequencies of DQA1*0501 and DQB1*0502 which were not as common in the
Table 5.2  Hardy-Weinberg equilibrium at DQA1, DQB1 and DRB1 loci.

<table>
<thead>
<tr>
<th></th>
<th>obs.#</th>
<th>exp.#</th>
<th>((o-e)^2/e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQA1-DQA1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0101/0101</td>
<td>0</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>0101/0102</td>
<td>0</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>0101/0501</td>
<td>2</td>
<td>1.13</td>
<td>0.67</td>
</tr>
<tr>
<td>0102/0102</td>
<td>6</td>
<td>9.13</td>
<td>1.07</td>
</tr>
<tr>
<td>0102/0501</td>
<td>32</td>
<td>24.90</td>
<td>2.02</td>
</tr>
<tr>
<td>0501/0501</td>
<td>13</td>
<td>16.98</td>
<td>0.93</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>52.99</td>
<td>5.54(^a)</td>
</tr>
<tr>
<td>DQB1-DQB1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0301/0301</td>
<td>15</td>
<td>11.77</td>
<td>0.89</td>
</tr>
<tr>
<td>0301/0502</td>
<td>7</td>
<td>15.85</td>
<td>4.94</td>
</tr>
<tr>
<td>0301/0503</td>
<td>0</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>0301/0601</td>
<td>12</td>
<td>9.13</td>
<td>0.90</td>
</tr>
<tr>
<td>0502/0502</td>
<td>11</td>
<td>5.43</td>
<td>0.75</td>
</tr>
<tr>
<td>0502/0503</td>
<td>0</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>0502/0601</td>
<td>4</td>
<td>6.14</td>
<td>0.75</td>
</tr>
<tr>
<td>0503/0503</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>0503/0601</td>
<td>1</td>
<td>0.19</td>
<td>3.48</td>
</tr>
<tr>
<td>0601/0601</td>
<td>1</td>
<td>1.78</td>
<td>0.33</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>50.98</td>
<td>18.10(^b)</td>
</tr>
<tr>
<td>DRB1-DRB1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1502/1502</td>
<td>1</td>
<td>2.12</td>
<td>0.59</td>
</tr>
<tr>
<td>1502/1602</td>
<td>5</td>
<td>4.24</td>
<td>0.14</td>
</tr>
<tr>
<td>1502/11</td>
<td>11</td>
<td>9.89</td>
<td>0.12</td>
</tr>
<tr>
<td>1502/12</td>
<td>2</td>
<td>2.02</td>
<td>0.00</td>
</tr>
<tr>
<td>1502/14</td>
<td>1</td>
<td>0.61</td>
<td>0.26</td>
</tr>
<tr>
<td>1602/1602</td>
<td>3</td>
<td>2.12</td>
<td>0.37</td>
</tr>
<tr>
<td>1602/11</td>
<td>5</td>
<td>9.89</td>
<td>2.42</td>
</tr>
<tr>
<td>1602/12</td>
<td>3</td>
<td>2.02</td>
<td>0.48</td>
</tr>
<tr>
<td>1602/14</td>
<td>2</td>
<td>0.61</td>
<td>3.20</td>
</tr>
<tr>
<td>11/11</td>
<td>16</td>
<td>11.50</td>
<td>1.72</td>
</tr>
<tr>
<td>11/12</td>
<td>1</td>
<td>4.71</td>
<td>2.92</td>
</tr>
<tr>
<td>11/14</td>
<td>0</td>
<td>1.41</td>
<td>1.41</td>
</tr>
<tr>
<td>12/12</td>
<td>2</td>
<td>0.48</td>
<td>4.80</td>
</tr>
<tr>
<td>12/14</td>
<td>0</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>14/14</td>
<td>0</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
<td>51.99</td>
<td>18.77(^c)</td>
</tr>
</tbody>
</table>

\(^{a}:\text{for df} = 3, \ p < 0.1; \ \text{b}:\text{for df} = 6, \ p < 0.001; \ \text{c}:\text{for df} = 10, \ p < 0.02.\)

Degrees of freedom were calculated as the number of possible allele combinations less the number of alleles eg. at DQA1 6-3 = 3df, DQB1 10-4 = 6df, DRB1 15-5 = 10df.
Table 5.3 Comparisons of HLA-DQ, -DQB and -DRB gene frequencies in 3 PNG populations

<table>
<thead>
<tr>
<th>Allele</th>
<th>Wosera 2n</th>
<th>Wosera %</th>
<th>Madang a 2n</th>
<th>Madang a %</th>
<th>PNG Highlands a 2n</th>
<th>PNG Highlands a %</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQA1*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0101</td>
<td>2</td>
<td>1.9</td>
<td>4</td>
<td>3.1</td>
<td>26</td>
<td>22.8</td>
</tr>
<tr>
<td>0102</td>
<td>44</td>
<td>41.5</td>
<td>50</td>
<td>38.5</td>
<td>53</td>
<td>46.5</td>
</tr>
<tr>
<td>0103</td>
<td>0</td>
<td>0.0</td>
<td>2</td>
<td>1.5</td>
<td>11</td>
<td>9.7</td>
</tr>
<tr>
<td>0301</td>
<td>0</td>
<td>0.0</td>
<td>6</td>
<td>4.6</td>
<td>17</td>
<td>14.9</td>
</tr>
<tr>
<td>0501</td>
<td>60</td>
<td>56.6</td>
<td>68</td>
<td>52.3</td>
<td>7</td>
<td>6.1</td>
</tr>
<tr>
<td>Total</td>
<td>106</td>
<td>100.0</td>
<td>130</td>
<td>100.0</td>
<td>114</td>
<td>100.0</td>
</tr>
</tbody>
</table>

| DQB1*   |           |          |             |            |                   |                  |
| 0501    | 0         | 0.0      | 1           | 0.8        | 0                 | 0.0              |
| 0502    | 33        | 32.4     | 22          | 16.9       | 5                 | 4.4              |
| 0503    | 1         | 1.0      | 1           | 0.8        | 29                | 25.4             |
| 0601    | 19        | 18.6     | 30          | 23.1       | 22                | 19.3             |
| 0602    | 0         | 0.0      | 2           | 1.5        | 34                | 29.8             |
| 0301    | 49        | 48.0     | 68          | 52.3       | 7                 | 6.1              |
| 0302    | 0         | 0.0      | 2           | 1.5        | 0                 | 0.0              |
| 0303    | 0         | 0.0      | 1           | 0.8        | 0                 | 0.0              |
| 0401    | 0         | 0.0      | 2           | 1.5        | 9                 | 7.9              |
| 0402    | 0         | 0.0      | 1           | 0.8        | 8                 | 7.0              |
| Total   | 102       | 100.0    | 130         | 100.0      | 114               | 100.0            |

| DRB1*   |           |          |             |            |                   |                  |
| 1501    | 0         | 0.0      | 6           | 4.6        | 34                | 29.8             |
| 1502    | 21        | 20.2     | 26          | 20.0       | 15                | 13.2             |
| 1602    | 21        | 20.2     | 19          | 14.6       | 4                 | 3.5              |
| 04†     | 0         | 0.0      | 5           | 3.8        | 17                | 14.9             |
| 11†     | 49        | 47.1     | 67          | 51.6       | 3                 | 2.6              |
| 12†     | 10        | 9.6      | 1           | 0.8        | 4                 | 3.5              |
| 14†     | 3         | 2.9      | 3           | 2.3        | 26                | 22.8             |
| 08†     | 0         | 0.0      | 2           | 1.5        | 11                | 9.7              |
| 09†     | 0         | 0.0      | 1           | 0.8        | 0                 | 0.0              |
| Total   | 104       | 100.0    | 130         | 100.0      | 114               | 100.0            |

| DRB5*   |           |          |             |            |                   |                  |
| 0101    | 21        | 20.2     | 31          | 23.8       | 49                | 43.0             |
| 0102    | 0         | 0.0      | 1           | 0.8        | 0                 | 0.0              |
| 02      | 18        | 17.3     | 19          | 14.6       | 4                 | 3.5              |

NB: Madang and PNG highland gene frequencies from Gao et al (1992)
†DR4, 8, 9, 11 and 14 alleles were not subtyped in the Wosera.
highlands. Although some DQ alleles were shared, e.g. DQA1*0102 and DQB1*0601, the most frequent alleles in the highlands (DQA1*0101, DQA1*0301, DQB1*0503 and DQBl*0602) were not found at appreciable frequencies in Madang or Wosera. DRB1*1502, DRB1*1602 and DR11 were very similar in frequency in Wosera and Madang (Table 5.3). The characteristic highland DR2 (DRB1*1501), and DRB1*08 alleles were not detected in Wosera. HLA-DR11, uncommon in the highlands, was the most frequent DRB1 allele in Wosera. The Wosera population also had a high frequency of DR12. Apart from DR14, there was an apparent lack of low frequency alleles, although DR11 and DR14 subtyping might reveal further heterogeneity.

5.3.3 Linkage disequilibrium of class II alleles in Wosera

HLA haplotypes that are in linkage disequilibrium (i.e. combinations of alleles associating together more than would be expected at random) can be used as markers of migrational relationships between Papua New Guineans and other Oceanic populations (Serjeantson et al. 1982). Anthropological studies have shown that some HLA haplotypes in the Pacific region are population specific (Serjeantson 1989c). For example, the A11,B40 haplotype is found in appreciable frequencies in Australian Aboriginal populations and also persists in Melanesians, whereas the A9,Bw22 haplotype is present commonly found in peoples of Polynesian descent (Serjeantson et al. 1982). A similar situation prevails for a number of class II haplotypes (Gao and Serjeantson 1991b).

Linkage disequilibrium between alleles of different loci can be used to demonstrate how genetically similar two populations are. For example, two populations that had recently diverged from one another would be expected to share many haplotypes, as insufficient time would have passed for linkage disequilibrium between all alleles to have been lost. However, if two populations have been separated for a long time only very closely linked allelic association (e.g. alleles of the B and C loci) will remain. Significant linkage between HLA-B and C alleles is due to the closer proximity of their loci (see Figure 5.1), with a recombination distance of only 0.2% (cf. 0.8% for HLA-A,B) (Bodmer and Bodmer 1978). Five hundred generations (or 20,000 years) are
required before linkage disequilibrium between alleles of the HLA B and C loci will decrease by a factor of 5. In contrast only 200 generations (or 5,000 years) is required for the linkage between the A and B loci to decrease by the same amount (Serjeantson 1989a). The PNG highland population of Asaro speakers have no apparent linkage disequilibrium in the HLA-A,B and HLA B,DR haplotypes, but still have allele haplotypes between their B and C loci (Bhatia et al. 1984). This suggests that this population has been isolated for between 5,000-20,000 years.

Two locus linkage disequilibrium has been calculated in Wosera to identify the presence of linked HLA class II haplotypes. The DRB1*1502 allele showed significant linkage disequilibrium with DQA1*0102 (p >0.01) and DQB1*0601 (p >0.001) alleles (Table 5.4). The DRB1*1602 allele was also shown to be in strong linkage disequilibrium with DQA1*0102 (p >0.001) and with DQB1*0601 (p >0.001) and the DRB1*1101 allele was most often found in linkage with DQA1*0501 (p >0.001) and DQB1*0301 (p >0.001) alleles. Significant two locus negative associations are also shown in Table 5.4 to confirm the significance of the positive associations.

5.4 Discussion

The HLA class II antigens present foreign peptides to the immune system and consequently play a role in protection from malarial infection. Although the sample size of the study population is too small to allow any direct comparisons of allele frequency with decreased parasitaemia, the similar allele distribution between the endemic malaria regions of Madang and Wosera might suggest the presence of a common factor that provides protection from malarial infection. There are several proposed mechanisms which might suggest that the HLA system provides protection from malarial infection; the presence of a protective allele or an increase in the number of HLA heterozygotes either by frequency-dependent or overdominant selection.

Studies in Sardinia, Tanzania and The Gambia have identified several HLA alleles that provide protection from malaria (Piazza et al. 1973; Osoba et al. 1979; Hill et al. 1991a). Piazza et al. (1973) identified an increased frequency of A2 and Bw17 antigens in two
Table 5.4  HLA-DR,DQ two-locus haplotypes with significant linkage disequilibrium

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Haplotype frequency</th>
<th>Coefficient of linkage disequilibrium (Δ)</th>
<th>Chi-squared (1df)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive associations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRB1-DQA1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1502-0102</td>
<td>0.159</td>
<td>0.062</td>
<td>8.9 *</td>
</tr>
<tr>
<td>1602-0102</td>
<td>0.179</td>
<td>0.099</td>
<td>23.9 **</td>
</tr>
<tr>
<td>1101-0501</td>
<td>0.380</td>
<td>0.113</td>
<td>20.8 **</td>
</tr>
<tr>
<td>DRB1-DQB1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1502-0601</td>
<td>0.214</td>
<td>0.166</td>
<td>81.3 **</td>
</tr>
<tr>
<td>1602-0502</td>
<td>0.161</td>
<td>0.124</td>
<td>54.2 **</td>
</tr>
<tr>
<td>1101-0301</td>
<td>0.374</td>
<td>0.177</td>
<td>45.7 **</td>
</tr>
<tr>
<td><strong>Negative associations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRB1-DQA1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1502-0501</td>
<td>0.070</td>
<td>-0.072</td>
<td>13.04 **</td>
</tr>
<tr>
<td>1602-0501</td>
<td>0.028</td>
<td>-0.144</td>
<td>56.72 **</td>
</tr>
<tr>
<td>1101-0102</td>
<td>0.060</td>
<td>-0.124</td>
<td>25.13 **</td>
</tr>
<tr>
<td>DRB1-DQB1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1602-0301</td>
<td>0.098</td>
<td>-0.181</td>
<td>80.93 **</td>
</tr>
<tr>
<td>1101-0502</td>
<td>0.012</td>
<td>-0.101</td>
<td>21.02 **</td>
</tr>
</tbody>
</table>

NB: d.f. = 1, * p < 0.01; ** p < 0.001.
lowland populations of Sardinia experiencing frequent episodes of malaria infection. The Tanzanian study showed that titres of immunofluorescent antibody (IFA) to falciparum antigens were much higher in individuals with both A2 and Aw30. They also demonstrated that individuals with A2 and Bw17 also had higher IFA titres, although this finding was not statistically significant. They hypothesized that individuals with these antigen combinations have a greater capacity for mounting an antibody response to malaria and therefore have an increased survival advantage (Osoba et al. 1979).

A study by Hill and colleagues in The Gambia showed that the Bw53 and DRB1*1302 alleles are associated with reduced susceptibility to severe malaria and that this protective effect is more pronounced in Bw53 and DRB1*1302 homozygotes than heterozygotes, contradicting the arguments of overdominant selection. The much higher allele frequency of Bw53 in African populations affected by malaria is thought to be a result of its selective advantage (Hill et al. 1991a). The authors went on to suggest that the Bw53 allele provides protection from severe \textit{P. falciparum} infection by more adequately presenting parasite epitopes (LSA-1) to cytotoxic T lymphocytes (Hill et al. 1992).

There are a number of HLA class II alleles (e.g. DRB1*11 and *1602) that are found at significantly higher frequencies in Madang and Wosera compared with the malaria free PNG highland population. These antigens may present malarial antigens more efficiently and thus their genes might have been preferentially selected in the lowland populations due to the selective pressure of malaria, in much the same way that some HLA-A antigens in PNG have been proposed to (Smith et al. 1994). However, without a parallel measure of decreased morbidity, parasite rate or increased antibody titre in individuals with these antigens, their protective role against malarial infection cannot be confirmed.

Carter et al. (1992) suggests that, because protection from malaria morbidity in Bw53 homozygotes in The Gambia is greater than in Bw53 heterozygotes, the protective effect of Bw53 might be a result of protection from severe disease via reduced immunopathology rather than increased immune response. They
argue that, as the human immune response to malarial infection leads to an increased production of TNFα (Clark et al. 1989a) and clinical and experimental studies have shown that high TNFα levels can promote cerebral malaria and increase disease severity (Clark 1987; Grau et al. 1987, 1989; Kwaitkowski et al. 1990), the poor response of the immune system to infection would therefore aid the survival of an individual. In other words, the immune response in individuals with Bw53 is less effective, resulting in decreased immunopathology during malarial infection and an increased survival advantage. Therefore it was proposed that the poorly immunogenic Bw53 homozygotes are doubly protected.

In Wosera there are a greater number of individuals homozygous at their HLA-DQB alleles than expected under Hardy Weinberg equilibrium. This might also be the result of a selective advantage for HLA homozygotes due to decreased immunopathology. However, as this phenomenon is not also seen in the linked HLA-DRB and DQA alleles, the increased number of homozygotes might represent difficulties involved in allele typing. As no other studies of HLA class II frequencies in PNG have shown an increased number of individuals homozygous at their class II loci, the latter conclusion is the most likely explanation. Incorrect allele typing has obvious serious implications in the conclusions that can be drawn from subsequent analysis e.g. of two-locus haplotype linkage disequilibrium data. However, the positive linkage between alleles at the DR and DQ loci is highly significant (Table 5.4) and the linkage associations are mostly probably still representative of the Wosera population.

The main difference between the Madang and Wosera populations at the DRB1, DQA1 and DQB1 loci is the limited number of alleles in the latter population. The universal paucity of diversity at DRB1, DQA1 and DQB1 is not surprising considering the linkage disequilibrium between the three loci. The fact that the Wosera population contains only a subset of the coastal class II genes is probably because Wosera is located inland and therefore has been less influenced by migrations along the coast (Wurm 1983).

Despite similarities in HLA allele frequencies, the Madang and Wosera populations are distinct from one another. The Madang
people are a mixture of non-Austronesian (NAN) speakers and Austronesian (AN) speakers, whereas the Wosera people are of NAN stock (Laycock 1973). The mixing of peoples from both language groups (AN and NAN) in Madang might account for the greater degree of class II heterogeneity than in Wosera, although this heterogeneity has also been detected in other AN speaking coastal populations (Gao et al. 1992). Both Madang and Wosera populations are affected by endemic malaria which may explain the high frequencies of certain class II alleles (e.g. DRB1*1101 and *1602) that are only found at low frequencies in the unaffected highland populations.

The Wosera two locus haplotypes are a subset of the coastal Melanesian linked haplotypes (Gao et al. 1992). As several of these alleles (e.g. DRB1*1502-DQA1*0102) are also linked in PNG highlanders, the linkage disequilibrium data in this population does not provide any additional information about the ancestral origin of the Wosera people. Class II haplotypes that are specific to either the Madang or highland populations (e.g. DRB1*0403-DQA1*0301 and DRB1*1408-DQA1*0101 respectively) are not present in Wosera.

The apparent limited allelic diversity at the class II DRB1, DQA1 and DQB1 loci, lower than expected number of heterozygotes and virtual absence of low frequency alleles does not support the hypothesis that HLA polymorphism is being maintained in this population as a result of preferentially selection of heterozygotes by either frequency-dependent or overdominant selection.

In summary, as all the detectable Wosera HLA class II alleles are found at similar frequencies to Madang populations, it is assumed that the Wosera population shares a closer genetic relationship with this population than the PNG highland populations. It is possible that some of the alleles that these two populations share might have come under selection from malaria but, as discussed above, analysis of the association of HLA alleles with a measure of protection from malarial infection on large population samples is required to confirm this.
Chapter 6

GERBICH NEGATIVE IN WOSERA
6.1 Introduction

The Gerbich blood antigens (Ge:1, :2, :3, and :4) are presented by glycophorins expressed on the erythrocyte cell surface. Absence of these antigens is rare in most populations; however, in Papua New Guinea the frequency of Gerbich negatives (Ge-) is high, particularly in its highly malarious lowland areas along the north coast (Booth et al. 1970; Booth and McLoughlin 1972). The geographical correspondence between the distribution of Gerbich-negative blood groups and malarial endemicity in PNG provides indirect evidence that it may offer some protection against the disease. Additional epidemiological evidence supports this suggestion; Serjeantson (1989) reported the Gerbich-negative individuals to be significantly less parasitaemic for *P. falciparum* and *P. vivax*.

The mode of selection maintaining the high frequency of the phenotype has not been well understood, however. Earlier investigations have been limited by the inability of serological techniques to distinguish heterozygotes from homozygous Gerbich-positive genotypes. The recent cloning of the human glycophorin C (GPC) gene, which codes for the Gerbich antigens, has facilitated the distinction of these genotypes. A more clear insight into the malaria resistance mechanisms is now possible.

6.1.1 Gerbich blood antigens

Human erythrocytes are known to express five membrane spanning glycophorins (GP): GPA; GPB; GPC; GPD and GPE. These glycoproteins, characterised by their high carbohydrate content, constitute 2% of the total RBC membrane protein. All but GPE are specifically identified by periodic-acid-Schiff (PAS) staining. GPA and GPB constitute 95% of PAS-positive material on the cell surface and carry the MN and Ss blood group systems respectively (Cartron et al. 1993). GPE is a newly defined glycophorin that shares sequence homology with GPA and GPB and also carries the M blood group antigen (Kudo and Fukuda 1990; Vignal et al. 1990). Glycoproteins GPC and GPD carry determinants of the Gerbich blood group systems (Anstee et al. 1984b).
The Gerbich antigens were first described by Rosenfield et al. (1960) when three women made antibodies to one of the antigens (anti-Ge\textsuperscript{a}) during pregnancy. Serological studies have defined a number of different variants, denoted as the Leach- (Ge\textsuperscript{-1,-2,-3,-4}), Gerbich- (Ge\textsuperscript{-1,-2,-3,4}), and Yus- (Ge\textsuperscript{-1,-2,3,4}) and Melanesian- (-1,2,3,4) (Booth and McLoughlin 1972; Booth and Hornabrook 1973), types. The Gerbich-negative phenotype (erythrocytes lacking some or all of the Ge antigens) is rare in most ethnic groups (Dahr et al. 1985) including the European, Mediterranean, Mexican, South American, Indian, Black American, and Oriental populations (Booth et al. 1970). However, the phenotype is found frequently in Papua New Guinean populations (Booth et al. 1970; Booth et al. 1972; Booth and Hornabrook 1973) with an allele frequency as high as 0.8 in some population groups (Booth and McLoughlin 1972).

Erythrocytes from the Leach phenotype lack all Ge antigens, glycophorins C and D, and show marked elliptocytosis (Anstee et al. 1984a). Other examples of Ge negative cells, the Gerbich-type and Yus-type (Cartron et al. 1993), also lack glycophorins C & D, but are not elliptocytic, and carry a new glycophorin which exhibits features of both GPC and GPD (Anstee et al. 1984b; Dahr et al. 1985).

6.1.2 Molecular characterisation of Gerbich negative phenotypes

The cDNA of GPC was first sequenced by Colin and colleagues (1986); genomic cloning and restriction analysis in 1989 allowed the organisation of the normal, Gerbich-type and Yus-type variants to be determined (Colin et al. 1989; High et al. 1989). The GYPC locus is 13.5kb in length, contains four exons and is located on chromosome 2q14-q21. A single 1.1kb mRNA is transcribed for the translation of GPC. GPD is translated from the same mRNA as GPC, but at a second initiation site (amino acid (aa) position 22) by "leaky" transcription (Cartron et al. 1993). The resultant protein products for GPC and GPD are 128aa and 107aa, respectively (Figure 6.1). The Ge:2 antigen is expressed at the N-terminal of GPD and Ge:3 by exon 3 of both GPC and GPD. The coding locations of Ge:1 and :4 have not yet been determined.
Figure 6.1 Schematic representation of normal and Gerbich negative products of proteins synthesis of the Gerbich gene.

Met1 denotes initiation site of GPC translation. Met22 denotes GPD translation initiation start site. Amino acid (aa) sizes of protein products are given in brackets. Ge:2 and Ge:3 map the known regions of Gerbich antigenic determinants; Ge:2 and Ge:3. PstI denotes the position that GPC cDNA was cleaved to create the pGFC22 N-terminal cDNA probe. Black boxes indicate the membrane spanning regions of normal and variant proteins. Modified from Cartron et al. (1993)
GYPC LOCUS
2q14-q21

mRNA (1.1 kb)

GPD - normal (107aa)

GPC - normal (128aa)

GPC - Gerbich (100aa)

GPC - Yus (109aa)

pGFC22 - 5' cDNA probe
6.1.2.1 Gerbich gene variants

The GPC gene contains two 3.4kb regions of sequence homology spanning exon 2 and exon 3. The Gerbich-type gene is characterised by a 3.4kb deletion of one of these regions in the GYPC gene and is proposed to be the result of misalignment and unequal crossover between these two regions. This leads to the elimination of exon 3 as well as most of intron 2 and the 5' region of intron 3 (Colin et al. 1989). The variant protein lacks expression of the Ge:1,2 and 3 antigens, but retains Ge:4 expression (Dahr et al. 1987). The Yus-type variant gene is also probably the product of unequal crossover, but an altered recombination break point leads to the deletion of exon 2, rather than exon 3 (High et al. 1989). The Yus-type variant proteins lack Ge:1 and 2 but retain Ge:3 and 4. Figure 6.2 illustrates the proposed mechanism of homologous recombination leading to the Gerbich- and Yus-type variants.

The Leach-type variant lacks all Gerbich antigens and is the result of a deletion in the GPC gene spanning the 3' region of the gene (including exons 3 and 4). A single nucleotide deletion leading to a premature stop codon in exon 3 can also give rise to this variant (Telen et al. 1991). The variant protein products lack the 3' membrane-spanning sequence of the gene and therefore are unable to associate with the erythrocyte membrane. However, the Leach variant gene does encode a mRNA which has been detected in circulating reticulocytes (Winardi et al. 1993). The molecular basis of the Gerbich negative Melanesian-type variant has not been determined.

6.1.3 Gerbich negatives in PNG

The majority of Gerbich negative samples in Papua New Guinea appear to be the Gerbich (Ge -1,-2,-3,+4) type; the uniquely Melanesian Gerbich phenotype (Ge -1,2,3,4) has been found at much lower allele frequencies (0-0.3) (Booth and McLoughlin 1972). The distribution of Gerbich negative phenotypes in PNG is confined largely to the Sepik, Madang, Morobe, and Northern provinces (Booth and McLoughlin 1972) and is largely absent from the Highlands and South Coast regions (Booth and McLoughlin 1972; Booth et al. 1972). The molecular characterisation of the Gerbich negative phenotype
Figure 6.2 Schematic representation of misalignment and homologous recombination at the GYPc locus.

Misalignment of the two regions of homology (indicated by arrows) results in the deletion of 3.4 kb of the GPC gene. Two Gerbich negative variants; Gerbich-type and Yus-type are created, the Yus-type due to the proposed breakage point at a) and the Gerbich-type due to breakage at b). The variant genes lack exons 2 and 3 respectively.
in Oceania confirm the presence a 3.4kb deletion in the GPC gene (Serjeantson et al. 1994). Although both the Yus- and Gerbich-type variants result from a deletion of this size, serological studies of the Gerbich antigens (Booth and Hornabrook 1973) suggest that this is the Gerbich-type variant.

6.1.4 Gerbich negative and resistance to malarial infection

It has been previously observed that the distribution of the Gerbich negative phenotype is restricted to malaria endemic populations in PNG. This observation led Serjeantson to suggest that the Gerbich negative phenotype may offer protection from malaria infection. Her hypothesis was further strengthened by the observation that Gerbich negative individuals have a lower rate of *P. vivax* and *P. falciparum* infection *in vivo* than Gerbich positive individuals (Serjeantson 1989e). This study extends this work by allowing the comparison of Gerbich negative genotypes, as well as phenotypes, with the incidence of malarial parasitaemias in a PNG population.

6.2 Materials and Methods

6.2.1 Gerbich negative screening

Fifty-one individuals from Wosera were chosen at random and screened for the 3.4kb Gerbich-type deletion. Briefly, 10μg DNA from these individuals was digested with *Bgl*II, electrophoresed on agarose and blotted onto nylon membranes. These membranes were hybridised to the pGCF22 full length cDNA probe obtained from Dr L.P. Cotten (la Van Kim et al. 1987). To increase the sample size of the Ge -/- haplotype for later comparisons of genotype vs presence of parasitemia, children of parents carrying a Ge +/- or Ge -/- haplotype were typed. Of the 21 additional DNA samples digested by PstI, Southern blotted as before, and hybridised with a 385bp 5' PstI fragment of the pGCF22 cDNA probe (see Figure 6.2 for cDNA probe) only one individual was found to be Ge -/- . Ge +/- and Ge +/-/ haplotypes could not be differentiated due to a very similar banding pattern and were consequently excluded from subsequent analysis. samples and 13kb + 9kb tor individuals _... .
3.3kb and 1.25kb (for Ge +/+), 3.3, 1.3, and 1.25kb (Ge +/-), and 1.3kb (Ge -/-) (Tanner et al. 1988; Serjeantson et al. 1994).

6.2.2 Parasitaemia rates

The presence of the four malarial parasites that infect humans (Plasmodium falciparum, P. vivax, P. ovale, and P. malariae) was detected on thin blood slides taken during July, 1990. The number of parasite-infected erythrocytes per 2000 erythrocytes was recorded. The % of individuals affected by each parasite species present (P. falciparum, P. vivax and P. malariae), as well as by any or a mixed plasmodium infection, was recorded for Gerbich genotypes and phenotypes.

6.3 Results

6.3.1 Gerbich negative in the Wosera

Initial Gerbich screening of the Woserans revealed 11 Ge -/- homozygotes, 18 +/- heterozygotes and 22 Ge +/+ homozygotes. Hardy-Weinberg analysis of these results suggests that although the number of heterozygotes is slightly lower than expected, the Ge- gene is at equilibrium in this population. These results are summarised in Table 6.1. An example of the BglII restriction analysis results are shown in Figure 6.3. Results from the PstI digested DNA samples identified an additional Ge -/- sample. The remaining 20 genotypes could not be differentiated due to the difficulty discerning the resultant 1.3kb and 1.25kb bands of the Ge +/- and +/- samples. As the PstI digested samples were not collected in a random manner the Ge -/- sample was not included in the Hardy-Weinberg analysis, but was included in the subsequent analysis of genotype and phenotype versus malarial parasitaemia.
Table 6.1 Hardy-Weinberg equilibrium of the Gerbich genotypes in the Wosera population.

<table>
<thead>
<tr>
<th>Wosera population</th>
<th>Gerbich status</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+/+</td>
<td>+/−</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td>Observed numbers (o)</td>
<td>22.0</td>
<td>18.0</td>
<td>11.0</td>
<td>51.0</td>
</tr>
<tr>
<td>Expected numbers (e)</td>
<td>19.3</td>
<td>24.7</td>
<td>7.9</td>
<td>51.0</td>
</tr>
<tr>
<td>$\chi^2(1\text{df}) = \sum(o-e)^2/e$</td>
<td>0.47</td>
<td>1.63</td>
<td>1.31</td>
<td>3.42</td>
</tr>
</tbody>
</table>

NB: "-" represents a GPC gene carrying the 3.4kb deletion. * p > 0.05

6.3.2 Correlation of Gerbich negative phenotypes and genotypes with malarial parasitaemias

Samples were first sorted according to their Gerbich phenotypes to assess whether Gerbich-negativity was significantly associated with a decreased incidence of plasmodium infection. As some of the numbers in 2x2 contingency tables fell below 5, Fisher's exact test was used to determine the statistical significance of differences of plasmodium infections in Gerbich-negative and Gerbich-positive phenotypes (Fisher and Yates 1957) (see Table 6.2). The decreased incidence of *P. malariae* infected Gerbich-negative phenotyped individuals was marginally significant (p<0.1). The number of individuals affected by any plasmodium infection was significantly decreased in individuals with the Gerbich-negative phenotype (p<0.05).

The presence or absence of plasmodium infection for each Gerbich genotype was also determined. A 3x2 contingency table was used to calculate the chi-squared association between Ge genotypes and the incidence of plasmodium infection. Reduced *P. falciparum* infection in Ge (+/-) heterozygotes was marginally significant (p<0.1). A decreased incidence of any plasmodium infection was also marginally significant in Ge (-/-) homozygotes (p<0.1) (Table 6.2).
Table 6.2 % of individuals infected by malarial parasitaemias with respect to Gerbich phenotypes and genotypes in a Wosera population sample (July 1990).

<table>
<thead>
<tr>
<th>Gerbich phenotype</th>
<th>Plasmodium sp.</th>
<th>positive (N=40)</th>
<th>negative (N=12)*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P. falciparum</td>
<td>22.5</td>
<td>16.7</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>P. vivax</td>
<td>20.0</td>
<td>8.3</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>P. malariae</td>
<td>22.5</td>
<td>0.0</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Mixed species</td>
<td>12.5</td>
<td>0.0</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>Any species</td>
<td>50.0</td>
<td>16.7</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gerbich genotype</th>
<th>Plasmodium sp.</th>
<th>+/+ (N=22)</th>
<th>+/- (N=18)</th>
<th>-/- (N=12)*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P. falciparum</td>
<td>36.4</td>
<td>5.6</td>
<td>16.7</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>P. vivax</td>
<td>22.7</td>
<td>16.7</td>
<td>8.3</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>P. malariae</td>
<td>18.2</td>
<td>27.8</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Mixed species</td>
<td>18.2</td>
<td>5.6</td>
<td>0.0</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Any species</td>
<td>54.5</td>
<td>44.4</td>
<td>16.6</td>
<td>0.09</td>
</tr>
</tbody>
</table>

NB: An additional GYPC (-/-) was added to the sample set (see text for details). P values determined by Fishers Exact test. * marginally significant, ** significant

Gerbich phenotype - Any Species” is significant as it was rounded up to 0.05 and 0.05 in Gerbich genotype - P. falciparum was marginally significant as it was rounded down to 0.05.
Figure 6.3 Restriction analysis ($Bg$II) of normal (+/+), and Gerbich negative heterozygote (+/-) and homozygote (-/-) samples.

M denotes the $\lambda$ HindIII/EcoRI & HindIII molecular weight marker.
6.4 Discussion

By using RFLP analysis to screen a sample of the Wosera population for the presence of the Ge- gene variant, this study was able to evaluate the possible role of both Gerbich negative phenotype and genotype in protection against malarial parasitaemias. The Ge- gene was found at high frequency (0.39) in the Wosera. Although the number of Ge (+/-) heterozygotes (18) was slightly lower than expected (24.7) under Hardy-Weinberg equilibrium (see Table 6.1), this association was not statistically significant (p>0.05). No atypical banding patterns were observed in BglII and PstI RFLPs, suggesting that if the Melanesian-type (Ge -1,2,3,4) variant is present in this population, then the molecular lesion characterising this variant may be small and therefore undetected by restriction analysis.

Individuals with a Gerbich-negative phenotype were shown to have a decreased overall parasite rate (% individuals with any plasmodium infection) compared with individuals with the Gerbich-positive phenotypes (p=0.05). When the incidence of parasite was divided by each plasmodium species, a trend of decreased incidence of plasmodium infection in Gerbich-negative individuals was noted. This association only reached marginal significance for *P. malariae* infected individuals (p<0.1).

When the incidence of malarial parasitaemia was compared with Gerbich genotypes, the decreased incidence of overall parasite rate in Gerbich-negative (-/-) individuals was only marginally significant (p<0.1). The absence of *P. malariae* in Ge (-/-) individuals did not reach significance. A decrease in the incidence of *P. falciparum* infections in Ge heterozygotes (+/-) was marginally significant. These associations may have reached significance if the sample sizes available was larger. These results tentatively suggest that the presence of the Ge (-/-) homozygote provides protection from plasmodium infection. This protection is not obvious for Ge (+/-) heterozygotes, except possibly for individuals with *P. falciparum* infections.

The Gerbich negative phenotype is largely confined to the northern regions of PNG, where malaria is endemic. Serjeantson *et al* (1989,
1992) suggested that the Gerbich negative phenotype may have been selected in these populations as it provides some protection from malarial infection (Serjeantson 1989e; Serjeantson et al. 1992). Serjeantson's 1989 study showed that there was a significantly decreased percentage of Gerbich negative *P. falciparum* and/or *P. vivax* infected individuals compared with Gerbich positives (Serjeantson 1989e), and that no single *P. falciparum* infections were noted in Gerbich negative individuals. However, this serological study was unable to discriminate heterozygote carriers of the deleted gene from Gerbich negative individuals.

Data from the present study supports the previous conclusion by Serjeantson that Gerbich negative phenotype provides protection from plasmodium infection (see Table 6.2). This association does not, however, reach significance when the incidence of infection by each plasmodium species is considered in isolation. As the population sample size of the present study is small, it is possible that any parasite associations present may have been missed. However, contrary to the previous suggestion by Serjeantson this study was able to demonstrate that single *P. falciparum* infections do, in fact, occur in Gerbich negative individuals.

There are two proposed mechanisms by which protection against malarial infection might be mediated in Gerbich negative erythrocytes; by altered merozoite invasion due to the presence of a variant erythroid receptor or by increased phagocytosis of parasitised cells. If GPC or GPD act as a receptor for merozoite invasion, the Gerbich negative phenotype might confer resistance to malarial infection by lacking these receptors. Several examples of this type of resistance have already been demonstrated. The Duffy antigen is coded by the human erythrocyte chemokine receptor (Chaudhuri et al. 1994). Duffy negative (Fy a-, b-) erythrocytes that lack this receptor are resistant to invasion by *P. vivax* and *P. knowlesi* (Miller et al. 1975, 1976) but not *P. falciparum* (Miller et al. 1977). Merozoite invasion is inhibited due to the lack of merozoite/erythrocyte membrane junction formation in Duffy negative erythrocytes (Miller et al. 1979). The lack of GPA, in En(a-) cells, has been shown to inhibit the rate of *P. falciparum* invasion in vitro (Pasvol et al. 1982b). Similarly, S-s-U cells that lack GPB are also shown to exhibit a decreased rate of *P.
falciparum invasion, and elimination of both GPA and GPB leads to complete inhibition of parasite invasion (Pasvol et al. 1982a).

In all of these cases, both alleles coding for the variant receptors must be absent (i.e. homozygous negative). If glycophorin C or glycophorin D are the normal merozoite receptors, then only Ge (-/-) erythrocytes should provide protection against malarial infection. To test this hypothesis, the parasitaemia in individuals with variant GPC genes was compared with the parasitaemia in those that carry the normal and heterozygote GPC genes (i.e. phenotype vs. parasitaemia). Each analysis was divided by parasite, on the assumption that the variant glycophorin C may resist invasion of one, but not all, Plasmodium species.

Considered on their own, none of the individual plasmodium species were demonstrated to be less prevalent in Gerbich negatives, although a decreased incidence of P. malariae infection in Gerbich negative individuals was marginally significant (see Table 6.2). To date, no erythroid receptor for P. malariae merozoite recognition, binding and internalisation has been identified. It is possible that glycophorin C or glycophorin D may play a role in its binding and/or invasion into the erythrocyte. However, a previous study by Serjeantson showed no difference in P. malariae invasion in Gerbich-positive and Gerbich-negative individuals (Serjeantson 1989e).

The effect of the Gerbich-negative phenotype has also been assessed by in vitro studies. Although Leach-type Gerbich-negative erythrocytes have been demonstrated to have decreased P. falciparum invasion in vitro (Pasvol et al. 1984), Ge(a-) erythrocytes are normally invaded (Miller et al. 1977). It is possible that the Gerbich-type variants confer resistance independently to the plasmodium invasion pathway. As erythrocytes lacking glycophorin C and D have decreased membrane stability (Reid et al. 1987), plasmodium protection may be mediated by increased deformability of Gerbich-negative compared with Gerbich-positive erythrocytes (Serjeantson 1989e). In much the same way that HbS heterozygotes resist P. falciparum invasion by selective phagocytosis of parasitized cells due to low oxygen present in the capillaries where P. falciparum matures (Luzzatto et
al. 1970), Ge- homozygotes might be protected from plasmodium infection via the preferential sickling of parasitized cells. If the presence of the variant glycophorin C and D proteins does result in this intracellular instability, then deformability, and therefore protection, would be most pronounced in Ge (-/-) homozygotes but may also occur to a lesser extent in Ge (+/-) heterozygotes.

Due to the small sample size of the present study, no definitive conclusions can be drawn about the role of the Ge- gene in protection from malarial infection can be drawn. This exploratory study does suggest that decreased plasmodium invasion is more pronounced for Gerbich-negative homozygotes. However, the precise mechanism by which the Gerbich-negative gene provides protection from malarial infections awaits larger sample sizes, the analysis of non-immune children (in which the effects of genotype on parasite density can be more appropriately assessed) and further in vitro analysis with erythrocytes deficient for the Gerbich-type (rather than Leach-type) variant.
Chapter 7

α THALASSAEMIA AND OVALOCYTOSIS IN WOSERA
Within PNG, both α thalassemia and ovalocytosis, genetic disorders that affect the red blood cell, are largely confined to regions of the country where malaria is endemic. However, the gene frequencies of these variants vary considerably within these malarious regions. However, the frequency of these traits varies across the country. The genetic bases of these defects have been investigated to determine their frequency in the Wosera population and their possible role in protection from malarial parasitaemias.

7.1 α thalassaemia

7.1.1 Molecular organisation of α globin gene

Haemoglobin is a protein complex consisting of two α globin subunits, two β globin subunits and four associated heme groups. It constitutes the major protein component of erythrocytes and is responsible for the reversible binding of oxygen and carbon dioxide for their circulation around the body. The α globin proteins are coded by genes in the α globin gene cluster on chromosome 16. This gene cluster consists of two α globin genes (α2 and α1), an embryonic α globin-like gene (ξ2), three pseudo genes (ψζ1, ψα2, ψα1) and a gene of unknown function (θ1) arranged in a 5'- ζ2 - ψζ1-ψα2 - ψα1 - α2 -α1- θ1- 3' order (see Higgs et al. 1989 for review). Adult β globin is coded by a gene in the β globin gene cluster on the short arm of chromosome 11. Other genes in this cluster are the δ globin gene (δ), an embryonic globin gene (ε), two fetal globin genes (Gγ, Aγ) and a β globin pseudo gene (ψβ1) arranged in a 5'- ε - Gγ - Aγ -ψβ1 - δ - β - 3' order (Antonarakis et al. 1985).

Disorders affecting globin gene synthesis are broadly known as haemoglobinopathies. These can be divided into two groups; those that produce a variant protein product (e.g. HbS or HbJ) and those that lead to the underproduction of globin chains - the thalassaemias (Hill et al. 1989). α thalassaemia results from the dysfunction of one or more of the four functional α globin genes. There are various forms of α thalassaemia. The milder phenotype (α+ thalassaemia) involves the deletion of a single α globin gene (-α/). Homozygotes for this deletion (-α/-α) are clinically characterised by mild hypochromic anaemia and heterozygotes (-α/αα) often do not present any clinical complications. α° thalassaemia denotes deletion of both the α globin genes of a
single chromosome (--/). The homozygous form of this condition (--/-) is lethal, but the heterozygous form (--/αα) is clinically indistinguishable from -α/-α (Flint et al. 1986). Due to the similar clinical presentations, various forms of α thalassaemia were difficult to distinguish prior to their characterisation by DNA analysis (Flint et al. 1986).

There are four common α globin deletion variants, a 4.2 kb deletion (-α^4.2) and three 3.7 kb deletions (-α^3.7I-III) respectively (Yenchitsomanus et al. 1985). Due to the sequence homology of the two α globin genes (in the X, Y, and Z boxes), these deletions appear to have occurred during meiosis probably as a result of misalignment and unequal crossing-over (Dickerson and Geis 1983). The variant α globin genes that result from these misalignments are summarised in Figure 7.1.

7.1.2 α thalassaemia and protection from malaria

α^+ thalassaemia has been found extensively and almost exclusively in regions affected by malaria. α^0 is found in the Mediterranean region and in Southeast Asia (Weatherall 1987), but is rare in Melanesia (Flint et al. 1986). In 1949, J.B.S. Haldane proposed a connection between the parallel distributions of malaria endemicity and thalassaemia. Since then others have gone on to confirm this correlation in the Mediterranean, Africa, South East Asia and Melanesia (Piazza et al. 1972; Oppenheim et al. 1984; Yenchitsomanus et al. 1985). However, the correlation is not absolute as in some areas, such as Nigeria, malaria is endemic but thalassaemia is uncommon (Luzzatto 1979).

There have been several suggested mechanisms by which thalassaemic erythrocytes might provide increased protection from plasmodium infection. These include; resistance to invasion due to decreased parasite binding (Luzzi and Pasvol 1990), impaired parasite development as a result of increased oxidative stress (Clark et al. 1989b, Serjeantson 1989c) and increased immune clearance due to the increased expression of malarial neoantigens on the surface of infected erythrocytes (Marsh et al. 1989, Yuthavong et al. 1990, Luzzi et al. 1991). However, experimental demonstration of these mechanisms have been largely
Figure 7.1 Schematic representation of \( \alpha \) globin gene cluster organisation and recombination.

A. A schematic representation of the genomic organisation of the \( \alpha \) globin gene cluster. Functional genes are indicated by black boxes, pseudogenes by white boxes. Regions of variable length (hypervariable regions or HVRs) are denoted by jagged lines. X, Y and Z boxes denote regions of sequence homology surrounding the \( \alpha_2 \) and \( \alpha_1 \) genes.

B. Interchromosomal misalignment and crossover of the \( \alpha \) globin genes leading to four common \( \alpha \) thalassaemia deletion variants. The Z box contains three regions within which crossover occurs (denoted by the shaded, clear and bold boxes) to produce the (rightward) \( \alpha_{3.7I-II} \) deletion variants and a \( \alpha_2/\alpha_1 \) hybrid gene. Recombination within the X box results in a (leftward) 4.2kb deletion including the \( \alpha_2 \) gene. Regions in which crossover occurs in each case are shown in black. Recipical triplicated (\( \alpha\alpha\alpha / \)) \( \alpha \) globin genes also result from these recombination events. Adapted from Flint et al. (1993)
unsuccessful. *In vitro* cell cultures designed to determine whether thalassaemic erythrocytes are more resistant to merozoite invasion than normal erythrocytes failed to do so (Pasvol and Wilson 1982). Also, decreased binding of *P. falciparum* to thalassaemic cells could not be demonstrated *in vitro* (Luzzi and Pasvol 1990).

7.1.2  $\alpha$ thalassaemia in PNG

Within Melanesia the correlation between $\alpha^+$ thalassaemia gene frequencies and malaria has been shown to be both latitude and altitude dependent (Hill *et al.* 1985; Flint *et al.* 1986). Restriction fragment length polymorphism (RFLP) technology, used to identify the $\alpha$ thalassaemia variants found in PNG, has shown that although the subtype ($-\alpha^3.7III$) is rare elsewhere, it is common in island Melanesia and Polynesia. The $-\alpha^4.2$ variant predominates in the northern coast of PNG and some areas of Island Melanesia (Hill *et al.* 1989), whereas the $-\alpha^3.7I$ subtype is common to the southern coastal populations of PNG (Flint *et al.* 1993). In Madang, the frequency of $\alpha^+$ genes is near fixation at 97% (Yenchitsomanus *et al.* 1985). Selection due to malaria, rather than genetic drift or founder effects, is believed to be responsible for the high frequency of $\alpha^+$ thalassaemia deletions in PNG (Flint *et al.* 1986).

7.2 Ovalocytosis

7.2.1 Ovalocytosis in PNG

Hereditary ovalocytosis was originally defined as a condition in which >50% of erythrocytes appear oval in shape with a length/width ratio greater than 1:1 but less than the 2:1 ratio of elliptic erythrocytes (Amato and Booth 1977). The condition is also characterised by the failure of erythrocytes to sediment, the presence of oval macrophages, a tendency for the ovalocytes to be stomacytic or knizocytic in shape and absence of rouleaux formation. Ovalocytosis is uncommon in most populations, but has been documented in appreciable frequencies in parts of Southeast Asia (Husain and Ruff 1991; Jones *et al.* 1991) and Melanesia (Booth *et al.* 1977).
Within PNG the distribution of ovalocytosis is mainly, but not exclusively, confined to coastal regions that are affected by endemic malaria (see Figure 7.2). Two studies along the western border of PNG (with Irian Jaya) have detected significant frequencies of ovalocytosis in the non-malarious Sepik/Fly divide (Holt et al. 1981; Schuurkamp et al. 1989). This divide has been suggested as a possible route for ovalocytosis migration from the Northern to the Southern coasts of PNG, as no single migration would otherwise account for the north/south distribution of ovalocytosis (Wurm et al. 1975, Holt et al. 1981).

7.2.2 Ovalocytosis and protection from malaria

Epidemiological studies of ovalocytosis in PNG have shown that persons with ovalocytosis have fewer *P. vivax* and *P. malariae* infections and lower *P. falciparum* densities *in vivo*, suggesting it may offer resistance to malarial infection (Serjeantson et al. 1977; Cattani et al. 1987). *In vitro* studies have also shown that erythrocytes from individuals with ovalocytosis do not support *P. falciparum* (Kidson et al. 1981) or *P. knowlesi* growth (Hadley et al. 1983).

The decreased parasite invasion of ovalocytes *in vitro* is now thought to be due to the rapid depletion of intracellular ATP to levels below the threshold for parasite invasion (Serjeantson et al. 1992). An equivalent depletion in ATP levels does not occur in normal erythrocytes (Dluzewski et al. 1992). Earlier studies had suggested that the increased rigidity of ovalocytes might result in decreased plasmodium infection by limiting the proliferation of trophozoites and consequently the severity of the disease (Mohandas et al. 1984). Dluzewski's study demonstrated that the restoration of intracellular ATP levels resulted in the return of parasite invasion, without any alteration in membrane rigidity.

An alternative mechanism for decreased plasmodium invasion *in vivo* is that band 3 is an erythrocyte receptor for *P. falciparum* invasion (Okoye and Bennett 1985, Husain and Ruff 1991), as the antigens that recognised monoclonal antibodies to erythrocytes infected with *P. falciparum* were found to be structurally related to the band 3 protein (Winograd and Sherman 1989). It is suggested
Figure 7.2 Ovalocytosis in Papua New Guinea.

The percentage of ovalocytosis-affected individuals in each population is indicated in brackets.
NCD = National Capital District.
that the variant band 3 protein, characteristic of ovalocytes, may inhibit parasite binding (Jones et al. 1991).

7.2.3 Molecular organisation of band 3 gene

The molecular variation underlying ovalocytosis is a nine amino acid deletion (aa400-408) in band 3 (Jarolim et al. 1991; Tanner et al. 1991). Band 3 is a membrane spanning protein responsible for anion transport across the membrane (Lux et al. 1989). The deletion occurs in the boundary of the cytoplasmic and first membrane spanning domain (see Figure 7.3). Although ovalocytes from SE Asia and Melanesia were originally believed to be different, due to the dominant and recessive modes of inheritance, respectively (Booth et al. 1977), ovalocytes from Malaysia, the Philippines, PNG, Mauritius and Indonesia have all been shown to express the band 3 deletion in a codominant fashion (Tanner et al. 1991; Schofield et al. 1992; Takeshima et al. 1994). A band 3 polymorphism known as Band 3 Memphis (A->G at nt166) (Jarolim et al. 1992) is found in 6-7% of the general population (Tanner, 1993). Despite its comparatively low frequency, it is found in tight linkage disequilibrium with the 9aa band 3 deletion characterising ovalocytosis, suggesting that this deletion may have had a single origin.

To date no individuals have been shown to carry two copies of the band 3 deleted gene and it is assumed that the homozygous state of this gene is lethal (Jarolim et al. 1991). This is consistent with the importance of the anion exchange function of the normal band 3 protein, which would be abolished in variant band 3 proteins.

In this study, molecular genetic techniques have been employed to assess the frequency of the α⁺ thalassaemia and variant band 3 genotypes in a sub population from the Wosera. The presence of these genotypes is compared to the presence of plasmodium infection to assess whether they might provide increased protection in vivo.
Figure 7.3 Comparison of band 3 cDNA sequence in normal and ovalocyte band 3.

The band 3 cDNA nucleotide sequence from nt1189-1233 is shown. Proposed amino acid codes are shown above, including the amino acid numbers indicating the deleted region in ovalocyte band 3. The functional protein domains are shown by arrows (modified from Tanner et al. 1992).
7.3 Materials and Methods

7.3.1 α thalassaemia typing by RFLP analysis

Fifty-two DNA samples from the Wosera were digested with BglII and electrophoresed on 0.8% agarose alongside HindIII + HindIII/EcoRI digested lambda marker DNAs (Bresatec). The digested DNA was transferred to Biotrace HP membranes by Southern blotting using 0.4M NaOH (Reed and Mann 1985). A zeta (ζ) globin probe was prepared using the Multiprime DNA labeling system (Amersham). The labeled probe was hybridized to the BglII digested DNA membranes using Nasmyths hybridization solution and incubated overnight at 65°C. Membranes were washed in increasing stringencies of SSC/SDS solutions and probe-hybridized DNA fragments were detected by exposure of the labeled membranes to autoradiographic (Fuji) film.

The -α^4.2^ variant gene was detected by the presence of an 8.4kb fragment, the -α^3.7^ variant by a 16kb fragment and the normal α globin gene by a 12.6kb fragment (see Figure 7.4). Triplicated α-globin genes (ααα/) arising as the reciprocal in the rearrangement that produces the single α-globin deletion variant were not detected. The ζ globin probe also hybridizes to a hypervariable region of the gene. In this study the -α^3.7^ and αα genes associated with a 10.5kb fragment, and the -α^4.2^ gene with a 11.3kb fragment, in this region. A constant band of 1.6kb was seen in all genotypes.

7.3.2 PCR detection of band 3 deletion

The presence of the band 3 deletion in the Wosera population was detected by PCR-based screening. Oligonucleotides flanking the coding regions of the deletion were used to prime the amplification reaction (Jarolim et al. 1991). Briefly, using 10pmoles of primers p198 (5'-GGGCCAGATGACCCCTCTG-3'; nt1098-1117) and p199 (5'-GCCGAAGGTGATGGCGGGTG-3'; nt1272-1253) 200ng of genomic DNA was amplified by 30 cycles of 96°C (70s), 58°C (50s), and 72°C (50s) in a reaction mixture of 1x Taq Polymerase Buffer (Promega), 30uM dNTPs, and 1 U Taq Polymerase (Promega) in a final volume of 50ul. Ten percent of the PCR product volume was electrophoresed
Figure 7.4 *Bgl*II (Bg) restriction site analysis of normal (*α*α/), rightward (-α3.7) and leftward (-α4.2) deletion variants of the *α* globin gene cluster

The site of hybridization of the *ζ* globin probe is indicated. Hypervariable regions of the gene are denoted by jagged lines, functional genes by black boxes and pseudogenes by white boxes. The *ψα*2 and θ1 genes are not depicted on the maps.
on 2.5% TAE agarose alongside an *MspI* digested pBR322 DNA molecular weight marker and stained with ethidium bromide.

The normal band 3 gene (N) yields an 175bp PCR product, whereas an 148bp product is obtained from the amplification of the variant band 3 gene (O). The presence of two additional bands of higher molecular weight was detected in N/O heterozygotes and these are thought to be the products of loop-out hybrids between the two PCR products.

7.3.3 Collection of parasitaemia data

Thin film stains of the four human malarial parasitaemias were scored for each of the individuals tested for α thalassaemia and/or ovalocytosis. The presence or absence of human malaria parasitaemias (*Plasmodium falciparum, P. vivax, P. malariae*) was measured against the presence or absence of α thalassaemia or ovalocytosis. The presence of any (plasmodium) infection and the presence of mixed plasmodium infections was also compared between individuals of different genotypes. Chi squared values using Yates correction were determined using the INSTAT software computer package.

7.4 Results

7.4.1 α thalassaemia in Wosera

α⁺ thalassaemia in the Wosera population was almost exclusively the -α⁴.²/ subtype (Table 7.1). A single example of -α³.⁷ was detected in a -α⁴.²/-α³.⁷ heterozygote (Figure 7.5), however the -α³.⁷ subtype was not determined. The majority of samples (84.6%) were homozygous for the -α⁴.² deletion, 13.5% were heterozygotes for the -α⁴.² deletion and the non-deletion chromosome and one individual (1.9%) was a -α⁴.²/-α³.⁷ heterozygote. None of the individuals studied were found to carry the normal complement of four α globin genes. No significant deviation from the expected Hardy-Weinberg equilibrium frequencies was found (χ²=0.46 for 1df; p>0.3).
Table 7.1 Gene frequencies of α globin and band 3 deletion genotypes in Wosera

<table>
<thead>
<tr>
<th>α globin genotypes</th>
<th>obs. #</th>
<th>genotype frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-α^4.2/-α^4.2</td>
<td>44</td>
<td>85.0</td>
</tr>
<tr>
<td>-α^4.2/-α^3.7</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>-α^4.2/α α</td>
<td>7</td>
<td>13.5</td>
</tr>
<tr>
<td>α α/α α</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

| total              | 52     |

Gene frequencies; -α^4.2 (0.923), -α^3.7 (0.007), α (0.067)
χ^2=0.46 for 1df; p > 0.3

<table>
<thead>
<tr>
<th>band 3 genotypes</th>
<th>obs. #</th>
<th>genotype frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/N</td>
<td>51</td>
<td>100.0</td>
</tr>
<tr>
<td>N/O</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>O/O</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

| total              | 51     |

NB: N= normal band 3 gene, O= variant band 3 gene.
Gene frequencies; N (1.00), O (0.00)
Figure 7.5 Autoradiograph of $Bgl_{II}$ digested genomic DNA probed with a $\zeta$ globin gene probe.

Expected restriction fragments of $\alpha$ globin variants are summarised in Figure 7.4.
7.4.2 Ovalocytosis in Wosera

All Wosera samples amplified by p198 and p199 primers showed a 175bp band indicating the presence of a normal band 3 gene and the absence of the 27bp deletion (characterising ovalocytosis) from this population (Table 7.1). A band 3 deletion control sample was amplified concurrently with Wosera samples and shows the characteristic 148bp/175bp bands of a band 3 deletion heterozygote (Figure 7.6).

7.4.3 Decreased parasitaemia vs genotype

As no band 3 deletion was detected in the Wosera population, the effect of this genotype on malarial infection could not be assessed. The high frequency of -α^{4.2}/ suggests that -α^{4.2} thalassaemia may provide protection from malarial infection. To test whether the homozygous state may have an added advantage, the number of homozygotes infected by each malarial parasite was compared with the number of heterozygotes affected (see Table 7.2).

<table>
<thead>
<tr>
<th>Plasmodium sp.</th>
<th>-α^{4.2}/ deletion</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+/+</td>
<td>+/−</td>
<td>χ²</td>
</tr>
<tr>
<td>(N=44)</td>
<td>(N=8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| P. falciparum | 22.7    | 12.5   | 0.42 | 0.52 |
| P. vivax      | 20.5    | 12.5   | 0.11 | 0.74 |
| P. malariae   | 15.9    | 25.0   | 0.06 | 0.81 |
| Any infection | 45.5    | 50.0   | 0.05 | 0.82 |
| Mixed infection | 13.6  | 0.00   | 1.23 | 0.27 |
Figure 7.6 PCR amplification of normal (N) and ovalocyte affected (O) individuals.

The 148bp product indicates the presence of a 27bp deletion of the band 3 gene, and the 175bp product a normal band 3 gene (see text for details). A heteroduplex band is formed by the loop-out hybrid of N and O PCR products.
genomic DNA

heteroduplex 175bp 148bp

heteroduplex 175bp 148bp
7.5 Discussion

Epidemiological studies of $\alpha^+$ thalassaemia in Oceania have noted its shared distribution with malaria endemicity (Oppenheim et al. 1984). The $\alpha^+$ thalassaemia gene frequency is particularly high in populations from the northern coast of Papua New Guinea and appears to be nearing fixation in those populations (Yenchitmanus et al. 1985, 1986). It has been hypothesized that the $-\alpha/$ gene is being actively selected in these populations as it provides protection from *P. falciparum* infection (Flint et al. 1986). This is supported by the observation that the reciprocal product of crossover leading to the formation of the $-\alpha/$ gene, i.e. the triplicated $\alpha$ globin gene ($\alpha\alpha\alpha/$), is only found at low frequencies in these same populations (Higgs et al. 1983).

*P. falciparum* is known to exhibit selective pressure on the sickle haemoglobin gene (HbS) in Africans. However, as the HbS homozygote does not survive to adulthood, the gene is maintained in balanced polymorphism because of survival advantage of heterozygotes. As the phenotype of $\alpha^+$ thalassaemia is mild, even in homozygotes, selection is most probably directional and will lead to fixation of the $-\alpha/$ gene in the population in due course, assuming that insufficient time has lapsed for this to have already occurred. The $-\alpha^{4.2}$ gene frequency is high (0.92) in Wosera. A similar study of $-\alpha/$ genes in the Sepik also indicated a very high gene frequency, although the sample size of this study was very small (N=6) (Yenchitmanus et al. 1986).

Near fixation of $\alpha^+$ thalassaemia has also been seen in the Tharu people of Nepal. The high frequency of this malaria-related gene was suggested to explain the decreased incidence of malaria in the native Tharu people compared to the recent immigration of a non-Tharu population. However, as has been the case in other *in vitro* assays of merozoite invasion, there was no significant difference in *P. falciparum* invasion and growth between $\alpha$ thalassaemic and normal erythrocytes (Modiano et al. 1991).

Ideally, it was hoped that this study could have compared the presence of parasites in individuals possessing the normal $\alpha$ globin genes ($\alpha\alpha/$) with those heterozygous or homozygous for the $-\alpha/$
gene in vivo. However, the -α/ gene was present in all individuals tested. Restriction fragment analysis determined that majority of these individuals were homozygous for the -α4.2/ gene (85%) and the remaining 15% were heterozygous, including one individual who was a compound heterozygote for the -α3.7/ and -α4.2/ genotypes (-α3.7/-α4.2). As no individuals carried the normal compliment of α globin genes, i.e. αα/αα, the frequency of plasmodium infection was compared between -α/ hetero- and homo-zygotes, assuming that a double dose of the protective -α4.2 gene provided increased protection in vivo. This analysis was carried out for each of the Plasmodium sp. present in the Wosera and the results are summarised in Table 7.2. As the sample sizes involved in this type of analysis were small, the effect of -α4.2 genotype on any plasmodium infection was also determined. An earlier study of the protective effect of the Gerbich-negative phenotype on the presence of plasmodium infection in vivo suggested that the protection from P. falciparum infection was masked in a mixed P. falciparum/P. vivax infection (Serjeantson et al. 1989). To determine whether mixed plasmodium infections might obscure a protective effect, this study also compared α globin genotypes in individuals with mixed plasmodium infections.

Chi squared analysis (using Yates correction) could not demonstrate any significant difference in number of individuals infected with any of the plasmodium species in -α4.2 hetero- and homozygotes. There was also no significance in heterozygotes and homozygotes with any or mixed plasmodium infections (see Table 7.2). Although these results appear inconclusive, the near fixation of the gene argues that this gene is being selected in this population, although without the presence of αα/αα genotypes for in vivo comparison, the involvement of plasmodium infection as a selector cannot be assumed.

The protective effect of the -α4.2/ gene may have been more pronounced in a study of non-immune children, as the effect of protective genotypes is thought to be most marked before adult immunity is acquired. Even so, the absence of any normal α globin genotypes could indicate that individuals carrying the -α4.2 gene have an increased probability of reaching adulthood, i.e. of surviving the non-immune stage of childhood. If this were the
case, the number of $\alpha\alpha/\alpha\alpha$ genotypes would be lower than expected under Hardy-Weinberg equilibrium. The absence of the $\alpha\alpha/\alpha\alpha$ genotype in this population, however, does not reach statistical significance ($\chi^2 = 0.46$ for 1df; $p > 0.3$).

An interesting aspect of these results is that the $-\alpha^4.2$, but not $-\alpha^3.7$, is found at high frequency in this population. The shared distribution of $\alpha^+$ thalassaemia and malaria in Oceania ends to the east of the Vanuatu Archipelago, where $\alpha^+$ thalassaemia is present, but malaria is absent. Only the $-\alpha^3.7$ variant is found in these nonendemic populations (Hill et al. 1985, Hertzberg et al. 1988). A study by Bowden et al. (1987) has shown that when compared to the $-\alpha^3.7$ thalassaemia deletion variants, the $-\alpha^4.2$ thalassaemia variant has a more severe phenotype, characterised by an increased % of Hb Barts ($\gamma_4$) in neonates and a decreased production of $\alpha$ globin. As the $\alpha 2$ globin gene is deleted in the $-\alpha^4.2$ variant $\alpha$ globin gene cluster (Figure 7.1B), this suggests that the $\alpha 2$ gene has a higher $\alpha$ globin output than the $\alpha 1$ gene. The authors also suggest that as all non-deletion $\alpha$ thalassaemias detected to date are also associated with mutations in the $\alpha 2$ gene, the more severe $-\alpha^4.2$ phenotype may be under greater selective pressure than its $-\alpha^3.7$ counterparts (Higgs and Weatherall 1983). It is possible that the phenotypically more severe $-\alpha^4.2$ variant, rather than $\alpha^+$ thalassaemia per se, has been selected by malaria and that the presence of the $-\alpha^3.7$ variant in Fiji and Polynesia is a result of gene migration from Melanesia (Hill et al. 1985).

The $-\alpha^4.2$ gene has been detected in PNG highland populations, where malaria is absent (Yenchitsomanus et al. 1986). However the low frequency of this gene and can also be attributed to the influences of gene flow or migration from the northern coastal regions where $-\alpha^4.2$ gene frequency is high.

The presence of ovalocytosis to the Austronesian (AN) speaking people of Malaysia, Indonesia, Mauritius and the Philippines, suggests that the band 3 variant gene may be restricted to populations of AN-speakers (Bellwood 1989). Ovalocytosis in PNG has previously been shown to be largely confined to the malarial endemic coastal regions, where AN-speaking populations are found. If the variant band 3 gene is indeed an Austronesian gene, this
would explain why ovalocytosis was not found in the non-AN speaking population of Wosera.

Although there appears to be outstanding epidemiological support of the role of α thalassaemia in protection from malarial infection, no direct experimental evidence has thus far been produced to support this conclusion. The near fixation of the -α^{4.2} globin gene in this population may be the result of a protective advantage from malarial infection. However, the absence of αα/αα and ovalocytosis, and the sampling of immune adults of the Wosera population, prevents the accurate assessment of the protective effect of these genetic variants in this study.
Chapter 8

DISCUSSION
8.1 Project aim

It is now well established that some variant erythrocyte proteins provide protection against malarial infection (see Nagel and Roth 1989; Flint et al. 1993 and Miller et al. 1994 for recent reviews). A large number of these studies have investigated the role that genes encoding these variant proteins play. The initial suggestion that these genes may play a role in protection from malaria came from their otherwise unexplained high gene frequencies in populations affected by endemic malaria (Livingstone 1967; Serjeantson et al. 1977; Serjeantson 1989e). With the rise in antimalarial drug and insecticide resistance, and in light of the slow road to the development of an effective antimalarial vaccine, a closer understanding of the mechanisms by which these genes prevent or decrease malarial infection is now necessary.

The frequencies of the proposed malaria protective genotypes have been extensively studied in Papua New Guinea (see Serjeantson et al. 1992 for review), but to date each study has only concentrated on one, or a few, genotypes within a given population. No study has determined the frequency of all of the presently known malaria-protective genotypes in a single population.

The purpose of this study was to determine the simultaneous distribution of $\alpha$ thalassaemia, Gerbich negative phenotype, G6PD deficiency, and ovalocytosis defects in a single village with the aim of determining the balance achieved by various combinations of genetic traits, both intra-locus and multiple-locus, in maximising malaria protection. The Wosera population was chosen because it experiences high malaria mortality and morbidity. Therefore, the genes present in this population will have experienced strong selective pressure from malaria infection thus providing an ideal opportunity for studying genotype-malaria interactions.

In this study the latest molecular genetic techniques were used to accurately discriminate variant genes from normal ones. Parasite rate data was collected to assess the relationship between the presence of each genotype and the presence of malarial parasites in each individual. As the genes of the human immune system are also subject to selective pressures from malarial infection, the degree
of heterogeneity and heterozygosity of the class II alleles of the HLA system was determined.

8.2 Experimental results

8.2.1 Frequencies of protective genes in Wosera

The gene frequencies of α thalassaemia, Gerbich negative, G6PD deficiency, and ovalocytosis in adult members of Wosera village of Apusit are given in Table 8.1.

The -α4.2 gene was found in every tested member of the population. The frequency of -α4.2 in Wosera (92%) is the highest yet recorded (Higgs and Weatherall 1983; Oppenheimer et al. 1984; Yenchitomanus et al. 1985, 1986a; Flint et al. 1986), although a significantly high frequency for the deletion (83%) was previously noted in a small Sepik sample (Yenchitsomanus et al. 1986a). In the absence of any obvious homozygous disadvantage to balance this gene variant, it appears to be approaching fixation in Wosera. The fact that this gene is only found at high frequency in populations exposed to endemic malaria suggests that malaria is the selective force driving this gene to fixation.

The frequency of Gerbich negative variant (Ge-) in Wosera was estimated to be 0.39. The apparent restriction of Gerbich negative to populations exposed to endemic malaria, and an in vivo decreased parasitaemia in Ge negative individuals (Serjeantson 1989e), suggest that the variant glycophorin C protein plays a role in protection from malarial infection.

The frequency of G6PD deficiency in Wosera males was surprisingly low (0.01), especially as an earlier study among Abelam speaking males of the East Sepik Province had recorded 8% G6PD deficiency (Kidson and Gorman 1962). It is likely that G6PD deficiency may have only recently been introduced to Wosera and therefore has not yet had the opportunity to rise in frequency in response to malarial selection.

The band 3 variant characterising ovalocytosis was also absent from this population. However, as this variant appears to be
### Table 8.1 Summary of the frequency of malaria protective genes in Wosera

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Variant</th>
<th>Sample size (2n)</th>
<th>Gene frequency (GF)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>α Thalassaemia</td>
<td>-α&lt;sup&gt;4.2&lt;/sup&gt; deletion</td>
<td>104</td>
<td>0.08 0.92</td>
<td>near fixation</td>
</tr>
<tr>
<td>Gerbich negative</td>
<td>glycophorin C deletion</td>
<td>102</td>
<td>0.61 0.39</td>
<td>high GF</td>
</tr>
<tr>
<td>G6PD deficiency</td>
<td>single base substitution</td>
<td>119*</td>
<td>0.99* 0.01*</td>
<td>low GF</td>
</tr>
<tr>
<td>Ovalocytosis</td>
<td>band 3 deletion</td>
<td>102</td>
<td>1.00 0.00</td>
<td>absent</td>
</tr>
</tbody>
</table>

NB: 2n = number of chromosomes; + = normal gene; - = variant gene; * gene frequency in males
largely confined to AN-speaking populations (Serjeantson et al. 1977), it not surprising that it is not found in the NAN-speakers of the Wosera.

A subset of the Wosera population was typed at their HLA class II loci. Comparison of class II allele frequencies in Wosera and Madang which have endemic malaria, with the allele frequencies in Goroka, a non-endemic area showed that the Wosera alleles and their frequencies more closely matched those seen in Madang. As both Wosera and Madang populations are under selective pressure from malaria, it is possible that the high frequency alleles common to both populations (e.g. DR11 and DR*1602) may provide some protection from malarial infection. Increased class II allele frequency due to protection from malarial infection has previously been seen in The Gambia (Hill et al. 1991a). However, without adequate analysis of the association of the high frequency class II alleles in Madang and Wosera with decreased malaria morbidity, their role in protection from infection remains to be proven.

The increased presence of HLA allele diversity and heterozygosity have been suggested as alternative mechanisms by which HLA alleles might provide protection from malarial infection (Bodmer 1972, Serjeantson 1989d, Hill et al. 1991b). The Wosera population showed a rather limited number of alleles (diversity) and these alleles did not appear to associate as heterozygotes any more than expected under Hardy Weinberg equilibrium. If selection by malaria does promote HLA diversity or heterozygosity, it does not appear to do so in Wosera.

8.2.2 Effect of host genotype on parasitaemia rates

As the molecular basis of the Gerbich phenotype has only recently been elucidated, this study was the first to investigate whether glycophorin C variant homozygotes and/or heterozygotes are protected from malarial infection in vivo. Gerbich negative individuals showed significantly reduced parasite rates compared to individuals with a Gerbich positive phenotype (see Table 6.2, p116). This difference was statistically significant (p=0.05) when all parasite species were pooled (i.e. phenotype versus any plasmodium infection). The decrease of P. malariae infections in
Gerbich negative phenotyped individuals was marginally significant. Although there was an apparent trend, the small sample sizes available in this study failed to demonstrate a definitive relationship between Gerbich phenotype and reduced *P. vivax* and *P. falciparum* infection.

An earlier study of *in vivo* protection of Gerbich negatives from malarial infection proposed that Gerbich negative homozygotes were refractory to single infections of *P. falciparum* (Serjeantson 1989e). The present study was able to demonstrate that single infections of *P. falciparum* do occur. Therefore, the absence of the normal glycophorin C protein may inhibit, but not prevent, *P. falciparum* infection of Gerbich negative individuals.

The effect of the -α^4.2^ phenotype on parasite rates could not be accurately assessed due to an absence of αα homozygotes in the Wosera population. Instead, an attempt was made to determine whether -α^4.2^ homozygotes provided greater protection from malarial infection than -α^4.2^- heterozygotes. No significant decrease in infection rate was detected in the -α^4.2^ homozygotes. As the -α^4.2^ gene has not yet reached fixation in the Western Pacific it has been suggested that any selective advantage is relatively small (Hill *et al.* 1987), therefore it is unlikely that any selective difference between -α^4.2^ homozygotes and heterozygotes would have been detected in a study of this sample size.

8.2.3 G6PD deficiency in Papua New Guinea - a genetic trail

Although the gene frequency of G6PD deficiency was not high in Wosera, the molecular analysis of this variant was able to give an indication of how the deficiency may have arisen in this region. Biochemical and molecular analysis of G6PD deficiencies in other ethnic groups have already proven that G6PD deficient variants have arisen independently on many separate occasions (Beutler 1992). Whenever this has occurred in a population exposed to endemic malaria the variant has apparently been selected, thus raising its gene frequency in the population (Allison 1961). Indeed, G6PD deficiency is highly heterogeneous in PNG at the biochemical level, suggesting that it has arisen there on many separate occasions (Chockkalingam and Board 1980; Chockkalingam *et al.*
Since G6PD deficiency appears to be less heterogeneous when studied at the molecular level than when studied at the biochemical level (Vulliamy et al. 1993) it was of interest to determine whether the degree of G6PD deficiency heterogeneity in PNG was as great at the molecular level as suggested by the biochemical studies.

Molecular analysis of the Wosera variant, and a G6PD deficient variant from Kalo on the southern coast of PNG, established that there was indeed G6PD heterogeneity in PNG, as each deficiency was defined by a different nucleotide substitution. A subsequent screening study of G6PD deficiency samples from individuals of seven additional PNG populations with SSO probes specific for the Kalo and Wosera nucleotide substitutions was able to identify two additional points of interest. First, while the Wosera G6PD deficient variant (nt1388A) was not found in any of the samples screened, the Kalo variant (nt1360T) was also found in samples from neighbouring southern coast populations as well in an individual from the urban population of Madang. Each of these variants was also found linked to an uncommon NlaIII polymorphism in intron 11 of the G6PD genes. This evidence suggests that the Kalo variant had a single origin and has been introduced to these populations by gene migration. Second, a significant proportion (89%) of the G6PD deficient samples did not hybridize to either of the variant probes, suggesting the presence of further molecular heterogeneity of G6PD deficiency in PNG.

It was also interesting to discover that both the Wosera and Kalo variants had been previously detected in populations of Asian origin (Du et al. 1992; Perng et al. 1992). The Kalo variant has been detected in a large variety of ethnic populations and this nucleotide substitution is thought to be a mutational hotspot (Schorderet and Gartler 1992). There have also been several documented cases of sporadic G6PD deficiency confirming the high mutation rate of this gene (Vulliamy et al. 1988; Cappellini et al. 1994). Its presence in PNG is probably the result of an independent mutational event, although its detection in the neighbouring island group of Vanuatu (M. Ganczakowski & L. Luzzatto, personal communication) also suggests a degree of regional distribution.
The Wosera variant, on the other hand, has only been described in populations of Asian ethnic origin (Chui et al. 1991, 1993; Du et al. 1992; Chang et al. 1992; Tang et al. 1992), and it is possible that, in light of the NAN-speakers original migration from Southern Asia (Bellwood 1989), this variant predates the arrival of the Wosera ancestors in PNG. However, the G6PD deficiency variant in Wosera is present in only one individual from this population. Therefore conclusions drawn from the genetic similarity of this sample and those of Asian origin might not necessarily be representative of the Wosera population as a whole.

8.3 Origins and Affinities of Protective Genotypes in Wosera

The presence of protective genotypes in Wosera is attributable to several factors: their presence in the original population, their introduction by migration, or their spontaneous presence due to gene mutation. Once present in a population, the frequency of protective genes can be profoundly influenced by the effects of genetic drift or founder effects, as well as by selection by malaria, especially in the near-isolate population groups of PNG (Allen 1983, Wurm 1983).

Traits such as Gerbich negative, which are known to be relatively homogenous, are most unlikely to be recent mutations. Such traits must have been either present in the ancestral Wosera population or introduced through genetic admixture. G6PD deficiencies and $\alpha$ globin variants, on the other hand, have multiple foci both within and outside Papua New Guinea, probably as a result of their high mutation rates (Weatherall, 1987) with $de$ $novo$ genetic mutations often amplified locally by malaria selection, and could therefore have had local Woseran origins. [There are several documented cases of sporadic G6PD deficiency variants (Vulliamy et al. 1988; Beutler 1992; Cappellini et al. 1994) and $\alpha$ globin deletions on unrelated haplotypes to support this assertion (Yenchitomanus et al. 1985, 1986b; Flint et al. 1986).]

The Gerbich variant, characterised by a single molecular variant in Papua New Guinea (Colin et al. 1989; High et al. 1989; Serjeantson et al. 1994), is confined to speakers of Papuan Torricelli phylum
languages, in regions of the Sepik Provinces adjacent or close to the Torricelli phylum language speakers, amongst present-day Austronesian speakers of the Markham Valley and further east (Booth et al. 1971; Laycock 1973; Serjeantson 1989e). The restricted distribution suggests a pre-Austronesian (NAN) origin to this gene variant (Wurm 1983). In contrast, ovalocytosis is present in Malaysian, Philippine, Melanesian, Mauritian and Indonesian populations (Tanner et al. 1991; Schofield et al. 1992; Takeshima et al. 1994) and was introduced to PNG within the last 5,000 years by Austronesian (AN) speakers. The absence of the band 3 from Wosera reflects the non-arrival of Austronesian genetic elements in Wosera.

The $-\alpha^{4.2}$ globin variant is ubiquitous in Wosera; RFLP analysis reveals its homogeneity pointing to a single origin ($Bgl\text{II}$ digested $\zeta$-globin RFLPs all showed an 11.3kb allelic fragment in association with the $-\alpha^{4.2}$ deletion, the most common haplotype in PNG; Yenchitsomanus et al. 1986a). The high frequency of this deletion can be attributed to its presence in the ancestral population and its selective advantage against malaria. The absence of other $\alpha$ globin deletions ($-\alpha^{3.7\text{I}}$ and $-\alpha^{3.7\text{III}}$) which are significantly common in other parts of Papua New Guinea, is indicative of the genetic isolation of the Woserans.

The low frequency of Wosera G6PD variant - a single example was found in the present study - suggests its recent introduction either from outside or through a local mutation. The variant has been described widely from populations of Asian extraction (Chui et al. 1991, 1993; Du et al. 1992; Chang et al. 1992; Tang et al. 1992) and may have been brought to Papua New Guinea by Austronesian speakers traveling through Southeast Asia. However, in view of the limited exchange of genes between largely coastal AN-speakers in Sepik and inland NAN-speaking Woserans, as in the case of band 3 deletion, a recent mutation within the Sepik region may be the case.

8.4 Combined effect of protective genotypes

The presence and frequency of protective genotypes does not occur evenly in every population. For example, HbC is restricted to parts
of West Africa, HbE to Asia (including India), and the duffy negative phenotype to descendants of East Africans (see Table 1.1, p9). The gene frequency of protective genotypes also varies from country to country. HbS gene frequencies range from 0-20% in Africa and 0-15% in Southern Europe and India (Livingstone 1985; Bunn and Forget 1986). The -α^4.2 gene frequency is 0.92 in Wosera, but is absent in other populations of PNG. Although not every protective gene is found in every country exposed to endemic malaria, there is some degree of overlap. For example, the distribution of G6PD deficiency appears to be similar to that of the haemoglobin variants (Allison 1961; Livingstone 1985). G6PD deficiency also occurs in populations carrying either HbS or β thalassaemia genes without apparent detrimental effects to individuals carrying both genes (Motulsky 1960). A Sardinian study suggested that the combination of G6PD deficiency and β thalassaemia provides a greater survival rate due to the decreased occurrence of haemolytic crises compared with sole G6PD deficiency (Siniscalco et al. 1961). However, in vitro studies have not shown altered growth of *P. falciparum* in G6PD deficient erythrocytes affected by β thalassaemia (Roth et al. 1983).

α thalassaemia, β thalassaemia, Gerbich negative, G6PD deficiency and ovalocytosis are all thought to confer protection from malarial infection in PNG (Serjeantson et al. 1992). Each of these defects results from single gene abnormalities that segregate independently from one another. Therefore, it is conceivable that individuals in a populations where many of these genes are present may carry more than one defect. The frequency of any given protective gene in a population, assuming that each is under a similar selective pressure, is also influenced by the disadvantage of any detrimental effects of the combined genes.

Yenchitomanus and colleagues (1986a) suggest that β thalassaemia is uncommon in some parts of PNG where α thalassaemia is prevalent as the coinheritance of these two defects leads to less globin chain imbalance, reducing the protective effect of oxidative stress and metabolic changes. As β thalassaemia homozygotes suffer a more severe disease and shortened life span compared to α thalassaemia homozygotes, α thalassaemia appears to have been more effectively selected than β thalassaemia in many populations.
For example, in Madang, where α thalassaemia is approaching fixation, the frequency of β thalassaemia is unexpectedly low (Hornabrook et al. 1972; Beaven et al. 1974; Yenchitomanus et al. 1985) compared to other regions of PNG (Curtain et al. 1962; Giles et al. 1967). In contrast, high frequencies of ovalocytosis, G6PD deficiency and α thalassaemia are known to occur together (Booth et al. 1977; Serjeantson et al. 1977; Chockkalingam 1981, Oppenheimer et al. 1984; Yenchitsomanus et al. 1985, 1986b).

In the Wosera the presence of protective genotypes is dominated by the 100% penetrance of α thalassaemia. There does not appear to be an adverse interaction between the Gerbich negative genotype and α thalassaemia as the Ge- allele is found at an appreciably high frequency (0.39). As the glycophorin C molecule normally functions in regulating membrane shape and mechanical properties (Cartron et al. 1993), and α thalassaemia is suggested to provide protection from plasmodium infection via extraerythrocytic pathways (Nagel and Roth 1989), it is possible that these two phenotypes work in conjunction with each other by effecting both the parasite invasion and development pathways. Although not present in Wosera, ovalocytosis has previously been detected in the presence of Gerbich negative (Serjeantson et al. 1994). These two genotypes do not appear to be mutually exclusive.

8.5 Mechanisms of intervention by protective genotypes

The research carried out on the genetic mechanisms of evasion of plasmodium infection is important for many reasons. First, understanding of the molecular nature of these mechanisms can aide in the development of effective targets for antimalarial drug development. Second, understanding of how the parasite evades the human host immune system is important in the development of effective vaccine candidates. Knowledge of the mechanisms of genetic protection can help to identify the host factors that are targeted by the parasite and can lead to better understanding of the basic biology of parasite/host interaction.

Genetic variation in the structure and function of the erythrocyte appears to be the main avenue of human host evasion of parasite
infection. This stage involves the invasion of merozoites and the internal growth and development of the trophozoites and subsequent release of merozoite progeny. Merozoite invasion is inhibited by the absence of cell surface receptors that are recognised for initial binding (e.g. glycophorin A) (Hadley et al. 1987) as well as the absence of antigens essential for merozoite junction formation (e.g. Duffy antigen) (Miller et al. 1975). Trophozoite development can be inhibited by altered conditions within the erythrocytes. For example, G6PD deficient erythrocytes do not clear products of oxidative metabolism as quickly as unaffected erythrocytes and trophozoite growth is initially inhibited by the toxic intracellular environment (Friedman 1979). Ovalocytes, erythrocytes with a variant band 3 protein, have an increased metabolic rate which is thought to drain the cell of the energy stores required for parasite development (Serjeantson et al. 1992). Merozoite release appears to be inhibited in HbCC erythrocytes (Olson and Nagel 1986).

Some protective mechanisms rely on the rapid identification of parasitized erythrocytes leading to a more effective immune response. Increased sickling of erythrocytes infected by *P. falciparum in vitro* and their subsequent increased clearance by phagocytosis is the proposed mechanism by which HbS carriers are protected from infection (Luzzatto et al. 1970). Similarly, thalassaemic erythrocytes infected by plasmodium have been shown to express neoantigens which lead to their phagocytosis and increased immune clearance (Marsh et al. 1989; Yuthavong et al. 1990).

*In vitro* studies have demonstrated that parasite protection can be twofold. Protection is absolute in the case of Duffy negative erythrocytes that cannot be invaded by *P. vivax* merozoites (Miller et al. 1975) whereas, glycophorin A deficiency leads to a decrease in, but not inhibition of, *P. falciparum* binding (Pasvol et al. 1982b) and therefore acts by suppressing rather than inhibiting malarial infection. This suggests that although recognition of glycophorin A is important in parasite invasion, it is not the only pathway (Sim et al. 1994).
Some protective genotypes are species-specific. *In vitro* studies have shown that the absence of glycophorin A in En(a-) erythrocytes causes a decreased rate of *P. falciparum* invasion, and yet it does not affect the rate of *P. knowlesi* (a homologue of *P. vivax*) infection (Pasvol *et al.* 1982b). On the other hand, the lack of duffy antigens (or chemokine receptor protein) specifically prevents *P. knowlesi*, and not *P. falciparum*, invasion (Miller *et al.* 1977). In accordance with this experimental finding, *P. vivax* infections are absent in African populations lacking the duffy antigens (Young 1955; Bray 1958). As yet no protective variant has been identified that is species-specific for *P. malariae* or *P. ovale*, although this is probably due to their limited distribution (Montanari *et al.* 1992).

As mentioned previously, *P. falciparum* is the most common form of the malarial parasite in PNG (Montanari *et al.* 1992) and mortality due to *P. falciparum* infection can be a powerful selective mechanism for the maintenance of protective genes. *P. vivax*, although not as common as *P. falciparum* is still found in appreciable frequencies (Montanari *et al.* 1992), and although infections by *P. vivax* are not usually fatal, they can lead to increased morbidity and indirect mortality due to complications such as anaemia. However, the selection pressure to increase the frequency of *P. vivax*-specific protective genotypes may not be as great as occurs in *P. falciparum* infections. Therefore, most effective, and highly selected, protective genotypes appear to be those that guard against *P. falciparum* infection.

This study has shown a significant decrease in the number of Gerbich negative individuals infected with plasmodium infection. Although there appears to be a general decrease in parasite rates of each plasmodium species in Gerbich negatives, this trend only reaches marginal significance for one species, *P. malariae*. This contrasts with *in vivo* results obtained by Serjeantson (1989e) demonstrating a significant decrease in the number of Gerbich negative individuals infected with *P. falciparum* and *P. vivax*, but not *P. malariae*. As both studies indicate an overall decrease in plasmodium infection of Gerbich negatives, the level of protection
provided may not be species specific. This is consistent with the suggestion that membrane instability is decreased and subsequent phagocytosis is increased in Gerbich negative parasitised erythrocytes (Reid et al. 1987).

8.6 Advantages and disadvantages of DNA-based analysis

DNA based analysis was used in this study for a number of reasons. As technical analysis was not practical in the field, and transportation of samples was required, techniques requiring viable cells were not appropriate. Buffy coats could easily be collected in the field and transported to Canberra, Australia, without a significant loss in the quality of DNA sample obtained. DNA analysis is able to detect the presence and subtype α+ thalassaemia, whose mild phenotype makes biochemical detection of -α/ heterozygotes difficult (Hill et al. 1989). DNA analysis allows the identification of Gerbich negative heterozygotes, who cannot be differentiated from Gerbich normal homozygotes by serological analysis (Serjeantson et al. 1989).

DNA analysis can also accurately determine the genetic mode of inheritance of certain protective genes. The presence of Ge- and -α4.2/ homozygotes in Wosera confirms that the resultant variant protein products still allow viable erythrocyte function. On the other hand, homozygotes of the band 3 deletion characterising ovalocytosis have not been detected (Tanner et al. 1991; Schofield et al. 1992; Takeshima et al. 1994; Serjeantson et al. 1994) and, by default, this genotype is believed to be lethal (Tanner 1993). The role of the normal band 3 protein can therefore be concluded as essential for erythrocyte viability.

DNA analysis does have some limitations. Most DNA based screening strategies require a minimum quantity of DNA sequence information (e.g. cDNA probe for RFLP analysis or sequence information for PCR primer and SSO design). This relies on the identification of the protein product associated with a variant phenotype and subsequent identification of the normal gene and its variant forms. Genotype analysis relies on the ability of the screening techniques used to detect all variant polymorphisms present in a population. The nature of mutations characterising
G6PD deficiencies (mostly single nucleotide substitutions) makes this gene a difficult one to study. Such single base changes can only be detected by very detailed, and labour intensive, DNA analysis methods, e.g. sequence analysis. Specific screening of polymorphisms, as demonstrated by the SSO screening for G6PD deficient variants in Chapter 4, was limited by its inability to accurately identify new variants.

Some of the strengths of genetic analysis can also be inherent weaknesses. For example, genetic information does not provide any information of how a variant protein may react under physiological conditions of stress. As the structure of the normal G6PD protein remains elusive, the potential structural and functional changes produced in a variant protein are difficult to predict. As is the case for any type of in vitro research, it is limited by its inability to provide a complete picture of the environmental, genetic and immune factors may interact to provide protection from severe malarial infection.

The biochemical analysis of the Wosera G6PD deficient variant was carried out to act as a bridge between data collected on the biochemical characteristics of G6PD deficiencies in PNG and the molecular data collected in this study. The biochemical properties of the Wosera variant closely matched those of two biochemical variants, G6PD Amboin and G6PD Bukitu (Chockkalingam et al. 1982), described in neighbouring populations of the East Sepik Province. Thus, these variants may share the nt1388A mutation. Molecular analysis of G6PD variants in recent years has shown that, even though some G6PD variants are not identical at the biochemical level, they may share the same molecular basis. The phenomenon of biochemically distinct but molecularly identical G6PD deficient variants has been attributed to interlaboratory variation or possible post-translational and post-transcriptional modifications (De Vita et al. 1989; Cappellini et al. 1994). The unstable nature of the deficient enzyme may also account, to a degree, for these discrepancies.

Biochemical studies have been used to detect protein variants. DNA analysis of protein variants can identify the coding regions of the protein that, when altered, affect the normal function of the
protein. In conjunction, biochemical and DNA based analyses provide a powerful means of assessing the genes involved in protection from malarial infection. The mechanisms of protection from malarial infection are potentially our best clues to acquiring a thorough understanding of the basic biology of parasite invasion and evasion.

8.7 Limitations of present study

There are some other recognised limitations of a study of this nature. Conclusions drawn from population genetic analysis are dependent upon the population sample tested being an accurate representation of the population. As all the genotypes tested fell, more or less, within the bounds of Hardy Weinberg distribution, it is assumed that this is the case. However the sample sizes analysed in this study are small, and any minor protective associations present may not have reached statistical significance.

The analysis of G6PD deficiency was limited by the time taken to process the data. The detailed molecular analysis of G6PD deficiencies in PNG, although valuable to the understanding of the evolution and distribution of the enzyme deficiency within the country, is time limiting in a population based study. This demonstrates the requirement of population based studies to utilize well characterized screening techniques that provide rapid and informative data.

The study was also limited by its analysis of malaria immune adults. Any decrease in parasitaemia due to protective genotypes during the non-immune window of life may be masked by the effects of the acquired immunity of adults (McGregor 1986; Sehgal et al. 1989). This type of analysis would be more appropriate for the analysis of children aged between 6 months and 5 years, after their protection due to maternally derived immunity and foetal haemoglobin (Pasvol et al. 1982) and before their acquisition of acquired immunity.

Although mechanisms of protection for thalassaemic erythrocytes from plasmodium infection have been suggested (see 7.1.2 p123), the mechanism of protection has not yet been experimentally
confirmed. The proposed role of thalassaemia in protection from malarial infection is still largely based on its parallel distribution with endemic malaria (Hill et al. 1987). Even assuming a role in protection, the near fixation of the -α^4.2 gene in Wosera makes it difficult to assess the level of infection in individuals carrying this variant gene from those carrying normal α globin genes.

8.8 Conclusions

The work outlined in this thesis has helped to build a more comprehensive understanding of the presence and distribution of malaria protective genotypes in PNG. The molecular characterisation of two G6PD deficiencies has confirmed the heterogeneity of this deficiency in the country. The previous detection of the Wosera variant in Asian populations suggests its possible origin from Asian ancestors. The biochemical similarity of this deficiency with other biochemical variants in Sepik suggest that the Wosera G6PD deficiency variant may have a localised distribution.

The molecular basis defining the second G6PD deficiency variant, from a population on the southern coast of PNG, has been detected in a wide range of ethnic groups. The substitution characterising this variant is indicative of a mutational hotspot in the G6PD gene, explaining its presence in so many diverse population groups. A screening study of G6PD deficiency samples from seven additional PNG populations identified that the Kalo nt1360T variant also has a regional distribution. Its presence in the urban population of Madang and the neighbouring island group of Vanuatu, is evidence of its more widespread distribution. Further molecular heterogeneity was indicated by the 89% of G6PD deficiency samples screened that did not hybridize to either the Wosera or Kalo variant probes.

A high frequency of the variant glycophorin C gene, characterising the Gerbich negative blood group, was present in Wosera. Analysis of the prevalence of malarial infection in Gerbich negatives and positives showed a decrease in the number of Ge- homozygotes with plasmodium infections (p=0.05). This level of protection did not appear to be species specific as suggested by a previous study (Serjeantson 1989e) although the decrease prevalence of P.
malariae infection in Ge- homozygotes was marginally significant. Protection from infection may be at a level common to the infective processes of all human plasmodium species. As there is no other obvious advantage for the glycophorin C variant protein, it is presumed that the Ge- variant is present in this population at a high frequency as it provides a degree of protection to its carriers, possibly from malarial infection.

The frequency of the α thalassaemia variant gene was the highest recorded for any PNG population. The -α^4.2 gene does not appear to be balanced by a disadvantage to homozygotes (Yenchitomanus et al. 1986b) and is approaching fixation in this population. This is believed to be the result of directional selection by malaria due to the protection the phenotype provides from malarial infection.

The presence of multiple protective genotypes in a single population suggests that these genotypes (α thalassaemia, Gerbich negative, G6PD deficiency and ovalocytosis) can provide a phenotypically combined affect on protection from malaria. The high frequency of Gerbich negative and α thalassaemia in Wosera suggests that these two genotypes may provide the main genetic protection from malaria in the Wosera population. The variant band 3 gene characterising the ovalocytosis defect, associated with AN-speaking populations, is not found in the NAN-speaking population of Wosera, and G6PD deficiency occurs only rarely.

Finally, analysis of the class II alleles of the HLA system in Wosera suggests the presence of high frequency alleles found commonly in populations exposed to endemic malaria may be the result of their role in protection from malarial infection. The limited allele diversity and heterozygosity in Wosera does not support previous hypotheses that suggest that these mechanisms are promoted under the selective pressure of malaria.

8.9 Future directions

The next obvious step in this type of analysis is to genotype non-immune children and compare the in vivo parasitaemia readings in those with and without protective genotypes. Increasing the population sample size would also increase the probability of
detecting an *in vivo* response to parasitaemias. Further population studies investigating the combined presence of protective genotypes may provide insight into possible combined protective effects of these genotypes in protection from malarial infection.

As this study is one of only two to describe the molecular basis of G6PD deficiencies in PNG, it would be interesting and informative to extend this aspect of the project further. The determination of the molecular basis of other G6PD variants in PNG may further support the hypothesis presented in this study that the presence of G6PD deficient variants in PNG is influenced by migration of individuals with variant G6PDs from populations of Asian origin.
REFERENCES


and immunogenicity of the synthetic malaria vaccine SPf66 in a large field trial. J Infect Dis 166: 139-144


Anstee DJ, Ridgwell K, Tanner MJA, Daniels GL, Parsons SF (1984b) Individuals lacking the Gerbich blood-group antigen have alterations in the human erythrocyte membrane sialoglycoproteins β and γ. Biochem J 221: 97-104


Betke K, Beutler E, Brewer GJ, Kirkman HN, Luzzatto L (1967) Standardization of procedures for the study of glucose-6-


Bowden DK, Hill AVS, Higgs DR, Oppenheimer SJ, Weatherall DJ, Clegg JB (1987) Different hematologic phenotypes are associated with the leftward (-α4.2) and rightward (-α3.7) α+-thalassemia deletions. J Clin Invest 79: 39-43


Boyd MF, Kitchen SF (1936) Is the acquired homologous immunity to Plasmodium vivax equally effective against sporozoites and trophozoites? Am J Trop Med 16: 317-322


Chang JG, Chen TP, Lo YS, Sy WD, Liu HJ, Chiou SS (1993) Molecular and clinical characterization of G6PD deficiency in the Chinese
infant with or without severe neonatal hyperbilirubinemia. Blood 82: 463a

Chang JG, Chiou SS, Perng LI, Chen TC, Liu TC, Lee LS, Chen PH, Tang TK (1992) Molecular characterization of glucose-6-phosphate dehydrogenase (G6PD) deficiency by natural and amplification created restriction sites: five mutations account for most G6PD deficient cases in Taiwan. Blood 80: 1079-1082


Clyde DF, McCarthy VC, Miller RM, Hornick RB (1975) Immunization of man against *falciparum* and *vivax* malarial by use of attenuated sporozoites. Am J Trop Med Hyg 24: 397-401

Cohen N, Brautbar C, Font M-P, Dausset J, Cohen D (1986) HLA-DR2-associated Dw subtypes correlate with RFLP clusters: most DR2 IDDM patients belong to one of these clusters. Immunogen 23: 84-89


Dagert M, Erhlich SD (1977) Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene 6: 23-28


Doherty PC, Zinkernagel RM (1975) A biological role for the major histocompatibility antigens. Lancet i: 1406-19


Gao X (1992) Nucleotide sequence diversity of HLA class II genes in Australian aborigines and populations of Asia-Oceania. PhD, The Australian National University


Gao X, Serjeantson SW (1991b) Heterogeneity in HLA-DR2-related DR,DQ haplotypes in eight populations of Asia-Oceania. Immunogen 34: 401-408


Greenwood BM (1990) Immune responses to sporozoite antigens and their relationship to naturally acquired immunity to malaria. WHO Bull 68 (Suppl.): 184-190


Haldane JBS (1949) Disease and evolution. Ric Sci 19: 68-76


protein caused by substitution AAG -> GAG (Lys -> Glu) in codon 56. Blood 80: 1592-1598


Laycock DC (1973) Sepik Languages. Checklist and preliminary classification. The Australian National University, Canberra


electrophoresis as single-strand conformation polymorphisms.
Proc Natl Acad Sci U S A 86: 2766-2770

Osoba D, Dick H, Voller A, Goosen TJ, Goosen T, Draper CC, de the G (1979) Role of the HLA complex in the antibody response to malaria under natural conditions. Immunogen 8: 323-338


Ramakrishnan SP (1954) Studies on Plasmodium berghei XVII. Effect of different quantities of the same diet on the course of blood-induced infections in rats. Ind J Malarialogy 8: 89-96

Reed KC, Mann DA (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucl Acids Res 13: 7207-7221


Steel CM (1984) DNA in medicine, the tools, Parts I and II. Lancet ii: 908-911, 966-968


Tang TK, Huang CS, Huang MJ, Tam KB, Yeh CH, Tang CJ (1992) Diverse point mutations result in glucose-6-phosphate dehydrogenase (G6PD) polymorphism in Taiwan. Blood 79: 2135-2140


Tiwari JL, Terasaki PI (1985) HLA and disease associations. Springer Verlag, New York


Vulliamy T, Beutler E, Luzzatto L (1993) Variants of glucose-6-phosphate dehydrogenase are due to missense mutations spread throughout the coding region of the gene. Hum Mutat 2: 159-167


Yenchitsomanus PT (1986) Molecular genetics of thalassemias in Papua New Guinea and neighbouring regions. PhD, The Australian National University


