The spatial epidemiology of the Duffy blood group and G6PD deficiency

A thesis submitted for the degree of *Doctor of Philosophy*

Rosalind Elisabeth Howes

Worcester College, University of Oxford

Michaelmas 2012

The spatial epidemiology of the Duffy blood group and G6PD deficiency

A thesis submitted for the degree of Doctor of Philosophy
Michaelmas 2012

Rosalind Elisabeth Howes
Worcester College, University of Oxford

Over a third of the world's population lives at risk of potentially severe *Plasmodium vivax* malaria. Unique aspects of this parasite's biology and interactions with its human host make it harder to control and eliminate than the better studied *Plasmodium falciparum* parasite. Spatial mapping of two human genetic polymorphisms were developed to support evidence-based targeting of control interventions and therapies.

First, to enumerate and map the population at risk of *P. vivax* infection (*Pv*PAR), the prevalence of this parasite's human blood cell receptor – the Duffy antigen – was mapped globally. Duffy negative individuals are resistant to infection, and this map provided the means to objectively model the low endemicity of *P. vivax* across Africa. The Duffy maps helped resolve that only 3% of the global *Pv*PAR was from Africa.

The second major research focus was to map the spatial distribution of glucose-6-phosphate dehydrogenase enzyme deficiency (G6PDd), the genetic condition which predisposes individuals to potentially life-threatening haemolysis from primaquine therapy. Despite this drug's vital role in being the only treatment of relapsing *P. vivax* parasites, risks of G6PDd-associated haemolysis result in significant under-use of primaquine. G6PDd was found to be widespread, with an estimated frequency of 8.0% (50% CI: 7.4-8.8%) across malarious regions.

Third, it was important to represent more detailed descriptions of the genetic diversity underpinning this enzyme disorder, which ranges in phenotype from expressing mild to life-threatening primaquine-induced haemolysis. These variants' spatial distributions were mapped globally and showed strikingly conspicuous distributions, with widespread A- dominance across Africa, predominance of the Mediterranean variant from the Middle East across to India, and east of India diversifying into a different and diverse array of variants, showing heterogeneity both at regional and community levels.

Fourth, the G6PDd prevalence and severity maps were synthesised into a framework assessing the spatial variability of overall risk from G6PDd to primaquine therapy. This found that risks from G6PDd were too widespread and potentially severe to sanction primaquine treatment without prior G6PDd screening, particularly across Asia where the majority of the population are Duffy positive and G6PDd was common and severe.

Finally, the conclusions from these studies were discussed and recommendations made for essential further research needed to support current efforts into *P. vivax* control.

Statement of Contribution and Associated Publications

The research chapters presented in this thesis (excluding Chapters 1 & 7) have either been published or present data in preparation for submission to peer-review journals. As such, these chapters represent collaborative efforts. I summarise here my personal contributions to each of these chapters. My main collaborators have been members of the Malaria Atlas Project, particularly my supervisors Simon I. Hay (SIH) and Fred B. Piel (FBP), and thesis advisors J. Kevin Baird (JKB), Peter W. Gething (PWG) and Anand P. Patil (APP). Each author's contributions are acknowledged and detailed in a summary statement at the end of each chapter. My own contribution to each chapter is summarized here.

Chapter 2: The global distribution of the Duffy blood group.

This chapter has been published and is included here in its final form. I was responsible for overseeing all aspects of this work: conceived the study (with SIH and FBP), assembled the data (with OAN and CWK), helped implement the modelling and computational tasks (with APP and PWG), and wrote the first draft of the manuscript. The final manuscript text was edited and approved by all authors.

Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Ménard, D., Williams, T.N., Weatherall, D.J. and Hay, S.I. (2011). The global distribution of the Duffy blood group. *Nature Communications*. 2:270 doi: 10.1038/ncomms1265.

Chapter 3: Duffy negativity as an indicator of *Plasmodium vivax* transmission potential.

This chapter is an overview of four studies in which the Duffy blood group maps have been applied within larger studies. I did not lead these studies, but contributed the data and knowledge, and generated the figures relating to the spatial distribution of the Duffy variants. I commented on and edited the manuscripts. The overview given in this thesis chapter is my own interpretation of these studies.

Guerra, C.A., Howes, R.E., Patil, A.P., Gething, P.W., Van Boeckel, T.P., Temperley, W.H., Kabaria, C.W., Tatem, A.J., Manh, B.H., Elyazar, I.R.F., Baird, J.K., Snow, R.W. and Hay, S.I. The international limits and population at risk of *Plasmodium vivax* transmission in 2009. (2010) *PLoS Neglected Tropical Diseases*, **4**(8): e774.

The Duffy negativity map was an important component of this study. I contributed it pre-publication and wrote the relevant sections in the manuscript and supplementary information.

Gething, P.W., Elyazar, I.R.F., Moyes, C.L., Smith, D.L., Battle, K.E., Guerra, C.A., Patil, A.P., Tatem, A.J., Howes, R.E., Myers, M.F., George, D.B., Horby, P., Wertheim, H.F.L., Price, R.N., Mueller, I., Baird, J.K., Hay, S.I. (2012) A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Neglected Tropical Diseases*, **6**(9): e1814

I contributed the Duffy data to this study and edited the full manuscript, but particularly the sections of the manuscript pertaining to the Duffy maps.

King, C.L., Adams, J.H., Xianli, J., Grimberg, B., McHenry, A., Greenberg, L., Siddiqui, A., Howes, R.E., da Silva-Nunes, M., Ferreira, M.U., Zimmerman, P.A. (2011). Fy^a/Fy^b polymorphism in human erythrocyte Duffy antigen affects susceptibility to *Plasmodium vivax* malaria. *Proceedings of the National Academy of Sciences of the United States of America.* doi:10.1073/pnas.1109621108

I generated the map of the dominant Duffy alleles and the population estimates of each phenotype. I wrote the relevant methodological sections and edited the full manuscript.

Zimmerman, P.A., Ferreira, M.U., Howes, R.E., Puijalon, O.M. Red blood cell polymorphism and susceptibility to *Plasmodium vivax*. *Advances in Parasitology*. In press (due 1st Feb 2013)

I contributed the dominant Duffy variants map and *Pv*PAR plots, and edited the full manuscript, but particularly the section pertaining to the spatial distribution of the Duffy variants.

Chapter 4: G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map

This chapter has been published and is included here in its final form. I was responsible for overseeing all aspects of this work: conceived the study (with SIH and FBP), assembled the data (with FBP, OAN, MD, MMH and KEB), helped conceive the modelling (with APP, PWG and FBP), implemented all computational tasks (with FBP) and wrote the first draft of the manuscript. The final manuscript text was edited and approved by all authors.

Howes, R.E., Piel, F.B., Patil, A.P., Nyangiri, O.A., Gething, P.W., Dewi, M., Hogg, M.M., Battle, K.E., Padilla, C.D., Baird, J.K. and Hay, S.I. (2012) G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Medicine*. 9(11): e1001339

Chapter 5: Distinct spatial trends in G6PD deficiency variants across malaria endemic regions

This chapter has been prepared for submission to a peer-review journal. I was responsible for all aspects of this work: conceived the study, assembled the data (with MD), generated the maps and wrote the first draft of the manuscript. The final manuscript text was edited by all authors.

Howes, R.E., Dewi, M., Piel, F.B., Baird, J.K., Hay, S.I. Geographic gradients of clinically significant G6PD deficiency variants within *P. vivax* malaria endemic countries. In prep.

Chapter 6: Towards a risk framework for *P. vivax* relapse therapy

Parts of this chapter have been discussed in Chapter 4, but are presented more fully in this chapter. Some of the background information is also in press as part of a larger review of G6PD deficiency, for which I wrote the first draft, and KEB helped prepare the figures. The final manuscript text was edited and approved by all authors.

Howes, R.E., Battle, K.E., Satyagraha, A.W., Baird, J.K., Hay, S.I. G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Advances in Parasitology*. In press (due 1st Feb 2013)

* * *

Much of my technical knowledge about geostatistical modelling of inherited blood polymorphisms came from working with my supervisors on mapping models for sickle-cell disease and HbC. These three papers are therefore commonly cited in the methodological chapters of this thesis, as they provided a guide to the analyses presented in Chapters 2 and 4 of this thesis.

Piel, F.B., Patil, A.P., Howes, R.E., Nyangiri, O.A., Gething, P.W., Williams, T.N., Weatherall, D.J. and Hay, S.I. (2010). Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications*. 1:104 doi: 10.1038/ncomms1104.

Piel, F.B., Patil, A.P., Howes, R.E., Nyangiri, O.A., Gething, P.W., Dewi, M., Temperley, W.H., Williams, T.N., Weatherall, D.J. and Hay, S.I. (2012) Global estimates of sickle haemoglobin in newborns. *The Lancet*. doi:10.1016/S0140-6736(12)61229-X

Piel, F.B., Howes, R.E., Patil, A.P., Nyangiri, O.A., Gething, P.W., Williams, T.N., Weatherall, D.J. and Hay, S.I. The distribution of haemoglobin C and its prevalence in newborns. *Scientific Reports* (in resubmission)

* * *

Finally, parallel publications which I have been involved with but which are not directly integrated into this thesis include:

Hay, S.I., Sinka, M.E., Okara, R.M., Kabaria, C.K., Mbithi, P.M., Tago, C.C., Benz, D., Gething, P.W., Howes, R.E., Patil, A.P., Temperley, W.H., Bangs, M.J, Chareonviriyaphap, T., Elyazar, I.R.F., Harbach, R.E., Hemingway, J., Manguin, S., Mbogo, C. M., Rubio-Palis, Y. and Godfray, H.G.H. (2010). Developing global maps of the dominant *Anopheles* vectors of human malaria. *PLoS Medicine*, 7(2): e1000209.

Battle, K.E., Gething, P.W., Elyazar, I.R.F., Moyes, C.L., Sinka, M.E., Howes, R.E., Guerra, C.A., Price, R.N., Baird, J.K., Hay, S.I. The global public health significance of *Plasmodium vivax*. *Advances in Parasitology*. 80:1-111

Piel, F.B., Howes, R.E., Moyes, C., Hay, S.I. Online biomedical resources for inherited blood disorders: genetics and epidemiology. *Human Mutation* (under review)

von Seidlein, L., Auburn, S., Espino, E., Shanks, D., Cheng, Q., McCarthy, J., Baird, J.K., Moyes, C., Howes, R.E., Menard, D., Bancone, G., Satyagraha, A.W., Vestergaard, L., Green, J., Domingo, G., Yeung, S., and Price, R. Review of key knowledge gaps in G6PD deficiency with regard to the safe clinical deployment of 8-aminoquinolone treatment regimens: a workshop report. In prep

* * *

As her supervisors, we certify that the statements of contribution listed here are a fair representation of Rosalind Howes' work.

Prof Simon Hay Dr Fred Piel

11th December 2012

Table of Contents

Figures

Tables

Acknowledgements

My supervisors, Simon Hay and Fred Piel, and a rather good Thai dinner are entirely to blame for this venture, and I'm enormously indebted to them for it. This work is the product of their tireless encouragement and support. I'm grateful to Simon for all his backing and patience, and the exciting opportunities he has given me. And to Fred, for his friendship and life-coaching which I value deeply, and for his saint-like stamina reviewing my work.

Pete Gething and Anand Patil have been key players in all aspects of the modelling developed here which are central to this thesis, I'm grateful to both for their genius. My brilliant officemates – Oli, David & particularly Katherine – have provided much needed moral support, friendship and understanding, and been stringent proof-readers.

Further afield, I owe the obsession I've developed with *P. vivax* and G6PDd in large part to Kevin Baird who has mentored me through those aspects of this thesis, and who hosted me at EOCRU for a couple of months in 2010. My Indonesian colleagues, Mewahyu Dewi and Iqbal Elyazar, have been great supports and fun guides to Jakarta. I would also like to thank Pete Zimmerman's expertise in relation to the Duffy aspects of this thesis, and for the collaborations which we have entered into.

I also acknowledge Kirk Rockett and Dominic Kwiatowski from the MalariaGEN Consortium for sharing Duffy variant data, and Carmencita Padilla for her contributions from the Filippino G6PDd newborn screening programme. Finally, I acknowledge the Wellcome Trust for the Biomedical Resources Grant (#085406) which supported this research.

The other essential team-members of this project have been my family and friends. I look to Jane & Jeffery, Kai Ma & Robin for their kindness and their examples in living life to the full. Sonya for her constant support. And Mummy, Daddy, Clare & Pick – thank you for all the nurturing getting me here, and particularly for the late-night office party. Alex, Alex, Miranda, Sarah and Siena: you're all brilliant.

Abbreviations

ACT	Artemisinin combination therapy
ADMIN0	National administrative level
ADMIN1	First administrative level
ADMIN2	Second administrative level
ADMIN3	Third administrative level
Africa+	Africa, Saudi Arabia and Yemen
AFRO	WHO African Regional Office
APMEN	Asia Pacific Malaria Elimination Network
CNSHA	Chronic non-spherocytic haemolytic anaemia
CSE Asia	Central and southeast Asia
DARC	Duffy antigen receptor for chemokines
DDT	Dichloro-diphenyl-trichloroethane
DRC	Democratic Republic of the Congo
Fy	Duffy antigen
G6PD	Glucose-6-phosphate dehydrogenase
G6PDd	Glucose-6-phosphate dehydrogenase deficiency
GMEP	Global Malaria Elimination Programme
GIS	Geographic information system
GRUMP	Global rural-urban mapping project
Hb	Haemoglobin
IQR	Interquartile range
Lao PDR	Lao People’s Democratic Republic
MAP	Malaria Atlas Project
MBG	Model-based geostatistics
MCMC	Markov chain Monte Carlo
MEC	Malaria endemic country
n	Number
NADP	Nicotinamide adenine dinucleotide phosphate

NADPH	Reduced nicotinamide adenine dinucleotide phosphate
PAR	Population at risk
PCR	Polymerase chain reaction
PPD	Posterior predictive distribution
PPP	Pentose phosphate pathway
*Pv*API	*Plasmodium vivax* annual parasite incidence
*Pv*CSP	*Plasmodium vivax* circumsporozoite protein
*Pv*DBP	*Plasmodium vivax* Duffy-binding protein
*Pv*PAR	*Plasmodium vivax* population at risk
*Pv*PR	*Plasmodium vivax* parasite rate
Q25	25^{th} percentile quartile
Q75	75^{th} percentile quartile
R_0	Basic reproductive number
R&D	Research and development
RBC	Red blood cell
RDT	Rapid diagnostic test
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism
TPP	Target product profile
UN	United Nations
WHO	World Health Organization

Chapter 1 – Introduction

1.1. The Malaria Footprint

Malaria has stamped a multi-faceted footprint on the human race. A footprint which tells of great death and disease, of evolutionary selection, of stunted economic development, of failed political and scientific endeavour, but also of humanity's resolve in the face of a quite brilliant adversary. The 20th century saw all of these.

At the peak of its distribution, thought to be around the year 1900, the threat of malaria was truly global (Figure 1.1). Although the disease was most debilitating in tropical regions, such as swathes of sub-Saharan Africa where more than 75% of young children would be infected at any given time, the "*mala aria*" was also impacting temperate regions such as Italy where its disease toll threatened the stability of the fragile newly unified state (Hay *et al.*, 2004; Capanna, 2006).

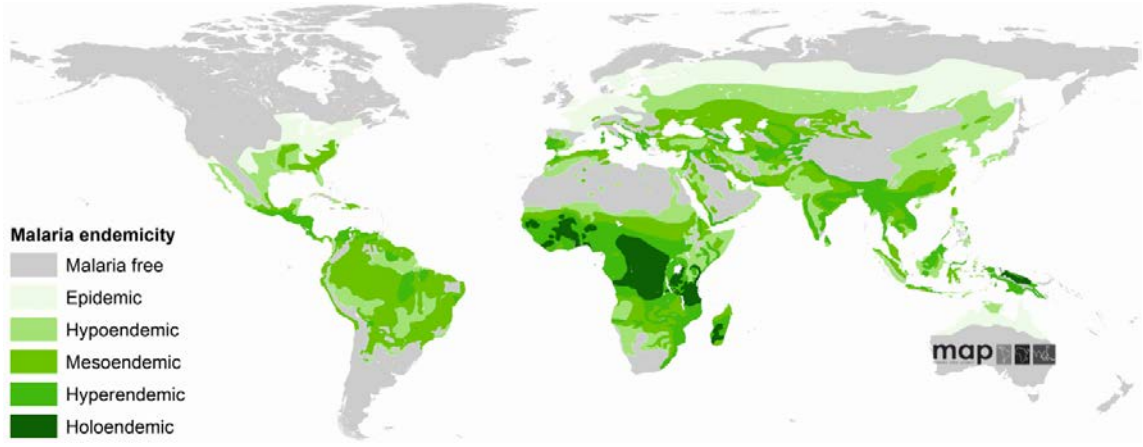


Figure 1.1. The pre-control map of malaria transmission at the peak of its distribution *circa* 1900. The endemicity classes refer to transmission endemicity (quantified as 'parasite rate', or the proportion of individuals found infected with the *Plasmodium* parasite; in this case in 2-10 year olds, PR_{2-10}): hypoendemic: PR_{2-10} <0.10; mesoendemic: PR_{2-10} = 0.10 - <0.50; hyperendemic: PR_{2-10} = 0.50 - <0.75; holoendemic: PR_{0-1} >0.75 (this class was measured in 0 to 1 year olds). This map was generated by Lysenko & Semashko (1968) and subsequently digitised by Hay *et al.* (2004).

At about this time, Ronald Ross and Battista Grassi demonstrated the role of *Anopheles* mosquitoes in vectoring the malaria *Plasmodium* parasites between human hosts (Ross, 1898; Bynum, 1999; Fantini, 1999; Capanna, 2006). The practical implications of their crucial 'mosquito theory' were rapidly put into practice and the era of malaria control was born. Widespread vector 'sanitation' programmes, together with an ongoing social and economic transition towards urbanisation, dramatically reduced malaria cases in many areas in the first half of the 20th century (Harrison, 1978; Kitron, 1987; Atmosoedjono *et al.*, 1992; Litsios, 2002). Large scale successes were particularly evident in the Middle East, South Asia and Western Pacific regions (Carter and Mendis, 2002) (Figure 1.2). Despite reductions in mortality, it is estimated that malaria claimed at least 3 million deaths annually for much of the first half of the 20th century, corresponding to around 10% of deaths globally (Carter and Mendis, 2002), though possibly up to 50% in India (Christophers, 1924).

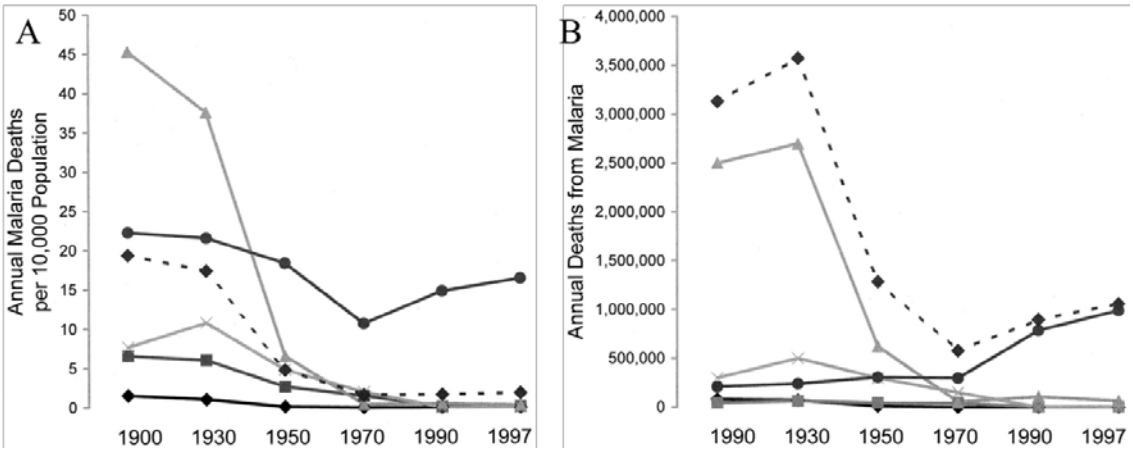


Figure 1.2. Malaria mortality in the 20th century. Panel A shows the number of malaria deaths per 10,000 population per year and Panel B represents the total number of deaths due to malaria per year. The regions are Europe and North America (⧫—⧫); the Caribbean and Central and South America (▪—▪); sub-Saharan Africa (•—•); China and Northeast Asia (×—×); the Middle East, South Asia, and the Western Pacific (▲—▲); and worldwide (⧫- - -⧫). Figure reproduced from Carter and Mendis (2002).

It was probably the alarming loss of European and North American lives through colonial and military involvements, such as in India (Ross, 1923; Christophers, 1924) and the

Second World War (Downs *et al.*, 1947; Coatney *et al.*, 1948; Condon-Rall, 1992), however, which drove the technological and pharmaceutical development of the tools which were to become the centre pieces of the United Nations' Global Malaria Eradication Programme (GMEP), launched in 1955 (WHO, 1955; WHO, 1973). Residual insecticide spraying of dichloro-diphenyl-trichloroethane (DDT) together with highly effective chloroquine treatment successfully reduced cases in many areas during the early 1960s (Najera *et al.*, 2011). By 1969, however, the programme was abandoned, and the GMEP's impact as an attempt at eradication had become synonymous with failure. The root of the downfall was probably two-fold. First, widespread use of chloroquine mono-therapy and mono-insecticide spraying selected resistant parasites and mosquitoes, and left the programme without efficacious tools (Harrison, 1978). Second, in areas where control did successfully reduce transmission and disease, political and economic will failed and investment in control measures dwindled (Najera, 2001; Najera *et al.*, 2011). Both resulted in huge relapses of infection in areas previously close to elimination. A particularly alarming example of this was in Sri Lanka, where as few as 17 cases reported in 1963 exploded into an epidemic of half a million cases four years later (Ministry of Health Sri Lanka and the World Health Organization and the University of California-San Francisco, 2012). The state of malaria control through the 1970s and 80s has been epitomized as 'Resurgence' and 'Chaos', respectively (Bradley, 1992).

The GMEP's legacy varied geographically, and although its overall target was never met, its positive contributions and lessons for contemporary programmes have been hailed (Gething *et al.*, 2010; Najera *et al.*, 2011). Malaria was successfully eliminated from most of Europe and North America, where the public health infrastructure meant that the programme's success could be sustained in the long-term (De Zulueta, 1998). Although the goal of elimination was not reached in any country in Asia, mortality was dramatically reduced (Figure 1.2), despite resistance having emerged from, and being widespread across, this region. In contrast, control of the very high intensity transmission across Africa was barely attempted (Litsios, 1966). Death rates in Africa were not substantially different between the start and the close of the 20th century (Figure 1.2A) (Carter and Mendis, 2002).

The great burden of malaria on this continent has been felt not only in its relentless death toll, but also as one of the severest impediments to social and economic development (Gallup and Sachs, 2001; Carter and Mendis, 2002).

The turn of the 21st century provided the much needed impetus to address the failures of the GMEP and get a grip on malaria's rampage: estimated to have cost 150-300 million lives during the 20th century (Carter and Mendis, 2002), and despite a century's efforts, an estimated 48% of the world's population still lived at risk of infection (Hay *et al.*, 2004). Since then, a remarkable resurgence of political will has been backed by substantial economic support from many governmental and philanthropic sources (President's Malaria Initiative; The Global Fund to Fight AIDS Tuberculosis and Malaria; Snow *et al.*, 2008; Pigott *et al.*, 2012). A number of international forums have been set up and commitments been made to control the clinical incidence of malaria (Sachs, 2002): Roll Back Malaria in 1998, the Abuja Declaration in 2000 (The Abuja Declaration and the Plan of Action, 2000; Snow and Marsh, 2010), and the Millennium Development Goals in 2002 (MDG, 2012), to name a few. However, the major policy-changing commitments were initiated in October 2007 by Bill & Melinda Gates' endorsement and substantial financial backing for a final attempt at eradication (Bill & Melinda Gates Foundation; Bill and Melinda Gates Foundation Malaria Forum). Although their announcement at the time took the research community by surprise, prompting headlines such as 'Did They Really Say…Eradication?' (Roberts and Enserink, 2007), this major gear-shift has rapidly permeated the malaria community – both at the national and regional control programme levels (APMEN; Feachem and The Malaria Elimination Group, 2009; Hsiang *et al.*, 2010; Eziefula *et al.*, 2012) and research and development (R&D) focusses (Baird, 2010; Feachem *et al.*, 2010; Alonso *et al.*, 2011). The results of this invigorated focus are evident: between 2000 and 2009, 50 of the 99 malaria endemic countries reported a decrease of at least 25% of malaria cases and mortality, and 43 reported a decrease of more than 50% (WHO, 2010; WHO, 2011). Nevertheless, in 2010 malaria was still in the top ten killers globally, with mortality estimates ranging from 655,000 (539,000-906,000) (WHO, 2011) to 1,238,000 (929,000-1,685,000) (Murray *et al.*, 2012) deaths that year.

1.2. *Plasmodium vivax*: the neglected parasite

This chapter has thus far deliberately referred to "malaria", reflecting the long-standing attitude of many. Although there are five species of the protozoan *Plasmodium* genus which are known to infect humans (*P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and the recently recognised zoonotic *P. knowlesi* (Cox-Singh *et al.*, 2008)), little effort has been made until recently to distinguish between parasites. This is still evident from reading the 2011 World Malaria Report (WHO, 2011), for instance, which does not distinguish between species in its maps or basic epidemiological statistics. This short-sighted approach overlooks important differences in the biology of the various *Plasmodium* species which have far-reaching implications for their control. A single parasite species has been considered responsible for virtually all mortality, *P. falciparum*, and as such has dominated the malaria agenda with little regard for the other parasites. *P. malariae*, *P. ovale* and *P. knowlesi* are thought to be relatively limited in their transmission and their clinical incidence much less common than that of the other two (WHO, 2011; Baird, 2013). Although *P. vivax* is the most widely distributed human malaria parasite (Guerra *et al.*, 2010), it has been largely ignored due to its alleged "benign" and rarely lethal virulence. The strong focus on *P. falciparum* has allowed this parasite to pass relatively unnoticed for the past 50 years (Galinski and Barnwell, 2008; Mueller *et al.*, 2009; Carlton *et al.*, 2011). This is still evident from a contemporary financial perspective: of an estimated $1.68bn spent on malaria R&D between 2007 and 2009, just 3.1% of expenditures were committed to *P. vivax*, compared to 44.6% on *P. falciparum* (the remaining 52.3% of investment was not reportedly targeted to a specific parasite) (PATH, 2011). The downstream effect of that investment skew, for example, is that of the 41 vaccines in preclinical and clinical development in 2011, the overwhelming majority (95%) targeted *P. falciparum*, with only two specific to *P. vivax* (PATH, 2011).

Compelling evidence is accumulating of the clinical significance of *P. vivax*, countering the long-held designation of "benign" vivax malaria; its potential to cause severe disease and death is being increasingly frequently diagnosed and reported (Tjitra *et al.*, 2008; Anstey *et*

al., 2009; Kochar *et al.*, 2009; Price *et al.*, 2009; Singh *et al.*, 2011). This misnomer is attributed to the era of *P. vivax* inoculation as a therapeutic agent for neurosyphilis patients. The fevers brought on by *P. vivax* infection were the treatment of choice for this otherwise lethal disease until the advent of antibiotics (Fong, 1937; Baird, 2013; Snounou and Perignon, 2013). The mortality data from these treatments have been recently re-examined, and together with the recent clinical evidence, demonstrate a significant mortality burden associated with *P. vivax* malaria (Baird, 2013). A wide range of severe symptoms is described, essentially similar to those traditionally attributed solely to *P. falciparum* infections: cerebral malaria, hepatic dysfunction with severe jaundice, acute lung injury, shock, renal failure, splenic rupture, severe thrombocytopenia and haemorrhage, and severe anaemia (Baird, 2007; Anstey *et al.*, 2009; Mueller *et al.*, 2009).

The public health significance of *P. vivax* is also increasingly evident. The *P. falciparum* bias which has dominated malaria R&D, means that current control interventions are not suited to the challenges presented by the distinct *P. vivax* lifecycle (Bockarie and Dagoro, 2006; Baird, 2010; Bousema and Drakeley, 2011; Gething *et al.*, 2012). Existing interventions are therefore most effective against *P. falciparum*, and areas where overall endemicity is successfully dropping are experiencing an increase in the relative proportion of *P. vivax* infections (Phimpraphi *et al.*, 2008; Feachem *et al.*, 2010; Ministry of Health Sri Lanka and the World Health Organization and the University of California-San Francisco, 2012). This is well illustrated by Figure 1.3 which shows that of the 35 malaria eliminating countries in 2010, 29 have a predominantly *P. vivax* problem (Feachem *et al.*, 2010; UCSF Global Health Group and Malaria Atlas Project, 2011). Outside Africa, where *P. vivax* endemicity is reportedly very low, the challenge of malaria elimination is therefore tantamount to the challenge of *P. vivax* (Feachem *et al.*, 2010). The unique biological characteristics of *P. vivax* to which this may be attributable are discussed below.

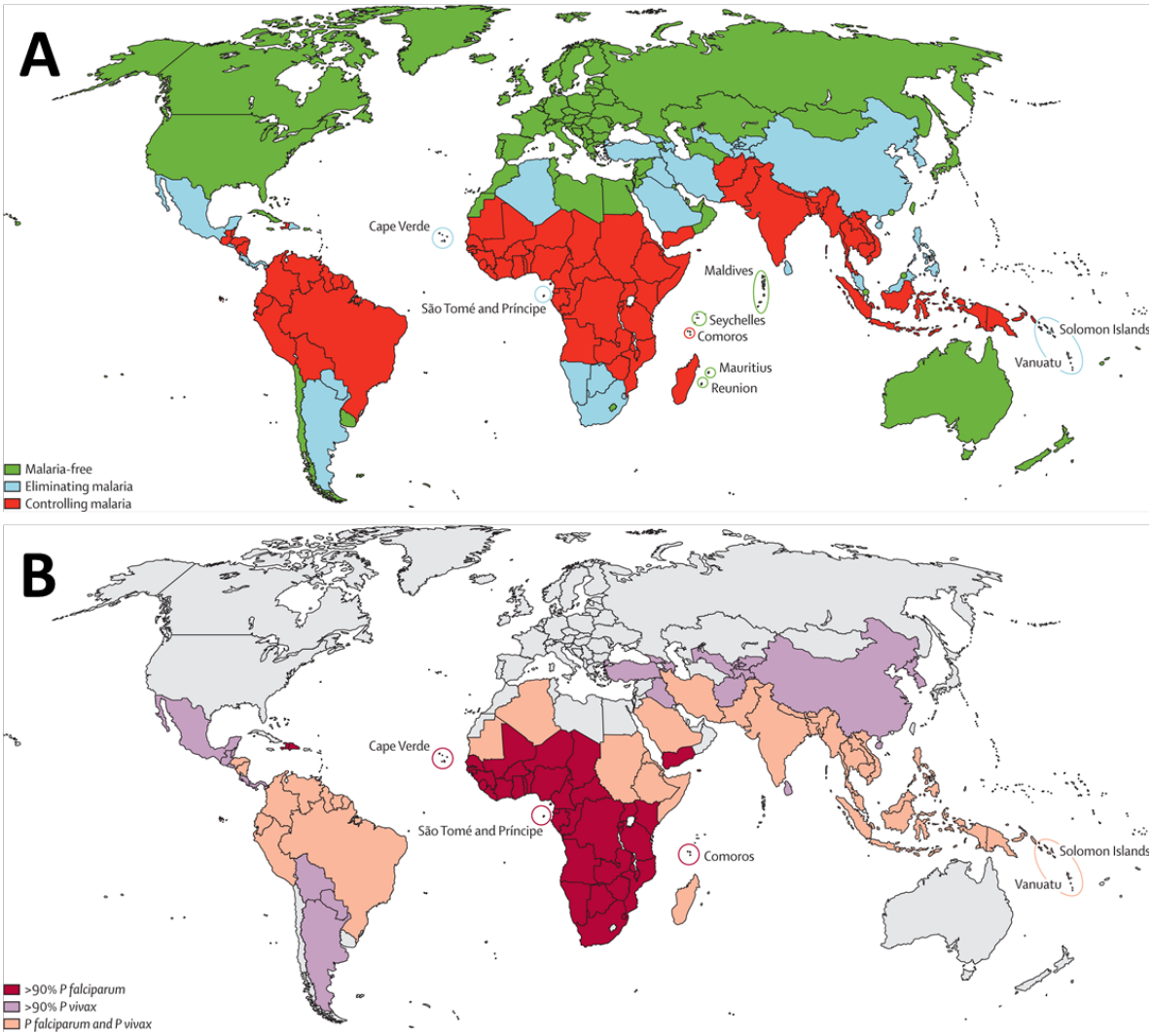


Figure 1.3. The challenge of malaria elimination. Panel A shows the status of national malaria programmes as malaria free (green), eliminating malaria (blue), or controlling malaria (red). Panel B shows the relative composition of malaria infections by parasite: predominantly *P. falciparum* (>90%) in red, predominantly *P. vivax* (>90%) in purple, or both in orange. Figures reproduced from Feachem *et al*. (2010).

1.2.1. A poor evidence-base of epidemiological data to understand P. vivax *transmission*

The neglect of *P. vivax* has permeated all aspects of our knowledge of this parasite. Basic epidemiological data, which are essential to underpin any attempt at control, are less complete than for *P. falciparum* (Hay *et al.*, 2009; Marsh, 2010). This may be partly due to the unrecognised clinical significance of this parasite, but also to its infections having a characteristically lower parasitaemia than *P. falciparum* infections (Baird, 2013). Diagnosing *P. vivax*, particularly in mixed-infections, requires considerably more effort.

Gething and colleagues report that many in-country surveillance systems are not attuned to *P. vivax*; between 2002 and 2010 only 53 of 95 *P. vivax* endemic countries provided *P. vivax*-specific routine case reporting data reliably distinct from their *P. falciparum* data (Gething *et al.*, 2012). Community malaria parasite rate surveys are also much less common and spatially homogenous for *P. vivax* relative to *P. falciparum*, presenting an important challenge to endemicity mapping (Gething *et al.*, 2011; Gething *et al.*, 2012).

Despite being highly genetically diverse, *P. vivax* parasites appear to be dependent on only a single antigen to successfully infect red blood cells (RBCs): the Duffy antigen (Miller *et al.*, 1976; Carlton *et al.*, 2008; Neafsey *et al.*, 2012) (Figure 1.4♠). No other invasion pathway has been described to date. This polymorphic human antigen shows distinct geographic patterns, and for instance is not commonly expressed by individuals of African origin, making them resistant to *P. vivax* infection (Miller *et al.*, 1976). These findings provided a simple solution to the apparent absence of *P. vivax* from Africa, and until recently, little further study has been conducted of the prevalence of this parasite on this continent. Knowledge of the underlying human Duffy landscape therefore represents a proxy susceptibility map of human populations to *P. vivax* infection. This information could support an improved understanding of this parasite's transmission potential at the population level and help fill this epidemiological knowledge gap.

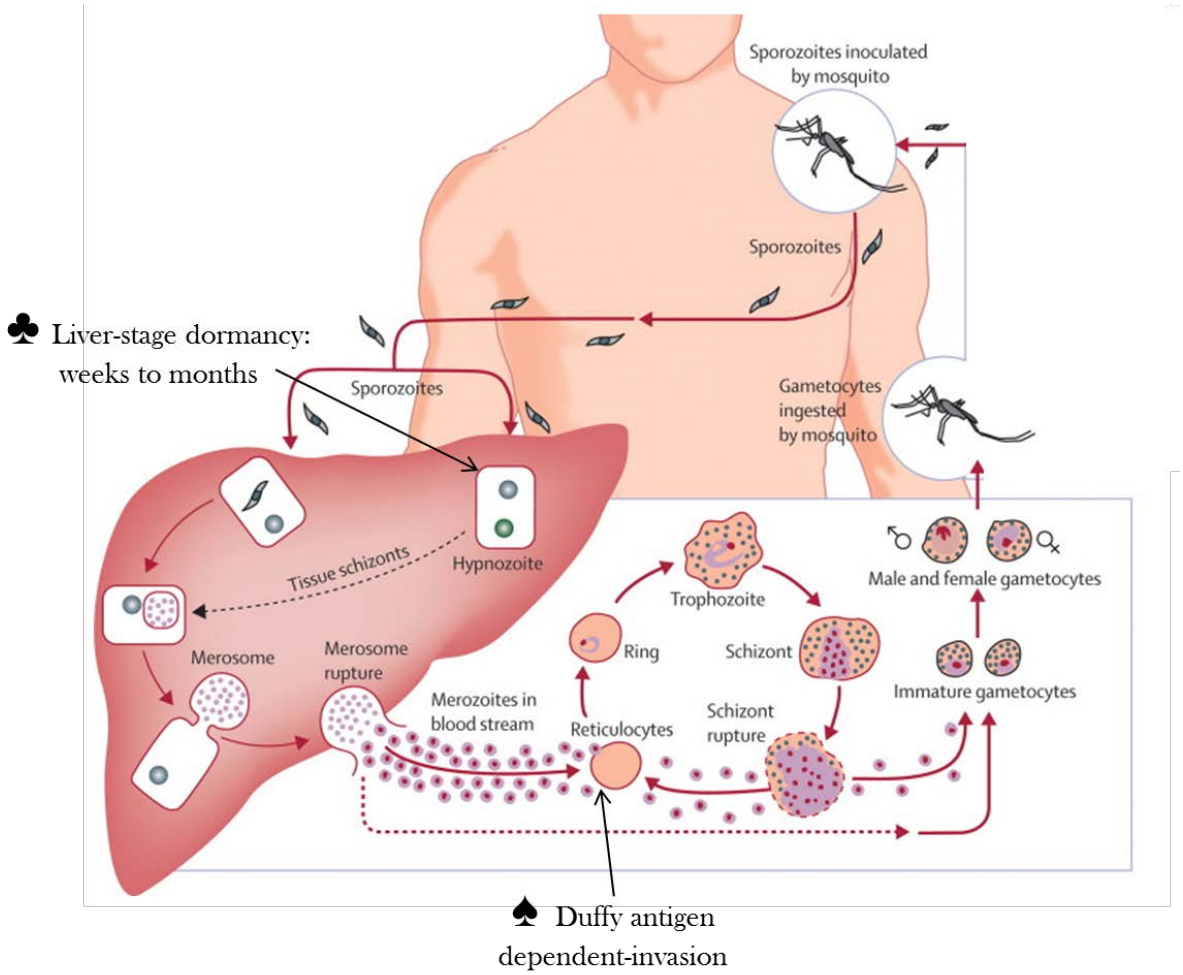


Figure 1.4. *Plasmodium vivax* lifecycle in the human host. The figure highlights two aspects of the lifecycle which differ from *P. falciparum*: ♠ indicates the dependency on the Duffy antigen for establishing blood-stage infection; ♣ indicates the dormant hypnozoites which may be reactivated weeks to months after an initial infection and cause clinical relapse. Figure adapted from Mueller *et al* (2009).

1.2.2. A poorly targeted therapeutic arsenal

The importance of distinguishing between the different *Plasmodium* parasites for malaria therapy was recognised in the 1950s. In 1952, L.T. Coggeshall warned (Coggeshall, 1952):

> "*The greatest misconceptions about the treatment of malaria… have arisen from the fact that too many considered it a single disease. Malaria is not a disease – it is a variety of diseases. …it is too commonly believed that any standard form of therapy is sufficient. This is far from true. It is essential first to know the species involved because they differ markedly in their behavior and response to different drugs.*"

This warning, however, was not heeded during the GMEP era and succeeding decades, and the most damaging impact of the neglected significance of *P. vivax* has perhaps been in its therapeutic development. The lifecycle distinction which makes the *Plasmodium falciparum* control toolkit most redundant for *P. vivax* is the latter's ability to form persistent liver-stage parasites: hypnozoites (Figure 1.4♣). Although all *Plasmodium* species have a developmental step in liver hepatocytes, *P. vivax* and *P. ovale* are unique in forming a sub-population of dormant hypnozoites which can trigger relapses of infection anytime between a few weeks and years after the initial infection (Battle *et al.*, 2011; White, 2011). Population-level malaria elimination requires this silent reservoir of parasites to be successfully treated, so as to prevent continuous relapse of clinical cases and infected mosquitoes (Baird, 2010; Wells *et al.*, 2010). At present, these silent parasites remain diagnostically invisible (The malERA Consultative Group on Diagnoses and Diagnostics, 2011).

Pharmaceutical developments during the 20th century were scant for *P. vivax* radical cure. The current portfolio of antimalarial drugs mostly targets the blood stages of *P. falciparum* and are not active against *P. vivax* hypnozoites (Baird, 2010). A single hypnozoiticide, primaquine, has been in use since the 1950s, and only one other drug is currently in development, tafenoquine (GSK-MMV; Phase IIb/III trials) (Medicines for Malaria Venture; The malERA Consultative Group on Drugs, 2011; John *et al.*, 2012). Both these drugs are 8-aminoquinolines and highly oxidative (Brueckner *et al.*, 2001). Individuals with a glucose-6-phosphate dehydrogenase deficiency (G6PDd), a congenital enzyme disorder, are at risk of life-threatening haemolysis from this family of drugs; no point-of-care diagnostics exist for this predisposing condition (Kim *et al.*, 2011; The malERA Consultative Group on Diagnoses and Diagnostics, 2011).

The demand for hypnozoiticidal therapy is broader than solely against symptomatic *P. vivax* infections. Strong evidence points to silent *P. vivax* infections being much more prevalent than the numbers of symptomatic infections alone would suggest. A review of 10,549 naturally acquired acute *P. falciparum* infections in Thailand treated with standard

blood-stage therapy found that within a 63 day follow-up period in an area of low endemicity (where chances of re-infection were low), 51% had suffered an attack of *P. vivax* (Douglas *et al.*, 2011), suggesting a high prevalence of undiagnosed *P. vivax* infection. It has been argued that *P. vivax* hypnozoite therapy ought be given in all cases of malaria, regardless of the diagnosed pathogen (Baird, 2011). The aforementioned low parasitaemia associated with *P. vivax* adds further difficulty to diagnosing this species: a study in the Solomon Islands found that fewer than 30% of PCR-diagnosed blood infections could be detected by expert microscopists (Harris *et al.*, 2010). The threat which asymptomatic, sub-microscopic, latent *P. vivax* infection poses to elimination efforts demands widespread use of radical cure treatment. Furthermore, primaquine is the only drug active against the infectious gametocyte lifestages of all *Plasmodium* parasites which infect mosquitoes and ensure onward transmission. The use of primaquine for transmission-blocking has great potential as a tool to reduce endemicity, and is thus a key tool for elimination (Bousema and Drakeley, 2011; White, 2012).

The fear of triggering potentially severe haemolysis prevents widespread use of the only drug available for *P. vivax* radical cure and *P. falciparum* transmission blocking. G6PD deficiency therefore represents a major hurdle towards the current resolve for malaria elimination. A better knowledge of the distribution, prevalence, genetic diversity and haemolytic risks associated with this enzyme deficiency among the populations where primaquine is needed would be an important evidence-base for maximising safe use of this unique drug.

1.3. The role of maps

An important lesson from the history of malaria control is that one size does not fit all. Interventions developed against one parasite species are not effective against all others, and similarly the effectiveness of control measures varies between regions. The epidemiology of disease – and of vector-borne parasites in particular – is determined by a range of factors,

many of which are spatially variable such as environmental characteristics. Spatial epidemiology has been defined to be:

> "*the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioural, socioeconomic, genetic, and infectious risk factors"* (Elliott and Wartenberg, 2004)

In a time of financial austerity, appropriately targeting resources according to need and intervention suitability is essential. This underpins the "*investing for impact*" motto of major funding agencies (The Global Fund to Fight AIDS Tuberculosis and Malaria; Pigott *et al.*, 2012). Spatial epidemiology and geographic information systems (GIS) are tools which allow the visualisation of spatially variable datasets, as well as being powerful analytical platforms bringing together disparate datasets into single frameworks (Gordon and Womersley, 1997). The application of these spatial approaches to the two knowledge gaps previously described in relation to improving our evidence-base of *P. vivax* transmission limits and characterising the risks associated with primaquine therapy are evident. *P. vivax* transmission is determined by a range of biological, environmental and anthropological factors (such as spread of urbanisation), as well as the human genetic landscape determining susceptibility to infection (Hay *et al.*, 2004; Guerra *et al.*, 2010). Attempts to enumerate the public health significance of *P. vivax* must be considered in a spatial framework which reflects this parasite's transmission heterogeneity. Similarly, targeting the distribution of appropriate diagnostics and drugs requires prior knowledge of the local epidemiology of malaria in affected populations.

1.4. Aims of the thesis and description of chapters

As discussed, the neglect of *P. vivax* during the 20th century trails a legacy of knowledge gaps and poorly targeted tools including diagnostics, surveillance systems and treatments. This thesis attempts to address two important aspects of this neglect by providing policy-makers with evidence-based information to target their public health control efforts: first,

by contributing towards establishing the transmission potential of *P. vivax* and the population at risk of infection, and second, by providing a framework to support decisions regarding the use of primaquine therapy against *P. vivax* relapsing infections. Recognising the spatially heterogeneous nature of these aspects of malaria, I address both these issues from a spatial perspective.

The specific objectives of this thesis are:

1. To map the frequencies of the human Duffy blood group variants and establish the population at risk of *P. vivax* infection.
2. To map the prevalence and diversity of G6PD deficiency and quantify the population susceptible to primaquine-induced haemolysis.

Existing maps of these spatially variable human polymorphisms are poor, as discussed in detail through this thesis. I aim to use state of the art mapping methods to synthesise the diverse array of available data into mapping frameworks which quantify their uncertainty. Knowledge of the distribution of the population at risk of *P. vivax* infection and of the population susceptible to primaquine-induced haemolysis would facilitate appropriate diagnostic and drug deployment to the right regions. Maps also provide an overview of the comprehensiveness and robustness of current knowledge, and an identification of where additional data are most needed.

Chapter 2 of this thesis describes the methodological steps involved with mapping the Duffy antigen alleles (*FY*A*, *FY*B*, *FY*B*^{ES}). The application of these maps to estimating the population at risk of *P. vivax* infection, together with an overview of how they support the study of *P. vivax* in Africa specifically, is discussed in Chapter 3. A description of the methodological steps involved in modelling the prevalence of the G6PD deficient phenotype, together with estimates of the population affected by this disorder, is given in Chapter 4. Maps of the common genetic variants of G6PD deficiency are presented in Chapter 5. Next, Chapter 6 considers the implications of these two suites of G6PD deficiency maps on public health haemolytic risks associated with administering primaquine for *P. vivax* radical cure, and the important limitations to quantifying this risk

are discussed. Finally, Chapter 7 is a general discussion of the overall conclusions of this research, its strengths and limitations, as well as its implications for further study.

The two core chapters of this thesis, Chapters 2 and 4, are included here in their published form. Due to the space constraints associated with publishing these substantial research outputs, extensive additional information about all aspects of these studies is included here in the Appendix. These are co-authored publications and each author's contributions are detailed at the end of these chapters.

1.5. References

Alonso, P.L., Brown, G., Arevalo-Herrera, M., et al. (2011). A research agenda to underpin malaria eradication. *PLoS Medicine* **8**(1): e1000406.

Anstey, N.M., Russell, B., Yeo, T.W., et al. (2009). The pathophysiology of vivax malaria. *Trends in Parasitology* **25**(5): 220-227.

APMEN. Asia Pacific Malaria Elimination Network. Accessed: 23 Nov 2012. URL: http://apmen.org/.

Atmosoedjono, S., Arbani, P.R. and Bangs, M.J. (1992). The use of species sanitation and insecticides for malaria control in coastal areas of Java. *Buletin Penelitian Kesehatan* **20**(3): 1-15.

Baird, J.K. (2007). Neglect of *Plasmodium vivax* malaria. *Trends in Parasitology* **23**(11): 533-539.

Baird, J.K. (2010). Eliminating malaria--all of them. *Lancet* **376**(9756): 1883-1885.

Baird, J.K. (2011). Radical cure: the case for anti-relapse therapy against all malarias. *Clinical Infectious Diseases* **52**(5): 621-623.

Baird, J.K. (2013). Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. *Clinical Microbiology Reviews* **26**(1): 1-22.

Battle, K.E., Van Boeckel, T.P., Gething, P.W., et al. (2011). A review of the geographical variation in *Plasmodium vivax* relapse rate. *American Journal of Tropical Medicine and Hygiene* **85**(Suppl 451-A-516): S470.

Bill & Melinda Gates Foundation. Bill & Melinda Gates Foundation. Accessed: 22 Nov 2012. URL: http://www.gatesfoundation.org/Pages/home.aspx.

Bill and Melinda Gates Foundation Malaria Forum Day 2, 17 October 2007 [Transcript].

Bockarie, M.J. and Dagoro, H. (2006). Are insecticide-treated bednets more protective against *Plasmodium falciparum* than *Plasmodium vivax*-infected mosquitoes? *Malaria Journal* **5**: 15.

Bousema, T. and Drakeley, C. (2011). Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. *Clinical Microbiology Reviews* **24**(2): 377-410.

Bradley, D.J. (1992). Malaria: old infections, changing epidemiology. *Health Transition Review* **2**(suppl): 137-152.

Brueckner, R.P., Ohrt, C., Baird, J.K., et al. (2001). 8-Aminoquinolines. In: Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery. P.J. Rosenthal (eds). Totowa, NJ, Humana Press.

Bynum, W.F. (1999). Ronald Ross and the malaria-mosquito cycle. *Parassitologia* **41**(1-3): 49-52.

Capanna, E. (2006). Grassi versus Ross: who solved the riddle of malaria? *International Microbiology* **9**(1): 69-74.

Carlton, J.M., Adams, J.H., Silva, J.C., et al. (2008). Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**(7214): 757-763.

Carlton, J.M., Sina, B.J. and Adams, J.H. (2011). Why is *Plasmodium vivax* a neglected tropical disease? *PLoS Neglected Tropical Diseases* **5**(6): e1160.

Carter, R. and Mendis, K.N. (2002). Evolutionary and historical aspects of the burden of malaria. *Clinical Microbiology Reviews* **15**(4): 564-594.

Christophers, S.R. (1924). What disease costs India. *Indian Medical Gazette* **59**: 196-200.

Coatney, G.R., Cooper, W.C. and Ruhe, D.S. (1948). Studies in human malaria; the organization of a program for testing potential antimalarial drugs in prisoner volunteers. *American Journal of Hygiene* **47**(1): 113-119.

Coggeshall, L.T. (1952). The treatment of malaria. *American Journal of Tropical Medicine and Hygiene* **1**(1): 124-131.

Condon-Rall, M.E. (1992). U.S. Army medical preparations and the outbreak of war: the Philippines, 1941-6 May 1942. *The Journal of Military History* **56**: 35-56.

Cox-Singh, J., Davis, T.M., Lee, K.S., et al. (2008). *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clinical Infectious Diseases* **46**(2): 165-171.

De Zulueta, J. (1998). The end of malaria in Europe: an eradication of the disease by control measures. *Parassitologia* **40**: 245-246.

Douglas, N.M., Nosten, F., Ashley, E.A., et al. (2011). *Plasmodium vivax* recurrence following falciparum and mixed species malaria: risk factors and effect of antimalarial kinetics. *Clinical Infectious Diseases* **52**(5): 612-620.

Downs, W.G., Harper, P.A. and Lisansky, E.T. (1947). Malaria and other insect-borne diseases in the South Pacific campaign. 1942-1945. II. Epidemiology of insect-borne diseases in Army troops. *American Journal of Tropical Medicine* **27**: 69-89.

Elliott, P. and Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives* **112**(9): 998-1006.

Eziefula, A.C., Gosling, R., Hwang, J., et al. (2012). Rationale for short course primaquine in Africa to interrupt malaria transmission. *Malaria Journal* **11**: 360.

Fantini, B. (1999). The concept of specificity and the Italian contribution to the discovery of the malaria transmission cycle. *Parassitologia* **41**(1-3): 39-47.

Feachem, R.G., Phillips, A.A., Hwang, J., et al. (2010). Shrinking the malaria map: progress and prospects. *Lancet* **376**(9752): 1566-1578.

Feachem, R.G. and The Malaria Elimination Group (2009). Shrinking the Malaria Map: A Guide on Malaria Elimination for Policy Makers. San Francisco. The Global Health Group, University of California, San Francisco.

Fong, T.C.C. (1937). A study of the mortality rate and complications following therapeutic malaria. *Southern Med J* **30**: 1084-1088.

Galinski, M.R. and Barnwell, J.W. (2008). *Plasmodium vivax*: who cares? *Malaria Journal* **7 Suppl 1**: S9.

Gallup, J.L. and Sachs, J.D. (2001). The economic burden of malaria. *American Journal of Tropical Medicine and Hygiene* **64**(1-2 Suppl): 85-96.

Gething, P.W., Elyazar, I.R., Moyes, C.L., et al. (2012). A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Neglected Tropical Diseases* **6**(9): e1814.

Gething, P.W., Patil, A.P., Smith, D.L., et al. (2011). A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal* **10**: 378.

Gething, P.W., Smith, D.L., Patil, A.P., et al. (2010). Climate change and the global malaria recession. *Nature* **465**(7296): 342-345.

Gordon, A. and Womersley, J. (1997). The use of mapping in public health and planning health services. *Journal of Public Health Medicine* **19**(2): 139-147.

Guerra, C.A., Howes, R.E., Patil, A.P., et al. (2010). The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Neglected Tropical Diseases* **4**(8): e774.

Harris, I., Sharrock, W.W., Bain, L.M., et al. (2010). A large proportion of asymptomatic *Plasmodium* infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. *Malaria Journal* **9**: 254.

Harrison, G. (1978). Mosquitoes, Malaria and Man: a History of Hostilities since 1880. London, John Murray.

Hay, S.I., Guerra, C.A., Gething, P.W., et al. (2009). A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Medicine* **6**(3): e1000048.

Hay, S.I., Guerra, C.A., Tatem, A.J., et al. (2004). The global distribution and population at risk of malaria: past, present, and future. *Lancet Infectious Diseases* **4**(6): 327-336.

Hsiang, M.S., Abeyasinghe, R., Whittaker, M., et al. (2010). Malaria elimination in Asia-Pacific: an under-told story. *Lancet* **375**(9726): 1586-1587.

John, G.K., Douglas, N.M., von Seidlein, L., et al. (2012). Primaquine radical cure of *Plasmodium vivax*: a critical review of the literature. *Malaria Journal* **11**: 280.

Kim, S., Nguon, C., Guillard, B., et al. (2011). Performance of the CareStart G6PD deficiency screening test, a point-of-care diagnostic for primaquine therapy screening. *PLoS One* **6**(12): e28357.

Kitron, U. (1987). Malaria, agriculture, and development: lessons from past campaigns. *International Journal of Health Services* **17**(2): 295-326.

Kochar, D.K., Das, A., Kochar, S.K., et al. (2009). Severe *Plasmodium vivax* malaria: a report on serial cases from Bikaner in northwestern India. *American Journal of Tropical Medicine and Hygiene* **80**(2): 194-198.

Litsios, S. (1966). The Tomorrow of Malaria. Wellington, Pacific Press.

Litsios, S. (2002). Malaria control and the future of international public health. In: The Contextual Determinants of Malaria. E. Casman and H. Dowlatabadi (eds). Washington DC, Resources for the Future: 292-328.

Lysenko, A.J. and Semashko, I.N. (1968). Geography of malaria. A medico-geographic profile of an ancient disease. In: Itogi Nauki: Medicinskaja Geografija. A.W. Lebedew, Academy of Sciences, USSR, Moscow: 25-146.

Marsh, K. (2010). Research priorities for malaria elimination. *Lancet* **376**(9753): 1626-1627.

MDG. UN Millennium Development Goals. Accessed: 20 Nov 2012. URL: http://www.un.org/millenniumgoals/.

Medicines for Malaria Venture. MMV Research & Development. Accessed: 26 Nov 2012. URL: http://www.mmv.org/research-development.

Miller, L.H., Mason, S.J., Clyde, D.F., et al. (1976). The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *New England Journal of Medicine* **295**(6): 302-304.

Ministry of Health Sri Lanka and the World Health Organization and the University of California-San Francisco (2012). Eliminating Malaria: Case-study 3 | Progress towards elimination in Sri Lanka. Geneva. The World Health Organization.

Mueller, I., Galinski, M.R., Baird, J.K., et al. (2009). Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infectious Diseases* **9**(9): 555-566.

Murray, C.J., Rosenfeld, L.C., Lim, S.S., et al. (2012). Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet* **379**(9814): 413-431.

Najera, J.A. (2001). Malaria control: achievements, problems and strategies. *Parassitologia* **43**(1-2): 1-89.

Najera, J.A., Gonzalez-Silva, M. and Alonso, P.L. (2011). Some lessons for the future from the Global Malaria Eradication Programme (1955-1969). *PLoS Medicine* **8**(1): e1000412.

Neafsey, D.E., Galinsky, K., Jiang, R.H., et al. (2012). The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nature Genetics* **44**(9): 1046-1050.

PATH (2011). Staying the course? Malaria research and development in a time of economic uncertainty. Seattle. PATH.

Phimpraphi, W., Paul, R.E., Yimsamran, S., et al. (2008). Longitudinal study of *Plasmodium falciparum* and *Plasmodium vivax* in a Karen population in Thailand. *Malaria Journal* **7**: 99.

Pigott, D.M., Atun, R., Moyes, C.L., et al. (2012). Funding for malaria control 2006-2010: a comprehensive global assessment. *Malaria Journal* **11**: 246.

President's Malaria Initiative. President's Malaria Initiative. Accessed: 22 Nov 2012. URL: http://www.pmi.gov/.

Price, R.N., Douglas, N.M. and Anstey, N.M. (2009). New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. *Current Opinion in Infectious Diseases* **22**(5): 430-435.

Roberts, L. and Enserink, M. (2007). Malaria. Did they really say ... eradication? *Science* **318**(5856): 1544-1545.

Ross, R. (1898). The rôle of the mosquito in the evolution of the malarial parasite. *Lancet* **152**(3912): 488-490.

Ross, R. (1923). Memoires. London, John Murray.

Sachs, J.D. (2002). A new global effort to control malaria. *Science* **298**(5591): 122-124.
Singh, H., Parakh, A., Basu, S., et al. (2011). *Plasmodium vivax* malaria: is it actually benign? *Journal of Infection and Public Health* **4**(2): 91-95.
Snounou, G. and Perignon, J.-L. (2013). Malariatherapy - Insanity at the service of malariology. *Advances in Parasitology* **81**: in press.
Snow, R.W., Guerra, C.A., Mutheu, J.J., et al. (2008). International funding for malaria control in relation to populations at risk of stable *Plasmodium falciparum* transmission. *PLoS Medicine* **5**(7): e142.
Snow, R.W. and Marsh, K. (2010). Malaria in Africa: progress and prospects in the decade since the Abuja Declaration. *Lancet* **376**(9735): 137-139.
The Abuja Declaration and the Plan of Action. The African Summit on Roll Back Malaria, Abuja, 25 April 2000 (WHO/CDS/RBM/2000.17). Accessed: 22 Nov 2012. URL: http://www.rbm.who.int/docs/abuja_declaration_final.htm.
The Global Fund to Fight AIDS Tuberculosis and Malaria. The Global Fund to Fight AIDS, Tuberculosis and Malaria. Accessed: 22 Nov 2012. URL: http://www.theglobalfund.org/en/.
The malERA Consultative Group on Diagnoses and Diagnostics (2011). A Research Agenda for Malaria Eradication: Diagnoses and Diagnostics. *PLoS Medicine* **8**(1): e1000396.
The malERA Consultative Group on Drugs (2011). A Research Agenda for Malaria Eradication: Drugs. *PLoS Medicine* **8**(1): e1000402.
Tjitra, E., Anstey, N.M., Sugiarto, P., et al. (2008). Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. *PLoS Medicine* **5**(6): e128.
UCSF Global Health Group and Malaria Atlas Project (2011). Atlas of Malaria-Eliminating Countries. San Francisco. Unversity of California.
Wells, T.N., Burrows, J.N. and Baird, J.K. (2010). Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination. *Trends in Parasitology* **26**(3): 145-151.
White, N.J. (2011). Determinants of relapse periodicity in *Plasmodium vivax* malaria. *Malaria Journal* **10**: 297.
White, N.J. (2012). Primaquine to prevent transmission of falciparum malaria. *Lancet Infectious Diseases*: doi:10.1016/S1473-3099(1012)70198-70196.
WHO (1955). Eighth World Health Assembly (Mexico, D.F., 10-27 May 1955). Official records of the World Health Organization, N^{o} 63. 236-240. Geneva. World Health Organization.
WHO (1973). Malaria. Handbook of resolutions and decisions of the World Health Assembly and the Executive Board. Volume 1, 1948-1972, 1st to 25th WHA and 1st to 50th EB. 66-81. Geneva. World Health Organization.
WHO (2010). World Malaria Report 2010. Geneva. World Health Organization.
WHO (2011). World Malaria Report 2011. Geneva. World Health Organization.

Chapter 2 – The global distribution of the Duffy blood group

This first research chapter describes the assembly of the Duffy blood group variant maps. This work has been published in *Nature Communications* and is included here in its final form. The additional information referred to in this chapter is included in the Appendix of this thesis. The application of these maps to understanding the spatial characteristics of *P. vivax* transmission is discussed in Chapter 3.

nature COMMUNICATIONS

OPEN

ARTICLE

Received 9 Aug 2010 | Accepted 3 Mar 2011 | Published 5 Apr 2011

DOI: 10.1038/ncomms1265

The global distribution of the Duffy blood group

Rosalind E. Howes[1], Anand P. Patil[1], Frédéric B. Piel[1], Oscar A. Nyangiri[2], Caroline W. Kabaria[3], Peter W. Gething[1], Peter A. Zimmerman[4], Céline Barnadas[5], Cynthia M. Beall[6], Amha Gebremedhin[7], Didier Ménard[8], Thomas N. Williams[2], David J. Weatherall[9] & Simon I. Hay[1]

Blood group variants are characteristic of population groups, and can show conspicuous geographic patterns. Interest in the global prevalence of the Duffy blood group variants is multidisciplinary, but of particular importance to malariologists due to the resistance generally conferred by the Duffy-negative phenotype against *Plasmodium vivax* infection. Here we collate an extensive geo-database of surveys, forming the evidence-base for a multi-locus Bayesian geostatistical model to generate global frequency maps of the common Duffy alleles to refine the global cartography of the common Duffy variants. We show that the most prevalent allele globally was *FY*A*, while across sub-Saharan Africa the predominant allele was the silent $FY*B^{ES}$ variant, commonly reaching fixation across stretches of the continent. The maps presented not only represent the first spatially and genetically comprehensive description of variation at this locus, but also constitute an advance towards understanding the transmission patterns of the neglected *P. vivax* malaria parasite.

[1] Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [2] Kenya Medical Research Institute (KEMRI)/Wellcome Trust Programme, Centre for Geographic Medicine Research, Coast, PO Box 230, Kilifi District Hospital, Kilifi 80108, Kenya. [3] Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, KEMRI—University of Oxford—Wellcome Trust Collaborative Programme, Kenyatta National Hospital Grounds (behind NASCOP), PO Box 43640-00100, Nairobi, Kenya. [4] Center for Global Health and Diseases, Case Western Reserve University, Wolstein Research Building, 2103 Cornell Road, Cleveland, Ohio 44106-7286, USA. [5] Vector Borne Diseases Unit, Papua New Guinea Institute for Medical Research, PO BOX 60, Goroka, EHP 441, Papua New Guinea. [6] Anthropology Department, Case Western Reserve University, 238 Mather Memorial Building, 11220 Bellflower Road, Cleveland, Ohio 44106-7125, USA. [7] Department of Internal Medicine, PO Box 14227, Faculty of Medicine, Addis Ababa University, Addis Ababa, Ethiopia. [8] Molecular Epidemiology Unit, Pasteur Institute of Cambodia, 5 Boulevard Monivong, PO Box 983, Phnom Penh, Cambodia. [9] Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK. Correspondence and requests for materials should be addressed to S.I.H. (email: simon.hay@zoo.ox.ac.uk).

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms1265

First described 60 years ago in a multiply transfused haemolytic patient who lent his name to the system[1], the Duffy blood group has since been of interest in diverse fields from anthropology[2] to genetics[3] and malariology[4,5]. Being of only occasional clinical significance[6], much of the research into this weakly immunogenic blood group has been concerned with establishing characteristic expression patterns among populations. Easily diagnosable, the Duffy blood group quickly became part of the package of commonly investigated blood groups used to characterize the world's populations[7] and assess relatedness between communities. Interest in the Duffy blood group rose substantially, however, following experimental demonstration of the malaria parasite *Plasmodium vivax*'s dependency on the Duffy antigen for establishing erythrocytic infection[8–10], and therefore that erythrocytes lacking the antigen were refractory to this parasitic infection. This Duffy negativity phenotype, long known to be common among sub-Saharan African populations[11], provided an explanation for the apparent absence of *P. vivax* among these populations and their diaspora[12]. To date, no other erythrocyte receptor has been described for *P. vivax*, although some cases of infection have been reported in Duffy-negative individuals[13–15]. Furthermore, the universal expression of the Duffy antigen binding protein (*Pv*DBP) has made this merozoite invasion ligand protein a prime *P. vivax* vaccine target[16]. The role of the Duffy receptor in *P. vivax* infection, therefore, allows the Duffy-negative phenotype to be a proxy of host resistance to blood stage infection[17].

Recognizing its physiological function as a chemokine receptor involved in inflammation, the Duffy antigen is also known as the Duffy antigen receptor for chemokines (DARC). Although specific mechanisms underlying its functions remain uncertain, there is interest in DARC as an explanatory variable for population-specific differences in disease susceptibility[18], as demonstrated by ongoing research into its role in inflammation-associated pathology and malignancy[18,19], and the recent, though highly controversial[20], surge in interest around the antigen's role in HIV infection[21].

The monogenic Duffy system was the first human blood group assigned to a specific autosome: position q21–q25 on chromosome 1 (ref. 22). The gene product has two main variant forms: Fy^a and Fy^b antigens, which differ by a single amino acid (Gly42Asp), encoded by alleles *FY*A* and *FY*B*, which are differentiated by a single base substitution (G125A)[23,24]. Duffy expression is disrupted by a T to C substitution in the gene's promoter region at nucleotide −33, preventing transcription and resulting in the null 'erythrocyte silent' (ES) phenotype. This promoter-region variant is commonly haplotypically associated with the *FY*B* coding region (corresponding to the $FY*B^{ES}$ allele)[24], although occasional reports of association with the *FY*A* sequence have been published ($FY*A^{ES}$ allele)[25,26]. These four alleles combine to ten possible genotypes (Fig. 1), with *FY*A* and *FY*B* alleles expressed codominantly over the null variants $FY*B^{ES}$ and $FY*A^{ES}$. Genotypes therefore correspond to four phenotypes: Fy(a+b+), Fy(a+b−), Fy(a−b+) and Fy(a−b−). Further details about the genetics and molecular aspects of the Duffy system, including other rare variants such as the weakly expressed *FY*X* allele[27] (expressed as the $Fy(b+^{weak})$ phenotype), are fully discussed by Langhi and Bordin[23] and Zimmerman[24].

The common Duffy alleles present striking patterns of geographic differentiation, which have only once been mapped spatially, as part of Cavalli-Sforza *et al.*'s[28] efforts to unravel the genetic

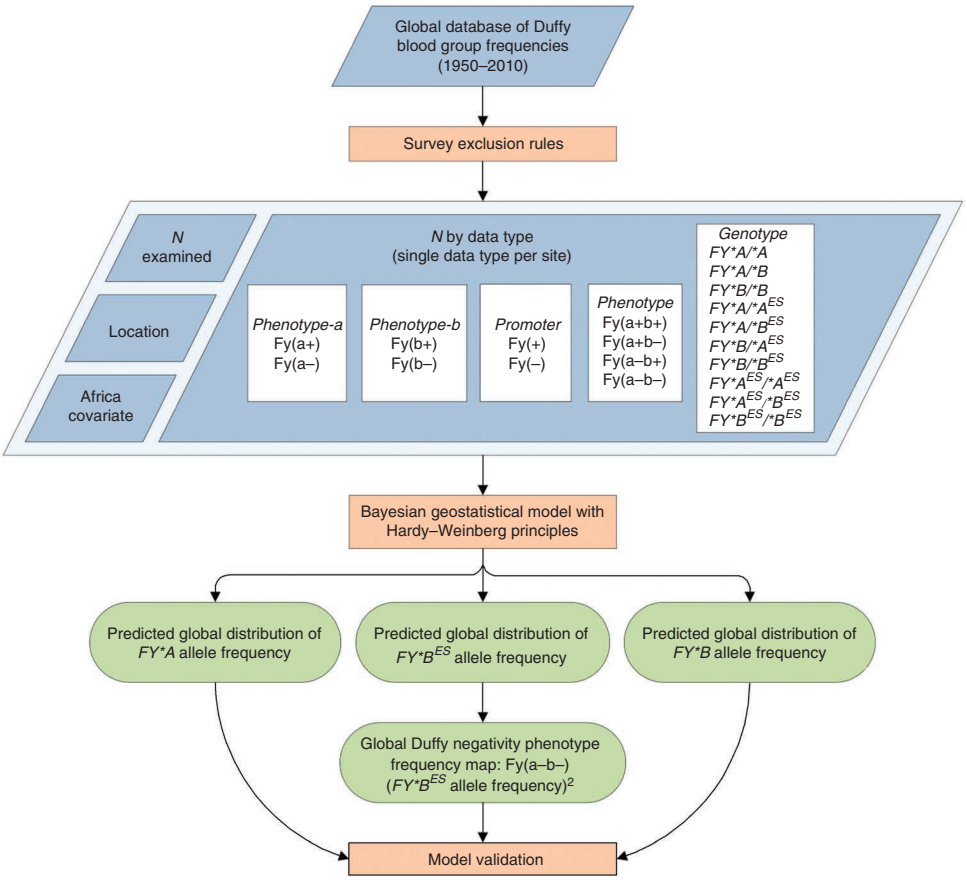


Figure 1 | Schematic overview of the procedures and methods. Blue diamonds describe input data. White boxes within the '*N* by data type' diamond represent different possible data types, with each spatially unique survey being represented by only one white box. Orange boxes denote models and experimental procedures. Green rods indicate model outputs.

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms1265

history of human populations. Alleles were mapped in isolation of each other (*FY*A*, *FY*B* and *FY*B*^{ES}) using limited subsets of data which directly informed the frequency of each particular variant. Previous methodology, which used a sequential inverse distance weighting algorithm, was unable to estimate uncertainty in the mapped predictions. Since this publication in 1994, new data have been generated, much of which have benefitted from full genotyping. In addition, significant advances have been made to geostatistical mapping techniques allowing more rigorous predictions of spatially continuous variables using data obtained at a limited number of spatial locations[29], and varied data inputs to inform all output predictions simultaneously.

Here we first assembled an updated data set based on a thorough review of published and unpublished data, then used this to generate global maps of the Duffy alleles and the Duffy negativity phenotype using a bespoke Bayesian geostatistical model. This generated for every location on a gridded surface a posterior distribution of all predicted values from the model's thousands of iterations, representing a complete model of uncertainty from which median values were derived to generate point-estimate maps. In addition to providing fundamental biomedical descriptions of human populations with potential applications for explaining population-level variation to a range of clinical conditions[19], these maps are intended to support contemporary analyses of *P. vivax* transmission risk[17].

Results

The survey database. Literature searches identified 821 spatially unique data points of Duffy blood type prevalence matching the database inclusion criteria for representativeness (Supplementary Data). These represented a total of 131,187 individuals sampled, 17.8% of whom were surveyed on the African continent (Table 1). Of this total, 536 surveys were geopositioned as points (≤25 km^2), and 285 mapped as polygon centroids. Polygons were digitized and centroid coordinates calculated in GIS software (ArcView GIS 3.2 and ArcMap 9.3, ESRI). Surveys reported only to province or country level were considered to lack sufficient geographical specificity and were thus excluded. A total of 89 additional data points were excluded, as they could not be located with sufficient precision. The selected data points were relatively evenly distributed between regions, with 32% in the Americas, 25% in Africa, 26% in Asia and 17% in Europe (Fig. 2). Survey sample sizes were highly variable: ranging from 1 to 2,470. The mean size was 160 and the median 99.

Serological techniques were the only methods used for blood typing until the 1990s, (Supplementary Fig. S1). Half of the surveys (49%) used anti-Fy^a antiserum only, thus were recorded as '*Phenotype-a*' data types (402 of 821 surveys). Complete '*Phenotype*' data were provided in 247 surveys (30%). Molecularly diagnosed '*Genotype*' and '*Promoter*' data (9 and 12%, respectively) were only commonly reported in post-2000 surveys, and mainly across Africa (71% of the 168 DNA-based records were from Africa). The five categories of data types are summarized in Table 2, and the spatial distribution of each is represented by the colour-coded data point map in Figure 2. Relative proportions of each variant-type reported by data type and continent are displayed in Table 1; further summaries of the data are presented graphically by decade in Supplementary Figures S1–S2.

The maps. To generate continuous global maps from the assembled database, a Duffy-specific geostatistical model was developed (Supplementary Methods). Its key features are as follows: first, to incorporate all genotypic variants and data types simultaneously; second, to predict any genotype or phenotype frequency desired; third, to allow for local heterogeneity; and fourth, to take into account sampling error through sample size, while also generating uncertainty estimates with the prediction at each spatial unit (pixel).

Table 1 | Summary of input data.

Total individuals sampled	Africa	Americas	Asia	Europe	World
	23,349 17.8%	**37,410** 28.5%	**32,971** 25.1%	**37,457** 28.6%	**131,187**
Genotype	**1,720**	**1,107**	**5,993**	**336**	**9,156**
*FY*A/*A*	40	183	5,805	42	6,070
*FY*A/*B*	88	341	24	174	627
*FY*B/*B*	49	217	6	117	389
*FY*A/*A*^{ES}	0	0	157	0	157
*FY*A/*B*^{ES}	226	122	0	0	348
*FY*B/*A*^{ES}	0	0	0	0	0
*FY*B/*B*^{ES}	190	122	1	1	314
*FY*A*^{ES}*/*A*^{ES}	0	0	0	0	0
*FY*A*^{ES}*/*B*^{ES}	4	0	0	0	4
*FY*B*^{ES}*/*B*^{ES}	1,123	122	0	2	1,247
Phenotype	**11,370**	**10,939**	**11,143**	**17,126**	**50,578**
Fy(a+b+)	1,448	3,841	2,471	7,355	15,115
Fy(a+b−)	1,338	3,904	6,821	4,872	16,935
Fy(a−b+)	2,493	2,595	1,493	4,873	11,454
Fy(a−b−)	6,091	599	358	26	7,074
Promoter	**7,290**	**803**	**187**	**104**	**8,384**
Fy(+)	7,266	406	0	0	7,672
Fy(−)	24	397	187	104	712
Phenotype-a	**2,821**	**24,420**	**15,648**	**17,891**	**60,780**
Fy(a+)	589	19,950	12,436	11,616	44,591
Fy(a−)	2,232	4,470	3,212	6,275	16,189
Phenotype-b	**148**	**141**	**0**	**2,000**	**2,289**
Fy(b+)	1	55	0	1,651	1,707
Fy(b−)	147	86	0	349	582
Total sites surveyed	**203** 24.7%	**265** 32.3%	**217** 26.4%	**136** 16.6%	**821**

The table shows the total number of individuals sampled by continent, broken down by variant within each data type category. Totals are shown in bold. The number of spatially unique sites in each continent is given in the bottom row.

Allele frequencies. Continuous global frequency maps of each of the three common Duffy alleles (*FY*A*, *FY*B* and *FY*B*^{ES}) were generated simultaneously, along with summaries of uncertainty in the predictions quantified by the 50% interquartile range (IQR; Fig. 3). Full statistical summaries of the model parameters at each locus are provided in Supplementary Table S1. The silent *FY*A*^{ES} allele could not be modelled spatially due to its rarity (see Methods).

The allele frequency maps reveal strong geographic patterns, the most conspicuously focal being the distribution of the silent *FY*B*^{ES} allele across sub-Saharan Africa. Allelic frequencies across 30 countries in this region are characterized by >90% *FY*B*^{ES} and frequencies of 0–5% for *FY*A* and *FY*B* (Fig. 3a–c). Frequencies indicate fixation (that is, frequencies of 100% (ref. 30)) in parts of west, central and east Africa, suggesting total refractoriness of the local population to *P. vivax* infection[3,30]. The *FY*B*^{ES} allele, however, is not confined to the mainland African sub-continent, with frequencies predicted above 80% across Madagascar and above 50% through the Arabian Peninsula (Fig. 3c). Low allelic frequencies have also spread into the Americas, notably along the Atlantic coast and in the Caribbean. Median frequencies of 5–20% are predicted across India and up to 11% in South-East Asia.

Allelic heterogeneity is greatest in the Americas, with all three alleles predicted as being present and with only localized patches of predominance of single alleles. The *FY*A* allele (Fig. 3a) is close to fixation across pockets of eastern Asia, and remains high with

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms1265

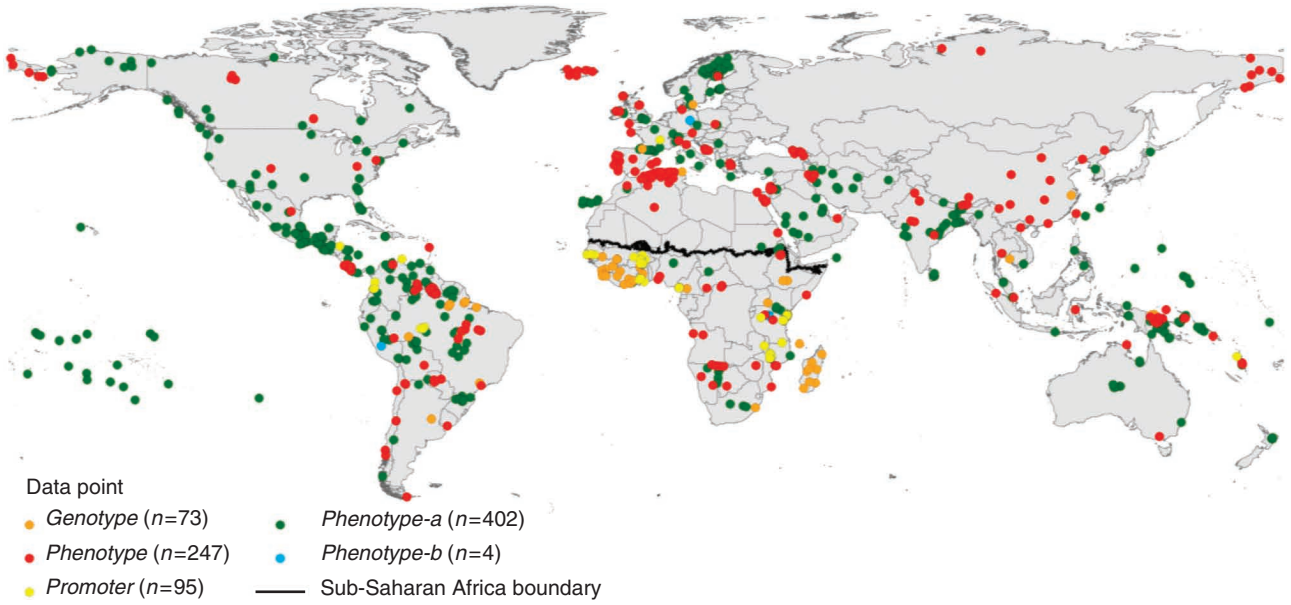


Figure 2 | Spatial distribution of the input data points categorized by data type. Symbol colours represent the type of information in the survey: orange when full genotypes were detected (*Genotype*); red for full phenotype diagnosis (*Phenotype*); yellow for expression/non-expression of Duffy antigen (*Promoter*); green and blue for partial phenotypic data, about expression of Fy^a (*Phenotype-a*) and Fy^b (*Phenotype-b*) respectively. Total data points are $n = 821$; totals by data type are listed in the legend. The sub-Saharan Africa covariate boundary is shown in black.

Table 2 | Diagnostic methods and corresponding classification data type categories.

Diagnostic method	Diagnostic type	Data type	Description	Information given	Homo/heterozygote status
Serological	Phenotype	*Phenotype*	Study tested for Fy^a and Fy^b	Four phenotypes	No
		Phenotype-a	Study only tested for Fy^a	Two data types (Fya+/−): cannot distinguish Fy(a−b+) from Fy(a−b−)	No
		Phenotype- b	Study only tested for Fy^b	Two data types (Fyb+/−): cannot distinguish Fy(a+b−) from Fy(a−b−)	No
DNA-based	Genotype	*Promoter*	Study only looked at promoter region SNP	Distinguishes expression from non-expression: cannot distinguish *FY*A* from *FY*B* coding region	Yes (promoter SNP only)
		Genotype	Study looked at promoter and a/b SNP	Fully distinguishes all individual alleles	Yes

SNP, single-nucleotide polymorphism.
During the abstraction process, data points were classified into data types according to the diagnostic methodology used.

median frequencies above 80% predicted across large extents of south Asia, Australia and in populations from Mongolia and eastern parts of China and Russia. The allele is also at high frequencies (>90%) in Alaska and northwest Canada. Outside these regions of highest predominance, *FY*A* remains relatively common outside the African continent, with median frequencies >50% predicted across 67.7% of the global surface. The *FY*B* allele (Fig. 3b) is the allele least prevalent globally, with a maximum predicted median frequency of 83% (thus fixation is never reached). Reflecting its reduced prevalence, the distribution of *FY*B* matches areas of highest allelic heterogeneity, where its presence increases to frequencies similar to, or greater than, *FY*A* and FY^*B^{ES}. Frequencies above 50% are restricted to Europe and pockets of the Americas, notably along the east coast of the United States of America. *FY*B* frequencies decrease from their European epicentre eastwards, as the *FY*A* allele becomes predominant in Asia. *FY*B* is also prevalent in the buffer zones around the region of FY^*B^{ES} predominance, in northern, north-eastern and southern Africa.

Duffy negativity phenotype. The phenotype map of Duffy negativity (Fig. 4) reveals very low frequencies of the homozygous null genotype ($FY^*B^{ES}/^*B^{ES}$) from much of the predicted non-African FY^*B^{ES} distribution (Fig. 3c). Despite being present, the low allelic frequencies mean that homozygotic inheritance is too low to feature in the phenotype map. Therefore, even more pronounced than the allele's distribution, the Duffy negativity phenotype is highly constrained to sub-Saharan African populations, and localized patches in the Americas. Across the African continent, the phenotype's median frequency is greatest (98–100%) in western, central and south-eastern regions from The Gambia to Mozambique, buffered by a high median frequency region of ≥90% frequency covering 22 countries (Fig. 4a). Around this high frequency region, steep clines into the Sahel in the north, and Namibia and South Africa in the south lead to median phenotypic frequencies of <10% in parts of these extremities.

Frequencies of Duffy negativity increase by ~10% south of the sub-Saharan desert boundary, as defined in the model by the

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms1265

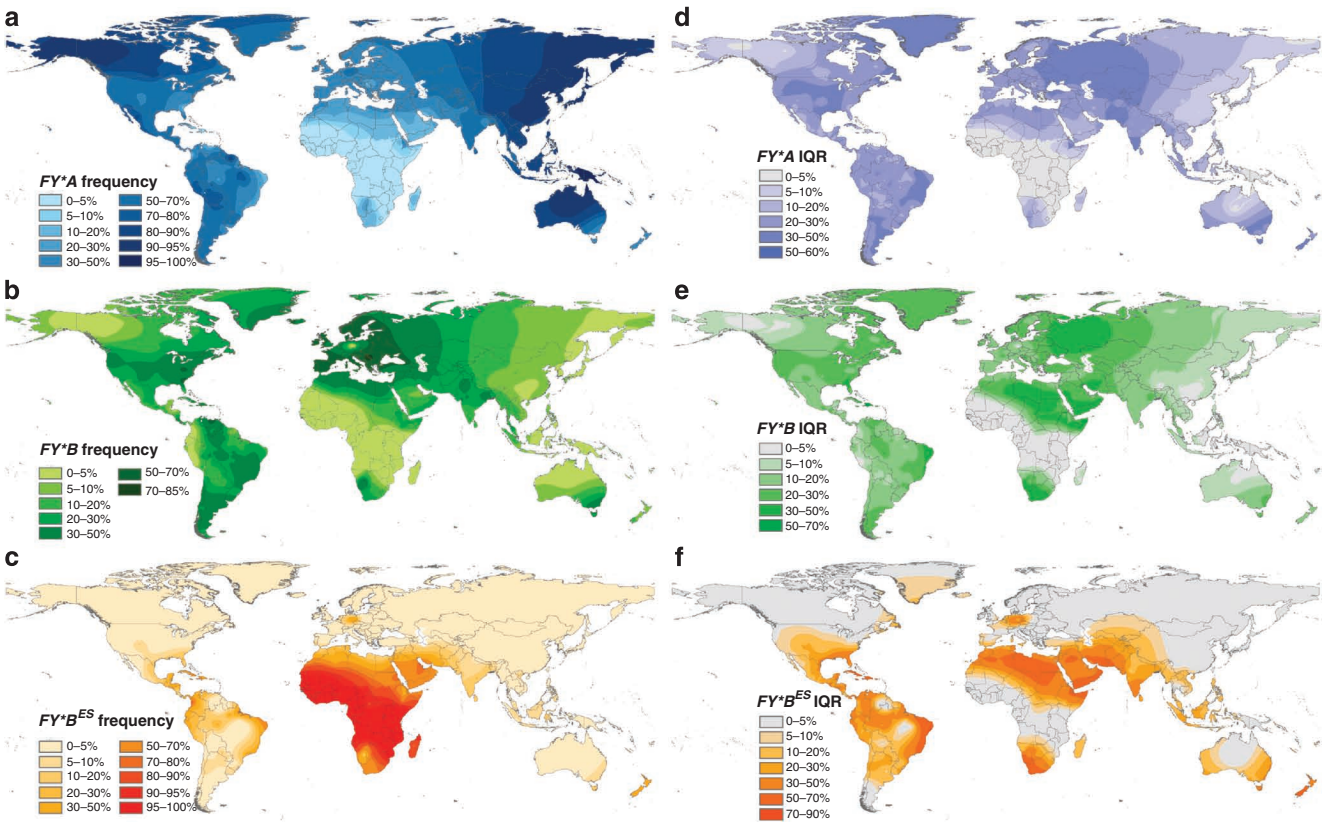


Figure 3 | Global Duffy blood group allele frequencies and uncertainty maps. (**a**–**c**) Correspond to *FY*A*, *FY*B* and *FY*B*^{ES} allele frequency maps, respectively (median values of the prediction posterior distributions); (**d**–**f**) show the respective interquartile ranges (IQR) of each allele frequency map (25–75% interval). Predictions are made on a 5×5 km grid in Africa and 10×10 km grid elsewhere. Supplementary Figure S5 is a greyscale image of this figure.

GlobCover bare ground data set[31] (Fig. 5). This trend of increased frequencies is reflected by the positive values associated with the sub-Saharan Africa covariate, fully described in Supplementary Table S2. The increase is sharpest where there are few underlying data points, contrasting with the prediction in data-rich regions, such as West Africa, which is smoother across the desert boundary due to the abundance of input data overriding its influence.

Prediction uncertainty. The output generated by the Bayesian framework is a predictive posterior distribution for each modelled variable for each 10×10 km pixel on the global grid (and 5×5 km across Africa). The posterior quantifies the probabilities associated with every candidate value of each modelled variable and therefore represents a complete description of uncertainty in the model output[32]. The outputs, summarized in Figures 3a–c and 4a, are the median values of these distributions. The uncertainties around these predictions, represented by the intervals between the 25 and 75% quartiles of the posterior distributions (or IQRs), are shown for the allele maps in Figures 3d–f and for Duffy negativity in Figure 4b. Remarkable certainty in the prediction for high Duffy negativity in sub-Saharan Africa is reflected by IQRs of 0–5% for all outputs. The absence of data points from the Democratic Republic of the Congo leads to a slightly elevated level of uncertainty relative to the surrounding region (Figs 4b and 5). Certainty in the prediction of the highest frequencies of Duffy negativity is illustrated by the hatched areas of ≥95% Duffy negativity prevalence, determined with 75 and 95% confidence (Fig. 5). As would be expected from the heterogeneity in allelic make-up (Fig. 3a,b), the greatest uncertainty in the global predictions of *FY*A* and *FY*B* prevalence is associated with the predictions across Europe, western Asia and the Americas (Fig. 3d,e).

Validation statistics. Bespoke validation procedures were developed to quantify the model's ability to predict frequencies of each allele, as well as the validity in the underlying assumption of Hardy–Weinberg equilibrium: first, by validating the Duffy-negative phenotype surface; second, by assessing rates of heterozygosity between *FY*A* and *FY*B*. The model's predictive ability was quantified by assessing the disparity between model predictions and held-out subsets of data excluded for validation analyses[33]. The validations were summarized using simple statistical measures: mean error (assesses overall model bias) and mean absolute error (quantifies overall prediction accuracy as the average magnitude of the errors, Supplementary Methods).

First, the mean error in the prediction of the Duffy negativity phenotype revealed a slight positive bias in the posterior predictive distribution of Duffy negativity (mean error: 1.3%), while the mean absolute error revealed relatively high typical precision in the predictions (mean absolute error: 5.8%). Second, the heterozygosity validation process identified an overall positive bias in the posterior predictive distribution of rates of heterozygosity (frequencies of the *FY*A/*B* genotype) with mean error of 5.5% and mean absolute error of 7.8%. Overall, therefore, the model's predictive ability for the clinically significant Duffy-negative phenotype was relatively high, although the assumption of Hardy–Weinberg equilibrium was not strongly supported, indicating that better predictions might be achieved in future iterations by modelling the fixation index as a third geostatistical random field. However, the resulting model would be substantially more complicated than the current one, which is already a major advance beyond the state of the art, and commensurately more difficult to fit. Given the relatively small size of the heterozygote deficiency in the holdout data set, we decided

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms1265

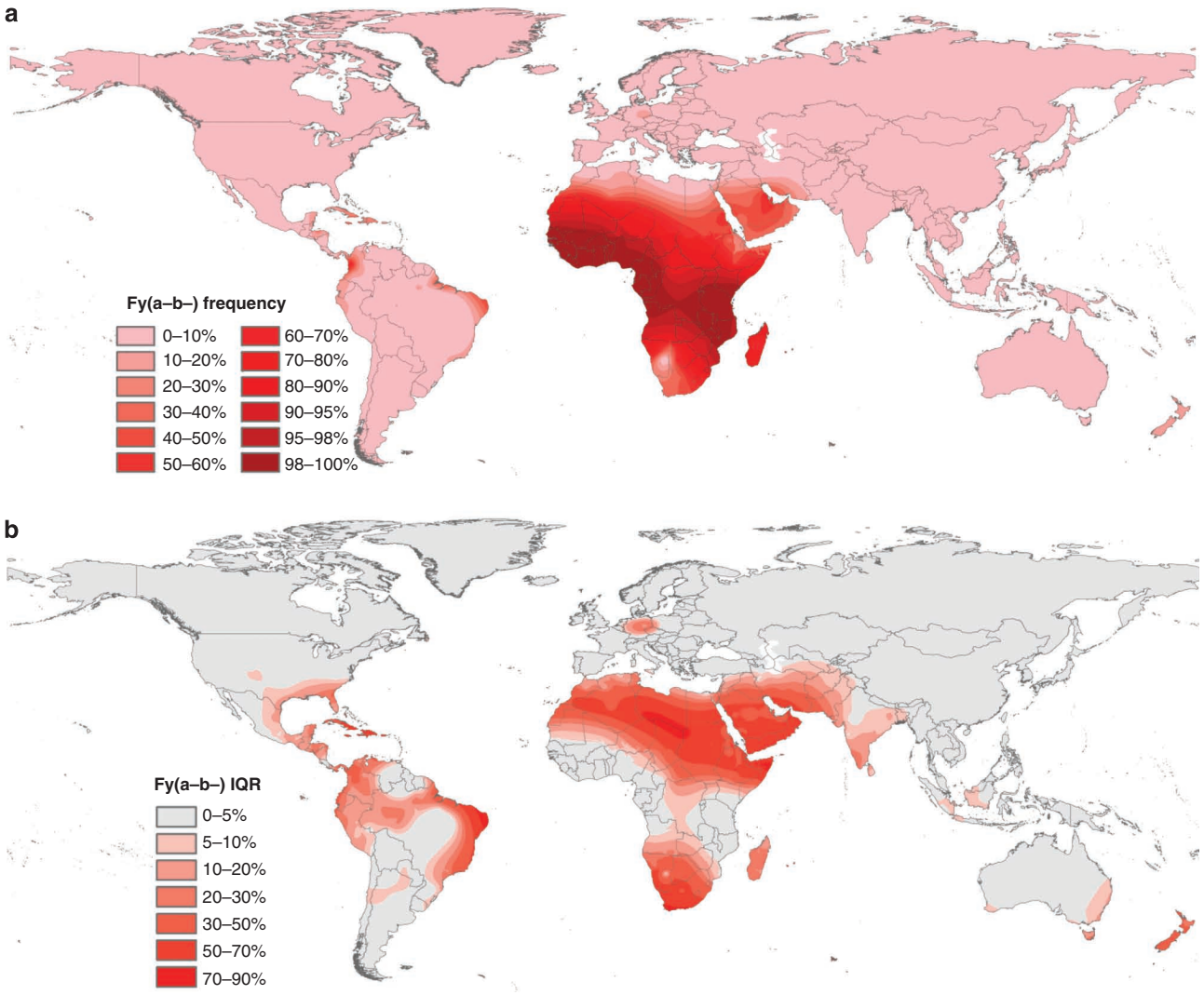


Figure 4 | Global distribution of the Duffy negativity phenotype. (**a**) Global prevalence of Fy(a−b−); (**b**) associated uncertainty map. Uncertainty is represented by the interval between the 25 and 75% quartiles of the posterior distribution (IQR). Supplementary Figure S6 is a greyscale image of this figure.

against elaborating the model in the current study. More validation results are given in Supplementary Figure S3.

Discussion

The spatial distribution of the Duffy blood group variants has been of interest since its discovery 60 years ago because of its link to the pathology of both infectious and non-communicable diseases, including most notably with *P. vivax* infection. We have assembled an up-to-date database of Duffy phenotypic and genotypic data, from which we identified 821 geographically unique community surveys, and developed a geostatistical model to generate global frequency maps for the main Duffy alleles, as well as the first map of the Duffy-negative phenotype. These refined maps and associated uncertainty measures allow both an assessment of the quality and distribution of existing data as well as a discussion of how the maps may help direct further research into the interactions between Duffy negativity and *P. vivax* malaria. A detailed comparison with the existing maps from Cavalli-Sforza *et al.*[28] is presented in the Supplementary Discussion and Supplementary Figures S8–S10.

The summary median maps presented reveal relatively smooth global-scale patterns of geographic differentiation among populations. Despite being considered the ancestral allele[34], our maps show a remarkable restriction in the distribution and frequency of the *FY*B* allele, with highest prevalence found in Europe and parts of the Americas, with further patches of increased prevalence in areas buffering the region of $FY*B^{ES}$ predominance in sub-Saharan Africa. Frequencies of *FY*A* prevalence increase with distance from Africa and Europe, becoming dominant across south-east Asia, including those areas where *P. vivax* endemicity is highest[17]. Although the $FY*B^{ES}$ allele map predicts presence outside the African continent and the Arabian Peninsula, its frequencies remain too low for the Duffy-negative phenotype frequencies to exceed 10%. Although these static contemporary representations of allelic frequencies cannot alone be interpreted to advance current speculation regarding the causative mechanisms of selection of the high frequencies of the $FY*B^{ES}$ allele[4,5,35,36], the Duffy negativity map does reflect visually the historical areas of malaria transmission, as defined by Lysenko's pre-control era malaria map[37] (Supplementary Fig. S4, recently republished by Piel *et al.*[38])

A major challenge in this study was synthesizing the results of surveys, which used a range of diagnostic methods with potentially different reliabilities, particularly between genotyping and phenotyping methods. The possible influence of such variability on the model input is reviewed in detail in the Supplementary Methods,

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms1265

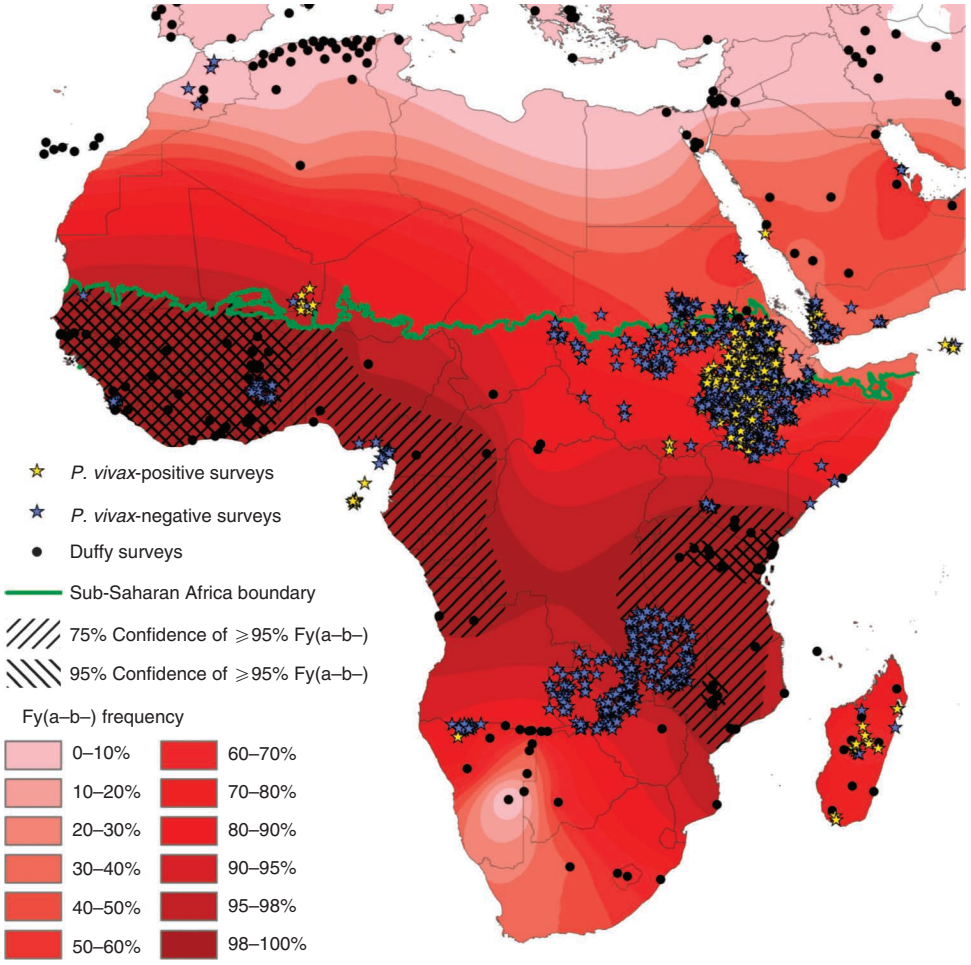


Figure 5 | Characteristics of the Duffy negativity phenotype in Africa. This figure shows the covariate line (in green), which separates sub-Saharan African populations from the rest of the continent; hatched areas indicate areas of confidence in the distribution of ≥95% Duffy negativity frequency: with 75%and 95% confidence. Black data points correspond to the input Duffy data points ($n = 821$). Yellow stars indicate locations of *P. vivax*-positive community surveys ($n = 354$), and blue stars *P. vivax*-negative surveys ($n = 1405$) (data assembled by the Malaria Atlas Project[17,46]). Supplementary Figure S7 is a greyscale image of this figure.

but is not considered to have major influence on the final output. By categorizing results into five data types (Table 2) and developing a versatile geostatistical model, we were able to draw information from the differing data types in our full data set to generate each allele frequency map simultaneously. The *Genotype* data, generated from molecular diagnostic methods only widely available after the previous maps[28] were published, were most informative for the model. Despite a generally good global spread of survey data points (Fig. 2), the uncertainty maps allow identification of areas where additional data would have proportionally greatest impact on our understanding of the distributions. Both the quality (data type) and quantity (data distribution) of the data affect the uncertainty measures. Uncertainty is increased by both scarcity of input data (exemplified across the Arabian Peninsula where only *Phenotype-a* data were available) and heterogeneity (characteristic of the Americas where populations of diverse origins coexist; Figs 3d–f and 5). In contrast, areas of lowest uncertainty match data-rich regions and areas of near-fixation, illustrated by the hatched areas of 95% confidence in the prediction shown in Figure 5. Scarcity of input data also leaves us uncertain about possible fine-scale variation of allelic heterogeneity. This is demonstrated by the relatively high uncertainty in the predictions of the patchily distributed $FY*B^{ES}$ allele across the Americas, where spatial heterogeneity is expected to be high and perhaps not fully represented by the data set. As well as improving reliability in the current predictions, additional molecularly diagnosed data would allow refinements of the model to include additional polymorphic variants, such as the low-frequency weak *FY*X* variant[39]. This is discussed in detail in the Supplementary Discussion.

Reflecting the growing appreciation of *P. vivax*'s public health significance and the realization that it is not 'benign'[16,40,41], the parasite's relationship with the Duffy receptor is the primary focus of contemporary studies of the Duffy antigen. However, two lines of evidence, both from a community and an individual standpoint, support the need for further research into the Duffy–parasite association. First, contrary to expectation, there is evidence of *P. vivax* transmission in areas mapped with highest Duffy negativity frequencies. Although widespread surveys have failed to identify the parasite in this region (including a continental-wide survey by Culleton *et al.*[42], and the data set of community parasite rate surveys displayed in Fig. 5), reports of infected mosquitoes[13], travellers[17] and exposed individuals[43] suggest low level transmission. Across this predominantly Duffy-negative region, very low numbers of Duffy-positive individuals were identified (0.6% of individuals in 123 surveys across the 98–100% Fy(a–b–) region; Supplementary Table S3). To see whether these two observations can be reconciled to explain transmission, mathematical modelling is needed to estimate the basic reproductive number (R_0) of *P. vivax* (as done for *Plasmodium falciparum*[44]) to

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms1265

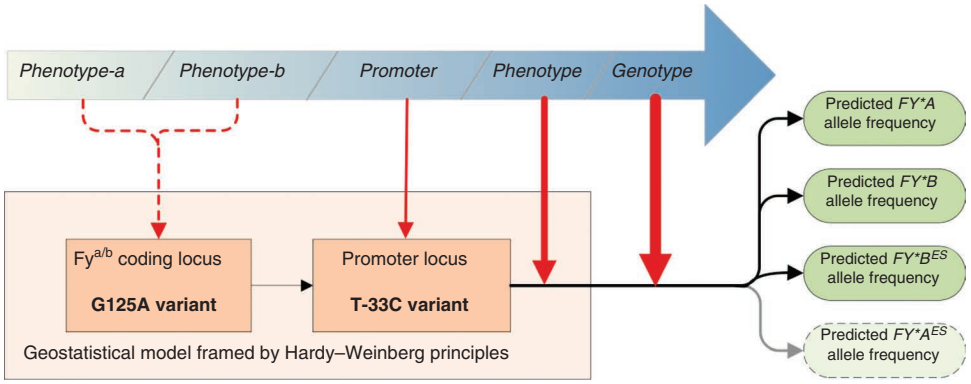


Figure 6 | Relationship between data types and the information conveyed to the model. Left to right along the large arrow, the deepening colour intensity represents each data type's relative influence on the model output. Dashed vertical arrows denote information about only one locus. Thickness of vertical lines emphasizes the completeness of the data type. Orange boxes represent the Bayesian model. Green rods indicate output data. The grey horizontal arrow and greyed $FY*A^{ES}$ prediction rod indicates that this allele was accounted for in the model structure, but not one of the final outputs.

help assess whether the very low predicted frequencies of susceptible Duffy-positive hosts could sustain transmission in populations mapped as predominantly Duffy negative.

Second, from areas mapped with high Duffy phenotypic heterogeneity, *P. vivax* infections have been identified in Duffy-negative hosts (in Madagascar[15] and Brazil[14]). If this phenomenon of infected Fy(a–b–) individuals is associated with local Duffy heterogeneity, as hypothesized by Ménard *et al.*[15], the Duffy maps presented here could be used to target further studies in other heterogeneous *P. vivax* endemic areas[17], including southern Africa, Ethiopia, southern Sudan and pockets of the Brazilian and Colombian coasts. Investigation of *P. vivax* transmission in these areas particularly, but also across regions with a spectrum of characteristic Duffy phenotypes, could provide vital public health insights into *P. vivax* populations at risk, particularly when coupled with host-level data on Duffy types.

In this era of increasing concern about the *P. vivax* parasite, we believe that a contemporary spatial description of the prevalence of the Duffy antigen receptor is essential for optimizing our understanding of the parasite's clinical burden. The geopositioned database and maps represent a new effort to document the spatial characteristics of a fundamental biomedical trait implicated in haematological and other clinical contexts. The versatile geostatistical model developed was adapted to a multiple-locus trait, informed by a range of input data types to generate a suite of output products. Such methods are uncommonly used by the genetics community, but we believe could have an important role in the current era of large-scale spatial genomic analyses. Although we present a cartographic suite which we believe constitutes a significant improvement from previously published attempts[28] (see Supplementary Discussion and Supplementary Figs S8–S10), this study highlights limitations to our current knowledge of the Duffy blood group: both in terms of the scarcity of data from many areas, and in relation to the *P. vivax* invasion pathway. All collated data and model code will be made openly accessible.

Methods

Analysis outline. The methodological steps of this work were threefold: first, to assemble a library of full-text references describing Duffy blood group surveys, complemented with unpublished data; second, to abstract the Duffy frequency data from each source and to georeference survey locations; and third, to develop a spatial model which uses the full heterogeneous data set assembled to predict continuous global frequency maps of the Duffy variants. A schematic overview of the methodological process is given in Figure 1, and each component is now discussed in more detail.

Library assembly. Systematic searches, adapted from those developed by the Malaria Atlas Project (MAP, http://www.map.ox.ac.uk)[45,46], were conducted in an attempt to assemble a comprehensive database of Duffy blood group surveys dating from 1950, the publication year of Cutbush's description of the blood group[1]. Keyword searches for 'Duffy' and 'DARC' were conducted in online bibliographic archives PubMed (http://www.pubmed.gov), ISI Web of Knowledge (http://isiwebofknowledge.com) and Scopus (http://www.scopus.com). Searches were last performed on 08 December 2009. Manual duplicate removal and abstract reviews of the amalgamated search results identified 303 references likely to contain data, in addition to the 296 and 60 references from existing databases published by Mourant *et al.*[47] and Cavalli-Sforza *et al.*[28], respectively. Full-text searches were then conducted for each of these 659 unique references. Following direct contact with researchers, 15 additional unpublished data sets were also included. All sources from which data met the criteria for inclusion are cited in the Supplementary References.

Data abstraction and inclusion criteria. The library of assembled references was reviewed to identify location-specific records of Duffy variant frequencies representative of local populations. Data were abstracted into a customized database, including population descriptions and ethnicities as reported by authors, methodological details and Duffy variant frequencies. Potentially biased samples of hospital patients with malaria symptoms or recently transfused individuals were excluded, as were family-based investigations and studies focussing on selected subgroups of larger mixed communities (for example, African-American communities in American cities). No constraints were placed on sample size, as the geostatistical framework downweighted the information content of very small surveys in accordance with a binomial sampling model[33].

Geopositioning. The geographic location of each survey was determined as precisely as possible using the georeferencing protocol previously described by Guerra *et al.*[45]. Author descriptions of survey sites were used to verify locations identified in digital databases including Microsoft Encarta (Microsoft Corporation), and online databases such as Geonames (National Geospatial-Intelligence Agency, http://geonames.nga.mil/ggmagaz/, accessed June–December 2009) and Global Gazetteer Version 2.2 (Falling Rain Genomics, http://www.fallingrain.com/world/index.html, accessed June–December 2009). Surveys were categorized according to the area they represented: points (≤25 km^2), and small (≥25 and ≤100 km^2) or large polygons (>100 km^2).

Duffy blood group data. In addition to prevalence data of specific variants, details of diagnostic methodology were recorded to classify the type of information provided from the survey (Table 2). According to the range of possible serological and molecular diagnostic methods, data points were classified into five data types: '*Genotype*', where full genotypes were reported; '*Phenotype*', if full serological diagnoses were performed (with both anti-Fy^a and anti-Fy^b antisera); '*Promoter*', if results reported only antigen expression/non-expression without distinguishing Fy^a from Fy^b (data were mainly from molecular studies examining only the promoter-region locus, but also occasionally from serological tests not distinguishing between antigenic variants); '*Phenotype-a*', if only the Fy^a antigen was tested for (meaning that presence of Fy^b antigen could not be distinguished from the negativity phenotype); and '*Phenotype-b*', if the study was only concerned with Fy^b expression (Fig. 6).

Modelling. To accommodate the five input data types described and to model the multiallelic system, the two primary loci differentiating the Duffy variants were considered simultaneously. These were position −33 in the promoter region, which determines expression/non-expression, and base position 125 of exon 2, differen-

tiating Fy^a from Fy^b coding regions[23]. Including the full data set while modelling each variant optimizes the model predictions, as each data type informs, either directly or indirectly, the frequency of variants at both loci, by ruling out certain genotypes. This feature was not possible with previously used mapping models[28].

Genetic loci modelled. The genetic loci were considered as two spatially independent random fields, but modelled in association: first, the random field representing the coding region variant modelled the frequency of Fy^a or Fy^b expression; and second, a random field represented the probability of the promoter 'ES' variant being associated with the Fy^b coding variant, thus determining prevalence of *FY*B* versus $FY*B^{ES}$ alleles. Reports in the data set of the 'ES' variant in association with the Fy^a variant (that is, the $FY*A^{ES}$ allele) were too infrequent and, when identified, were too rare to be modelled as a spatial random field. Therefore, this variant was modelled by a small constant. Further details about the model are given in the Supplementary Methods.

Sub-Saharan Africa covariate. Preliminary examination of the '*Genotype*' data set confirmed the assumption that the haplotypic association between the Fy^b coding variant and the 'ES' promoter variant, together corresponding to the $FY*B^{ES}$ allele, was very high within sub-Saharan African populations, but rare outside the region. To allow the model to reflect this high probability of association across sub-Saharan Africa, we used a generalized version of the GlobCover Land Cover V2.2 bare ground surface (channel 200)[31] to differentiate the sub-Saharan populations (presence), including those living in Madagascar and on other nearby islands (decision informed by the '*Genotype*' data), from the other populations (absence; Fig. 2). This binary descriptor was the only covariate used in the model.

Model implementation. The analyses were implemented in a Bayesian model-based geostatistical framework[29], the principal aspects of which have been previously described[33,48]. In brief, the geostatistical model uses Gaussian random fields to represent the spatial heterogeneity observed in the data and to predict values at unsampled locations. Repeated sampling of the random fields ensures that a representative sample of all the possibilities consistent with the input data set is used in predicting pixel values at sites where there is no data (Supplementary Methods). Estimates for pixels distant from any input data points or in areas of high spatial heterogeneity are inherently more difficult to predict precisely, and so are associated with greater prediction uncertainty. The posterior median values and associated uncertainty (IQR (25th to 75th percent quartile ranges) of the posterior distribution) are used to summarize the model's predictions for each pixel[33,49].

Generating the map surfaces. The model's predictions for allele frequencies were mapped at all pixels on the global grid (at 5×5 km in Africa and 10×10 km elsewhere). Median values of the posterior distribution were chosen for the maps as these were considered more appropriate than mean values, due to the long-tailed distributions of the predictions that could strongly skew mean estimates. From these allele frequency surfaces, genotype frequencies could be obtained using the standard Hardy–Weinberg formula[50,51]. Thus, the Duffy negativity phenotype was expressed by the squared frequency of the silent $FY*B^{ES}$ allele (the $FY*A^{ES}$ allele being too rare to occur in homozygous form).

Model validation. To validate the three allele frequency surfaces and cross-examine the model's assumption of Hardy–Weinberg equilibrium, both the frequency of Duffy negativity and frequency of *FY*A/*B* heterozygosity were validated. For each validation procedure, the model was run with a random subset of the data set left out and predictions at these locations were compared with the observed frequencies. Estimates of the model's overall bias and precision were quantified as mean error and mean absolute error values, respectively (Supplementary Methods).

Availability of data. The survey database and maps are publicly accessible through the MAP website (http://www.map.ox.ac.uk) in line with the MAP's open-access policy and the terms of the Wellcome Trust Biomedical Resources Grant (#085406) funding this work.

References

1. Cutbush, M. & Mollison, P. L. The Duffy blood group system. *Heredity* **4,** 383–389 (1950).
2. Mourant, A. E. *Blood Relations: Blood Groups and Anthropology* (Oxford University, 1983).
3. Hamblin, M. T., Thompson, E. E. & Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* **70,** 369–383 (2002).
4. Carter, R. Speculations on the origins of *Plasmodium vivax* malaria. *Trends Parasitol.* **19,** 214–219 (2003).
5. Rosenberg, R. *Plasmodium vivax* in Africa: hidden in plain sight? *Trends Parasitol.* **23,** 193–196 (2007).
6. Klein, H. G. & Anstee, D. J. (eds.) *Mollison's Blood Transfusion in Clinical Medicine* (Blackwell Publishing, 2005).
7. Mourant, A. E. & Domaniewska-Sobczak, K. The use in anthropology of blood groups and other genetical characters. *J. Afr. Hist.* **3,** 291–296 (1962).
8. Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N. Engl. J. Med.* **295,** 302–304 (1976).
9. Barnwell, J. W., Nichols, M. E. & Rubinstein, P. *In vitro* evaluation of the role of the Duffy blood group in erythrocyte invasion by *Plasmodium vivax*. *J. Exp. Med.* **169,** 1795–1802 (1989).
10. Wertheimer, S. P. & Barnwell, J. W. *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. *Exp. Parasitol.* **69,** 340–350 (1989).
11. Miller, L. H., Mason, S. J., Dvorak, J. A., McGinniss, M. H. & Rothman, I. K. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* **189,** 561–563 (1975).
12. Barber, M. A. & Komp, W. H. W. The seasonal and regional incidence of types of malaria parasites. *Public Health Rep. (1896–1970)* **44,** 2048–2057 (1929).
13. Ryan, J. R. *et al.* Evidence for transmission of *Plasmodium vivax* among a Duffy antigen negative population in Western Kenya. *Am. J. Hum. Genet.* **75,** 575–581 (2006).
14. Cavasini, C. E. *et al. Plasmodium vivax* infection among Duffy antigen-negative individuals from the Brazilian Amazon region: an exception? *Trans. R. Soc. Trop. Med. Hyg.* **101,** 1042–1044 (2007).
15. Ménard, D. *et al. Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc. Natl Acad. Sci. USA* **107,** 5967–5971 (2010).
16. Galinski, M. R. & Barnwell, J. W. *Plasmodium vivax*: who cares? *Malar. J.* **7,** S9 (2008).
17. Guerra, C. A. *et al.* The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl. Trop. Dis.* **4,** e774 (2010).
18. Anstee, D. J. The relationship between blood groups and disease. *Blood* **115,** 4635–4643 (2010).
19. Horne, K. & Woolley, I. J. Shedding light on DARC: the role of the Duffy antigen/receptor for chemokines in inflammation, infection and malignancy. *Inflamm. Res.* **58,** 431–435 (2009).
20. Walley, N. M. *et al.* The Duffy antigen receptor for chemokines null promoter variant does not influence HIV-1 acquisition or disease progression. *Cell Host Microbe* **5,** 408–410; author reply 418–409 (2009).
21. He, W. *et al.* Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. *Cell Host Microbe* **4,** 52–62 (2008).
22. Donahue, R. P., Bias, W. B., Renwick, J. H. & McKusick, V. A. Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc. Natl Acad. Sci. USA* **61,** 949–955 (1968).
23. Langhi, D. M. Jr. & Bordin, J. O. Duffy blood group and malaria. *Hematology* **11,** 389–398 (2006).
24. Zimmerman, P. A. in *Infectious Disease and Host-Pathogen Evolution* (ed. Dronamraju, K. R.) 141–172 (Cambridge University, 2004).
25. Zimmerman, P. A. *et al.* Emergence of $FY*A^{null}$ in a *Plasmodium vivax*-endemic region of Papua New Guinea. *Proc. Natl Acad. Sci. USA* **96,** 13973–13977 (1999).
26. Sellami, M. H. *et al.* Duffy blood group system genotyping in an urban Tunisian population. *Ann. Hum. Biol.* **35,** 406–415 (2008).
27. Olsson, M. L. *et al.* The Fy(x) phenotype is associated with a missense mutation in the Fy(b) allele predicting Arg89Cys in the Duffy glycoprotein. *Br. J. Haematol.* **103,** 1184–1191 (1998).
28. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University, 1994).
29. Diggle, P. J. & Ribeiro, P. J. Jr. *Model-based Geostatistics* (Springer, 2007).
30. Hamblin, M. T. & Di Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.* **66,** 1669–1679 (2000).
31. Bicheron,, P. *et al. GLOBCOVER: Products Description and Validation Report* (MEDIAS, 2008).
32. Hogg, R. V. & Craig, A. *Introduction to Mathematical Statistics* (Pearson Education, 2005).
33. Hay, S. I. *et al.* A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.* **6,** e1000048 (2009).
34. Tournamille, C. *et al.* Sequence, evolution and ligand binding properties of mammalian Duffy antigen/receptor for chemokines. *Immunogenetics* **55,** 682–694 (2004).
35. Carter, R. & Mendis, K. N. Evolutionary and historical aspects of the burden of malaria. *Clin. Microbiol. Rev.* **15,** 564–594 (2002).
36. Livingstone, F. B. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. *Hum. Biol.* **56,** 413–425 (1984).
37. Lysenko, A. J. & Semashko, I. N. in *Itogi Nauki: Medicinskaja Geografija* (ed. Lebedew, A. W.) 25–146 (Academy of Sciences, 1968).
38. Piel, F. B. *et al.* Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat. Commun.* **1,** 104 (2010).
39. Chown, B., Lewis, M. & Kaita, H. Duffy blood group system in Caucasians - evidence for a new allele. *Am. J. Hum. Genet.* **17,** 384–389 (1965).
40. Price, R. N. *et al.* Vivax malaria: neglected and not benign. *Am. J. Hum. Genet.* **77,** 79–87 (2007).

41. Baird, J. K. Severe and fatal vivax malaria challenges 'benign tertian malaria' dogma. *Ann. Trop. Paediatr.* **29,** 251–252 (2009).
42. Culleton, R. L. *et al.* Failure to detect *Plasmodium vivax* in West and Central Africa by PCR species typing. *Malaria J.* **7,** 174 (2008).
43. Culleton, R. *et al.* Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. *J. Infect. Dis.* **200,** 1465–1469 (2009).
44. Smith, D. L., McKenzie, F. E., Snow, R. W. & Hay, S. I. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol.* **5,** e42 (2007).
45. Guerra, C. *et al.* Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malaria J.* **6,** 17 (2007).
46. Hay, S. I. & Snow, R. W. The Malaria Atlas Project: developing global maps of malaria risk. *PLoS Med.* **3,** e473 (2006).
47. Mourant, A. E., Kopeć, A. C. & Domaniewska-Sobczak, K. *The Distribution of the Human Blood Groups and other Polymorphisms* (Oxford University, 1976).
48. Diggle, P. J., Tawn, J. A. & Moyeed, R. A. Model-based geostatistics. *J. R. Stat. Soc. Ser. C Appl. Stat.* **47,** 299–326 (1998).
49. Gething, P. W., Patil, A. P. & Hay, S. I. Quantifying aggregated uncertainty in *Plasmodium falciparum* malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. *PLoS Comput. Biol.* **6,** e1000724 (2010).
50. Hardy, G. H. Mendelian proportions in a mixed population. *Science* **28,** 49–50 (1908).
51. Weinberg, W. Über den nachweis der vererbung beim menschen. *Jahresh. Wuertt. Verh. Vaterl. Naturkd.* **64,** 369–382 (1908).

Acknowledgments

Data from the MalariaGEN Consortium (http://www.malariagen.net) have been shared for inclusion in our database. We thank all its contributing collaborators and members for collecting, preparing and genotyping the samples. We also thank the following people for sharing unpublished data: Anabel Arends and Gilberto Gòmez for Venezuela data; Marcelo Urbano Ferreira for Brazil data; Rick Fairhurst, Carole Long and Mahamadou Diakite for Mali data; Rick Fairhurst and Duong Socheat for Cambodia data. We acknowledge Kevin Baird, Carlos Guerra, Kevin Marsh, Robert Snow and William Wint for comments on the manuscript, and Anja Bibby for editing the manuscript. This work was supported by a Wellcome Trust Biomedical Resources Grant (#085406), which funds R.E.H., F.B.P. and O.A.N., a Senior Research Fellowship to S.I.H. from the Wellcome Trust (#079091), which also supports P.W.G., and a Wellcome Trust Principal Research Fellowship (#079080) to Professor Robert Snow, which funds A.P.P. T.N.W. is funded by a Senior Clinical Fellowship (#076934) from the Wellcome Trust. D.J.W. is funded by the Wellcome Trust. This paper is published with the permission of the director of KEMRI. This work forms part of the output of the MAP (http://www.map.ox.ac.uk), principally funded by the Wellcome Trust, UK.

Author contributions

R.E.H. assembled the data and wrote the first draft of the manuscript with S.I.H. and F.B.P., who conceived the study and advised on all aspects of the project; O.A.N. and C.W.K. helped assemble and geoposition the data; A.P.P. and P.W.G. conceived and helped to implement the modelling and all computational tasks. T.N.W. and D.J.W. had advisory roles throughout the project. P.A.Z. contributed data for the West Africa region and Papua New Guinea, C. Beall and A.G. contributed unpublished data from Ethiopia, and C. Barnadas and D.M. contributed data from Madagascar. All authors contributed to the revision of the final manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Howes, R.E. *et al.* The global distribution of the Duffy blood group. *Nat. Commun.* 2:266 doi: 10.1038/ncomms1265 (2011).

Chapter 3 – Duffy negativity as an indicator of *P. vivax* transmission potential

Having described the methodological assembly of the Duffy blood group maps in Chapter 2, I now turn to consider these in their epidemiological context. I discuss how the Duffy negativity map has supported the assessment of *P. vivax* transmission and of the population at risk of infection, particularly in Africa. Although I did not lead the studies discussed here, I made active contributions towards them and they were substantively enabled by the Duffy maps. My exact contributions to each publication are detailed in the Statement of Contribution. These studies are collated here to demonstrate the application of the Duffy maps to supporting our evolving understanding of *P. vivax* epidemiology.

Miller's demonstration of the Duffy antigen as a trans-membrane receptor for *P. vivax* infection of red blood cells (Miller *et al.*, 1975; Miller *et al.*, 1976) abated further research into the epidemiology of *P. vivax* in Africa as populations across this continent were known to rarely express the Duffy antigen, and the existing dogma of *P. vivax* absence from Africa was reinforced (Rosenberg, 2007). The WHO malaria treatment guidelines in 2010 for the African region (WHO African Regional Office: AFRO) make no mention of *P. vivax* transmission for most countries (exceptions include: Algeria, Eritrea, Ethiopia, Namibia, South Africa) (WHO, 2011). Suspected malaria patients are not consistently diagnosed: of the cases reported to public sector health facilities across the AFRO region in 2010, only 45% were tested, and only a "*relatively low*" number with microscopy (approx. 30 million patients across AFRO) (WHO, 2011). This indicates that non-*P. falciparum* cases would most likely escape undiagnosed to species level and simply be considered another case of 'malaria'. Where diagnostics are available, *P. vivax* cases have commonly been mistakenly characterised as *P. ovale*, due to their morphological similarities and the widespread assumption that *P. vivax* is absent from this continent (Rosenberg, 2007).

Until recently, no substantive evidence brought into doubt the universality of Miller's findings and the absence of *P. vivax* from Africa, despite reports accumulating of *P. vivax* transmission from almost all countries across the continent (Guerra *et al.*, 2010). Since 2010, landmark studies have also robustly demonstrated *P. vivax* infection of Duffy negative red blood cells, although these reports are sporadic and of limited scope, both spatially and in terms of sample sizes. While *P. vivax* is evidently not a major co-endemic parasite in sub-Saharan Africa, the recent evidence brings into question the long-held dogma of its absence and throws these assumptions back into question.

This chapter discusses how knowledge of the spatial distribution of Duffy negativity prevalence can support our understanding of *P. vivax* epidemiology in Africa, and perhaps provide some insight into the implications of a shifting understanding of the dependency by *P. vivax* on the Duffy antigen. I first consider how the Duffy negativity map has supported efforts to map the global transmission limits and endemicity of *P. vivax,* and from these, estimates of the population at risk of *P. vivax* (*Pv*PAR). Next, the recent evidence of Duffy-independent transmission is reviewed, and consideration of how the Duffy negativity map may be used to project scenarios of the impact of Duffy-independent transmission are considered. Finally, I consider how the Duffy variant maps (*FY*A* and *FY*B*) are supporting investigations of vaccine targets. In each of these cases, brief overviews of these studies are given, and the role of the Duffy maps explained. The full publications from which these studies are taken are cited in the relevant sections (Guerra *et al.*, 2010; King *et al.*, 2011; Gething *et al.*, 2012; Zimmerman *et al.*, 2013) and included in my thesis Appendix.

3.1. The global population at risk of *P. vivax* in 2010

Strategic planning, monitoring and evaluation of disease control require basic information on the spatial distribution of the PAR of infection. The first global map of *P. vivax* transmission was generated by the Malaria Atlas Project in 2010 (Guerra *et al.*, 2010), using a raft of evidence from medical and environmental sources elaborated further below.

These were combined to define areas of stable and unstable transmission, differentiated due to each requiring different monitoring and intervention strategies. Within areas of stable transmission (defined as having one or more cases per 10,000 population per annum) it was important to map transmission intensity (endemicity). Endemicity is an epidemiological measure with important implications for optimising control interventions (Hay *et al.*, 2009), as well as for assessing the feasibility of control or local elimination (Moonen *et al.*, 2010; Tatem *et al.*, 2010). Furthermore, important clinical relationships such as the incidence and clinical severity of infections vary non-linearly in relation to parasite endemicity (Patil *et al.*, 2009).

The Duffy negativity map was incorporated into the *P. vivax* endemicity-mapping model to account for individuals who were innately immune to infection. Endemicity was quantified in this study as the parasite rate (*Pv*PR), a measure of the proportion of a sample population carrying parasites (Guerra *et al.*, 2007). Integrating the two data sources (*Pv*PR population surveys and the Duffy negativity prevalence map) reduced the level of uncertainty in the *P. vivax* endemicity map in areas where parasite data was scarce. From this Duffy-integrated map, it was possible to estimate the *Pv*PAR. Each of these steps is discussed in more detail now.

3.1.1. Mapping the limits of P. vivax *transmission in 2010*

A cumulative step-wise approach using a wide range of data was taken to refine the transmission limits of *P. vivax* infection, fully described by Guerra *et al*. (2010) and subsequently updated by Gething *et al.* (2012) (Figure 3.1). Sources of international travel and health guidelines identified 95 countries globally as being endemic with *P. vivax* malaria in 2010. Sub-national extents of transmission risk were defined using *P. vivax* annual parasite incidence (*Pv*API) data aggregated by administrative units, which were classified as: malaria free, unstable (i.e. extremely low transmission: where *Pv*API reported <1 case per 10,000 people/annum), and stable (i.e. where *Pv*API reported ≥1 case per 10,000 people/annum). Biological masks were then applied to exclude areas where

environmental characteristics would prevent transmission, including low temperatures and high aridity. For transmission to be biologically feasible, a cohort of *Anopheles* vectors infected with the parasite must survive long enough for sporogony to complete. This was modelled based on sporogonic rate as

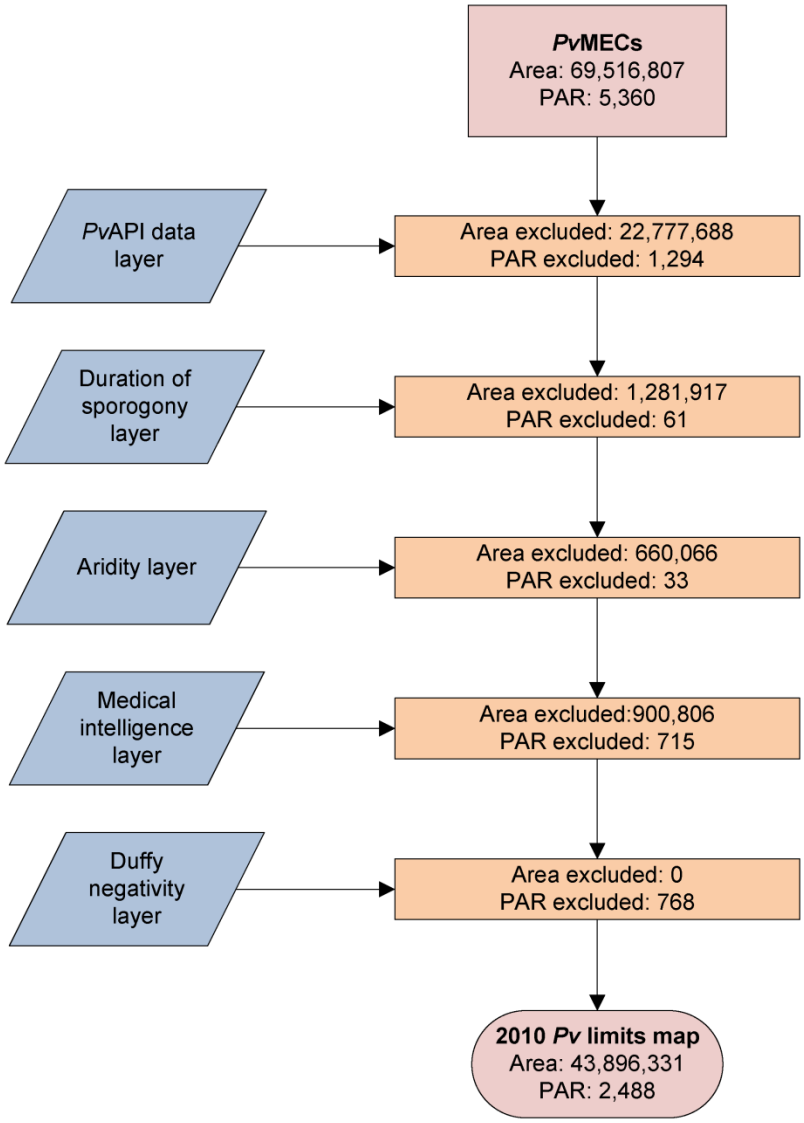


Figure 3.1. Flow chart of the various exclusion layers used to derive the final map of *P. vivax* transmission limits. Area (expressed in km^2) and population at risk (*Pv*PAR; expressed in millions) excluded are shown at each step to illustrate how these were reduced progressively. Figure published by Gething *et al*. (2012), updating the iteration by Guerra *et al*. (2010).

determined by temperature in relation to vector lifespan (Gething *et al.*, 2011b). Areas where this would not be possible at any time of year were excluded as 'malaria free'. Similarly, areas of high aridity would prevent mosquito development by restricting oviposition sites and their survival chances. This aridity mask was derived from the bare ground areas in the GlobCover land-cover imagery (ESA/ESA GlobCover Project, led by MEDIAS-France/POSTEL) (Bicheron *et al.*, 2008). To allow for potential human adaptations to arid environments inadvertently creating vector breeding habitats, arid areas were only downgraded by one transmission class (rather than fully excluded, as with the temperature mask). This meant that areas previously classified as having stable transmission became unstable; and unstable became malaria free. Finally, medical intelligence of localised transmission risk was incorporated, such as urban areas and sub-national territories (commonly islands) known to be malaria free. Most vectors are known to be poorly adapted to urban environments, with the exception of *An. stephensi* which is common across the Indian sub-continent (Sinka *et al.*, 2011); urban areas outside the range of *An. stephensi* were therefore classified as malaria free, and risk within urban areas where *An. stephensi* was likely to be present were downgraded by one risk level (stable to unstable; unstable to malaria free).

These different exclusion layers were progressively applied in a geographical information system, leading to incremental reductions in estimated areas of transmission and *Pv*PAR (Figure 3.1). The limits of stable and unstable transmission are mapped in Figure 3.2A. These steps involved no direct exclusion of transmission areas based on prevalence of Duffy negativity, thus the whole of sub-Saharan Africa was classified as having the potential to support transmission. The Duffy negative population was excluded in the final step of the *Pv*PAR calculation (Figure 3.1). This is further discussed in Section 3.1.3.

3.1.2. Modelling the endemicity map of P. vivax *in 2010*

The Duffy negativity prevalence map was integrated into the global model for mapping *P. vivax* endemicity for two reasons (Gething *et al.*, 2012). First, the additional source of data

could reduce prediction uncertainty in areas where other input data were sparse; second, it ensured consistency between the Duffy negativity prevalence and *P. vivax* endemicity predictions, thus avoiding biologically implausible pairs of results being generated.

As with the previously published global endemicity map of *P. falciparum* (Hay *et al.*, 2009; Gething *et al.*, 2011a) and the Duffy blood group maps (Howes *et al.*, 2011), model-based geostatistics were used in a Bayesian framework to predict the continuous prevalence map of *P. vivax* endemicity (Gething *et al.*, 2012), each model bespoke to the particular intricacies of the biological systems being modelled (Patil *et al.*, 2011). The endemicity mapping was restricted to areas of stable transmission. *Pv*PR surveys were the primary evidence-base to the *P. vivax* endemicity map, these quantified endemicity as the proportion of infected individuals surveyed. The distribution of these surveys was highly uneven across regions, particularly across sub-Saharan Africa (Figure 3.2A). Of the 9,970 *Pv*PR surveys recorded globally, 79.7% (n = 7,942) were from Central and South-East Asia (CSE Asia) and only 16.4% (n = 1,640) were identified from the Africa+ region (comprising Africa, Yemen and Saudi Arabia). Survey distribution across Africa+ was highly clustered across the 18 countries from which data were identified. 86.0% of data points were from only three countries: Ethiopia (50.4% overall; n = 826), Zimbabwe (18.0% overall; n = 295) and Sudan (17.7% overall; n = 290). Across the Africa+ region, 79% of the data were absence records. The large swathes of sub-Saharan Africa where data were especially scarce are where Duffy negativity prevalence was at its highest. The *P. vivax* endemicity model predictions could therefore borrow strength from the Duffy negativity map by restricting plausible predictions of *Pv*PR to a much narrower range of possible values.

The median predicted Duffy negativity prevalence map was integrated within the *P. vivax* mapping model so that *Pv*PR was only mapped within the proportion of the Duffy positive population. The proportion of Duffy negative individuals was excluded from the denominator of the *Pv*PR survey data, such that any *P. vivax* positive individuals were considered to have arisen from the Duffy positive population subset. Thus in a location with

90% Duffy negativity, five positive individuals in a survey of 100 would give an assumed prevalence of 50% amongst Duffy positives. Correspondingly, prediction of *Pv*PR was then restricted to the Duffy positive proportion at each pixel, with the final prevalence estimate re-converted to relate to the total population. This approach ensured that the predicted *Pv*PR at each location could never exceed the Duffy positive proportion. *Pv*PR predictions were standardised for the 1-99 year age distribution ($PvPR_{1\text{-}99}$) (Gething *et al.*, 2012).

Some areas were predicted to have extremely low endemicity, either due to a high density of survey data reporting zero infections, or due to the very high prevalence of Duffy negativity. Such low transmission levels are not appropriately described as being at stable transmission risk, so a rule was defined whereby pixels on the map grid (5×5 km) with a high probability (>0.9) of being less than 1% $PvPR_{1\text{-}99}$ were assigned to unstable transmission, thereby changing the original transmission limits.

Figure 3.2 summarises the progression from the original transmission limits derived from the step-wise exclusion criteria (Figure 3.1; Figure 3.2A), through to the modelled *P. vivax* endemicity prediction map which incorporated the Duffy negativity population surface (Figure 3.2B), and finally the updated transmission limits which reflect the re-assigned areas of very low $PvPR_{1\text{-}99}$ endemicity as unstable (Figure 3.2C). It is this final map which was used to extract the PAR of *P. vivax* infection.

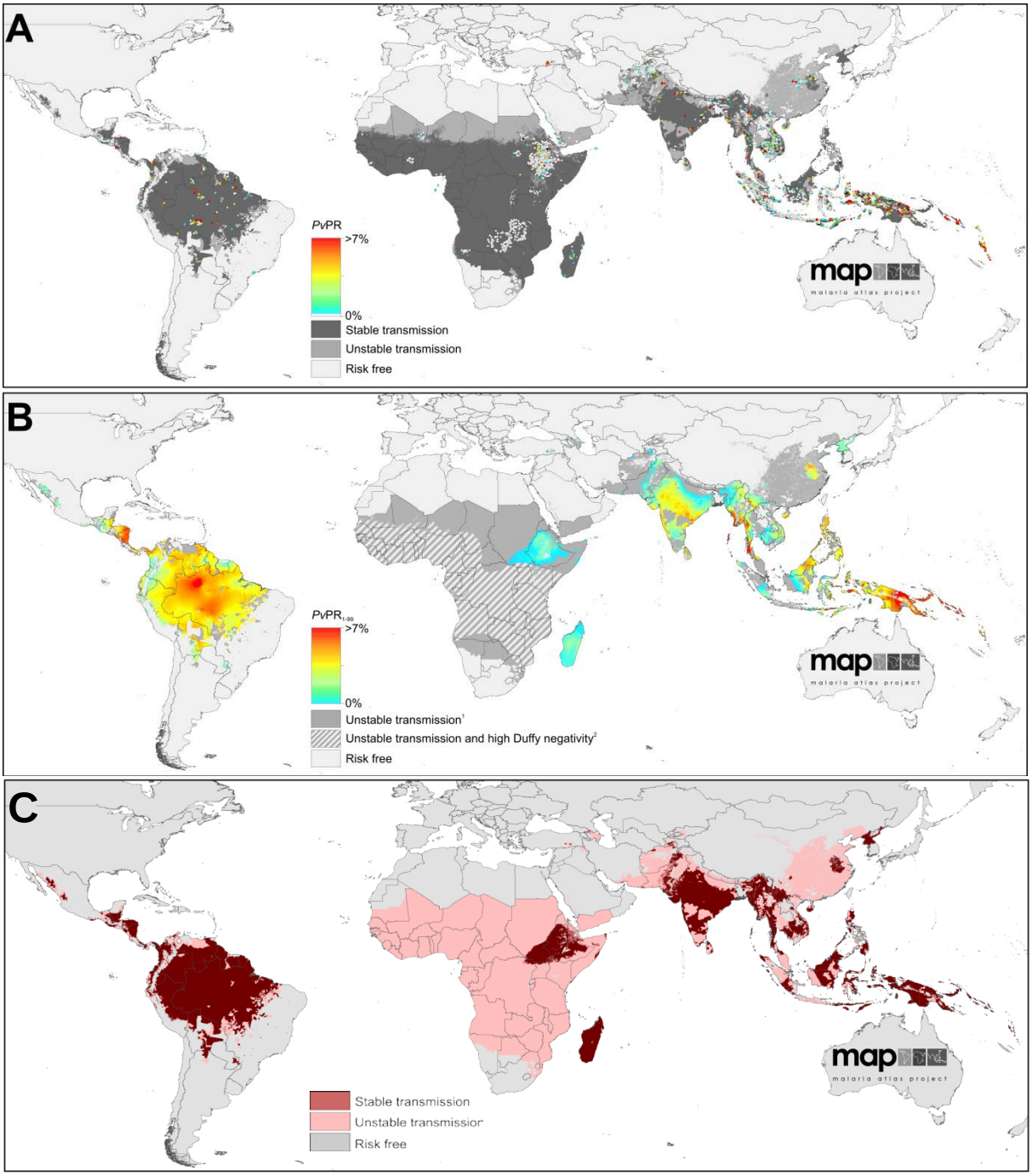


Figure 3.2. The spatial distribution of *Plasmodium vivax* malaria endemicity in 2010. Panel A shows the 2010 spatial limits of *P. vivax* malaria risk, distinguishing malaria free regions from areas of stable (≥1 case per 10,000 people/annum) and unstable transmission (<1 case per 10,000 people/annum). Parasite rate surveys are plotted on a continuous colour scale (see map legend), with *P. vivax* absence surveys shown in white. Panel B shows the mean prediction of *P. vivax* endemicity in 2010 (standardised for the 1-99 years age distribution), within the stable limits of transmission. Areas where Duffy negativity median prevalence was predicted to exceed 90% (Howes *et al.*, 2011) are hatched. Panel C shows the adjusted transmission limits of stable (red) and unstable (pink) transmission accounting for the downgrading of risk in areas where there was a high probability of (>0.9) of low endemicity (<1% *Pv*PR$_{1\text{-}99}$). Figures published by Gething *et al*. (2012)

3.1.3. Estimating the population at risk of P. vivax *infection in 2010*

Estimates of the *Pv*PAR of stable and unstable transmission (Figure 3.2C) had to account for the human genetic landscape at the Duffy locus (Howes *et al.*, 2011). Duffy negative individuals were excluded from these estimates as they were not considered to be at risk of infection (Gething *et al.*, 2012). The Global Rural Urban Mapping Project (GRUMP) *beta* version provides gridded population counts at 1×1 km spatial resolution for the year 2000 adjusted to the United Nations' national population estimates (Balk *et al.*, 2006). These were projected to the year 2010 by applying the relevant urban and rural national growth rates by country (U.N.P.D., 2007) (Figure 3.3). The Duffy negative population was then excluded, and the map of Duffy positive individuals was overlaid. The *Pv*PAR could then be extracted using GIS software (ArcMap 10.0, ESRI Inc., Redlands, CA, USA).

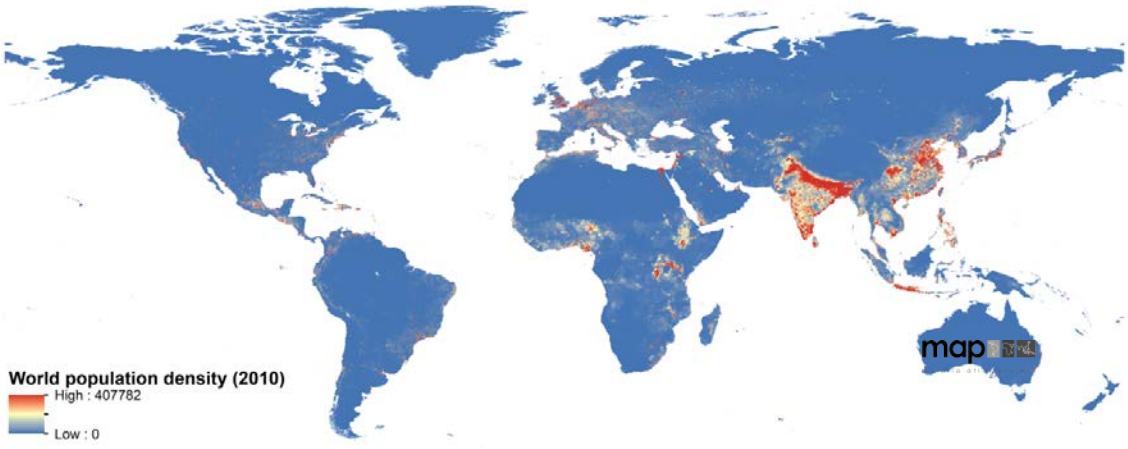


Figure 3.3. Global population density in 2010. Number of individuals per 1×1 km pixel from the GRUMP *beta* version. Figure is reproduced from Piel *et al.* (2010)

The global focus of the *Pv*PAR is clearly centred in the CSE Asia region, a product both of very high population density across this region but also of the protective effect of the Duffy negative phenotype among African populations (Table 3.1). Of the 2.48 billion individuals at risk of *P. vivax* in 2010, 91% were from the CSE Asia region, 6% in the Americas, and only 3% from the Africa+ region. Without the Duffy negativity exclusion, these numbers would put 25% of the global *P. vivax* PAR in the Africa+ region. The public health significance of the protective Duffy negativity effect is discussed below, but it is evident

from a cartographic perspective that the Duffy surface has considerably extended and refined the use of the *P. vivax* epidemiological data.

Exclusion	Region	Americas	Africa+	CSE Asia	Global
Without Duffy exclusions	*Unstable*	99.9	724.8	1,393.6	2,218.3
	Stable	59.8	92.0	882.8	1,034.6
	Total at Risk	159.7	816.8	2,276.4	3,252.9
With Duffy exclusions	*Unstable*	87.6	48.7	1,384.3	1,520.6
	Stable	49.8	37.7	876.6	964.1
	Total at Risk	137.5	86.4	2,260.9	2,484.7

Table 3.1. Population at risk of *Plasmodium vivax* malaria in 2010. Population estimates are regionally stratified and shown with and without exclusion of the Duffy negative population. The 'unstable' category includes individuals in areas which were re-classified from stable to unstable for having a high probability of low endemicity.

3.2. Bringing into question the dependency of *P. vivax* on the Duffy antigen: implications for global *P. vivax* epidemiology

3.2.1. P. vivax *is present in Africa*

In spite of the very high prevalence of Duffy negativity across Africa and the probability of *P. vivax* presence being too low to represent in the endemicity map (Figure 3.2), there is compelling evidence that *P. vivax* is nevertheless present across this region. First, clinical cases have been reported from almost every country (exceptions are Guinea-Bissau and Swaziland from which no published references could be identified (Guerra *et al.*, 2010)), usually from returning travellers (Rubio *et al.*, 1999; Gautret *et al.*, 2001; Muhlberger *et al.*, 2004). Second, a study in the Republic of Congo, where 96.2% of the population were predicted to be Duffy negative (King *et al.*, 2011), 13% of individuals (n = 409) tested were found to be positive for pre-erythrocytic *P. vivax*-specific antibodies (Culleton *et al.*, 2009). Third, *P. vivax*-infected *Anopheles* have been reported from western Kenya (32 of 4,901 mosquitoes) (Ryan *et al.*, 2006). Attempts have been made to quantify the extent of

infection in sub-Saharan Africa. For instance, a study of 2,588 individuals across nine countries in west and central Africa (where Duffy negativity prevalence was always ≥95% (Howes *et al.*, 2011; King *et al.*, 2011)), reported only a single *P. vivax*-positive case by PCR-diagnosis (Culleton *et al.*, 2008). The case was a Duffy-positive individual. Population screening surveys reviewed by the Malaria Atlas Project identified *P. vivax* infections in communities on the fringes of the areas with highest prevalence of Duffy negativity (e.g. in southern Sudan, eastern Mali and northern Namibia), though none across areas of highest prevalence of Duffy negativity (Figure 3.4).

Evidently, the parasite is present in communities across Africa, but the relative rarity of blood-stage infections makes these hard to quantify. Although these observations of *P. vivax* might be considered unexpected, they do not bring into question the fundamental biology of *P. vivax* transmission. *P. vivax* in Africa could be being propagated by the minority of Duffy positive individuals across the region; the parasite's relapsing nature would facilitate small host populations sustaining the parasites. Strong evidence, however, points to a more complex situation.

3.2.2. Duffy-independent P. vivax *transmission*

Ménard and colleagues provided the first robust evidence of *P. vivax*-infected Duffy negative blood cells, an indication of Duffy-independent transmission (Ménard *et al.*, 2010). Furthermore, in the high transmission areas of Madagascar where this study was conducted, the prevalence of infection was not significantly different between Duffy positive and negative individuals, although prevalence of clinical *P. vivax* was significantly lower (>15-fold reduction) among Duffy negative individuals. Further studies have reported infection in Duffy negative individuals across Africa (Figure 3.4), but with data not as compelling as the Malagasy results. Surveys in Ethiopia reported an unspecified number of *P. vivax*-infected Duffy negative hosts, diagnosed both with microscopy and PCR. However, these results are so far unpublished in the peer-reviewed literature (Woldearegai *et al.*, 2011). Microscopic analysis of Kenyan *P. vivax* patients could not be confirmed with

certainty as being distinct from *P. ovale* (Ryan *et al.*, 2006); and reports from Mauritania (Wurtz *et al.*, 2011), Angola and Equatorial Guinea (Mendes *et al.*, 2011) used PCR analysis alone, which may have identified pre-erythrocytic stage parasites (Culleton and Carter, 2012) and therefore not be evidence of Duffy-independent transmission.

The admixture of Duffy positive and negative phenotypes in the Malagasy population (Figure 3.4) led Ménard and colleagues to hypothesise that this very admixture may have fuelled selection of Duffy-independent transmission (Ménard *et al.*, 2010; Zimmerman *et al.*, 2013). They found that communities with the highest frequencies of Duffy negativity had very low or no *P. vivax* infection in any hosts; whereas where Duffy positive individuals were more common, *P. vivax* prevalence increased in individuals of both phenotypes. These results suggest that high population levels of Duffy negativity may provide herd immunity to reduce transmission and consequently protect Duffy positives from *P. vivax* infection. In populations with higher frequencies of Duffy positivity, Duffy positive hosts may act as a source of infection, increasing the opportunities for the parasite to infect hepatocytes of Duffy negative hosts and attempt erythrocyte invasion. The same heterogeneous landscape of Duffy phenotypes is present in Ethiopia, though no details were available about the prevalence of Duffy-independent transmission (Woldearegai *et al.*, 2011). Similarly, observations of two cases of Duffy-independent transmission were reported from Brazil (Cavasini *et al.*, 2007) where the Duffy landscape is similarly heterogeneous (national-level average Duffy phenotype prevalence: Fy(a+b+): 21.77%; Fy(a+b-): 36.54%; Fy(a-b+): 28.51%; Fy(a-b-): 13.18% (King *et al.*, 2011)). The same hypothesis cannot be applied to the other reports of Duffy-negative *P. vivax* infections in Africa however, as these are all in areas of extremely high Duffy negativity prevalence (94 to 98%; illustrated by the pie charts in Figure 3.4).

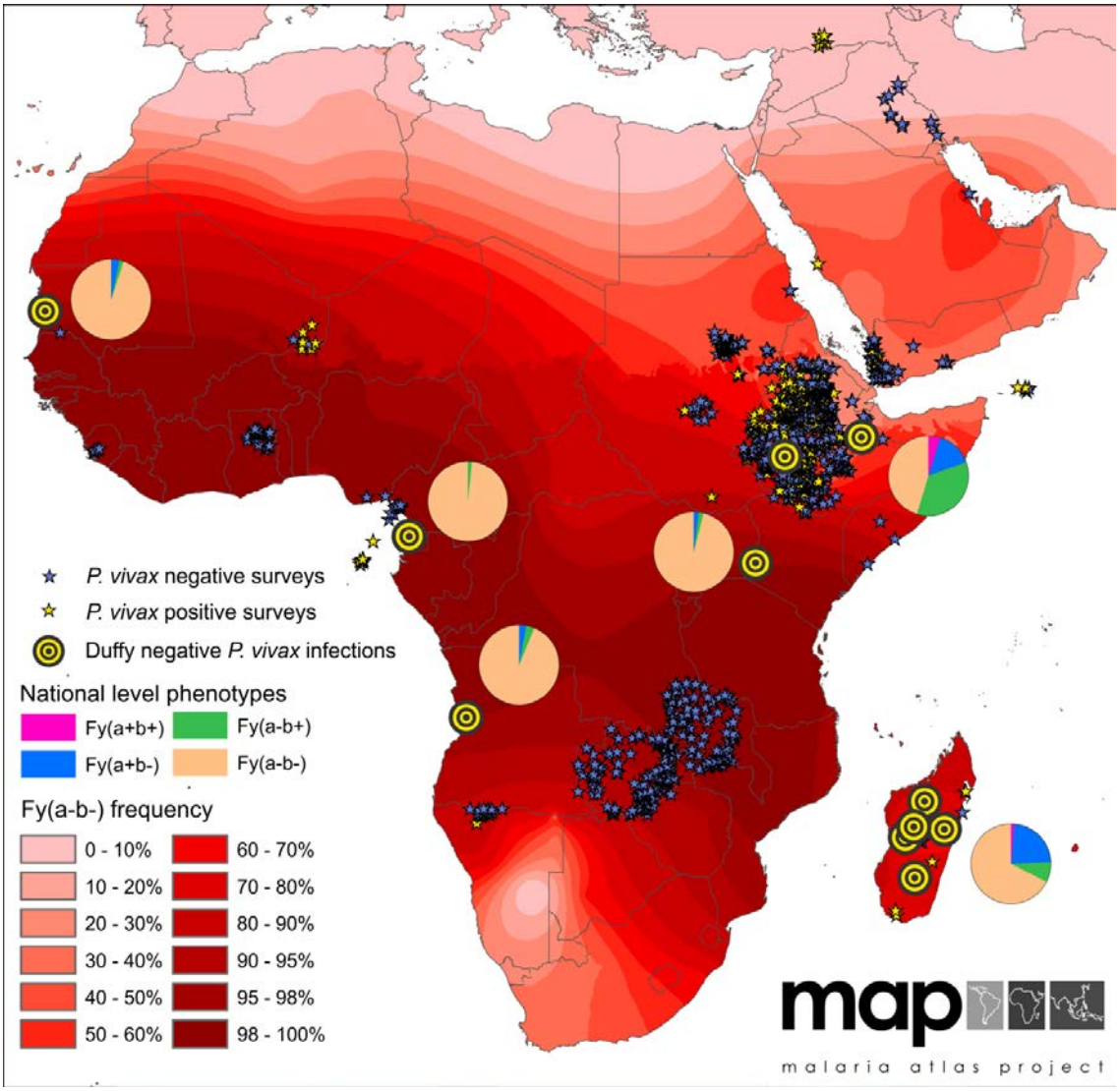


Figure 3.4. Observations of *P. vivax* transmission in Africa. Yellow bull's-eye icons represent reports of *P. vivax* infections in Duffy negative individuals: Angola (Mendes *et al.*, 2011), Equatorial Guinea (Mendes *et al.*, 2011), Ethiopia (Woldearegai *et al.*, 2011), Kenya (Ryan *et al*, 2006), Madagascar (Ménard *et al.*, 2010), Mauritania (Wurtz *et al.*, 2011); the pie-charts summarise the predicted prevalence of Duffy phenotypes in each country (King *et al.*, 2011). Yellow stars indicate locations of *P. vivax*-positive community surveys (n = 352) and blue stars *P. vivax*-negative surveys (n = 1,288) (data from Malaria Atlas Project used in *P. vivax* endemicity mapping (Gething *et al.*, 2012)). The background map is the predicted prevalence of Duffy negativity (Howes *et al.*, 2011).

Full genome sequencing of the parasites could allow some insight into the mechanisms enabling these infections. Limited genetic examination of the Malagasy data, however, provided no obvious clues from the parasite genome. Circumsporozoite (*Pv*CSP) protein and *P. vivax*-specific microsatellite analysis identified at least two strains able to infect Duffy negative cells, though with no significant differences at microsatellite loci between parasites isolated from Duffy negative or positive individuals (Ménard *et al.*, 2010). These results suggest that multiple *P. vivax* strains were able to infect cells with and without the Duffy receptor, and open the door to further investigation into the underlying mechanisms enabling Duffy-independent infection.

To assess the potential implications of widespread Duffy-independent transmission on the global distribution of *Pv*PAR, the Duffy negativity prevalence map was adjusted to different scenarios of Duffy-independent transmission. The estimates described in Section 3.4 earlier in this chapter assumed that the Duffy negative phenotype was 100% refractory to *P. vivax* infection (Guerra *et al.*, 2010; Gething *et al.*, 2012). Given that this assumption may – to an as yet unknown extent – be overly-conservative, the Duffy negativity map was used to make projections under different scenarios of *Pv*-protection to assess how different levels of protection from Duffy negativity would shift the global distribution of the *Pv*PAR (Figure 3.5; Zimmerman *et al.*, 2013). These figures indicate that even if the degree of protection afforded by Duffy negativity were only 50%, the overall global PAR of *P. vivax* would remain heavily focused in Asia given the high population density across this region. Not surprisingly, the most significant changes in *Pv*PAR would be in the Africa+ region, with a projected five-fold increase. Interestingly, this level of Duffy-independent infection would increase the *Pv*PAR in Africa to three-fold that of the Americas. Establishing where along this spectrum the true infection rate lies is necessary to more accurately estimate the population at risk of *P. vivax* and adjust control approaches accordingly. A series of malaria patient surveys diagnosing *P. vivax* infections with PCR methods, accompanied by PCR identification of Duffy expression would shed important light on this. It would be valuable to conduct these studies across a range of settings with different proportions of Duffy positive and negative individuals, and in areas at different levels of *P. vivax* endemicity.

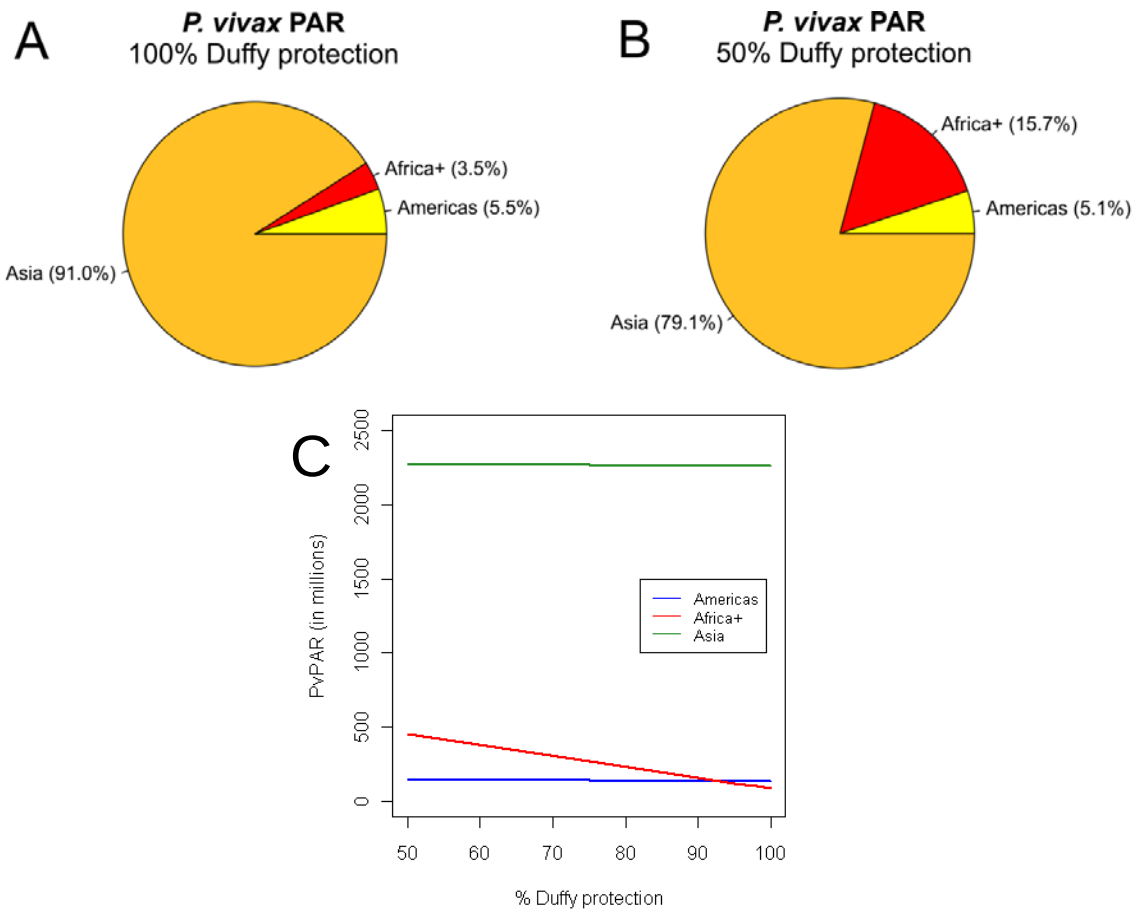


Figure 3.5. Scenarios of global *P. vivax* epidemiology under different frequencies of Duffy-independent transmission. Panels A and B represent the relative repartition of the *Pv*PAR across regions under different scenarios of protection conferred by Duffy negativity against infection (Zimmerman *et al.*, 2013). Panel C plots the absolute *Pv*PAR which would at risk under different levels of protection afforded by the Duffy negativity phenotype.

3.3. *P. vivax* binding affinity to Fy^a/Fy^b

3.3.1. Evidence of differential Duffy antigen binding

It is important to remember that the Duffy positive phenotype is not of a single type. The two dominant antigens, Fy^a and Fy^b, are distinguished by a single nucleotide polymorphism at locus 125 (G -> A) which encodes an amino acid substitution (Gly42Asp). *Plasmodium vivax* expresses a ligand, termed the Duffy binding protein (*Pv*DBP) which interacts with the host Duffy antigen to successfully invade host erythrocytes (Wertheimer and Barnwell, 1989; Singh *et al.*, 2005). It is the efficacy and specificity of this parasite-host interaction which defines the parasite's ability to establish blood-stage infections. A recent suite of

evidence demonstrated an increased binding affinity of the *Pv*DBP to the Fy^b antigen compared to Fy^a (King *et al.*, 2011). *In vitro* studies identified a 41-50% decreased binding affinity to Fy(a+b-) phenotypes relative to Fy(a-b+). This was reflected in a cohort study in the Brazilian Amazon which identified a 29-80% reduced risk of clinical *P. vivax* malaria in Fy(a+b-) individuals relative to Fy(a+b+) hosts. Further, Fy(a-b+) individuals (including heterozygotes) had 220-270% greater risk of *P. vivax* malaria compared to Fy(a+b+) individuals. No such effects were observed in relation to *P. falciparum*. The Fy^a antigen was therefore relatively more protective against clinical infection than its ancestral Fy^b form (Tournamille *et al.*, 2004).

3.3.2. Global spatial patterns and population estimates of dominant Duffy variant expression

A composite map of dominant Duffy alleles was generated to support these observations (Figure 3.6), and facilitate identification of the dominant Duffy variants between different populations. Areas where a single allele predominates ≥50% are shown in colour, while populations with heterogeneous variant expression are mapped in greyscale with darker areas corresponding to where phenotype diversity is greatest.

National-level population estimates of the frequency of each Duffy phenotype Fy(a+b+), Fy(a+b-), Fy(a-b+), Fy(a-b-) – were calculated to provide summary statistics of the dominant phenotypes across all countries. To generate fully additive population counts, the Bayesian model outputs (PPDs) described in Chapter 2 were summarised as the mean of the PPD. Allele frequencies were assumed to be in Hardy-Weinberg equilibrium and genotype frequencies calculated accordingly to generate the six possible allelic combinations which together sum to represent all combinations of expression at the population level (Equation 3.1):

$$1 = FY{*}A^2 + FY{*}B^2 + FY{*}B^{(ES)2} + 2(FY{*}A \times FY{*}B) + 2\,(FY{*}A \times FY{*}B^{ES}) + 2(FY{*}B \times FY{*}B^{ES})$$

(Equation 3.1)

These calculations were computed using the 10 × 10km gridded map surfaces in GIS software (ArcMap 9.3, ESRI Inc., Redlands, CA, USA). Phenotype surfaces were then derived by summing the surfaces where required (Table 3.2).

Genotype	Phenotype
*FY*A/FY*A* *FY*A/FY*B*^{ES}	Fy(a+b-)
*FY*B/FY*B* *FY*B/FY*B*^{ES}	Fy(a-b+)
*FY*A/FY*B*	Fy(a+b+)
*FY*B*^{ES}*/FY*B*^{ES}	Fy(a-b-)

Table 3.2. Duffy genotype and phenotype relationships.

To derive population estimates for each Duffy phenotype, a high resolution population density surface was combined with the Duffy phenotype maps at 1 × 1 km grid resolution. The *beta* version of the Global Rural Urban Mapping Project (GRUMP) gridded population database (http://sedac.ciesin.columbia.edu/gpw/) was projected to the year 2010 using separate urban and rural growth rates estimated by the 2007 United Nations World Urbanization Prospects (http://esa.un.org/unup/) (Balk *et al.*, 2006; Hay *et al.*, 2009). This population surface was overlaid on each of the Duffy phenotype frequency maps to derive gridded population counts for each phenotype, which were then aggregated by country to estimate the relative proportions of each phenotype (see Appendix).

3.3.3. Evolutionary significance of global Duffy variant distributions

Given the strong spatial gradients of *FY*A* and *FY*B* across Asia (Figure 3.4) and the reduced binding to Fy^{a} antigen, it may be that the *FY*A* variant has been under selection from *P. vivax* infection. Recent correlation analysis of *P. vivax* incidence in relation to Duffy allele frequencies across India, identified deviance from Hardy-Weinberg equilibrium leading authors to conclude that the *FY*A* allele was under strong positive natural selection (Chittoria *et al.*, 2012). While this study was very limited in its scope, using samples of only 250 individuals aggregated into seven zones across India, it considers interesting epidemiological questions about these interactions which demand further

examination. Knowledge of the selective power of *P. vivax* would lend further insight into the relative clinical significance of this parasite, the severity of which is likely to have been strongly underestimated over the past half-century (Price *et al.*, 2007; Baird, 2013).

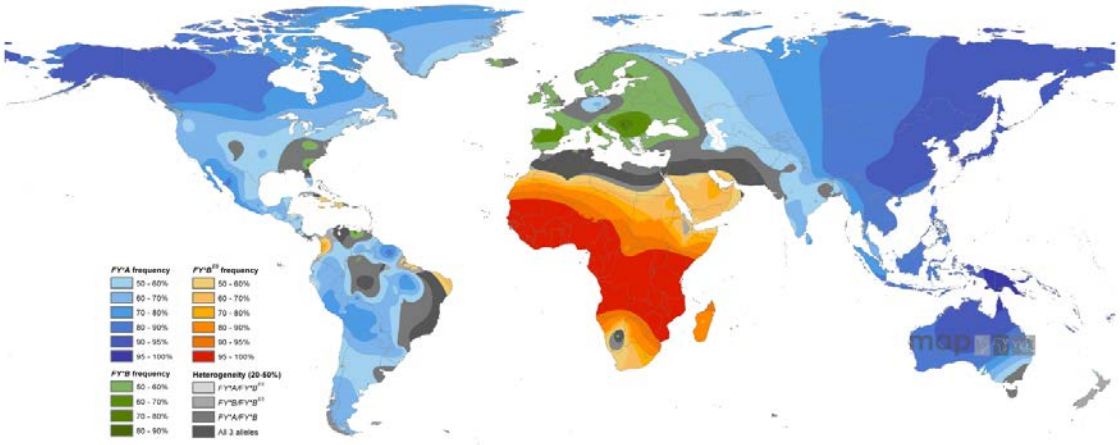


Figure 3.6. Composite map of dominant Duffy allele frequencies (>50%). Areas predominated by a single allele (frequency ≥50%) are represented by a colour gradient (blue: *FY*A*; green: *FY*B*; red/yellow: *FY*B*^{ES}). Areas of allelic heterogeneity where no single allele predominates, but two or more alleles each have frequencies ≥20%, are shown in grayscale: palest for heterogeneity between the silent *FY*B*^{ES} allele and either *FY*A* or *FY*B* (when co-inherited, these do not generate new phenotypes); and darkest being co-occurrence of all three alleles (and correspondingly the greatest genotypic and phenotypic diversity).

3.3.4. Significance of global Duffy variant distributions for vaccine development

The major implication of differing *Pv*DBP binding affinities for Fy^{a}/Fy^{b} antigens is for vaccine development. Understanding how to inhibit, disrupt or block the intimate *Pv*DBP-Fy$^{a/b}$ interaction creates potential strategies for a vaccine against blood-stage infection. Being usually so central to establishing infection, the *Pv*DBP is an important vaccine candidate. Antibodies against this molecule have been observed to prevent binding and cell invasion *in vitro*, and to protect against blood-stage infection (Michon *et al.*, 2000; Grimberg *et al.*, 2007; Nobrega de Sousa *et al.*, 2011; Ntumngia *et al.*, 2012). The investigations by King and colleagues identified differential binding of these inhibitory antibodies between variants, with greater blocking of *Pv*DBP to Fy^{a} than to Fy^{b} antigen

(King *et al.*, 2011). Artificially-induced antibodies were 200-300% more inhibitive against Fy^a homozygotes than Fy^b homozygotes. These findings therefore suggest that a *Pv*DBP vaccine would be more efficacious among populations where *FY*A* frequencies were higher. It will also be important to test *Pv*DBP-based vaccines in populations that carry combinations of both *FY*A* and *FY*B* alleles.

3.4. Concluding thoughts

It is evident that *P. vivax* is not absent from Africa. The degree to which it is present, however, and how its transmission is sustained, is not known. The implications of *P. vivax* transmission are critical to malaria epidemiology in Africa. From a control perspective, it has important implications for: burden, diagnostics, drug therapy, and prospects for future malaria epidemiology as *P. falciparum* endemicity drops. In many areas outside Africa, control interventions have successfully reduced *P. falciparum* endemicity with *P. vivax* concurrently growing in relative significance and presenting a much tougher challenge for control and elimination (Ministry of Health and Quality of Life Mauritius and the World Health Organization and the University of California San Francisco, 2012; Ministry of Health Sri Lanka and the World Health Organization and the University of California San Francisco, 2012; Shanks, 2012). Assessing the status of clinical *P. vivax* malaria in Africa must be a higher priority: both through increased awareness among clinicians and appropriate diagnoses of patients, as well as with targeted community screening with appropriate diagnostic tools. Understanding the mechanisms of Duffy-independent transmission in areas of differing Duffy blood group characteristics is vital to re-evaluating the public-health scale burden of this relapsing parasite.

The findings discussed in this chapter, however, should not be misinterpreted. Current epidemiological data overwhelmingly indicates that *P. falciparum* is the predominant malaria pathogen across most parts of sub-Saharan Africa (Culleton *et al.*, 2008; Gething *et al.*, 2011a; WHO, 2011). Available evidence suggests that, where present, *P. vivax* prevalence is low, and for now, Fy-DBP binding remains the only characterised invasion

pathway. Further, clinical *P. vivax* was significantly lower among Duffy negatives in Madagascar (Ménard *et al.*, 2010), suggesting a partial impairment of *P. vivax* invasion of Duffy negative cells, even if parasitaemia is present. However, the evidence-base for assessing the significance of *P. vivax* in relation to *P. falciparum* in Africa is poor, and the potential implications of Duffy-independent transmission, together with their natural history, are highly significant and justify further investigation.

I have described how the map of Duffy negativity has supported this evolving understanding of *P. vivax* epidemiology from several perspectives. Given the assumption that Duffy negativity is a protective phenotype against *P. vivax* infection, the map has been used to support current estimates of the limits and endemicity of *P. vivax* transmission, and the PAR of *P. vivax* infection; with the value of the Duffy negativity map increasing in areas where *P. vivax* prevalence data were scarce. Knowledge of the underlying Duffy blood group landscape will allow assessments of *P. vivax* transmission to be refined as additional data emerges quantifying the extent of Duffy-independent transmission, as well as the extent of transmission among Duffy positive hosts. While the main application of the Duffy maps discussed in this chapter has been in supplementing the scarce parasitological evidence-base in Africa, maps of Duffy positive variants have also been used alongside evidence of differential invasion of Fy^a- and Fy^b-positive cells to ascertain infection susceptibility and vaccine efficacy globally. Uncertainties around the evolutionary significance of the very clear spatial patterns in the human Duffy polymorphism – $FY*B^{ES}$ near-fixation in Africa, the ancestral *FY*B* allele predominance in Europe transitioning to near-fixation of *FY*A* in Asia – have been mentioned and must be further examined in relation to *P. vivax* phylogenetic studies (Carter, 2003; Rosenberg, 2007; Culleton and Carter, 2012). Recognition of the evolutionary significance of *P. vivax* as an agent of selection would provide insight into the clinical severity of the parasite as an agent of disease.

This chapter aimed to provide an overview of some scenarios in which an understanding of the human genetic landscape can support spatial epidemiological studies of *P. vivax* transmission. I have demonstrated the plausibility and additive value of bringing together maps of human genetic traits to help improve our understanding of infectious disease transmission potential. In the next chapters, I turn to consider how spatial maps of another human genetic polymorphism – glucose-6-phosphate dehydrogenase deficiency (G6PDd) – can support insights into the risks associated with *P. vivax* therapy.

3.5. References

Baird, J.K. (2013). Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. *Clinical Microbiology Reviews* **26**(1): 1-22.

Balk, D.L., Deichmann, U., Yetman, G., et al. (2006). Determining global population distribution: methods, applications and data. *Advances in Parasitology* **62**: 119-156.

Bicheron, P., Defourny, P., Brockmann, C., et al. (2008). GLOBCOVER: Products Description and Validation Report. Tolouse. MEDIAS-France.

Carter, R. (2003). Speculations on the origins of *Plasmodium vivax* malaria. *Trends in Parasitology* **19**(5): 214-219.

Cavasini, C.E., Mattos, L.C., Couto, A.A., et al. (2007). *Plasmodium vivax* infection among Duffy antigen-negative individuals from the Brazilian Amazon region: an exception? *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**(10): 1042-1044.

Chittoria, A., Mohanty, S., Jaiswal, Y.K., et al. (2012). Natural selection mediated association of the Duffy (FY) gene polymorphisms with *Plasmodium vivax* malaria in India. *PLoS One* **7**(9): e45219.

Culleton, R. and Carter, R. (2012). African *Plasmodium vivax*: Distribution and origins. *International Journal for Parasitology* **42**(12): 1091-1097.

Culleton, R., Ndounga, M., Zeyrek, F.Y., et al. (2009). Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. *Journal of Infectious Diseases* **200**(9): 1465-1469.

Culleton, R.L., Mita, T., Ndounga, M., et al. (2008). Failure to detect *Plasmodium vivax* in West and Central Africa by PCR species typing. *Malaria Journal* **7**: 174.

Gautret, P., Legros, F., Koulmann, P., et al. (2001). Imported *Plasmodium vivax* malaria in France: geographical origin and report of an atypical case acquired in Central or Western Africa. *Acta Tropica* **78**(2): 177-181.

Gething, P.W., Elyazar, I.R., Moyes, C.L., et al. (2012). A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Neglected Tropical Diseases* **6**(9): e1814.

Gething, P.W., Patil, A.P., Smith, D.L., et al. (2011a). A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal* **10**: 378.

Gething, P.W., Van Boeckel, T.P., Smith, D.L., et al. (2011b). Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasites & Vectors* **4**: 92.

Grimberg, B.T., Udomsangpetch, R., Xainli, J., et al. (2007). *Plasmodium vivax* invasion of human erythrocytes inhibited by antibodies directed against the Duffy binding protein. *PLoS Medicine* **4**(12): e337.

Guerra, C., Hay, S., Lucioparedes, L., et al. (2007). Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malaria Journal* **6**(1): 17.

Guerra, C.A., Howes, R.E., Patil, A.P., et al. (2010). The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Neglected Tropical Diseases* **4**(8): e774.

Hay, S.I., Guerra, C.A., Gething, P.W., et al. (2009). A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Medicine* **6**(3): e1000048.

Howes, R.E., Patil, A.P., Piel, F.B., et al. (2011). The global distribution of the Duffy blood group. *Nature Communications* **2**: 266.

King, C.L., Adams, J.H., Xianli, J., et al. (2011). Fy^a/Fy^b antigen polymorphism in human erythrocyte Duffy antigen affects susceptibility to *Plasmodium vivax* malaria. *Proceedings of the National Academy of Sciences of the United States of America* **108**(50): 20113-20118.

Ménard, D., Barnadas, C., Bouchier, C., et al. (2010). *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proceedings of the National Academy of Sciences of the United States of America* **107**(13): 5967-5971.

Mendes, C., Dias, F., Figueiredo, J., et al. (2011). Duffy negative antigen is no longer a barrier to *Plasmodium vivax*--molecular evidences from the African West Coast (Angola and Equatorial Guinea). *PLoS Neglected Tropical Diseases* **5**(6): e1192.

Michon, P., Fraser, T. and Adams, J.H. (2000). Naturally acquired and vaccine-elicited antibodies block erythrocyte cytoadherence of the *Plasmodium vivax* Duffy binding protein. *Infection and Immunity* **68**(6): 3164-3171.

Miller, L.H., Mason, S.J., Clyde, D.F., et al. (1976). The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *New England Journal of Medicine* **295**(6): 302-304.

Miller, L.H., Mason, S.J., Dvorak, J.A., et al. (1975). Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* **189**(4202): 561-563.

Ministry of Health and Quality of Life Mauritius and the World Health Organization and the University of California San Francisco (2012). Eliminating Malaria: Case-study 4 | Preventing reintroduction in Mauritius. Geneva, The World Health Organization.

Ministry of Health Sri Lanka and the World Health Organization and the University of California San Francisco (2012). Eliminating Malaria: Case-study 3 | Progress towards elimination in Sri Lanka. Geneva, The World Health Organization.

Moonen, B., Cohen, J.M., Tatem, A.J., et al. (2010). A framework for assessing the feasibility of malaria elimination. *Malaria Journal* **9**: 322.

Muhlberger, N., Jelinek, T., Gascon, J., et al. (2004). Epidemiology and clinical features of vivax malaria imported to Europe: sentinel surveillance data from TropNetEurop. *Malaria Journal* **3**: 5.

Nobrega de Sousa, T., Carvalho, L.H. and Alves de Brito, C.F. (2011). Worldwide genetic variability of the Duffy binding protein: insights into *Plasmodium vivax* vaccine development. *PLoS One* **6**(8): e22944.

Ntumngia, F.B., King, C.L. and Adams, J.H. (2012). Finding the sweet spots of inhibition: Understanding the targets of a functional antibody against *Plasmodium vivax* Duffy binding protein. *International Journal for Parasitology*.

Patil, A.P., Gething, P.W., Piel, F.B., et al. (2011). Bayesian geostatistics in health cartography: the perspective of malaria. *Trends in Parasitology* **27**(6): 246-253.

Patil, A.P., Okiro, E.A., Gething, P.W., et al. (2009). Defining the relationship between *Plasmodium falciparum* parasite rate and clinical disease: statistical models for disease burden estimation. *Malaria Journal* **8**: 186.

Piel, F.B., Patil, A.P., Howes, R.E., et al. (2012). Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet*.

Price, R.N., Tjitra, E., Guerra, C.A., et al. (2007). Vivax malaria: neglected and not benign. *American Journal of Tropical Medicine and Hygiene* **77**(Suppl 6): 79-87.

Rosenberg, R. (2007). *Plasmodium vivax* in Africa: hidden in plain sight? *Trends in Parasitology* **23**(5): 193-196.

Rubio, J.M., Benito, A., Roche, J., et al. (1999). Semi-nested, multiplex polymerase chain reaction for detection of human malaria parasites and evidence of *Plasmodium vivax* infection in Equatorial Guinea. *American Journal of Tropical Medicine and Hygiene* **60**(2): 183-187.

Ryan, J.R., Stoute, J.A., Amon, J., et al. (2006). Evidence for transmission of *Plasmodium vivax* among a duffy antigen negative population in Western Kenya. *American Journal of Tropical Medicine and Hygiene* **75**(4): 575-581.

Shanks, G.D. (2012). Control and elimination of *Plasmodium vivax*. *Advances in Parasitology* **80**: 297-337.

Singh, A.P., Ozwara, H., Kocken, C.H., et al. (2005). Targeted deletion of *Plasmodium knowlesi* Duffy binding protein confirms its role in junction formation during invasion. *Molecular Microbiology* **55**(6): 1925-1934.

Sinka, M.E., Bangs, M.J., Manguin, S., et al. (2011). The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis. *Parasites & Vectors* **4**: 89.

Tatem, A.J., Smith, D.L., Gething, P.W., et al. (2010). Ranking of elimination feasibility between malaria-endemic countries. *Lancet* **376**(9752): 1579-1591.

Tournamille, C., Blancher, A., Le Van Kim, C., et al. (2004). Sequence, evolution and ligand binding properties of mammalian Duffy antigen/receptor for chemokines. *Immunogenetics* **55**(10): 682-694.

U.N.P.D. (2007). World urbanization prospects: population database. New York, United Nations Population Division (U.N.D.P.).

Wertheimer, S.P. and Barnwell, J.W. (1989). *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. *Experimental Parasitology* **69**(4): 340-350.

WHO (2011). World Malaria Report 2011. Geneva.

Woldearegai, T.G., Kremsner, P.G. and Kun, J.F.J. (2011). *P. vivax* infection in Duffy-negative individuals in Ethiopia: Indications against an old paradigm. *American Journal of Tropical Medicine and Hygiene* **85**(Suppl 452-A-516): S463.

Wurtz, N., Mint Lekweiry, K., Bogreau, H., et al. (2011). Vivax malaria in Mauritania includes infection of a Duffy-negative individual. *Malaria Journal* **10**: 336.

Zimmerman, P.A., Ferreira, M.U., Howes, R.E., et al. (2013). Red blood cell polymorphism and susceptibility to *Plasmodium vivax*. *Advances in Parasitology* **81**: (in press).

Chapter 4 – G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map

Following on from the discussions of *Plasmodium vivax* transmission and the role of the Duffy antigen in supporting this, I now turn to consider the treatment options for this very widespread parasite.

As discussed in Chapter 1, therapeutic options against the relapsing forms of *P. vivax* are limited to a single drug. The use of this drug, however, is severely constrained by a human genetic predisposition to haemolysis triggered by this drug. No evidence-base currently exists to assess the spatial prevalence and characteristics of this human disorder, glucose-6-phosphate dehydrogenase deficiency (G6PDd), in spite of its far-reaching significance for the treatment of *P. vivax*. A trio of chapters discusses these issues. First, Chapter 4 maps the prevalence of phenotypic G6PDd across malaria endemic countries. Second, Chapter 5 considers the spatial patterns of the genetic variability and clinical severity of this disorder. Finally, Chapter 6 brings these together into a framework assessing the relative risks of haemolysis due to G6PDd associated with using primaquine.

This first chapter has been published in *PLoS Medicine* and is included here in its final form. Additional information relating to the methods and results of this study is included in the Appendix.

OPEN ACCESS Freely available online

PLOS | MEDICINE

G6PD Deficiency Prevalence and Estimates of Affected Populations in Malaria Endemic Countries: A Geostatistical Model-Based Map

Rosalind E. Howes[1]*, Frédéric B. Piel[1], Anand P. Patil[1], Oscar A. Nyangiri[2], Peter W. Gething[1], Mewahyu Dewi[3], Mariana M. Hogg[1], Katherine E. Battle[1], Carmencita D. Padilla[4,5], J. Kevin Baird[3,6], Simon I. Hay[1]*

1 Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, United Kingdom, **2** Kenya Medical Research Institute/Wellcome Trust Programme, Centre for Geographic Medicine Research-Coast, Kilifi District Hospital, Kilifi, Kenya, **3** Eijkman-Oxford Clinical Research Unit, Jakarta, Indonesia, **4** Department of Pediatrics, College of Medicine, University of the Philippines Manila, Manila, Philippines, **5** Newborn Screening Reference Center, National Institutes of Health (Philippines), Ermita, Manila, Philippines, **6** Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom

Abstract

Background: Primaquine is a key drug for malaria elimination. In addition to being the only drug active against the dormant relapsing forms of *Plasmodium vivax*, primaquine is the sole effective treatment of infectious *P. falciparum* gametocytes, and may interrupt transmission and help contain the spread of artemisinin resistance. However, primaquine can trigger haemolysis in patients with a deficiency in glucose-6-phosphate dehydrogenase (G6PDd). Poor information is available about the distribution of individuals at risk of primaquine-induced haemolysis. We present a continuous evidence-based prevalence map of G6PDd and estimates of affected populations, together with a national index of relative haemolytic risk.

Methods and Findings: Representative community surveys of phenotypic G6PDd prevalence were identified for 1,734 spatially unique sites. These surveys formed the evidence-base for a Bayesian geostatistical model adapted to the gene's X-linked inheritance, which predicted a G6PDd allele frequency map across malaria endemic countries (MECs) and generated population-weighted estimates of affected populations. Highest median prevalence (peaking at 32.5%) was predicted across sub-Saharan Africa and the Arabian Peninsula. Although G6PDd prevalence was generally lower across central and southeast Asia, rarely exceeding 20%, the majority of G6PDd individuals (67.5% median estimate) were from Asian countries. We estimated a G6PDd allele frequency of 8.0% (interquartile range: 7.4–8.8) across MECs, and 5.3% (4.4–6.7) within malaria-eliminating countries. The reliability of the map is contingent on the underlying data informing the model; population heterogeneity can only be represented by the available surveys, and important weaknesses exist in the map across data-sparse regions. Uncertainty metrics are used to quantify some aspects of these limitations in the map. Finally, we assembled a database of G6PDd variant occurrences to inform a national-level index of relative G6PDd haemolytic risk. Asian countries, where variants were most severe, had the highest relative risks from G6PDd.

Conclusions: G6PDd is widespread and spatially heterogeneous across most MECs where primaquine would be valuable for malaria control and elimination. The maps and population estimates presented here reflect potential risk of primaquine-associated harm. In the absence of non-toxic alternatives to primaquine, these results represent additional evidence to help inform safe use of this valuable, yet dangerous, component of the malaria-elimination toolkit.

Please see later in the article for the Editors' Summary.

Citation: Howes RE, Piel FB, Patil AP, Nyangiri OA, Gething PW, et al. (2012) G6PD Deficiency Prevalence and Estimates of Affected Populations in Malaria Endemic Countries: A Geostatistical Model-Based Map. PLoS Med 9(11): e1001339. doi:10.1371/journal.pmed.1001339

Academic Editor: Lorenz von Seidlein, Menzies School of Health Research, Australia

Received February 22, 2012; **Accepted** October 4, 2012; **Published** November 13, 2012

Funding: This work was supported by a Wellcome Trust Biomedical Resources Grant (#085406), which funded REH, FBP, OAN, and MMH; SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066) that also supports PWG and KEB; APP was funded by a Biomedical Resources Grant from the Wellcome Trust (#091835). MD is funded by the Oxford University-Li Ka Shing Foundation Global Health Programme. This work forms part of the output of the Malaria Atlas Project (MAP, http://www.map.ox.ac.uk/), principally funded by the Wellcome Trust, UK. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: G6PDd, glucose-6-phosphate dehydrogenase deficiency; GRUMP, Global Rural-Urban Mapping Project; IQR, interquartile range; MEC, malaria endemic country; PPD, posterior predictive distribution; UN, United Nations; WHO, World Health Organization.

* E-mail: rosalind.howes@zoo.ox.ac.uk (REH); simon.hay@zoo.ox.ac.uk (SIH)

Introduction

A third of malaria endemic countries (MECs, 35/99) now plan for malaria elimination [1–3]. This strategy is very distinct from routine malaria control, requiring not only the reduction of clinical burden, but complete depletion of the parasite reservoir by attacking the gametocytes responsible for transmission and killing the silent hypnozoites that may otherwise relapse [4–8]. Primaquine, an 8-aminoquinoline, is the only drug available for each of those therapeutic compartments [9,10], and is thus key to any elimination strategy [11]. However, this drug can also be dangerously toxic to individuals with a genetic deficiency in glucose-6-phosphate dehydrogenase (G6PDd), usually a clinically silent condition [12]. Tafenoquine (GSK) is a new drug in phase IIb/III clinical trials intended to replace primaquine, but is likely to retain haemolytic toxicity in G6PDd patients [13]. No alternative non-toxic drugs with these unique modes of action are currently close to clinical trials [8].

The 2010 World Health Organization (WHO) guidelines for uncomplicated *P. falciparum* malaria treatment recommend a single dose of primaquine alongside artemisinin-based combination therapy (ACT) to prevent parasite transmission, particularly as a component of pre-elimination or elimination programmes [14,15] and as part of artemisinin resistance containment programmes [16]. This gametocytocidal therapy has been shown to be effective in low endemicity settings in combination with an ACT [17], and in theory could significantly reduce transmission levels [18]. However, evidence for a derived community benefit is poor and a recent Cochrane review finds little support for these WHO treatment guidelines [19]. Transmission may be sustained by submicroscopic gametocyte levels [7,20], meaning that effective blocking of community transmission may require wider drug administration beyond symptomatic cases [4,21].

Key to sustaining progress towards malaria elimination is the prevention of parasite reintroduction from the relapsing malarias *P. vivax* and *P. ovale* [8]. This therapeutic target is complicated by the absence of diagnostic testing for liver-stage parasites [22], and recent studies suggest high prevalence of hypnozoites, even in areas considered to have relatively low transmission intensity [23]. Although recommended dosages vary regionally, 14-d regimens of primaquine (either 15 or 30 mg daily adult doses) are advised for successful hypnozoite treatment [14]. The key impediment to attacking hypnozoite reservoirs among endemic populations in this way is the risk of potential harm from primaquine [24].

Primaquine can cause mild to severe haemolysis in G6PDd patients. The mechanism of primaquine-induced haemolysis is not fully understood. Reduced G6PD enzyme activity levels are likely to create a redox equilibrium within red blood cells that favours oxidised species of highly reactive primaquine metabolites. In one hypothesis, the 5-hydroxyprimaquine metabolite would be dominated by its oxidised quinoneimine species in G6PDd red blood cells, which may then react with the haem moiety of haemoglobin and cause its displacement to the lipid bilayer of red blood cells [25]. The resulting acute intravascular haemolysis may be mild and self-limiting, or very severe and threaten life [26,27]. Freely circulating haemoglobin may cause the most severe clinical symptoms, such as renal failure [28]. There is currently no practical point-of-care field test for G6PDd [29], leaving most primaquine treatment decisions blind to haemolytic risk. There is a difficult ethical balance for weighing the benefits of transmission reduction and relapse prevention against poorly defined haemolytic risks [24].

Understanding the distribution and prevalence of this genetic risk factor in any given area may substantially inform risk and thus better equip policy makers and practitioners alike in designing and implementing primaquine treatment practices. We respond here to demands from the malaria community for a prevalence map of this genetic condition [22,30]. Existing published maps of G6PDd have important limitations. They either present average frequency data summarised to national levels thereby masking sub-national variation [31,32] and enabling mapping only for countries from where surveys were identified, leaving gaps in the maps [33]; or use broad categorical classes to present basic data extrapolation [34]. None exclude potentially skewed or unrepresentative survey samples (such as malaria patients), none consider prevalence in females, none have a framework for assessing statistical uncertainty, and none have mechanisms for incorporating G6PDd spatial heterogeneity into population affected estimates.

In addition to the public health importance of G6PDd in the context of malaria elimination, the clinical burden of this genetic condition includes a range of haematological conditions, including neonatal jaundice and acute haemolytic anaemia in adults triggered by a range of foods, infections, and other drugs [26,35]. Across the Asia-Pacific region, risk of neonatal complications due to G6PDd already justifies significant investment through inclusion in neonatal screening programmes in Malaysia, the Philippines, Taiwan, and Hong Kong [35].

In this study, we compile data from available sources of G6PDd prevalence surveys, and use these as the evidence-base to inform a Bayesian geostatistical model specifically adapted to the gene's X-chromosome inheritance mechanism. This model generates spatially continuous G6PDd prevalence predictions, and allows quantification of prediction uncertainty. The model predictions are then matched with high-resolution population data to estimate numbers of deficient individuals within MECs, accounting for the predicted sub-national heterogeneity in deficiency rates. Finally, we assemble a database of G6PDd variant occurrences and propose here an index for how the prevalence map could be used to stratify haemolytic risk at the national level.

Methods

This study's methodological objectives involved the assembly of representative community G6PDd prevalence surveys and the development of a Bayesian geostatistical model used to derive (i) maps of G6PDd prevalence within MECs, (ii) sex-specific estimates of the populations affected by this deficiency, and (iii) associated uncertainty metrics. These results were then combined with information on the distribution of the underlying G6PDd variants to generate an index for stratifying haemolytic risk from G6PDd. Each of these aspects is discussed briefly here, and in more detail in Protocols S1, S2, S3, S4, S5, S6. A schematic overview of the methodology is given in Figure 1.

Prevalence Survey Database Assembly and Inclusion Criteria

A literature search of online bibliographic databases was conducted to identify published community surveys of G6PDd. Existing databases published by Singh et al. in 1973 [36], Mourant et al. in 1976 [37], Livingstone in 1985 [38], and Nkhoma et al. in 2009 [33] were reviewed for any further sources. Direct contact with national screening programmes and researchers in the field was also undertaken to identify additional unpublished data. All identified surveys were reviewed for suitability for informing the G6PDd prevalence mapping analysis (Protocol S1).

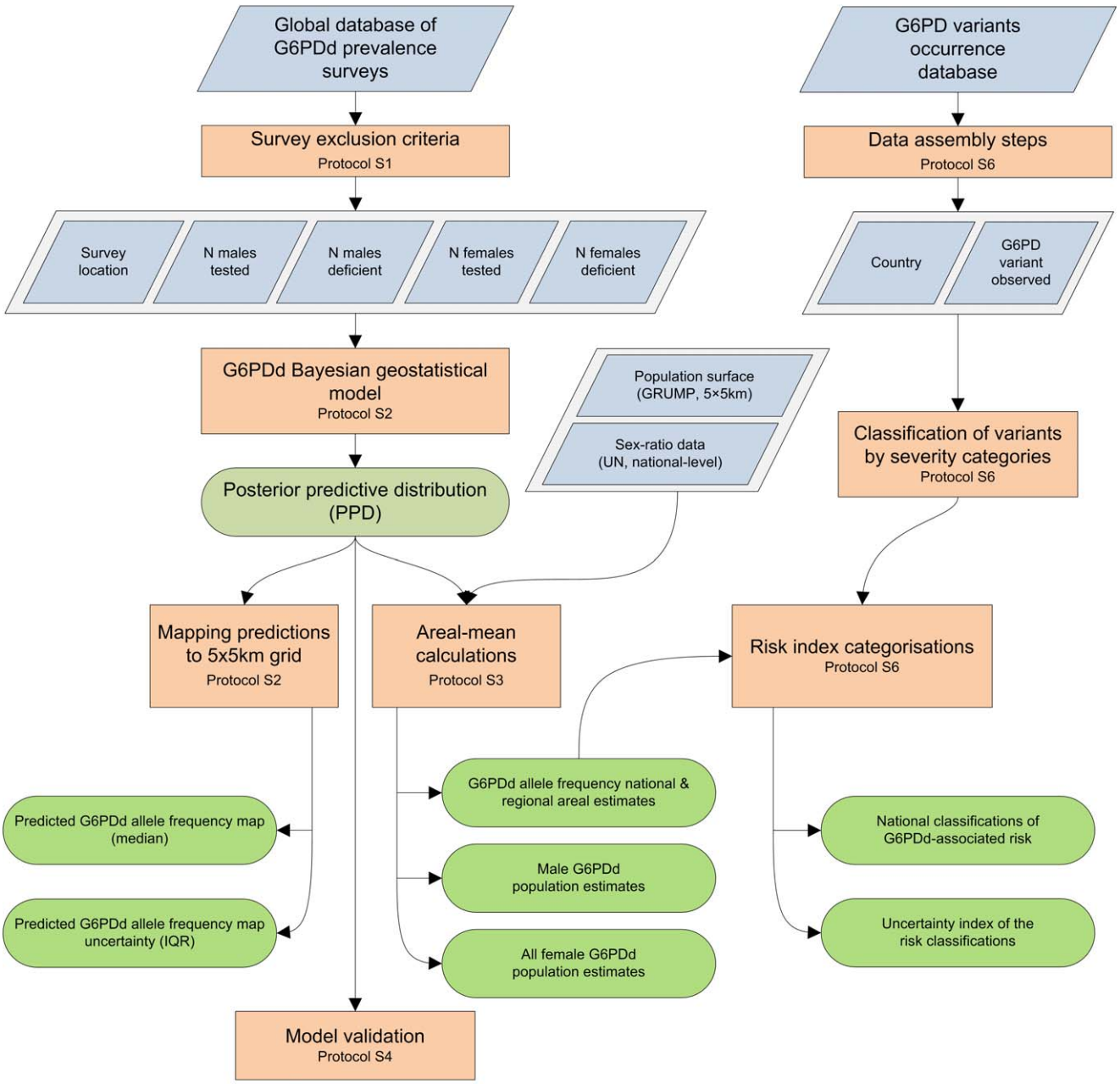


Figure 1. Schematic overview of the procedures and model outputs. Blue diamonds describe input data. Orange boxes denote data selection methods and analytical models. Green rods indicate model outputs.
doi:10.1371/journal.pmed.1001339.g001

Inclusion criteria were applied to ensure: (i) community representativeness: all potentially biased samples were excluded (e.g., any patient groups including malaria patients, ethnically selected samples, and family-based studies); (ii) gender representativeness: only surveys reporting sex-specific raw data were included; (iii) spatial representativeness: only surveys that could be mapped with relatively confined extents ($\leq$3,867 km^2) were included to ensure that sub-national variation could be represented [39,40]; (iv) clinically significant deficiency: only phenotypic diagnoses were considered. Because of the narrow range of primers usually used in molecular investigations, DNA-based diagnoses were excluded as they are susceptible to underestimating deficiency rates (Protocol S1) [24]. This study focused on G6PDd prevalence within MECs (corresponding to 99 countries, as defined in Protocol S1.5), with a particular focus on countries eliminating malaria (35 countries), but imposed no spatial restrictions to the dataset in order to make maximal use of existing information, particularly around the edges of the MEC limits.

The WHO uses mild and severe categorisations for G6PDd [31], with different treatment recommendations for each in relation to primaquine regimens [14]. Only through specific individual level G6PD testing can these be differentiated. The community level G6PD deficiency map presented here represents the prevalence of all clinically significant enzyme deficiency, as

would be diagnosed by the common phenotypic diagnostic tests. Additional resolution into the severity of the deficiency is derived from the G6PDd variant database described below.

The Model

A Bayesian geostatistical framework [41–46] was adopted to model the global prevalence of G6PDd. This framework used the evidence-base of surveys to generate predictions for G6PDd frequencies across the MECs, together with quantified uncertainty estimates for the predictions. This framework, developed for mapping the prevalence of a range of inherited blood disorders [40,47,48] was adapted to the X-linked inheritance mechanism of the G6PD gene [12]. Unlike females who have two copies, males inherit only a single copy of the G6PD gene, thus frequencies of deficiency in males correspond to the population-level allele frequency. Assuming populations to be at Hardy-Weinberg equilibrium [49,50], squaring the deficiency allele frequency (q) gives an estimate of the expected prevalence of homozygous females (q^2). Phenotypic expression of female heterozygous ($2q(1-q)$) deficiency ranges across a spectrum of enzyme activity levels. Expression is variable due to irregular Lyonization rates [51] and inconsistent cut-off points of phenotypic diagnostic methods (Protocols S1, S2, and S5). Thus only a proportion of heterozygotes are diagnosed as phenotypically "deficient" [51,52]. As no clear genotype-phenotype relationship could be identified from the observed survey data (Protocol S5), the model was given the flexibility to determine this relationship empirically, directly from the input data. The deviance of expected genetic heterozygotes from observed phenotypic deficiency cases (h) varied between surveys; h was modelled as a spatial variable, with values learned from the data, but not modelled as a spatially structured variable. The deviance value represents both the proportion of heterozygotes diagnosed as phenotypically normal, as well as actual deviance from expected Hardy-Weinberg equilibrium due to factors such as selection, consanguinity, migration, or small population sizes.

The model framework is thus $p(d) = q+q^2+2q(1-q)h$; where $p(d)$ is the probability of an individual being phenotypically deficient, and q is the allele frequency for deficiency. From this equation, frequencies of hemizygotes (males, q), homozygotes (females, q^2), and all deficient females (homozygotes and phenotypically deficient heterozygotes: $q^2+2a(1-q)h$ could be estimated. The model was fitted to the data and 1 million Markov chain Monte Carlo (MCMC) iterations [53] were used to generate full posterior predictive distributions (PPDs). The PPDs are summarised by the median value of the predictions and mapped continuously at 5×5 km resolution. Prediction uncertainty was quantified as the interquartile range (IQR) of the PPD. The model and its implementation are fully described in Protocol S2.

To validate the model predictions, an independent model iteration was implemented with a 95% subset of the dataset, allowing comparison of the predicted frequencies with observed frequencies from the 5% hold-out data. The hold-out data sample was selected to preferentially include spatially isolated data points, so as to ensure that the full prediction surface was included in the validation. Moreover, isolated areas are harder to make predictions for, and are therefore a conservative assessment of model reliability. Further details about validation methodology and derived statistics are given in Protocol S3.

Estimating Populations Affected

To quantify the prevalence of G6PDd across national and regional populations, areal estimates (regional aggregates that account for uncertainty) [47] were calculated by relating the model predictions to high resolution population density data from the Global Rural Urban Mapping Project (GRUMP) *beta* version, adjusted to United Nations (UN) population estimates for the year 2010 [41,54]. The areal-prediction model [47] was implemented to repeatedly sample G6PDd PPDs from selected locations, weighted according to population density, at a 5×5 km resolution. So, for each area of interest, the model generated an areal frequency PPD adjusted to the population density distribution across the area of interest. Multiplying the resulting aggregated G6PDd frequencies from the areal PPDs by UN 2010 national level population data adjusted for national-level sex ratio [55] gave estimates of the population numbers affected by each phenotype. To account for the stochasticity of the sampling, this process was repeated ten times for the national estimates, and five times for aggregated regional estimates (because of computational constraints) in order to calculate the Monte Carlo standard error associated with the estimates. This process is fully described in Protocol S4.

Stratifying National G6PDd Severity

In order to stratify the potential haemolytic risk associated with G6PDd, a simple index was developed that incorporated both the national prevalence of the trait and the severity of the local genetic variants.

Predicted national prevalence was stratified into three categories ($\leq$1%, $>$1–10%, and $>$10%). Stratifying the severity of the local forms of G6PDd was more involved. A second online literature review was conducted to assemble all reports of genetic and biochemical variants, using the same search methods as for assembling the prevalence data. All occurrences of named G6PDd variants were abstracted into a database and mapped to the country where they had been observed. Variants were then grouped according to their severity: the only severity classification widely applied to all variants is that proposed by Yoshida et al. [56], and endorsed by the WHO [31], which classifies variants according to their residual enzyme activity levels, their polymorphic/sporadic occurrence in populations, and the severity of their clinical symptoms (Protocol S6). Limitations to this classification system are reviewed in the Discussion. Only variants of class II (residual enzyme activity $<$10%) and class III (10%–60%) were relevant to this study. A score based on the relative composition of variants from these classes was assigned to each country to represent the relative proportions of class II and III variants: a proxy indicator of the severity of local variants. If no data were available from a country, a conservative approach was followed which took the highest score (most severe) from any neighbouring country.

The prevalence and variant severity scores were then multiplied to give a stratified measure of the relative haemolytic risk of G6PDd in each country. A similar uncertainty index was determined on the basis of the uncertainty in the prevalence estimates, and the availability and heterogeneity of variant data in each country. The variant data, risk scoring tables, and uncertainty estimates are presented in more detail in Protocol S6.

Results

The Prevalence Survey Database

Literature searches were conducted to collate all available reports of representative community G6PDd prevalence. A total of 17,272 G6PD abstracts were identified from online bibliographic databases, together with 472 potential data sources found in existing G6PDd databases [33,36–38] and unpublished reports. Following careful review, 1,601 abstracts were considered suitable

for our study and their full texts were reviewed for data. The Filipino Newborn Screening Reference Center (National Institutes of Health, Philippines) also contributed their universal screening results since 2004 to this study, adding 636 spatially unique locations to the database.

The total number of surveys identified that met the inclusion criteria was 1,734 globally, with 74% from MECs ($n = 1,289$) (Figure 2A). Surveys were unevenly distributed, some areas having been examined in micro-mapping studies (such as Sri Lanka) and universal screening (Philippines) while large extents of other areas remain unstudied (e.g., extensive parts of Indonesia, Madagascar, and central Africa). Within the MECs, 85% of surveys ($n = 1,101$) were from 23 Asian countries; 10% of surveys ($n = 132$) represented 23 African countries; data from only nine countries in the Americas were identified, corresponding to 4% of surveys ($n = 56$). Male data were reported from 99% of the surveys, while 62% presented female data. Overall numbers of individuals sampled were 2.4 million males and 2.0 million females.

The database is described in more detail in Protocols S1 and S5, with additional discussion about the influence of diagnostic methodology on test outcome in males and females. Female diagnosis is known to depend on numerous factors; however, in the absence of any standardised or established mathematical relationships for modelling the genotype-phenotype association in females, we decided to use the input dataset as the evidence-base, and the mapping model was given the freedom to determine this spatially variable relationship according to the raw data (Protocols S1, S2, S5).

G6PDd Prevalence Predictions: Overview

The survey database formed the evidence-base for the geostatistical model, which predicted both the spatially continuous map of G6PDd allele frequency (Figure 2) and the estimates of G6PDd populations (Figure 3); all model predictions are summarised with median values [53]. Model outputs indicated G6PDd to be widespread across malarious regions, with lowest frequencies in the Americas and highest in tropical Africa; an overall allele frequency of 8.0% (IQR: 7.4–8.8) was predicted across all MECs (Table 1). High population density in Asia meant that the highest numbers of G6PDd individuals were predicted to be from this continent (Table S1). The database and resulting model outputs indicated heterogeneity in G6PDd prevalence, with considerable variation across relatively short geographical distances in many areas (Figure 2B). All model predictions must be considered in relation to their associated uncertainty metrics (IQR; Figure 2C, Tables 1, S1 and S2). Model uncertainty is greatest where data points are scarce (Figure 2A) or where available data indicates heterogeneity (Protocol S2). Limitations to the database and the weaknesses that these lead to in the predictions are considered in the Discussion.

G6PDd Allele Frequency Map

Large swathes of the American MECs were predicted to have median G6PDd frequencies ≤1% (40.8% land area), with G6PDd being virtually absent from northern Mexico, Costa Rica, Peru, Bolivia, and much of Argentina (Figure 2). Prevalence increased towards coastal regions, peaking in Venezuela where the majority of the continent's predictions of >5% were located. Model uncertainty was relatively low across most of the Americas (IQR: <5%), with the IQR increasing to 5%–10% across the Amazon region where data were extremely scarce, and peaking between 15%–20% across Venezuela.

At the continental level, G6PDd was most prevalent across sub-Saharan Africa: 65.9% of the land area was predicted to have median G6PDd prevalence ≥5%, and 37.5% a median prevalence ≥10%. Predictions ranged from <1% at the continental extremities (western Sahel, Horn of Africa, and southern Africa) to >20% in isolated pockets of Sudan, coastal west Africa, and around the mouth of the river Congo. These broad patterns were interspersed with some striking sub-national variation within countries with deficiency hotspots, including Nigeria (range: 2% [IQR: 1–6] to 31% [22–42]), Sudan (1% [0–2] to 29% [19–41]) and Democratic Republic of Congo (DRC) (4% [1–11] to 32% [23–41]). These areas were also associated with the highest levels of model uncertainty—a reflection of this sub-national heterogeneity and also of the scarcity of input data from these areas. Highest prediction uncertainty across the continent was found in Sudan, Chad, and central Africa between DRC and Madagascar.

The highest median predicted prevalence of G6PDd across the entire MEC region was 32.5% in the Eastern Province of Saudi Arabia (specifically, around the urbanised coastal areas of Al-Qatif and Ad-Dammam). More broadly, rates across this disparately populated peninsula as a whole were heterogeneous, for example, dropping to prevalence of 3% (IQR: 2–4) in the central Al-Kharj and Riyadh area of Saudi Arabia. Further east, predicted prevalence remained high into southern Pakistan.This region had the highest uncertainty of the entire map (IQR exceeding 30%). No surveys were available from the south of Pakistan, and the closest neighbouring surveys in southern Iran, Oman, and western India reported prevalence of >20%, contrasting data from northern Pakistan. Prediction uncertainty dropped across central and southeast Asia, and predicted prevalence remained largely <10%, with three notable G6PDd prevalence hotspots in the central and southeast Asia regions peaking to >20%: (i) among the tribal, endogamous groups of Orissa province in east India, (ii) a patch along the northern Lao/Thai border, and (iii) much of the Solomon Islands archipelago. Underlying the broadly smooth continental-level variation, some areas were predicted to have highly heterogeneous sub-national G6PDd prevalence. Across Lao People's Democratic Republic (PDR), for instance, frequencies were predicted to range from 1% (IQR: 0–2) to 23% (16–32); predictions in Indonesia were from 0% (0–1) to 15% (10–21) in Nusa Tenggara; in Papua New Guinea, frequencies ranged from 1% (0–2) along the southern coast to 15% (10–22) along the East Sepik northern coast (Figure 2B–2C).

Validation Statistics

The predicted allele frequency surface was evaluated against a hold-out subset of the data selected with spatially declustered randomization that preferentially selected data sparse sites where model predictions would be inherently most difficult [41]. Differences between predicted and observed prevalence returned a mean error of 1.45% and a mean absolute error of 4.07%. These indicate a slight tendency of the model to overestimate prevalence, and relatively more substantial error in the magnitude of prediction precision. Full validation results are given in Protocol S3.

G6PDd Prevalence Predictions: Population Affected Estimates

The second modelling process related the allele frequency predictions to population distribution, generating sex-specific aggregated estimates of G6PDd populations, weighted by population distribution across the spatial regions of interest: national, malaria endemic, and the subset of 35 MECs targeting malaria elimination. These population-weighted estimates were modelled separately from the mapping process, and used the full model predictions, not just the summary median allele frequency map (Figure 2B). As with the map, these areal predictions and their

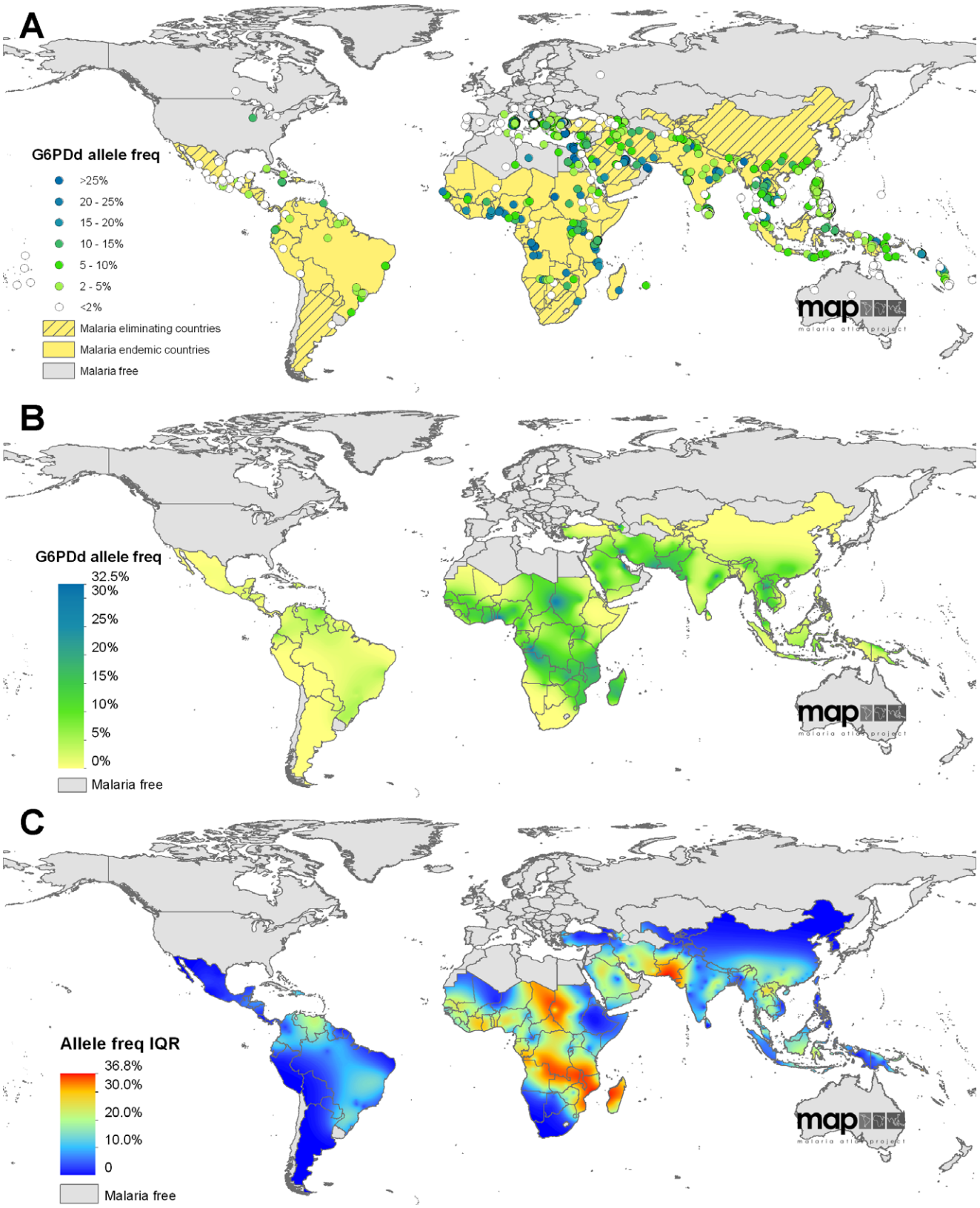


Figure 2. The global distribution of G6PDd. (A) shows the global assembly of G6PDd community surveys included in the model dataset; data points are coloured according to the reported prevalence of deficiency in males ($n = 1,720$). Background map colour indicates the national malaria status (malaria free/malaria endemic/malaria eliminating). (B) is the median predicted allele frequency map of G6PDd. (C) presents the associated prediction uncertainty metrics (IQR); highest uncertainty is shown in red and indicates where predictions are least precise.
doi:10.1371/journal.pmed.1001339.g002

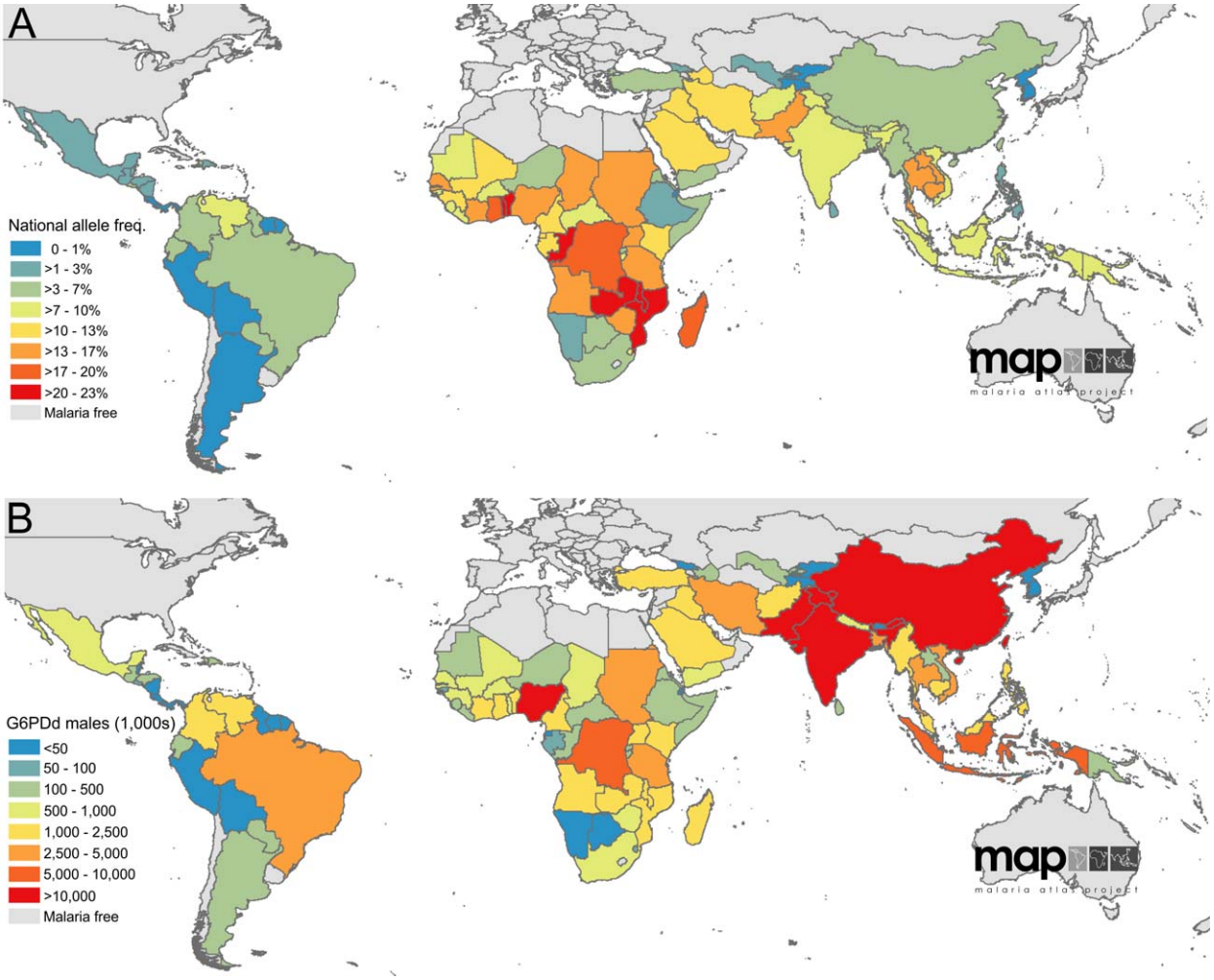


Figure 3. Population-weighted areal estimates of national G6PDd prevalence predictions. (A) summarises national-level allele frequencies, while (B) displays national-level population estimates of G6PDd males. Values are in thousands.
doi:10.1371/journal.pmed.1001339.g003

Table 1. G6PDd allele frequency and G6PDd population estimates across malaria endemic countries ($n = 99$) and the subset of malaria eliminating countries ($n = 35$).

G6PDd Allele Frequency and Population Estimates	Median (SE)		Q25(SE)		Q75(SE)	
	MEC[a]	Eliminating[b]	MEC[a]	Eliminating[b]	MEC[a]	Eliminating[b]
Allele frequency	8.04%(0.02%)	5.30%(0.01%)	7.44%(0.02%)	4.43%(0.02%)	8.81%(0.03%)	6.68%(0.02%)
G6PDd males	220,130(669)	61,227(96)	203,729(597)	51,200(184)	241,114(847)	77,223(251)
G6PDd females (homozygotes only[c])	17,115(n/a)	3,100(n/a)	(n/a)	(n/a)	(n/a)	(n/a)
G6PDd females (all females)	132,932(467)	35,205(71)	121,618(550)	28,862(96)	147,814(693)	45,608(144)

All figures are in thousands. Q25 and Q75 refer to the low and high limits of the IQR of the model predictions. Numbers in brackets represent the Monte Carlo standard error (SE) of the estimates; presented in the same units as the associated estimate. Full explanations are given in Protocol S4.
[a]Total regional male population: 2,736,515; Total regional female population: 2,644,975. Source: GRUMP-adjusted projected UN 2010 population estimates and sex-ratio data from UN World Population Prospects 2010 Revision.
[b]Total regional male population: 1,156,300; Total regional female population: 1,105,603. Source: GRUMP-adjusted projected UN 2010 population estimates and sex-ratio data from UN World Population Prospects 2010 Revision.
[c]Figures derived from the allele frequency estimates so do not have specific model-derived uncertainty metrics.
n/a, not available.
doi:10.1371/journal.pmed.1001339.t001

associated uncertainty were summarised with median and IQR values (Tables 1 and S1).

We estimated overall G6PDd allele frequency across MECs to be 8.0% (IQR: 7.4–8.8); using 2010 population data (Protocol S4), this corresponded to 220 million males (IQR: 203–241) and an estimated 133 million females (122–148), including 17 million homozygous females (assuming Hardy-Weinberg equilibrium). Across the subset of malaria eliminating countries (Figure 1), prevalence was lower, at 5.3% (4.4–6.7). Population estimates for 2010 across this subset of eliminating countries were 61 million G6PDd males (51–77) and an expected 35 million G6PDd females (29–46), including 3 million homozygous females.

National frequency estimates ranged from 0.1% in Cape Verde (IQR: 0.0–0.5) and the Democratic People's Republic of Korea (0.0–0.4) to 22.3% in the Solomon Islands (15.7–30.9), 22.5% in the Congo (17.3–29.6) and 23.0% in Benin (17.0–30.1). Reflecting the prevalence map, national allele frequency estimates were generally lowest in the Americas and highest in Africa (Figure 3A). Converting these national-level allele frequency estimates to G6PDd population numbers (G6PDd males; Figure 3B), however, shifts attention away from Africa towards the highly populous Asian countries, notably China and India where 41.3% of G6PDd males within MECs were predicted to be. Overall, the Americas contributed only 4.5% of the MEC G6PDd male population, sub-Saharan Africa 28.0%, and Asia an estimated 67.5%.

Index of National G6PDd Severity

Data searches for reports of G6PDd variants identified 527 occurrences of class II variants and 405 class III variants from a total of 54 countries out of 99 MECs (Table S3). Occurrences of these data points were used to score the severity of the overall composition of variants in each country, with scores inferred from neighbouring countries in instances where no data points had been reported (Figure 4A). Once combined with a rank of G6PDd prevalence, an overall score of the severity of risk from G6PDd was derived for each country (Figures 4B–4C). A similar scoring was used to determine the relative confidence in the severity scores, shown in Figures 4D–4E. Further figures and the table of all variant occurrences by country are given in Protocol S6 and Table S3.

This index of risk is predicated on the current state of knowledge of G6PDd variant occurrence and the relationship between variants and haemolysis, as outlined in the Discussion. From the present dataset, we see strong regional patterns in the distribution of variants, with sub-Saharan Africa being predominantly ranked as having mildly severe variants (class III), predominantly A−, though some class II variants were reported from Sudan and South Africa, and Senegal and the Gambia in west Africa (Table S3). Relatively few data were available from the Americas, but these included a greater diversity of variants including a minority of class II variants. In contrast, variant reports were more heterogeneous across Asia, a majority of which were class II (most commonly Mediterranean, then Canton and Kaiping), though certain class III variants were also widely reported (Mahidol, then Chinese-5 and Gaohe being most frequently identified); the predominance of class II variants put the classification of all Asian countries as having severe variants.

Combining these variant severity scores with the scores of G6PDd prevalence gave an index of overall risk from G6PDd for each MEC. Greatest haemolytic risk from G6PDd was found in the Arabian Peninsula and across west Asia, where both prevalence and variant severity (dominated by the class II Mediterranean variant) were high. Across the Asian continent, risk remained high (level 5 of 6, increasing to level 6 in the Mekong region where prevalence was at its highest). In contrast, despite high prevalence, the low severity of the variants reported from sub-Saharan Africa resulted in the lowest risk categorizations from G6PDd globally, which was a moderate risk (mostly levels 2 to 3 of 6, though increasing to level 5 in countries where class II variants had been reported).

The uncertainty inherent in this synthesis is considerable; however, the index indicated that according to the metrics employed in this study, uncertainty ranked highest in many sub-Saharan countries and most countries in the Americas (where 19 of 21 countries had uncertainty ranked 5–6 out of 6). Further data from these regions would substantially improve reliability both of the modelled prevalence predictions, as well as of the variant severity categorisations, many of which had to be inferred from neighbouring countries.

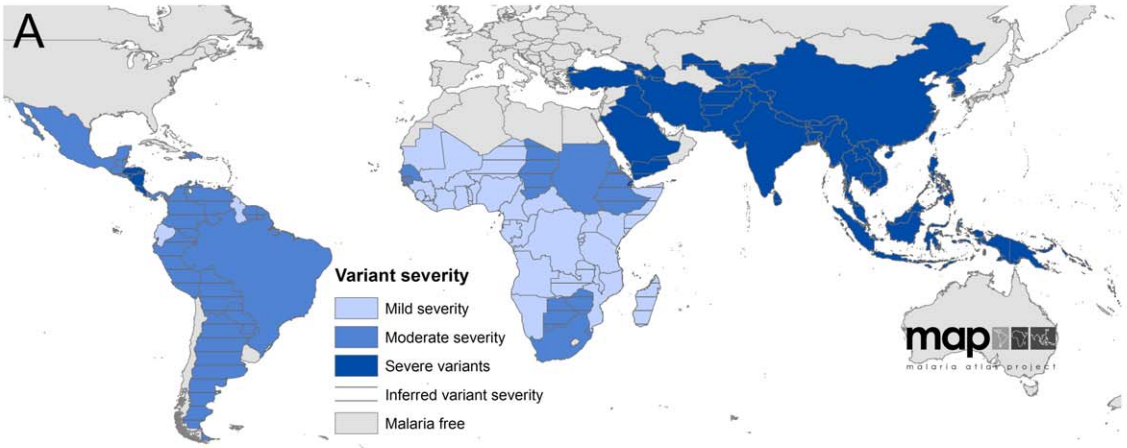
The framework proposed here can be updated and refined as new data about variant occurrence and haemolytic risk become available.

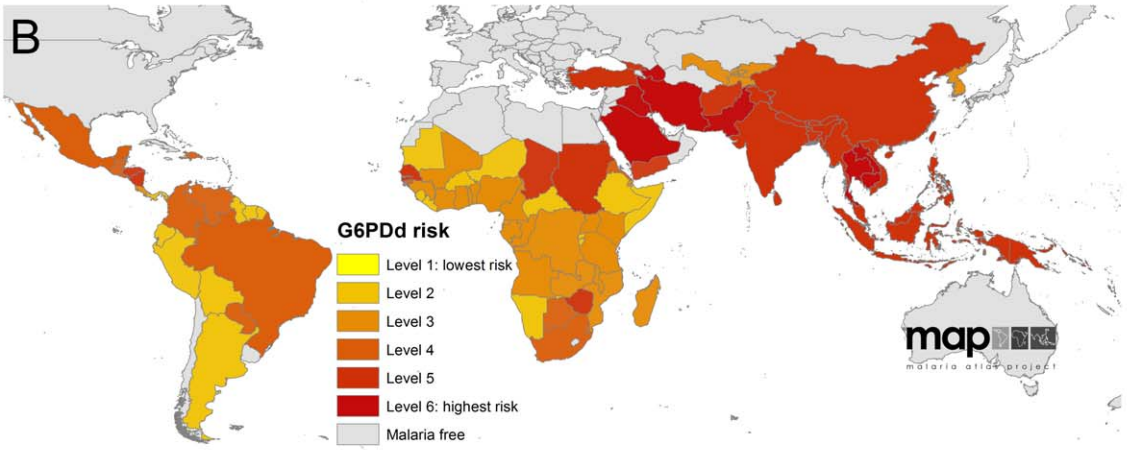
Discussion

G6PDd is widespread across malarious regions, where we estimated the deficiency to have an overall allele frequency of 8.0%. We have developed here an evidence-based, geostatistically modelled, and spatially continuous prevalence map of G6PDd, together with uncertainty metrics and population estimates of affected individuals. Although highest levels of G6PDd frequency are predicted in sub-Saharan Africa, high population density makes Asia the centre of weight of G6PD deficiency-burdened populations. We discuss our results first in relation to existing G6PDd maps, and then in their public health context in relation to the coincident severity of local variants. Important limitations to the maps and population estimates stem from weaknesses in the underlying database of surveys. These are also discussed, in relation to the difficulties of predicting deficiency in females, in assessing the robustness of the model predictions, and in overcoming the barriers to predicting the severity of primaquine-induced haemolysis.

Comparison with Existing Maps and Population Estimates

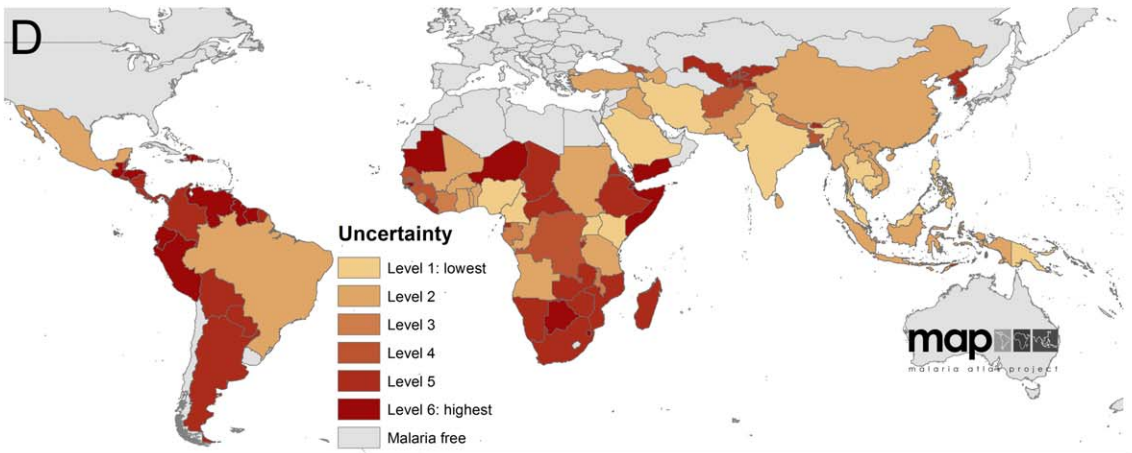
Previous G6PDd maps have been published by the WHO G6PD Working Group in 1989 [31], Cavalli-Sforza et al. in 1994 [34], and more recently in 2009 by Nkhoma et al. [33]. Both the WHO and Nkhoma et al. maps present data averages at national levels, thus masking all sub-national variation and making direct comparisons with our continuous prevalence map difficult. Further, Nkhoma et al.'s map has many gaps for countries from where no data could be found. However, all maps show broadly similar patterns, with lowest frequencies in the Americas, highest rates predicted across the tropical belt of sub-Saharan Africa, and generally heterogeneous distributions across Asia ranging from virtually absent to relatively high. Comparison of the national-level, population-weighted allele frequency estimates generated here with the WHO categories showed no obvious trends, with estimates for 29% of MECs predicted higher here than by WHO, frequencies in 36% of countries being predicted lower than those predicted by WHO, and 35% having consistent values. Reasons for these disparities relate both to the criteria imposed on the survey evidence-base (with both WHO and Nkhoma et al. including surveys that were excluded from this current study for risk of bias or lack of spatial specificity, corresponding to 108 and 17 surveys, respectively) and the statistical methods involved (accounting for the sample size and spatial distribution of data





C **G6PDd risk index**

		Variant severity		
		Class III only	Class II uncommon	Class II common
National G6PDd prevalence	**Rare**: ≤1%	Level 1 (n = 1)	Level 2 (n = 7)	Level 3 (n = 7)
	Common: >1 - 10%	Level 2 (n = 13)	Level 4 (n = 15)	Level 5 (n = 20)
	High: >10%	Level 3 (n = 20)	Level 5 (n = 5)	Level 6 (n = 11)



E **Uncertainty index**

		Reliability of severity score		
		Low uncertainty	Medium uncertainty	High uncertainty
Prevalence uncertainty (IQR/Median)	**Low**: ≤ 50%	Level 1 (n = 13)	Level 2 (n = 4)	Level 3 (n = 0)
	Medium: >50 - 100%	Level 2 (n = 19)	Level 4 (n =6)	Level 5 (n = 9)
	High: >100%	Level 3 (n = 7)	Level 5 (n = 21)	Level 6 (n = 20)

Figure 4. Index of severity risk from G6PDd. (A) shows the national score of variant severity, determined by the ratio of class II to class III variant occurrences reported from each country; (B) maps the risk index from G6PDd, accounting for both the severity of variants (A) and the overall prevalence of G6PDd (Figure 3A); the scoring matrix describing these scores is given in (C), specifying the different categories of risk determined by the scores of national-level prevalence of phenotypic deficiency (rows) multiplied by severity scores of the variants present (columns). (D) represents the uncertainty in the assembly of the risk index based on the prevalence scores (E rows) and in the assessment of variant severity (E columns). These uncertainties relate specifically to the analysis of these data into the risk index, and do not account for the underlying uncertainty in their interpretation in relation to haemolysis (see Discussion).
doi:10.1371/journal.pmed.1001339.g004

points, and relating G6PDd prevalence to spatial patterns of population density). The new map also has the benefit of two decades of additional surveys since the publication of the WHO map, and more than six times the number of surveys (in spite of the stricter inclusion criteria) than were used by Nkhoma et al. (280 surveys versus 1,734). Globally, the WHO study estimates 2.6% of male newborns to be hemizygous for G6PDd alleles. As our study focused on the subset of countries with highest G6PDd prevalences (MEC versus non-MEC [31]), our MEC regional estimate (8.0%; IQR: 7.4–8.8) cannot be directly compared to the global WHO figure. However, the considerably higher regional estimate predicted here is more consistent with the recent estimate of 7.3% (95%; confidence interval: 7.0–7.6) of the global population by Nkhoma et al. [33]. Disparity between estimates may result from the population weighting used in this present study, which ensures that prevalence in densely populated regions contributes proportionally more in the regional estimate than through simple national estimate averages. Finally, this study is the first to model G6PDd prevalence in females. The previous studies discussed here, selected that 10% of heterozygous females would be diagnosed as phenotypically deficient. The flexible Bayesian model developed for the current study, and the extensive database of female survey data, enabled an empirical assessment of this spatially variable threshold. The resulting estimates, however, are subject to the same limitations as the original diagnostic tests used (Protocol S5). Diagnosing heterozygotes, who express two populations of red blood cells—normal and deficient—is highly sensitive to the enzyme activity level thresholds imposed, as the deficiency can be masked by cells expressing normal activity. The population of G6PDd cells, however, is as vulnerable to haemolytic stress as the deficient cells of hemizygotes or female homozygotes. This source of diagnostic uncertainty should be considered when interpreting these predictions of deficient females, which are based directly upon the diagnostic results.

Model Uncertainty

The evidence-based nature of the analysis leaves the model predictions vulnerable to weaknesses in the underlying database. While some of these limitations can be quantified, such as prediction uncertainty in areas with very scarce data, others cannot. The current study presents a methodological advance over previously published maps for being the first to quantify any aspect of prediction uncertainty. In brief, our mapping procedure involved 500 repeated predictions being made from the optimised Markov chain Monte Carlo (MCMC) algorithm (Protocol S2). The median of all predicted values for each pixel is displayed in Figure 2B, and the IQR (50% confidence interval) of the repeated predictions was used to quantify model uncertainty (Figure 2C). Where model uncertainty is lowest, the 500 repeated predictions will fall within a small range, and the IQR will be correspondingly small; less straightforward predictions are associated with larger IQR values. In general, model uncertainty increases where fewer data are available and sample sizes are smaller, and where observed prevalence values are heterogeneous. This same principle applies to the population affected estimates.

Not all sources of uncertainty, however, could be accounted for by the model, which is dependent on the input dataset to represent the underlying G6PDd prevalence patterns. No global resource of genetic relatedness among populations was available, thus differences in prevalence between geographically close but genetically distant communities could only be represented in the map through the inclusion of surveys, thus, a scarcity of data may mask significant heterogeneity. For example, high prevalence of G6PDd among populations such as the endogamous groups of Orissa could not have been predicted by the model without data points from those communities. While the final dataset provides relatively good coverage, there are some large expanses lacking data where additional surveys are most needed to improve confidence in our knowledge of G6PDd prevalence, as indicated in the uncertainty map. These include several South American countries, large parts of central and southern Africa, and some highly populous Indonesian islands; the careful geopositioning of all surveys in this study allows specific gaps in the datasets to be identified that are masked in nationally aggregated maps. However, uncertainty in some of the data point geopositioning was also unaccounted for. While 80% of surveys could be mapped as points (<10 km^2), 20% were less specific and mapped as polygons up to 35 km in radius of which centroid coordinates were used in the model (Protocol S1). The relative uncertainty introduced from this level of geopositioning uncertainty was deemed acceptable relative to the level of uncertainty, which would have been introduced by excluding those 20% of data points altogether. Finally, uncertainty in the prevalence estimates themselves stemming from the diagnostics is discussed in Protocols S1 and S5. In brief, the binary expression of normal activity versus deficiency is generally considered to be relatively reliably detected in males by most diagnostics [26,31], though the quality of reagents and the practical difficulties of field-based settings for instance will produce some errors. As discussed previously, diagnostics for heterozygous females are altogether more complex and uncertain. The most ambiguous diagnostics for assessing the deficiency phenotype—molecular-based methods, due to the gene's extensive genetic variability—were excluded (Protocol S1).

G6PDd Applications to Malaria Treatment

G6PDd is of pertinence to malaria treatment due to the potentially dangerous consequences of exposing G6PD deficient individuals to the vitally important anti-malarial drug primaquine. An endemicity map of *P. vivax* has recently been developed [57] indicating where this anti-relapse drug is likely to be most needed, with greatest demand being in countries targeting elimination [14]. The G6PDd map presented here can contribute to the evidence-base for weighing risk and benefit in formulating primaquine treatment strategies that could greatly accelerate the elimination of malaria transmission. We predict here that within countries targeting malaria elimination, G6PDd had an allele frequency of 5.3%, corresponding to an estimated 61 million G6PDd males and 35 million G6PDd females, with most of those occurring in Asia. However, there is evidence of a protective role for G6PDd against severe *P. falciparum* malaria [58,59], and an

effect has recently been reported against *P. vivax* parasitaemia as well [60,61]. This being so, the prevalence of G6PDd in clinical cases of malaria may be lower than among the general population, though the precise nature of the protective effect (including which genotypes benefit) remains controversial [62,63]. In any event, G6PDd prevalence in the broader population, as we present, remains a useful measure of the risks incurred with prescribed primaquine therapy. This may be particularly true where mass drug administration that includes primaquine is considered.

G6PDd Severity

The diagnostic tests commonly used in community surveys determine a binary deficient/non-deficient classification; the prevalence map presented here corresponds to this binary classification, an indicator of whether primaquine may or may not be tolerated. Such diagnostics, however, cannot predict clinical severity of primaquine-induced harm, which is known to range from clinically inconsequential to life threatening [24]. More than 186 mutations have been described to the gene [64], which encode proteins expressing a spectrum of residual enzyme activity. In an attempt to encapsulate a measure of that variability in deficiency severity, we devised a simple index accounting for the relative prevalence and severity of G6PDd variants, which is intended as a guide to stratify broad categories of G6PDd-associated risk between countries and regions. However, interpretation of this analysis is constrained by major knowledge gaps. First, in relation to the evidence-base: there were no data from almost half of MECs (45 of 99) meaning that severity scores had to be inferred for many of them. Further, it is likely that reporter bias and preconceptions regarding which mutations are common, and thus worthwhile testing for, will have a strong effect on the collated database. Second, relating this index to primaquine-induced haemolytic risk assumes an inverse correlation between variant enzyme activity levels and primaquine sensitivity. Although this relationship has been found with the three variants in which the primaquine sensitivity phenotype has been characterised (A−, Mediterranean, and Mahidol [24]), further research into the association between the numerous other genetic variants and their susceptibility to primaquine is essential to substantiate this assumption. Third, the classification used here to distinguish "more severe" from "less severe" variants, in other words, the enzyme classifications into classes II and III, uses an arbitrary cut-off of 10% enzyme activity, which is not founded on clinical evidence of significance to haemolytic severity [31,56]. It has been suggested that the distinction between these classes is blurred and may no longer be useful [65]. Fourth, the mechanism of haemolytic trigger by primaquine remains to be determined: this basic biochemical research would offer a rational basis for all of the above, and enable much more robust predictions of haemolytic risk using the datasets already collated here (of G6PDd prevalence and of the distribution of G6PDd variants).

In the absence of evidence supporting robust predictions of relative risk of severe haemolysis, residual enzyme activity is an easily obtained, albeit as yet not validated, surrogate. While such a surrogate could help inform the risk and benefit for using primaquine in any given population, in clinical practice with patients it is the dichotomy of normal versus deficient that guides primaquine treatment decisions. No treatment recommendations refer to residual enzyme activity [66]. As such, the current map of phenotypic deficiency prevalence remains the most detailed, robust, and appropriate risk assessment of overall G6PDd-associated harm, whether mild or severe, relevant to public health policies of mass primaquine administration. The insight offered by the severity index presented here corroborates the high G6PDd-associated risk that the majority of the global population at risk of *P. vivax* [57] faces.

G6PDd in African Malaria Endemic Countries

At the continental level, highest prevalence of G6PDd is predicted across sub-Saharan Africa, where prevalence drops below 5% only on the edges of its distribution in eastern and southern Africa. In spite of being so common, the implications of G6PDd-associated primaquine reactions are not currently of major concern due to the present status of malaria control across much of the continent. High *P. falciparum* endemicity [41] means that drug policy almost exclusively targets the clinical stages. Transmission blocking therapies in such settings have not proven effective or sustainable [67]. Furthermore, the continent has relatively few people at risk of *P. vivax* [57,68] due to the predominance of the Duffy negativity blood group [40], which is generally refractory to *P. vivax*. Thus, despite endemicity of the other relapsing human malaria, *P. ovale*, primaquine for anti-relapse is not applied in Africa [69]. However, this basis for not applying primaquine may well disappear as malaria control programmes reduce endemicity to sustainably low transmission levels, thus increasing the feasibility of elimination. When low transmission intensity is reached, policy in Africa will need to consider the treatment and practice questions now being faced in Asian and American MECs. Any primaquine treatment policy will have to account for the high prevalence of G6PDd across this continent. The G6PDd variant causing deficiency across the African population is commonly attributed to the "mild" A− mutant (Table S3) [70], and thus primaquine-associated risk of harm is thought to be minor and self-limiting [71], reflected by the moderate risk levels predicted across most of the continent (Figure 4B). However, recent evidence of low primaquine dosage triggering severe anaemia in an A– type individual (a genotype commonly considered very mildly deficient) [72], and findings from extensive DNA sequencing identifying a greater diversity of G6PD mutations than previously acknowledged [70,73], calls for caution when using primaquine in these areas of high G6PDd prevalence, in spite of the relatively mild nature of primaquine sensitivity of the A− variants, as determined in otherwise healthy adults (rather than in children with malaria).

G6PDd in Countries Targeting Malaria Elimination

Malaria eliminating countries (Figure 2) face steep challenges in achieving their ambitions. Prominent among these many challenges include: (i) endemic *P. vivax* malaria, and emerging resistance to chloroquine, previously the drug of choice for treating acute attacks, and recently arteminisin resistance also; (ii) high prevalence of carriers of the clinically silent and diagnostically invisible *P. vivax* hypnozoite; and (iii) the predominance of asymptomatic carriers of sexual and asexual blood stages despite low transmission intensity. The problem of *P. vivax* resistance to chloroquine is discussed elsewhere [74], but is most prevalent and threatening in south and southeast Asia [75], where its emergence greatly compounds the difficulty of the therapeutic problem [76]. A recent study along the Thai-Myanmar border [23] documented very high prevalence of *P. vivax* parasitaemia in the 63 d following therapy for acute *P. falciparum* malaria (20%–51%; correlated with drug half-life). Those rates seem to support rational and pragmatic use of anti-hypnozoiticidal primaquine treatment for all malaria patients where these parasites occur together [77]. Further, another study in the hypo-endemic Solomon Islands found that fewer than 30% of PCR-diagnosed blood infections were detected by expert microscopists, and only about 5% of infected individuals were symptomatic (overall prevalence was 9% according to PCR

diagnostics but only 2.7% with microscopy) [78]. Both of these studies demonstrate the important parasite reservoir represented by asymptomatic, sub-microscopic, and latent infections, and the WHO now reconsiders its long-standing recommendation against mass drug administration as an element of malaria control [79].

Primaquine is the only chemotherapeutic tool currently available for attacking hypnozoites and mature gametocytes. As explained elsewhere [24], the available data on the safety of any regimen of primaquine may be considered almost completely inadequate by any contemporary clinical and pharmacological standards. Any MEC considering a strategy for attacking the silent hypnozoite and gametocyte reservoirs would greatly benefit from an adequate evidence-base for rational weighing of clinical risk and benefit in their areas of operations. Such an assessment may require evaluation of local G6PDd variants for vulnerability to primaquine and, ideally, point-of-care G6PDd screening to exclude those at risk of harm. Such strategies would come with substantial financial and logistical outlays, but could be most usefully directed to areas with highest potential benefit with minimal risk of harm, as indicated by the many national maps of G6PDd prevalence embedded within the global map presented here. Additional information about the severity of local variants would help support this decision-making process. The map in this study, and any subsequent iterations (worthwhile if substantial numbers of new surveys become available), provides one of the many pieces of evidence to consider when strategizing for chemotherapeutic policy aimed at elimination of transmission and relapse.

Future Prospects and Conclusions

There is no immediate prospect of relief from the serious constraints to chemotherapeutics for malaria elimination. A new drug in phase IIb/III trials in 2012, Tafenoquine, is strategized as a successor to primaquine, but it is also likely to come with haemolytic toxicity in G6PDd patients, and thus the same constraints would apply [8]. The very brief dosing with Tafenoquine, combined with its relatively long plasma half-life, will require even greater caution in individuals affected by severe variants; though risks will be similar for patients with mild variants that lead to self-limiting haemolysis. Minimising treatment duration of primaquine from the standard 14 d has also been discussed as a means to promote course adherence and reduce risk of resistance emergence [30]. In other words, the stakes in 8-aminoquinoline therapies will increase as the commitment to elimination rises alongside a determination to attack the parasite stages that threaten success. Evaluation of risk informed by the G6PDd maps and population estimates presented here may guide appropriate investments in measures that will minimise the harm incurred by hypnozoites and gametocytes chemotherapeutics. For instance, an important potential tool in minimizing harm is a point-of-care diagnostic capable of excluding those at risk of harm caused by 8-aminoquinoline therapies. One such rapid diagnostic test in laboratory development showed promise in its first field evaluation [80]. As well as directly improving individual-level safety, such a kit may also vastly expand the available data to refine prevalence maps like that presented here, improving its resolution and margins of error. Areas where additional data would be most informative are those with highest uncertainty in the current map (Figure 3) where no, or only very few, surveys were found. Furthermore, a single diagnostic test could contribute towards standardising diagnoses and removing the potential variation between diagnostic kits, which is inherent within the current database. Although diagnosis in males is generally considered consistent with existing kits (Protocol S1), a single test would ensure this.

The prominence of G6PDd represents a barrier to current options for malaria elimination therapy. Nevertheless, the unique properties of primaquine are increasingly in demand as communities target depletion of their parasite reservoirs. It is evident that no measures are currently in place to ensure safe delivery of primaquine within the context of G6PDd risk. The complexity and diversity of both malaria epidemiology and G6PDd mean that no single solution will be applicable for ensuring safe and effective primaquine treatment. The maps and population estimates presented here represent one component of this treatment decision-making framework, and pave the way for further data collection and refinement of mapping studies of G6PDd severity. The relative urgency of this important component to determining appropriate elimination therapy may be determined by the relative prevalence of G6PDd and malaria endemicity in any given area [57,81].

All maps at national and regional scales and in GIS and image formats, population estimates, as well as the input surveys database are freely available on the Malaria Atlas Project website (MAP; http://www.map.ox.ac.uk/).

Supporting Information

Dataset S1 Bibliography of sources from which surveys included in the model were identified.
(RTF)

Protocol S1 Assembling a global database of G6PD deficiency (G6PDd) prevalence surveys. (S1.1) Overview of database requirements. (S1.2) Library assembly. (S1.3) Dataset inclusion criteria. (S1.4) Survey diagnostic methods. (S1.5) The final G6PDd survey dataset. (S1.6) Defining MECs' limits.
(DOCX)

Protocol S2 Model based geostatistical framework for predicting G6PDd prevalence maps. (S2.1) Model requirements in relation to G6PD genetics. (S2.2) The model. (S2.3) Model implementation. (S2.4) Overview of mapping procedure. (S2.5) Uncertainty.
(DOCX)

Protocol S3 Model validation procedures and results. (S3.1) Creation of the validation datasets. (S3.2) Model validation methodology. (S3.3) Validation results.
(DOCX)

Protocol S4 Demographic database and population estimate procedures. (S4.1) GRUMP-beta human population surface. (S4.2) Areal prediction procedures.
(DOCX)

Protocol S5 Mapping the prevalence of G6PDd in females. (S5.1) Overview of G6PDd in females. (S5.2) Heterozygous G6PDd expression and diagnosis. (S5.3) Overview of female data in the G6PD database. (S5.4) Modelling phenotypic G6PDd prevalence in females. (S5.5) Maps of G6PDd in females and population estimates. (S5.6) Improving the map of G6PDd in females.
(DOCX)

Protocol S6 Developing an index of overall national-level risk from G6PD deficiency. (S6.1) G6PDd variants database. (S6.2) Generating an index of national-level risk from G6PDd. (S6.3) Generating an uncertainty index of the national-level risk index categories.
(DOCX)

Table S1 National-level demographic metrics and G6PDd allele frequency and population estimates.
(PDF)

Table S2 National areal prediction summary statistics and Monte Carlo standard error (SE) for each model output.
(PDF)

Table S3 Reported observations of class II and III G6PD variants from malaria endemic countries.
(PDF)

Acknowledgments

We thank the large number of people who have generously contributed their unpublished data to this initiative; these individuals are listed on the Malaria Atlas Project website (MAP: http://www.map.ox.ac.uk/). We particularly acknowledge the contribution of the Filipino national screening data from the Newborn Screening Reference Center, National Institutes of Health, Philippines. We thank Harriet Dalrymple, Suzanne Phillips, and Jennie Charlton for help with the library assembly. We gratefully acknowledge Oliver Brady, Justin Green, Lucio Luzzatto, Catherine Moyes, David Pigott, Ric Price, and Dennis Shanks for comments on the manuscript.

Author Contributions

Conceived and designed the experiments: REH FBP SIH. Performed the experiments: REH FBP OAN MD MMH KEB APP PWG. Analyzed the data: REH FBP APP PWG SIH. Contributed reagents/materials/analysis tools: CDP. Wrote the first draft of the manuscript: REH. Contributed to the writing of the manuscript: FBP SIH JKB. ICMJE criteria for authorship read and met: REH FBP APP OAN PWG MD MMH KEB CDP JKB SIH. Agree with manuscript results and conclusions: REH FBP APP OAN PWG MD MMH KEB CDP JKB SIH.

References

1. Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B, et al. (2010) Shrinking the malaria map: progress and prospects. Lancet 376: 1566–1578.
2. Das P, Horton R (2010) Malaria elimination: worthy, challenging, and just possible. Lancet 376: 1515–1517.
3. The Global Health Group and the Malaria Atlas Project (2011) Atlas of malaria-eliminating countries. San Francisco: The Global Health Group, Global Health Sciences, University of California, San Francisco.
4. Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, et al. (2010) Operational strategies to achieve and maintain malaria elimination. Lancet 376: 1592–1603.
5. Carlton JM, Sina BJ, Adams JH (2011) Why is *Plasmodium vivax* a neglected tropical disease? PLoS Negl Trop Dis 5: e1160. doi:10.1371/journal.pntd.0001160
6. Gosling RD, Okell L, Mosha J, Chandramohan D (2011) The role of antimalarial treatment in the elimination of malaria. Clin Microbiol Infect 17: 1617–1623.
7. Karl S, Gurarie D, Zimmerman PA, King CH, St Pierre TG, et al. (2011) A sub-microscopic gametocyte reservoir can sustain malaria transmission. PLoS One 6: e20805. doi:10.1371/journal.pone.0020805
8. Wells TN, Burrows JN, Baird JK (2010) Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination. Trends Parasitol 26: 145–151.
9. White NJ (2008) The role of anti-malarial drugs in eliminating malaria. Malar J 7 Suppl 1: S8.
10. Baird JK, Schwartz E, Hoffman SL (2007) Prevention and treatment of vivax malaria. Curr Infect Dis Rep 9: 39–46.
11. Baird JK (2010) Eliminating malaria - all of them. Lancet 376: 1883–1885.
12. Cappellini MD, Fiorelli G (2008) Glucose-6-phosphate dehydrogenase deficiency. Lancet 371: 64–74.
13. Shanks GD, Kain KC, Keystone JS (2001) Malaria chemoprophylaxis in the age of drug resistance. II. Drugs that may be available in the future. Clin Infect Dis 33: 381–385.
14. WHO (2010) Guidelines for the treatment of malaria, second edition. Geneva: World Health Organization.
15. WHO (2010) World malaria report 2010. Geneva: World Health Organization.
16. WHO (2011) Global plan for artemisinin resistance containment (GPARC). Geneva: World Health Organization.
17. Song J, Socheat D, Tan B, Dara P, Deng C, et al. (2010) Rapid and effective malaria control in Cambodia through mass administration of artemisinin-piperaquine. Malar J 9: 57.
18. Lawpoolsri S, Klein EY, Singhasivanon P, Yimsamran S, Thanyavanich N, et al. (2009) Optimally timing primaquine treatment to reduce *Plasmodium falciparum* transmission in low endemicity Thai-Myanmar border populations. Malar J 8: 159.
19. Graves PM, Gelband H, Garner P (2012) Primaquine for reducing *Plasmodium falciparum* transmission. Cochrane Database Syst Rev 9: CD008152.
20. Shekalaghe SA, Bousema JT, Kunei KK, Lushino P, Masokoto A, et al. (2007) Submicroscopic *Plasmodium falciparum* gametocyte carriage is common in an area of low and seasonal transmission in Tanzania. Trop Med Int Health 12: 547–553.
21. Global Malaria Programme (2007) Malaria elimination: A field manual for low and moderate endemic countries. Geneva: World Health Organization.
22. The malERA Consultative Group on Diagnoses Diagnostics (2011) A research agenda for malaria eradication: diagnoses and diagnostics. PLoS Med 8: e1000396. doi:10.1371/journal.pmed.1000396
23. Douglas NM, Nosten F, Ashley EA, Phaiphun L, van Vugt M, et al. (2011) *Plasmodium vivax* recurrence following falciparum and mixed species malaria: risk factors and effect of antimalarial kinetics. Clin Infect Dis 52: 612–620.
24. Baird JK, Surjadjaja C (2011) Consideration of ethics in primaquine therapy against malaria transmission. Trends Parasitol 27: 11–16.
25. Brueckner RP, Ohrt C, Baird JK, Milhous WK (2001) 8-Aminoquinolines. Rosenthal PJ, editor. Antimalarial chemotherapy: mechanisms of action, resistance, and new directions in drug discovery. Totowa (New Jersey): Humana Press.
26. Beutler E (1994) G6PD deficiency. Blood 84: 3613–3636.
27. Abeyaratne KP, Halpe NL (1968) Sensitivity to primaquine in Ceylonese children due to deficiency of erythrocytic glucose-6-phosphate dehydrogenase. Ceylon Med J 13: 134–138.
28. Burgoine KL, Bancone G, Nosten F (2010) The reality of using primaquine. Malar J 9: 376.
29. The malERA Consultative Group on Drugs (2011) A research agenda for malaria eradication: drugs. PLoS Med 8: e1000402. doi:10.1371/journal.pmed.1000402
30. APMEN Vivax working group. Annual Business and Technical meeting; 2011; Kota Kinabalu, Malaysia. Available: http://apmen.org/storage/apmen-iii/Dr%20Ric%20Price.pdf. Accessed 8 February 2012.
31. WHO Working Group (1989) Glucose-6-phosphate dehydrogenase deficiency. Bull World Health Organ 67: 601–611.
32. Luzzatto L, Notaro R (2001) Malaria. Protecting against bad air. Science 293: 442–443.
33. Nkhoma ET, Poole C, Vannappagari V, Hall SA, Beutler E (2009) The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. Blood Cells Mol Dis 42: 267–278.
34. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton (New Jersey): Princeton University Press. 1088 p.
35. Padilla CD, Therrell BL (2007) Newborn screening in the Asia Pacific region. J Inherit Metab Dis 30: 490–506.
36. Singh S (1973) Distribution of certain polymorphic traits in populations of the Indian peninsula and South Asia. Isr J Med Sci 9: 1225–1237.
37. Mourant AE, Kopec AC, Domaniewska-Sobczak K (1976) The distribution of the human blood groups and other polymorphisms. London: Oxford University Press.
38. Livingstone FB (1985) Frequencies of hemoglobin variants: thalassemia, the glucose-6-phosphate dehydrogenase deficiency, g6pd variants and ovalocytosis in human populations. New York: Oxford University Press.
39. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, et al. (2007) Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. Malar J 6: 17.
40. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, et al. (2011) The global distribution of the Duffy blood group. Nat Commun 2: 266.
41. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6: e1000048. doi:10.1371/journal.pmed.1000048
42. Diggle PJ, Ribeiro PJ, Jr (2007) Model-based Geostatistics: Springer.
43. Diggle P, Moyeed R, Rowlingson B, Thomson M (2002) Childhood malaria in the Gambia: a case-study in model-based geostatistics. J Roy Stat Soc C-App 51: 493–506.
44. Clements AC, Moyeed R, Brooker S (2006) Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. Parasitology 133: 711–719.
45. Magalhaes RJ, Clements AC (2011) Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. PLoS Med 8: e1000438. doi:10.1371/journal.pmed.1000438
46. Raso G, Matthys B, N'Goran EK, Tanner M, Vounatsou P, et al. (2005) Spatial risk prediction and mapping of *Schistosoma mansoni* infections among schoolchildren living in western Cote d'Ivoire. Parasitology 131: 97–108.

47. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. (2012) Global estimates of sickle haemoglobin in newborns. Lancet. Published online 25 Oct 2012. doi: http://dx.doi.org/10.1016/S0140-6736(12)61229-X.
48. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. (2010) Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. Nat Commun 1: 104.
49. Hardy GH (1908) Mendelian proportions in a mixed population. Science 28: 49–50.
50. Weinberg W (1908) Über den nachweis der vererbung beim menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg 64: 368–382.
51. Peters AL, Van Noorden CJ (2009) Glucose-6-phosphate dehydrogenase deficiency and malaria: cytochemical detection of heterozygous G6PD deficiency in women. J Histochem Cytochem 57: 1003–1011.
52. Abdulrazzaq YM, Micallef R, Qureshi M, Dawodu A, Ahmed I, et al. (1999) Diversity in expression of glucose-6-phosphate dehydrogenase deficiency in females. Clin Genet 55: 13–19.
53. Patil AP, Gething PW, Piel FB, Hay SI (2011) Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol 27: 246–253.
54. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, et al. (2006) Determining global population distribution: methods, applications and data. Adv Parasitol 62: 119–156.
55. United Nations Department of Economics and Social Affairs (2011) World population prospects, the 2010 Revision. New York: United Nationals Population Division.
56. Yoshida A, Beutler E, Motulsky AG (1971) Human glucose-6-phosphate dehydrogenase variants. Bull World Health Organ 45: 243–253.
57. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, et al. (2012) A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. PLoS Negl Trop Dis 6: e1814. doi:10.1371/journal.pntd.0001814
58. Guindo A, Fairhurst RM, Doumbo OK, Wellems TE, Diallo DA (2007) X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. PLoS Med 4: e66. doi:10.1371/journal.pmed.0040066
59. Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, et al. (1995) Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376: 246–249.
60. Leslie T, Briceno M, Mayan I, Mohammed N, Klinkenberg E, et al. (2010) The impact of phenotypic and genotypic G6PD deficiency on risk of *Plasmodium vivax* infection: a case-control study amongst Afghan refugees in Pakistan. PLoS Med 7: e1000283. doi:10.1371/journal.pmed.1000283
61. Louicharoen C, Patin E, Paul R, Nuchprayoon I, Witoonpanich B, et al. (2009) Positively selected G6PD-Mahidol mutation reduces *Plasmodium vivax* density in Southeast Asians. Science 326: 1546–1549.
62. Hedrick PW (2011) Population genetics of malaria resistance in humans. Heredity (Edinb) 107: 283–304.
63. Luzzatto L (2012) G6PD deficiency and malaria selection. Heredity (Edinb) 108: 456.
64. Minucci A, Moradkhani K, Hwang MJ, Zuppi C, Giardina B, et al. (2012) Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. Blood Cells Mol Dis 48: 154–165.
65. Luzzatto L (2009) Glucose-6-phosphate dehydrogenase deficiency. Orkin SH, Nathan DG, Ginsburg D, Look AT, Fisher DE, et al., editors. Nathan and Oski's hematology of infancy and childhood. 7th ed. Philadelphia: Saunders.
66. Baird JK (2012) Chemotherapeutics challenges in developing effective treatments for the endemic malarias. Int J Parasitol. In press.
67. Bousema T, Drakeley C (2011) Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. Clin Microbiol Rev 24: 377–410.
68. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. PLoS Negl Trop Dis 4: e774. doi:10.1371/journal.pntd.0000774
69. WHO (2011) Country antimalarial drug policies: by region. Geneva: WHO.
70. Clark TG, Fry AE, Auburn S, Campino S, Diakite M, et al. (2009) Allelic heterogeneity of G6PD deficiency in West Africa and severe malaria susceptibility. Eur J Hum Genet 17: 1080–1085.
71. Dern RJ, Beutler E, Alving AS (1954) The hemolytic effect of primaquine. II. The natural course of the hemolytic anemia and the mechanism of its self-limited character. J Lab Clin Med 44: 171–176.
72. Shekalaghe SA, ter Braak R, Daou M, Kavishe R, van den Bijllaardt W, et al. (2010) In Tanzania, hemolysis after a single dose of primaquine coadministered with an artemisinin is not restricted to glucose-6-phosphate dehydrogenase-deficient (G6PD A-) individuals. Antimicrob Agents Chemother 54: 1762–1768.
73. Johnson MK, Clark TD, Njama-Meya D, Rosenthal PJ, Parikh S (2009) Impact of the method of G6PD deficiency assessment on genetic association studies of malaria susceptibility. PLoS One 4: e7246. doi:10.1371/journal.pone.0007246
74. Baird JK (2009) Resistance to therapies for infection by *Plasmodium vivax*. Clin Microbiol Rev 22: 508–534.
75. Douglas NM, Anstey NM, Angus BJ, Nosten F, Price RN (2010) Artemisinin combination therapy for vivax malaria. Lancet Infect Dis 10: 405–416.
76. Baird JK (2011) Resistance to chloroquine unhinges vivax malaria therapeutics. Antimicrob Agents Chemother 55: 1827–1830.
77. Baird JK (2011) Radical cure: the case for anti-relapse therapy against all malarias. Clin Infect Dis 52: 621–623.
78. Harris I, Sharrock WW, Bain LM, Gray KA, Bobogare A, et al. (2010) A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. Malar J 9: 254.
79. WHO (2011) Consideration of mass drug administration for the containment of artemisinin resistant malaria in the greater Mekong subregion. Geneva: WHO.
80. Kim S, Nguon C, Guillard B, Duong S, Chy S, et al. (2011) Performance of the CareStart^{TM} G6PD deficiency screening test, a point-of-care diagnostic for primaquine therapy screening. PLoS One 6: e28357. doi:10.1371/journal.pone.0028357
81. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, et al. (2011) A new world malaria map: *Plasmodium falciparum* endemicity in 2010. Malar J 10: 378.

Chapter 5 – Distinct spatial trends in G6PD deficiency variants across malaria endemic regions

Having established the spatial distribution and prevalence of glucose-6-phosphate dehydrogenase deficiency (G6PDd) across malaria endemic countries, this next chapter turns to consider in more detail the underlying genetic variants of this disorder. G6PDd variants make up a spectrum of enzyme deficiency severity levels and have important implications for the assessing relative safety of primaquine therapy regimens.

5.1. Background

The discovery of G6PDd occurred during the clinical development of primaquine against the relapse of *Plasmodium vivax* malaria in American prisoner volunteers during and immediately following the Second World War (Carson *et al.*, 1956). This inherited abnormality and primaquine are inextricably linked in malaria therapy, control and elimination as the drug's potential lethal toxicity to G6PDd individuals sharply limits its effective use (Baird and Surjadjaja, 2011; Baird, 2013). Nevertheless, the unique therapeutic activities of primaquine render it potentially extremely useful in combating endemic malaria (White, 2008; Baird, 2012b). This drug is currently the only licenced therapy active against the liver-stages of *P. vivax*. If untreated, these dormant hypnozoites will relapse after a latency period of three weeks to ten months (White, 2011), reactivating blood stage infection which may cause severe disease and mortality (Baird, 2013). Furthermore, primaquine is the only drug with activity against the sexual transmission-stages of all *Plasmodium* species (Baird and Surjadjaja, 2011; Bousema and Drakeley, 2011; White, 2012), a role of undeniable importance in reducing transmission levels, most particularly in containing the spread of artemisinin-resistant *P. falciparum* (WHO, 2010; WHO, 2011). Despite these advantages, the absence of a practical field-based diagnostic test for G6PDd impedes widespread use of primaquine (Recht *et al.*, unpublished). Our

study of evidence-based screening data has shown that G6PDd is especially common across malaria endemic countries (MECs) with an estimated global median allele frequency of 8.0% (50% CI: 7.4-8.8) (Howes *et al.*, 2012), thought to be driven by a selective advantage against life-threatening malaria (Greene, 1993; Beutler, 1996; Kwiatkowski, 2005).

G6PD is a housekeeping enzyme essential for red blood cell (RBC) survival. G6PD maintains reserves of reducing power in the form of NADPH, which are necessary for detoxifying oxidative challenges to RBCs such as hydrogen peroxide and oxygen radicals (Cappellini and Fiorelli, 2008; Howes *et al.*, 2013); this is further discussed in Chapter 6. Unlike other cells, RBCs lack mitochondria so have no alternative mechanisms for generating reducing power (Mason *et al.*, 2007). Mutations in the G6PD gene can destabilise the enzyme and reduce its activity levels, leaving cells vulnerable to oxidative damage. Exogenous triggers, including certain foods, infections, and a range of drugs can trigger damage and result in RBC death. The resulting clinical symptoms range in severity from negligible to potentially lethal and may include neonatal jaundice, favism (triggered by fava bean ingestion) and acute haemolytic anaemia (Luzzatto, 2009). Severe symptoms tend to be most common in males, as the G6PDd trait is X-linked, meaning that females must inherit two deficient alleles to express the same overall reductions in enzyme activity as deficient males (who inherit a single X chromosome). Heterozygous females may express anywhere between all or none of the G6PDd phenotype, depending on the relative proportions of normal and deficient cell populations resulting from the random process of Lyonization. This spectrum of expression in heterozygous females imposes a challenge to their diagnosis (Luzzatto, 2009; Howes *et al.*, 2012). A detailed description of this gene's population genetics is given in Chapter 4.

G6PDd may be diagnosed in several ways (Figure 5.1). In Chapter 4, I considered G6PD as a binary deficiency/normal phenotype (Figure 5.1A). This classification is diagnosed with qualitative or semi-quantitative enzyme-based methods commonly used in population screening surveys, and occasionally fully quantitative methods. However, to distinguish the different variants of G6PDd, a more detailed diagnosis is required. These variant-

characterising methods are those referred to in this chapter (Figure 5.1B). To distinguish different G6PDd variants, either the enzyme itself must be thoroughly described with a suite of biochemical investigations of the purified enzyme (including enzyme kinetics, electrophoretic mobility, heat stability, activity-pH curves and Michaelis constant measurements (Betke *et al.*, 1967)) or the gene's DNA must be examined directly to identify the mutations to the G6PD gene.

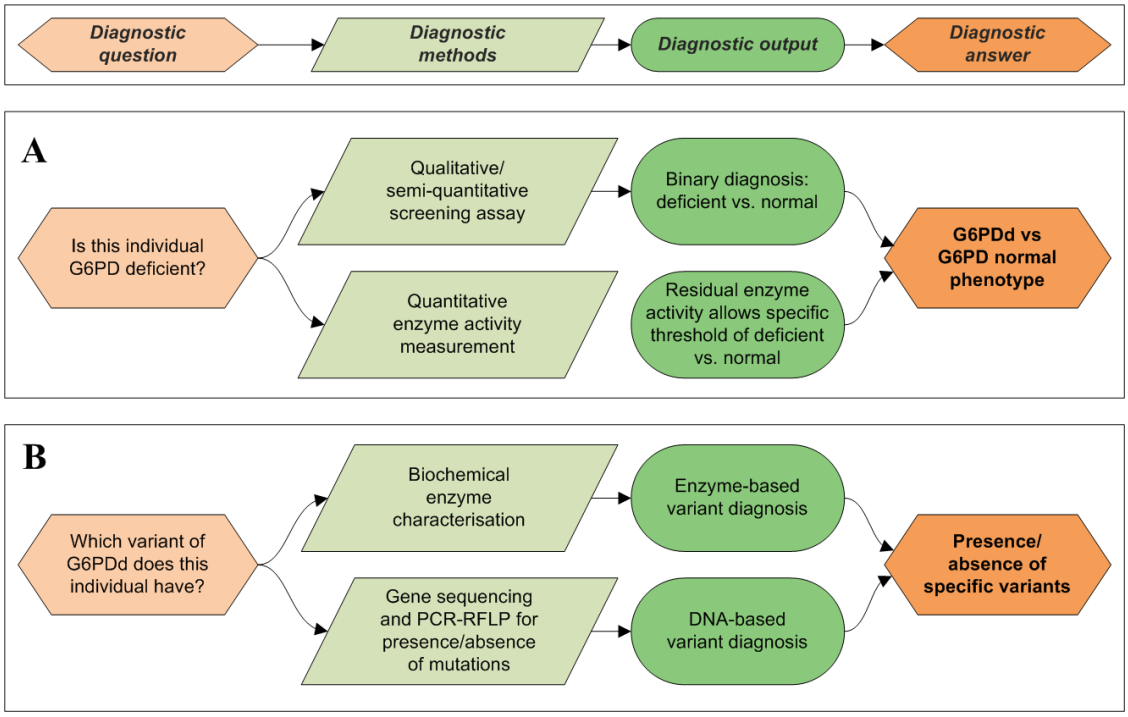


Figure 5.1. G6PDd diagnostic methods and common laboratory techniques associated with different types of diagnostic questions. Panel A summarises diagnostics related to identifying deficient from normal G6PD activity. Panel B indicates the methods required to characterise the variants of G6PDd. The orange hexagons indicate the question and answers associated with the different methods. The different diagnostic methods associated with each are shown in the pale green boxes, and the diagnostic outcomes of each are shown in the bright green ellipses.

At least 186 mutations have been genetically characterised in the G6PD gene (Minucci *et al.*, 2012), though not all are polymorphic and of clinical significance. About half of these variants (Mason and Vulliamy, 2005) appear to be sporadic mutations identified in only a handful of patients. These rare variants usually express very severe, chronic symptoms, a pathology known as chronic non-spherocytic haemolytic anaemia (CNSHA) which can

result in lifelong dependency on blood transfusion (Luzzatto, 2009). Though numerous by type, these variants never reach polymorphic frequencies and are thus not of widespread public health concern. Previous designations of "common" polymorphic variants differ, but include at least 15 and up to 26 variants (Beutler, 1994; Luzzatto and Notaro, 2001; Luzzatto, 2006; Mason *et al.*, 2007); although the evidence base and rationale for these classifications is unclear. Current knowledge of G6PDd variants is skewed towards a handful of variants, both with respect to their clinical characteristics in response to primaquine and to their spatial distributions.

The G6PDd A- variant has been longest studied as all of the earliest evidence about the haemolytic risk of G6PDd came from African American volunteers, in whom this variant was known to be very common (Hockwald *et al.*, 1952). This variant usually reduces enzyme activity to around 5-10% of normal levels (Beutler, 1991). These early primaquine toxicity studies noted that primaquine-induced haemolysis associated with this A- variant was usually self-limiting (Dern *et al.*, 1954). It was found that G6PDd A- individuals could tolerate primaquine doses over an extended period of eight weeks, rather than the usual 14 days, thereby significantly reducing the drug's toxicity (Alving *et al.*, 1960). In contrast, the Mediterranean G6PDd variant, common across southern Italy and Sardinia, as well as among populations inhabiting the Persian Gulf and the Arabian Sea (Cavalli-Sforza *et al.*, 1994), is best known for predisposing individuals to favism (Luisada, 1940; Meloni *et al.*, 1983). This variant is commonly considered the most clinically severe (Beutler and Duparc, 2007), expressing barely detectable levels of enzyme expression (<1% (Piomelli *et al.*, 1968)). Anti-malarial therapy with primaquine is contra-indicated for individuals with this variant (WHO, 2010) as even low doses of the drug may trigger highly severe haemolysis which would require transfusion therapy (Clyde, 1981). Across Asia, the Mahidol variant is best characterised, and is often considered the predominant variant across Myanmar and Thailand (Buchachart *et al.*, 2001; Matsuoka *et al.*, 2004). Enzyme activity in G6PDd Mahidol individuals is reduced to 5-32% of normal levels (Louicharoen *et al.*, 2009). Although low dosing of primaquine following G6PDd screening is recommended in this region (WHO, 2010), severe pathologies associated with G6PDd have been reported

(Laosombat *et al.*, 2006; Burgoine *et al.*, 2010). Although a far greater diversity of G6PDd variants than these three exists, knowledge of their characteristics is anecdotal and limited (Baird and Surjadjaja, 2011).

No evidence-based global spatial analysis of all the common G6PDd variants exists. Cavalli-Sforza and colleagues have mapped frequencies of the Mediterranean and A- alleles (Cavalli-Sforza *et al.*, 1994), and Luzzatto and Notaro presented an occurrence map of 15 common variants (Luzzatto and Notaro, 2001). However, neither of these provides any insight into the relative dominance of local variants, nor any measure of the proportions of unknown variants causing G6PDd in different areas.

This diversity of clinical phenotypes and genotypes in G6PDd, and our limited knowledge thereof, compounds the difficulty of addressing the technical and practical limitations which G6PDd imposes on primaquine treatment for attacking the endemic malarias (Baird, 2012a). National authorities responsible for the prevention, control and treatment of endemic malaria naturally strive for evidence-based practices that maximize benefit and minimize risk. The present study begins the complex task of characterizing G6PDd variant distribution and diversity with an aim to informing a framework of primaquine-associated risk, and providing a vital tool in controlling, eliminating, and, ultimately, eradicating malaria.

5.2. Methods

The aim of this study was to assemble an evidence-base of surveys reporting the prevalence of common G6PDd variants which are of widespread public health concern. Two types of data informed this goal:

- **Variant proportion data (series 1)**: data which reported the relative proportion of different variants among G6PDd individuals. These were population samples previously diagnosed as having deficiency enzyme activity using the methods

illustrated in Figure 5.1A. The variant analyses (represented in Figure 5.1B) were used to characterise the underlying mutations causing the deficiency.

- **Variant frequency data (series 2)**: population surveys which diagnosed variants directly without prior G6PDd screening (i.e. used the methods in Figure 5.1B only, without the enzyme screening methods from Figure 5.1A). These surveys had no absolute denominator of deficiency cases, meaning that individuals who were G6PD normal could not be distinguished from those with a deficiency caused by an undiagnosed variant. These surveys provided measures of the allelic frequencies of particular variants across the overall population.

Both data types had to be considered for this study as the different types of surveys were common to different parts of the world. We first describe the methodological steps common to both types of data, and then detail the inclusion criteria specific to each mapping protocol. The methodology and data inclusion criteria are summarised in Figure 5.2.

5.2.1. Library assembly

The first methodological step was a literature search to identify sources of representative population surveys of G6PDd, using the protocol previously described (Howes *et al.*, 2012). In summary, these used systematic keyword searches ("G6PD", "glucose-6-phosphate dehydrogenase" or "glucose 6 phosphate dehydrogenase") of major online biomedical literature databases (PubMed, ISI Web of Science and Scopus; last conducted 26 March 2012) and cross-checks with existing databases (Singh, 1973; Mourant *et al.*, 1976; Livingstone, 1985; Mason *et al.*, 2007; Nkhoma *et al.*, 2009; Minucci *et al.*, 2012). The study was limited to data from malaria endemic countries (MECs), thus corresponding to those areas where primaquine therapy is needed.

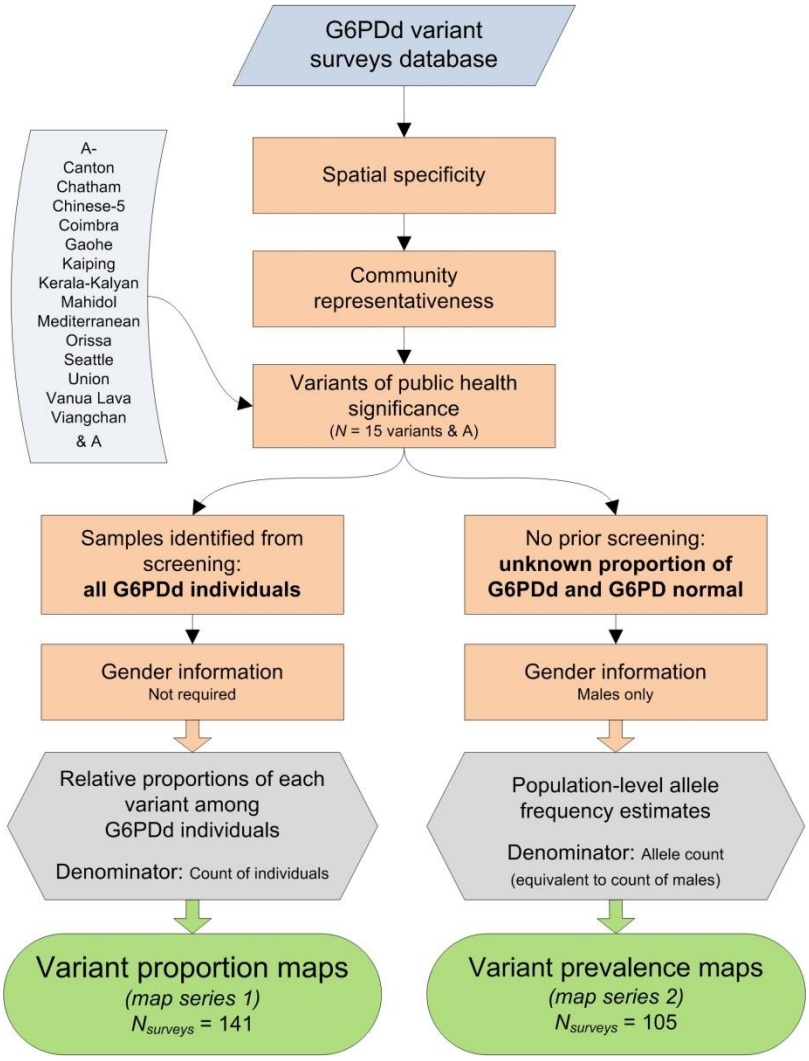


Figure 5.2. Survey inclusion criteria and G6PDd variant map outputs. Orange rectangles indicate the exclusion criteria, grey hexagons summarise the two final input data, and green rods represent the two map types. The A variant is included in the maps despite not being a variant of clinical significance; this variant is commonly associated with mutations encoding the A- variant.

5.2.2. Survey selection criteria

Two initial inclusion criteria were imposed. First, only population surveys which could be geopositioned to at least the national level were included. Where possible, surveys were mapped to the highest resolution spatial scale as point locations (e.g. villages). Second, to ensure that population samples were representative of the communities being surveyed, only studies which provided unbiased prevalence estimates were included. Case studies or

other patient groups, particularly those with symptoms of severe G6PDd (such as hyperbilirubinaemia, kernicterus or kidney failure), were excluded on account of being more likely to include individuals with severe variants of G6PDd. Malaria patients were excluded due to a potential advantage conferred by G6PDd, which would underestimate frequencies of the most protective variants. Similarly, family studies were excluded for being unrepresentative of the wider community due to their high degree of consanguinity. Finally, studies which included only individuals from a minority of ethnic backgrounds were also excluded to ensure that the data collated would be widely representative (Figure 5.2).

5.2.3. Variant inclusion criteria

Given the genetic diversity of G6PDd variants, it was necessary to identify those variants which presented a significant public health threat. The variant inclusion criteria were: (i) that the residual enzyme activity level should be significantly reduced (<25% normal expression) and thus diagnosable as deficient by standard qualitative diagnostics, and (ii) that the variants be reported from at least five localities across the malaria endemic region. Across the database, 15 variants met these criteria and were included in the mapping: A-, Canton, Chatham, Chinese-5, Coimbra, Gaohe, Kaiping, Kerala-Kalyan, Mahidol, Mediterranean, Orissa, Seattle, Union, Vanua Lava, Viangchan. These largely corresponded to those previously designated as "common variants" (Beutler, 1994; Luzzatto and Notaro, 2001), with the exception of the Aures, Chinese-4, Chinese-5, Cosenza, Honiara, Santamaria, Quing Yan, Taipei and Ube-Konan variants which were not widely reported, including from some non-malaria endemic areas including Japan and the Mediterranean region.

An exception was made for the A variant. Although this variant (A376G) does not meet the criteria of having significantly reduced enzyme activity (G6PDd A variant expression is barely reduced at approx. of 85% normal levels), it was included in the maps as it was almost always inherited alongside mutations in other loci (Clark *et al.*, 2009), together

expressing the A- (G202A/A376G or T968C/A376G or G680T/A376G) and Santamaria (A542T/A376G) variants, and therefore associated with detectably reduced enzyme activity.

G6PDd variants were diagnosed with either DNA-based methods (e.g. PCR-RFLP and sequencing) or enzyme-based methods (e.g. enzyme electrophoresis and biochemical characterisation of purified enzyme samples, as per the WHO standardised procedures (Betke *et al.*, 1967)). Over time, some variants have been defined using biochemical diagnostic methods by several laboratories independently, resulting in a degree of duplicity when re-examined at the DNA-level (Beutler, 2008). To ensure congruence between these reports, it was necessary to remove this duplication and standardise the database according to the underlying genetic mutations. The mutation tables compiled by Mason *et al.* (Mason *et al.*, 2007) and Minucci *et al.* (Minucci *et al.*, 2012) were used to reconcile duplicated variants with their underlying mutations; for example, the Union variant (C1360T) includes the biochemical variants Maewo, Chinese-2 and Kalo, as well as Union; the Seattle variant (G844C) also includes Lodi, Modena, Ferrara II, Athens-like and Mexico biochemical variants. The G871A mutation is common to both Viangchan and Jammu variants, although these are distinguished by haplotype analysis of a non-coding locus which is not frequently examined (Beutler *et al.*, 1991); these two variants were therefore considered a single variant by this study. Although a range of molecular mutations (G202A/A376G or T968C/A376G or G680T/A376G) have been identified as encoding the A- phenotype, the underlying mutations are not consistently reported by studies, and biochemical or electrophoretic diagnostic methods targeting the phenotype cannot discriminate this level of genetic variation anyway. All mutations relating to this phenotype were therefore categorised as A-. In cases where only the 202 locus was examined, these were also classified as A-.

5.2.4. Mapping the data

Surveys satisfying the inclusion criteria were abstracted into a database and mapped. Pie charts were chosen to display the relative prevalence of the different variants identified in each of the surveys, with each colour-coded segment proportional to the relative frequency of the variants reported. Surveys which could only be mapped to the national level were indicated by a star in the centre of the pie charts; pie charts without stars were therefore mapped with greater precision, from the village- to province-level. Spatial duplicates from independent studies, where multiple surveys had been conducted among the same communities, were mapped with a "jitter" of 0.5° in their latitude or longitude decimal degree coordinates to allow visualisation of multiple charts for the same location. Study sample size was incorporated in the maps through the radius of the pie charts, with larger pie charts representing bigger sample sizes. These had to be illustrated on a logarithmic scale to account for the large range in sample sizes. The MEC limits were those previously described (Gething *et al.*, 2011; Gething *et al.*, 2012; Howes *et al.*, 2012), corresponding to 99 *P. vivax* and *P. falciparum* endemic countries in 2010. The geographic regions used were selected for consistency with previous Malaria Atlas Project subdivisions, based on malaria epidemiological characteristics (Gething *et al.*, 2011; Gething *et al.*, 2012). These are: Americas, Africa+ (Africa, Saudi Arabia, Yemen), Asia (subdivided into West and East Central Asia, and the Pacific region). All mapping was performed in ArcMap 10 (ESRI, Redlands, CA, USA). The two series of maps previously described differed in the following respects.

5.2.4.1. Variant proportion maps (map series 1)

If the study samples had been previously identified as G6PDd from a population screening survey (using binary qualitative or quantitative enzyme-activity based methods, as shown in Figure 5.1A), these data were included in map series 1. In this series, the total number of G6PDd individuals in each study is known. For any of these individuals for whom a successful variant diagnosis was unavailable, the variant was classified as 'Other'.

The pie charts in these maps represent the relative proportions of each variant in the sample of G6PDd individuals examined (irrespective of gender), without providing any estimate of their overall population-level frequencies. The unit of these pie charts was the number of G6PDd individuals in the study. Sample size was therefore a further inclusion criterion for these maps and allowed the relative confidence in the data to be represented in the maps.

5.2.4.2. Variant frequency maps (map series 2)

Surveys which investigated G6PDd variants in individuals from cross-sectional population samples with no prior G6PDd screening were included in map series 2. These studies had no baseline information about the number of samples which were G6PDd, but instead, quantified the allele frequencies of selected G6PDd variants at the population level.

Given the X-linked genetics of the G6PD gene, deriving estimates of allele frequency required the sex of the individuals to be taken into account. As males carry only a single copy of the X allele, numbers of affected/non-affected males translated directly into frequency estimates. Precision in the terminology around female diagnostics could be unclear due to the variable thresholds of heterozygous deficiency (described in Chapter 4). Not all methods reliably differentiated heterozygous from homozygous G6PDd. For consistency and reliability therefore, only data from males were included in these variant frequency maps. Sample size corresponded to the total number of alleles (equivalent to total male individuals) tested. Data informing these maps therefore carried the additional inclusion criterion of providing results according to sex, as well as total sample sizes to allow the maps to represent relative confidence in the surveys.

It is important to note that in both types of maps, the studies did not always attempt to identify all the variants represented in the legends of these maps. For instance, if only the Mediterranean variant was tested for among a sample of deficient individuals (map series 1), those individuals who tested negative would be attributed to the "Other" category, regardless of whether they expressed one of the other variants listed in the legend or a

completely different one. The absence of a specific variant from a sample can only be inferred if no "Other" samples are reported. Similarly, with the variant frequency maps (map series 2), individuals with G6PDd variants would only have been identified if the specific variants they had were tested for in the study. Differentiating G6PD normal samples from undiagnosed G6PDd samples was not possible, and thus a large proportion of samples were classified as "Unidentified" in this second suite of maps.

5.3. Results

5.3.1. The database

A total of 18,939 bibliographic sources were identified from the broad keyword searches, of which 2,176 were considered likely to include spatial information about G6PDd variants and reviewed in detail for inclusion in the database. From these sources, 3,501 occurrences of any named G6PDd variant were identified; 3,221 of which could be mapped to a country. A total of 2,156 variant occurrences met the criteria of community representativeness, which excluded all potentially biased case studies (589 variant occurrences) and patients (298 occurrences reported from patients with G6PDd-associated symptoms, 48 occurrences from patients with malaria, and 130 variant occurrences from other patients). Of these representative community variant occurrences, 1,353 were from malaria endemic countries, and could therefore be included in the present mapping study. More than half of the occurrences of these variants (n = 823) were reported from community samples which had undergone prior screening for phenotypic G6PDd, thus meeting the inclusion criteria for the variant proportion maps (map series 1). These variant occurrences were reported from 141 population surveys. Excluding "Other" classifications, the mean number of variant occurrences reported per survey was 2.3 (range 1-10; SD: 1.8). The remaining variant occurrences were from population samples which had not undergone prior screening and therefore informed the frequencies of the variants at the population level (map series 2). Sixty occurrences were excluded for not specifying data by sex. The

remaining data were from 105 population surveys. Excluding "Unidentified" and "Other" classifications, the mean number of variants reported per survey was 1.8 (range 0-4; SD: 0.8). Overall, these two datasets pertaining to each map series were assembled from 145 published bibliographic sources, which are listed in the Appendix.

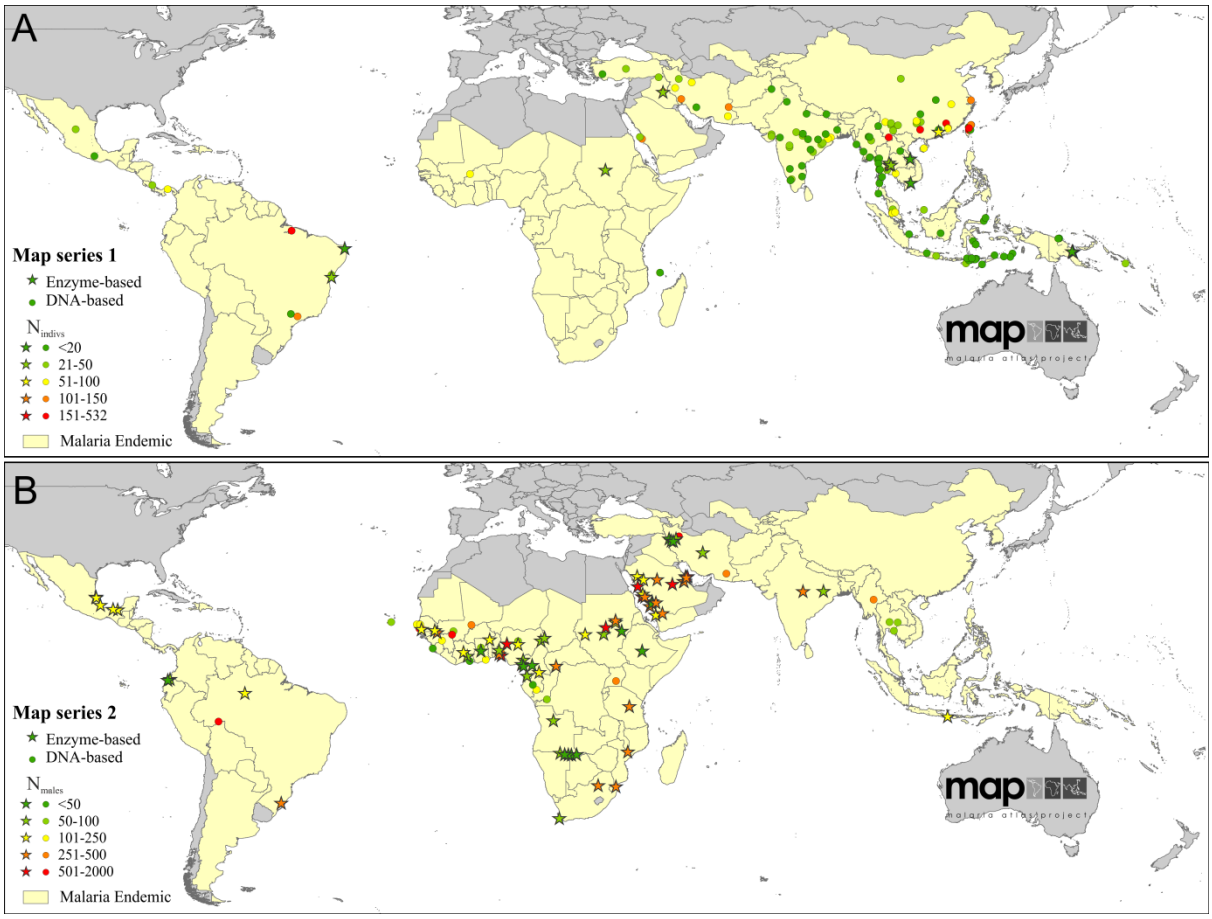


Figure 5.3. Distribution of the map input data for the (A) variant proportion maps and (B) variant frequency maps. Symbol shapes indicate their method of diagnosis: enzyme-based diagnoses are represented by starts, and circles indicate DNA-based diagnosis. Symbol colours reflect survey sample size: (A) total number of individuals and (B) total number of males.

The spatial distribution, sample size and diagnostic type (biochemical *vs*. molecular) of the surveys informing the two series of maps are summarised in Figure 5.3 and Table 5.1. The variant proportion data (map series 1; 141 surveys) were predominantly from Asian populations (126/141 surveys; 89%), and diagnosed with molecular methods (125/141 surveys; 89%). In contrast, the variant frequency data (map series 2; 105 surveys) were mostly from the Africa+ region (81/105 surveys; 77%) and used electrophoresis and other

biochemical diagnostic methods (77/105 surveys; 73%). Sample sizes also differed between data types, with a mean of 40 individuals in the variant proportion data which tested only G6PDd individuals (range: 1-532; SD: 61), and a mean of 281 males (alleles) tested in the G6PDd variant frequency data which considered all individuals indiscriminately (range: 17-2,000; SD: 367). Most surveys could be mapped with sub-national precision, but 11% of surveys in each map series (15 and 12 surveys in the proportion and frequency datasets, respectively) were mapped only to the national level.

		Africa+	**Americas**	**Asia**	**Global (MECs)**
Data series 1 *Variant proportion maps*	$N_{surveys}$	5	10	126	141
	$N_{countries}$	4	4	17	25
	$N_{G6PDd\ indivs}$	272	573	4,852	5,697
	Mean sample size	54.4	57.3	38.5	40.4
	Diagnosis:				
	Enzyme-based	1	2	13	16
	DNA-based	4	8	113	125
Data series 2 *Variant frequency maps*	$N_{surveys}$	81	10	14	105
	$N_{countries}$	24	3	6	33
	N_{indivs}	24,464	1,934	3,148	29,546
	Mean sample size	302.0	193.4	224.9	281.4
	Diagnosis:				
	Enzyme-based	60	9	8	77
	DNA-based	21	1	6	28

Table 5.1. Summary of input data according to map type.

5.3.2. G6PDd variants global patterns

The maps reveal conspicuously distinct geographical patterns in the distribution and prevalence of G6PDd variants across regions. The two series of maps represent both the relative proportions of the variants responsible for phenotypic G6PDd (Figure 5.4) and the

allele frequencies of some common variants at the population level (Figure 5.5). Together, these maps show three clear patterns: (i) low diversity of G6PDd variants reported from populations of the Americas and Africa+ regions, among whom the A- variant is predominantly reported; similarly, the Mediterranean variant was predominant in west Asia (Saudi Arabia and Turkey to India); (ii) a sharp shift in variants identified east of India, showing very little admixture with the common variants of west Asia; (iii) high variant diversity across populations of east Asia and the Asia-Pacific region, with multiple variants commonly co-occurring and no single variant predominating in any area. The characteristics of the variants in each region are discussed in more detail in the following sections.

5.3.3. G6PDd variants in the Americas

Only a relatively small number of surveys were available from the Americas (ten of each data type), which were mainly from central America and coastal regions of south America. Figure 5.4A indicates that among G6PDd individuals, the predominant variant was A- (G202A/A376G or T968C/A376G or G680T/A376G), identified in 90% of deficient individuals surveyed across the region (513 of 573 total G6PDd individuals surveyed). Other variants identified included the Mediterranean (C563T) and Seattle (G844C) variants; the latter was only reported from Brazil. A small minority of deficient variants remained unidentified or were too rare to be of public health significance; a survey among Mexicans reported 61% of deficient cases as being due to A-, but did not test for any other variant (e.g. Mediterranean), so the remaining variants remain "Other" (Vaca *et al.*, 2002). The allele frequency surveys (map series 2), examining a total of 1,934 males across ten different sites (Figure 5.5A), corroborated this picture of the dominant alleles, with A- being the most common variant searched for and identified. At the population level, this variant ranged in allele frequency from ≤2.5% in five surveys in Mexico to 13.8% in Ecuador.

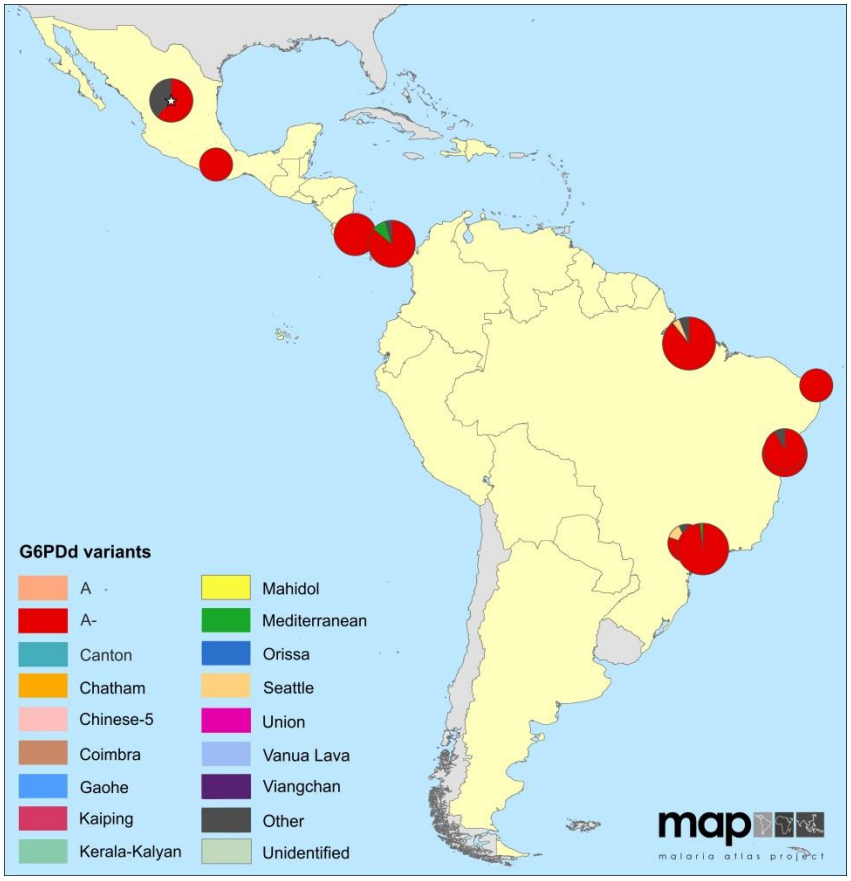


Figure 5.4. G6PDd variant proportion maps (map series 1). Pie charts represent individuals previously identified as G6PDd. Sample size is reflected in the size of the pie charts, which is normalised on a logarithmic scale. Surveys which could only be mapped to the country-level are indicated by a white star. MECs in the region mapped are shown with a yellow background; white backgrounds indicate MECs outside the region in focus; grey backgrounds represent malaria free countries. Variants which could not be diagnosed were reported as "Other".

Figure 5.4.A. G6PDd variant proportion maps (map series 1): Americas. 10 surveys with a mean sample size of 57 (range: 8-196; for reference, the most easterly survey in Brazil included 8 individuals). 1 survey was mapped at the national-level.

Figure 5.5. G6PDd variant frequency maps (map series 2). Pie charts represent allele frequencies. Sample size is reflected in the size of the pie charts, which is normalised on a logarithmic scale. Surveys which could only be mapped to the country-level are indicated by a white star. MECs in the region mapped are shown with a yellow background; white backgrounds indicate MECs outside the region in focus; grey backgrounds represent malaria free countries. Surveys in whom rare G6PDd variants which did not meet the variant inclusion criteria are classified as "Other"; "Unidentified" cases represent to individuals whose G6PD status remains uncertain: they may either be G6PD normal, or have an unidentified G6PDd variant.

Figure 5.5A. G6PDd variant frequency maps (map series 2): Americas. 10 surveys with a mean sample size of 193 alleles (range: 29-90; for reference, the sample in Porto Alegre, Brazil was of 462 alleles). No surveys were mapped at the national-level.

5.3.4. G6PD variants in Africa, Yemen and Saudi Arabia (Africa+)

Very few studies (n = 5) investigating the variants of known G6PDd cases were identified from the Africa+ region (Figure 5.4B). Instead, frequencies of specific G6PDd variants were commonly investigated across the continent (Figure 5.5B), identifying in particular the A, A- and Mediterranean variants. Frequencies of the A- variant ranged from <1% across Saudi Arabia, Sudan, Ethiopia and South Africa to >20% across West Africa (Figure 5.5B(2)). A large survey of 1,451 males diagnosed using electrophoretic methods in Western Nigeria reported 21.6% males carrying the A- variant (Luzzatto and Allan, 1968). The Mediterranean (C563T) variant was only reported from Saudi Arabia, where it was the predominant variant identified, reported by three surveys to be at frequencies above 35% on the Persian Gulf coast (sample sizes: 305 to 515 individuals) (Figure 5.5B(3)).

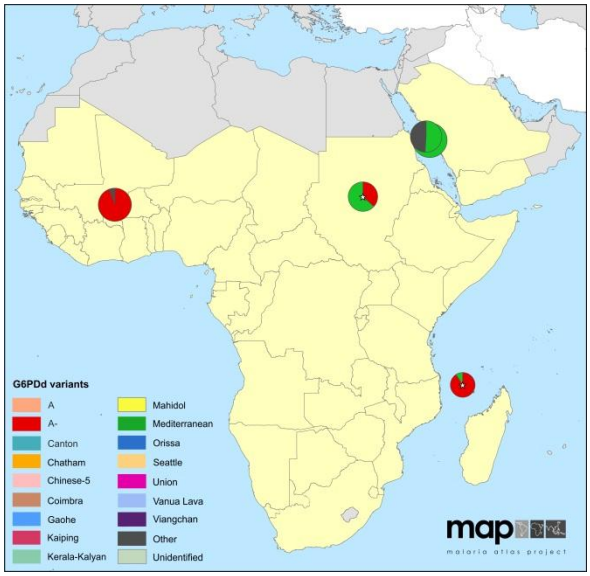


Figure 5.4B. G6PDd variant proportion maps (map series 1): **Africa+**. 5 surveys with a mean sample size of 54 (range: 11-110; for reference, the survey in Sudan included 30 individuals). 2 surveys were mapped at the national-level.

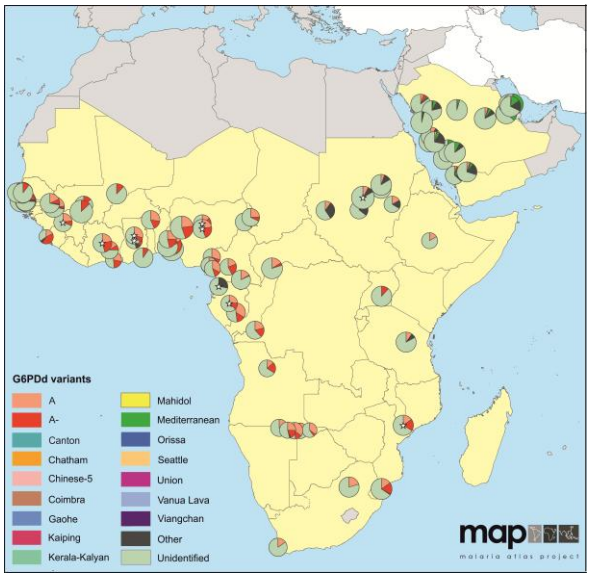


Figure 5.5B(1). G6PDd variant frequency maps (map series 2): Africa+. 81 surveys with a mean sample size of 302 alleles (range: 17-2000; for reference, the survey in Ethiopia was of 36 alleles and Uganda was of 311 alleles). 10 surveys were mapped at the national-level.

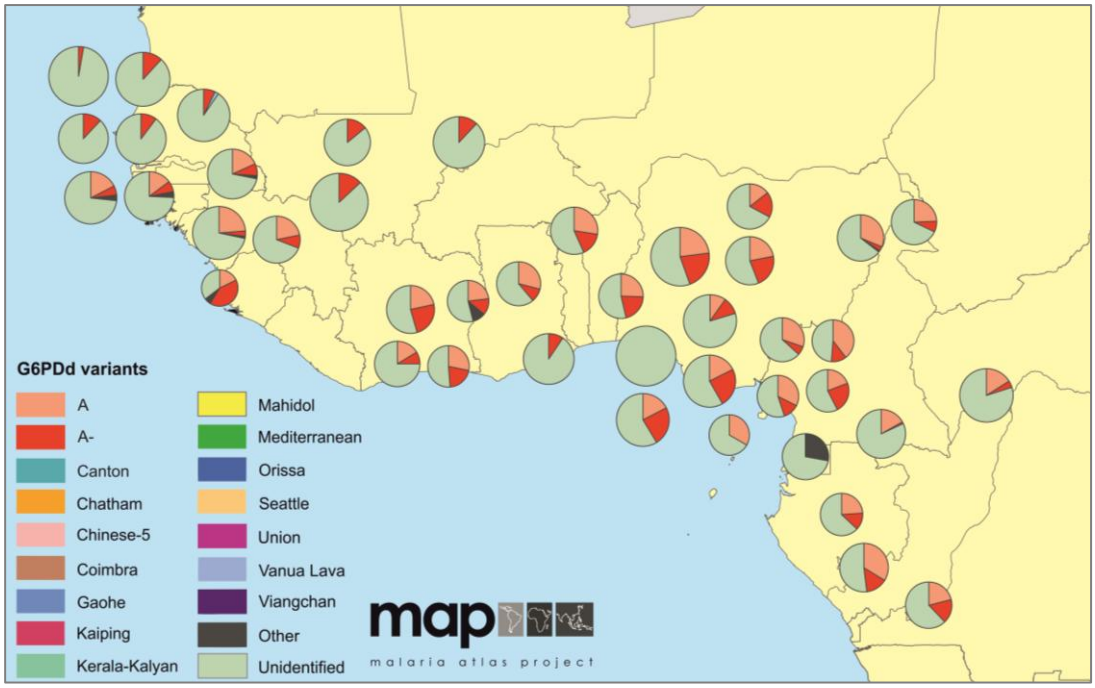


Figure 5.5B(2). G6PDd variant frequency maps (map series 2): West Africa. A higher resolution map of Figure 3B(1), but with pie charts spread out to avoid overlap.

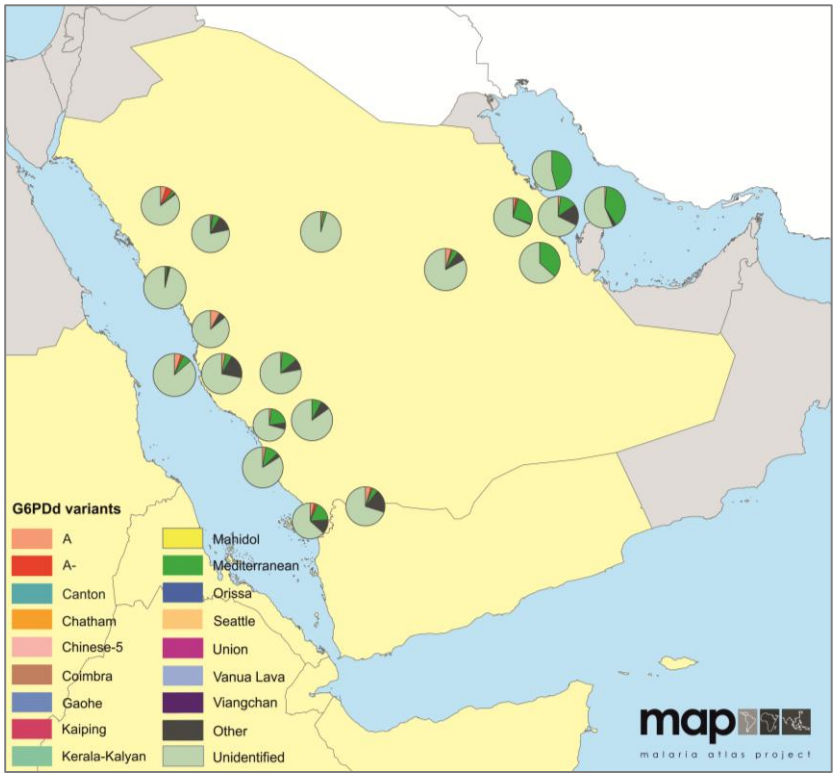


Figure 5.5B(3). G6PDd variant frequency maps (map series 2): Saudi Arabia. A higher resolution map of Figure 3B(1), but with pie charts spread out to avoid overlap.

5.3.5. G6PD variants in Asia and Asia-Pacific

The highest G6PDd variant diversity globally was across the Asia and Asia-Pacific regions (Figure 5.4C-D & 5.5C-D), where up to ten G6PDd variants were reported to co-occur within single populations at polymorphic frequencies. Furthermore, significant proportions of "Other" cases were frequently reported in the variant proportion maps (map series 1), indicating that genetic diversity is even greater than represented by the pie charts in these maps. This "Other" classification peaked at 100% in surveys from Papua New Guinea where all variants reported were local, and were not reported from sufficient surveys to meet the inclusion criterion of "public health significance". Only a small number of allele frequency surveys (map series 2) were recorded from the Asia region (14 of 105 globally), so we focus the discussion here on the variant proportion maps (Figure 5.4D). The majority of variant proportion studies across Asia used molecular diagnostics (113 of 126 surveys).

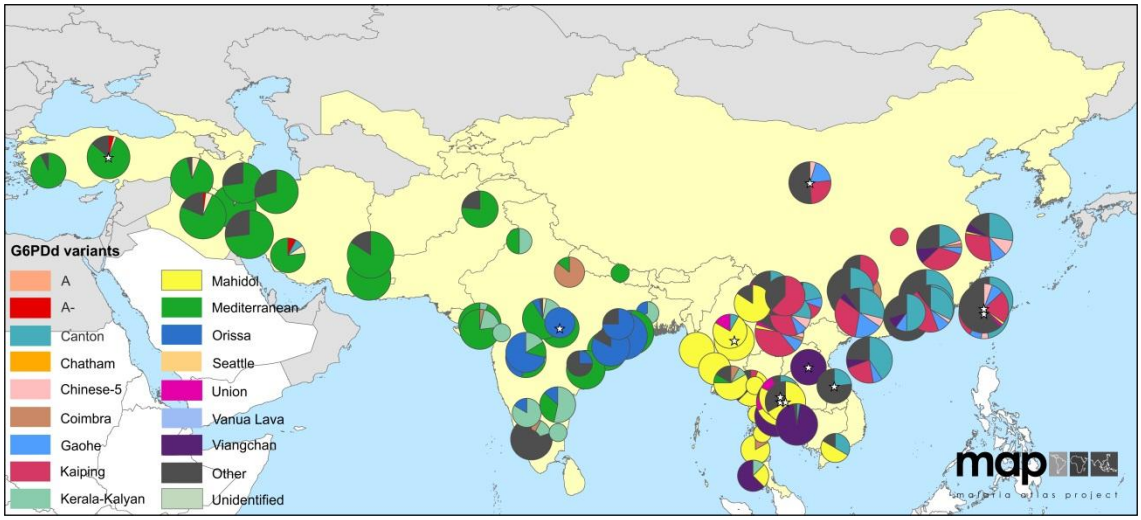


Figure 5.4C(1). G6PDd variant proportion maps (map series 1): Asia. 90 surveys with a mean sample size of 47 (range: 1-532; for reference, the survey in Nepal included 2 individuals and the survey mapped to the national-level in China was of 43 individuals). 12 surveys were mapped to the national-level.

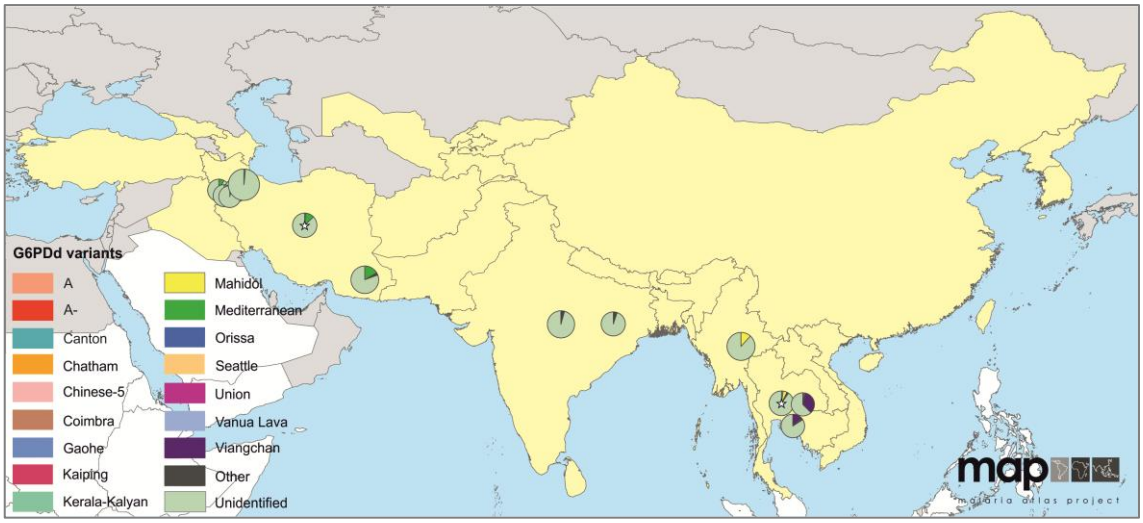


Figure 5.5C. G6PDd variant frequency maps (map series 2): Asia. 13 surveys with a mean sample size of 229 (range: 34-1500; for reference, the sample in Myanmar was of 353 alleles). 2 surveys were mapped at national-level.

From Turkey to Pakistan, the Mediterranean variant was predominant among G6PDd individuals, identified in 580 of the 754 G6PDd individuals examined (77%) (Figure 5.4C(2)). Two variants, Kerala-Kalyan (G949A) and Orissa (C131G), were reported only from Indian populations. On the Indian sub-continent, these two variants and the Mediterranean variant represented the majority of deficiency cases, though notable proportions of "Other" cases were also reported from eastern and southern India.

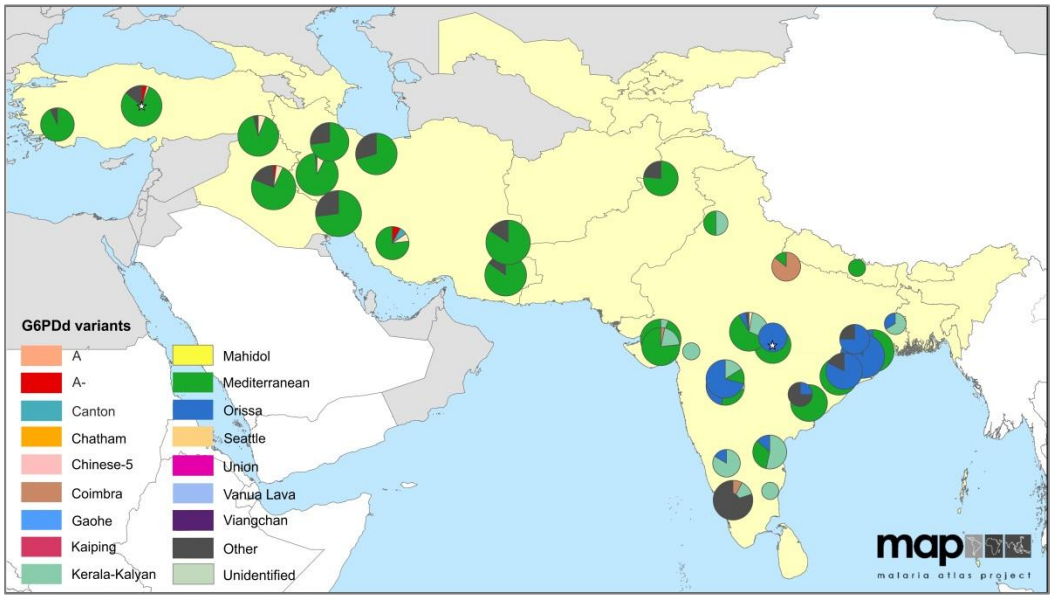


Figure 5.4C(2). G6PDd variant proportion maps (map series 1): West Asia. A higher resolution map of Figure 5.4C(1).

East of India, a completely different set of variants appeared (Figure 5.4C(3)). The Mahidol (G487A) variant predominates across Myanmar, with 98 of 117 G6PDd individuals (84%) diagnosed across 14 surveys as carrying this variant. Variants common among G6PDd individuals in south-east China were largely unique to these populations. Of the 1,906 G6PDd cases diagnosed from China, commonly identified variants included Kaiping (G1388A) (total G6PDd cases: 593; 31% across China), Canton (G1376T) (472 cases; 25%), Gaohe (A95G) (164 cases; 9%) and Chinese-5 (C1024T) (46 cases; 2%) variants. Surveys often reported variation beyond these common variants, including important proportions of "Other" variants (533 cases; 28%). The distribution of the Viangchan (G871A) variant was diffuse, reportedly common from Laos (where examination of 15 G6PDd individuals all carried this variant, (Iwai *et al.*, 2001)) and Cambodia (reported from 61 of 64 G6PDd individuals (Kim *et al.*, 2011)) to Papua New Guinea (where a sample of 13 G6PDd individuals included nine with this variant (Hung *et al.*, 2008)).

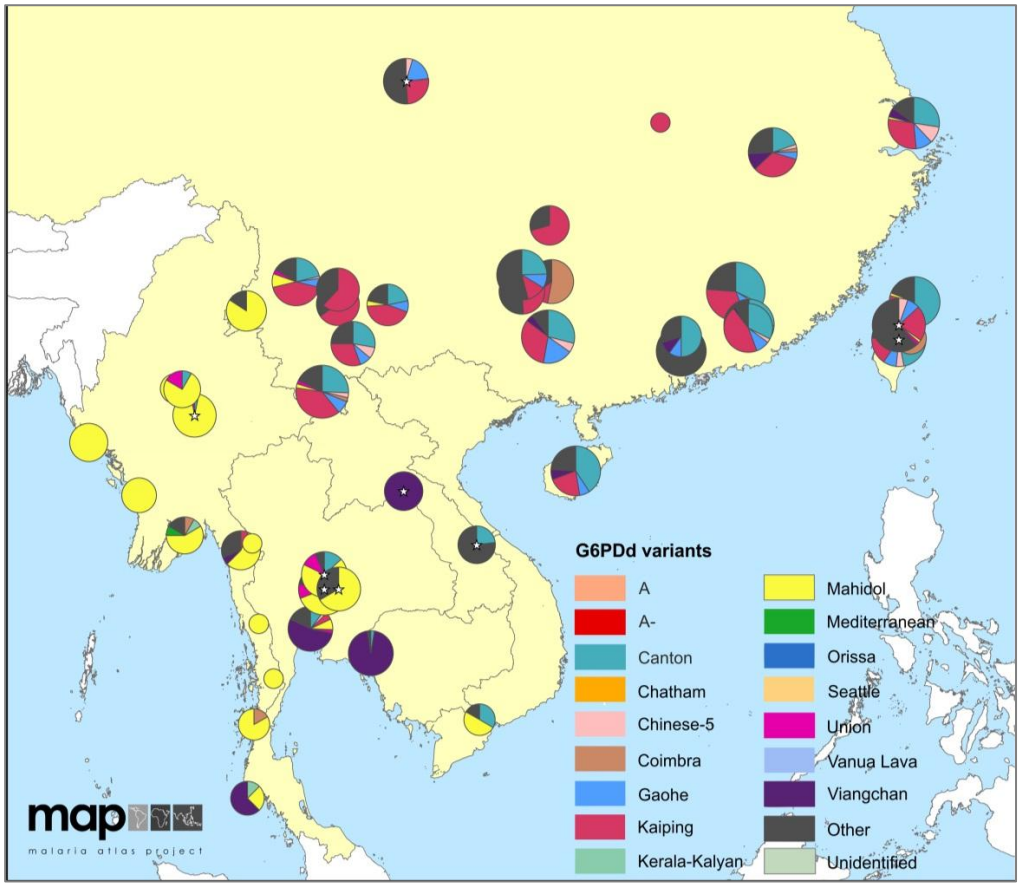


Figure 5.4C(3). G6PDd variant proportion maps (map series 1): East Asia. A higher resolution map of Figure 5.4C(1).

Variants across the Asia-Pacific region were highly heterogeneous (Figure 5.4D), with the only common pattern emerging from the 36 surveys of deficient individuals across this region (n_{indivs} = 629) being the diversity of variants and the heterogeneity of their prevalence among different populations. It should be noted, however, that 21 of these surveys

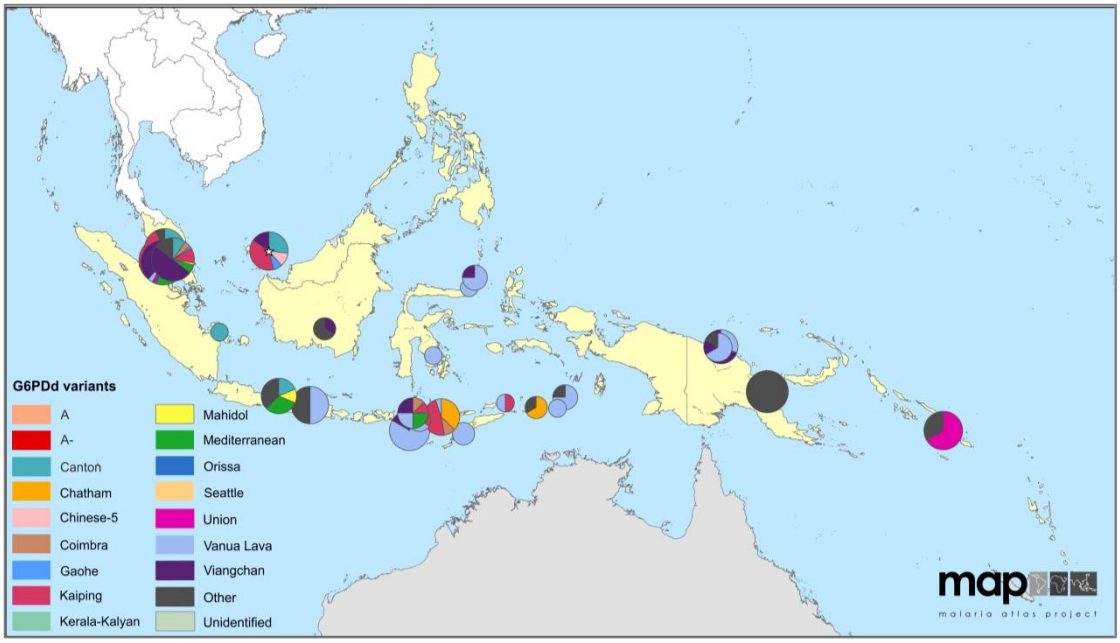


Figure 5.4D. G6PDd variant proportion maps (map series 1): Asia-Pacific: 36 surveys with a mean sample size of 17 (range: 1-128; for reference, the survey in the Solomon Islands was of 27 individuals and Kalimantan, Indonesia, was of 3 individuals). 1 survey was mapped at the national-level.

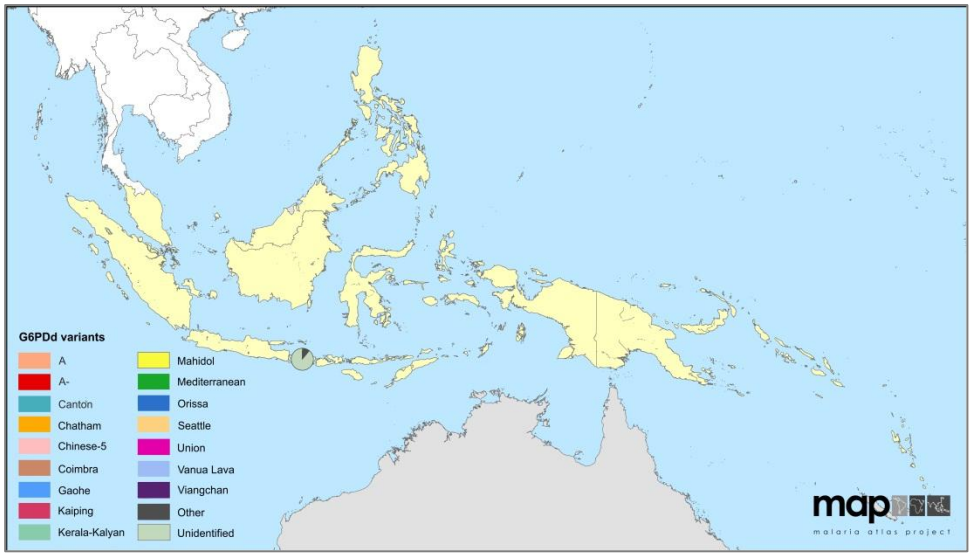


Figure 5.5D. G6PDd variant frequency maps (map series 2): Asia-Pacific: 1 survey was identified from this region with a sample size of 166 alleles.

tested fewer than ten G6PDd individuals, limiting the diversity which would be captured and thus the representativeness of these reports. The greatest diversity identified in a single survey was from the Malaysian neonatal screening programme in Kuala Lumpur, which recorded ten variants from an investigation of 86 G6PDd newborns. The Vanua Lava (T383C) variant was commonly identified from Indonesian studies, with seven other variants also reported across the archipelago.

5.4. Discussion

Over a third of the world's population lives at risk of *Plasmodium vivax* infection (Gething *et al.*, 2012). Very limited evidence underpins estimates of clinical cases, but these have been estimated at 70 to 400 million annually (Mendis *et al.*, 2001; Hay *et al.*, 2004). *Plasmodium vivax* causes potentially severe illness and death (Singh *et al.*, 2011; Mahgoub *et al.*, 2012; Baird, 2013). The only drug that can prevent relapsing clinical attacks is primaquine, a treatment contra-indicated for an estimated 8.0% of the population at risk of infection due to G6PDd (Howes *et al.*, 2012). The half-century of neglect of this malaria species has seriously impeded chemotherapeutic advances that addressed the G6PDd toxicity problem (Mendis *et al.*, 2001; Baird and Surjadjaja, 2011; Baird, 2012b; Baird, 2012a). Access to safe and effective therapy where most malaria patients live will require a new non-haemolytic drug or a practical means of identifying G6PDd malaria patients prior to administering therapy.

5.4.1. G6PDd haemolytic risk

The severity of haemolysis in G6PDd individuals following exposure to primaquine is determined by drug dose and the time period over which it is taken, the age distribution of red cells (which is altered by states of anaemia), concurrent infection, and the nature of the G6PDd variant (Cappellini and Fiorelli, 2008; Luzzatto, 2009; Howes *et al.*, 2013). Early primaquine studies established that a total dose of approximately 200 mg combined with

standard chloroquine therapy against the acute attack was needed to achieve *P. vivax* radical cure (Edgcomb *et al.*, 1950; Alving *et al.*, 1953; Coatney *et al.*, 1953); dosing over 14 days was found to be the optimal compromise between safety and acceptable compliance in relation to pamaquine (Most *et al.*, 1946) (the precursor drug to primaquine), and was adopted as standard primaquine therapy. Safety studies were subsequently conducted among "primaquine-sensitive" individuals who had been observed to haemolyse following exposure to this dosing (Alving *et al.*, 1948; Edgcomb *et al.*, 1950; Hockwald *et al.*, 1952; Carson *et al.*, 1956; Flanagan *et al.*, 1958; Beutler, 1959). Early on, investigators documented that the same total dose of primaquine had equal efficacy whether administered as a single dose, daily doses for 14 days, or weekly doses for 8 weeks (Mihaly *et al.*, 1985). This total dose effect enabled extended drug regimens with apparently good safety profiles to be used for G6PDd individuals. Current WHO recommendations for primaquine are based on the primaquine sensitivity phenotypes of three G6PDd variants: A-, Mediterranean and Mahidol (WHO, 2010; Baird and Surjadjaja, 2011; Howes *et al.*, 2013). Discussion of these variant-specific regimens is organised here on a geographic basis, together with key messages pertinent from the different regional maps.

5.4.2. G6PDd variants in Africa

The early investigations of "primaquine sensitivity" were conducted on African Americans (Hockwald *et al.*, 1952). Given their ancestral origin, it is likely that they were expressing the G6PDd A- variant, which reduces G6PD enzyme expression to 5-10% of normal levels (Beutler, 1991). An intermittent regimen of eight weekly 45 mg doses was found to avert haemolytic risk by avoiding dangerous drops in haemocrit (Alving *et al.*, 1960), while remaining an efficacious *P. vivax* radical cure. This intermittent dosing remains the current WHO recommended schedule for individuals with "mild" G6PD deficiency (WHO, 2010). Nevertheless, its safety has been questioned (Hill *et al.*, 2006; Shekalaghe *et al.*, 2010).

Historically, the use of primaquine has been very limited across Africa. The reasons for this are diminishing, however, and use is likely to increase. First, *P. vivax* has been thought

absent from populations of sub-Saharan Africa as they do not commonly express the Duffy antigen receptor needed by *P. vivax* to establish infection (Howes *et al.*, 2011). However, there is increasing evidence of *P. vivax* endemicity across this region, including infected travellers returning from most countries in sub-Saharan Africa (Guerra *et al.*, 2010), infected *Anopheles* vectors (Ryan *et al.*, 2006), relatively common prevalence of antibodies to pre-erythrocytic *P. vivax*-specific antigens in the Republic of the Congo (Culleton *et al.*, 2009), and – most worryingly – evidence of *P. vivax* infected Duffy negative individuals, previously thought to be refractory to infection (Menard *et al.*, 2010; Mendes *et al.*, 2011; Wurtz *et al.*, 2011). Primaquine demand will also increase in Africa following new recommendations from the WHO for low, single-dose primaquine (0.25 mg/kg) for confirmed *P. falciparum* cases as a transmission-blocking agent (Bousema and Drakeley, 2011; Eziefula *et al.*, 2012; White, 2012) without prior testing for G6PDd (*Pers. Comm*. R. Newman, Challenges in Malaria Research conference, Basel, Switzerland, 12 Oct. 2012).

The haemolytic susceptibility of the A- variant, sometimes considered "mild", has been associated with cases of transfusion-dependent haemolysis (Shekalaghe *et al.*, 2010), and caused the failure of the Lapdap antimalarial trials (Luzzatto, 2010; Pamba *et al.*, 2012). This "mild" variant has also regularly been associated with haemolysis due to ingestion of fava beans (Galiano *et al.*, 1990), previously thought to only be triggered by the more severe variants (Mehta, 1994). Haemolysis associated with this variant, while perhaps less severe than with other variants, is evidently not "mild" and the risks associated with G6PDd in African populations are an important concern. These risks were recently reviewed in relation to 0.75 mg/kg single-dose primaquine applications and considered to outweigh any community benefit which could be derived from its *P. falciparum* transmission-blocking activity (where individuals gain no direct benefit) (Eziefula *et al.*, 2012; Graves *et al.*, 2012).

Discerning the true diversity of G6PDd variants across sub-Saharan Africa is not possible from the available surveys assembled here as only three surveys used prior screening for deficiency so were able to determine the overall proportion of G6PDd cases attributable to

each variant (Figure 5.4A and 5.5A), and the remaining 81 surveys tested unscreened population samples. These later surveys had a strong bias towards a single variant (A-: encoded by G202A/A376G), ignoring the variant recently found to be most common across populations of West Africa (T968C/A376G) (De Araujo *et al.*, 2006; Clark *et al.*, 2009). Furthermore, given that electrophoretic methods were commonly used, these did not allow the relative contributions of each of the A- genotypes to be identified. Given the increasing call for primaquine use in Africa, the potentially serious haemolytic reactions to this drug, and the high prevalence of deficiency in Africa (found to average between 10-20% for most African populations (Howes *et al.*, 2012)), there is a serious need for an inventory of G6PDd variant diversity in African populations. Prior screening for deficiency allows the denominator of individuals with a significant deficiency to be identified; full gene sequencing is the only reliable way to identify with certainty all variants present in the population. Finally, large areas across the continent are not represented in the current maps (particularly central and southern Africa, and Madagascar), despite their high prevalence of G6PDd; additional surveys would be highly insightful from these areas.

5.4.3. G6PDd variants in West Asia

The Mediterranean variant, predominant across west Asia and the Arabian Peninsula and common in India, is the most severe variant to reach frequencies of public health concern. This variant's enzyme activity is virtually undetectable and commonly at less than 1% of normal activity levels (Piomelli *et al.*, 1968). Exposure of affected cells to primaquine carries a mortal risk, with reports of even single low-doses of primaquine (0.75 mg/kg) requiring transfusion to overcome the haemolytic reaction (Clyde, 1981), let alone regimens for *P. vivax* radical cure which require extended dosing of primaquine over four times that quantity (3.5 mg/kg total dose) (Graves *et al.*, 2012). The WHO guidelines therefore state that no primaquine should be given to individuals with such severe variants (WHO; WHO, 2010). The wide distribution of the severe Mediterranean variant across west Asia, where a number of countries are targeting malaria elimination (Azerbaijan, Georgia, Iran, Iraq, Kyrgyzstan, Saudi Arabia, Tajikistan, Turkey, Uzbekistan (UCSF Global Health Group and

Malaria Atlas Project, 2011)) is a major hindrance to preventing parasite re-introduction from *P. vivax* relapses. The deployment of G6PDd diagnostic capacity across these countries is essential to permit primaquine therapy.

Although the Mediterranean variant is common across Indian populations, two other variants are also commonly reported which are indigenous to populations from this sub-continent: the Kerala-Kalyan and Orissa variants. Little is known of their susceptibility to primaquine, but given this country's high malaria endemicity (Hay *et al.*, 2010; Gething *et al.*, 2012) investigation into their primaquine-sensitivity phenotypes is necessary to potentially increase access to low dosing primaquine regimens.

5.4.4. G6PDd variants in East Asia and Asia Pacific

The maps of east Asia and the west Pacific islands present the most complex picture of G6PDd variants globally, which coincides with the highest population at risk of *P. vivax* infection (Guerra *et al.*, 2010; Gething *et al.*, 2012). Virtually all the common variants of public health concern globally are reported from this region. Reasons for this high diversity are unclear, but it is interesting to note that *P. falciparum* parasites (postulated to be selective agents of G6PDd (Greene, 1993)) have been found to show a greater degree of population structure with lower genetic relatedness between populations in Asia than across Africa (Manske *et al.*, 2012). As well as this overall diversity, the structure of G6PDd variant heterogeneity is starkly different from other areas where single variants predominate. Instead, most populations were reported to have multiple variants co-occurring, with no single variants dominating.

Despite the large diversity of variants across this region where many countries are now targeting elimination (thus increasingly requiring primaquine radical cure) (UCSF Global Health Group and Malaria Atlas Project, 2011), only one variant has been examined in relation to haemolytic risk from primaquine. The Mahidol variant, found predominantly across Myanmar and parts of Thailand (Figure 5.4C(3)), reduces G6PD enzyme activity to 5-32% of normal levels (Louicharoen *et al.*, 2009). A handful of small studies have been

conducted in Thailand, reporting tolerance of G6PDd individuals to 14 day and eight weekly primaquine regimens (Myat Phone *et al.*, 1994; Buchachart *et al.*, 2001; Takeuchi *et al.*, 2010), as well as to single-dose regimens for *P. falciparum* transmission blocking (Song *et al.*, 2010). Further study of the primaquine sensitivity phenotypes of the numerous other common variants represented in these maps is urgently required to ensure safe dosing; this was also prioritised by the WHO Expert Review Group on primaquine earlier this year (Malaria Policy Advisory Committee Meeting, 2012). In the absence of these data, the Mahidol phenotype cannot be assumed to be representative of all the region's variants. Individuals identified as phenotypically deficient will not necessarily all have the same tolerance to primaquine. In populations of high heterogeneity, a dual approach of phenotypic screening followed by variant analysis may be required if moderate and severe variants are co-occurring. The danger of under-diagnosing deficiency by using molecular identification alone is well illustrated by comparison of the two types of maps (Figure 5.4C and 5.4C). If only a handful of 'common' variants are used, a proportion of deficient individuals may be missed, and put at risk from primaquine therapy, as only the variants which are looked for will be identified. The maps presented here also repeatedly highlight considerable proportions of "Other" G6PDd variants. These correspond to an unknown haemolytic risk, and hint towards an ever greater diversity of variants than currently acknowledged. Only full gene sequencing will allow full characterisation of the diversity in this gene.

5.5. Conclusions

Both the failure to treat *P. vivax* infections and the treatment itself carry risk of severe clinical complications (Malaria Policy Advisory Committee Meeting, 2012; Recht *et al.*, unpublished). Each repeated episode of acute *P. vivax* malaria carries risks of delayed or improper diagnosis, improper treatment, onward transmission, and serious illness and death (Baird, 2013). Likewise, primaquine-induced acute intravascular haemolysis in G6PDd patients may provoke renal failure and require multiple transfusions for recovery (Burgoine *et al.*, 2010). Evidence-based assessment of the risks incurred with primaquine therapy in

any given region is essential for rational strategies to minimize harm caused by the drug and the parasite. The maps presented in this chapter offer the possibility of examining these risks in relation to G6PDd prevalence across malaria endemic regions. Developing a framework for representing the haemolytic risks associated with G6PDd forms the objective of the following chapter (Chapter 6).

5.6. Acknowledgements

A modified version of this chapter will soon be submitted to a peer-review journal. The authors would like to thank Harriet Dalrymple, Suzanne Phillips and Jennie Charlton for help with the library assembly. The co-authors of the manuscript are Mewahyu Dewi[1], Fred B. Piel[2], J. Kevin Baird[1,3] and Simon Hay[4.]

1. Eijkman-Oxford Clinical Research Unit, Jalan Diponegoro No. 69, Jakarta, Indonesia,
2. Evolutionary Ecology of Infectious Disease Group, Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom,
3. Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom
4. Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom,

5.7. Author contributions

R.E.H. conceived the study and oversaw its design and implementation with guidance from F.B.P., J.K.B. and S.I.H.; R.E.H. wrote the first draft of the chapter, and assembled the data with assistance from M.D.; all authors participated in the interpretation of results and in the writing and editing of the chapter.

5.8. References

Alving, A.S., Craige, B., Pullman, T.N., et al. (1948). Procedures used at Stateville Penitentiary for the testing of potential antimalarial agents. *Journal of Clinical Investigation* **27**(3 Pt 2): 2-5.

Alving, A.S., Hankey, D.D., Coatney, G.R., et al. (1953). Korean vivax malaria. II. Curative treatment with pamaquine and primaquine. *American Journal of Tropical Medicine and Hygiene* **2**(6): 970-976.

Alving, A.S., Johnson, C.F., Tarlov, A.R., et al. (1960). Mitigation of the haemolytic effect of primaquine and enhancement of its action against exoerythrocytic forms of the Chesson strain of *Plasmodium vivax* by intermittent regimens of drug administration: a preliminary report. *Bulletin of the World Health Organization* **22**: 621-631.

Baird, J.K. (2012a). Elimination therapy for the endemic malarias. *Current Infectious Disease Reports* **14**(3): 227-237.

Baird, J.K. (2012b). Reinventing primaquine for endemic malaria. *Expert Opinion on Emerging Drugs* **17**(4): 439-444.

Baird, J.K. (2013). Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. *Clinical Microbiology Reviews* **26**(1): 1-22.

Baird, J.K. and Surjadjaja, C. (2011). Consideration of ethics in primaquine therapy against malaria transmission. *Trends in Parasitology* **27**(1): 11-16.

Betke, K., Brewer, G.J., Kirkman, H.N., et al. (1967). Standardization of procedures for the study of glucose-6-phosphate dehydrogenase. Report of a WHO Scientific Group. *World Health Organization Technical Report Series* **No. 366**: 1-53.

Beutler, E. (1959). The hemolytic effect of primaquine and related compounds: a review. *Blood* **14**(2): 103-139.

Beutler, E. (1991). Glucose-6-phosphate dehydrogenase deficiency. *New England Journal of Medicine* **324**(3): 169-174.

Beutler, E. (1994). G6PD deficiency. *Blood* **84**(11): 3613-3636.

Beutler, E. (1996). G6PD: population genetics and clinical manifestations. *Blood Reviews* **10**(1): 45-52.

Beutler, E. (2008). Glucose-6-phosphate dehydrogenase deficiency: a historical perspective. *Blood* **111**(1): 16-24.

Beutler, E. and Duparc, S. (2007). Glucose-6-phosphate dehydrogenase deficiency and antimalarial drug development. *American Journal of Tropical Medicine and Hygiene* **77**(4): 779-789.

Beutler, E., Westwood, B. and Kuhl, W. (1991). Definition of the mutations of G6PD Wayne, G6PD Viangchan, G6PD Jammu, and G6PD 'LeJeune'. *Acta Haematologica* **86**(4): 179-182.

Bousema, T. and Drakeley, C. (2011). Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. *Clinical Microbiology Reviews* **24**(2): 377-410.

Buchachart, K., Krudsood, S., Singhasivanon, P., et al. (2001). Effect of primaquine standard dose (15 mg/day for 14 days) in the treatment of vivax malaria patients in Thailand. *Southeast Asian Journal of Tropical Medicine and Public Health* **32**(4): 720-726.

Burgoine, K.L., Bancone, G. and Nosten, F. (2010). The reality of using primaquine. *Malaria Journal* **9**: 376.

Cappellini, M.D. and Fiorelli, G. (2008). Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**(9606): 64-74.

Carson, P.E., Flanagan, C.L., Ickes, C.E., et al. (1956). Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science* **124**(3220): 484-485.

Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1994). The History and Geography of Human Genes. Princeton, New Jersey, Princeton University Press.

Clark, T.G., Fry, A.E., Auburn, S., et al. (2009). Allelic heterogeneity of G6PD deficiency in West Africa and severe malaria susceptibility. *European Journal of Human Genetics* **17**(8): 1080-1085.

Clyde, D.F. (1981). Clinical problems associated with the use of primaquine as a tissue schizontocidal and gametocytocidal drug. *Bulletin of the World Health Organization* **59**(3): 391-395.

Coatney, G.R., Alving, A.S., Jones, R., Jr., et al. (1953). Korean vivax malaria. V. Cure of the infection by primaquine administered during long-term latency. *American Journal of Tropical Medicine and Hygiene* **2**(6): 985-988.

Culleton, R., Ndounga, M., Zeyrek, F.Y., et al. (2009). Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. *Journal of Infectious Diseases* **200**(9): 1465-1469.

De Araujo, C., Migot-Nabias, F., Guitard, J., et al. (2006). The role of the G6PD A-376G/968C allele in glucose-6-phosphate dehydrogenase deficiency in the seerer population of Senegal. *Haematologica* **91**(2): 262-263.

Dern, R.J., Beutler, E. and Alving, A.S. (1954). The hemolytic effect of primaquine. II. The natural course of the hemolytic anemia and the mechanism of its self-limited character. *Journal of Laboratory and Clinical Medicine* **44**(2): 171-176.

Edgcomb, J.H., Arnold, J., Yount, E.H., Jr., et al. (1950). Primaquine, SN 13272, a new curative agent in vivax malaria; a preliminary report. *Journal of the National Malaria Society* **9**(4): 285-292.

Eziefula, A.C., Gosling, R., Hwang, J., et al. (2012). Rationale for short course primaquine in Africa to interrupt malaria transmission. *Malaria Journal* **11**: 360.

Flanagan, C.L., Schrier, S.L., Carson, P.E., et al. (1958). The hemolytic effect of primaquine. VIII. The effect of drug administration on parameters of primaquine sensitivity. *Journal of Laboratory and Clinical Medicine* **51**(4): 600-608.

Galiano, S., Gaetani, G.F., Barabino, A., et al. (1990). Favism in the African type of glucose-6-phosphate dehydrogenase deficiency (A-). *BMJ: British Medical Journal* **300**(6719): 236.

Gething, P.W., Elyazar, I.R., Moyes, C.L., et al. (2012). A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Neglected Tropical Diseases* **6**(9): e1814.

Gething, P.W., Patil, A.P., Smith, D.L., et al. (2011). A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal* **10**: 378.

Graves, P.M., Gelband, H. and Garner, P. (2012). Primaquine for reducing *Plasmodium falciparum* transmission. *Cochrane Database of Systematic Reviews* **9**: CD008152.

Greene, L.S. (1993). G6PD deficiency as protection against *falciparum*-malaria: an epidemiologic critique of population and experimental studies. *Yearbook of Physical Anthropology* **36**: 153-178.

Guerra, C.A., Howes, R.E., Patil, A.P., et al. (2010). The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Neglected Tropical Diseases* **4**(8): e774.

Hay, S.I., Gething, P.W. and Snow, R.W. (2010). India's invisible malaria burden. *Lancet* **376**(9754): 1716-1717.

Hay, S.I., Guerra, C.A., Tatem, A.J., et al. (2004). The global distribution and population at risk of malaria: past, present, and future. *Lancet Infectious Diseases* **4**(6): 327-336.

Hill, D.R., Baird, J.K., Parise, M.E., et al. (2006). Primaquine: report from CDC expert meeting on malaria chemoprophylaxis I. *American Journal of Tropical Medicine and Hygiene* **75**(3): 402-415.

Hockwald, R.S., Arnold, J., Clayman, C.B., et al. (1952). Toxicity of primaquine in Negroes. *JAMA: The Journal of the American Medical Association* **149**(17): 1568-1570.

Howes, R.E., Battle, K.E., Satyagraha, A.W., et al. (2013). G6PD deficiency: Global distribution, genetic variants and primaquine therapy. *Advances in Parasitology* **81**: In press.

Howes, R.E., Patil, A.P., Piel, F.B., et al. (2011). The global distribution of the Duffy blood group. *Nature Communications* **2**: 266.

Howes, R.E., Piel, F.B., Patil, A.P., et al. (2012). G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Medicine* **9**(11): e1001339.

Hung, N.M., Eto, H., Mita, T., et al. (2008). Glucose - 6 - phosphate dehydrogenase (G6PD) variants in East Sepik province of Papua New Guinea: G6PD Jammu, G6PD Vanua Lava, and a novel variant (G6PD Dagua). *Tropical Medicine and Health* **36**(4): 163-169.

Iwai, K., Hirono, A., Matsuoka, H., et al. (2001). Distribution of glucose-6-phosphate dehydrogenase mutations in Southeast Asia. *Human Genetics* **108**(6): 445-449.

Kim, S., Nguon, C., Guillard, B., et al. (2011). Performance of the CareStart G6PD deficiency screening test, a point-of-care diagnostic for primaquine therapy screening. *PLoS One* **6**(12): e28357.

Kwiatkowski, D.P. (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *American Journal of Human Genetics* **77**(2): 171-192.

Laosombat, V., Sattayasevana, B., Chotsampancharoen, T., et al. (2006). Glucose-6-phosphate dehydrogenase variants associated with favism in Thai children. *International Journal of Hematology* **83**(2): 139-143.

Livingstone, F.B. (1985). Frequencies of Hemoglobin Variants: Thalassemia, the Glucose-6-Phosphate Dehydrogenase Deficiency, G6PD Variants and Ovalocytosis in Human Populations. New York, Oxford University Press.

Louicharoen, C., Patin, E., Paul, R., et al. (2009). Positively selected G6PD-Mahidol mutation reduces *Plasmodium vivax* density in Southeast Asians. *Science* **326**(5959): 1546-1549.

Luisada, A. (1940). Favism. *JAMA: The Journal of the American Medical Association* **115**(8): 632-632.

Luzzatto, L. (2006). Glucose 6-phosphate dehydrogenase deficiency: from genotype to phenotype. *Haematologica* **91**(10): 1303-1306.

Luzzatto, L. (2009). Glucose-6-phosphate dehydrogenase deficiency. In: Nathan and Oski's Hematology of Infancy and Childhood. S.H. Orkin, D.G. Nathan, D. Ginsburg, et al. (eds). Philadelphia, Saunders.

Luzzatto, L. (2010). The rise and fall of the antimalarial Lapdap: a lesson in pharmacogenetics. *Lancet* **376**(9742): 739-741.

Luzzatto, L. and Allan, N.C. (1968). Relationship between the genes for glucose-6-phosphate dehydrogenase and for haemoglobin in a Nigerian population. *Nature* **219**(5158): 1041-1042.

Luzzatto, L. and Notaro, R. (2001). Malaria. Protecting against bad air. *Science* **293**(5529): 442-443.

Mahgoub, H., Gasim, G.I., Musa, I.R., et al. (2012). Severe *Plasmodium vivax* malaria among Sudanese children at New Halfa Hospital, Eastern Sudan. *Parasites & Vectors* **5**(1): 154.

Malaria Policy Advisory Committee Meeting (2012). WHO Evidence Review Group Report: the safety and effectiveness of single dose primaquine as a *P. falciparum* gametocytocide. Geneva. WHO.

Manske, M., Miotto, O., Campino, S., et al. (2012). Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**(7407): 375-379.

Mason, P.J., Bautista, J.M. and Gilsanz, F. (2007). G6PD deficiency: the genotype-phenotype association. *Blood Reviews* **21**(5): 267-283.

Mason, P.J. and Vulliamy, T.J. (2005). Glucose-6-phosphate dehydrogenase (G6PD) deficiency: genetics. Encyclopedia of Life Sciences, John Wiley & Sons, Ltd.

Matsuoka, H., Wang, J., Hirai, M., et al. (2004). Glucose-6-phosphate dehydrogenase (G6PD) mutations in Myanmar: G6PD Mahidol (487G>A) is the most common variant in the Myanmar population. *Journal of Human Genetics* **49**(10): 544-547.

Mehta, A.B. (1994). Glucose-6-phosphate dehydrogenase deficiency. *Postgraduate Medical Journal* **70**(830): 871-877.

Meloni, T., Forteleoni, G., Dore, A., et al. (1983). Favism and hemolytic anemia in glucose-6-phosphate dehydrogenase-deficient subjects in North Sardinia. *Acta Haematologica* **70**(2): 83-90.

Menard, D., Barnadas, C., Bouchier, C., et al. (2010). *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proceedings of the National Academy of Sciences of the United States of America* **107**(13): 5967-5971.

Mendes, C., Dias, F., Figueiredo, J., et al. (2011). Duffy negative antigen is no longer a barrier to *Plasmodium vivax*--molecular evidences from the African West Coast (Angola and Equatorial Guinea). *PLoS Neglected Tropical Diseases* **5**(6): e1192.

Mendis, K., Sina, B.J., Marchesini, P., et al. (2001). The neglected burden of *Plasmodium vivax* malaria. *American Journal of Tropical Medicine and Hygiene* **64**(1-2 Suppl): 97-106.

Mihaly, G.W., Ward, S.A., Edwards, G., et al. (1985). Pharmacokinetics of primaquine in man. I. Studies of the absolute bioavailability and effects of dose size. *British Journal of Clinical Pharmacology* **19**(6): 745-750.

Minucci, A., Moradkhani, K., Hwang, M.J., et al. (2012). Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. *Blood Cells Molecules and Diseases* **48**(3): 154-165.

Most, H., Kane, C.A. and et al. (1946). Combined quinine-plasmochin treatment of vivax malaria; effect of relapse rate. *The American Journal of the Medical Sciences* **212**(5): 550-560.

Mourant, A.E., Kopec, A.C. and Domaniewska-Sobczak, K. (1976). The Distribution of the Human Blood Groups and other Polymorphisms. London, Oxford University Press.

Myat Phone, K., Myint, O., Aung, N., et al. (1994). The use of primaquine in malaria infected patients with red cell glucose-6-phosphate dehydrogenase (G6PD) deficiency in Myanmar. *Southeast Asian Journal of Tropical Medicine and Public Health* **25**(4): 710-713.

Nkhoma, E.T., Poole, C., Vannappagari, V., et al. (2009). The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. *Blood Cells Molecules and Diseases* **42**(3): 267-278.

Pamba, A., Richardson, N.D., Carter, N., et al. (2012). Clinical spectrum and severity of hemolytic anemia in glucose 6-phosphate dehydrogenase-deficient children receiving dapsone. *Blood* **120**(20): 4123-4133.

Piomelli, S., Corash, L.M., Davenport, D.D., et al. (1968). In vivo lability of glucose-6-phosphate dehydrogenase in Gd^{A-} and $Gd^{Mediterranean}$ deficiency. *Journal of Clinical Investigation* **47**(4): 940-948.

Recht, J., Ashley, E.A. and White, N.J. (unpublished). 8-aminoquinolines safety review for WHO primaquine ERG.

Ryan, J.R., Stoute, J.A., Amon, J., et al. (2006). Evidence for transmission of *Plasmodium vivax* among a duffy antigen negative population in Western Kenya. *American Journal of Tropical Medicine and Hygiene* **75**(4): 575-581.

Shekalaghe, S.A., ter Braak, R., Daou, M., et al. (2010). In Tanzania, hemolysis after a single dose of primaquine coadministered with an artemisinin is not restricted to glucose-6-phosphate dehydrogenase-deficient (G6PD A-) individuals. *Antimicrobial Agents and Chemotherapy* **54**(5): 1762-1768.

Singh, H., Parakh, A., Basu, S., et al. (2011). *Plasmodium vivax* malaria: is it actually benign? *Journal of Infection and Public Health* **4**(2): 91-95.

Singh, S. (1973). Distribution of certain polymorphic traits in populations of the Indian peninsula and South Asia. *Israel Journal of Medical Sciences* **9**(9): 1225-1237.

Song, J., Socheat, D., Tan, B., et al. (2010). Rapid and effective malaria control in Cambodia through mass administration of artemisinin-piperaquine. *Malaria Journal* **9**: 57.

Takeuchi, R., Lawpoolsri, S., Imwong, M., et al. (2010). Directly-observed therapy (DOT) for the radical 14-day primaquine treatment of *Plasmodium vivax* malaria on the Thai-Myanmar border. *Malaria Journal* **9**: 308.

UCSF Global Health Group and Malaria Atlas Project (2011). Atlas of Malaria-Eliminating Countries. San Francisco. Unversity of California.

Vaca, G., Arambula, E. and Esparza, A. (2002). Molecular heterogeneity of glucose-6-phosphate dehydrogenase deficiency in Mexico: overall results of a 7-year project. *Blood Cells Molecules and Diseases* **28**(3): 436-444.

White, N.J. (2008). The role of anti-malarial drugs in eliminating malaria. *Malaria Journal* **7 Suppl 1**: S8.

White, N.J. (2011). Determinants of relapse periodicity in *Plasmodium vivax* malaria. *Malaria Journal* **10**: 297.

White, N.J. (2012). Primaquine to prevent transmission of falciparum malaria. *Lancet Infectious Diseases*: doi:10.1016/S1473-3099(1012)70198-70196.

WHO. Country antimalarial drug policies: by region. Accessed: 1 May 2012. URL: http://www.who.int/malaria/am_drug_policies_by_region_afro/en/index.html.

WHO (2010). Guidelines for the treatment of malaria, second edition. Geneva: World Health Organization.

WHO (2011). Global plan for artemisinin resistance containment (GPARC).

Wurtz, N., Mint Lekweiry, K., Bogreau, H., et al. (2011). Vivax malaria in Mauritania includes infection of a Duffy-negative individual. *Malaria Journal* **10**: 336.

Chapter 6 – Towards a haemolytic risk assessment framework for primaquine therapy

In the preceding two chapters, I have synthesised available knowledge on the spatial epidemiology of G6PD deficiency (G6PDd) in malaria endemic countries (MECs). In Chapter 4, I assessed the magnitude of the public health problem of G6PDd through spatial modelling of the prevalence of deficiency. G6PDd was found to be widespread, with a median predicted allele frequency of 8.0% (IQR: 7.4-8.8%), corresponding to an estimated 300 million affected individuals across the 99 malaria endemic countries. In Chapter 5, I mapped the genetic diversity of this polymorphic disorder. G6PDd variants were shown to have striking geographic patterns, with large areas of Africa predominated by a single variant, contrasting with as many as ten different variants reported from a single population survey in Southeast Asia. G6PDd variants range in their enzyme activity across a spectrum from normal to virtual none. Correspondingly, red blood cells (RBCs) affected by G6PDd differ in their susceptibility to primaquine-induced haemolysis according to the variant expressed. Symptoms vary in severity from mild and self-limiting (e.g. with the A- or Mahidol variants), to severe and requiring transfusion (e.g. Mediterranean variant) (WHO Working Group, 1989). This present chapter considers how these two evidence-bases – the prevalence and variant maps – may be used in an assessment of the relative risks of primaquine therapy: a vital, yet imperfect, drug.

Although primaquine therapy is fraught by significant risks of severe adverse events, its advantages as a therapeutic agent are also important, and are particularly valuable as part of elimination therapy through its two unique applications (Baird, 2012a):

1. Primaquine is the only drug effective against the mature gametocytes of *Plasmodium falciparum*, a therapeutic target to prevent onward parasite transmission. Although artemisinin combination therapies (ACTs) and other commonly used antimalarial drugs are active against early stage gametocytes, these

are not effective against the mature infectious stages (Bousema and Drakeley, 2011). Relatively low doses of primaquine are effective against this blood-stage target. For individuals with normal G6PD activity, current WHO guidelines recommend a single 0.75 mg/kg dose, to be administered alongside a blood-stage schizontocide (WHO, 2010). Transmission-blocking primaquine dosing is not recommended in G6PDd individuals. The benefits of preventing transmission are particularly far-reaching in areas of emerging drug-resistance, and transmission-blocking primaquine is specifically recommended by the ACT-resistance containment programme (WHO, 2011). The application of primaquine to block *P. falciparum* transmission has recently received increased attention (Eziefula *et al.*, 2012; Graves *et al.*, 2012; Malaria Policy Advisory Committee Meeting, 2012; White, 2012; Recht *et al.*, unpublished), including from a WHO Expert Review Group convened in August 2012. Their review of the safety and efficacy of single dose transmission blocking determined that a lower single dose (0.25 mg/kg dose) would be safer whilst remaining effective. This dosing is now recommended for more widespread use, particularly in areas of resistance emergence, without the prior requirement for G6PDd testing (Malaria Policy Advisory Committee Meeting, 2012).

2. Primaquine is currently the only available option for eliminating the reservoir of relapsing *Plasmodium vivax* hypnozoites, which otherwise present a major challenge to achieving elimination (Baird, 2012b). This application of primaquine requires a total dose of 200 mg, usually administered as 14 daily 0.25 mg/kg doses in G6PD normal patients (increasing to 0.5 mg/kg in Southeast Asia and Oceania where the *P. vivax* Chesson strain occurs). For patients with mild G6PDd, WHO guidelines recommend eight weekly doses of 0.75 mg/kg of primaquine to reduce the drug's toxic effects (WHO, 2010). On the other hand, primaquine is contraindicated by the WHO for patients with severe deficiency. The WHO

guidelines do not define what enzyme activities thresholds or genetic variants correspond to ‘mild’ and ‘severe’ G6PDd (WHO, 2010).

Given that consideration for G6PDd status is no longer required for *P. falciparum* transmission blocking regimens of primaquine, I instead focus here on the risks associated with primaquine as a radical cure of *P. vivax*. G6PDd screening is pre-requisite to this application, as the high primaquine doses necessary for effective treatment are potentially very severe in G6PDd patients. The lack of G6PDd screening capacity in many areas prevents widespread use of the drug, as the risk of inducing haemolysis following treatment is frequently perceived to be too great. Given the far-reaching benefits of ensuring *P. vivax* radical cure, understanding how to appropriately balance the risks and benefits of using primaquine has important implications: the severity of both primaquine-induced haemolysis and untreated *P. vivax* can be severe, even lethal (Burgoine *et al.*, 2010; Baird, 2013; Recht *et al.*, unpublished). First, I explore primaquine-induced haemolysis: how its severity is thought to be influenced, the mechanism of primaquine-induced haemolysis, and its clinical manifestations. Next, I consider how the overall risk of primaquine-induced adverse events may be considered at large public-health scales. Finally, I consider the knowledge gaps and studies required to improve the resolution and public health applications of the framework discussed here.

The analysis included in Section 6.2 was also briefly discussed in Chapter 4 (this was a requirement for publication of that chapter), but is described fully in this chapter. Parts of this chapter are also set for publication in February 2013 in a special issue of *Advances in Parasitology* dedicated to *Plasmodium vivax* (Chapter 4, Volume 81 (Howes *et al.*, 2013)).

6.1. Primaquine-induced haemolysis

6.1.1. G6PD enzyme as an anti-oxidant defence

RBCs at risk of haemolysis are those which cannot protect themselves from oxidative challenges. The G6PD enzyme plays a critical role in maintaining RBC integrity through catalysing a key step in the pentose phosphate pathway (PPP) (Figure 6.1). This pathway generates the cell's anti-oxidant power by reducing NADP to NADPH, which in turn sustains the reserves of reduced glutathione which are essential to neutralizing oxidative challenges (Pandolfi *et al.*, 1995). The PPP is particularly important in RBCs as the absence of mitochondria in these cells means that they have no alternative pathways for reducing NADPH. The reducing power supplied by the PPP is necessary to neutralise oxidative challenges, such as hydrogen peroxide or free radicals, which would otherwise cause irreversible damage to the cell (especially the cell membrane), leading to cell apoptosis and clinical episodes of acute haemolytic anaemia (Greene, 1993). Given the common circulation of oxidative challenges in the blood, the G6PD enzyme, which catalyses the rate-limiting step of the PPP is therefore essential for RBC survival. Enzyme activity decays naturally with cell age, and it is estimated that in normal blood, reticulocytes have about five times higher G6PD activity levels than that of the oldest 10% of RBCs (Luzzatto, 2006). The oldest cells are therefore those most vulnerable to oxidative challenges.

6.1.2. Effect of reduced G6PD enzyme activity

Mutations to the G6PD gene encode enzyme variants with disrupted enzyme structure and therefore reduced activity levels (in some rare exceptions, the mutations increase enzyme activity). In these cells, the ageing process is effectively sped up, with a larger proportion of cells having lower enzyme levels and being at increased risk of oxidative damage. Depending upon the severity of the impact of the mutations, the proportion of cells at risk of haemolysis varies. The clinical symptoms resulting from exposure of cells to oxidative challenges range from being asymptomatic and self-limiting (in cases where haemolysed

cells can be rapidly replaced) to highly severe (when haemolysed cells cannot be regenerated fast enough to compensate for the high proportion of lost cells) (Alving *et al.*, 1960; Howes *et al.*, 2013). All individuals with a deficiency in enzyme activity are therefore at risk of haemolysis, but it is the severity of that haemolysis which is of concern, and which is determined by the proportion of cells affected (Malaria Policy Advisory Committee Meeting, 2012) – if small, the haemolysis will be clinically negligible. As discussed in the following sections, the mutations are the major determinant of the severity of haemolysis. The variant maps presented in Chapter 5 represent an attempt to encapsulate the variability in the spatial distribution of these predisposing mutations to haemolysis.

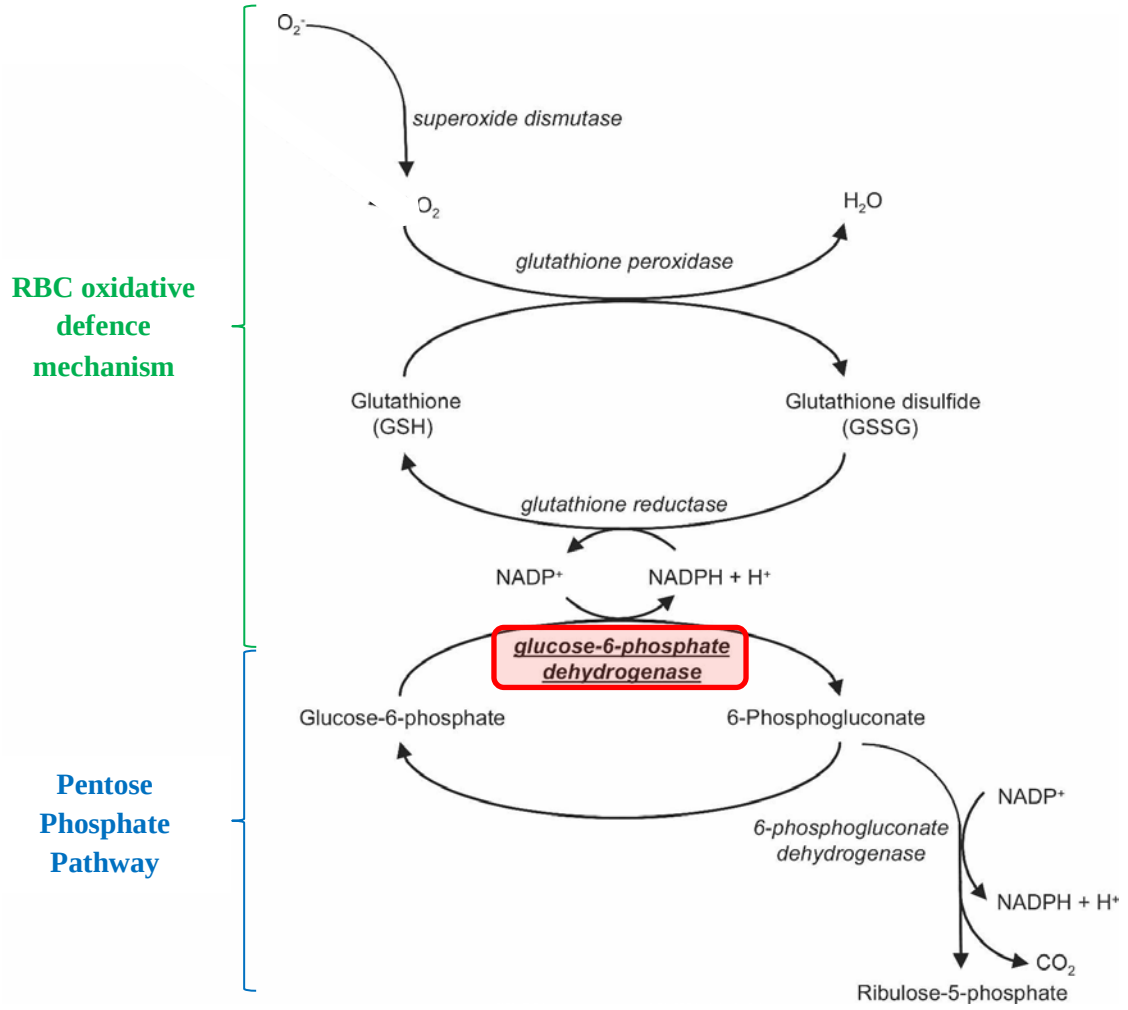


Figure 6.1. Section of the pentose phosphate pathway (PPP) and the role of the G6PD enzyme as a driver of RBC oxidative defence. NADP: nicotinamide adenine dinucleotide phosphate; NADPH: reduced form of NADP; O_2^- represents an oxidative stress (e.g. hydrogen peroxide or free radicals); enzymes are named in italics. Figure modified from Beutler and Duparc (2007).

6.1.3. Determinants of haemolytic severity in G6PDd cells

At the individual level, severity of primaquine-induced haemolysis is highly variable, influenced by the residual enzyme activity (encoded by different G6PDd variants), as well as by a combination of other influences including exogenous and endogenous factors, as discussed below (Beutler, 1994; Luzzatto, 2009). It is important to recall that the primaquine-sensitivity phenotype of most genetic variants has so far been poorly documented (Baird and Surjadjaja, 2011), as discussed in Chapter 5. Other major factors which affect haemolytic severity are briefly discussed here.

First, the dose-dependency of the severity of primaquine-induced haemolysis has been long recognised, with higher doses of primaquine triggering more severe haemolytic reactions. This is made evident by many cases of severe adverse reactions to primaquine being due to drug overdoses (Burgoine *et al.*, 2010; Recht *et al.*, unpublished). Haemolysis can be averted to an extent by extending the dosing regimens over a long period of time (Alving *et al.*, 1960), without compromising it's therapeutic activity (Mihaly *et al.*, 1985). Second, pre-existing infections or anaemia will alter the age distribution of RBCs, with a potentially greater proportion of reticulocytes expressing high G6PD activity (for this reason, phenotypic diagnoses during episodes of haemolytic anaemia when reticulocytosis is upregulated are likely to give false-negative results). Third, as explained in Chapter 4, the sex-linked inheritance of the G6PD gene means that males are more likely to suffer severe adverse reactions to primaquine, though homozygous females will be at the same risk as affected males, and the proportion of cells carrying the deficient gene in heterozygotes (which may be a large proportion) will also be at haemolytic risk. The severity of clinical symptoms will be determined by the proportion of haemolysed cells. Finally, it has also been noted that severe haemolysis is more life-threatening in children (Recht *et al.*, unpublished).

It is important to remember the complex interplay of factors affecting the outcome of primaquine therapy at the individual level. However, in the public health context of assessing relative risks associated with primaquine use between populations, I consider the

influence of enzyme activity (as determined by the characteristics of the local G6PDd variants) to be the primary predictor of severe or mild adverse events. The influence of the other factors in differentiating the level of risk between populations is likely to be relatively constant, and not the major determinant of the spatial variability of haemolytic risk at this large scale.

6.1.4. Mechanism of primaquine-induced haemolysis

Despite the haemolytic threats associated with primaquine having been recognised for six decades, the intracellular mechanisms of haemolysis remain uncertain. Understanding the mechanisms of primaquine toxicity to G6PDd RBCs is important to understanding the relationship between the residual enzyme activity level of different variants and their susceptibility to severe haemolysis. Furthermore, understanding the molecular events leading to haemolysis is also necessary for future therapeutic developments of non-toxic drugs in which it may be possible to disassociate primaquine's toxicity from its therapeutic properties (Pybus *et al.*, 2012).

Primaquine has a short half-life of only about four hours (Greaves *et al.*, 1980; Carson *et al.*, 1981), rapidly metabolised into a complex array of a dozen or so distinct moieties (primaquine-induced haemolysis usually appears after two days, however (Reeve *et al.*, 1992)). Several of these metabolites reach plasma concentrations 10-times that of primaquine (e.g. carboxy-primaquine (Mihaly *et al.*, 1985)), and may be much more potent than primaquine as oxidative agents in stimulating the PPP (e.g. 5-hydroxy-6-methoxy-8-aminoquinoline was 2500-times more potent (Baird *et al.*, 1986)). It is therefore likely to be one of several metabolic products of primaquine which is the active agent against the parasites, rather than the parent compound itself (Beutler, 1969; Carson *et al.*, 1981; Fletcher *et al.*, 1988); however, it remains for any of these to be conclusively implicated with activity against the *Plasmodium* parasite (Baird and Hoffman, 2004; Myint *et al.*, 2011).

A number of mechanisms for primaquine-induced haemolysis have been proposed. One long-held hypothesis is that haemolysis results from direct oxidative stress caused by primaquine metabolites. Investigation of oxidized glutathione, which accumulates during oxidative stress, was found to be a strong intracellular mediator activating membrane cation channels (Koliwad *et al.*, 1996). The opening of these Ca^{2+}-permeable channels triggered cell death (apoptosis) during oxidative stress (Lang *et al.*, 2003; Lang *et al.*, 2006), though these observations are not conclusive (Ganesan *et al.*, 2012). Oxidative activity from primaquine metabolites has also been found to cause injury to the erythrocyte cytoskeleton (Bowman *et al.*, 2005b), accelerating the process of cell phagocytosis (Bowman *et al.*, 2005a). In a second theory, the build-up of methaemoglobin (met-Hb) can result in tissue hypoxia, hypoemia and cyanosis due to met-Hb's low affinity for oxygen (Percy *et al.*, 2005; Hill *et al.*, 2006). Oxidized primaquine derivatives have been found to be strongly associated with the formation of met-Hb (Link *et al.*, 1985), which results from the oxidation of haemoglobin iron ($Fe^{2+} \rightarrow Fe^{3+}$). However, methaemoglobinaemia is not a pronounced feature of acute haemolytic anaemia in G6PDd patients. A third set of evidence points away from direct damage mediated through oxidative degradation of the RBC cytosol or membrane. Baird and colleagues observed that the increased PPP activity stimulated by primaquine metabolites occurred independently of glutathione redox activity. They suggested that redox equilibrium between a reduced and oxidized species of primaquine would, in a G6PDd cell, strongly favour the oxidized species, without necessarily prompting a broad oxidative degradation of cytosol proteins (Brueckner *et al.*, 2001). The oxidised species could therefore be the agent of haemolysis. Any mechanism of haemolysis must ultimately be reconciled with what is highly likely to be a very brief and quantitatively insubstantial oxidative challenge to the RBC by primaquine. It would therefore appear that a general oxidative stress would be insufficient, but that an irreversible accumulation of the harmful primaquine species (and its damage) is being captured within the RBC. This is consistent with evidence that haemolysis usually starts in earnest after the third or fourth daily dose. For example, an accumulation of displaced haem

molecules in the RBC membrane, manifest as Heinz bodies, is compatible with all these features.

Whatever the mechanism, it seems to be that the degree of damage in the cell is influenced by the rate of the PPP: the level of direct oxidative damage, the amount of met-Hb accumulated, or the position of equilibrium between reduced and oxidised species in the cell. G6PD enzyme activity, which is the rate limiting step of the PPP, would appear to be directly associated with the severity of the haemolysis. Correspondingly, it is well established that primaquine-induced haemolysis does not occur in individuals with normal levels of G6PD activity (Edgcomb *et al.*, 1950; Baird *et al.*, 2001). The limited pharmacokinetic data available appears to support an inverse correlation of residual enzyme activity with severity of potential primaquine-induced harm (Baird and Surjadjaja, 2011). As explained in Chapter 5, three G6PDd variants have been characterised in detail in relation to their susceptibility to primaquine-induced haemolysis. The Mediterranean variant (expressing ca. <1% activity levels (Piomelli *et al.*, 1968)) is associated with very severe symptoms (Beutler, 1991), while the Mahidol (ca. 5-32% residual enzyme activity (Louicharoen *et al.*, 2009)) and the A- variants (ca. 10% residual enzyme activity (Beutler, 1991)) are usually associated with milder, self-limiting haemolysis (Dern *et al.*, 1954). Although this small subset of variants fits this correlation, a universal relationship across all variants requires investigation of a greater number of common G6PDd variants.

6.1.5. Clinical manifestations of primaquine-induced haemolysis

The physiological damage caused by exposing G6PDd RBCs to primaquine leads to intravascular haemolysis, making acute haemolytic anaemia the main clinical symptom. Haemolytic attacks are typically characterised by malaise, weakness, and abdominal or lower back pain. After a few hours or days, patients will develop jaundice and dark urine due to haemoglobinuria (Luzzatto, 2006). Freely circulating haemoglobin from haemolysed cells causes the most severe and potentially lethal conditions, including haemoglobinuria and acute renal failure (Burgoine *et al.*, 2010; Recht *et al.*, unpublished). A full review of

severe adverse reactions from primaquine, and primaquine-induced death, has recently been conducted (Recht *et al.*, unpublished). It is important to recall, that all G6PDd individuals will be haemolysed to an extent by exposure to primaquine, but it is the severity of that haemolytic event (determined by the proportion of haemolysed cells) which is of clinical concern. Measures of residual enzyme activity are the only proxy indication of the severity of haemolytic risk.

6.2. Assessing national-level haemolytic risk from primaquine therapy

Bringing together the various streams of evidence about primaquine-associated risks into a framework applicable to public health questions is important to applying the research outputs of Chapters 4 and 5. This section explores how the epidemiological datasets of the prevalence and variants of G6PDd may be applied in classifying relative haemolytic risk between areas. I focus here on large regional scales for public health perspectives. Risk at the individual level should always be assessed directly by clinicians, and large scale maps can never supersede the need for individual level drug supervision. As discussed later in this chapter, underlying data simply does not exist for modelling quantitative estimates of adverse haemolytic events under different scenarios of primaquine policy. Instead, qualitative national level risk maps are generated to answer questions of relative risk: whether the requirement for G6PDd screening could be lifted in some areas; or whether testing needs to be mandatory everywhere; whether molecular diagnostics in addition to G6PDd phenotypic screening out to be used to differentiate the severity of the deficiency. This chapter makes no direct attempt to answer these questions, but instead provides the evidence-base to support local medical and public health experts. The level of acceptable risk will vary, for instance, according to the local laboratory facilities, medical infrastructure and ability to provide emergency transfusions, to the baseline health status of the population, and to the level of malaria endemicity (to gauge the relative frequency with which primaquine would need to be used, and the impact that a primaquine policy would have).

Although there are multiple factors which determine the severity of haemolysis at the individual patient level, in making national-level predictions, I consider the main spatially variable predictor of haemolysis to be the underlying G6PDd variants. The relative influence of the other factors (e.g. age, concurrent infection) would be relatively constant across all areas. As described in the previous section, current hypotheses indicate that the rate of the PPP (itself directly determined by the G6PD enzyme activity rate) relates to the severity of haemolysis. The main assumption made in this classification of risk, therefore, is that enzyme activity levels are an indicator of the severity of haemolytic risk. There are certainly exceptions to this rule. For instance, an Iranian boy with 19.5% residual activity (usually considered 'mild deficiency') required a transfusion after a single 45 mg primaquine dose (Ziai *et al.*, 1967), and a 5-year old heterozygote Tanzanian experienced a severe adverse reaction (Hb level <5 g/dL) also following a single 15 mg dose (Shekalaghe *et al.*, 2010). These are important reminders that large public-health scale policies do not supersede the need for careful medical monitoring of primaquine therapy.

In an attempt to classify overall national-level risk from G6PDd, I propose a simple framework synthesising the two epidemiological datasets presented in Chapters 4 and 5. This is a coarse-scaled attempt highly constrained in its scope by the important limitations to the data informing this analysis: one which must be refined as more data becomes available about risk of haemolysis.

6.2.1. Proposed framework for ranking national-level risk from G6PDd

Current WHO treatment guidelines consider haemolytic risk to differ between "mild/moderate" and "severe" classes of G6PDd. "Mild/moderate" cases may be treated with the eight weekly 0.75 mg/kg dosing regimen, while "severe" patients should not be administered any primaquine (WHO; WHO, 2010). In the absence of more specific categorisation of risk for most variants, the G6PDd population at risk of haemolysis resulting from treatment with primaquine may also be considered to be at these two levels of risk. The overall public health risk presented to a population by the use of primaquine

may be considered contingent on two factors: (i) the prevalence of deficiency, and (ii) the relative composition of "mild" and "severe" G6PDd variants reported. The dataset used to inform the first factor was the population-weighted, national-level frequency map presented in Chapter 4. The second dataset, informing the severity of local variants was more involved, and is described in the following section.

6.2.2. A database of G6PDd variants

A database of occurrences of named variants was assembled from the same literature search as described in Chapter 5. The objective of this dataset was to reflect the most commonly reported variants from each country. This meant that all the criteria imposed on the dataset described in Chapter 5 were not necessary to apply, and patient and case reports could be included. All occurrences of specified variants from named malaria endemic countries were recorded, generating a total of 1,468 positive occurrence reports across malarious regions. In the same way as in Chapter 5, duplicated biochemical variants were aggregated according to their genetic mutations using the tables by Mason *et al*. (Mason *et al.*, 2007) in 2007 and recently updated by Minucci *et al*. (Minucci *et al.*, 2012).

6.2.3. A variant severity classification system

A simple classification system of G6PDd variants (Yoshida *et al.*, 1971), endorsed by the WHO (WHO Working Group, 1989), groups variants into five classes according to their residual enzyme activity levels, their clinical characteristics, and their frequency within populations (polymorphic/sporadic) (Luzzatto *et al.*, 2001) (Table 6.1). I used this classification system to distinguish "mild" from "severe" variants. Class I variants are associated with the most severe clinical symptoms (CNSHA, see Chapter 5). The chronic anaemia associated with Class I variants means that they never reach polymorphic frequencies (defined as ≥1% prevalence) and as such are not of major public health concern and could therefore be excluded from this analysis. Class IV and V variants do not express significantly reduced enzyme activity, so are therefore not relevant to this study.

Class	Degree of deficiency	Residual enzyme activity	Clinical characteristics
I	Severe	Virtually none	Chronic non-spherocytic haemolytic anaemia
II	Severe	<10%	Acute haemolytic anaemia
III	Moderate	10 – 60%	Occasional haemolysis
IV	Normal	60 – 150%	None
V	Increased activity	>150%	None

Table 6.1. G6PDd variant classifications, based on residual enzyme activity levels, the severity of clinical symptoms, and the frequency at the population-level. Table adapted from WHO Working Group (1989) and Cappellini and Fiorelli (2008).

Class II and III variants do express a significant reduction of enzyme activity (<10% and 10-60% residual activity levels, respectively) and represent the most commonly inherited variants, reaching polymorphic frequencies across communities. The clinical phenotypes of Class II variants (e.g. Mediterranean) are more severe than those of Class III variants (e.g. A- and Mahidol). These are the classes of variants mapped in Chapter 5. A series of published databases assign G6PDd variants into these WHO classes (Dr Andrew C. R. Martin's Group; Yoshida *et al.*, 1971; Beutler, 1993; Vulliamy *et al.*, 1997; Luzzatto *et al.*, 2001; Kwok *et al.*, 2002; Minucci *et al.*, 2012). Where there was discrepancy between classifications of variants, these were reviewed on a case-by-case basis from the original publications and categorised according to either the most frequently designated class, or as the most severe type when several classes were commonly assigned. Of the genetically characterised variants, classifications did not exist for a small number (n = 5 variants in our database: Dagua, Gond, Laibin, Yunan) of rarely reported variants (n = 9 total occurrences) which had to be excluded. A much more significant number of reports had to be excluded due to their poorly defined biochemical characteristics which precluded them from being reliably assigned to severity classes; these included diagnoses such as “Gd-1”, “B-slow” and, for example, 21 different variants each reported only up to three times from populations in Papua New Guinea in the 1970-80s.

Following this classification of variant occurrences, the final dataset included 932 positive occurrences, 527 of “severe” Class II variants, and 405 of “mild/moderate”, from 54 MECs.

The variants reported from each country are listed in Table S3 of the Appendix to Chapter 4. The mapped distribution of this dataset, by severity class (II and III) is shown in Figure 6.2.

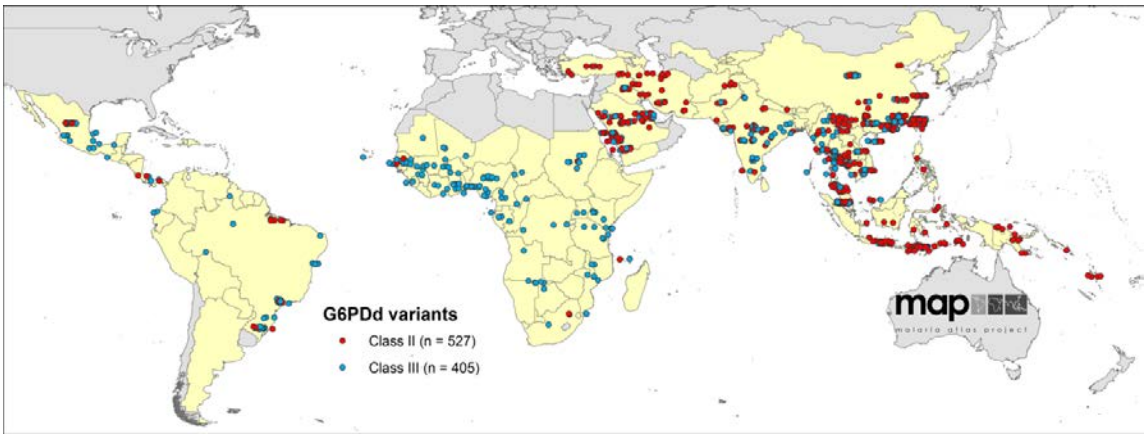


Figure 6.2. Distribution of G6PDd variant occurrences, by severity class. Class II variants (red data points) are the more severe, with <10% residual enzyme activity; Class III variants (blue data points) are the milder, with 10-60% residual enzyme activity. Data points are mapped with a jitter to show spatial duplicates (R jitter function; factor = 100), so their exact position is only approximate.

6.2.4. Generating an index of national-level risk from G6PDd

The simple risk framework brought together both the prevalence of G6PDd and the relative severity of G6PDd variants at the national level. The prevalence score was based on the population-weighted national estimate presented in Chapter 4 (Figure 3A). Countries with a median predicted national G6PDd allele frequency of ≤1% were scored 1 (rare); >1 - <10% (common) were scored 2; and national prevalence ≥10% (high) was scored 3 (see Table 6.2). Next, the variant severity score reflected the relative proportion of Class II and Class III variant occurrences. Score 1 (mildest severity) was given to countries from which only Class III variants were reported; score 2 (moderate severity) to countries where two thirds or more of the data points were Class III, but Class II variants were nevertheless reported; countries where Class II variants made up more than a third of occurrences were scored 3 (severe). If no data were available from a country, a conservative approach was followed which took the highest score from any neighbouring country. Four island nations were lacking data: Haiti and Dominican Republic were scored 2 (moderate) based on the score of the large majority of surrounding mainland countries; and Madagascar was scored 1 (mild) based on Mozambique

and all other central African countries; Mayotte was assigned score 3 (severe) in accordance with data from the nearby Comoros islands.

The index of overall risk was calculated by multiplying the prevalence and severity scores, resulting in six categories across a spectrum of risk from mild and rare G6PDd to common and severe G6PDd (Table 6.2). The resulting risk maps are shown in Figure 6.3.

G6PDd risk index		**Variant severity**		
		Class III only	**Class II uncommon**	**Class II common**
National G6PDd prevalence	**Rare**: ≤1%	Level 1 (n = 1)	Level 2 (n = 7)	Level 3 (n = 7)
	Common: >1 - 10%	Level 2 (n = 13)	Level 4 (n = 15)	Level 5 (n = 20)
	High: >10%	Level 3 (n = 20)	Level 5 (n = 5)	Level 6 (n = 11)

Table 6.2. Scoring table for determining an index of overall national-level risk from G6PDd, accounting for the severity of the commonly reported mutations and the overall prevalence of deficiency.

6.2.5. Generating an uncertainty index of national-level risk from G6PDd.

The starting premise of the risk analysis is that all predictions of primaquine-induced haemolytic events are inherently uncertain, constrained by the relatively poor knowledge of haemolysis. Additionally, I have assembled an uncertainty framework to capture the additional uncertainty introduced specifically through this analysis. I attempt here to qualitatively assess the relative uncertainty from the two data sources used in the risk analysis.

Using the same framework as for the risk index, scores were devised to account for the level of uncertainty with which the risk classifications were made. These accounted for uncertainty in the national prevalence prediction as well as uncertainty in the estimate of local variant severity. Uncertainty in the prevalence estimate was scored as the size of the IQR around the prevalence prediction relative to the median estimate. Countries where the prediction was most certain (IQR ≤50% median estimate) were given score 1; if the IQR was 50-100% the size of the median estimate, countries were scored 2; finally a score of 3 was given when the IQR was >100% the size of the median estimate. IQR values for all national-level prevalence estimates are given in Table S1 of the Appendix to Chapter 4.

To stratify uncertainty in the variant severity score, I used two factors: the number of occurrence points within each country, and the local heterogeneity in variant scores. If all neighbouring countries had the same severity scores, then uncertainty was decreased. Conversely, nearby heterogeneity increased uncertainty. The categories for numbers of data points per country were defined as: ≥3 data points (score = 1), 1-2 data points (score = 2), and 0 data points (score = 3). Heterogeneity was scored as Low (score = 1) if a country's variant severity score was the same as all neighbouring countries, Medium (score = 2) if a country's variant severity score was the same as more neighbours than not, and High (score = 3) if most neighbours had a different variant severity score. Based on these two scores, each country was allocated an overall score from 1 to 3 for variant severity uncertainty using Table 6.3.

Variant severity scoring uncertainty		**Number of data points per country**		
		3+ (score = 1)	**1 – 2** (score = 2)	**0** (score = 3)
Nearby heterogeneity	**Low** (score = 1)	Low uncertainty (n = 22)	Low uncertainty (n = 10)	Medium uncertainty (n=22)
	Medium (score = 2)	Low uncertainty (n = 7)	Medium uncertainty (n = 4)	High uncertainty (n = 10)
	High (score = 3)	Medium uncertainty (n = 5)	High uncertainty (n = 6)	High uncertainty (n = 13)

Table 6.3. Scoring table for determining the uncertainty of variant severity scores, based on numbers of data points per country, and regional heterogeneity in variant severity scores. These uncertainty classes in the variant severity scores are mapped in Figure 6.4B.

The variant severity uncertainty scores were combined with the prevalence uncertainty scores for each country in a multiplicative table (Table 6.4) to generate a final uncertainty score for each country, as mapped in Figure 6.4.

Overall uncertainty index		**Variant severity uncertainty**		
		Low uncertainty	**Medium uncertainty**	**High uncertainty**
Prevalence uncertainty (IQR/Median)	**Low**: 0 - 50%	Level 1 (n = 13)	Level 2 (n = 4)	Level 3 (n = 0)
	Medium: 50 - 100%	Level 2 (n = 19)	Level 4 (n =6)	Level 5 (n = 9)
	High: >100%	Level 3 (n = 7)	Level 5 (n = 21)	Level 6 (n = 20)

Table 6.4. Scoring table for determining the index of overall uncertainty in the national-level risk classifications. Final categories of the risk scores are shown, with total number of MECs belonging to each category. These are mapped in Figure 6.4C.

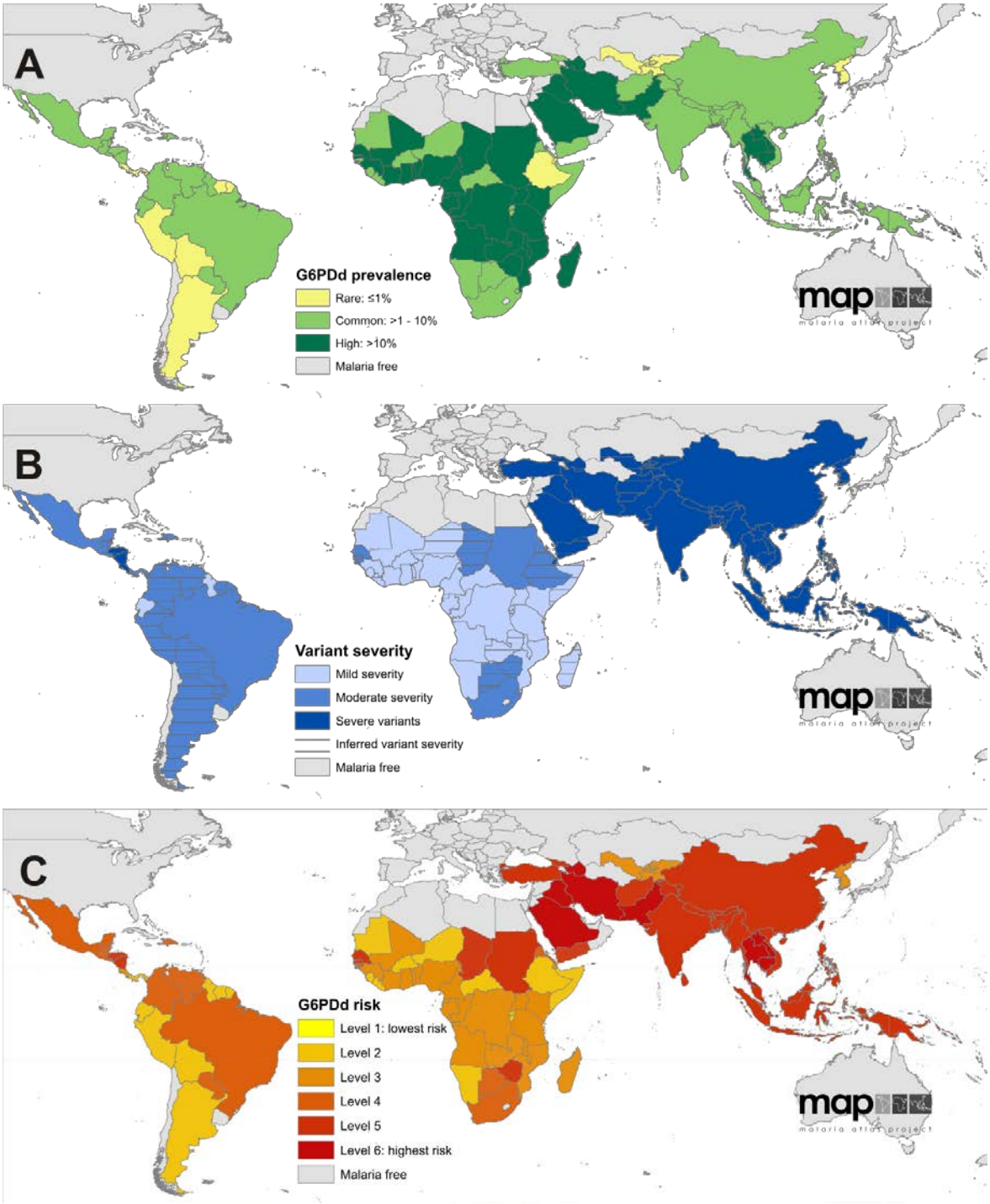


Figure 6.3. National-level prevalence scores, variant severity scores and the culminating index of overall national-level risk from G6PDd. Panel A shows the scored prevalence estimates (score = 1: if national prevalence is estimated as ≤1%; score = 2 if national prevalence is estimated as >1-10%; and score = 3 if the national prevalence is >10%); Panel B gives the three variant severity scores: lowest severity (score = 1) for countries with only Class III G6PDd variants, moderate variant severity (score = 2) for countries where a minority (≤⅓) of Class II prevailed among Class III variants; and the most severe (score = 3) for countries where Class II G6PDd variants were common (>⅓ records). Panel C shows the final six categories of overall national-level risk from G6PDd: the scores in Panels A and B were multiplied.

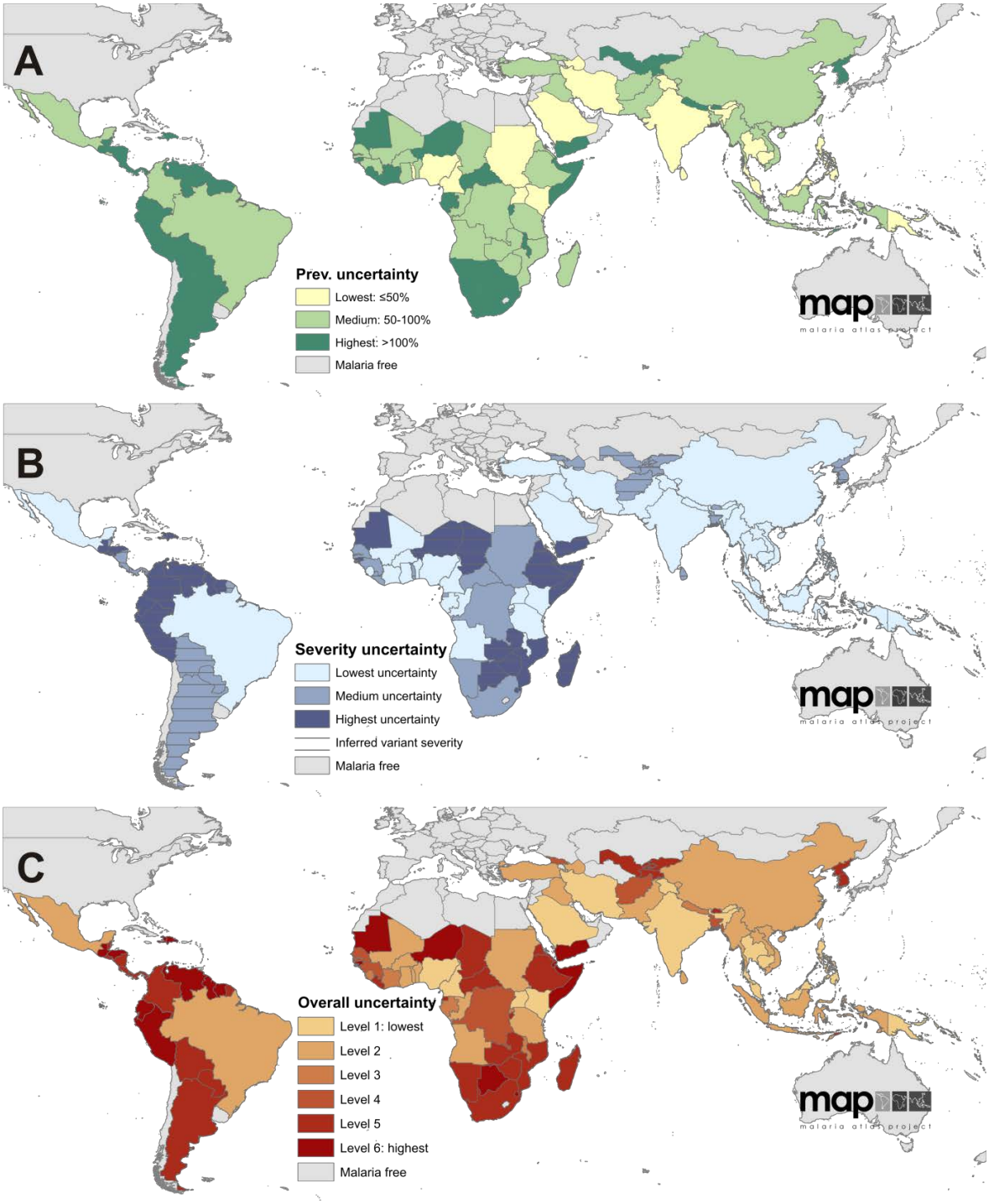


Figure 6.4. National-level scores of prevalence uncertainty, variant severity uncertainty and overall uncertainty in national-level risk from G6PD. Panel A shows the stratified prevalence uncertainty based on the proportion of IQR relative to median predictions (score = 1: if the IQR of the prevalence prediction was ≤50% of median prediction; score = 2: if the IQR was >50-100% of median prevalence prediction; score = 3 if the IQR was >100% of the median prevalence prediction for that country); Panel B gives the estimated variant severity uncertainty: scores were determined by both the number of data points in each country and the local heterogeneity in variant severity scores (fully described in Section 6.3.4); and Panel C maps the final scores from multiplying Panels A and B into an index of overall uncertainty in the national-level classifications (Table 6.3).

6.2.6. Global distribution of G6PDd-associated haemolytic risk from primaquine therapy

The simple qualitative framework proposed here ranks overall risk (from 1: lowest risk; to 6: highest risk) of using primaquine across populations with spatially variable prevalence of deficiency as well as underlying variants causing the deficiency. These maps may be interpreted as the overall public health risk of G6PDd on a population, and the relative likelihood of triggering severe adverse reactions to primaquine within populations. There is significant uncertainty in these risk predictions (Figure 6.4), which is further discussed in the following section, and important to bear in mind when considering the main findings of the analysis. The significance of these maps was also previously discussed in Chapter 4.

Risk associated with primaquine therapy was found to be high across Asia and the Asia-Pacific region, with highest risk predicted across the countries targeting malaria elimination in the Persian Gulf (Saudi Arabia, Iraq, Iran) and the Mekong region where a high diversity of severe variants are present at high frequencies (>10%) (Figure 6.3). Risk across the rest of the Asian continent and west Pacific island nations was only slightly lower (Level 5 of 6), due to the widespread prevalence of severe variants and their relatively common prevalence (>1-10% across the Asia Pacific and Indian sub-continent).

Data were scarce from many countries of the Americas, making risk assessment highly uncertain. However, moderate risk (Levels 2-4 of 6) was predicted across most of the continent, which increased in countries of central America where severe variants (usually Mediterranean) were more commonly reported, such as Costa Rica. Variants were found to be of ‘Moderate severity’ across the continent – an indication of the relative admixture of Class II and III variants.

Uncertainty was high across most sub-Saharan African countries, with variant occurrence data lacking from many areas. However, where available, the variants reported were generally Class III (A- variant), putting the variant severity rank as ‘Mild’. The high prevalence of deficiency, however (>10%), meant that Level 3 was the median rank across the continent. This rose in countries around Sudan and South Africa where Class II variants (Mediterranean) were reported.

G6PDd-associated risks are therefore prevalent across all malaria endemic regions, and very high in a third of countries (in 36 of 99 countries, G6PDd was common and caused by severe variants). The benefits of administering primaquine can only be justified where G6PDd-associated risks are low. Twenty-one countries, mainly located on the fringes of *P. vivax* transmission (Pacific coastal regions of the Americas, and Sahel countries in sub-Saharan Africa) were classified as being at the lowest two classes of risk. With the exception of Peru, these are not areas where high dose primaquine will be commonly needed as *P. vivax* endemicity in these countries is low (Gething *et al.*, 2012). The results of this analysis (Figure 6.3) indicate that in no areas can the risks of primaquine therapy be deemed 'low', and that prior screening for G6PDd must continue to be pre-requisite to treatment. The importance of semi- or fully-quantitative screening to distinguish mild from severe deficiency is vital in areas where Class II variants are reported – these areas are mapped as Moderate or Severe in Figure 6.3B, and correspond to 65 of 99 countries.

6.2.7. Important limitations to predicting national-level haemolytic risk

Uncertainty was introduced into this framework from two sources. First, from the two imperfect evidence-bases informing the 'true' prevalence and variant distributions, which included large areas with no data. Countries from which no data could be identified had to have their scores inferred from neighbouring countries. The uncertainty framework attempted to represent these gaps, summarising the relative confidence in the underlying maps of prevalence and variant severity (Figure 6.4). The more fundamental source of uncertainty, however, stemmed from the uncertainty in the interpretation of the maps. Interpreting the significance of the different variants is currently constrained by the lack of clinical data on the relative haemolytic risk associated with each of the variants and different primaquine doses. The data which would be necessary to improve our interpretation of the maps and to move towards a quantitative assessment of risk are discussed below.

A main assumption of the simple framework proposed here is that the proxy measures of primaquine sensitivity – the WHO-endorsed severity classes – are adequate indicators of the relative risk of primaquine-induced haemolysis. This was the only classification which was universally available for all variants. However, the rationale for the distinction between Class II and III variants has been deemed blurry and no longer useful (Luzzatto, 2009), given that any degree of haemolysis will always be detrimental. And there appears to be no empirical clinical evidence supporting the 10% residual enzyme activity threshold by which Class II and III are distinguished. There does appear to be a differential tolerance between variants with differing levels of deficiency, with milder G6PDd phenotypes able to tolerate lower doses of primaquine safely in a majority of cases. This is the basis of the WHO's current primaquine treatment guidelines (WHO; WHO, 2010). More uncertain, is the evidence base upon which variants are classified as Class II or III. For many variants, this has only been assessed with small sample sizes, and inconsistently normalised laboratory techniques (*Pers. Comm*., Lucio Luzzatto; 23 May 2012). Furthermore, in many countries, there were no reports of named G6PDd variants, so the severity scores of 45 countries had to be inferred from neighbouring countries. Countries classified as having high uncertainty in their assignment of variant severity would be important targets for carrying out surveys of the local variants.

6.3. Towards a quantitative haemolytic risk framework for primaquine therapy

Primaquine is a necessary drug for malaria control and elimination. Knowing how to manage it properly is the cornerstone of safe *P. vivax* therapy. Looking ahead, the only plausible alternative drug to primaquine for at least the next decade is tafenoquine, a GSK-MMV partnership drug currently in Phase IIb/III trials (Shanks *et al.*, 2001; Crockett and Kain, 2007). However, also being an 8-aminoquinoline, this drug presents similar oxidative challenges to G6PDd individuals, which overshadows its prospects for licencing:

tafenoquine has already been in development for three decades. Understanding how to safely deploy 8-aminoquinolines is a major challenge facing the *P. vivax* community.

Elimination of *P. vivax* is likely to be dependent upon primaquine, although it has been suggested that elimination may still be feasible in some areas without widespread use of primaquine (Lasse Verstergaard, WHO WPRO representative, APMEN Vivax Working Group, Incheon 2012), as is reported from Vanuatu where both *P. falciparum* and *P. vivax* endemicity are dropping without primaquine. However, the feasibility of 'finishing the job' and achieving elimination status without primaquine to clear the reservoirs of relapsing hypnozoites is not yet demonstrated, and neither is the scalability of the progress from Vanuatu to non-island regions.

The operational inadequacy and severe threats posed by primaquine to G6PDd patients leave this drug unfit for use in endemic zones in its current form. Large-scale screening of G6PDd – together with a robust understanding of how reduced enzyme activity predicts haemolytic risk – is essential to allow widespread and safe use of this drug. The current uncertainties for assessing many aspects of primaquine therapy in G6PDd individuals mean that the coarse, qualitative assessment of relative risk presented here may be the best framework that can be produced for now. Knowing how to interpret the G6PDd variant data more fully will allow this risk framework to be substantially improved and its uncertainties reduced. In the meantime, the only certain predictor of risk – a phenotypic deficiency in G6PD activity – may be the safest and most reliable indicator of primaquine suitability, regardless of G6PDd variant severity.

Fundamental questions need to be answered, such as what constitutes an acceptable degree of haemolysis? Given that all G6PDd individuals suffer some degree of haemolysis, it is necessary to establish the threshold at which this becomes clinically unacceptable and primaquine ought not to be administered. This threshold can then be used to determine where the 'mild' and 'severe' cut-offs ought to lie, and what regimens would be suitable for these different levels of deficiency. The G6PDd variant maps can then be used to discern which regimens would be best suited to different areas. In the absence of practical field-

based tests, it is also necessary to determine how low the probability of adverse haemolytic events to primaquine has to be to justify the use of primaquine without prior phenotypic and genotypic testing. Using the thresholds applied in the risk assessment presented here, only Cape Verde emerged as having the lowest category of risk: with low prevalence (<1% allele frequency) of low risk variants. The widespread prevalence of G6PDd, combined with the common occurrence of severe variants makes primaquine therapy for *P. vivax* radical cure without prior G6PDd screening highly unadvisable.

The main knowledge gap preventing a quantitative prediction of adverse events to primaquine is sensitivity data for all the common variants in relation to specific primaquine doses. The further study required to obtain this risk data could address the following questions:

1. What are the biochemical and genetic characteristics of individuals who suffer primaquine-induced haemolysis? Retrospective analyses of G6PDd individuals in whom primaquine was erroneously prescribed resulting in adverse events. For example, the follow-up investigations of the collapsed Dapsone trials have allowed valuable insight into the widespread susceptibility of haemolysis triggered by this drug in individuals with the A- G6PDd variant (Pamba *et al.*, 2012). Similarly thorough studies must be conducted on all reported cases of primaquine-induced haemolysis. A central repository of these adverse event reports would also greatly facilitate their analysis.
2. A robust database of paired genotype and phenotype data for a broad range of variants. It is widely reported that enzyme expression levels are heterogeneous even within individuals of the same genotypes (Kim *et al.*, 2011; Shah *et al.*, 2012). As such, can genotypes be reliable predictors of enzyme expression?
3. What is the relationship between residual enzyme activity and the degree of haemolysis which results from certain doses of primaquine? Given the obvious ethical difficulties in investigating this question, is it possible to study this *in vitro*? Closely monitored and well equipped environments may allow *in vivo*

investigations by administering primaquine in gradually ascending doses. One such study is being carried out by the MMV-GSK Tafenoquine trials in Thailand on G6PDd heterozygotes with the Mahidol variant (Medicines for Malaria Venture/GlaxoSmithKline). These data would indicate whether it is necessary to adapt diagnostic tests to predicting haemolysis, rather than providing a measure of residual enzyme activity as an acceptable indicator of risk.

4. What is the relative contribution of the different factors determining haemolysis in individual patients? These are currently thought to include co-occurring infection, anaemia, patient age etc, as discussed earlier in this chapter. Individual-level data from patients with a range of variants would allow insight into these questions.
5. How can the primaquine dosing regimen be altered to reduce haemolysis to acceptable levels, whilst maintaining plausible adherence rates? Could individuals with 'Mild' deficiency be safely administered a shorter regimen than the extended 8-week therapy currently recommended? An understanding of the factors affecting these relationships would provide an evidence-base for tailoring primaquine therapy.
6. What are the molecular mechanisms by which primaquine triggers haemolysis? This knowledge would allow insight into the aforementioned relationship between variants and their primaquine sensitivity.
7. Are there spatially-variable patterns of primaquine-susceptibility to improve safe access to *P. vivax* radical cure? To apply knowledge of risk at the individual level to spatial scales of public health pertinence, it is necessary to fill the gaps in the existing maps and reduce uncertainty in the evidence-bases of the existing maps. Countries for which variant severity had to be inferred (hashed out in Figure 6.2B and 6.3B) need studies to investigate the common variants. Similarly, countries in which the prevalence of deficiency could only be predicted with high uncertainty (Figure 6.3A) would benefit from additional population screening surveys. Robust data about the prevalence of G6PDd variants would allow sub-national level assessments.

As data become available to answer these questions, it will be important to refine the risk framework described here and provide the malaria public health community with more detailed analyses of the spatially variable risks surrounding *P. vivax* radical cure.

Improving safe access to primaquine requires: (i) better management of the risks associated with this drug, as discussed above, but also (ii) drug developments to reduce primaquine toxicity whilst maintaining its therapeutic efficacy. Drug management requires drug trials which determine primaquine sensitivity phenotypes of a broad range of G6PDd variants. Knowledge of drug regimens which can be tolerated by individuals with different variants could enable broader access to safe therapy. In turn, this has important implications for the types of diagnostic methods required (qualitative/semi-quantitative/quantitative/molecular). A thorough understanding of the relative risks associated with different variants would also provide a more detailed landscape within which to assess the relative risks of primaquine spatially and quantitatively, at the public health decision-making scale. In terms of drug developments, the early studies which identified the total-dose effect of primaquine, whereby the periodicity of treatment does not affect its efficacy but can make the drug safe for A- G6PDd patients, opened up a large population able to tolerate the drug. Similar studies with other variants are needed to establish whether other regimens could be developed to further increase its tolerability. Further, there is evidence that primaquine interaction with co-administered schizonticides may affect the drug's toxicity (Myint *et al.*, 2011). If confirmed, this could prove a relatively straightforward solution to safer therapy. Understanding the mechanisms of primaquine-induced haemolysis would support this goal.

The next chapter – Chapter 7 – is a general discussion of the overall thesis. In this, I discuss in more detail the practical options for increasing safe access to primaquine in the short-term, and propose studies to increase evidence-based assessments of risk in the longer term.

6.4. References

Alving, A.S., Johnson, C.F., Tarlov, A.R., et al. (1960). Mitigation of the haemolytic effect of primaquine and enhancement of its action against exoerythrocytic forms of the Chesson strain of *Plasmodium vivax* by intermittent regimens of drug administration: a preliminary report. *Bulletin of the World Health Organization* **22**: 621-631.

Baird, J.K. (2012a). Elimination therapy for the endemic malarias. *Current Infectious Disease Reports* **14**(3): 227-237.

Baird, J.K. (2012b). Primaquine toxicity forestalls effective therapeutic management of the endemic malarias. *International Journal for Parasitology* **42**(12): 1049-1054.

Baird, J.K. (2013). Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. *Clinical Microbiology Reviews* **26**(1): 1-22.

Baird, J.K. and Hoffman, S.L. (2004). Primaquine therapy for malaria. *Clinical Infectious Diseases* **39**(9): 1336-1345.

Baird, J.K., Lacy, M.D., Basri, H., et al. (2001). Randomized, parallel placebo-controlled trial of primaquine for malaria prophylaxis in Papua, Indonesia. *Clinical Infectious Diseases* **33**(12): 1990-1997.

Baird, J.K., McCormick, G.J. and Canfield, C.J. (1986). Effects of nine synthetic putative metabolites of primaquine on activity of the hexose monophosphate shunt in intact human red blood cells in vitro. *Biochemical Pharmacology* **35**(7): 1099-1106.

Baird, J.K. and Surjadjaja, C. (2011). Consideration of ethics in primaquine therapy against malaria transmission. *Trends in Parasitology* **27**(1): 11-16.

Beutler, E. (1969). Drug-induced hemolytic anemia. *Pharmacological Reviews* **21**(1): 73-103.

Beutler, E. (1991). Glucose-6-phosphate dehydrogenase deficiency. *New England Journal of Medicine* **324**(3): 169-174.

Beutler, E. (1993). Study of glucose-6-phosphate dehydrogenase: history and molecular biology. *American Journal of Hematology* **42**(1): 53-58.

Beutler, E. (1994). G6PD deficiency. *Blood* **84**(11): 3613-3636.

Beutler, E. and Duparc, S. (2007). Glucose-6-phosphate dehydrogenase deficiency and antimalarial drug development. *American Journal of Tropical Medicine and Hygiene* **77**(4): 779-789.

Bousema, T. and Drakeley, C. (2011). Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. *Clinical Microbiology Reviews* **24**(2): 377-410.

Bowman, Z.S., Jollow, D.J. and McMillan, D.C. (2005a). Primaquine-induced hemolytic anemia: role of splenic macrophages in the fate of 5-hydroxyprimaquine-treated rat erythrocytes. *Journal of Pharmacology and Experimental Therapeutics* **315**(3): 980-986.

Bowman, Z.S., Morrow, J.D., Jollow, D.J., et al. (2005b). Primaquine-induced hemolytic anemia: role of membrane lipid peroxidation and cytoskeletal protein alterations in the hemotoxicity of 5-hydroxyprimaquine. *Journal of Pharmacology and Experimental Therapeutics* **314**(2): 838-845.

Brueckner, R.P., Ohrt, C., Baird, J.K., et al. (2001). 8-Aminoquinolines. In: Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery. P.J. Rosenthal (eds). Totowa, NJ, Humana Press.

Burgoine, K.L., Bancone, G. and Nosten, F. (2010). The reality of using primaquine. *Malaria Journal* **9**: 376.

Cappellini, M.D. and Fiorelli, G. (2008). Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**(9606): 64-74.

Carson, P.E., Hohl, R., Nora, M.V., et al. (1981). Toxicology of the 8-aminoquinolines and genetic factors associated with their toxicity in man. *Bulletin of the World Health Organization* **59**(3): 427-437.

Crockett, M. and Kain, K.C. (2007). Tafenoquine: a promising new antimalarial agent. *Expert Opinion on Investigational Drugs* **16**(5): 705-715.

Dern, R.J., Beutler, E. and Alving, A.S. (1954). The hemolytic effect of primaquine. II. The natural course of the hemolytic anemia and the mechanism of its self-limited character. *Journal of Laboratory and Clinical Medicine* **44**(2): 171-176.

Dr Andrew C. R. Martin's Group. Andrew C. R. Martin's Bioinformatics Group at UCL. Accessed: 20 April 2012. URL: http://www.bioinf.org.uk/g6pd/db/.

Edgcomb, J.H., Arnold, J., Yount, E.H., Jr., et al. (1950). Primaquine, SN 13272, a new curative agent in vivax malaria; a preliminary report. *Journal of the National Malaria Society* **9**(4): 285-292.

Eziefula, A.C., Gosling, R., Hwang, J., et al. (2012). Rationale for short course primaquine in Africa to interrupt malaria transmission. *Malaria Journal* **11**: 360.

Fletcher, K.A., Barton, P.F. and Kelly, J.A. (1988). Studies on the mechanisms of oxidation in the erythrocyte by metabolites of primaquine. *Biochemical Pharmacology* **37**(13): 2683-2690.

Ganesan, S., Chaurasiya, N.D., Sahu, R., et al. (2012). Understanding the mechanisms for metabolism-linked hemolytic toxicity of primaquine against glucose 6-phosphate dehydrogenase deficient human erythrocytes: evaluation of eryptotic pathway. *Toxicology* **294**(1): 54-60.

Gething, P.W., Elyazar, I.R., Moyes, C.L., et al. (2012). A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Neglected Tropical Diseases* **6**(9): e1814.

Graves, P.M., Gelband, H. and Garner, P. (2012). Primaquine for reducing *Plasmodium falciparum* transmission. *Cochrane Database of Systematic Reviews* **9**: CD008152.

Greaves, J., Evans, D.A., Gilles, H.M., et al. (1980). Plasma kinetics and urinary excretion of primaquine in man. *British Journal of Clinical Pharmacology* **10**(4): 399-404.

Greene, L.S. (1993). G6PD deficiency as protection against *falciparum*-malaria: an epidemiologic critique of population and experimental studies. *Yearbook of Physical Anthropology* **36**: 153-178.

Hill, D.R., Baird, J.K., Parise, M.E., et al. (2006). Primaquine: report from CDC expert meeting on malaria chemoprophylaxis I. *American Journal of Tropical Medicine and Hygiene* **75**(3): 402-415.

Howes, R.E., Battle, K.E., Satyagraha, A.W., et al. (2013). G6PD deficiency: Global distribution, genetic variants and primaquine therapy. *Advances in Parasitology* **81**: In press.

Kim, S., Nguon, C., Guillard, B., et al. (2011). Performance of the CareStart G6PD deficiency screening test, a point-of-care diagnostic for primaquine therapy screening. *PLoS One* **6**(12): e28357.

Koliwad, S.K., Elliott, S.J. and Kunze, D.L. (1996). Oxidized glutathione mediates cation channel activation in calf vascular endothelial cells during oxidant stress. *The Journal of Physiology* **495 (Pt 1)**: 37-49.

Kwok, C.J., Martin, A.C., Au, S.W., et al. (2002). G6PDdb, an integrated database of glucose-6-phosphate dehydrogenase (G6PD) mutations. *Human Mutation* **19**(3): 217-224.

Lang, F., Lang, K.S., Lang, P.A., et al. (2006). Mechanisms and significance of eryptosis. *Antioxidants & Redox Signaling* **8**(7-8): 1183-1192.

Lang, P.A., Kaiser, S., Myssina, S., et al. (2003). Role of Ca^{2+}-activated K^{+} channels in human erythrocyte apoptosis. *American Journal of Physiology. Cell Physiology* **285**(6): C1553-1560.

Link, C.M., Theoharides, A.D., Anders, J.C., et al. (1985). Structure-activity relationships of putative primaquine metabolites causing methemoglobin formation in canine hemolysates. *Toxicology and Applied Pharmacology* **81**(2): 192-202.

Louicharoen, C., Patin, E., Paul, R., et al. (2009). Positively selected G6PD-Mahidol mutation reduces *Plasmodium vivax* density in Southeast Asians. *Science* **326**(5959): 1546-1549.

Luzzatto, L. (2006). Glucose 6-phosphate dehydrogenase deficiency: from genotype to phenotype. *Haematologica* **91**(10): 1303-1306.

Luzzatto, L. (2009). Glucose-6-phosphate dehydrogenase deficiency. In: Nathan and Oski's Hematology of Infancy and Childhood. S.H. Orkin, D.G. Nathan, D. Ginsburg, et al. (eds). Philadelphia, Saunders.

Luzzatto, L., Mehta, A. and Vulliamy, T.J. (2001). Glucose-6-phosphate dehydrogenase deficiency. In: The metabolic and molecular bases of inherited disease, 8th ed. C.R. Scriver, A.L. Beaudet, W.S. Sly, et al. (eds). New York, McGraw-Hill Inc. iii: 4517-4553.

Malaria Policy Advisory Committee Meeting (2012). WHO Evidence Review Group Report: the safety and effectiveness of single dose primaquine as a *P. falciparum* gametocytocide. Geneva. WHO.

Mason, P.J., Bautista, J.M. and Gilsanz, F. (2007). G6PD deficiency: the genotype-phenotype association. *Blood Reviews* **21**(5): 267-283.

Medicines for Malaria Venture/GlaxoSmithKline. MMV/GSK Tafenoquine Phase IIb/III trials. Accessed: 26 Nov 2012. URL: http://www.mmv.org/research-development/project-portfolio/tafenoquine.

Mihaly, G.W., Ward, S.A., Edwards, G., et al. (1985). Pharmacokinetics of primaquine in man. I. Studies of the absolute bioavailability and effects of dose size. *British Journal of Clinical Pharmacology* **19**(6): 745-750.

Minucci, A., Moradkhani, K., Hwang, M.J., et al. (2012). Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. *Blood Cells Molecules and Diseases* **48**(3): 154-165.

Myint, H.Y., Berman, J., Walker, L., et al. (2011). Review: Improving the therapeutic index of 8-aminoquinolines by the use of drug combinations: review of the literature and proposal for future investigations. *American Journal of Tropical Medicine and Hygiene* **85**(6): 1010-1014.

Pamba, A., Richardson, N.D., Carter, N., et al. (2012). Clinical spectrum and severity of hemolytic anemia in glucose 6-phosphate dehydrogenase-deficient children receiving dapsone. *Blood* **120**(20): 4123-4133.

Pandolfi, P.P., Sonati, F., Rivi, R., et al. (1995). Targeted disruption of the housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD): G6PD is dispensable for pentose synthesis but essential for defense against oxidative stress. *EMBO Journal* **14**(21): 5209-5215.

Percy, M.J., McFerran, N.V. and Lappin, T.R. (2005). Disorders of oxidised haemoglobin. *Blood Reviews* **19**(2): 61-68.

Piomelli, S., Corash, L.M., Davenport, D.D., et al. (1968). In vivo lability of glucose-6-phosphate dehydrogenase in Gd^{A-} and $Gd^{Mediterranean}$ deficiency. *Journal of Clinical Investigation* **47**(4): 940-948.

Pybus, B.S., Sousa, J.C., Jin, X., et al. (2012). CYP450 phenotyping and accurate mass identification of metabolites of the 8-aminoquinoline, anti-malarial drug primaquine. *Malaria Journal* **11**: 259.

Recht, J., Ashley, E.A. and White, N.J. (unpublished). 8-aminoquinolines safety review for WHO primaquine ERG.

Reeve, P.A., Toaliu, H., Kaneko, A., et al. (1992). Acute intravascular haemolysis in Vanuatu following a single dose of primaquine in individuals with glucose-6-phosphate dehydrogenase deficiency. *Journal of Tropical Medicine and Hygiene* **95**(5): 349-351.

Shah, S.S., Diakite, S.A., Traore, K., et al. (2012). A novel cytofluorometric assay for the detection and quantification of glucose-6-phosphate dehydrogenase deficiency. *Scientific Reports* **2**: 299.

Shanks, G.D., Oloo, A.J., Aleman, G.M., et al. (2001). A new primaquine analogue, tafenoquine (WR 238605), for prophylaxis against *Plasmodium falciparum* malaria. *Clinical Infectious Diseases* **33**(12): 1968-1974.

Shekalaghe, S.A., ter Braak, R., Daou, M., et al. (2010). In Tanzania, hemolysis after a single dose of primaquine coadministered with an artemisinin is not restricted to glucose-6-phosphate dehydrogenase-deficient (G6PD A-) individuals. *Antimicrobial Agents and Chemotherapy* **54**(5): 1762-1768.

Vulliamy, T., Luzzatto, L., Hirono, A., et al. (1997). Hematologically Important Mutations: Glucose-6-Phosphate Dehydrogenase. *Blood Cells, Molecules, and Diseases* **23**(2): 302-313.

White, N.J. (2012). Primaquine to prevent transmission of falciparum malaria. *Lancet Infectious Diseases*: doi:10.1016/S1473-3099(1012)70198-70196.

WHO. Country antimalarial drug policies: by region. Accessed: 1 May 2012. URL: http://www.who.int/malaria/am_drug_policies_by_region_afro/en/index.html.

WHO (2010). Guidelines for the treatment of malaria, second edition. Geneva: World Health Organization.

WHO (2011). Global plan for artemisinin resistance containment (GPARC).

WHO Working Group (1989). Glucose-6-phosphate dehydrogenase deficiency. *Bulletin of the World Health Organization* **67**(6): 601-611.

Yoshida, A., Beutler, E. and Motulsky, A.G. (1971). Human glucose-6-phosphate dehydrogenase variants. *Bulletin of the World Health Organization* **45**(2): 243-253.

Ziai, M., Amirhakimi, G.H., Reinhold, J.G., et al. (1967). Malaria prophylaxis and treatment in G-6-PD deficiency. An observation on the toxicity of primaquine and chloroquine. *Clinical Pediatrics* **6**(4): 242-243.

Chapter 7 – Discussion

This thesis' primary objectives were to map the spatial prevalence of the Duffy blood group variants and of G6PD deficiency (G6PDd) in order to support malaria control programmes globally. I have given examples of how these suites of maps are being applied to current public health problems, and through this have attempted to demonstrate the positive contribution that integrating spatial epidemiological human genetic data can make in improving the evidence-base for strategic planning for control of an infectious disease, thereby attempting to bridge the gap between basic biological research and the health sciences (Weatherall, 2010).

This Discussion opens by summarising the salient results from each chapter to pave the way through a general discussion of the strengths, limitations, and main conclusions of this work. The research objectives of the thesis presented different mapping challenges. These methodological challenges are reiterated, and I discuss both the value and the limitations of the geostatistical models developed. Next, I consider the status of these two bodies of work as milestones in the scientific progression towards ever-optimised malaria control and I propose future research priorities which emerge off the back of these results. In particular, I consider further studies to help understand the status of *P. vivax* transmission, particularly in Africa. In relation to G6PDd, I suggest priorities for increasing safe access to primaquine, and argue that investment in diagnostic methods must be at the forefront of short-term research and development (R&D) targets.

7.1. Chapter summary

This thesis' first two research chapters concern the Duffy blood group antigens – the only currently known red blood cell (RBC) receptors enabling *P. vivax* infection. Chapter 2 described the allele frequency maps of the Duffy antigen variants (*FY*A*, *FY*B*, *FY*B*^{ES})

and of the Duffy negative phenotype (Howes *et al.*, 2011). These maps showed highly conspicuous spatial patterns, with the Duffy negative phenotype predominant among populations of sub-Saharan Africa (>95%) and common to certain coastal communities of the Americas. *FY*B* was most common among European populations and in the Americas (ca. 50% prevalence). Away from the European epicentre of the *FY*B* allele, frequencies of the *FY*A* allele increased with distance, reaching frequencies greater than 90% across much of East and Southeast Asia. The significance of these results was considered in Chapter 3, in which the contribution of the Duffy negativity map to a range of analyses examining the global public health significance of *P. vivax* was discussed. In attempting to map the distribution of the population at risk of *P. vivax* (*Pv*PAR) infection, the Duffy negative map was used to represent the population resistant to infection (Guerra *et al.*, 2010). The high prevalence of Duffy negativity across sub-Saharan Africa meant that despite the environmental and biological environment being suitable for transmission, only 86.4 million (or 3.5%) of the 2.5 billion individuals at risk of infection globally were from the African region (Africa+). Instead, the global focus of *P. vivax* risk was demonstrated to be across the central and south-eastern regions of Asia. The Duffy negativity map was also incorporated into a model-based geostatistical framework to predict transmission intensity of *P. vivax* globally, allowing the mapping model to borrow predictive strength from the Duffy map in areas where parasite prevalence surveys were lacking (Gething *et al.*, 2012). Evidence for the differential binding preference of *P. vivax* to Fy^a and Fy^b antigens emphasised the value of mapping the individual Duffy positive alleles as well as the Duffy negative phenotype (King *et al.*, 2011).

Chapter 4 turned to consider the distribution of G6PDd. Individuals with this genetic predisposition to primaquine-induced haemolysis were found to be widespread across all malaria endemic regions (Howes *et al.*, 2012). Frequencies were highest across the tropical regions of sub-Saharan Africa and the Arabian Peninsula, where prevalence rose above 30% in certain communities. Across all 99 malaria endemic countries, a median allele frequency of 8.0% was estimated (IQR: 7.4-8.8), and a frequency of 5.3% (4.4-6.7) was predicted across those countries targeting malaria elimination. These estimates

corresponded to an estimated 350 million affected individuals. Population-weighted median national-level frequencies of G6PDd were also estimated and mapped.

In mapping its prevalence, G6PDd was considered to represent any persons with reduced G6PD enzyme activity levels. Though informative and practical from a public health perspective, this binary classification masks a range of enzyme deficiency levels. This spectrum of enzyme activity is reflected in its associated clinical severity which ranges from mild to highly severe. Chapter 5 attempted to represent this diversity and mapped the prevalence of fifteen common polymorphic variants. These maps revealed striking geographic patterns, with apparent genetic homogeneity among the G6PDd populations of the Americas and sub-Saharan Africa diversifying into a very heterogeneous pool of variants reported from the East and Southeast Asian regions. This regional heterogeneity in Asia was also evident at the population level, where as many as ten genetic variants were reported from a single population survey in Malaysia. The public health risks associated with primaquine stem both from the overall prevalence of G6PDd individuals in a population and from the severity of the variants common to that population. The relative influence of these two spatially variable factors was formalised in a risk framework in Chapter 6. Across Asia, the prevalence of G6PDd was common (>1% allele frequency) and the pool of variants severe. These countries were therefore ranked as being at highest risk from G6PDd. Variants from sub-Saharan Africa were generally less severe (predominantly the A- variant reported), but the high prevalence of G6PDd (>10%) meant that this region was at moderate risk from G6PDd. Across the Americas, an admixture of severe and moderate G6PDd variants was coupled with variable prevalence estimates of G6PDd which ranged from rare (≤1%) to common (1-10%). These factors ranked overall risk as being heterogeneous across this continent, ranging from relatively low to high. The limitations of this framework and the inadequacies of the underlying evidence-base leave much room for development in ranking and predicting risk between areas. These limitations and suggestions for further study were discussed in Chapter 6. From this suite of G6PDd-associated maps and analyses, it is demonstrated that G6PDd is of major public health concern, and that its widespread distribution means that primaquine cannot be safely

dispensed for *P. vivax* radical cure without prior screening for this risk factor of drug-induced haemolysis.

7.2. Methodological discussion

The core methodological requirement of both the Duffy and the G6PDd studies was to generate a model which could use a disparate evidence-base to predict spatially continuous maps of a series of required outputs. Additionally, population estimates for each output were also needed. I discuss here the methods developed for each of these applications, their strengths and limitations. The specifics of each modelling assignment are summarised in Table 7.1. As these have already been discussed at length in the relevant chapters and appendixes, I reiterate here a brief and consolidated overview.

7.2.1. Model strengths

Broadly the same framework was used to model the distribution of both the Duffy variants and G6PDd. This model-based geostatistical framework was developed from that previously applied to mapping the frequencies of sickle-cell haemoglobin (HbS) (Piel *et al.*, 2010; Piel *et al.*, 2012). Two of the main strengths of this model are its probabilistic approach and its flexibility in terms of types of input data and model outputs, as well as the option of incorporating known biological relationships within a primarily empirical overarching model architecture.

The major novel advance in the methodology developed here compared to previous attempts to map the Duffy variants and G6PDd was the model's Bayesian framework for generating probabilistic outputs (Patil *et al.*, 2011). Within this framework, over a million iterations were generated using a Markov chain Monte Carlo (MCMC) algorithm to infer a fitted joint distribution of the model parameters from the input data. From this set, over a thousand iterations were selected at random from which to generate a posterior predictive distribution (PPD) of the target outputs (Duffy or G6PDd allele frequencies) for each

location in the map. In both these studies, initial validation exercises demonstrated that the median of the PPD was the most representative summary statistic, and was thus adopted as a point estimate for each pixel in order to generate the modelled maps. The spread of the PPD is an indication of the relative confidence with which the model predictions were made, and the interquartile range (IQR) of the PPD was used as an indicator of model uncertainty. These uncertainty metrics are important reminders that the mapping predictions are not perfectly precise estimates, but rather the most probable scenario given the available evidence-base. The IQR maps indicate how this uncertainty varies spatially, and thus where the predictions should be interpreted with more or less caution. High uncertainty represents a clear call for new surveys and a strengthening of the evidence-base in that particular region.

As well as the model's framework for quantifying uncertainty, the model was highly flexible in terms of its input and output data structure, allowing bespoke adaptations to be made to each biological system being mapped (Table 7.1). In the case of the Duffy model, five different input data types were used, each contributing distinct but complementary information, and none of the inputs related directly to the required outputs (allele frequencies). For the G6PDd model, sex-specific input information was required and the primary output, the G6PDd allele frequency, could be directly inferred from the input 'Male' data type. This model, however, had to account for the gene's X-linked inheritance and the difficulties of predicting numbers of deficient females. While the number of genetic heterozygotes is relatively straightforward to predict, knowing what proportion of them would be phenotypically diagnosed as clinically deficient is not straightforward. There is no simple genotype-phenotype relationship for heterozygote expression. The model used here was the first spatially variable and evidence-based attempt to address this challenge. The framework allowed the proportion of phenotypically deficient females to vary spatially according to the evidence-base of surveys provided.

The major methodological advance between the two mapping projects was in estimating the population affected by each phenotype. In the earlier Duffy modelling project, I simply

	Duffy blood group	**G6PD deficiency**
Data inputs	Representative community surveys Sample size Survey location (Latitude, Longitude) and spatial extent (point *vs*. polygon)	
	Five data types according to diagnostic method (serological *vs*. molecular): *Phenotype, Phenotype-a, Phenotype-b, Promoter, Genotype*	Data according to sex: *N males tested & deficient* *N females tested & deficient*
Total datapoints	*n* = 821	*n* = 1,734
Map outputs		
Mapping methodology	Bayesian model-based geostatistical mapping framework	
Output	Allele frequency maps & uncertainty: *FY*A, FY*B, FY*B^{ES}* & Duffy negativity phenotype frequency	Map of G6PD deficiency allele frequency & uncertainty
Population estimates		
Population surface	Global-Rural Mapping Project (GRUMP) projected to year 2010	
Methodology	Calculation in GIS framework ⇨ No uncertainty estimate	Areal mean predictions model ⇨ Quantified uncertainty
Output	National population numbers by phenotype: *Fy(a+b+), Fy(a+b-)* *Fy(a-b+), Fy(a-b-)*	National median and uncertainty estimates: - Allele frequency (males) - Homozygous females - All deficient females

Table 7.1. Summary of key methodological distinctions between the challenges of mapping the Duffy blood group frequencies and the prevalence of G6PDd. Full explanations of each of the terms used here are given in the original chapters (Duffy in Chapter 2; G6PDd in Chapter 4) and in the associated Appendixes and publications.

multiplied the population density map surface (GRUMP) by the mean prevalence map of each phenotype and aggregated the number of individuals with each phenotype to the national level. For the G6PDd population estimates, however, the nature of the maps as a summary surface of a full PPD was taken into account, and areal estimates of each phenotype were generated within a Bayesian framework. The areal estimate model sampled repeatedly the full PPD, taking into account the spatial covariance between each pixel,

rather than simply the median prediction summary map. This population estimate model's output was another PPD which in turn allowed uncertainty metrics to be quantified from it. The Duffy phenotype population estimates are described in the paper by King and colleagues (2011), summarised in Chapter 3 and provided in the Appendix. The equivalent methods for G6PDd population estimates are described both in Chapter 4 and the associated Appendix files.

7.2.2. Model limitations

Conceptual as well as practical limitations hinder the mapping models used in this thesis. An important difficulty encountered with mapping human polymorphisms, as opposed to insect vector or parasite distributions, is the absence of suitable covariates. Ideally, maps of genetic relatedness (possibly using ethnicity as a proxy of genetic relatedness) and of consanguinity would have been available to inform the model's predictions. No such comprehensive and reliable global resources for these, however, could be identified. The model predictions were therefore wholly dependent on the input database of surveys, and areas scarcely populated with data were predicted with greatest uncertainty. A practical upshot of knowing this uncertainty is that it represents a clear indication of where new surveys would be most informative to improving our current understanding of the spatial epidemiology of these disorders. As new data become available, it may be feasible to substantively improve the resolution and reliability of the first generation maps presented here. The falling costs and technical difficulties of high-throughput gene sequencing have allowed extensive datasets to be generated by a number of large genomic consortium studies which may prove valuable sources of such data (1000 Genomes - A Deep Catalog of Human Genetic Variation; International HapMap Project; MalariaGen Genomic Epidemiology Network).

The models had to assume that populations were in Hardy-Weinberg equilibrium to allow derivation of allele frequencies from the observed input data, and phenotypes from the allele frequencies. Hardy-Weinberg equilibrium assumes that gene inheritance between

generations is fully random, and is not affected by any of the inevitable biases and deviations caused by small population sizes and genetic drift, selection, non-random mating, mutation, migration rates and hybridisation (Hardy, 1908; Weinberg, 1908; Crawford, 2007). In the absence of any global database of these deviances, it was not possible to refine the model to predict the relative influence of these factors and I therefore had to assume allele inheritance was in Hardy-Weinberg equilibrium.

From a practical perspective, these models were highly computationally and financially demanding. For example, the final G6PDd mapping and population estimates took over two months to compute despite using high performance cloud computing processors from the Amazon Elastic Compute Cloud (EC2, http://aws.amazon.com/ec2/). More importantly, the modelling code was written in Python coding language and was built upon complex existing libraries, limiting its core flexibility for non-specialist users. A more accessible model package may increase its application to a wider range of mapping scenarios. The strengths described of the flexible Bayesian model-based geostatistical framework, however, will hopefully inspire the continued development of new generations of these models.

7.3. The spatial epidemiology of the Duffy blood group

Knowledge of the spatial extent of *P. vivax* transmission across Africa has important practical repercussions for surveillance systems, the choice of diagnostic methodology and treatment guidelines, for instance, as well as obvious implications for the clinical burden of malaria across a continent where endemicity of *P. falciparum* is dropping (Gething *et al.*, 2010; WHO, 2011). The evidence of *P. vivax* being harder to target with conventional control interventions, as discussed in Chapter 1, may be a forewarning of a hidden threat which could emerge in the wake of *P. falciparum* control successes. Given the accumulating evidence from Africa of *P. vivax* infections in returning travellers, infected mosquitoes, widespread *P. vivax* seropositivity, and of apparent Duffy-independent *P. vivax*

infections (references all given in Chapter 3), there is a pressing need to pursue our understanding of *P. vivax* epidemiology across this continent. A further complication to the potential problem of *P. vivax* in Africa is the high prevalence of G6PDd mapped in Chapter 4, which prevents widespread use of the only radical cure for *P. vivax* infections.

The Duffy maps presented here are an important component of the evidence-base for investigating the public health significance of *P. vivax*, as these represent the key determinant of blood-stage infection success. Insights into the significance of the Duffy negative phenotype, both at the public health level and at the individual level, will support our understanding of *P. vivax* epidemiology across Africa. First, at the public health population level, a series of surveys are required to assemble a core epidemiological knowledge-base of *P. vivax* prevalence across Africa. The surveys should include both investigations of blood-stage infection (as done by Culleton and colleagues (2008)) and serological screening (as conducted by Culleton and colleagues (2009)) to assess different populations' overall exposure to the parasite. Complementary entomological surveys (similar to those reported by Ryan *et al*. in Kenya (2006) would further corroborate the body of evidence of the parasite's presence on this continent. To assess how the findings of these studies relate to their associated human Duffy landscapes, these surveys should be targeted across populations of differing Duffy negativity prevalence and across a spectrum of epidemiological settings. Target areas for these surveys should therefore include areas of highest Duffy negativity prevalence, such as anywhere across West Africa, as well as areas bordering the high Duffy negativity regions where Duffy phenotypes are more heterogeneous, such as Angola (where Duffy-independent transmission was reported, as described in Chapter 3). Although a large number of *Pv*PR surveys have been conducted across Zambia (all were *Pv*-negative, see Chapter 3), no surveys have been documented from Mozambique. *Plasmodium vivax* across this country would be valuable to document, both because of a gradient of Duffy negativity prevalence predicted across the country, but also its proximity to Madagascar where *P. vivax* endemicity is at stable transmission levels (Gething *et al.*, 2012) and Duffy-independent transmission has been confirmed (Menard *et al.*, 2010).

The Duffy negativity map should also be used to map the density of Duffy positive hosts and relate these to theoretical transmission models of *P. vivax*. Given the parasite's relapsing behaviour, it is likely that low population densities of competent hosts could sustain *P. vivax* parasites in populations of predominantly Duffy negative individuals, irrespective of Duffy-independent transmission. This population of Duffy positive hosts not only suffer the burden of relapsing clinical cases themselves, but also represent a latent and persistent source of infectious parasites sustaining infectious mosquitoes.

Second, at the level of the individual, it is necessary to pursue in-depth studies of the phenomenon of apparent Duffy-independent transmission. Where possible, simple serological characterisation of the Duffy phenotype of confirmed *P. vivax* positive cases would further populate the map of occurrences of Duffy-independent transmission in Chapter 3 (Figure 3.4), and allow the significance of these occurrences, and their association with the local Duffy phenotypes to be better understood. Instances of Duffy-independent transmission will only be recorded if these are specifically searched for. Unravelling the mechanisms enabling Duffy-independent transmission and identifying an alternative receptor is a vital component to understanding the potential for *P. vivax* infection in areas of high Duffy negativity. The on-going difficulties of culturing *P. vivax in vivo* (Mueller *et al.*, 2009), however, do not facilitate these studies. The gene sequencing-based approach briefly described in Chapter 3 may provide alternative evidence to support these investigations.

Evidence from both these streams of research – the prevalence of *P. vivax* across populations, and the extent and mechanism of Duffy-independent transmission – would provide important information for refining current estimates of the population at risk of *P. vivax* infection and the targeting of appropriate control measures. As discussed in Chapter 3, the assumption of complete immunity in Duffy negative hosts may be a substantial over-estimate which needs reassessing. Knowing how to calibrate the degree of protection which Duffy negativity prevalence confers to populations would allow an evidence-based re-

assessment of the *Pv*PAR to replace the series of projected scenarios presented in Chapter 3.

Although this discussion focuses on the relationship between *P. vivax* and the Duffy receptor across Africa, the emerging evidence of preferential binding between the Fy^a and Fy^b antigen variants (discussed in Chapter 3) provides interesting insights to support ongoing vaccine targeting, as well as opening up a range of questions concerning the evolutionary origins of the polymorphic variants of the Duffy gene. Establishing the drivers of the strong gradients of *FY*A* and *FY*B* across Asia, and the stark contrasts between the alleles of African populations and elsewhere, present fascinating questions which merit multi-evidenced investigation, bringing together molecular, phylogenetic, clinical and spatial epidemiological data. Attempting to reconstruct the evolutionary history of *P. vivax* and the Duffy antigen variants to assess causative relationships requires knowledge of the age and origin of *P. vivax* to estimate the time period over which this parasite has been conferring a fitness pressure on populations in different areas (Culleton and Carter, 2012). Clinical data, such as the evidence of reduced *P. vivax* parasitaemias observed in Duffy negative heterozygotes (Kasehagen *et al.*, 2007), would allow estimates of the time period over which the allele frequencies mapped here may have taken to emerge, if *P. vivax* is the selective agent. Insights into these processes in relation to *P. vivax* would provide not only an understanding of evolutionary history, but potentially also an indication of the clinical significance of the *P. vivax* parasite.

The Duffy maps are a means to an end, not a direct informant of public health policy in their own right. However, these maps are an important component to understanding the range of unanswered questions discussed here, and the potentially far-reaching public health implications of the emerging evidence make this field of study an important candidate for future research.

7.4. The spatial epidemiology of G6PDd

The maps generated in this thesis indicate that the risks associated with G6PDd are too widespread and potentially severe to justify administering primaquine for *P. vivax* radical cure without prior screening for this haemolytic risk factor. The practical implication of this widespread threat is that primaquine is largely under-used across areas where *P. vivax* endemicity is high (Baird and Surjadjaja, 2011; Baird, 2012; Baird, 2013). Given this, I discuss here practical options for overcoming the risks presented by G6PDd – a necessary hurdle to overcome if *P. vivax* control is to be successfully brought about.

If G6PDd screening is pre-requisite to primaquine therapy, development of suitable diagnostic methods must be an immediate R&D priority. Although a number of methods exist and a rapid diagnostic test (RDT) similar to RDTs for malaria parasites has been recently evaluated (Kim *et al.*, 2011), no methods are currently suitable for field-based point-of-care applications. Currently available diagnostics fall into three distinct categories: (i) binary phenotype-based tests distinguish deficient from normal cases based on a subjective threshold; (ii) quantitative assays which measure the residual enzyme activity; (iii) molecular or very specific biochemical studies which identify the presence or absence of specific mutations. Although each of these is suited to specific requirements and clinical settings, the logistical and financial constraints of field-based environments mean that only the first category of diagnostics would be practical. A methodologically simple test with a binary outcome indicating whether primaquine can or cannot be safely given is necessary. A target product profile (TPP) for this diagnostic was recently discussed at the Asia Pacific Malaria Elimination Network (APMEN) Vivax Working Group (Incheon, May 2012 (APMEN)) and included low cost, stability across a range of temperatures and humidity levels, rapid diagnostic result, methodological simplicity and requiring only minimal training. The urgency for such a test is well illustrated by the difficulties inherent to current methods (see Box 7.1).

Box 7.1. ***A case study: the WST-8/PMS diagnostic in Pulau Bacan, Maluku Province, Indonesia***

In July 2010, I was involved in a G6PDd screening survey led by members of Dr Din Syafruddin's Malaria Research Team from the Eijkman Institute in Jakarta, Indonesia. This study was conducted on Bacan Island, in the eastern province of the Maluku (Figure 7.1). The high *P. vivax* (Elyazar *et al.*, 2012) and *P. falciparum* (Elyazar *et al.*, 2011) endemicity across this region put it in the Indonesian Ministry of Health's final phase for malaria elimination. The field setting was extremely remote and although the island's central health centre was equipped with microscopes and basic laboratory facilities, the rural clinics were not. This environment is likely to be characteristic of many of the remaining transmission foci across this region.

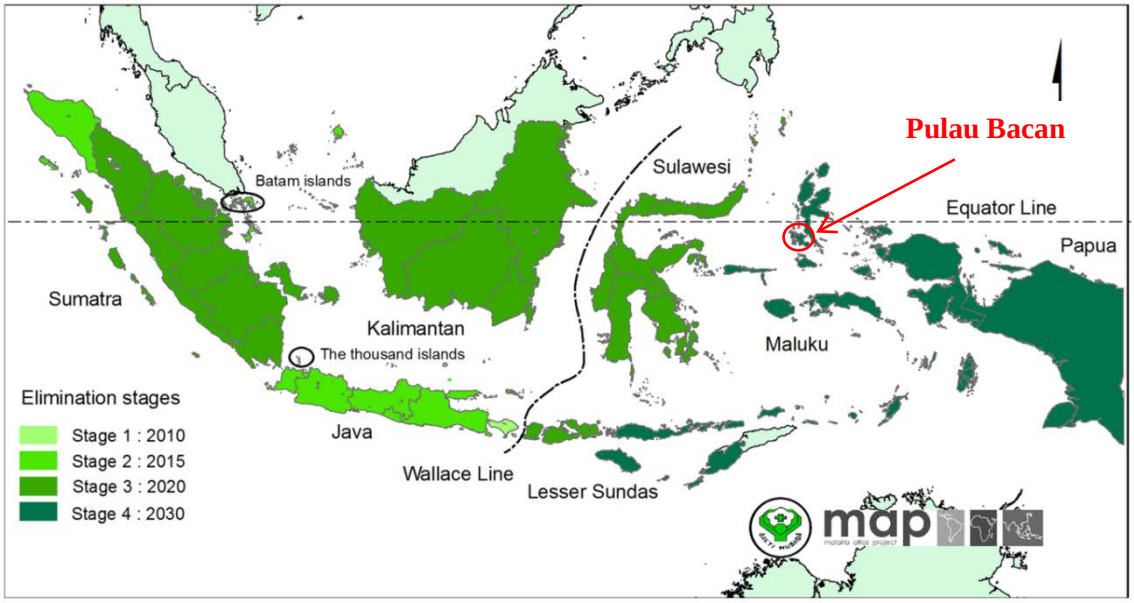


Figure 7.1. Map of Indonesian administrative boundaries and their elimination targets. Figure reproduced from Elyazar and colleagues (2011).

The diagnostic method selected for G6PDd screening was the WST-8/PMS diagnostic developed by Tantular and Kawamoto (2003) (Dojindo Laboratories) as a "simple" screening method. This dye decolourisation diagnostic required a team of two or three individuals, including a qualified nurse, to run the tests. It also depended on a cold-chain

Continued overleaf

Box 7.1. Cont.

for the reagents, and a full diagnostic apparatus filling several boxes. Although these requirements were certainly possible to meet in the context of a bespoke population screening survey, applying them in routine point-of-care scenarios would be prohibitive. Most constraining, however, was the difficulty of interpreting the results (Figure 7.2). The slight colour differences characteristic of normal and deficient expression made naked eye judgements of deficiency hard to interpret and required two independent blinded assessments. A third opinion was sought to settle discrepancies.

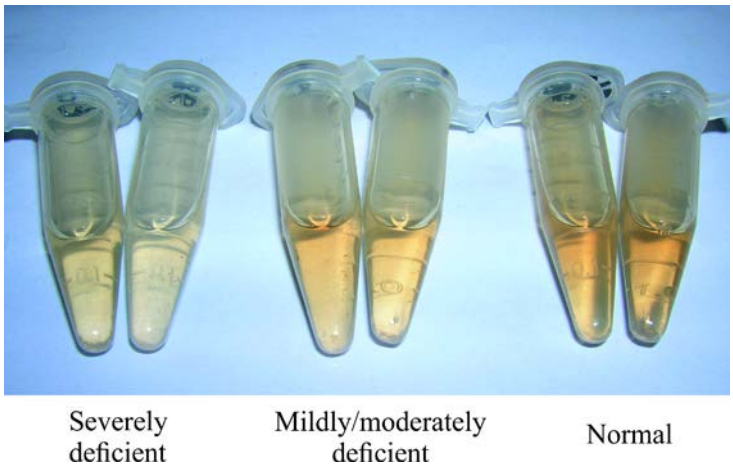


Figure 7.2. Results of the WST-8/PMS rapid screening method, showing colour differences after the recommended 20 minute reaction time: severely deficient (left), mildly or moderately deficient (centre) and normal (right) G6PD activity.

My direct involvement with this study emphasised to me the difficulties associated with existing diagnostic methods. Significant efforts are ongoing towards developing more suitable alternative diagnostics (Medicines for Malaria Venture/GlaxoSmithKline; Kim *et al.*, 2011; PATH, 2011; Eziefula *et al.*, 2012) and it seems reasonable to hope that such a product will be available relatively soon. As an aside, a G6PDd RDT is also likely to be a prerequisite to tafenoquine licencing (*Pers. Comm.*, Justin Green, GSK Tafenoquine Project Physician Lead; 6 May 2012). Difficult issues concerning female diagnoses (see the Appendix to Chapter 4), however, will need to be addressed. The G6PDd prevalence and *P. vivax* endemicity maps can support assessments of the relative need for these tests between areas. The immediate benefits from a simple G6PDd RDT would be in allowing much broader access to primaquine. However, as discussed in Chapter 6, the need for

demonstrating the relationship between residual enzyme activity and haemolytic risk remains to fully validate the suitability of a simple enzyme activity diagnostic.

In the longer term, the issues discussed at length in Chapter 6, namely the difficulties of characterising the primaquine sensitivity phenotypes of the common G6PDd variants and of predicting haemolytic risk, must be addressed. The heterogeneous maps of G6PDd prevalence and variants, however, indicate that a single binary diagnostic may not be optimal for ensuring maximal use of primaquine. For example, the extended dosing regimens discussed in Chapters 5 and 6 reduce the haemolytic risks associated with therapy whilst maintaining the drug's therapeutic efficacy. A more intricate understanding, both of the epidemiology and of the molecular mechanisms causing primaquine-induced haemolysis, would allow a more refined and targeted approach to increasing safe access to this important therapy.

Ultimately, the development of a non-toxic alternative to the 8-aminoquinolines (primaquine and tafenoquine) would provide the optimal solution to ensuring access to safe *P. vivax* radical cure. No such alternatives are currently in development stages (Medicines for Malaria Venture). The potential benefits of widespread use of such a drug, however, are very far-reaching and motivating towards this goal.

The different R&D steps which I envisage along the pathway to overcoming the difficulties presented by G6PDd to *P. vivax* radical cure are summarised in Figure 7.3.

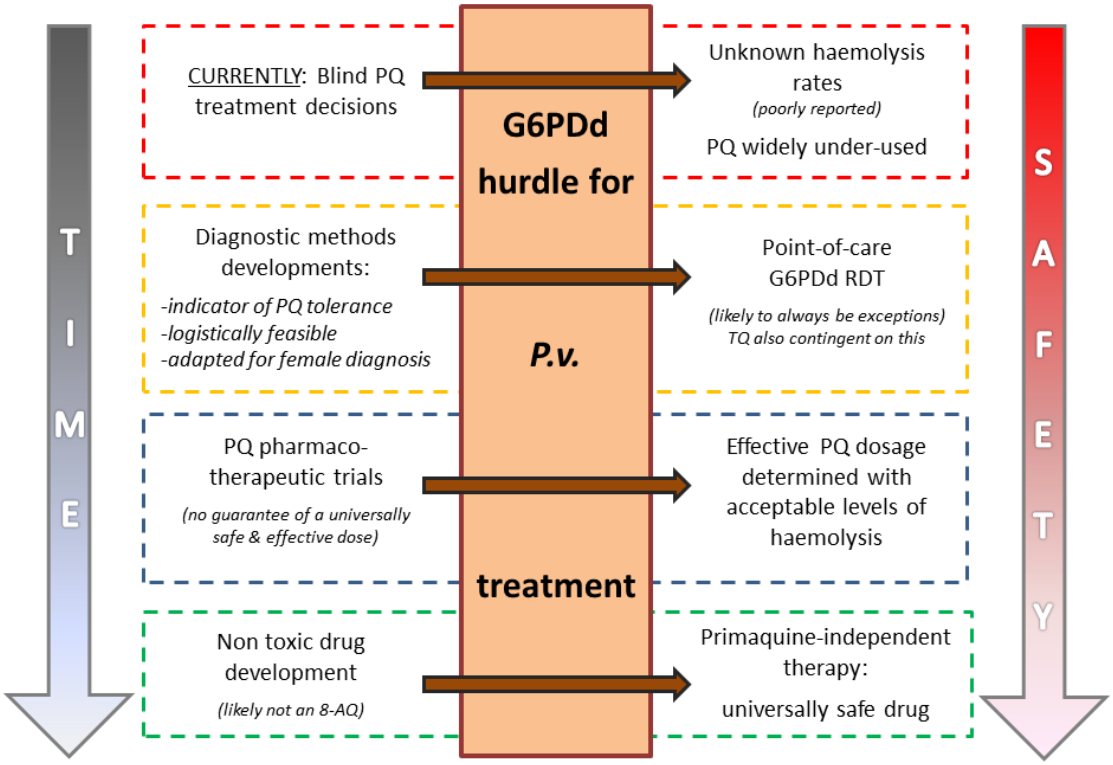


Figure 7.3. Necessary future R&D aims to increase safe access to *P. vivax* radical cure. The left side of the orange barrier indicate the types of studies needed, and the right side list the outputs required. Progression from top to bottom of the plot represents time and relative safety associated with *P. vivax* therapeutic options.

7.5. Conclusions

The maps developed through this research are certainly not perfect, and suffer important limitations from knowledge gaps, both in their underlying evidence-bases and in their interpretation as indicators of *P. vivax* immunity and of primaquine haemolytic risk. The maps do serve, nevertheless, a number of important purposes. First, they represent the first robust evidence-base to support assessments of the magnitude of the public health problems they address: the distribution and numbers of individuals at risk of *P. vivax* infection and the prohibitively widespread risks from primaquine therapy. Second, these large-scale assessments provide a basis for advocacy for further study. The potentially far-reaching negative repercussions of unrecognised *P. vivax* transmission in Africa and the potential benefits of a robust spatial framework assessing G6PDd-associated risk support calls for

further research into all the themes discussed in this chapter. Refining the analytical resolution of these maps will require a concerted inter-disciplinary effort bringing together information from numerous types of studies: molecular, biochemical, clinical and geographic.

This thesis opened by discussing the neglect of *P. vivax* as a legacy of the GMEP. The results generated through this research have served to support two specific knowledge gaps identified as particularly pertinent to understanding the spatial epidemiology of *P. vivax* with the target end-point of supporting its control and eventual elimination. Through this thesis, I hope to have demonstrated value of integrating spatial maps of human genetic traits into infectious disease models and public health decision making, and the need and nature of the future research required to sustain the increased attention which *P. vivax* has received in the years since starting this doctoral research.

7.6. References

1000 Genomes - A Deep Catalog of Human Genetic Variation. Accessed: 10 Dec 2012. URL: www.1000genomes.org.

APMEN. Asia Pacific Malaria Elimination Network. Accessed: 23 Nov 2012. URL: http://apmen.org/.

Baird, J.K. (2012). Primaquine toxicity forestalls effective therapeutic management of the endemic malarias. *International Journal for Parasitology* **42**(12): 1049-1054.

Baird, J.K. (2013). Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. *Clinical Microbiology Reviews* **26**(1): 1-22.

Baird, J.K. and Surjadjaja, C. (2011). Consideration of ethics in primaquine therapy against malaria transmission. *Trends in Parasitology* **27**(1): 11-16.

Crawford, M., Ed. (2007). Anthropological Genetic: Theory, Methods and Applications. Cambridge, Cambridge University Press.

Culleton, R. and Carter, R. (2012). African *Plasmodium vivax*: Distribution and origins. *International Journal for Parasitology* **42**(12): 1091-1097.

Culleton, R., Ndounga, M., Zeyrek, F.Y., et al. (2009). Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. *Journal of Infectious Diseases* **200**(9): 1465-1469.

Culleton, R.L., Mita, T., Ndounga, M., et al. (2008). Failure to detect *Plasmodium vivax* in West and Central Africa by PCR species typing. *Malaria Journal* **7**: 174.

Elyazar, I.R., Gething, P.W., Patil, A.P., et al. (2011). *Plasmodium falciparum* malaria endemicity in Indonesia in 2010. *PLoS One* **6**(6): e21315.

Elyazar, I.R., Gething, P.W., Patil, A.P., et al. (2012). *Plasmodium vivax* malaria endemicity in Indonesia in 2010. *PLoS One* **7**(5): e37325.

Eziefula, A.C., Gosling, R., Hwang, J., et al. (2012). Rationale for short course primaquine in Africa to interrupt malaria transmission. *Malaria Journal* **11**: 360.

Gething, P.W., Elyazar, I.R., Moyes, C.L., et al. (2012). A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Neglected Tropical Diseases* **6**(9): e1814.

Gething, P.W., Smith, D.L., Patil, A.P., et al. (2010). Climate change and the global malaria recession. *Nature* **465**(7296): 342-345.

Guerra, C.A., Howes, R.E., Patil, A.P., et al. (2010). The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Neglected Tropical Diseases* **4**(8): e774.

Hardy, G.H. (1908). Mendelian Proportions in a Mixed Population. *Science* **28**(706): 49-50.

Howes, R.E., Patil, A.P., Piel, F.B., et al. (2011). The global distribution of the Duffy blood group. *Nature Communications* **2**: 266.

Howes, R.E., Piel, F.B., Patil, A.P., et al. (2012). G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Medicine* **9**(11): e1001339.

International HapMap Project. Accessed: 10 Dec 2012. URL: http://hapmap.ncbi.nlm.nih.gov.

Kasehagen, L.J., Mueller, I., Kiniboro, B., et al. (2007). Reduced *Plasmodium vivax* erythrocyte infection in PNG Duffy-negative heterozygotes. *PLoS One* **2**(3): e336.

Kim, S., Nguon, C., Guillard, B., et al. (2011). Performance of the CareStart G6PD deficiency screening test, a point-of-care diagnostic for primaquine therapy screening. *PLoS One* **6**(12): e28357.

King, C.L., Adams, J.H., Xianli, J., et al. (2011). Fy(a)/Fy(b) antigen polymorphism in human erythrocyte Duffy antigen affects susceptibility to *Plasmodium vivax* malaria. *Proceedings of the National Academy of Sciences of the United States of America* **108**(50): 20113-20118.

MalariaGen Genomic Epidemiology Network. Accessed: 10 Dec 2012. URL: http://www.malariagen.net/.

Medicines for Malaria Venture. MMV Research & Development. Accessed: 26 Nov 2012. URL: http://www.mmv.org/research-development.

Medicines for Malaria Venture/GlaxoSmithKline. MMV/GSK Tafenoquine Phase IIb/III trials. Accessed: 26 Nov 2012. URL: http://www.mmv.org/research-development/project-portfolio/tafenoquine.

Menard, D., Barnadas, C., Bouchier, C., et al. (2010). *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proceedings of the National Academy of Sciences of the United States of America* **107**(13): 5967-5971.

Mueller, I., Galinski, M.R., Baird, J.K., et al. (2009). Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infectious Diseases* **9**(9): 555-566.

PATH (2011). Staying the course? Malaria research and development in a time of economic uncertainty. Seattle. PATH.

Patil, A.P., Gething, P.W., Piel, F.B., et al. (2011). Bayesian geostatistics in health cartography: the perspective of malaria. *Trends in Parasitology* **27**(6): 246-253.

Piel, F.B., Patil, A.P., Howes, R.E., et al. (2012). Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet*: doi:10.1016/S0140-6736(1012)61229-X.

Piel, F.B., Patil, A.P., Howes, R.E., et al. (2010). Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications* **1**: 104.

Ryan, J.R., Stoute, J.A., Amon, J., et al. (2006). Evidence for transmission of *Plasmodium vivax* among a duffy antigen negative population in Western Kenya. *American Journal of Tropical Medicine and Hygiene* **75**(4): 575-581.

Tantular, I.S. and Kawamoto, F. (2003). An improved, simple screening method for detection of glucose-6-phosphate dehydrogenase deficiency. *Tropical Medicine & International Health* **8**(6): 569-574.

Weatherall, D.J. (2010). Molecular medicine; the road to the better integration of the medical sciences in the twenty-first century. *Notes and Records of the Royal Society*: doi: 10.1098/rsnr.2010.0031.

Weinberg, W. (1908). Über den nachweis der vererbung beim menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* **64**: 368-382.

WHO (2011). World Malaria Report 2011. Geneva. World Health Organization.

Post-script

The timing of this work has made it particularly exciting. Both the Duffy blood group and G6PD deficiency were relatively unheard of among the wider malaria research community until recently. I recall a senior academic who works on the interactions of red blood cell polymorphisms with malaria in Africa telling me at the start of my doctorate that Duffy negativity was "one of those topics which I occasionally look into and forget all about within 24 hours". A few months later, I talked Pete Zimmerman through my poster of preliminary Duffy negativity maps at an ASTMH conference, and he told me about his team's recent findings of *P. vivax*-infected Duffy negative individuals. Their landmark publication has re-ignited the intrigue around *P. vivax* transmission in Africa. Similarly, *P. vivax* therapy has been in limbo for decades, but the significance of G6PD deficiency and primaquine was frequently brought home to me by the ever-enthusiastic valiant adversary of *P. vivax*, Kevin Baird. A couple of years later, unpublished copies of our G6PD deficiency maps had been requested as evidence for the WHO Expert Review Group examining single dose primaquine for transmission blocking applications. The upsurge of interest around *P. vivax*, both among the malaria research community and national control programme managers, has created the opportunity for witnessing practical applications for both avenues of the research presented in this thesis. While these results are only an imperfect beginning, they provide a starting platform which must be evolved and improved into the future.

APPENDIX

Appendix

This Appendix includes the following supplementary information:

Reference is also made throughout the thesis to two Advances in Parasitology chapters (Zimmerman *et al*. 2013 & Howes *et al*. 2013). Due to their length and space constraints in this Appendix, these are not included here but are available on request. They are in press and will be published in February 2013.

Appendix to Chapter 2 - *The global distribution of the Duffy blood group*

This Appendix includes:

1. Supplementary Figures:

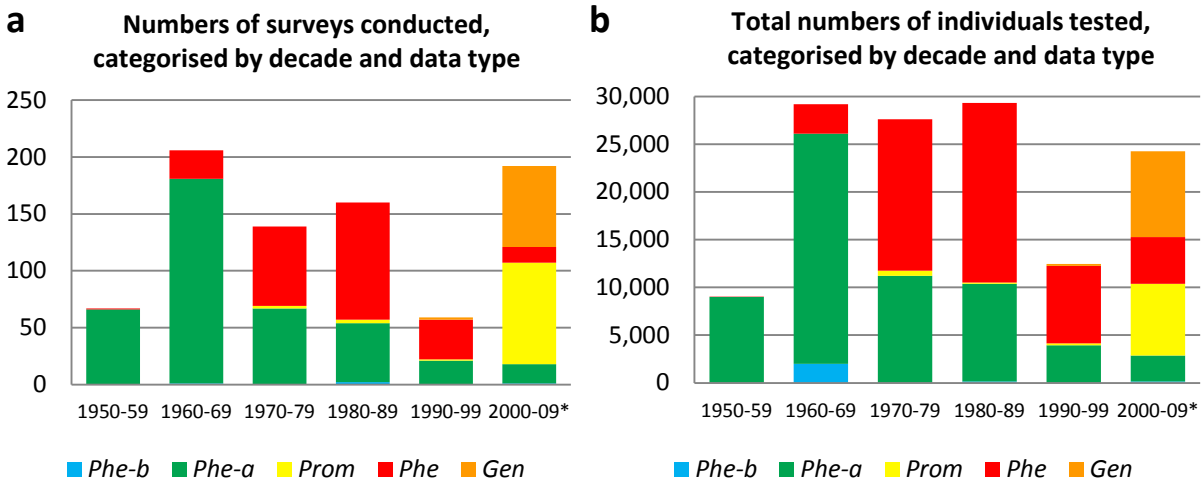
2. Supplementary Tables:

3. Supplementary Discussion:

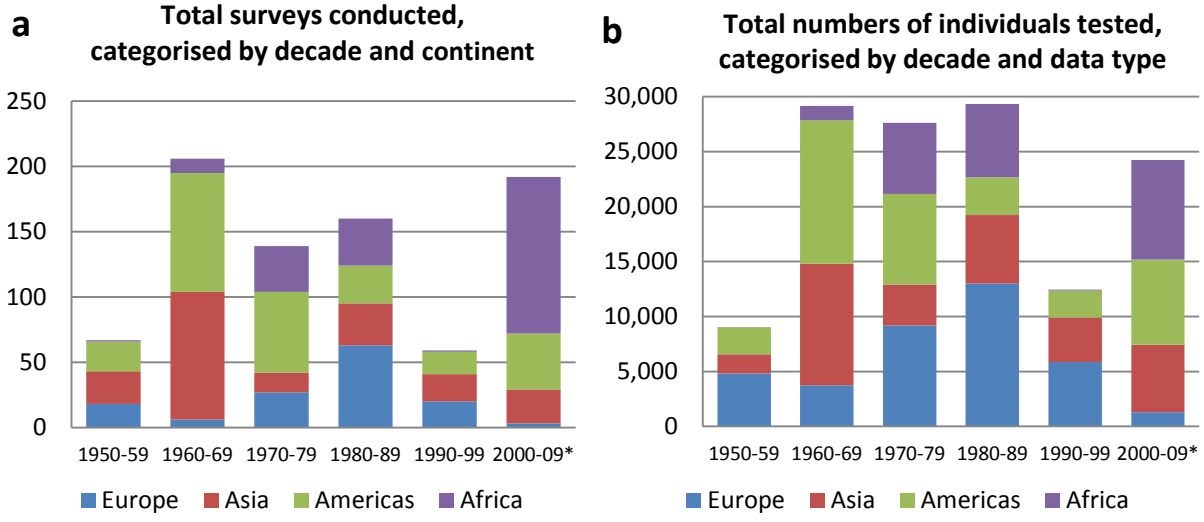
4. Supplementary Methods:

5. Supplementary References:

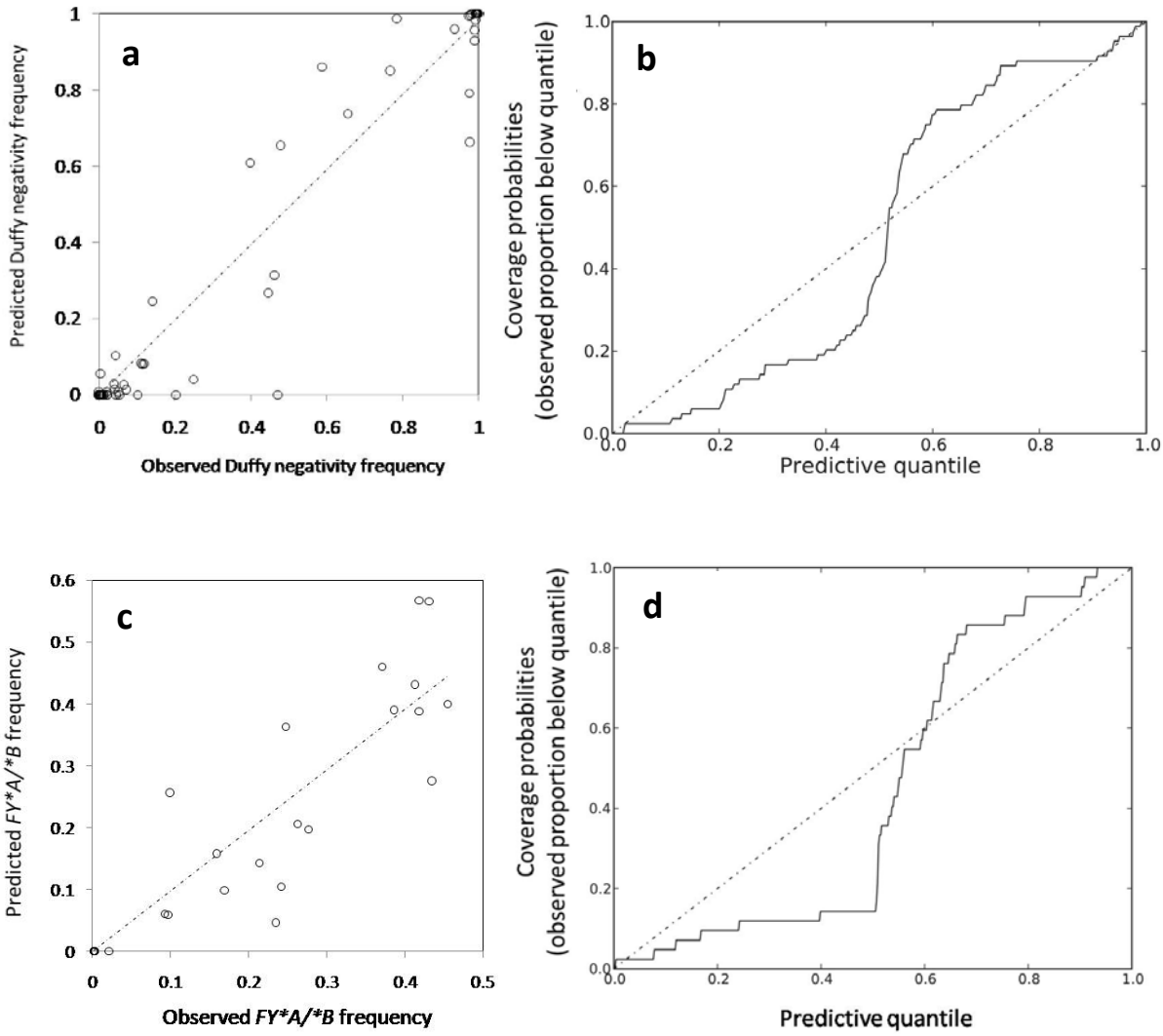
Supplementary Figures



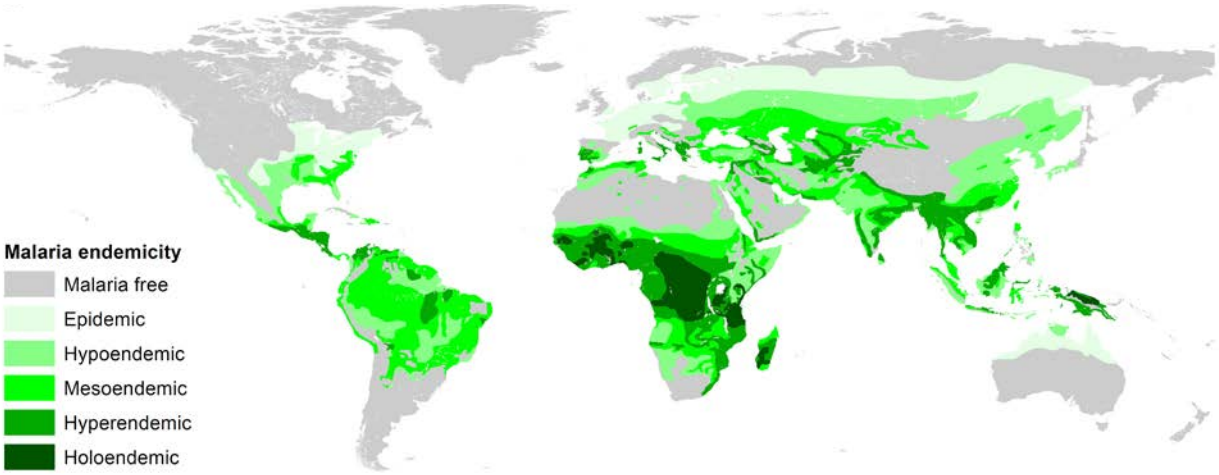
Supplementary Figure S1. Historical patterns of data types in the input data set. (a) Patterns according to survey numbers with colour-coded according to data points in Figure 2; (b) Patterns according to total individuals sampled. (*includes data from unpublished sources acquired in 2010).



Supplementary Figure S2. Historical patterns of survey locations by continent in the input data set. (a) Patterns according to survey numbers; (b) Patterns according to total individuals sampled. (*includes data from unpublished sources acquired in 2010)



Supplementary Figure S3. Model validation plots. (a) Scatter plot of actual versus predicted point-values of Duffy negativity prevalence. (b) Probability-probability plot comparing predicted probability thresholds with the actual proportion of true values below quantile for Duffy negativity; (c) and (d) show equivalent plots for *FY*A/*B* heterozygosity validation. In all plots the 1:1 line is also shown (dashed line) for reference.



Supplementary Figure S4. Historical malaria endemicity map, originally generated by Lysenko and Semashko[52], recently republished and fully described by Piel *et al*[53].The classes are defined by parasite rates (PR_{2-10}, the proportion of 2 to 10 years olds with parasites in their peripheral blood): malaria free, $PR_{2-10} = 0$; epidemic, $PR_{2-10} \approx 0$; hypoendemic, $PR_{2-10} < 0.10$; mesoendemic, $PR_{2-10} \geq 0.10$ and < 0.50; hyperendemic, $PR_{2-10} \geq 0.50$ and < 0.75; holoendemic $PR_{0-1} \geq 0.75$ (this class was measured in 0 to 1 year olds)[53].

Supplementary Tables

Parameter	Model term p_{ab}	Parameter	Model term p_0	Parameter	Model term p_1
ϕ_{ab} mean	3.57	ϕ_0 mean	7.02	p_1 mean	0.0119
ϕ_{ab} median	3.36	ϕ_0 median	6.26	p_1 median	0.0119
ϕ_{ab} std	1.24	ϕ_0 std	2.70	p_1 std	0.000921
ϕ_{ab} IQR	1.64	ϕ_0 IQR	3.36	p_1 IQR	0.00130
ϕ_{ab} 95 BCI	4.91	ϕ_0 95 BCI	10.5	p_1 95 BCI	0.00368
θ_{ab} mean	4.99	θ_0 mean	3.05		
θ_{ab} median	4.83	θ_0 median	2.23		
θ_{ab} std	3.04	θ_0 std	2.74		
θ_{ab} IQR	4.02	θ_0 IQR	3.11		
θ_{ab} 95 BCI	11.9	θ_0 95 BCI	10.7		
ν_{ab} mean	0.377	ν_0 mean	0.422		
ν_{ab} median	0.380	ν_0 median	0.427		
ν_{ab} std	0.0412	ν_0 std	0.0473		
ν_{ab} IQR	0.0640	ν_0 IQR	0.0769		
ν_{ab} 95 BCI	0.156	ν_0 95 BCI	0.152		
V_{ab} mean	0.152	V_0 mean	0.345		
V_{ab} median	0.149	V_0 median	0.323		
V_{ab} std	0.0322	V_0 std	0.123		
V_{ab} IQR	0.0492	V_0 IQR	0.163		
V_{ab} 95 BCI	0.114	V_0 95 BCI	0.478		

Supplementary Table S1. MCMC output parameter values at the three modelled loci. The first columns refer to the locus differentiating *FY*A* from *FY*B*. The association of the *FY*B* variant with the silencing Duffy negative mutation (-33C; i.e. the *FY*B*^{ES} allele) is considered in the middle columns. The third term, the *p1* variant, represents the constant modelling the frequency of association between the *FY*A* variant and the silencing promoter mutation (i.e. the *FY*A*^{ES} allele). Spatially variable parameters reported include amplitude (ϕ_0 and ϕ_{ab}), scale (θ_0 and θ_{ab}), degree of differentiability (ν_0 and ν_{ab}) and nugget variances (V_0 and V_{ab}). Summary statistics of the MCMC output include mean and median values, standard deviation ('std'), 50% interquartile range ('IQR') and 95% Bayesian credible intervals ('95 BCI'). Scale is measured in units of earth radii; other parameters are unitless. Values are presented to three significant figures.

Statistic	Sub-Saharan Africa covariate effect
Median	0.832
Mean	0.829
STD	0.310
IQR	0.420
95% BCI	1.214

Supplementary Table S2. Summary statistics for the sub-Saharan Africa (β_{africa}) covariate effect. The posterior is concentrated in positive values, indicating that the covariate increases likelihood of association between the silencing mutation and the *FY*B* locus in the sub-Saharan Africa region. This is reflected by increased frequencies of $FY*B^{ES}$ south of the covariate boundary. (Full model details are given in the Supplementary Methods 1, pages S2-4).

	N sites	N samples (all sites)	Variants		% Duffy positive†
Genotype	22	272	$FY*B^{ES}/FY*B^{ES}$	272	0%
Phenotype	11	3,738	Fy(a+b+) Fy(a+b-) Fy(a-b+) Fy(a-b-)	1 13 12 3,712	0.70%
Promoter	83	7290	Fy-pos Fy-neg	24 7,266	0.33%
Phenotype-a	5	508	Fy(a+) Fy(a-)	7 501	1.38% (excludes any Fy(a-b+))
Phenotype-b	2	148	Fy(b+) Fy(b-)	1 147	0.68% (excludes any Fy(a+b-))

Supplementary Table S3. Duffy positive samples in the predicted 98-100% Duffy Negative region. Total surveys conducted in this region: n=123.

† Of the 11,956 individuals surveyed across the predicted 98-100% Duffy negativity region, 58 Duffy positive individuals were identified at 22 sites across nine countries: Angola, Cameroon, Côte d'Ivoire, The Gambia, Kenya, Malawi, Mozambique, Nigeria and the United Republic of Tanzania.

Supplementary Discussion

Comparison with existing maps.

The only previously published maps of the Duffy alleles are those of Cavalli-Sforza *et al*, in their History and Geography of Human Genes (HGHG[54]). The following discussion centres on (i) the general and (ii) the specific cartographic differences between these two mapping efforts.

In general terms, both efforts employed a similar conceptual framework: an evidence-base of Duffy blood group surveys which informs a statistical model and generates a prediction map. The main methodological differences between the HGHG and the current mapping efforts can, therefore, be examined as (i) the evidence-base, and (ii) the statistical modelling method employed. The third aspect of this discussion will examine the specific differences between the resulting maps.

1. Database

The most obvious advantage that the MAP database has over the HGHG evidence base is twenty years of additional data collection, including all the genotyped data (the most informative data type). The data assembly effort, including reference gathering methods, criteria for survey inclusion and geopositioning protocols are documented in more detail in this effort than previously in the HGHG. In addition, from our review of their sources, it is apparent that non-representative samples of communities were included in the HGHG dataset, such as studies selecting specific ethnic groups from ethnically diverse communities, groups of related individuals and malaria patients, providing potentially biased allelic frequency estimates. These various limitations were addressed in the present study.

The methodological descriptions provided in this paper and the referenced sources for additional information are intended to provide sufficient detail to enable independent reproduction and thus objective evaluation. This addresses another significant limitation to the methods of the HGHG protocols, where only limited documentation is presented on the cartographic methodology employed. The full list of sources used here is given in the Supplementary References; the complete derived database will be made freely accessible online in mid-2011 (the model input extracted from the database is published here as Supplementary Data); the statistical code is freely available for download from the open-access github repository (https://github.com/malaria-atlas-project).

2. Statistical mapping model

The statistical mapping methods employed in the present effort benefit from two decades of development in the field of geostatistics, enabling better representation of small-scale variation in gene frequencies. The most significant development, however, relates to the Bayesian framework which allows the map surface to be interpreted according to its relative reliability by generating numerous iterations of predictions from which various summary indicators (e.g. mean or median) and uncertainty measures can be derived. An immediate advantage of this is enabling predictions for regions where data are absent, and quantifying the certainty in the predictions; in contrast, data

limitations, such as in Madagascar, prevented Cavalli-Sforza *et al* from making predictions in data-poor areas. Further, the importance of uncertainty metrics has recently been emphasised to the public health research community[55].

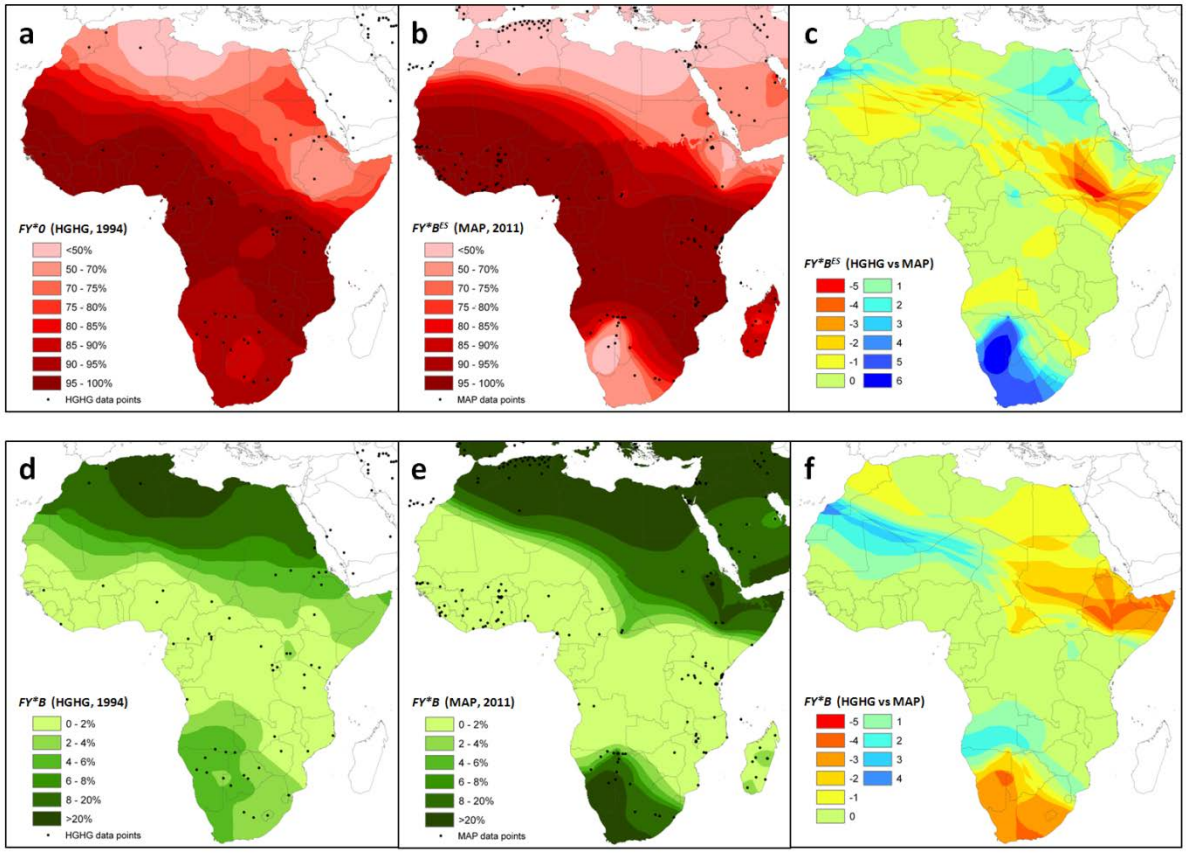
Furthermore, the HGHG model input was restricted to surveys specifically informing the frequency of each variant, with each variant being mapped in isolation of the others, in spite of their intricate associations. The multi-allelic framework employed here, adapted for the five data types previously described, allowed all data to inform the predictions of each allele simultaneously.

3. Spatial differences between maps

Comparisons of the predictions for all three alleles are discussed here. As Cavalli-Sforza *et al*[54] do not present global frequency maps for all alleles, predictions of *FY*B* and *FY*B*^{ES} frequencies focus on the African continent, while the *FY*A* maps are discussed on a global scale.

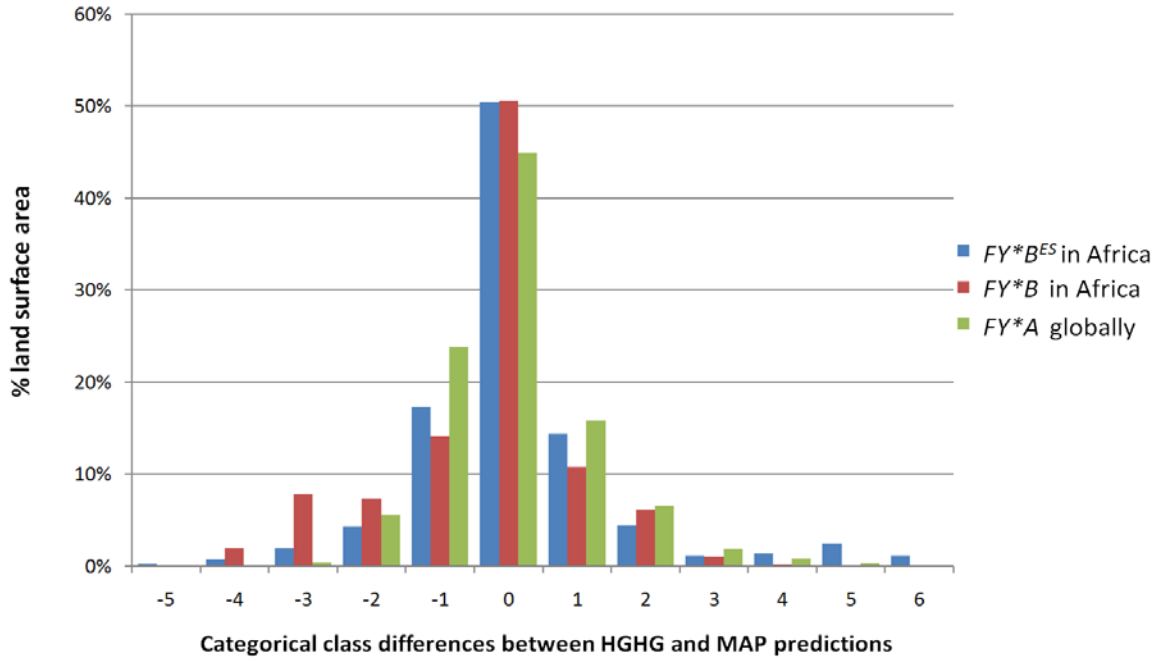
A limitation of the practical applicability of the HGHG maps is that their outputs include only gene frequency maps. We present here the first published Duffy negativity phenotype map, as we believe this output will be most informative to the anticipated end-user community. A further limitation of the HGHG maps derives from their presentation, specifically the categorical boundaries employed. The highest allele frequency band (95-100%) corresponds to a wide range of Duffy negative phenotype frequencies (90-100%), which could encompass a wide range of *P. vivax* epidemiological scenarios. Modelled as a continuous measure, the new version can be used as a continuous or categorical surface (available on request from the authors).

i. Null Duffy allele, *FY*B*^{ES}, and *FY*B* in Africa



Supplementary Figure S8. Comparative display of the HGHG and MAP allele frequency maps for *FY*B*^{ES} and *FY*B* in Africa. Supplementary Figures S8a-c represent the silent *FY*B*^{ES} allele (denoted *FY*0* in the HGHG), and S8d-f the *FY*B* allele frequencies. Differences between the HGHG maps (S8a and S8d) and the new maps (S8b and S8e) are shown in panels S8c and S8f, respectively. Discrepancies are represented by the difference in number of classes between the two versions, based on the categorical classes defined by the HGHG maps. Negative differences (yellow to red) indicate areas where the MAP version predicted higher frequencies than the HGHG version, positive differences (in blue) are where the HGHG version predicted frequencies higher than in the new version. Same class predictions in the HGHG and MAP versions appear in pale green. Black datapoints represent input data points (HGHG: n=41; MAP: n=203).

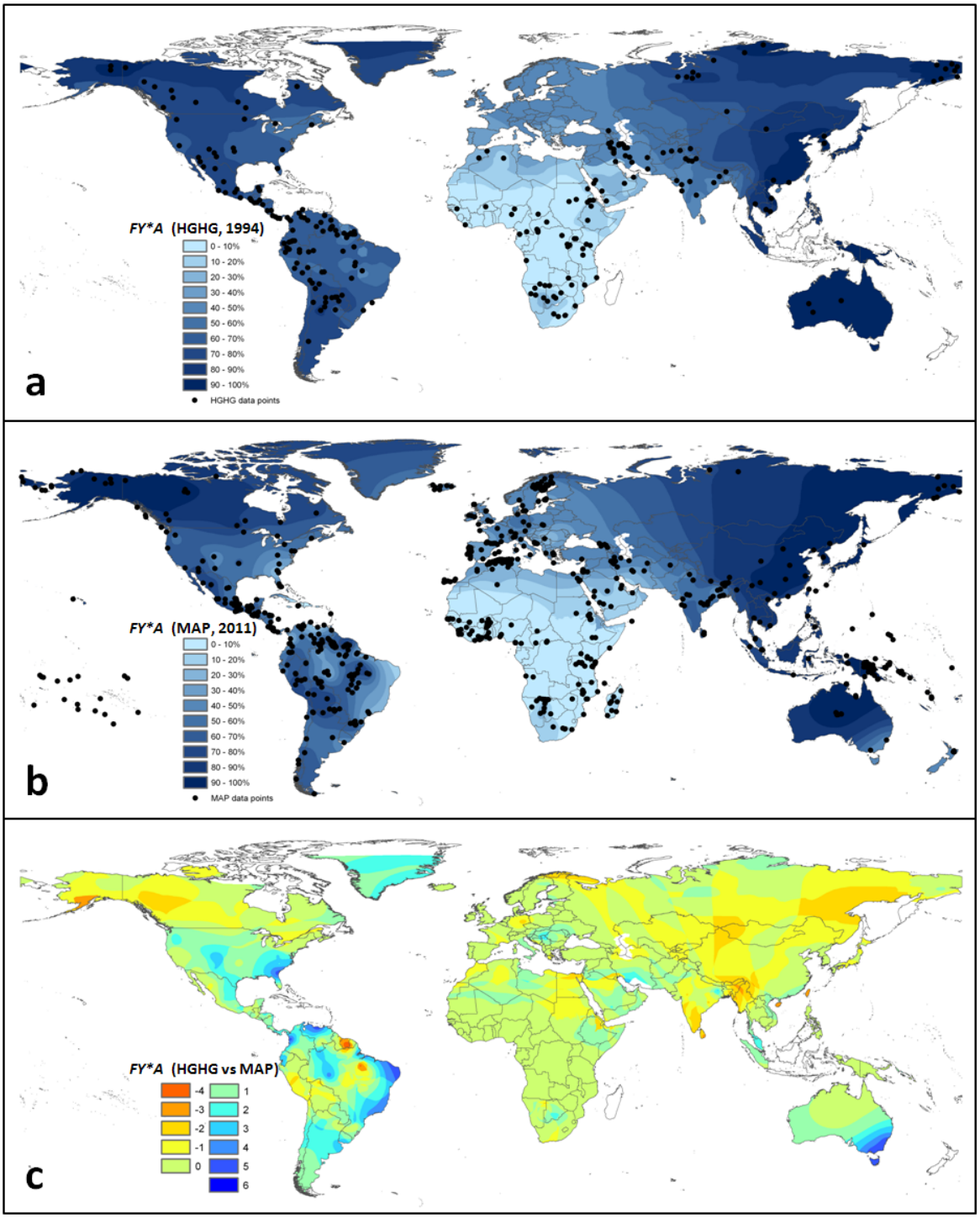
Visual comparison of the *FY*B*^{ES} and *FY*B* series of maps for Africa (Supp. Figure S8) reveals roughly comparable unskewed outputs, an observation supported by quantitative summaries of the difference maps (Supp. Figure S9). The categorical difference maps (Supp. Figure S8c and S8f) reveal 50% and 51% concordance for *FY*B*^{ES} and *FY*B* frequency predictions, respectively, between the HGHG and MAP maps (Supp. Figure S9). A quarter (24%) of the *FY*B* prediction surface area differed by two classes or more (≥4% difference in allelic frequency), and 18% of the *FY*B*^{ES} predicted area differed by 20% or more (2 frequency classes).



Supplementary Figure S9. Histogram of differences between frequency categories of the HGHG and MAP predictions for *FY*B*^{ES} and *FY*B* allele frequencies in Africa (HGHG – MAP prediction). Negative differences indicate higher frequency predictions in the MAP version than in the HGHG version; positive differences mean the HGHG version predicted frequencies higher than in the new version. The differences are quantified by proportions of land surface area.

Due to the additive nature of the variant frequencies, the location of the discrepancies overlaps for both alleles, namely the boundary zone around the region of highest Duffy negativity prevalence. The most striking area of discrepancy is the prediction for southern Africa, where the HGHG map predicts much higher frequencies of *FY*B*^{ES} and lower frequencies of *FY*B* than the new iteration does, a reflection of the new dataset which includes surveys reporting very low rates and even absence of Duffy negativity in the local population, an element not reflected in the HGHG map. At the eastern limit of the high frequency zone, the indent of lower *FY*B*^{ES} and higher *FY*B* frequencies into Sudan is less pronounced into the HGHG map than the current one. Similarly, predictions in Ethiopia for frequencies of both allele frequencies are generally lower in the HGHG map. The position of the northern boundary of high Duffy negativity differs by one or two classes between the maps, with higher frequencies of *FY*B*^{ES} stretching further north and correspondingly lower frequencies of *FY*B* predicted in the new maps.

The database updates yielded a five-fold increase in the input evidence-base, from n_{Africa}=41 in the HGHG version (though only 27 points could be digitised from their published maps) to n_{Africa}=203 in the current iteration (including 37 *Genotype* and 61 *Phenotype*), (Figure 2). The differences in the southern and eastern parts of the distributions are directly informed by the updated dataset. Notably, the highly informative genotype datapoints in southern Ethiopia enable the more spatially convoluted prediction in this area. However, being transition regions becoming increasingly heterozygous, both are also areas of high uncertainty (Figure 3e-f). The spectrum of frequencies across the western and central Sahara is informed by a much larger dataset in the current iteration than the HGHG one, both across the sub-Saharan countries (Nigeria and West Africa: 3 datapoints in



Supplementary Figure S10. Comparative display of the HGHG and MAP allele frequency maps of *FY*A* globally. Differences between the HGHG map (S10a) and the new map (S10b) are shown in panel S9c. Discrepancies are represented by the difference in number of classes between the two versions, based on the categorical classes defined by the HGHG maps. Negative differences (yellow to orange) indicate areas where the MAP version predicted higher frequencies than the HGHG version; positive differences (in blue) are shown where the HGHG version predicted frequencies higher than in the new version. Pale green indicates same class predictions in both versions. Blanks in the HGHG and difference maps indicate areas where no prediction was made, including many islands, most prominently Madagascar and large parts of Indonesia. Black datapoints represent input data points (HGHG: n=751; MAP: n=821). [*Although it was only possible to digitise 241 of the HGHG datapoints from their published map – many appear to be spatial duplicates, and the European datapoints were missing from their global* FY*A *map*].

the HGHG, 64 in the updated dataset) and the Maghreb region (west of Libya: 1 datapoint in the HGHG, 29 in the updated version). However, no surveys were identified within the Sahara region, likely a reflection of the low population densities in the area. The exact position of the frequency class boundaries therefore, as it corresponds to almost zero population levels, is of minor consequence. MAP predictions across the desert are moderated by increased uncertainty (Figure 3d-f), contrasting with the much more certain predictions across central Africa.

ii. *FY*A* global distribution

Only *FY*A* and *FY*B* are commonly found outside the Africa region. To avoid repetition, only the one set of maps are discussed here – *FY*A*, the only global Duffy allele frequency map presented in the HGHG.

Overall the *FY*A* HGHG and MAP maps reveal good correspondence (Supp. Figure S10), with 85% of the prediction surfaces being within 10% of the other. Largest differences are in the Americas, where the population composition is known to be highly heterogeneous with populations of diverse origins. Predictions from south eastern Australia also differ by up to 50% (5 prediction classes), with the MAP prediction falling to frequencies of 40 to 50% in areas predicted to be 90 to 100% by the HGHG map. The MAP prediction in this region is informed by two relatively large surveys from the SE Australian coast (n=304[56] and n=788[57]) and three from northern New Zealand, contrasting with the uniform HGHG prediction informed only by data from central Australia. The prediction across much of Africa and Europe is largely concordant between maps. The exact position of the class boundaries across Asia varies slightly, but both show the same general trend of increasing *FY*A* frequencies eastwards across the continent.

***FY*X* variant: potential further elaboration to the model.**

A number of mutations additional to those encoding the common Duffy variants discussed here (*FY*A*, *FY*B*, $FY*B^{ES}$, $FY*A^{ES}$) have been described[58], most notably the *FY*X* allele, which reaches polymorphic frequencies among populations of European origin[59]. Despite being relatively common, insufficient data prevented its inclusion in the current mapping project. Additional data on the prevalence of this variant would have allowed extension of the model to include the C265T and G298A loci which encode the *FY*X* allele in association with the *FY*B* variant[58]. These mutations cause reduced expression levels of Fy^b antigen, characterised as the Fy(b+^{weak}) phenotype.

The lack of reliable data on this variant is largely due to the low sensitivity of agglutination diagnostics. The low levels of Fy^b antigen expressed by *FY*X* (10% wildtype levels[60]) are not consistently or reliability detected by agglutination assays, due to differences in antiserum reactivity and experimental procedures. This inconsistency means that the Fy(b+^{weak}) phenotype may occasionally have been misclassified as the Fy(b-) phenotype, overlooking the presence of the Fy^b antigen.

The *FY*B* map presented here is, therefore, defined only in terms of expression at two loci for which sufficient data existed for reliable modelling: nucleotide -33 in the promoter GATA-box region and at

position 125 of exon 2. Where detected, expression of the *FY*X* allele is included in the *FY*B* map, due to its correspondence at the two loci considered. The availability of more multi-locus *Genotype* data would allow full refinement of the model.

Supplementary Methods

Mathematical description of the Bayesian geostatistical model and its implementation.

This supplement provides information on the geostatistical model used to map the allelic frequencies of the monogenic Duffy blood group variants. General principles of the geostatistical framework used have been previously described by Diggle and Ribeiro[61].

1. Random fields

To accommodate the five input data types previously detailed (Table 2), and to simultaneously model the distribution of both the positive alleles, *FY*A* and *FY*B*, and their corresponding negative variants, *FY*A*^{ES} and *FY*B*^{ES}, the model targets two spatially-varying allele frequencies: the frequency of the Fy^{b} variant (125A, which is present in both *FY*B* and *FY*B*^{ES}), and the frequency of the promoter region silencing variant (-33C) in association with the 125A coding region variant (thus the frequency of the *FY*B*^{ES} allele). These are denoted $p_{ab}(x)$ and $p_0(x)$ respectively, where x is a location. A third allele frequency, the frequency of the silencing variant occurring in association with the 125G variant (thus the *FY*A*^{ES} allele) is modelled as a small constant denoted p_1. This allele's highly restricted distribution, as reported in the assembled database, and its low frequencies where it was detected, prevent its distribution being modelled spatially as for the other alleles. The model for these allele frequencies is as follows:

$$
\begin{aligned}
m_{ab} &\sim \text{Normal}(0, 10000) \\
m_0 &\sim \text{Normal}(0, 10000) \\
\beta_{\texttt{africa}} &\sim \text{Normal}(0, 10000) \\
p_1 &\sim \text{Uniform}(0, .05)
\end{aligned}
$$

$$
\begin{aligned}
M_{ab}(x) &= \beta_{\text{africa}} 1_{x \text{ in africa}} + m_{ab} \\
M_0(x) &= m_0
\end{aligned}
$$

$$
\begin{aligned}
\phi_{ab} &\sim \text{Exponential}(.1) \\
\phi_0 &\sim \text{Exponential}(.1) \\
\theta_{ab} &\sim \text{Exponential}(.1) \\
\theta_0 &\sim \text{Exponential}(.1) \\
\nu_{ab} &\sim \text{Uniform}(0, 3) \\
\nu_0 &\sim \text{Uniform}(0, 3) \\
V_{ab} &\sim \text{Exponential}(.1) \\
V_0 &\sim \text{Exponential}(.1)
\end{aligned}
$$

$$
\begin{aligned}
C_{ab}(x, y) &= \phi_{ab} \text{Matèrn}(d(x, y)/\theta_{ab}; \nu_{ab}) + V_{ab} 1_{x=y} \\
C_0(x, y) &= \phi_0 \text{Matèrn}(d(x, y)/\theta_0; \nu_0) + V_0 1_{x=y}
\end{aligned}
$$

$$
\begin{aligned}
f_{ab} &\sim \text{GP}(M_{ab}, C_{ab}) \\
f_0 &\sim \text{GP}(M_0, C_0) \\
p_{ab}(x) &= \text{logit}^{-1}(f_{ab}(x)) \\
p_0(x) &= \text{logit}^{-1}(f_0(x))
\end{aligned}
$$

The mean function for f_0 simply returns a constant, m_0. The mean of f_{ab} takes presence in sub-Saharan Africa as a covariate with coefficient β_{africa}, and also a constant term m_{ab}.

Both fields use the Matèrn covariance function[62]. The range parameters of f_0 and f_{ab} are θ_0 and θ_{ab}, respectively; the corresponding amplitude parameters are ϕ_0 and ϕ_{ab}; the degree-of-differentiability parameters are denoted ν_0 and ν_{ab}; and the nugget variances are V_0 and V_{ab}. The distance function d gives the great-circle distance between its arguments.

The Gaussian random fields are converted to probabilities using the standard inverse logit link function.

2. Likelihood

The probability of the 125A variant being present (variant encoding Fy^b antigen) is $p_{ab}(x)$, and should be much higher for x in Africa. Given that $b = 1$, the promoter region "erythrocyte silent" -33C variant occurs with probability $p_0(x)$. Given that $b = 0$, the -33C variant happens with probability p_1, which is assumed to be a small, constant value (representing the *FY*A^{ES}* allele). Hardy-Weinberg assumptions apply to the genotype frequencies.

Input data falls into five categories: *gen, phe, prom, aphe, bphe* defined by the diagnostic method used (Table 1), and thus what phenotypic/genotypic information is discernable. The nomenclature system used includes a prefix denoting the data category (**gen***, **phe***, **prom***, **aphe***, **bphe***) followed by a symbol for each variant:***a** for *FY*A* or Fy^a variant, ***b** for *FY*B* or Fy^b variant, allelic variant ***0** for *FY*B^{ES}*, allelic variant ***1** for *FY*A^{ES}*, phenotypic variant ***0** to denote absence. These are fully described as follows:

gen* : Genotype data. Allelic frequencies are:

$$\begin{aligned}
&\mathbf{gena}\ \ (1 - p_{ab}(x))(1 - p_1)\\
&\mathbf{genb}\ \ p_{ab}(x)(1 - p_0(x))\\
&\mathbf{gen0}\ \ p_{ab}(x)p_0(x)\\
&\mathbf{gen1}\ \ (1 - p_{ab}(x))p_1
\end{aligned}$$

The genotype frequencies (**genaa** (or *FY*A/FY*A*), **genab** (or *FY*A/FY*B*), etc.) can be obtained using the standard Hardy-Weinberg formula. For example, the frequency of **genab** is twice the product of the frequencies of **gena** and **genb**, which is $2(1 - p_{ab}(x))(1 - p_1)p_{ab}(x)(1 - p_0(x))$[63,64].

phe* : Phenotype data. Studies where full phenotype resolution was provided.

pheab	This can only happen if the genotype is **genab**.
phea	This can only happen if the genotype is **gena0**, **gena1** or **genaa**.
pheb	This can only happen if the genotype is **genb0**, **genb1** or **genbb**.
phe0	This can only happen if the genotype is **gen00**, **gen01** or **gen11**.

prom* : Molecular data. Studies that considered only the promoter region variant (T-33C).

prom0	This can only happen if the genotype is **gen00**, **gen01** or **gen11**.
promab	This corresponds to the complement of **prom0**.

aphe* : Phenotype data. Only anti-Fya antibody was used in the diagnosis.

aphea	This can only happen if the genotype is **genaa**, **genab**, **gena1** or **gena0**.
aphe0	This corresponds to the complement of **aphea**.

bphe* : Phenotype data. Only anti-Fyb antibody was used in the diagnosis.

bpheb	This can only happen if the genotype is **genbb**, **genab**, **genb0** or **genb1**.
bphe0	This corresponds to the complement of **bpheb**.

The sampling distributions are assumed to be multinomial, conditional on the appropriate individual phenotype or genotype probabilities described above. This likelihood completes the Bayesian probability model.

3. Model implementation and output

As previously detailed[65], implementation of the modelling procedure was divided into two computational tasks: (i) the Bayesian inference stage which was implemented using the Markov Chain Monte Carlo (MCMC) algorithm[66] and was used to generate samples from the posterior distribution of the parameter set and the spatial random fields at the data locations; and (ii) a prediction stage in which samples were generated from the posterior distribution of allele frequencies at each prediction location on a global 10 x 10 km grid and 5 x 5 km grid across Africa.

Convergence of the MCMC tracefile was judged by visual inspection and verified using the Geweke convergence diagnostics[67]; 1.2 million MCMC iterations were run, with 10% recorded in the tracefile and the first 30,000 iterations excluded from the mapping stages. During the mapping process, the posterior distributions were thinned by 88, resulting in 1003 mapping iterations. MCMC dynamic traces[66] are available on request.

The model code was written in Python programming language (http://www.python.org), and is freely available from the MAP's code repository (http://github.com/malaria-atlas-project/duffy). MCMC algorithm[66], was used from the open-source Bayesian analysis package PyMC[68] (http://code.google.com/p/pymc). Maps were generated using Python and Fortran code, available from the MAP's code repository (http://github.com/malaria-atlas-project/generic-mbg).

Model validation procedure.

To assess the plausibility of the model's multiple outputs, two validation procedures were run to quantify the disparity between the model's predictions and hold-out subsets of the data[69]. First the frequency of the Duffy negativity phenotype was assessed, a measure determined directly by the *FY*B*^{ES} allele frequency map. Second, the frequency of heterozygosity was used to consider the reliability of all three allelic frequency predictions: *FY*A*, *FY*B*, *FY*B*^{ES}.

Selection of the validation sets:

1. Duffy negativity. Three of the five input data types (*Genotype*, *Phenotype*, *Promoter*) directly inform the frequency of the negative phenotype. A subset of these three data types corresponding to 10% of the overall dataset (n=84) were randomly selected as a hold-out dataset. The Bayesian geostatistical model was then implemented in full using the remaining 90% dataset.

2. Heterozygosity. Only molecularly diagnosed data, which assessed expression at both the promoter and coding-region loci can directly inform the frequency of allelic heterozygosity. This meant that only the *Genotype* data could be used in this hold-out dataset (n=73). Due to the small number of datapoints available, a smaller subset was held-back for validation (n=42), corresponding to 5% of the overall dataset.

Quantifying model performance

Simple statistical measures were used to quantify the model's ability to predict pixel values at unsampled locations by comparing model predictions with the held-out subset of observed values. The mean error (correlation coefficient between predicted and actual values) was used to assess overall model bias in the predictions, and mean absolute error (summary of the model's general tendency to over/underestimate frequencies) quantified the overall prediction accuracy as the average magnitude of errors in the predictions[69].

Variability of Duffy typing diagnostic methods.

Uncertainty in relation to diagnostic methodology may arise due to a number of factors, each is discussed in turn.

First, diagnoses have historically been hampered by shortages of Fy^b anti-serum, thus leading authors to 'guess' the unknown phenotypes. Across sub-Saharan Africa Fy(a-) samples were assumed to be Duffy negative samples; outside this region, Fy(a-) individuals were assumed to be Fy(b+) (e.g. refs [70-72]). The obvious uncertainty associated with such assumptions was addressed directly through the model design. According to the diagnostic methodology employed, data were categorised into five data types (Table 2) and informed the model accordingly. Our modelling techniques therefore allow use of the complete dataset for each map without requiring any such tenuous assumptions to be made.

Uncertainty may arise, however, due to the inconsistently diagnosed Fy(b+^{weak}) phenotype. This low copy-number variant may pass undetected by some antisera during agglutination assays[73]. Ideally, we would have included this additional locus as a spatially-variable term in the model. However, the allele's low frequencies, which are poorly and inconsistently reported, rendered it impossible to map this variant separately. For reasons discussed in Supplementary Discussion 2, the weakly agglutinating samples were treated as the Fy(b+) phenotype.

The second source of diagnostic uncertainty relates directly to the experimental procedures themselves. The main dichotomy between methods in terms of relative reliability is between serological methods (assessing phenotypes) and molecular methods (determining genotypes), corresponding to the greatest potential source of variation. Anticipated levels of variability between phenotypic and genotypic diagnoses can be ascertained from studies examining samples with multiple diagnostics. For example, this was recently done by Ménard *et al* on samples from Madagascar who found perfect concordance between a combination of phenotype-based methods (a microtyping kit and anti-sera, and flow cytometry) and molecular SNP analysis[74] (n = 661). Similar correspondence was reported from the range of diagnoses tested on a Brazilian sample: agglutination tests, flow cytometry analysis and PCR-RFLP DNA analysis and sequencing[75]; the same result was found among UK blood donors[76]. These diagnoses therefore support the commonly cited understanding of a tight phenotype-genotype association with Duffy blood types[76].

The third aspect of diagnostic uncertainty is introduced by experimental error. However, even poorly preserved blood samples need not necessarily be considered a major source a potential error, due to the remarkable stability of most blood antigens[70]. The ability of the antigens to maintain their integrity over time was demonstrated by the successful agglutination assay applied to six-month old samples from Papua New Guinea[77] and twelve-month old samples from the Maoris of New Zealand[78]. However, when sample degradation does arise, there is evidence from the literature of such samples being excluded from analyses (e.g. due to poor refrigeration during transport[78]). Furthermore, results described by authors as likely to contain false-positives (e.g. Livingstone *et al* in West Africa[79]) were excluded from the present study to conform with the conservative approach defined in our data abstraction protocols (main manuscript pages 15-17).

We decided not to include diagnostic methodology as a covariate in the model as the model structure already allows for the differences between antigen and DNA-based methods. These were the only two diagnostic types encountered: serological anti-serum agglutination tests (n=659) and DNA-molecular methods (n=162). We therefore considered that little additional information would have been derived from this addition to the model. It is therefore not possible to quantitatively determine the level of variation introduced through synthesis of the different methods; however, we hope to have demonstrated that, within the modelling framework used here, we believe its influence to be very low.

Supplementary References

52. Lysenko, A. J. & Semashko, I. N. in *Itogi Nauki: Medicinskaja Geografija* (ed. Lebedew, A. W.) 25-146 (Academy of Sciences, USSR, 1968).
53. Piel, F. B. *et al.* Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat. Commun.* **1**, 104 (2010).
54. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, 1994).
55. Horton, R. Science will never be the same again. *Lancet* **376**, 143-144 (2010).
56. Mitchell, R. J., Kosten, M. & Deacon, M. Population genetics of blood group polymorphisms in a sample of newborns from Melbourne, Australia. *Hum. Hered.* 36, 310-316 (1986).
57. Buckton, K., Lai, L. Y. C. & Gibson, J. B. A search for association between gene markers and serum cholesterol, triglyceride, urate and blood pressure. *Ann. Hum. Biol.* **8**, 39-48 (1981).
58. Castilho, L. *et al.* A novel FY allele in Brazilians. *Vox Sang.* **87**, 190-195 (2004).
59. Chown, B., Lewis, M. & Kaita, H. Duffy blood group system in Caucasians - evidence for a new allele. *Am. J. Hum. Genet.* **17**, 384-389 (1965).
60. Yazdanbakhsh, K. *et al.* Molecular mechanisms that lead to reduced expression of Duffy antigens. *Transfusion.* **40**, 310-320 (2000).
61. Diggle, P. J. & Ribeiro, P. J., Jr. *Model-based Geostatistics* (Springer, 2007).
62. Banerjee, S., Carlin, B. P. & Gelfand, A. E. *Hierarchical modeling and analysis for spatial data* (Chapman & Hall, 2003).
63. Hardy, G. H. Mendelian proportions in a mixed population. *Science* **28**, 49-50 (1908).
64. Weinberg, W. Über den nachweis der vererbung beim menschen. *Jahresh. Wuertt. Verh. Vaterl. Naturkd.* **64**, 369-382 (1908).
65. Hay, S. I. *et al.* A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.* **6**, e1000048 (2009).
66. Gilks, W. R., Richardson, S. & Spiegelhalter, D. *Markov chain Monte Carlo in Practice* (Chapman & Hall/CRC, 1995).
67. Geweke, J. in *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting* (eds Bernado, J. M., J. O. Berger, D. A. P., & A. F. M. Smith) (Oxford University Press, 1992).
68. Patil, A., Huard, D. & Fonnesbeck, C. PyMC: Bayesian stochastic modelling in Python. *J. Stat. Soft.* **35**, 1-81 (2010).
69. Hay, S. I. *et al.* A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.* **6**, e1000048 (2009).
70. Tills, D., Warlow, A., Kopec, A. C., Fridriksson, S. & Mourant, A. E. The blood groups and other hereditary blood factors of the Icelanders. *Ann. Hum. Biol.* **9**, 507-520 (1982).
71. Vyas, G. N., Bhatia, H. M., Banker, D. D. & Purandare, N. M. Study of blood groups and other genetical characters in six Gujarati endogamous groups in Western India. *Ann. Hum. Gen.* **22**, 185-199 (1958).
72. Matson, G. A. & Swanson, J. Distribution of hereditary blood antigens among the Maya and non-Maya Indians in Mexico and Guatemala. *Am. J. Phys. Anthrop.* **17**, 49-74 (1959).
73. Olsson, M. L. *et al.* The Fy(x) phenotype is associated with a missense mutation in the Fy(b) allele predicting Arg89Cys in the Duffy glycoprotein. *Br. J. Haematol.* **103**, 1184-1191 (1998).
74. Ménard, D. *et al. Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc. Natl. Acad. Sci. USA* **107**, 5967-5971 (2010).
75. Castilho, L. *et al.* A novel FY allele in Brazilians. *Vox Sang.* **87**, 190-195 (2004).
76. Mullighan, C. G., Marshall, S. E., Fanning, G. C., Briggs, D. C. & Welsh, K. I. Rapid haplotyping of mutations in the Duffy gene using the polymerase chain reaction and sequence-specific primers. *Tissue Antigens* **51**, 195-199 (1998).
77. Semple, N. M., Simmons, R. T., Graydon, J. J., Randmae, G. & Jamieson, D. Blood group frequencies in natives of the Central Highlands of New Guinea, and in the Bainings of New Britain. *Med. J. Aust.* **43**, 365-371 (1956).
78. Simmons, R. T., Graydon, J. J., Semple, N. M. & Fry, E. I. A blood group genetical survey in Cook Islanders, Polynesia, and comparisons with American Indians. *Am. J. Phys. Anthropol.* **13**, 667-690 (1955).

79. Livingstone, F. B., Gershowitz, H., Neel, J. V., Zuelzer, W. W. & Solomon, M. D. The distribution of several blood group genes in Liberia, the Ivory Coast and Upper Volta. *Am. J. Phys. Anthrop.* **18**, 161-178 (1960).

Sources from which data points were identified. (References 80 to 399)

Only sources reporting surveys included in the final model are listed. Population samples from these met the inclusion criteria for community representativeness and were geographically specific. References are ordered chronologically, and alphabetically by first author name within years; n=320.

80. Holländer, L. Über die Blutgruppe Duffy und ihre Verteilung in Basel. *Acta. Haematol.* **6**, 257-261 (1951).
81. Race, R. R., Holt, H. A. & Thompson, J. S. The inheritance and distribution of the Duffy blood groups. *Heredity* **5**, 103-110 (1951).
82. Allison, A. C., Hartmann, O., Brendemoen, O. J. & Mourant, A. E. The blood groups of the Norwegian Lapps. *Acta. Pathol. Microbiol. Scand.* **31**, 334-338 (1952).
83. Race, R. R. & Sanger, R. The inheritance of the Duffy blood groups: an analysis of 110 English families. *Heredity* **6**, 111-119 (1952).
84. Pantin, A. M. & Kallsen, R. The blood groups of the Diegueno Indians. *Am. J. Phys. Anthropol.* **11**, 91-96 (1953).
85. Polunin, I. & Sneah, P. H. A. Studies of blood groups in South-East Asia. *J. R. Anthrop. Inst.* **83**, 215-251 (1953).
86. Simmons, R. T., Graydon, J. J. & Sringam, S. A blood group genetical survey in Thais, Bangkok. *Am. J. Phys. Anthropol.* **12**, 407-412 (1954).
87. Matson, G. A., Koch, E. A. & Levine, P. A study of the hereditary blood factors among the Chippewa Indians of Minnesota. *Am. J. Phys. Anthropol.* **12**, 413-426 (1954).
88. Walsh, R. J., Kooptzoff, O., Dunn, D. & Atienza, R. Y. Blood groups of Filipinos. *Oceania* **25**, 61-67 (1954).
89. Simmons, R. T., Graydon, J. J., Semple, N. M. & Fry, E. I. A blood group genetical survey in Cook Islanders, Polynesia, and comparisons with American Indians. *Am. J. Phys. Anthropol.* **13**, 667-690 (1955).
90. Sanger, R., Race, R. R. & Jack, J. The Duffy blood groups of New York negroes: the phenotype Fy (a-b-). *Br. J. Haematol.* **1**, 370-374 (1955).
91. Chown, B. & Lewis, M. The blood groups genes of the Cree Indians and the Eskimos of the Ungawa district of Canada. *Am. J. Phys. Anthropol.* **14**, 215-224 (1956).
92. Nijenhuis, L. E. Blood group frequencies in French Basques. *Hum. Hered.* **6**, 531-535 (1956).
93. Semple, N. M., Simmons, R. T., Graydon, J. J., Randmae, G. & Jamieson, D. Blood group frequencies in natives of the Central Highlands of New Guinea, and in the Bainings of New Britain. *Med. J. Aust.* **43**, 365-371 (1956).
94. Nijenhuis, L. E. & Hoeven, J. A. v. d. Blood group frequencies in Papuans from Biak (Isles of Schouten). *Vox Sang.* **1**, 241-249 (1956).
95. Lewis, M., Kaita, H. & Chown, B. The blood groups of a Japanese population. *Am. J. Hum. Genet.* **9**, 274-283 (1957).
96. Simmons, R. T. & Graydon, J. J. A blood group genetical survey in Eastern and Central Polynesians. *Am. J. Phys. Anthropol.* **15**, 357-366 (1957).
97. Hulse, F. S. Linguistic barriers to gene-flow: the blood-groups of the Yakima, Okanagon and Swinomish Indians. *Am. J. Phys. Anthropol.* **15**, 235-246 (1957).
98. Bird, G. W. G., Jayaram, T. K., Ikin, E. W., Mourant, A. E. & Lehmann, H. The blood groups and haemoglobin of the Gorkhas of Nepal. *Am. J. Phys. Anthropol.* **15**, 163-169 (1957).
99. Junqueira, P. C., Kalmus, H. & Wishart, P. P.T.C. thresholds, colour vision and blood factors of Brazilian Indians, II: Carajas. *Ann. Hum. Genet.* **22**, 22-25 (1957).
100. Fernandes, J. L. *et al.* P.T.C. thresholds, colour vision and blood factors of Brazilian Indians, I: Kaingangs. *Ann. Hum. Genet.* **22**, 16-21 (1957).
101. Carcassi, U., Ceppellini, R. & Pitzus, F. Frequenza della talassemia in quattro popolazioni sarde e suoi

rapporti con la distribuzione dei gruppi sanguigni e della malaria. *Boll. 1st Sieroter. Milan* **36**, 206-218 (1957).
102. Alberdi, F., Allison, A. C., Blumberg, B. S., Ikin, E. W. & Mourant, A. E. The blood groups of the Spanish Basques. *J. R. Anthrop. Inst.* **87**, 217-221 (1957).
103. Ikin, E. W., Mourant, A. E., Kopec, A. C., Moor-Jankowski, J. K. & Huser, H. J. The blood groups of the western Walsers. *Vox Sang.* **2**, 159-174 (1957).
104. Juel, E. & Vogt, E. The frequency of the Duffy blood group antigens in 1000 Oslo blood donors as defined by anti-Fy. *Acta. Pathol. Microbiol. Scand.* **42**, 150-152 (1958).
105. Graydon, J. J., Semple, N. M., Simmons, R. T. & Franken, S. Blood groups in pygmies of the Wissellakes in Netherlands New Guinea: with anthropological notes by H. J. T. Bijlmer, University of Amsterdam. *Am. J. Phys. Anthropol.* **16**, 149-171 (1958).
106. Aksoy, M., Ikin, E. W., Mourant, A. E. & Lehmann, H. Blood groups, haemoglobins, and thalassaemia in Turks in southern Turkey and Eti-Turks. *Br. Med. J.* **2**, 937-939 (1958).
107. Ruffié, J. Etude séro-anthropologique des populations autochtones du versant nord des Pyrénées. *Bull. Mem. Soc. Anthropol. Paris* **1**, 3-89 (1958).
108. Hackett, W. & Dawson, G. The distribution of the ABO and simple rhesus (D) blood groups in the Republic of Ireland from a sample of 1 in 37 of the adult population. *Ir. J. Med. Sci.* **33**, 99-109 (1958).
109. Staveley, J. M. & Douglas, R. Blood groups in Maoris. *J. Polyn. Soc.* **67**, 239-247 (1958).
110. Matson, G. A. & Swanson, J. Distribution of hereditary blood antigens among the Maya and non-Maya Indians in Mexico and Guatemala. *Am. J. Phys. Anthropol.* **17**, 49-74 (1959).
111. Gershowitz, H. The Diego factor among Asiatic Indians, Apaches and West African Negroes: blood types of Asiatic Indians and Apaches. *Am. J. Phys. Anthropol.* **17**, 195-200 (1959).
112. Corcoran, P. A., Allen, F. H., Jr., Allison, A. C. & Blumberg, B. S. Blood groups of Alaskan Eskimos and Indians. *Am. J. Phys. Anthropol.* **17**, 187-193 (1959).
113. Chown, B. & Lewis, M. The blood group genes of the copper Eskimo. *Am. J. Phys. Anthropol.* **17**, 13-18 (1959).
114. Kout, M. Bestimmung der Blutgruppen A_1A_2BO, MNS, P, Rh/Hr, Kell und Duffy^a in der Prager Bevölkerung. *Blut* **5**, 205-209 (1959).
115. Staveley, J. M. & Douglas, R. Blood groups in Tongans (Polynesia). *J. Polyn. Soc.* **68**, 348-353 (1959).
116. Douglas, R. & Staveley, J. M. The blood groups of Cook Islanders. *J. Polyn. Soc.* **68**, 14-20 (1959).
117. Bories, S. Étude des groupes sanguins ABO, MN, des types Rh, des antigènes Kell et Duffy et de la sicklémie chez les Tahitiens. *Sang* **30**, 237-244 (1959).
118. Simmons, R. T., Gajdusek, D. C. & Larkin, L. C. A blood group genetical survey in New Britain. *Am. J. Phys. Anthropol.* **18**, 101-108 (1960).
119. Hulse, F. S. Ripples on a gene-pool: the shifting frequencies of blood-type alleles among the Indians of the Hupa reservation. *Am. J. Phys. Anthropol.* **18**, 141-152 (1960).
120. Chong Duk Won, Shin, H. S., Kim, S. W., Swanson, J. & Albin Matson, G. Distribution of hereditary blood factors among Koreans residing in Seoul, Korea. *Am. J. Phys. Anthropol.* **18**, 115-124 (1960).
121. Allen, F. H., Jr. & Corcoran, P. A. Blood groups of the Penobscot Indians. *Am. J. Phys. Anthropol.* **18**, 109-114 (1960).
122. Buettner-Janusch, J., Gershowitz, H., Pospisil, L. J. & Wilson, P. Blood groups of selected aboriginal and indigenous populations. *Nature* **188**, 153-154 (1960).
123. Hulse, F. S. & Firestone, M. M. in *2nd Int Congr. Hum. Genet.* 845-847.
124. Matson, G. A. & Swanson, J. Distribution of hereditary blood antigens among American Indians in Middle America: Lacandon and other Maya. *Am. Anthropol.* **63**, 1292-1320 (1961).
125. Lewis, M., Hildes, J. A., Kaita, H. & Chown, B. The blood groups of the Kutchin Indians at Old Crow, Yukon Territory. *Am. J. Phys. Anthropol.* **19**, 383-389 (1961).
126. Tejada, C., Sanchez, M., Guzman, M. A., Bregni, E. & Scrimshaw, N. S. Distribution of blood antigens among Guatemalan Indians. *Hum. Biol.* **33**, 319-334 (1961).
127. Douglas, R., Jacobs, J., Sherliker, J. & Staveley, J. M. Blood groups, serum genetic factors, and haemoglobins in Ellice Islanders. *N. Z. Med. J.* **60**, 259-261 (1961).
128. Cavallini, R. Distribuzione del fattore Duffy nella popolazione del Vercellese. *Trasfus. Sangue* **6**, 17-22 (1961).
129. Vyas, G. N., Bhatia, H. M., Sukumaran, P. K., Balkrishnan, V. & Sanghvi, L. D. Study of blood groups, abnormal hemoglobins and other genetical characters in some tribes of Gujarat. *Am. J. Phys. Anthropol.* **20**, 255-265 (1962).
130. Simmons, R. T., Tindale, N. B. & Birdsell, J. B. A blood group genetical survey in Australian Aborigines

of Bentinck, Mornington and Forsyth Islands, Gulf of Carpentaria. *Am. J. Phys. Anthropol.* **20**, 303-320 (1962).
131. Pollitzer, W. S. Blood types and anthropometry of the Kalmuck Mongols. *Am. J. Phys. Anthropol.* **20**, 11-15 (1962).
132. Layrisse, M., Layrisse, Z. E. & García, J. W. Blood group antigens of the Pemon Indians of Venezuela. *Am. J. Phys. Anthropol.* **20**, 411-420 (1962).
133. Corcoran, P. A., Rabin, D. L. & Allen, F. H., Jr. Blood groups of 237 Navajo school children at Pinon Boarding School, Pinon, Arizona (1961). *Am. J. Phys. Anthropol.* **20**, 389-390 (1962).
134. Best, W. R., Layrisse, M. & Bermejo, R. Blood group antigens in Aymara and Quechua speaking tribes from near Puno, Peru. *Am. J. Phys. Anthropol.* **20**, 321-329 (1962).
135. Silver, R. T., Haber, J. M. & Kellner, A. Blood group studies of jungle Indians of the Mato Grosso. *Transfusion* **2**, 110-114 (1962).
136. Douglas, R., Jacobs, J., Hoult, G. E. & Staveley, J. M. Blood groups, serum genetic factors and hemoglobins in Western Solomon Islanders. *Transfusion* **2**, 413-418 (1962).
137. Bartolo, M. d. Incidenza dei fattori Kell, Cellano e Duffy in un campione di popolazione romana. *Acta. Genet. Med. Gemellol. (Roma)* **12**, 291-297 (1963).
138. Layrisse, M., Layrisse, Z. & Wilbert, J. Blood group antigen studies of four Chibchan Tribes. *Am. Anthropol.* **65**, 36-55 (1963).
139. Cooper, A. J., Blumberg, B. S., Workman, P. L. & McDonough, J. R. Biochemical Polymorphic Traits in a U. S. White and Negro Population. *Am. J. Hum. Genet.* **15**, 420-428 (1963).
140. Matson, G. A. & Swanson, J. Distribution of hereditary blood antigens among Indians in Middle America, V: in Nicaragua. *Am. J. Phys. Anthropol.* **21**, 545-559 (1963).
141. Rodríguez, H., H. Rodríguez, A., Loría, A. & Lisker, R. Studies on several genetic hematological traits of the Mexican population, V: distribution of blood group antigens in Nahuas, Yaquis, Tarahumaras, Tarascos and Mixtecos. *Hum. Biol.* **35**, 350-360 (1963).
142. Johnson, R. H., Ikin, E. W. & Mourant, A. E. Blood groups of the Ait Haddidu Berbers of Morocco. *Hum. Biol.* **35**, 514-523 (1963).
143. Ellis, F. R., Cawley, L. P. & Lasker, G. W. Blood groups, hemoglobin types, and secretion of group-specific substance at Hacienda Cayalti, North Peru. *Hum. Biol.* **35**, 26-52 (1963).
144. Nakajima, H., Urano, M. & Jarumilinta, A. The ABO, MNSs, Q, Lewis, Rh, Kell, Duffy, Lutheran and Kidd blood groups of the Thais. *J. Anthrop. Soc. Nippon* **71**, 109-116 (1963).
145. Wickremasinghe, R. L., Ikin, E. W., Mourant, A. E. & Lehmann, H. The blood groups and haemoglobins of the Veddahs of Ceylon. *J. R. Anthrop. Inst.* **93**, 117-125 (1963).
146. Nijenhuis, L. E. Blood group frequencies and haemoglobin types in Tibetans and Nepalese. *Vox Sang.* **8**, 622-626 (1963).
147. Thomas, J. W. *et al.* Blood groups of the Haida Indians. *Am. J. Phys. Anthropol.* **22**, 189-192 (1964).
148. Salzano, F. M. Blood groups of Indians from Santa Catarina, Brazil. *Am. J. Phys. Anthropol.* **22**, 91-106 (1964).
149. Matson, G. A. & Swanson, J. Distribution of hereditary blood antigens among Indians in Middle America, VI: in British Honduras. *Am. J. Phys. Anthropol.* **22**, 271-284 (1964).
150. Buettner Janusch, J., Bove, J. R. & Young, N. Genetic traits and problems of biological parenthood in two Peruvian Indian tribes. *Am. J. Phys. Anthropol.* **22**, 149-154 (1964).
151. Fraser, G. R., Giblett, E. R., Stransky, E. & Motulsky, A. G. Blood groups in the Philippines. *J. Med. Genet.* **38**, 107-109 (1964).
152. Douglas, R., Jacobs, J., Greenbough, R. & Staveley, J. M. Blood groups, serum genetic factors, and hemoglobins in New Hebrides Islanders. *Transfusion* **4**, 177-184 (1964).
153. Anand, S. Filariasis in its relation to A_1A_2BO, MN, Kell, Duffy and Rhesus blood groups and secretor factor. *Acta. Genet. Med. Gemellol. (Roma)* **14**, 326-335 (1965).
154. Gawrzewski, W. & Kalczew, J. Studies on the Duffy blood group system (Fya) in the population of Cracow (Poland). *Acta. Med. Pol.* **6**, 255-256 (1965).
155. Nicholls, E. M., Lewis, H. B. M., Cooper, D. W. & Bennett, J. H. Blood group and serum protein differences in some Central Australian Aborigines. *Am. J. Hum. Genet.* **17**, 293-307 (1965).
156. Matson, G. A. & Swanson, J. Distribution of hereditary blood antigens among Indians in Middle America, VIII: In Panama. *Am. J. Phys. Anthropol.* **23**, 413-426 (1965).
157. Matson, G. A. & Swanson, J. Distribution of hereditary blood antigens among Indians in Middle America, VII: in Costa Rica. *Am. J. Phys. Anthropol.* **23**, 107-121 (1965).
158. Barnicot, N. A., Krimoas, C., McConnell, R. B. & Beaven, G. H. A genetical survey of Sphakhia, Crete.

Hum. Biol. **37**, 274-298. (1965).
159. Fraser, G. R., Giblett, E. R., Lee, T. C. & Motulsky, A. G. Blood and serum groups in Taiwan. *J. Med. Genet.* **42**, 21-23 (1965).
160. Lee, S. Y. Further analysis of Korean blood types. *Yonsei Med. J.* **6**, 16-25 (1965).
161. Nilsson, L. A., Ryttinger, L. & Tibblin, G. Distribution of the ABO, MN, Rh, Duffy and Kell blood groups in a random sample of Swedish men aged fifty. *Acta. Pathol. Microbiol. Scand.* **68**, 117-122 (1966).
162. Walker, M., Allen, F. H., Jr. & Newman, M. T. Blood groups of Cakchiquel Indians from Sumpango, Guatemala. *Am. J. Phys. Anthropol.* **25**, 239-242 (1966).
163. Plato, C. C., Rucknagel, D. L. & Kurland, L. T. Blood group investigations on the Carolinians and Chamorros of Saipan. *Am. J. Phys. Anthropol.* **24**, 147-154 (1966).
164. Matson, G. A., Swanson, J. & Robinson, A. Distribution of hereditary blood groups among Indians in South America, 3: in Bolivia. *Am. J. Phys. Anthropol.* **25**, 13-33 (1966).
165. Matson, G. A., Sutton, H. E., Swanson, J. & Robinson, A. Distribution of hereditary blood groups among Indians in South America, II: in Peru. *Am. J. Phys. Anthropol.* **24**, 325-349 (1966).
166. Walsh, R. J., Murrell, T. G. & Bradley, M. A. A medical and blood group survey of the Lake Kopiago natives. *Archaeol. Phys. Anthrop. Oceania* **1**, 57-66 (1966).
167. Giles, E., Ogan, E., Walsh, R. J. & Bradley, M. A. Blood group genetics of natives of the Morobe District and Bougainville, Territory of New Guinea. *Archaeol. Phys. Anthrop. Oceania* **1**, 135-154 (1966).
168. Sunderland, E. & Smith, H. M. The blood groups of the Shi'a in Yazd, Central Iran. *Hum. Biol.* **38**, 50-59. (1966).
169. Maranjian, G., Ikin, E. W., Mourant, A. E. & Lehmann, H. The blood groups and haemoglobins of the Saudi Arabians. *Hum. Biol.* **38**, 394-420 (1966).
170. Plato, C. H. C. & Cruz, M. Blood group and haptoglobin frequencies of the Trukese of Micronesia. *Hum. Hered.* **16**, 74-83 (1966).
171. Roberts, D. F., Evans, M., Ikin, E. W. & Mourant, A. E. Blood groups and the affinities of the Canary Islanders. *Man* **1**, 512-525 (1966).
172. Lister, R. W. *et al.* The blood groups and haemoglobin of the bedouin of Socotra. *Man* **1**, 82-86 (1966).
173. Simmons, R. T., Kidson, C., Gorman, J. G. & Rutgers, C. F. Blood group genetic studies in the Tolai and Sulka of New Britain. *Med. J. Aust.* **2**, 747-751 (1966).
174. Ikemoto, S., Watanabe, S., Ogawa, R. & Furuhata, T. Frequencies of blood groups among the Vietnamese. *Proc. Jpn. Acad.* **42**, 975-979 (1966).
175. Douglas, R., Jacobs, J., McCarthy, D. D. & Staveley, J. M. Blood group, serum genetic factors, and hemoglobins in Cook Islanders, I: Atiu Island. *Transfusion* **6**, 319-323 (1966).
176. Vetter, O. & Wegner, H. A further case of anti-Fyb and the frequency of Duffy-antigens in the population of the city of Leipzig. *Acta. Genet. Stat. Med.* **17**, 338-340 (1967).
177. Das, S. R., Mukherjee, D. P. & Bhattacharjee, P. N. Survey of the blood groups and PTC taste among the Rajbanshi caste of West Bengal (ABO, MNS, Rh, Duffy and Diego). *Acta. Genet. Stat. Med.* **17**, 433-445 (1967).
178. Gershowitz, H., Junqueira, P. C., Salzano, F. M. & Neel, J. V. Further studies on the Xavante Indians, III: blood groups and ABH-Lea secretor types in the Simoes Lopes and Sao Marcos Xavantes. *Am. J. Hum. Genet.* **19**, 502-513 (1967).
179. Doeblin, T. D. & Mohn, J. F. The blood groups of the Seneca Indians. *Am. J. Hum. Genet.* **19**, 700-712 (1967).
180. Woodd-Walker, R. B., Smith, H. M. & Clarke, V. A. The blood groups of the Timuri and related tribes in Afghanistan. *Am. J. Phys. Anthropol.* **27**, 195-204 (1967).
181. Pollitzer, W. S., Phelps, D. S., Waggoner, R. E. & Leyshon, W. C. Catawba Indians: morphology, genetics, and history. *Am. J. Phys. Anthropol.* **26**, 5-14 (1967).
182. Matson, G. A., Sutton, H. E., Etcheverry, R. B., Swanson, J. & Robinson, A. Distribution of hereditary blood groups among Indians in South America, IV: in Chile with inferences concerning genetic connections between Polynesia and America. *Am. J. Phys. Anthropol.* **27**, 157-193 (1967).
183. Cordova, M. S., Lisker, R. & Loria, A. Studies on several genetic hematological traits of the Mexican population, XII: Distribution of blood group antigens in twelve indian tribes. *Am. J. Phys. Anthropol.* **26**, 55-65 (1967).
184. Montenegro, L. Blood groups in Tucano Indians. *Hum. Biol.* **39**, 89-92 (1967).
185. McKusick, V. A., Bias, W. B., Norum, R. A. & Cross, H. E. Blood groups in two Amish demes. *Humangenetik* **5**, 36-41 (1967).
186. Robinson, J. C., Levene, C., Blumberg, B. S. & Pierce, J. E. Serum alkaline phosphatase types in North

American indians and negroes. *J. Med. Genet.* **4**, 96-101 (1967).
187. Nakajima, H. *et al.* The distribution of several serological and biochemical traits in East Asia, II: the distribution of ABO, MNSs, Q, Lewis, Rh, Kell, Duffy and Kidd blood groups in Ryukyu. *Jpn. J. Hum. Genet.* **12**, 29-37 (1967).
188. Simmons, R. T., Gajdusek, D. C., Gorman, J. G., Kidson, C. & Hornabrook, R. W. Presence of the Duffy blood group gene Fy-b demonstrated in Melanesians. *Nature* **213**, 1148-1149 (1967).
189. Etcheverry, R. *et al.* Blood groups in Indians of Tierra del Fuego. *Nature* **214**, 211-212 (1967).
190. Etcheverry, R. Blood groups in natives of Easter Island. *Nature* **216**, 690-691 (1967).
191. Chaudhuri, S., Mukherjee, B., Ghosh, J. & Roychoudhury, A. K. Blood groups of the Chinese in Calcutta. *Nature* **213**, 1245-1245 (1967).
192. Chandanayingyong, D., Sasaki, T. T. & Greenwalt, T. J. Blood groups of the Thais. *Transfusion* **7**, 269-276 (1967).
193. Matson, G. A. *et al.* Distribution of hereditary factors in the blood of Indians of the Gila River, Arizona. *Am. J. Phys. Anthropol.* **29**, 311-337 (1968).
194. Mourant, A. E. *et al.* The hereditary blood factors of some populations in Bhutan. *Anthropologist* **special vol.**, 29-43 (1968).
195. Simmons, R. T., Graydon, J. J., Curtain, C. C. & Baumgarten, A. Blood group genetic studies in Laiagam. and Mt. Hagen (lepers), New Guinea. *Archaeol. Phys. Anthrop. Oceania* **3**, 49-54 (1968).
196. Kariks, J. & Walsh, R. J. Some physical measurements and blood groups of the Bainings in New Britain. *Archaeol. Phys. Anthrop. Oceania* **3**, 129-142 (1968).
197. Booth, P. B. & Vines, A. P. Blood groups and other genetic data from the Bismarck Archipelago, New Guinea. *Archaeol. Phys. Anthrop. Oceania* **3**, 64-73 (1968).
198. Booth, P. B. & Oraka, R. E. Blood group frequencies along the South Coast of Papua. *Archaeol. Phys. Anthrop. Oceania* **3**, 146-155 (1968).
199. Johnston, F. E. *et al.* Red cell blood groups of the Peruvian Cashinahua. *Hum. Biol.* **40**, 508-516 (1968).
200. Lechat, M. F. *et al.* A study of various blood group systems in leprosy patients and controls in Cebu, Philippines. *Int. J. Lepr.* **36**, 17-31 (1968).
201. Bhattacharjee, P. N. Further study of Tibetan blood groups. *J. Indian Anthropol. Soc.* **3**, 57-66 (1968).
202. El Hassan, A. M. *et al.* The hereditary blood factors of the Beja of the Sudan. *Man* **3 New Series**, 272-283 (1968).
203. Misawa, S. & Hayashida, Y. On the blood groups among the Ainu in Shizunai, Hokkaido. *Proc. Jpn. Acad.* **44**, 83-88 (1968).
204. Glasgow, B. G. *et al.* The blood groups, serum groups and haemoglobins of the inhabitants of Lunana and Thimbu, Bhutan. *Vox Sang.* **14**, 31-42 (1968).
205. Fraser, G. R. *et al.* Gene frequencies at loci determining blood-group and serum-protein polymorphisms in two villages of northwestern Greece. *Am. J. Hum. Genet.* **21**, 46-60 (1969).
206. Cavalli-Sforza, L. L. *et al.* Studies on African Pygmies, I: a pilot investigation of Babinga Pygmies in the Central African Republic (with an analysis of genetic distances). *Am. J. Hum. Genet.* **21**, 252-274 (1969).
207. Matson, G. A., Sutton, H. E., Swanson, J. & Robinson, A. Distribution of hereditary blood groups among indians in South America, VII: in Argentina. *Am. J. Phys. Anthropol.* **30**, 61-83 (1969).
208. Lisker, R., Cordova, M. S. & Graciela Zarate, Q. B. Studies on several genetic hematological traits of the Mexican population, XVI: Hemoglobin, S and glucose-6-phosphate dehydrogenase deficiency in the east coast. *Am. J. Phys. Anthropol.* **30**, 349-354 (1969).
209. Alfred, B. M., Stout, T. D., Birkbeck, J., Lee, M. & Petrakis, N. L. Blood groups, red cell enzymes, and cerumen types of the Ahousat (Nootka) Indians. *Am. J. Phys. Anthropol.* **31**, 391-398 (1969).
210. Kumar, N. & Mukherjee, D. P. A genetic survey among the Desi Bhumij of Chota Nagpur in Bihar. *Anthropologist* **special vol**, 75-83 (1969).
211. Bhattacharjee, P. N. A genetical study of the Santals of Santal Parganas. *Anthropologist* **special vol**, 93-103 (1969).
212. Simmons, R. T. & Cooke, D. R. Population genetic studies in Australian aborigines of the Northern Territory: blood group genetic studies in the Malag of Elcho Island. *Archaeol. Phys. Anthrop. Oceania* **4**, 252-259 (1969).
213. Salazar-Mallén, M., Amezcua-Chavarría, E. & Mitrani-Levy, D. Estudios sobre la atopía en Mexico: los grupos eritrocitarios y el serológico Gm (1) en la población normal y en la atópica. *Gac. Med. Mex.* **99**, 730-735 (1969).
214. Harvey, R. G., Godber, M. J., Kopec, A. C., Mourant, A. E. & Tills, D. Frequency of genetic traits in the

Caribs of Dominica. *Hum. Biol.* **41**, 342-364 (1969).
215. Bajatzadeh, M. & Walter, H. Investigations on the distribution of blood and serum groups in Iran. *Hum. Biol.* **41**, 401-415 (1969).
216. Staveley, J. M., Tinielu, I. & Douglas, R. Blood groups and serum genetic factors in Tokelau islanders. *N. Z. Med. J.* **70**, 238-242 (1969).
217. Harrison, G. A. *et al.* The effects of altitudinal variation in Ethiopian populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **256**, 147-182 (1969).
218. Matznetter, T. & Spielmann, W. Blutgruppen moçambiquanischer Bantustämme. *Z. Morphol. Anthropol.* **61**, 57-71 (1969).
219. Skov, F., Eriksen, M. & Hagerup, L. Distribution of the ABO, MNS, P, Rhesus, Lutheran, Kell, Lewis and Duffy blood groups and frequency of irregular red cell antibodies in a population of Danes aged fifty years and a population of Danes aged seventy years. From the Glostrup population studies. *Acta. Pathol. Microbiol. Scand. B Microbiol. Immunol.* **78**, 553-559 (1970).
220. Niswander, J. D., Brown, K. S., Iba, B. Y., Leyshon, W. C. & Workman, P. L. Population studies on southwestern Indian tribes, I: History, culture, and genetics of the Papago. *Am. J. Hum. Genet.* **22**, 7-23 (1970).
221. Juberg, R. C. Blood-group gene frequencies in West Virginia. *Am. J. Hum. Genet.* **22**, 96-99 (1970).
222. Gershowitz, H. *et al.* Gene frequencies and microdifferentiation among the Makiritare Indians, I: eleven blood group systems and the ABH-Le secretor traits: a note on Rh gene frequency determinations. *Am. J. Hum. Genet.* **22**, 515-525 (1970).
223. Pollitzer, W. S. *et al.* The Seminole Indians of Florida: morphology and serology. *Am. J. Phys. Anthropol.* **32**, 65-81 (1970).
224. Pollitzer, W. S., Namboodiri, K. K., Elston, R. C., Brown, W. H. & Leyshon, W. C. The Seminole Indians of Oklahoma: morphology and serology. *Am. J. Phys. Anthropol.* **33**, 15-29 (1970).
225. Erickson, R. P., Nerlove, S., Creger, W. P. & Romney, A. K. Comparison of genetic and anthropological interpretations of population isolates in Aguacatenango, Chiapas, Mexico. *Am. J. Phys. Anthropol.* **32**, 105-120 (1970).
226. Alfred, B. M., Stout, T. D., Lee, M., Birkdeck, J. & Petrakis, N. L. Blood groups, phosphoglucomutase, and cerumen types of the Anahan (Chilcotin) Indians. *Am. J. Phys. Anthropol.* **32**, 329-338 (1970).
227. Stanhope, J. M. & Booth, P. B. The Kire people, Madang district, New Guinea: blood groups, haptoglobin and transferring types. *Archaeol. Phys. Anthrop. Oceania* **5**, 157-162 (1970).
228. Bonne, B. *et al.* The Habbanite isolate, I: Genetic markers in the blood. *Hum. Hered.* **20**, 609-622 (1970).
229. Spielmann, W., Teixidor, D., Renninger, W. & Matznetter, T. Blutgruppen und Lepra bei moçambiquanischen Völkerschaften. *Humangenetik* **10**, 304-317 (1970).
230. Cabannes, R. *et al.* Etude hémotypologique et biologique des Attié du village d'Atiékwa. *Med. Afr. Noire* **17**, 835-841 (1970).
231. Cunha, A. X. d. & Cunha, F. A. F. X. d. Grupos sanguineos da populacao Macua da Ilha de Mocambique (Sistema A_1A_2BO, MN, CDE e Duffy). *Rev. Ecuat. Med. Cienc. Biol.* **3**, 97-107 (1970).
232. Bonné, B., Godber, M., Ashbel, S., Mourant, A. E. & Tills, D. South-Sinai Beduin: a preliminary report on their inherited blood factors. *Am. J. Phys. Anthropol.* **34**, 397-408 (1971).
233. Lowe, R. F., Gadd, K. G., Chitiyo, M. E., Emmanuel, J. & Robertson, T. The MN, P, Kell and Duffy blood group systems of the Zezuru tribe of Rhodesia. *Cent. Afr. J. Med.* **17**, 207-209 (1971).
234. Salzano, F. M. *et al.* Blood groups and H-Le a salivary secretion of Brazilian Cayapo Indians. *Am. J. Phys. Anthropol.* **36**, 417-425 (1972).
235. Sorgo, G. & Piso, C. Das System Duffy: Genfrequenzen und Familienuntersuchung. *Blut* **24**, 89-93 (1972).
236. Warwick, R., Raynes, A. E., Ikin, E. W. & Mourant, A. E. The blood groups of the inhabitants of Lipari (Aeolian Islands, Italy). *Hum. Biol.* **44**, 649-654 (1972).
237. Spielmann, W., Teixidor, D. & Matznetter, T. Blutgruppen bei Bantu Populationen aus Angola zugleich ein Beitrag zur Berechnung der Vaterschaftswarscheinlichkeit bei Gutachten mit Negern als Eventualvätern. *Blut* **27**, 322-335 (1973).
238. Hawkins, B., Elliot, M., Kosasih, E. N. & Simons, M. J. Red cell genetic studies of the Toba Bataks of North Sumatra. *Hum. Biol. Oceania* **2**, 147-154 (1973).
239. Hakim, S. M. A., Baxi, A. J. & Balakrishnan, V. Blood groups, secretor status and ability to taste phenylthiocarbamide in some Muslim groups. *Hum. Hered.* **23**, 72-77 (1973).
240. Eriksson, A. W., Eskola, M.-R., Workman, P. L. & Morton, N. E. Population studies on the Aland Islands,

II: historical population structure: inference from bioassay of kinship and migration. *Hum. Hered.* **23**, 511-534 (1973).
241. Arends, T., Gallango, M. L., Muller, A., Gonzalez-Marroro, M. & Perez Bandez, O. in *International Congress on Anthropological and Ethnological Sciences.* 1.
242. Sunderland, E. & Coope, E. Biological studies of Yemenite and Kurdish Jews in Israel and other groups in south-west Asia, XII: genetic studies in Jordan. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **266**, 207-220 (1973).
243. Lehmann, H. *et al.* Biological studies of Yemenite and Kurdish Jews in Israel and other groups in south-west Asia, XI: the hereditary blood factors of the Kurds of Iran. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **266**, 195-205 (1973).
244. Geerdink, R. A., Nijenhuis, L. E., Van Loghem, E. & Li Fo Sjoe, E. Blood groups and immunoglobulin groups in trio and Wajana Indians from Surinam. *Am. J. Hum. Genet.* **26**, 45-53 (1974).
245. Kirk, R. L., McDermid, E. M. & Blake, N. M. Blood group, serum protein and red cell enzyme groups of Amerindian populations in Colombia. *Am. J. Phys. Anthropol.* **41**, 301-316 (1974).
246. Crawford, M. H., Leyshon, W. C. & Brown, K. Human biology in Mexico, II: A comparison of blood group, serum and red cell enzyme frequencies, and genetic distances of the Indian populations of Mexico. *Am. J. Phys. Anthropol.* **41**, 251-268 (1974).
247. Brown, S. M., Gajdusek, D. C. & Leyshon, W. C. Genetic studies in Paraguay: blood group, red cell, and serum genetic patterns of the Guayaki and Ayore Indians, Mennonite settlers, and seven other Indian tribes of the Paraguayan Chaco. *Am. J. Phys. Anthropol.* **41**, 317-343 (1974).
248. Roberts, D. F., Papiha, S. S. & Creen, C. K. Red cell enzyme and other polymorphic systems in Madhya Pradesh, Central India. *Ann. Hum. Biol.* **1**, 159-174 (1974).
249. Mourant, A. E. *et al.* The blood groups and haemoglobins of the Kunama and Baria of Eritrea, Ethiopia. *Ann. Hum. Biol.* **1**, 383-392 (1974).
250. Misawa, S., Ohno, N., Ishimoto, G. & Omoto, K. The distribution of genetic markers in blood samples from Okinawa, the Ryukyus, I: the distribution of red cell antigen groups in Ishigaki Island. *J. Anthrop. Soc. Nippon* **82**, 135-143 (1974).
251. Salmon, D. & Yvart, J. Fréquences géniques de 17 systèmes de polymorphisme: étude sur un échantillon de sujets vivant dans la région parisienne. *Rev. Franc. Trans.* **17**, 295-304 (1974).
252. Nurse, G. T. & Jenkins, T. The Griqua of Campbell, Cape Provice, South Africa. *Am. J. Phys. Anthropol.* **43**, 71-78 (1975).
253. Szathmary, E. J., Mohn, J. F. & Gershowitz, H. The Northern and Southeastern Ojibwa: blood group systems and the causes of genetic divergence. *Hum. Biol.* **47**, 351-368 (1975).
254. Chakraborty, R., Das, S. K. & Roy, M. Blood group genetics of some caste groups of Southern 24 Parganas, West Bengal. *Hum. Hered.* **25**, 218-225 (1975).
255. Nurse, G. T., Lane, A. B. & Jenkins, T. Sero-genetic studies on the Dama of South West Africa. *Ann. Hum. Biol.* **3**, 33-50 (1976).
256. Hiernaux, J. Blood polymorphism frequencies in the Sara Majingay of Chad. *Ann. Hum. Biol.* **3**, 127-140 (1976).
257. Godber, M. *et al.* The blood groups, serum groups, red-cell isoenzymes and haemoglobins of the Sandawe and Nyaturu of Tanzania. *Ann. Hum. Biol.* **3**, 463-473 (1976).
258. Cartwright, R. A. Unifactorially inherited attributes of the population of Holy Island, Northumberland. *Ann. Hum. Biol.* **3**, 351-362 (1976).
259. Basu, A. *et al.* Morphology, serology, dermatoglyphics, and microevolution of some village populations in Haiti, West Indies. *Hum. Biol.* **48**, 245-269 (1976).
260. Kojić, T. *et al.* Possible genetic predisposition for alcohol addiction. *Adv. Exp. Med. Biol.* **85**, 7-24 (1977).
261. Salzano, F. M., Neel, J. V., Gershowitz, H. & Migliazza, E. C. Intra and intertribal genetic variation within a linguistic group: the Ge-speaking indians of Brazil. *Am. J. Phys. Anthropol.* **47**, 337-347 (1977).
262. Tills, D., Teesdale, P. & Mourant, A. E. Blood groups of the Irish. *Ann. Hum. Biol.* **4**, 23-24 (1977).
263. Nurse, G. T. & Jenkins, T. Serogenetic studies on the Kavango peoples of South West Africa. *Ann. Hum. Biol.* **4**, 465-478 (1977).
264. Hiorns, R. W., Harrison, G. A. & Gibson, J. B. Genetic variation in some Oxfordshire villages. *Ann. Hum. Biol.* **4**, 197-210 (1977).
265. Levine, M. H., Von Hagen, V., Ruffie, J. & Darrasse, H. A hematological approach to Basque isolation in two French Basque Villages. *Ann. N. Y. Acad. Sci.* **293**, 185-193 (1977).
266. Neel, J. V. *et al.* Genetic studies of the Macushi and Wapishana Indians, II: data on 12 genetic

polymorphisms of the red cell and serum proteins: gene flor between the tribes. *Hum. Genet.* **37**, 207-219 (1977).
267. Nurse, G. T., Botha, M. C. & Jenkins, T. Sero-genetic studies on the San of South West Africa. *Hum. Hered.* **27**, 81-98 (1977).
268. Colino, F., Campillo, F. L., Toledano, J. A. & Senra, A. Distribución de los grupos del sistema sanguíneo Duffy en la población española. *Sangre (Barc)* **22**, 30-32 (1977).
269. Welch, S. G., McGregor, I. A. & Williams, K. The Duffy blood group and malaria prevalence in Gambian West Africans. *Trans. R. Soc. Trop. Med. Hyg.* **71**, 295-296 (1977).
270. Papiha, S. S., Roberts, D. F., Mukerjee, D. P., Singh, S. D. & Malhotra, M. A genetic survey in the Bhil tribe of Madhya Pradesh, Central India. *Am. J. Phys. Anthropol.* **49**, 179-185 (1978).
271. Spencer, H. C. *et al.* The Duffy blood group and resistance to *Plasmodium vivax* in Honduras. *Am. J. Trop. Med. Hyg.* **27**, 664-670 (1978).
272. Roberts, D. F. & Papiha, S. S. Les polymorphismes génétiques des Sukuma (Tanzanie). *Anthropologie* **82**, 565-574 (1978).
273. Salzano, F. M. *et al.* Unusual blood genetic characteristics among the Ayoreo Indians of Bolivia and Paraguay. *Hum. Biol. Oceania* **50**, 121-136 (1978).
274. Beckman, L., Cedergren, B., Perris, C. & Strandman, E. Blood groups and affective disorders. *Hum. Hered.* **28**, 48-55 (1978).
275. Sukernik, R. I., Karaphet, T. M. & Osipova, L. P. Distribution of blood groups, serum markers and red cell enzymes in two human populations from Northern Siberia. *Hum. Hered.* **28**, 321-327 (1978).
276. Le Petit, J. C. *et al.* Analyse génétique d'une population présentant des anomalies de fréquence génique. *Rev. Fr. Transfus. Immunohematol.* **21**, 921-933 (1978).
277. Nurse, G. T., Jenkins, T., David, J. H. & Steinberg, A. G. The Njinga of Angola: a serogenetic study. *Ann. Hum. Biol.* **6**, 337-348 (1979).
278. Beaumont, B., Nurse, G. T. & Jenkins, T. Highland and lowland populations of Lesotho. *Hum. Hered.* **29**, 42-49 (1979).
279. Martin, S. K. *et al.* Frequency of blood group antigens in Nigerian children with falciparum malaria. *Trans. R. Soc. Trop. Med. Hyg.* **73**, 216-218 (1979).
280. Sandler, S. G. *et al.* The Duffy blood group system in Israeli Jews and Arabs. *Vox Sang.* **37**, 41-46 (1979).
281. Salzano, F. M. *et al.* The Caingang revisited: blood genetics and anthropometry. *Am. J. Phys. Anthropol.* **53**, 513-524 (1980).
282. Saha, N. *et al.* Some blood genetic markers of selected tribes in western Saudi Arabia. *Am. J. Phys. Anthropol.* **52**, 595-600 (1980).
283. Kamel, K., Chandy, R., Mousa, H. & Yunis, D. Blood groups and types, hemoglobin variants, and G-6-PD deficiency among Abu Dhabians in the United Arab Emirates. *Am. J. Phys. Anthropol.* **52**, 481-484 (1980).
284. Black, F. L. *et al.* Restriction and persistence of polymorphisms of HLA and other blood genetic traits in the Parakana Indians of Brazil. *Am. J. Phys. Anthropol.* **52**, 119-132 (1980).
285. Wasfi, A. I., Saha, N., El Munshid, H. A., El Sheikh, F. S. & Ahmed, M. A. Genetic association in vitiligo: ABO, MNSs, Rhesus, Kell and Duffy blood groups. *Clin. Genet.* **17**, 415-417 (1980).
286. Armanet, L. *et al.* Frecuencia de los sistemas Rh, MNSs, Duffy, Diego, Kell, Lutheran y Xg en Chile. *Rev. Med. Chil.* **108**, 103-108 (1980).
287. Sukernik, R. I., Lemza, S. V., Karaphet, T. M. & Osipova, L. P. Reindeer Chukchi and Siberian Eskimos: studies on blood groups, serum proteins, and red cell enzymes with regard to genetic heterogeneity. *Am. J. Phys. Anthropol.* **55**, 121-128 (1981).
288. Ferrell, R. E., Chakraborty, R., Gershowitz, H., Laughlin, W. S. & Schull, W. J. The St. Lawrence Island Eskimos: genetic variation and genetic distance. *Am. J. Phys. Anthropol.* **55**, 351-358 (1981).
289. Crawford, M. H., Mielke, J. H., Devor, E. J., Dykes, D. D. & Polesky, H. F. Population structure of Alaskan and Siberian indigenous communities. *Am. J. Phys. Anthropol.* **55**, 167-185 (1981).
290. Buckton, K., Lai, L. Y. C. & Gibson, J. B. A search for association between gene markers and serum cholesterol, triglyceride, urate and blood pressure. *Ann. Hum. Biol.* **8**, 39-48 (1981).
291. Spedini, G. *et al.* Some genetic erythrocyte polymorphisms in the Mbugu and other populations of the Central African Republic with an analysis of genetic distances. *Anthropol. Anz.* **39**, 10-19 (1981).
292. Mourant, A. E. *et al.* Red cell antigen, serum protein, and red cell enzyme polymorphisms in inhabitants of the Jimi Valley, Western Highlands, New Guinea. *Hum. Genet.* **59**, 77-80 (1981).
293. Colauto, E. M. *et al.* Malária no município de Humaitá, estado do Amazonas, XII: freqüência de fatores

de resistência eritrocitária na população geral e em doentes: hemoglobina S e sistema sangüíneo Duffy. *Rev. Inst. Med. Trop. Sao Paulo* **23**, 72-78 (1981).
294. Barrantes, R., Smouse, P. E., Neel, J. V., Mohrenweiser, H. W. & Gershowitz, H. Migration and genetic infrastructure of the central American Guaymi and their affinities with other tribal groups. *Am. J. Phys. Anthropol.* **58**, 201-214 (1982).
295. Tills, D., Warlow, A., Kopec, A. C., Fridriksson, S. & Mourant, A. E. The blood groups and other hereditary blood factors of the Icelanders. *Ann. Hum. Biol.* **9**, 507-520 (1982).
296. Tills, D. *et al.* Blood group, protein, and red cell enzyme polymorphisms of the Hadza of Tanzania. *Hum. Genet.* **61**, 52-59 (1982).
297. Breguet, G. *et al.* Genetic survey of an isolated community in Bali, Indonesia, I: blood groups, serum proteins and hepatitis B serology. *Hum. Hered.* **32**, 52-61 (1982).
298. Booth, P. B. *et al.* Red cell antigen, serum protein and red cell enzyme polymorphisms in Karkar Islanders and inhabitants of the adjacent north coast of New Guinea. *Hum. Hered.* **32**, 385-403 (1982).
299. Rouger, P., Ruffie, J., Gueguen, A., Golmard, J. L. & Salmon, D. Human-blood groups of the Chinese population of Macau, I: blood-groups Abo, Rhesus, Mnss, Kidd, Duffy and Diego. *J. Hum. Evol.* **11**, 481-486 (1982).
300. Soustelle, B., Rubio, F., Coves, F. & Bizot, M. Phénotypage Kell et Duffy par une technique de coagglutination sur Groupmatic G 360. *Rev. Fr. Transfus. Immunohematol.* **25**, 309-319 (1982).
301. Szathmary, E. J. E. Dogrib Indians of the Northwest Territories, Canada: genetic diversity and genetic relationship among subarctic Indians. *Ann. Hum. Biol.* **10**, 147-162 (1983).
302. Woolley, V., Gill, P. S. & Sunderland, E. Blood groups, haptoglobins and red cell isoenzymes of the Jat Sikhs of Ludhiana District, Panjab, India. *Hum. Hered.* **33**, 44-51 (1983).
303. Wolanski, N., Nahar, R. A. & Roberts, D. F. Genetic studies in Poland. *Hum. Hered.* **33**, 270-276 (1983).
304. Cedergren, B., Nordenson, I. & Beckman, L. Population studies in northern Sweden, XI: the Duffy blood group polymorphism. *Hum. Hered.* **33**, 365-370 (1983).
305. Schliwa, R., Gilbert, K., Walter, H. & Dannewitz, A. Serological-genetic investigations on some populations of the northern Aegean Sea (Greece). *J. Hum. Evol.* **12**, 769-773 (1983).
306. Youinou, P. *et al.* An analysis of the blood group composition of a population in Brittany, the Bigoudens. *Rev. Fr. Transfus. Immunohematol.* **26**, 359-368 (1983).
307. Paul, B. Duffy blood group distribution in Malawi. *Trans. R. Soc. Trop. Med. Hyg.* **77**, 877 (1983).
308. Rothhammer, F., Goedde, H. W., Llop, E., Acuna, M. & Carvajal, P. Erythrocyte and HLA antigens of Atacameno Indians. *Am. J. Phys. Anthropol.* **65**, 243-247 (1984).
309. Yida, Y. *et al.* Distribution of eight blood-group systems and ABH secretion in Mongolian, Korean, and Zhuang nationalities in China. *Ann. Hum. Biol.* **11**, 377-388 (1984).
310. Sawhney, K. S., Sunderland, E. & Woolley, V. Genetic polymorphisms in the Kuwaiti Arabs. *Hum. Hered.* **34**, 303-307 (1984).
311. Ohkura, K. *et al.* Distribution of polymorphic traits in Mazandaranian and Guilanian in Iran. *Hum. Hered.* **34**, 27-39 (1984).
312. Salzano, F. M. Incidence, effects, and management of sickle cell disease in Brazil. *Am. J. Pediatr. Hematol. Oncol.* **7**, 240-244 (1985).
313. Ferrell, R. E., Salamatina, N. V., Dalakishvili, S. M., Bakuradze, N. A. & Chakraborty, R. A population genetic study in the Ochamchir region, Abkhazia, USSR. *Am. J. Phys. Anthropol.* **66**, 63-71 (1985).
314. Salzano, F. M. *et al.* Population structure and blood genetics of the Pacaas Novos Indians of Brazil. *Ann. Hum. Biol.* **12**, 241-249 (1985).
315. Clegg, E. J., Tills, D., Warlow, A., Wilkinson, J. & Marin, A. Blood group variation in the Isle of Lewis. *Ann. Hum. Biol.* **12**, 345-361 (1985).
316. Swart, J. & Pribilla, O. Das Duffy-System: eine populationsgenetische Untersuchung in Portugal, Guinea-Bissau und Brasilien. *Anthropol. Anz.* **43**, 285-297 (1985).
317. Kulkarni, A. G., Ibazebo, R. O., Dunn, D. T. & Fleming, A. F. Some red cell antigens in the Hausa population of northern Nigeria. *Hum. Hered.* **35**, 283-287 (1985).
318. Salzano, F. M. *et al.* Demography and genetics of the Satere-Mawe and their bearing on the differentiation of the Tupi tribes of South America. *J. Hum. Evol.* **14**, 647-655 (1985).
319. Bestetti, A., Cazzaniga, G., Ripamonti, M. & Assali, G. Risultati della tipizzazione antigenica eritrocitaria in un campione della popolazione del circondario di Monza, I: sistemi diallelici eritrocitari Kell, Lutheran, Duffy, Kidd, P. *Riv. Emoter. Immunoematol.* **32**, 17-40 (1985).
320. Singh, K. S. *et al.* Genetic markers among Meiteis and Brahmins of Manipur, India. *Hum. Hered.* **36**, 177-187 (1986).

321. Mitchell, R. J., Kosten, M. & Deacon, M. Population genetics of blood group polymorphisms in a sample of newborns from Melbourne, Australia. *Hum. Hered.* **36**, 310-316 (1986).
322. Harvey, R. G., Smith, M. T., Sherren, S., Bailey, L. & Hyndman, S. J. How Celtic are the Cornish? A study of biological affinities. *Man* **21**, 177-201 (1986).
323. Mahmoud, L. A., Ibrahim, A. A., Ghonem, H. R. & Jouvenceaux, A. Human blood groups in Dakahlya, Egypt. *Ann. Hum. Biol.* **14**, 487-493 (1987).
324. Jenkins, T., Speirs, J., Dunn, D. S. & Nurse, G. T. Serogenetic and haematological studies on the Kgalagadi of Botswana. *Ann. Hum. Biol.* **14**, 143-153 (1987).
325. Jin, F., Hao, L. & Du, R. Distribution of red cell blood group systems in Bai and Hani in China. *Gene Geogr.* **1**, 163-168 (1987).
326. Ai, Q., Yuan, Y., Zhao, H., Li, S. & Du, R. Distribution of red cell blood group systems in Yi, Tibetan and Manchu ethnic groups in China. *Gene Geogr.* **1**, 169-176 (1987).
327. Sistonen, P., Koistinen, J. & Aden Abdulle, O. Distribution of blood groups in the East African Somali population. *Hum. Hered.* **37**, 300-313 (1987).
328. Salzano, F. M. *et al.* Genetic variation within a linguistic group: Apalai-Wayana and other Carib tribes. *Am. J. Phys. Anthropol.* **75**, 347-356 (1988).
329. Nevo, S. Genetic blood markers in Arab Druze of Israel. *Am. J. Phys. Anthropol.* **77**, 183-190 (1988).
330. Lisker, R., Pérez-Briceno, R., Granados, J. & Babinsky, V. Gene frequencies and admixture estimates in the state of Puebla, Mexico. *Am. J. Phys. Anthropol.* **76**, 331-335 (1988).
331. Aireche, H. & Benabadji, M. Rh and Duffy gene frequencies in Algeria. *Gene Geogr.* **2**, 1-8 (1988).
332. Hoang, B. *et al.* An analysis of the blood group composition of the North Vietnam population. *Int. J. Anthropol.* **3**, 63-70 (1988).
333. Iyengar, S. *et al.* Genetic studies of type 2 (non-insulin-dependent) diabetes mellitus: lack of association with seven genetic markers. *Diabetologia* **32**, 690-693 (1989).
334. Yung, C. H. *et al.* Blood group phenotypes in Taiwan. *Transfusion* **29**, 233-235 (1989).
335. Schliwa, R., Dannewitz, A., Gilbert, K. & Walter, H. Genetic studies in four populations of the northern Aegean Sea, Greece. *Z. Morphol. Anthropol.* **78**, 89-106 (1989).
336. Saha, N., Tay, J. S., Tsoi, W. F. & Kua, E. H. Association of Duffy blood group with schizophrenia in Chinese. *Genet. Epidemiol.* **7**, 303-305 (1990).
337. Nasidze, I. S. *et al.* Genetika narodonaseleniia Kavkaza: raspredelenie nekotorykh immunologicheskikh i biokhimicheskikh markerov v Vostochnoi Gruzii. *Genetika* **26**, 936-945 (1990).
338. Insaridze, Z. P. *et al.* Genetika narodonaseleniia Kavkaza: raspredelenie nekotorykh immunologicheskikh i biokhimicheskikh markerov v Zapadnoi Gruzii. *Genetika* **26**, 1092-1101 (1990).
339. Lisker, R., Ramirez, E., Briceño, R. P., Granados, J. & Babinsky, V. Gene frequencies and admixture estimates in four Mexican urban centers. *Hum. Biol.* **62**, 791-801 (1990).
340. Picornell, A., Castro, J. A. & Ramon, M. M. Blood groups in the Chueta community (Majorcan Jews). *Hum. Hered.* **41**, 35-42 (1991).
341. Daher, V., Youlton, R., Nazer, J. & Cifuentes, L. Estudio genetico en gemelos. *Rev. Chil. Pediatr.* **62**, 23-28 (1991).
342. Acuña, M. *et al.* Evidencias genéticas corroboran la hipótesis de Neghme sobre la mayor benignidad de la tripanosomiasis Americana en Chile. *Rev. Med. Chil.* **120**, 233-238 (1992).
340. Walter, H. *et al.* Investigations on the variability of blood group polymorphisms among sixteen tribal populations from Orissa, Madhya Pradesh and Maharashtra, India. *Z. Morphol. Anthropol.* **79**, 69-94 (1992).
344. Salamatina, N. M. & Nasidze, I. S. Genetic polymorphisms in a rural Osetian community. *Gene Geogr.* **7**, 251-255 (1993).
345. Sans, M. *et al.* Blood-Group Frequencies and the Question of Race Admixture in Uruguay. *Interciencia* **18**, 29-32 (1993).
346. Verma, I. C. & Thakur, A. Duffy Blood-Group Determinants and Malaria in India. *J. Genet.* **72**, 15-19 (1993).
347. Hao, L., Liu, J., Li, M., Jin, S. & Du, R. Distribution of red cell blood groups in Han subpopulations of Inner Mongolia, Gansu and Henan Provinces, China. *Gene Geogr.* **8**, 175-184 (1994).
348. Brindle, P. M., Maitland, K., Williams, T. N. & Ganczakowski, M. E. A survey for the rare blood group antigen variants, En(a-), Gerbich negative and Duffy negative on Espiritu Santo, Vanuatu in the South Pacific. *Hum. Hered.* **45**, 211-214 (1995).
349. Callegari-Jacques, S. M. *et al.* The Wai Wai Indians of South America: history and genetics. *Ann. Hum. Biol.* **23**, 189-201 (1996).

350. Janicijevic, B., Bakran, M., Martinovic, I. & Roberts, D. F. Serogenetic polymorphisms of the four middle Dalmatian island and peninsular population isolates. *Coll. Antropol.* **20**, 47-54 (1996).
351. Pinto, F., Rando, J. C., Lopez, M., Morilla, J. M. & Larruga, J. M. Blood group polymorphisms in the Canary Islands. *Gene Geogr.* **10**, 171-179 (1996).
352. Picornell, A. *et al.* Genetic variation in the population of Ibiza (Spain): genetic structure, geography, and language. *Hum. Biol.* **68**, 899-913 (1996).
353. Salzano, F. M. *et al.* The Brazilian Xavante Indians revisited: new protein genetic studies. *Am. J. Phys. Anthropol.* **104**, 23-34 (1997).
354. Schnackenberg, L., Flesch, B. K. & Neppert, J. Linkage disequilibria between Duffy blood groups, Fc gamma IIa and Fc gamma IIIb allotypes. *Exp. Clin. Immunogenet.* **14**, 235-242 (1997).
355. Nabulsi, A. J., Cleve, H. & Rodewald, A. Serological analysis of the Abbad tribe of Jordan. *Hum. Biol.* **69**, 357-373 (1997).
356. Nanu, A. & Thapliyal, R. M. Blood group gene frequency in a selected north Indian population. *Indian J. Med. Res.* **106**, 242-246 (1997).
357. Santos, S. E. *et al.* New protein genetic studies in six Amazonian Indian populations. *Ann. Hum. Biol.* **25**, 505-522 (1998).
358. Harb, Z., Llop, E., Moreno, R. & Quiroz, D. Poblaciones costeras de Chile: marcadores genéticos en cuatro localidades. *Rev. Med. Chil.* **126**, 753-760 (1998).
359. Olsson, M. L. *et al.* A clinically applicable method for determining the three major alleles at the Duffy (FY) blood group locus using polymerase chain reaction with allele-specific primers. *Transfusion* **38**, 168-173 (1998).
360. Fernández-Santander, A. *et al.* Genetic relationships between southeastern Spain and Morocco: new data on ABO, RH, MNSs, and DUFFY polymorphisms. *Am. J. Hum. Biol.* **11**, 745-752 (1999).
361. Cerda-Flores, R. M., Barton, S. A., Marty-Gonzalez, L. F., Rivas, F. & Chakraborty, R. Estimation of nonpaternity in the Mexican population of Nuevo Leon: a validation study with blood group markers. *Am. J. Phys. Anthropol.* **109**, 281-293 (1999).
362. Hamblin, M. T. & Di Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.* **66**, 1669-1679 (2000).
363. Luiselli, D., Simoni, L., Tarazona-Santos, E., Pastor, S. & Pettener, D. Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am. J. Phys. Anthropol.* **113**, 5-17 (2000).
364. Shimizu, Y. *et al.* Sero- and molecular typing of Duffy blood group in Southeast Asians and Oceanians. *Hum. Biol.* **72**, 511-518 (2000).
365. Novaretti, M. C. Z., Dorlhiac-Llacer, P. E. & Chamone, D. A. F. Estudo de grupos sangüíneos em doadores de sangue caucasóides e negróides na cidade de São Paulo. *Rev. Bras. Hematol. Hemoter.* **22**, 23-32 (2000).
366. Goicoechea, A. S. *et al.* New genetic data on Amerindians from the Paraguayan Chaco. *Am. J. Hum. Biol.* **13**, 660-667 (2001).
367. Azofeifa, J., Ruiz, E. & Barrantes, R. Blood group, red cell, and serum protein variation in the Cabecar and Huetar, two Chibchan Amerindian tribes of Costa Rica. *Am. J. Hum. Biol.* **13**, 57-64 (2001).
368. Diedrich, B., Andersson, J., Sallander, S. & Shanwell, A. K, Fy(a), and Jk(a) phenotyping of donor RBCs on microplates. *Transfusion* **41**, 1263-1267 (2001).
369. Yan, L., Fu, Q., Jin, L. & Li, L. Duffy blood group phenotypes and genotypes in Chinese. *Transfusion* **41**, 970 (2001).
370. Velzing-Aarts, F. V., Muskiet, F. A., van der Dijs, F. P. & Duits, A. J. High serum interleukin-8 levels in Afro-Caribbean women with pre-eclampsia: relations with tumor necrosis factor-alpha, Duffy negative phenotype and von Willebrand factor. *Am. J. Reprod. Immunol.* **48**, 319-322 (2002).
371. Harich, N. *et al.* Classical polymorphisms in Berbers from Moyen Atlas (Morocco): genetics, geography, and historical evidence in the Mediterranean peoples. *Ann. Hum. Biol.* **29**, 473-487 (2002).
372. Weir-Medina, J. *et al.* Erythrocyte antigens and their relation to Bipolar Disease: a preliminary study. *Hum. Hered.* **43**, 25-34 (2002).
373. Morera, B., Barrantes, R. & Marin-Rojas, R. Gene admixture in the Costa Rican population. *Ann. Hum. Genet.* **67**, 71-80 (2003).
374. Bauduer, F. *et al.* Duffy blood group genotyping in French Basques using polymerase chain reaction with allele-specific primers (PCR-ASP). *Am. J. Hum. Biol.* **16**, 78-81 (2004).
375. Chiaroni, J. *et al.* Genetic characterization of the population of Grande Comore Island (Njazidja) according to major blood groups. *Hum. Biol.* **76**, 527-541 (2004).

376. Castilho, L. *et al.* A novel FY allele in Brazilians. *Vox Sang.* **87**, 190-195 (2004).
377. Herrera, S. *et al.* Antibody response to *Plasmodium vivax* antigens in Fy-negative individuals from the Colombian Pacific coast. *Am. J. Trop. Med. Hyg.* **73**, 44-49 (2005).
378. Wang, R. B. *et al.* Immune responses to *Plasmodium vivax* pre-erythrocytic stage antigens in naturally exposed Duffy-negative humans: a potential model for identification of liver-stage antigens. *Eur. J. Immunol.* **35**, 1859-1868 (2005).
379. Estalote, A. C., Proto-Siqueira, R., Silva, W. A., Jr., Zago, M. A. & Palatnik, M. The mutation G298A-->Ala100Thr on the coding sequence of the Duffy antigen/chemokine receptor gene in non-caucasian Brazilians. *Genet. Mol. Res.* **4**, 166-173 (2005).
380. Ferri, G. *et al.* Minisequencing-based genotyping of Duffy and ABO blood groups for forensic purposes. *J. Forensic Sci.* **51**, 357-360 (2006).
381. Perna, S. J., Cardoso, G. L. & Guerreiro, J. F. Duffy blood group genotypes among African-Brazilian communities of the Amazon region. *Genet. Mol. Res.* **6**, 166-172 (2007).
382. Sellami, M. H. *et al.* Duffy blood group system genotyping in an urban Tunisian population. *Ann. Hum. Biol.* **35**, 406-415 (2008).
383. Cotorruelo, C. M. *et al.* Distribution of the *FYB*^{ES} and RHCE*ce(733C>G) alleles in an Argentinean population: implications for transfusion medicine. *BMC Med. Genet.* **9**, 40 (2008).
384. Soares, S. C., Abé-Sandes, K., Nascimento Filho, V. B., Nunes, F. M. & Silva, W. A., Jr. Genetic polymorphisms in TLR4, CR1 and Duffy genes are not associated with malaria resistance in patients from Baixo Amazonas region, Brazil. *Genet. Mol. Res.* **7**, 1011-1019 (2008).
385. MalariaGEN Consortium. Duffy data from Burkina Faso. Further details can be found at http://www.malariagen.net/network/cp1burkinafaso. (2009)
386. MalariaGEN Consortium. Duffy data from Cameroon. Further details can be found at http://www.malariagen.net/network/cp1cameroon. (2009)
387. MalariaGEN Consortium. Duffy data from The Gambia. Further details can be found at http://www.malariagen.net/network/cp1thegambia. (2009)
388. MalariaGEN Consortium. Duffy data from Ghana. Further details can be found at http://www.malariagen.net/network/cp1ghananoguchi. (2009)
389. MalariaGEN Consortium. Duffy data from Kenya. Further details can be found at http://www.malariagen.net/home/science/localprojects.php. (2009)
390. MalariaGEN Consortium. Duffy data from Malawi. Further details can be found at http://www.malariagen.net/network/cp1malawi. (2009)
391. MalariaGEN Consortium. Duffy data from Tanzania. Further details can be found at http://www.malariagen.net/network/cp1tanzaniamoshi. (2009)
392. Arends A, Gòmez, G. Pers Comm. Duffy data from Venezuela. (2009)
393. Barnadas C, Ménard D, Zimmerman PA. Pers Comm. Duffy data from Madagascar. (2010)
394. Beall C, Gebremedhin A, Zimmerman PA. Pers Comm. Duffy data from Ethiopia. (2010)
395. Fairhurst RM, Long CA, Diakite M. Pers Comm. Duffy data from Mali. (2010)
396. Fairhurst RM, Socheat D. Pers Comm. Duffy data from Cambodia. (2010)
397. Ferreira MU. Pers Comm. Duffy data from Brazil. (2010)
398. Zimmerman PA. Pers Comm. Duffy data from West Africa. (2010)
399. Zimmerman PA. Pers Comm. Duffy data from Papua New Guinea. (2010)

Appendix to Chapter 4 – *G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map*

This Appendix includes:

Note: The original publication of this work also included “Protocol S6: Developing an index of overall national-level risk from G6PD deficiency”. This is not reproduced here as its contents are included in Chapter 6 of this thesis.

Protocol S1. Assembling a global database of G6PD deficiency (G6PDd) prevalence surveys

S1.1 Overview of database requirements

This document summarises the methodological steps involved in assembling the input dataset for the mapping model, documenting both the assembly of sources and data abstraction of survey details into a customised database. The aim of this literature search was to assemble a database of G6PD deficiency (G6PDd) surveys which would form the evidence base for the geostatistical mapping model (Protocol S2). A similar search strategy and data abstraction protocol has been previously described in detail in reference to the Duffy blood group variants by Howes et al. [1] and malaria parasite rate data by Guerra et al. [2]. The final input dataset is available for download from: http://www.map.ox.ac.uk/.

S1.2 Library assembly

Extensive efforts were invested in attempting to assemble all available surveys of G6PDd. These came both from the published and unpublished literature, dating from 1959. Systematic searches of the online biomedical literature databases PubMed (http://www.pubmed.gov), ISI Web of Science (http://wok.mimas.ac.uk/) and Scopus (http://www.scopus.com) were conducted for all articles including the terms 'G6PD', 'glucose-6-phosphate dehydrogenase' and 'glucose 6 phosphate dehydrogenase'. Following duplicate removal, a total of 17,272 unique sources were found to contain these terms. Titles and abstracts were reviewed conservatively for relevance to the project and, for example, clinical case reports, laboratory studies and animal studies were excluded. Searches were then conducted to identify full-text articles of the potentially relevant sources.

Unpublished sources were also identified through contact with the research and medical communities. In particular, the Filipino Newborn Screening Reference Center (NIH, Philippines) contributed their universal screening results since 2004 to this study, adding 636 locations to the database. Individuals who shared data with the project are gratefully acknowledged on the Malaria Atlas Project (MAP) website (http://www.map.ox.ac.uk/inherited-blood-disorders/acknowledgements/).

S1.3 Dataset inclusion criteria

All sources for which full text copies could be identified were reviewed in detail. Specific inclusion criteria for the final dataset are detailed below. A schematic breakdown of these criteria is illustrated in Figure S1.1.

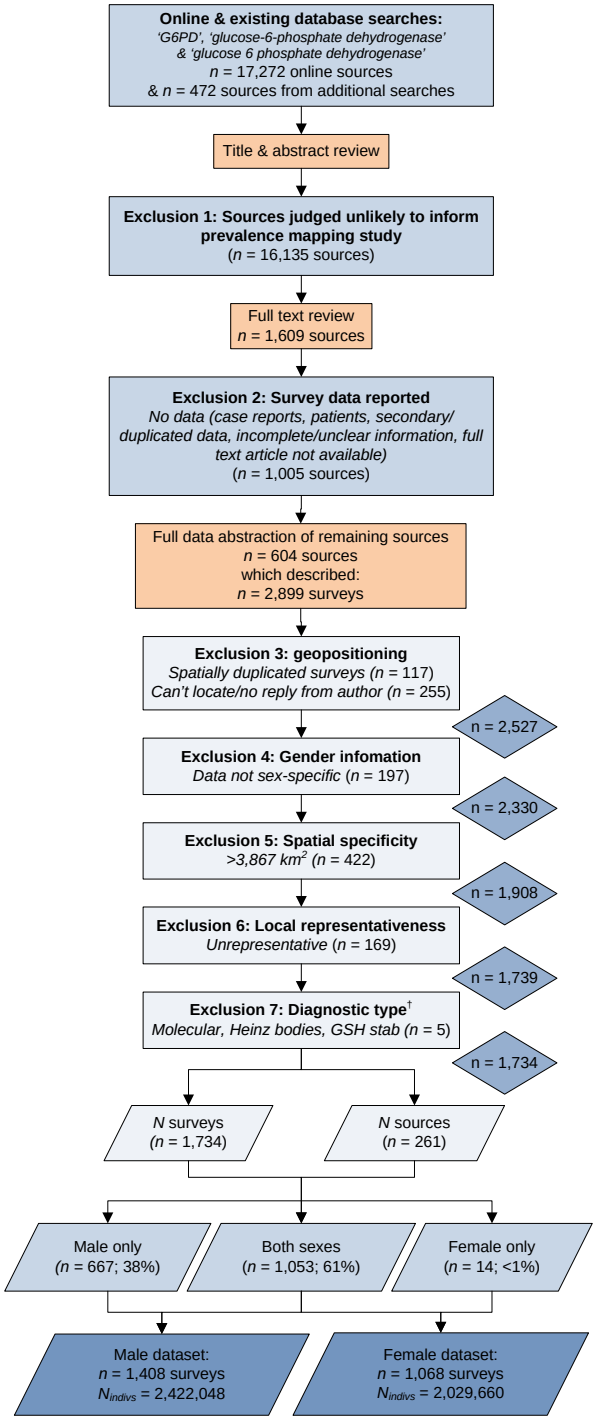


Figure S1.1. Breakdown of the exclusions applied in assembling the input dataset. Sources include published journal articles, book chapters, published and unpublished reports, etc; each source may report multiple surveys. Orange rectangles correspond to review steps; blue rectangles show the number of sources or surveys at each level of exclusion; pale blue indicating when the decisions are based on full-text review. Diamonds indicate the number of remaining surveys following each exclusion. The final input dataset, separated by sex, is represented by parallelograms.
†Most of the surveys reporting these methods had already been excluded before this step; this is not a representative estimate of the number of surveys available using those diagnostic methods.

A comprehensive protocol was developed to ensure objectivity and consistency in the abstraction decision-making process across the data-abstraction team [MMH, REH, OAN, FBP]. All data entered into the database were checked by a second individual and then reviewed by a senior member of the team [REH or FBP] before inclusion in the model input dataset.

Sex

To account for the G6PD gene's position on the X-chromosome and hence its sex-specific inheritance patterns [3], rates of deficiency were recorded separately for males and females. Data were then considered separately by the model (Protocol S2). Surveys which did not report sex-specific data were excluded.

Representativeness

To generate a map showing the overall population prevalence of G6PDd, only surveys of representative population samples were included. Surveys preferentially selecting individuals from specific ethnic groups were excluded as these would provide incomplete, and potentially skewed, information about the overall local frequency of G6PDd. Furthermore, samples of patients, both with minor ailments and all hospitalised cases, including malaria patients, were excluded for the same reasons of being potential biased samples.

Spatial specificity

Prerequisite to inclusion in the database is that surveys be geographically specific and possible to map. Both the latitude and longitude of all surveys, as well as the survey's spatial extent, were recorded. A specific mapping protocol has been developed within the Malaria Atlas Project (MAP) to ensure that a maximum of surveys can be positioned with the required specificity [1,2]. This involved both the use of online geopositioning gazetteers and direct contact with authors where online searches failed. Wherever possible, surveys were mapped to their specific recruitment sites. However, surveys were often reported to administrative level, rather than village or city levels; administrative level 1 corresponds to provinces or states, and administrative level 2 to districts. Globally, the sizes of these administrative divisions are highly variable. Previous MAP databases have used administrative region levels as a cut-off for inclusion: for instance allowing administrative level 2 data, but not administrative level 1 or national-level data. To refine the specificity of the data used here, the G6PD database included the spatial extent of all data points alongside geographic coordinates. We used ArcGIS Desktop (ArcMap 10.0, ESRI Inc., Redlands, CA, USA) to digitise and calculate surface areas of spatial polygons. Where data

presented were amalgamated from multiple sites, the maximum distance between sites was measured in Encarta (Microsoft Corporation, Redmond, WA, USA), and the area of the circle around these was calculated to obtain a conservative estimate of the maximal area represented. A conservative cut-off, based on the average size of the administrative level 2 regions, was used: 3,867 km^2. All surveys with an extent larger than this were excluded from the database as these were considered too spatially unspecific. This meant that national-level data from ten countries could be included; these were mostly Pacific islands, but also Bahrain and Singapore.

Only a single survey estimate could be entered for each pixel on the global grid. All spatial duplicates were therefore identified and only the survey considered to be most representative of the contemporary status of G6PD deficiency was included. In selecting between spatially duplicated surveys, the factors considered were survey date, sample size, whether both sexes had been tested and the diagnostic methodology. Beutler's SPOT test [4] and electrophoretic methods [5] were considered more reliable than Motulsky's cresyl blue dye decolourisation test [6], for example. Spatial overlap was notable in the Philippines, due to inclusion of the extensively distributed national screening data. To overcome this, all opportunistic community surveys from the Philippines (n = 27) were excluded, leaving only the gold-standard universal screening data.

S1.4 Survey diagnostic methods

The International Committee for Standardization in Haematology recommends the fluorescent spot test for population screening surveys [7]. However, this method requires some technical equipment and numerous alternative methods are also in wide use to overcome the constraints of the spot test. There is no single, standardised method in use today. Furthermore, the methods used are invariably modified by users (for example, in terms of the cut-off times imposed) and adapted to local conditions and survey constraints; there is therefore a broad range of methodology employed in diagnosing G6PDd. The database assembled here recorded key information about the methodology used, and these were categorised into ten methods (Figure S1.2). Where multiple methods were used, data from the initial screening method used to test the whole population was recorded; if several methods tested the full population (rare), then the method deemed most reliable was used.

Phenotypic vs. molecular diagnosis

The central dichotomy within diagnostics lies between molecular and phenotypic tests. As the objective of this study was to represent clinically significant cases, these were best detected by assessment of residual enzyme activity levels. In contrast, molecular methods

look for specific mutations, and their relationship with the expressed enzyme activity level, and thus their clinical severity, is poorly established in most cases [8]. A further limitation to molecular methods stems from the numerous mutations in the G6PD gene (186 mutations catalogued by Minucci et al. in 2012 [9]); most molecular methods search for only a subset of these, and hence cannot reliably identify all deficiency cases; significant discrepancies in this regard were recently demonstrated by a comparative study of enzymatic and molecular methods by Johnson et al. [10]. Surveys diagnosed with only molecular methods were therefore excluded from the current mapping analysis.

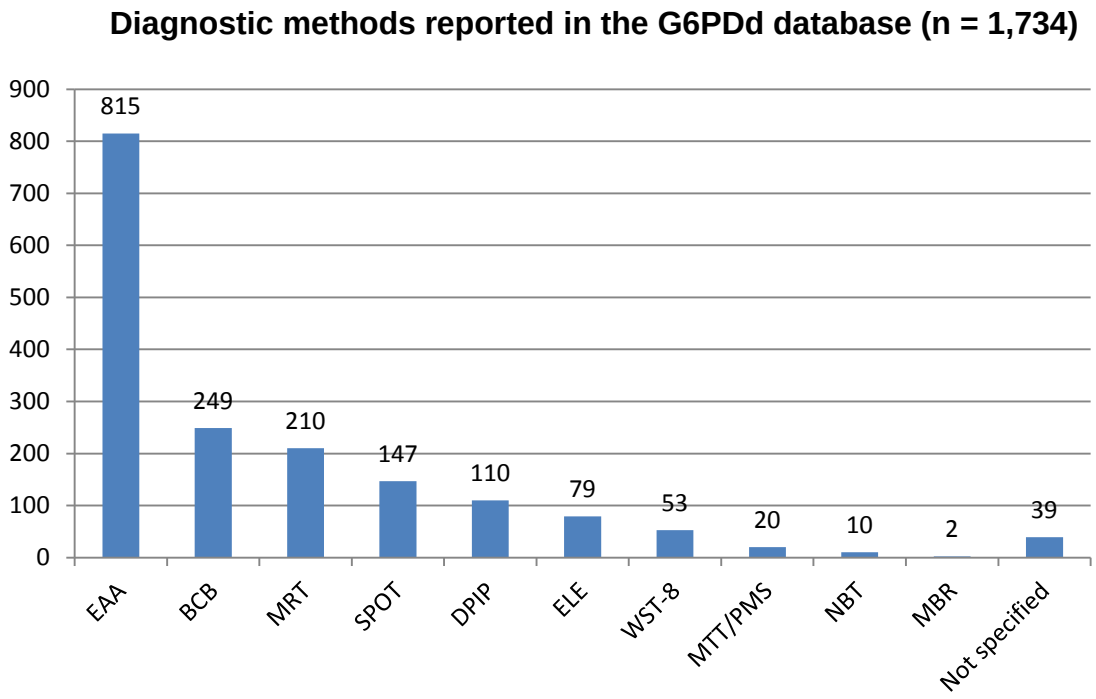


Figure S1.2. Diagnostic methodology use across the dataset. *EAA*: enzyme activity assay (quantitative or semi-quantitative measure of NAPD>NADPH conversion) (47%); *BCB*: brilliant cresyl blue test (14%); *MRT*: methaemoglobin reduction test (12%), *SPOT*: NADPH fluorescent spot test (9%), *DPIP*: 2,6-dicholorophenol indophenol dye test (6%), *ELE*: enzyme electrophoresis (5%), *WST-8*: WST-8/1-methoxy PMS method (3%); *MTT-PMS*: 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H tetrazolium bromide (MTT)-PMS (1%); NBT: nitro blue tetrazolium test (<1%); MBR: methylene blue reduction test (<1%). In 2% of surveys, the diagnostic method was not reported or was unclear from the author description.

Diagnosing deficiency

A number of simple, binary, qualitative/semi-quantitative/quantitative diagnostic kits were widely recorded across our database, with standardised protocols for many of these

originally set by the WHO [11], which are all considered adequate for diagnosing deficiency in males [5,12]. Most of these kits assess the rate of NAPD reduction to NADPH through dye decolourisation or fluorescence, which informs G6PD enzyme activity. Sensitivity and specificity of some of the recommended methods has been previously examined and were found to be very high for males [13,14]. Applying the principle of dye intensity to electrophoretic gels, detailed diagnoses of particular variants as well as residual enzyme activity can be ascertained. Enzyme electrophoresis is therefore another reliable diagnostic. Exceptions to this are the Heinz body test and the glutathione stability tests, which are indirect, proxy measures of G6PD activity and were not considered to be reliable indicators [12]. Thus all surveys using these tests were excluded. Enzyme activity levels are the same for homozygous deficient females as they are in deficient males – all red blood cells in a homozygous deficient female are affected by the deficiency. Thus diagnosis has been demonstrated to be equally reliable in homozygotes and deficient males.

Heterozygous deficient females are harder to diagnose reliably due to their mosaic populations of deficient and wild-type red blood cells. The relative proportions of each cell-type population are variable and a given genotype may be expressed as a spectrum of phenotypic levels of deficiency. As a result, the diagnostic cut-off point may therefore influence the proportion of females found to be deficient in a survey. This is discussed in further detail in Protocol S5. All model predictions made from these input datasets, however, must be evaluated with these constraints and uncertainties in mind.

Conclusions

With the exception of Heinz body diagnoses, GSH stability tests and molecular analyses, all diagnostic methods were included in the study. Diagnostic outcome is influenced by a number of factors, including variable sensitivity of methods used, variable reaction time cut-offs used by investigators, and differing levels of residual activity associated with the numerous genetic variants. A further confounding factor for diagnosis which has not yet been discussed, is the increased probability of false-positive diagnoses due to anaemia, which reduces the overall number of red blood cells, and therefore level of G6PD enzyme per volume of blood; G6PD diagnosis should account for this [15], though rarely does in most community screening surveys.

While the influence of these issues has, on the whole, been deemed relatively acceptable for diagnosing males and homozygotes, who tend to have a distinct deficient/non-deficient status, diagnosing heterozygotes across a spectrum of residual activity levels is less clear-cut. However, the absence of a single standardised methodological rapid-test suited to mass surveying, as well as the poor understanding of the genotype-phenotype relationship in

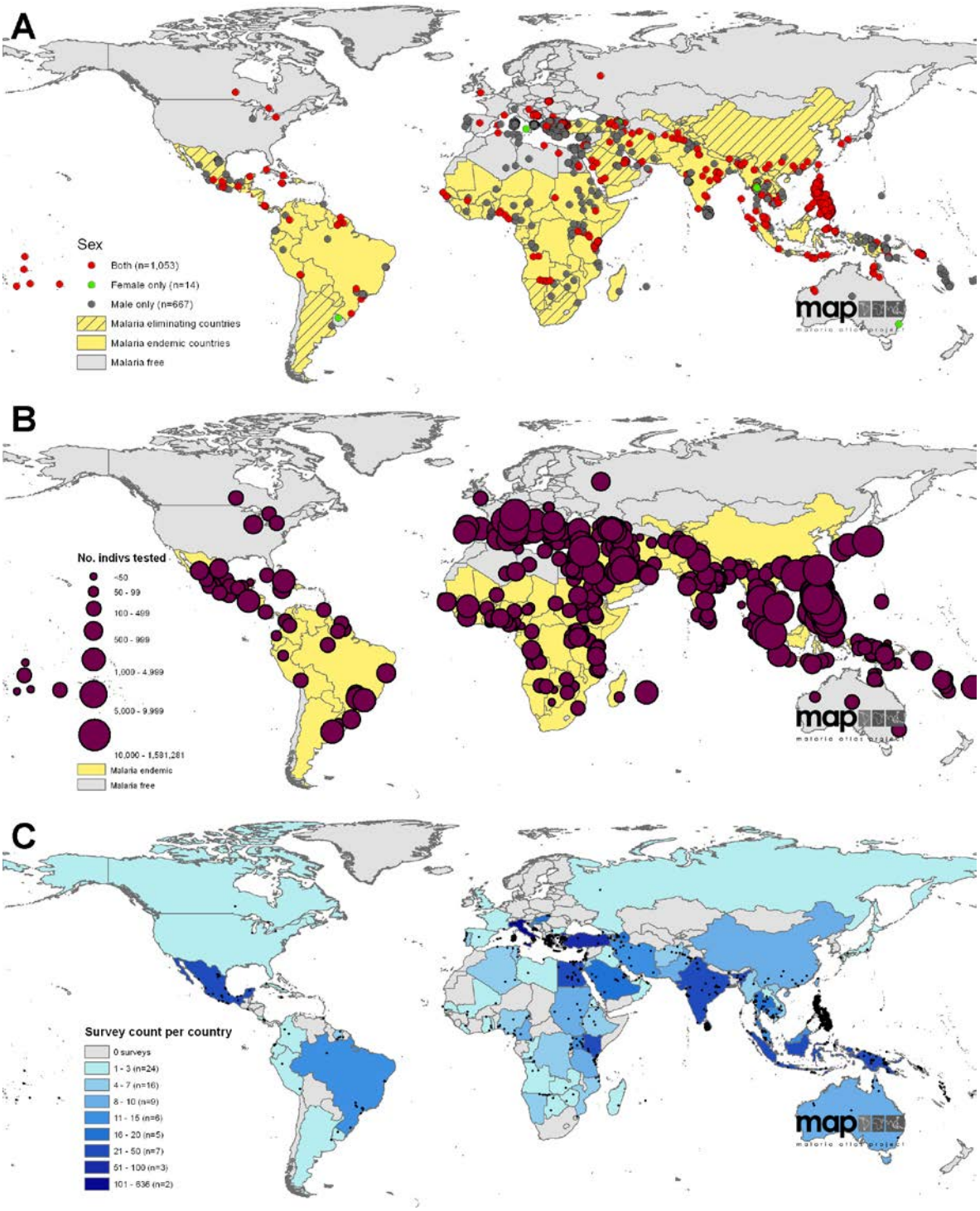
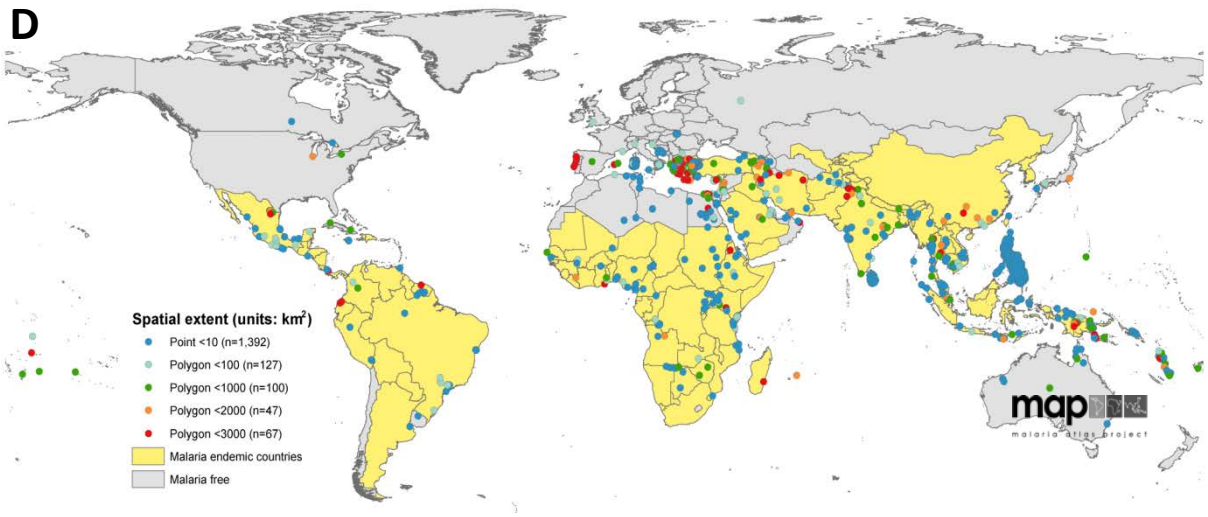


Figure S1.3. Global distribution of the assembled G6PDd dataset. Panel A shows the distribution of surveys according to which sex was tested: red where both sexes are included, green if female data only and black of male data only; Panel B maps the total numbers of individuals (both sexes) sampled in each survey; Panel C indicates the number of surveys identified from each country; Panel D maps the spatial extent of each of the surveys. Background map colours in Panels A, B and D represent the status of the national malaria programme (malaria free or malaria endemic).



heterozygotes and of the association with haemolytic risk, there is limited potential for imposing adjustments to the input dataset. The raw survey data was therefore used, with female data being treated separately from the more reliable male data. The map aims to represent individuals at greatest risk of G6PD-associated haemolysis. While some discrepancy risks being introduced due to the difficulties of diagnosing this complex system, this will only affect a small proportion of borderline heterozygote individuals, with intermediate deficiency, thus individuals who are not necessarily at greatest risk of deficiency. In the absence of any standardised methodology for this, we made the assumption that phenotypic diagnostic test cut-offs would have been calibrated to detect locally significant levels of G6PDd and that individuals at risk of primaquine-associated haemolysis (male and female) would be included in this group.

S1.5 The final G6PDd survey dataset

As detailed in Figure S1.1, the original set of references included 17,272 sources from online database searches and 472 sources from existing databases and personal communications. From these, 1,734 spatially-unique surveys were identified which met all our inclusion criteria, 1,289 of which were in malaria endemic countries (MECs) (74%); 97 countries, including 55 MECs, are represented in the dataset. These surveys were reported from 261 sources, which are listed in the Supplementary References. Table S1 summarises this final set of studies used to inform the mapping model analysis.

Surveying effort, in terms of number of surveys and number of individuals tested is strongly biased towards Asia & Europe, with 85% of surveys and 98.6% of individuals sampled in the Eurasian region. Surveys were tightly clustered across Sardinia, southern

Table S1. Summary of G6PD dataset characteristics by region. Numbers correspond to spatially unique community surveys meeting the study inclusion criteria and used in the mapping analysis. Italicised terms refer to the numbers within MECs.

	Africa		Americas		Asia & Europe		Global	
Total surveys	**Total**	***MEC***	**Total**	***MEC***	**Total**	***MEC***	**Total**	***MEC***
	169	*132*	67	*56*	1,498	*1,101*	1,734	*1,289*
Number of countries	28	*23*	15	*9*	54	*23*	97	*55*
Publication time								
1959 - 1969	42	*31*	22	*19*	204	*50*	268	*100*
1970 - 1979	41	*29*	14	*9*	77	*59*	132	*97*
1980 - 1989	21	*16*	8	*6*	143	*110*	172	*132*
1990 - 1999	12	*3*	12	*11*	224	*60*	248	*74*
2000 - 2011	53	*53*	11	*11*	850	*822*	914	*886*
Spatial extent (all surveys are ≤3,867km^2)								
Admin 0 centroids	1	*0*	0	*0*	9	*0*	10	*0*
Admin 1 centroids	8	*2*	0	*0*	16	*4*	24	*6*
Admin 2 centroids	3	*3*	9	*8*	62	*17*	74	*28*
Admin 3 centroids	0	*0*	0	*0*	8	*8*	8	*8*
Polygons (>25km^2 to ≤3,867km^2)	55	*52*	13	*9*	105	*51*	173	*112*
Wide areas (>10km^2 to ≤25km^2)	9	*6*	8	*8*	13	*8*	30	*22*
Points (≤10km^2)	88	*64*	36	*30*	1,265	*999*	1,389	*1,093*
Multiple points	5	*5*	1	*1*	20	*14*	26	*20*
Data type								
Male only	88	*56*	33	*30*	546	*225*	667	*311*
Female only	0	*0*	1	*0*	13	*11*	14	*11*
Male & Female	81	*76*	33	*26*	939	*865*	1,053	*967*
Total individuals sampled								
Male	24,528	*13,777*	24,979	*20,414*	2,372,541	*2,189,795*	2,422,048	*2,223,986*
Female	6,859	*5,172*	5,028	*3,374*	2,017,773	*1,973,873*	2,029,660	*1,982,419*
Total	31,387	*18,949*	30,007	*23,788*	4,390,314	*4,163,668*	4,451,708	*4,206,405*
Survey count by sample size (male + female)								
<50	57	*44*	7	*7*	258	*157*	322	*208*
50 - <100	36	*31*	13	*12*	307	*201*	356	*244*
100 - <500	63	*49*	33	*26*	495	*355*	591	*430*
500 - <1,000	9	*7*	9	*7*	126	*109*	144	*123*
1,000 - <5,000	4	*1*	5	*4*	195	*171*	204	*176*
5,000 - <10,000	0	*0*	0	*0*	49	*45*	49	*45*
>10,000	0	*0*	0	*0*	68	*63*	68	*63*
Mean	185.7	*143.6*	447.9	*424.8*	2,930.8	*3,782.0*	2,567.0	*3,263.3*
Median	83.0	*77.0*	147.0	*119.5*	165.0	*231.0*	145.0	*197.5*
IQR	43 - 160	*45 - 150*	90 - 479	*85 - 419*	70 - 662	*80 - 1009*	66 - 551	*77 - 759*
G6PD deficiency prevalence (male & female data)								
Surveys with no G6PDd	22	*15*	23	*20*	212	*98*	257	*133*
Surveys with G6PDd	147	*118*	44	*36*	1,286	*1,003*	1,477	*1,156*

mainland Italy, western Turkey, Sri Lanka, and especially the Philippines from where national screening data was available. Their distribution, datatype in terms of sex, sample size and prevalence values are shown in Figure S1.3. All three aspects influence each survey's relative influence on the model predictions (Protocol S2). Total numbers of males and females surveyed were roughly equal (2.4 million males and 2.0 million females). The distribution of these surveys through time is displayed in Figure S1.4.

A range of diagnostic methods were used to determine G6PD status, with the most common being enzyme activity assays reported in 47% of surveys (quantitative or semi-quantitative assessment of NAPD>NADPH breakdown [12], this method is used in the Philippines, which made up 36% of the dataset); other diagnoses were qualitative or semi-quantitative: Motulsky and Campbell-Kraut's [6] brilliant cresyl blue dye test (14% surveys globally); Brewer's [16] methaemoglobin reduction test (12% surveys globally); Beutler's [4] NADPH fluorescent spot test (9%); and Bernstein's [17] DPIP method (6% surveys globally) (Figure S1.2)

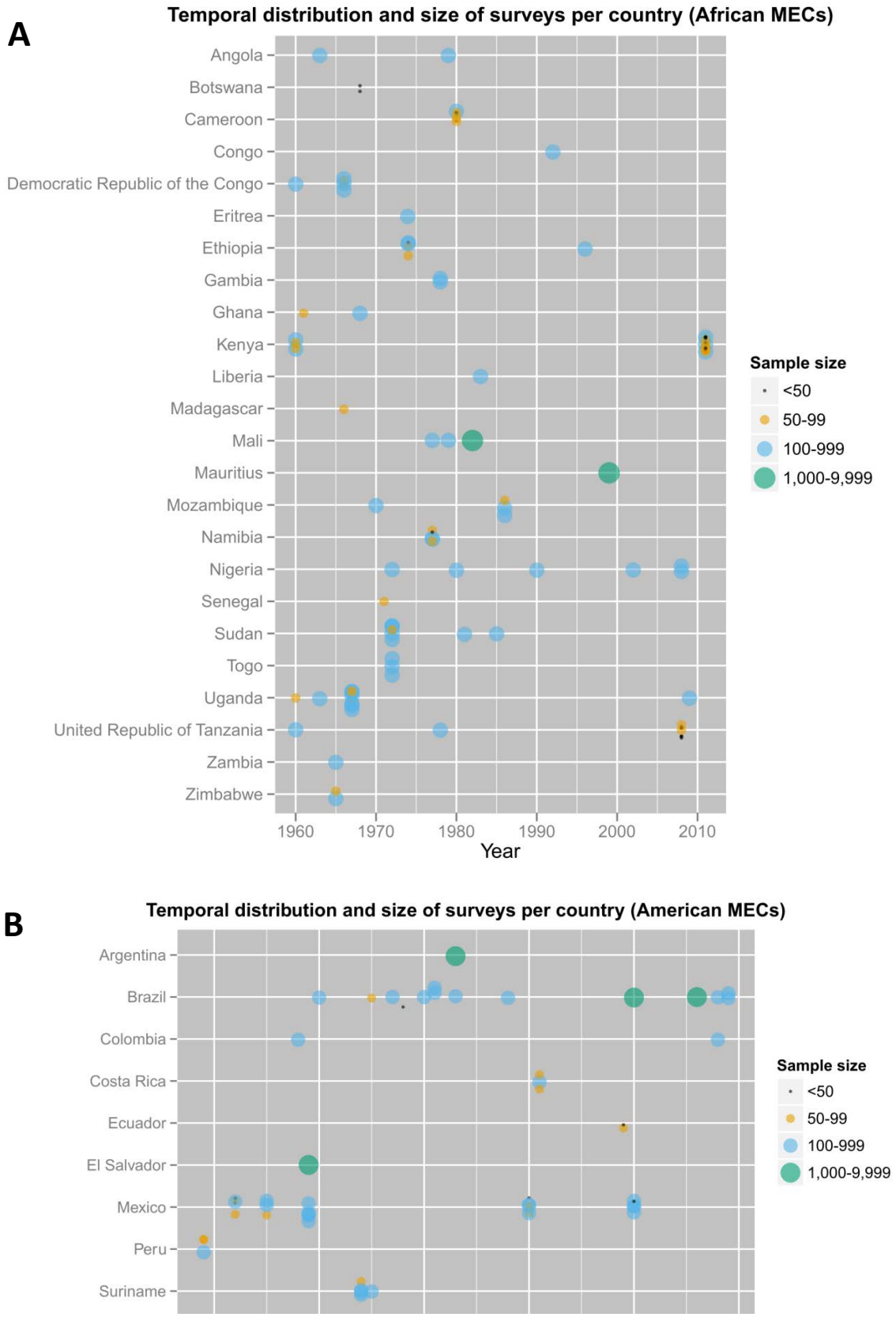
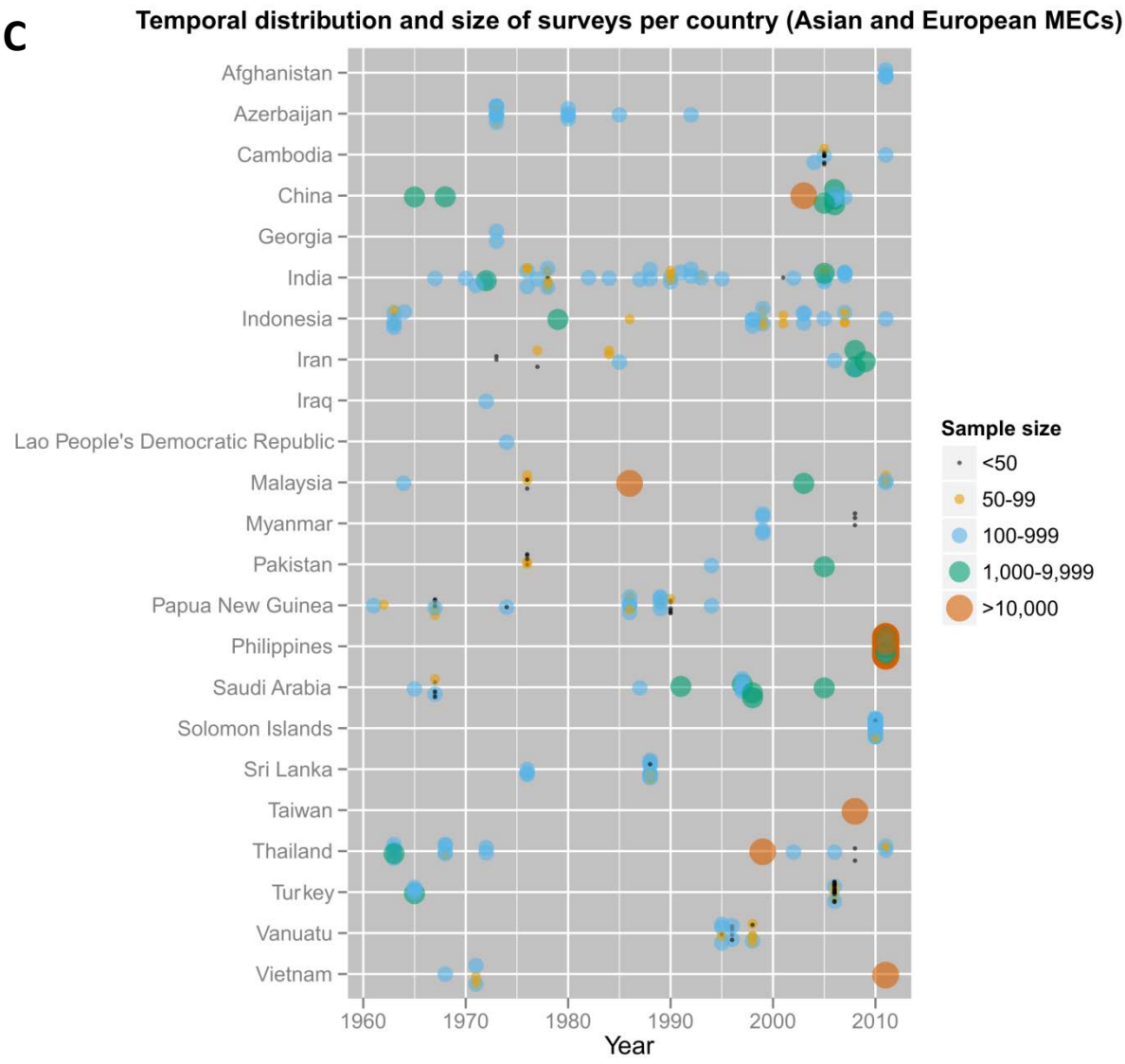


Figure S1.4. The temporal distribution of surveys by country, in African MECs (Panel A), MECs of the Americas (Panel B) and of Europe and Asia (Panel C). Data point colour and size reflect the total number of individuals tested in each survey. The date corresponds to the date of publication, as survey dates were not consistently reported.



S1.6 Defining Malaria Endemic Countries (MECs) limits

In considering the risk that primaquine carries for G6PDd individuals, this study is concerned with the condition's prevalence within areas of malaria endemicity. Outside these regions, primaquine will not be used routinely, and is most likely to be prescribed from hospitals, where facilities for G6PD activity testing are more likely to be available. The public health interest in this map is anticipated to be greatest in areas where point-of-care testing is unavailable. Furthermore, few G6PD surveys were available outside historically malaria endemic regions (including current MECs and the Mediterranean regions), which would have made predictions for these data-sparse regions much less reliable.

At the low end of transmission, classifying a country's malaria endemic status can be difficult and contentious [18], and is ultimately dependent upon meeting the administrative WHO certification criteria. The G6PDd map presented here is most applicable to countries with large-scale programmes, rather than targeted individual hospital-based treatment, as is likely to be the case in countries where malaria is very rare. The MAP's *Plasmodium* transmission mapping initiatives [19-21] do not consider countries with very low endemicity where transmission is limited to small, sporadic foci stimulated by imported cases (including Algeria, Armenia, Egypt, Jamaica, Russia and Syria), to be usefully included as endemic in mapping terms. The MAP therefore considers 99 MECs [21] (including Mayotte, a French overseas department). We focus this G6PDd mapping effort to the MAP MEC limits where large-scale primaquine use is still appropriate and point-of-care G6PDd testing likely to be rare. Countries targeting national elimination were identified from the 2011 Atlas of Malaria-Eliminating Countries [22] and included 35 of the MECs considered in this study (Figure S1.5).

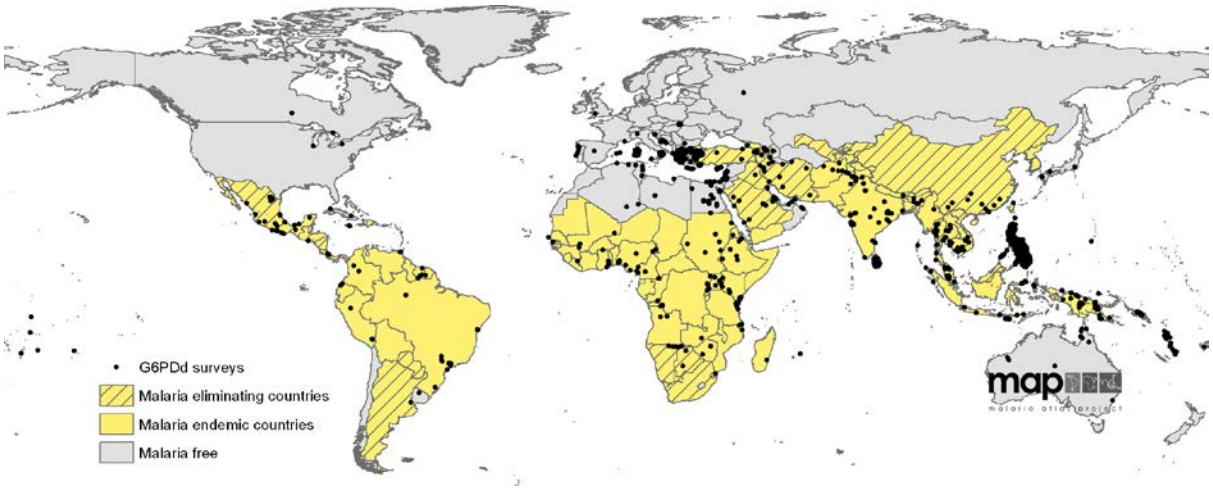


Figure S1.5. Malaria endemic country limits used for mapping G6PDd. Grey areas denote malaria-free regions; yellow areas are malaria endemic (n=99 countries), with hatches indicating those countries targeting malaria elimination (n=35) [22].

References

1. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, et al. (2011) The global distribution of the Duffy blood group. Nat Commun 2: 266.
2. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, et al. (2007) Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. Malar J 6: 17.
3. Cappellini MD, Fiorelli G (2008) Glucose-6-phosphate dehydrogenase deficiency. Lancet 371: 64-74.
4. Beutler E, Mitchell M (1968) Special modifications of the fluorescent screening method for glucose-6-phosphate dehydrogenase deficiency. Blood 32: 816-818.
5. WHO Working Group (1989) Glucose-6-phosphate dehydrogenase deficiency. Bull World Health Organ 67: 601-611.
6. Motulsky AG, Campbell-Kraut JM. Population genetics of glucose-6-phosphate dehydrogenase deficiency of the red cell. In: Blumberg BS, editor; 1961; New York, NY. Grune & Stratton. pp. 159.
7. Beutler E, Blume KG, Kaplan JC, Lohr GW, Ramot B, et al. (1979) International Committee for Standardization in Haematology: recommended screening test for glucose-6-phosphate dehydrogenase (G-6-PD) deficiency. Br J Haematol 43: 465-467.
8. Baird JK, Surjadjaja C (2011) Consideration of ethics in primaquine therapy against malaria transmission. Trends Parasitol 27: 11-16.
9. Minucci A, Moradkhani K, Hwang MJ, Zuppi C, Giardina B, et al. (2012) Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. Blood Cells Mol Dis 48: 154-165.
10. Johnson MK, Clark TD, Njama-Meya D, Rosenthal PJ, Parikh S (2009) Impact of the method of G6PD deficiency assessment on genetic association studies of malaria susceptibility. PLoS One 4: e7246.
11. Betke K, Brewer GJ, Kirkman HN, Luzzatto L, Motulsky AG, et al. (1967) Standardization of procedures for the study of glucose-6-phosphate dehydrogenase. Report of a WHO Scientific Group. World Health Organ Tech Rep Ser No. 366: 1-53.
12. Beutler E (1994) G6PD deficiency. Blood 84: 3613-3636.
13. Peters AL, Van Noorden CJ (2009) Glucose-6-phosphate dehydrogenase deficiency and malaria: cytochemical detection of heterozygous G6PD deficiency in women. J Histochem Cytochem 57: 1003-1011.
14. Beutler E, Duparc S (2007) Glucose-6-phosphate dehydrogenase deficiency and antimalarial drug development. Am J Trop Med Hyg 77: 779-789.
15. Matsuoka H, Nguon C, Kanbe T, Jalloh A, Sato H, et al. (2005) Glucose-6-phosphate dehydrogenase (G6PD) mutations in Cambodia: G6PD Viangchan (871G>A) is the most common variant in the Cambodian population. J Hum Genet 50: 468-472.
16. Brewer GJ, Tarlov AR, Alving AS (1962) The methemoglobin reduction test for primaquine-type sensitivity of erythrocytes. A simplified procedure for detecting a specific hypersusceptibility to drug hemolysis. JAMA 180: 386-388.
17. Bernstein RE (1962) A rapid screening dye test for the detection of glucose-6-phosphate dehydrogenase deficiency in red cells. Nature 194: 192-193.
18. Cohen JM, Moonen B, Snow RW, Smith DL (2010) How absolute is zero? An evaluation of historical and current definitions of malaria elimination. Malar J 9: 213.
19. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6: e1000048.
20. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. PLoS Negl Trop Dis 4: e774.
21. Tatem AJ, Smith DL, Gething PW, Kabaria CW, Snow RW, et al. (2010) Ranking of elimination feasibility between malaria-endemic countries. Lancet 376: 1579-1591.

22. The Global Health Group and the Malaria Atlas Project (2011) Atlas of Malaria-Eliminating Countries. San Francisco: The Global Health Group, Global Health Sciences, University of California, San Francisco.

Protocol S2. Model-based geostatistical framework for predicting G6PDd prevalence maps

This protocol provides information about the geostatistical model used to map the prevalence of G6PDd. The general principles of model-based geostatistics have been previously described by Diggle and Ribeiro [1]. The general framework of the model used here has been fully described by Piel et al. in application to HbS mapping [2], though specific adaptations to that model were made to reflect the G6PD gene's sex-linked inheritance mechanism.

S2.1 Model requirements in relation to G6PD genetics

The G6PD gene is carried on the X-chromosome, meaning that males inherit only a single copy of the gene ("hemizygotes"). In contrast, females have two copies so may carry two wild-type alleles or two deficient alleles ("homozygotes"), or a combination of one wild-type and one deficient ("heterozygotes"). The relative frequencies of each of these genotypes can be predicted if populations are assumed to be in Hardy-Weinberg equilibrium [3,4], where q represents the frequency of deficient alleles, making $(1-q)$ the allele frequency of normal G6PD expression. The overall population allele frequency corresponds to the frequency of male hemizygotes. As a consequence, the frequency of female homozygous deficiency expression corresponds to q^2, and $2q\,(1-q)$ in heterozygotes. However, for reasons discussed in Protocol S5, the heterozygous female genotype is often not expressed as phenotypically deficient, so only a proportion of genetically heterozygote females are identified in surveys and reported as deficient cases. There is therefore a genotype-phenotype disjunction between frequencies of observed deficient females and expected deficient females based upon Hardy-Weinberg derivations from frequencies of deficiency in males.

The model requirements were therefore to incorporate prevalence data relating to both males and females; and from this input dataset to generate continuous frequency estimates across malaria endemic countries (MECs; Protocol S1) with quantified uncertainty measures for rates of deficiency in (i) males, (ii) homozygous females, and (iii) all females – this final category being the combination of all homozygotes and the proportion of heterozygotes expected to be diagnosed as phenotypically deficient.

S2.2 The model

In this section, we describe our Bayesian spatial model for the G6PDd allele frequency surface $q(.)$. q takes as its argument an arbitrary location on the Earth's surface. The

posterior $[q]$ induces a posterior $[qq]$ for the homozygous frequency surface $qq(.)$. We computed summaries of $[q(x)]$ and $[qq(x)]$, such as the mean E and the variance V, at each location x, to produce the maps relating to G6PDd allele frequency. Predictions of expected genetic heterozygotes $[\{2q(1-q)\}(x)]$ were adjusted to reflect the observation that many genetic heterozygotes are not phenotypically deficient. A deviance term, h, was determined by the model from the input data to represent the predicted deviation from the expected rates of genetic heterozygotes (see Likelihood). National and regional level prevalence of G6PDd in males and females was then computed from the model posterior distributions, as described in Protocol S4.

Prior

We model q as a nonlinear transformation of a Gaussian random field [5,6] $f(.)$, plus a random field $\varepsilon(.)$ that associates an independent normally distributed value with each location on the earth's surface. Specifically,

$$q(x) = g\big(f(x) + \varepsilon(x)\big)$$

The link function g maps the random variable $f(x) + \varepsilon(x)$, which can be any real number, to the interval (0,1), so $q(x)$ can be used as a probability or prevalence. We used a non-standard link function, which is described below.

The prior for f is parameterized by the constant mean function $M(x) = m$, and the standard exponential covariance function $Cov(x, y) = \phi^2 \exp\frac{d(x-y)}{\theta}$ with amplitude parameter ϕ and range parameter θ. The distance function d gave the great-circle distance between x and y, unless x was on a different side of the Atlantic ocean to y, in which case it returned ∞. This modification prevented data in Africa from unduly influencing the east coast of south America. Suitable priors were assigned to the scalar parameters m, ϕ and θ:

$$p(m) \propto 1$$
$$\phi \sim Exponential(.1)$$
$$\theta \sim Exponential(.1)|(\theta < 0.5)$$
$$f \sim GP(M, Cov)$$

GP indicates a Gaussian process. The units of x, y and θ are earth radii, and m and ϕ are unitless. The unstructured component $\varepsilon(x)$ is modelled as normally distributed with unknown variance V:

$$V \sim Exponential(.1)$$
$$\varepsilon(x) \overset{iid}{\sim} Normal(0, V)$$

The distribution parameters for h, the deviance term from expected Hardy-Weinberg deficiency rates in heterozygous females were defined by α and β, where:

$$h \sim Beta(\alpha, \beta)$$
$$\alpha \sim Exponential(.01)$$
$$\beta \sim Exponential(.01)$$

Likelihood

Separate likelihoods were defined for males and females, in accordance with their different inheritance mechanism of the G6PD gene.

In males: The frequency of G6PDd in males corresponds to the population allele frequency of deficiency, q. If n_i male individuals are sampled at the i^{th} observation location o_i, the probability distribution for the number k_i of copies of the G6PDd allele that will be found is binomial, with probability $q(o_i)$:

$$k_i \sim Binomial(n_i, q(o_i))$$

In females: Under the assumptions of Hardy-Weinberg equilibrium [3,4], if n_i female individuals are sampled at the i^{th} observation location o_i (for a total of $2n_i$ chromosomes), the probability distribution for the number k_i of G6PDd females is binomial, with probability $q(o_i)$:

$$k_i \sim Binomial(2n_i, h_i(2q(o_i)\big(1 - q(o_i)\big)) + q(o_i)^2)$$

Flexible link function and empirical Bayesian analysis

The link function g for binomial data is usually taken to be the inverse logit function:

$$g(x) = logit^{-1}(x) = \frac{\exp x}{1 + \exp x}$$

Piel et al. [7] employed this model. Applying the change of variables formula, the induced prior for $q(x)$ is:

$$p\big(q(x)\big) = \frac{1}{q(x)\big(1-q(x)\big)} Normal\big(logit\big(q(x)\big); m, \phi^2 + V\big)$$

Note that this is essentially a two-parameter family of probability distributions, since ϕ and V appear only in the sum.

Under this model, in areas where data points are highly clustered the best fitting values of m and $\phi^2 + V$ result in long right-hand tails for the predictive distribution of prevalence in the next observation at x, thus skewing the summary statistics to give implausibly high predicted median values. To overcome this problem, we used an alternative flexible link function:

$$j(x) = \sum_{i=0}^{3} \mathcal{C}_i x^i$$

$$g = logit^{-1} \circ j$$

We were unable to infer $\mathcal{C}_i$ jointly with the other model parameters in a fully Bayesian manner due to poor MCMC mixing, so we adopted an empirical fitting approach inspired by data pre-processing steps employed in classical geostatistics. This improved the fitting of the model to the data and is described in more detail below.

Empirical Bayesian approach to fitting the polynomial coefficients

For each observation of n_i males tested and k_i deficient males identified, we first obtained the posterior expectation of the gene pool-wide prevalence of G6PDd with uniform prior density on [0,1]:

$$\hat{p}_i = \frac{k_i + 1}{n_i + 2}$$

We discarded values for which n_i was below 25. Then, we inferred the parameters $\tilde{m}$ and $\tilde{V}$ of the non-spatial Bayesian model:

$$p(\hat{p}) = \prod_i \frac{1}{\hat{p}_i(1-\hat{p}_i)} Normal\big(logit(\hat{p}_i); \tilde{m}, \tilde{V}\big)$$

We then plotted the posterior predictive cumulative distribution function (CDF) of $logit(\hat{p})$ against its empirical CDF, and fitted the coefficients $\mathcal{C}$ of the cubic polynomial function j to

the points using least squares, subject to the constraint that j must be invertible (or, equivalently, monotone).

The polynomial coefficients for such a function are specific to the dataset. The set of coefficients used, corresponding to an invertible function and fitting the empirical CDF was:

$$y = 0.15200304x^3 + 1.12465599x^2 + 0.03658898x + 0.00342442$$

In the Bayesian analysis of the full spatial model, the fitted values of $\mathbb{C}$ were taken as known and fixed. Although this empirical procedure is admittedly informal, the resulting non-standard link function did substantially improve the fit of the model to the data.

Prior predictive constraint

The G6PDd database was largely composed of opportunistic community surveys. As such, data points were unevenly distributed across the malaria endemic region (Figure 2.1A), and the model had to be adapted to this. Inclusion of the highly clustered datasets, particularly in the Philippines, did not appear to alter predictions across the rest of the map. In contrast, high and low latitude areas which were sparse in data (such as northern China and Argentina) were very hard for the model to predict, generating implausibly high population estimates of G6PDd prevalence. Based on our knowledge of the distribution of G6PDd from existing maps [8-10], it seemed reasonable to assume that G6PDd was at low prevalence in these areas, supporting the principle that most surveys will have been conducted where G6PDd was expected to be found. To introduce this into the model, we decided to constrain ϕ, m and V in such a way that the prior predictive distribution of $q(x)$, before the data are incorporated, puts probability mass of 1×10^{-4} or less on values in excess of 0.0001. In other words, we constrained 99.99% of the prior predictive probability mass to be between allele frequencies of 0% and 0.01%.This constraint arguably induces a lack of fit by forcing $f(x)$ to depart from its prior mean by many standard deviations in areas where G6PDd allele frequency is known to be high; but it does remedy the implausibly high predictive values in the data-sparse edges of the map, and does not seem to adversely affect the fit in areas of non-zero allele frequency. A detailed explanation of this process is given by Piel et al. [2].

S2.3 Model implementation

As previously described [11], implementation of the modelling and mapping procedure was divided into two computational tasks: (i) the Bayesian inference stage was implemented

using the Markov chain Monte Carlo (MCMC) algorithm [12,13] and was used to generate samples from the posterior predictive distribution (PPD) of the parameter set and the spatial random fields at the data locations; and (ii) a prediction stage in which samples were generated from the PPD of G6PDd frequencies at each pixel on the 5×5 km grid, as described below in Protocol 2.5.

The scalar parameters ϕ, m, θ and V were updated jointly using Haario, Saksman and Tamminen's adaptive Metropolis algorithm [14], as implemented by PyMC's AdaptiveMetropolis step method. Each value of $\varepsilon(o_i)$ at observation location o_i was updated separately using the standard one-at-a-time Metropolis algorithm. The distribution of the Gaussian random field at the observation locations, $\{f(o_i)\}$, is conjugate to the distribution of $\{\varepsilon(o_i) + f(o_i)\}$, so we updated $\{f(o_i)\}$ by sampling from its full conditional distribution. MCMC output parameter values are summarised in Table S2.1.

Convergence of the MCMC tracefile was judged by visual inspection; one million MCMC iterations were run, with 10% recorded in the output tracefile, the first 50,000 iterations of which were excluded from the mapping stages. During the mapping process, the posterior distributions were thinned by 100, resulting in 500 mapping iterations. MCMC dynamic traces are available on request.

The model code was written in Python programming language (http://www.python.org), and is freely available from the MAP's code repository (https://github.com/malaria-atlas-project and https://github.com/RosalindH/g6pd). The MCMC algorithm was used from the open-source Bayesian analysis package PyMC [15] (http://code.google.com/p/pymc).

S2.4 Overview of mapping procedure

The output algorithm from the MCMC was used to generate PPDs for each model output at all pixels in the MEC 5×5 km grids. The PPDs can be used to determine the most probable prevalence estimate at each pixel [5], these were summarised for each pixel as median values, together with the PPD's interquartile range (IQR). The median was determined to be more representative than the mean value, due to the PPD's right-hand skew, inflating the mean values of the predictions. Maps were generated using Python and Fortran code, available from the MAP's code repository (http://github.com/malaria-atlas-project/generic-mbg) and subsequently displayed in GIS software (ArcMap 10.0, ESRI Inc., Redlands, CA, USA).

S2.5 Uncertainty

Uncertainty metrics.

The MCMC and derived map predictions are directly informed by the evidence-base of surveys. However, there is usually some level of disparity within this database (i.e. in clustered areas) and between the database and the map predictions. The relative influence of each point's influence on the model predictions is moderated according to the data points' sample size (through a binomial model previously described [11]), their spatial distribution and the level of heterogeneity in the frequencies identified between neighbouring data points. These three factors influence the model's ability to predict frequencies, which are then reflected in the spread of the PPD. The precision of the PPD is an indicator of the representativeness of the summary statistic used (in this case median values) [5] and is quantified here by the interquartile range (IQR) of the PPD, thus corresponding to a 50% probability class. However, because the values of the IQR are affected by the underlying G6PDd prevalence levels, with IQR generally increasing with prediction values. The IQR was therefore also standardised against the median map to produce an uncertainty index less affected by the underlying prevalence levels and more illustrative of relative model performance driven by data densities in different locations ($\frac{IQR}{Median}$). Maps of both uncertainty metrics are presented for comparison (Figure S2.1).

Uncertainty maps.

Greatest absolute variation in the predictions (Figure S2.1B) was found from areas of relatively high predicted frequencies and low input data availability (Figure S2.1A), specifically southern Pakistan, the central Sahel region across Chad and Sudan, and southern central Africa (Democratic Republic of Congo (DRC), Zambia, Malawi, southern Tanzania and Mozambique) and Madagascar. Model predictions for these areas should be interpreted with caution; additional surveys from these areas would enable more robust predictions.

Adjusting the IQR values to the predicted median frequency values gives a representation of the relative variability in the PPD (Figure S2.2B). The greatest relative prediction uncertainty is in peripheral transition regions, such as the Horn of Africa and southern Africa – where prevalence drops relative to surrounding areas. The high IQR region across DRC and neighbouring areas is not so pronounced in the proportional map. Greater uncertainty is shown across the Americas, previously masked in the raw IQR map (Figure S2.2A). The large proportion of northern China which is uninformed by data has very high relative uncertainty, reflecting an uncertainty in the model about where exactly to bring down the

prevalence predictions to the very low levels predicted across the north of the country. An equivalent increase in uncertainty and associated data absence is seen on Borneo.

These two uncertainty metrics complement each other, allowing interpretation of confidence in the map predictions and can be used to identify areas where additional data would be most informative towards improving our understanding of G6PDd prevalence.

Parameter	Symbol	Mean	Median	St. Dev.	IQR	95% BCI
Nugget variance	V	0.239	0.226	0.080	0.037	0.190
Amplitude (or partial sill)	$\emptyset$	3.422	3.426	0.202	0.213	0.919
Scale (or range)	θ	0.498	0.499	0.003	0.002	0.007

Table S2.1. MCMC output parameter values. Summary values of the model parameters, as fully described in Protocol S2. Summary statistics of the MCMC output include mean, median, standard deviation (St. Dev.), interquartile range (IQR) and the 95% Bayesian credible interval (95% BCI). Scale is measured in units of Earth radii; other parameters are unitless.

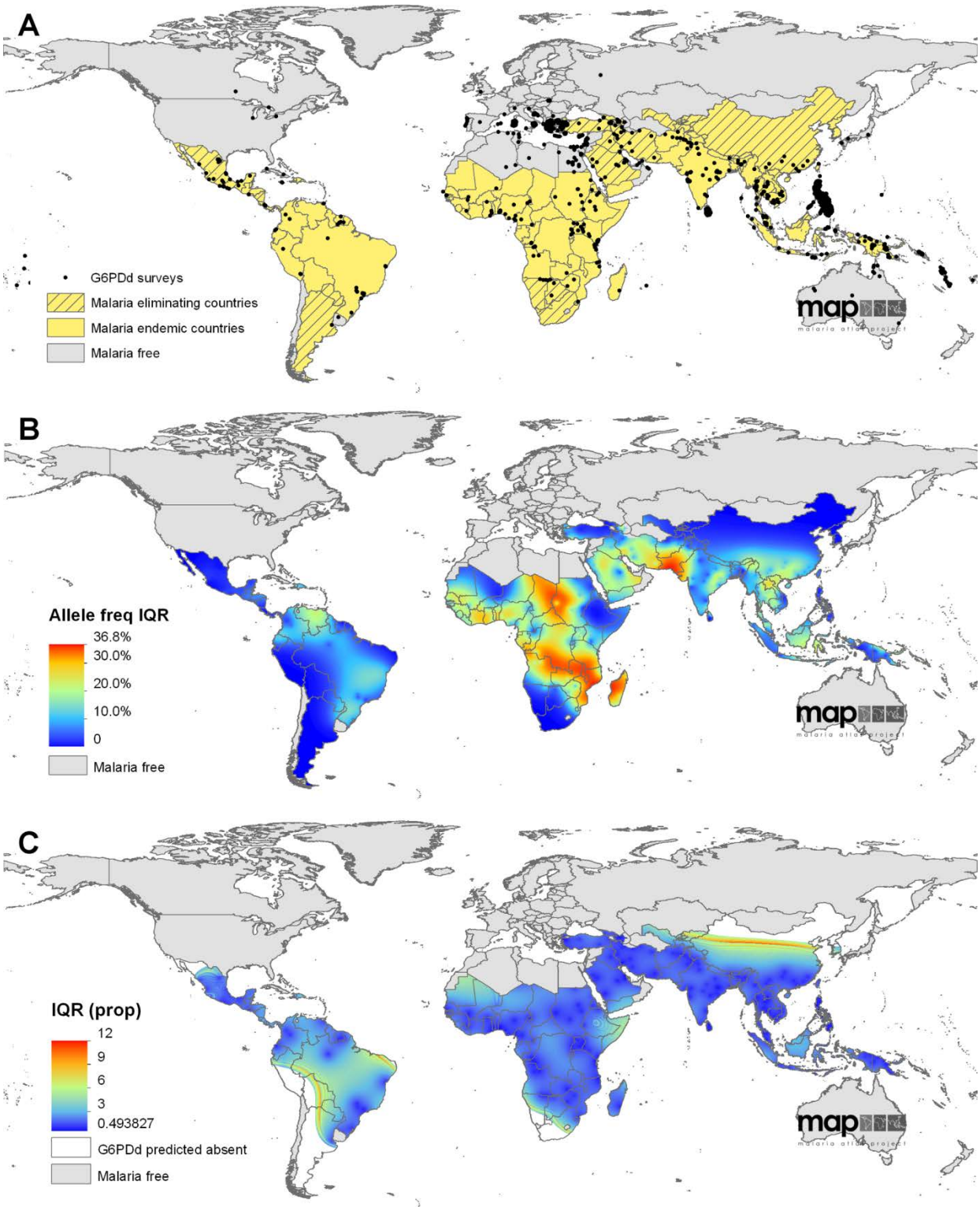


Figure S2.1. Prediction uncertainty metrics. Panel A shows the input surveys database. Panel B is the interquartile-range (IQR) of the PPD of male G6PDd prevalence, representing absolute values of map uncertainty. The map in Panel C was derived directly from the IQR map, adjusted to the median values.

References

1. Diggle PJ, Ribeiro PJ, Jr. (2007) Model-based Geostatistics: Springer.
2. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. Global estimates of sickle haemoglobin in newborns. The Lancet: In press.
3. Hardy GH (1908) Mendelian Proportions in a Mixed Population. Science 28: 49-50.
4. Weinberg W (1908) Über den nachweis der vererbung beim menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg 64: 368-382.
5. Patil AP, Gething PW, Piel FB, Hay SI (2011) Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol 27: 246-253.
6. Banerjee S, Carlin BP, Gelfand AE (2004) Hierachical modeling and analysis for spatial data. Monographs on Statistics and Applied Probability 101. Boca Randon, Florida, U.S.A.: Chapman & Hall / CRC Press LLC.
7. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. (2010) Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. Nat Commun 1: 104.
8. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. Princeton, New Jersey: Princeton University Press. 1088 p.
9. WHO Working Group (1989) Glucose-6-phosphate dehydrogenase deficiency. Bull World Health Organ 67: 601-611.
10. Nkhoma ET, Poole C, Vannappagari V, Hall SA, Beutler E (2009) The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. Blood Cells Mol Dis 42: 267-278.
11. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6: e1000048.
12. Gilks WR, Richardson S, Spiegelhalter D (1995) Markov chain Monte Carlo in Practice: Chapmand & Hall/CRC.
13. Gelman A, Carlin JB, Stern HS (2003) Bayesian Data Analysis. Texts in Statistical Science. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC.
14. Haario H, Saksman E, Tamminen J (2001) An Adaptive Metropolis Algorithm. Bernoulli 7: 223-242.
15. Patil A, Huard D, Fonnesbeck CJ (2010) PyMC: Bayesian Stochastic Modelling in Python. J Stat Softw 35: 1-81.

Protocol S3. Model validation procedures and results

To assess the model's predictive ability, a validation run was used to quantify the disparity between the model's predictions and a hold-out subset of the data [1,2,3,4]. The disparities were quantified to summarise general trends in the model's predictions.

S3.1 Creation of the validation datasets

To ensure that the validation metrics were representative of the overall model performance, it was necessary to ensure that the validation dataset was also representative of the full predicted surface. However, the dataset included some strong spatial clustering of surveys where small geographic areas had been heavily surveyed (e.g. the Philippines, Sri Lanka and the Kenyan coast). To ensure that data points from these areas (which are inherently more likely to be easier to predict, even with expected heterogeneity in allele frequency) were not over-represented in the validation procedure, a spatial-declustering sampling procedure was applied to the dataset. The details of this procedure have been previously described by Hay et al. [5]. This increased the probability that isolated samples will be included in the hold-out dataset, thus corresponding to areas which will be harder for the model to predict. A spatially-declustered 5% subset of the data (n=86) was therefore withheld from the validation MCMC, which was run with the remaining 95% (n=1,648) of the dataset (Figure S3.1A). It was not possible to use a hold-out sample larger than 5% with the declustering algorithm as the remaining dataset would have been too sparse to allow plausible mapping, thus making the validation process unrepresentative.

A second validation procedure was run with a randomly selected 10% subset of the data (n=173). Selecting the dataset at random meant that the geographical distribution of the hold-out dataset was more likely to be representative of the distribution of the overall dataset (Figure S3.1B).

S3.2 Model validation methodology

The Bayesian geostatistical G6PDd model was implemented in full with the thinned datasets (n=1,648 and n=1,561). This generated full PPDs, predictions from which could be compared with those in the hold-out datasets (from n=86 and n=173 locations, respectively), with differences summarised with simple statistical measures. Mean error was used to assess the model's overall bias, and the mean absolute error quantified the overall prediction accuracy as the average magnitude of errors in the predictions. These are fully described by Hay et al. [5]. A scatter plot was also generated as a visualisation of the correspondence between the predicted and actual values.

As well as considering the model's ability to predict point estimates, it was also important to assess the extent to which the model PPDs provided a suitable measure of uncertainty. A previously developed process [3,6,7] was used to test how well the validation sets of n=86 and n=173 PPDs captured the true uncertainty in the model output. Credible intervals (CIs) define a range of candidate values associated with a specified predicted probability of occurrence. Working through 100 progressively narrower CIs, from the 99% CI to the 1% CI, each was tested by computing the actual proportion of the hold-out prevalence observations that fell within the predicted CI. Plotting these actual proportions against each predicted CI level allowed the overall fidelity of the PPDs at the hold-out data locations to be assessed.

The bespoke dataset spatial declustering code and validation code are freely available from the MAP online code repository (https://github.com/malaria-atlas-project/generic-mbg).

S3.3 Validation results

The declustered validation data subset was a more stringent validation test than the randomly selected dataset, as the model's predictive ability was being preferentially tested in areas where fewer data were available to inform the predictions (Figure S3.2 and Table S3.1). The mean error values reveal a slight tendency to overestimate G6PDd frequency by 1.45% and 0.17% in the declustered and randomly selected validation runs respectively. So although the model has relatively low overall prediction bias, the magnitude of the variance between the predicted and observed prevalence can be more substantial, as indicated by mean absolute errors of 4.07 and 3.48%. These larger discrepancies may be due to the absence of nearby data points and to local heterogeneity in G6PDd values.

The probability-probability plots comparing predicted quantiles with observed coverage fractions (Figures 3.2C-D) show the fraction of the observations that were actually contained within each predicted CI. These plots show a good degree of fidelity in the predicted quantiles, indicating that the IQR measures used to represent uncertainty are a good representation of the model predictions.

Validation metric	Declustered 5%	Random 10%
Hold-out dataset	n=86	n=173
Thinned dataset	n=1,648	n=1,561
Mean error	1.45%	0.17%
Mean absolute error	4.07%	3.48%

Table S3.1. Summary of the validation statistics. Values are given in % frequency of G6PDd. Mean error statistics summarise the model's overall predictive bias (over/under-estimating); mean absolute errors indicate the magnitude of those errors.

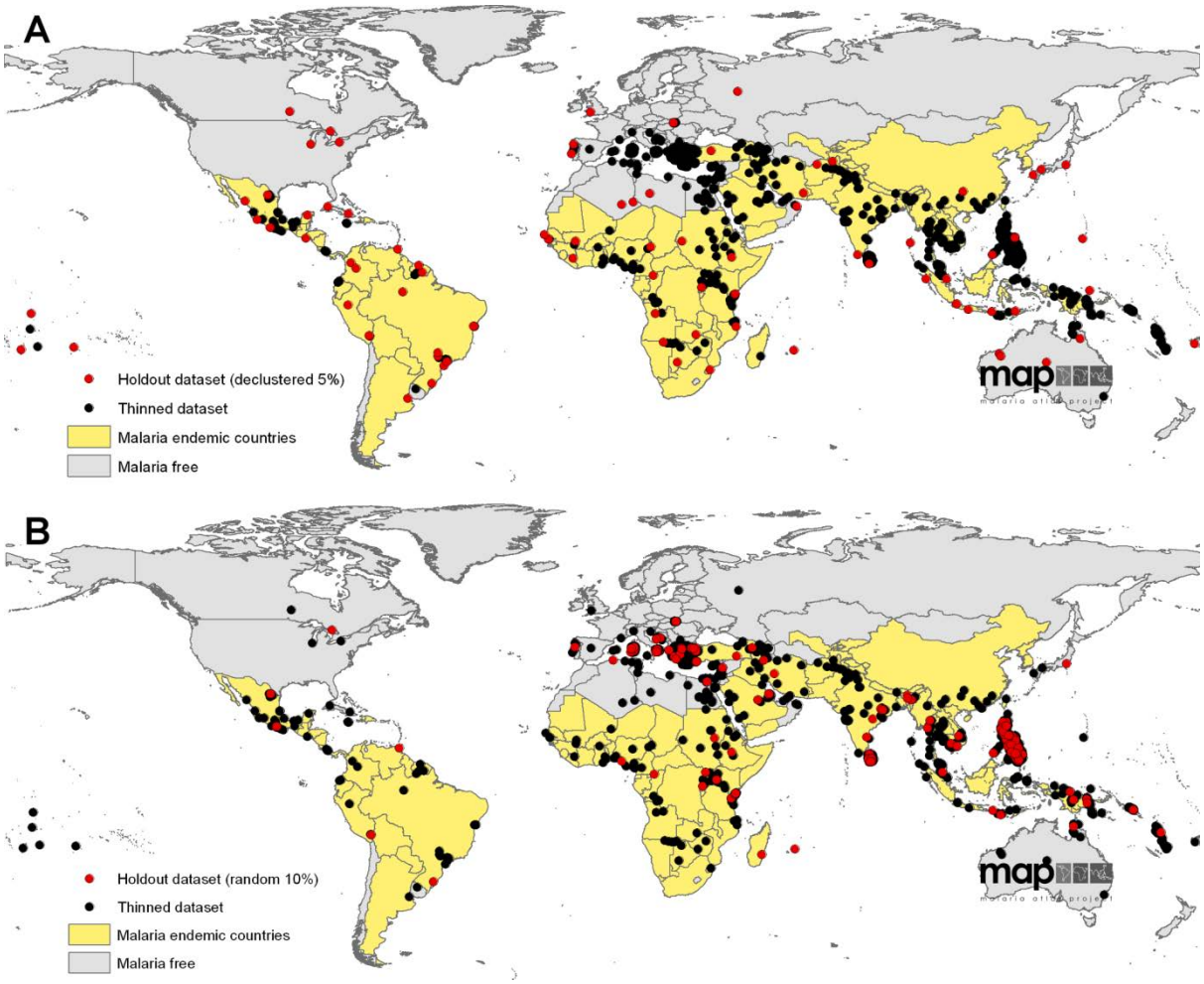


Figure S3.1. Distribution of the hold-out data subsets used for model validation. Red points are those held-out from the model run and used for validation of the model's predictions based on the remaining thinned dataset (black data points). Panel A shows the spatially declustered 5% hold-out dataset; Panel B shows the randomly selected 10% hold-out dataset.

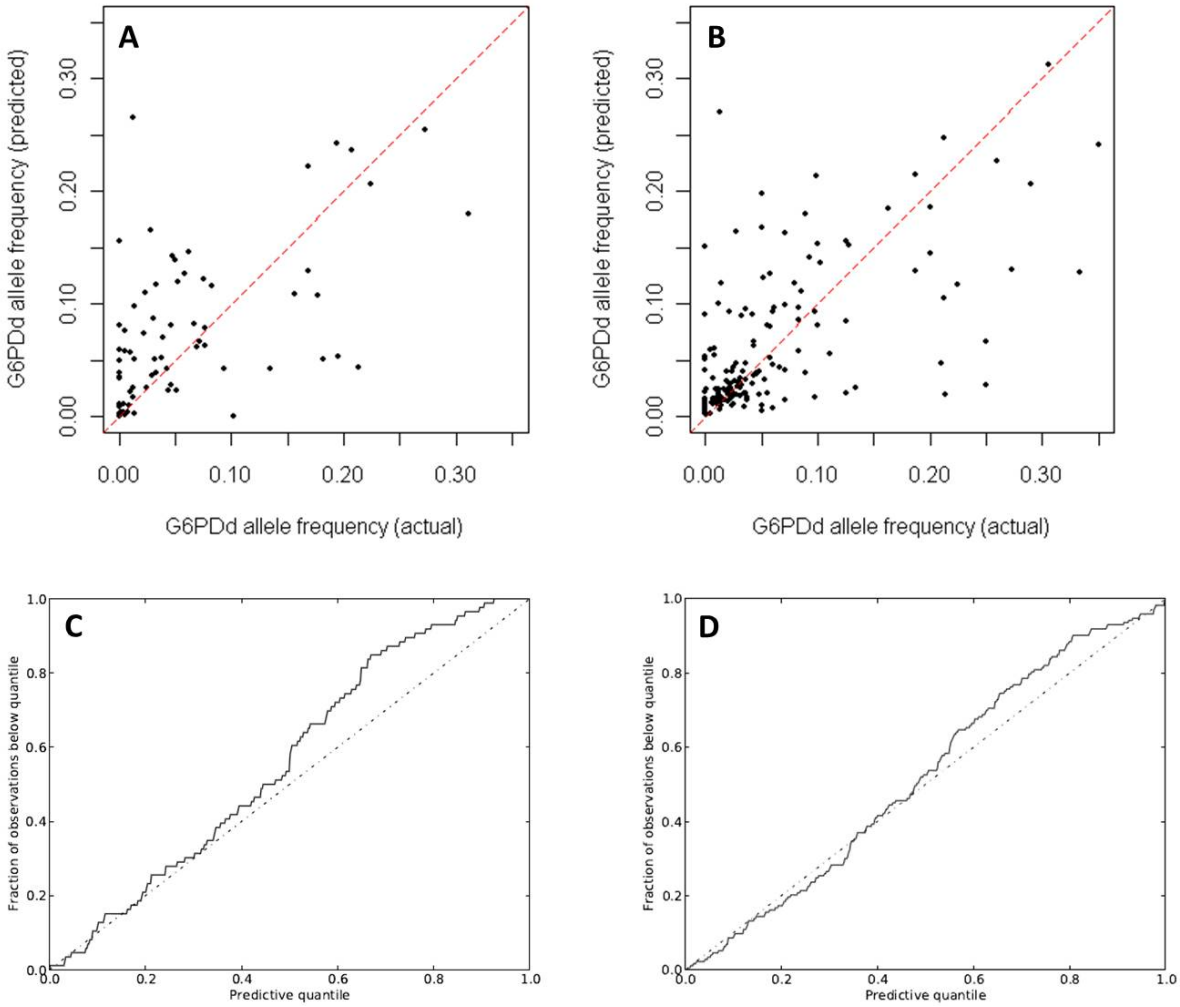


Figure S3.2. Model validation plots for both validation processes. Panels A and C correspond to the declustered 5% hold-out (n=86), B and D are the randomly selected 10% hold-out (n=173). Panels A and B are scatter plots of actual versus predicted point-values of G6PDd. Panels C and D show the probability-probability plots comparing predicted credible intervals with the actual percentage of true values lying in those intervals. The 1:1 line is also shown for reference.

References

1. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, et al. A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. Public Library of Science NTD: In press.
2. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, et al. (2011) A new world malaria map: *Plasmodium falciparum* endemicity in 2010. Malar J 10: 378.
3. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. Global estimates of sickle haemoglobin in newborns. The Lancet: In press.
4. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, et al. (2011) The global distribution of the Duffy blood group. Nat Commun 2: 266.
5. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6: e1000048.
6. Gething PW, Noor AM, Gikandi PW, Hay SI, Nixon MS, et al. (2008) Developing geostatistical space-time models to predict outpatient treatment burdens from incomplete national data. Geogr Anal 40: 167-188.
7. Moyeed RA, Papritz A (2002) An empirical comparison of kriging methods for nonlinear spatial point prediction. Mathematical Geology 34: 365-386.

Protocol S4. Demographic database and population estimate procedures

Both G6PDd prevalence and population density are heterogeneous in their distributions. To reflect both these sources of spatial variation in our estimates of G6PDd populations, we used high resolution population density grids to weight the G6PDd predictions and generate a single representative predictive posterior distribution (PPD) and summary prevalence estimates for each national and regional area of interest.

S4.1 GRUMP-beta human population surface

We used the Global Rural Urban Mapping Project (GRUMP) *beta* grids of population counts rescaled to a 5×5 km spatial resolution (from the original 1×1 km resolution), adjusted to UN national population total estimates for 2010 [1,2,3]. National-level sex-ratio population data were also taken from the UN World Population Prospects [4].

S4.2 Areal prediction procedures

To account for model uncertainty, the full MCMC tracefile, and not simply the mapped median values, was used to estimate aggregated population numbers affected by G6PDd [5]. The areal prediction model is fully described by Piel et al. [6], but a brief conceptual overview is given here. The model aimed to generate a summary description of G6PDd prevalence across each country and regional MEC aggregation of interest. The model sampled the MCMC tracefile repeatedly from the population-weighted selected sites across the areal region of interest (30,000 and 1,000 spatial points from the MEC regional and each national area, respectively). The model weighted the G6PDd estimate according to population density (at a 5×5 km grid resolution), thus generating a single summary PPD for the overall region. Each summary prevalence estimate was then related to the total population estimate for 2010 from that region, accounting for the national sex-ratio [4]. The PPDs allowed us to quantify the predictions' uncertainty as done with the summary maps.

As with the summary allele frequency map, summaries of the areal-prediction PPDs are given as median and IQR values (Supplementary Table S1 and S2). However, it is important to note that the sum of median values is not equivalent to the median of sums [5]. As a result, differences can be observed between the regional and the sums of national G6PDd population median estimates in the region.

Median values were also used to derive the homozygous frequencies directly from the areal allele frequency estimates (as opposed to from the full MCMC, as in for the male & all female estimates) as these corresponded to the estimates which would have been generated if we had used the full pixel-level method of estimation of homozygosity rates. Additionally, to generate Monte Carlo standard error estimates of the areal prediction model,

these were repeated ten times for each national spatial aggregate and five times for the regional aggregates. The variation around the mean of each set of repeated summary statistics (mean, median, 25% and 75% quartiles) is given as standard errors (SE) in Supplementary Table S2.

The code used to implement this analysis is freely available at https://github.com/malaria-atlas-project/.

References

1. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, et al. (2006) Determining global population distribution: methods, applications and data. Adv Parasitol 62: 119-156.
2. United Nations Department of Economics and Social Affairs (2008) World Population Prospects: the 2008 revision population database. New York: United Nationals Population Division (U.N.D.P.).
3. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6: e1000048.
4. United Nations Department of Economics and Social Affairs (2011) World Population Prospects, the 2010 Revision. New York: United Nationals Population Division.
5. Patil AP, Gething PW, Piel FB, Hay SI (2011) Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol 27: 246-253.
6. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. Global estimates of sickle haemoglobin in newborns. The Lancet: In press.

Protocol S5. Predicting the prevalence of G6PDd in females

S5.1 Overview of G6PDd in females

The implications of the G6PD gene's position on the X-chromosome on its genetics and inheritance mechanisms are particularly pertinent to estimates of the prevalence of G6PDd in heterozygous females. Reasons for these complexities have been previously reviewed [1-3], and are briefly discussed here.

S5.2 Heterozygous G6PDd expression and diagnosis

In a population at Hardy-Weinberg equilibrium where G6PDd allele frequency is <0.5, prevalence of heterozygous females will be higher than that of affected males [4] (frequencies >0.5 were not identified during our data searches apart from one instance where individuals tested was n<10). However, not all heterozygous females express a phenotypically significant deficiency; the difficulty of diagnosing deficiency in females, and extrapolating these figures to national levels has been previously discussed [5].

The gene's X-chromosome genetics means that heterozygotes have a mosaic-effect of G6PD expression, with two populations of red blood cells – those expressing the wild-type G6PD gene, and the rest bearing the deficiency [6]. This is due to the phenomenon of Lyonization whereby only one X-chromosome is actively expressed in each cell [7]. Lyonization is a random process and the resulting proportions of normal and deficient cells may deviate significantly away from the expected 50:50 ratio [3], leading some heterozygotes to have virtually normal expression, and others with expression levels comparable to female homozygotes (i.e. entirely deficient). Enzyme expression is therefore mixed, and inherently hard to predict based on genotype. Depending upon the ratio of the two cell populations, only a minority of heterozygotes will be clinically deficient and thus identified by the standard enzyme deficiency tests which usually detect deficiencies below about 20% of normal wild-type expression levels. Therefore, while deficiency may be readily diagnosed in homozygous females, heterozygote expression is variable and harder to determine.

As well as mosaic-expression of G6PD, the variability of expression among heterozygotes is partly attributable to the diversity of G6PD mutations, as many as 160 genetic variants have been described [1]. These cause a spectrum of deficiency levels, from <1% normal activity levels to having no discernible effect on enzyme activity [1]. As a result, the residual enzyme activity levels in females will reflect this range; the proportion of enzymatically deficient heterozygotes will therefore be spatially variable, determined by the local molecular variants [8].

Illustrating these issues, the diagnostic sensitivity of the WHO-recommended fluorescent spot test was found to be only 32% (with 99% specificity) for heterozygote detection [2]. Although a few biochemical tests have been described which are better suited to detecting heterozygosity, such as G6PD/6PDG and G6PD/PK ratio analysis, and the cytochemical G6PD staining assay [2,9], these are impractical for large-scale, field-based population surveys, and rarely used.

Whilst acknowledging that the common phenotypic tests have low sensitivity in detecting heterozygotes, and therefore that a large proportion of genotypically deficient individuals will escape identification, the aim of the population estimates presented in this paper is to represent the prevalence of individuals with clinically significant deficiency, as determined by the common phenotypic diagnostic tests.

S5.3 Overview of female data in the G6PD database

Of the 1,734 locations reporting rates of G6PDd in the database, 1,067 (62%) reported rates in females, with only 14 giving rates in females exclusively. The map in Figure 2A in the main paper shows their geographic distribution, and that of the male-only data. The prevalence values reported for females ranged from 0% (in 301 of the 1,067 female surveys) to 51% (from a survey of 100 females in the Solomon Islands [10]). Prevalence estimates in females are compared to those in males in Figure S5.1. As would be expected, there is a general correlation between prevalence in males and females in the same population (R^2 = 74.9%). Female estimates tend to be lower than those in corresponding male samples, though these patterns are heterogeneous.

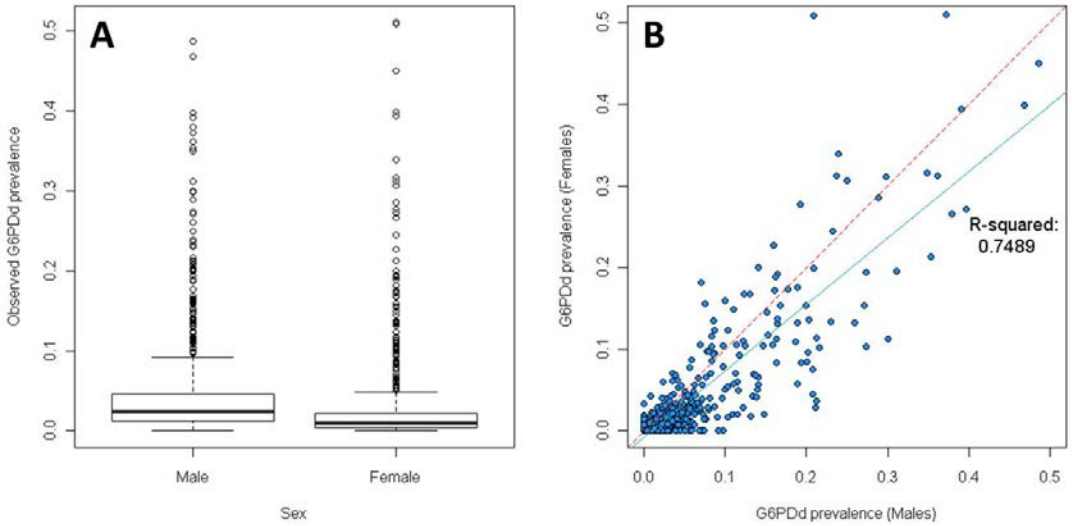


Figure S5.1. Observed G6PDd prevalence in males and females. Data points are from database raw data. Only surveys sampling both sexes and with sample sizes ≥50 are included in these plots (n = 725). Panel A summarises the overall deficiency prevalence estimates; Panel B plots the observed co-occurring male and female prevalence estimates. Panel B also shows the 1:1 line (dashed red line) and the line of best fit (continuous blue line).

For reasons previously discussed (Protocol S2), estimates of the prevalence of homozygosity can be calculated using simple rules of inheritance (q^2). It is therefore possible to subtract this predicted number of homozygous females, who are highly likely to be phenotypically deficient, from the overall observed number of deficient females, and compare prevalence of observed and expected genetically heterozygous females ($2pq$) (Figure S5.2). This provides an estimate of the observed deviance in heterozygous expression away from expected Hardy-Weinberg proportions. This deviance corresponds to h – an estimate of the proportion of heterozygous females who will not be diagnosed as phenotypically deficient (Protocol S2).

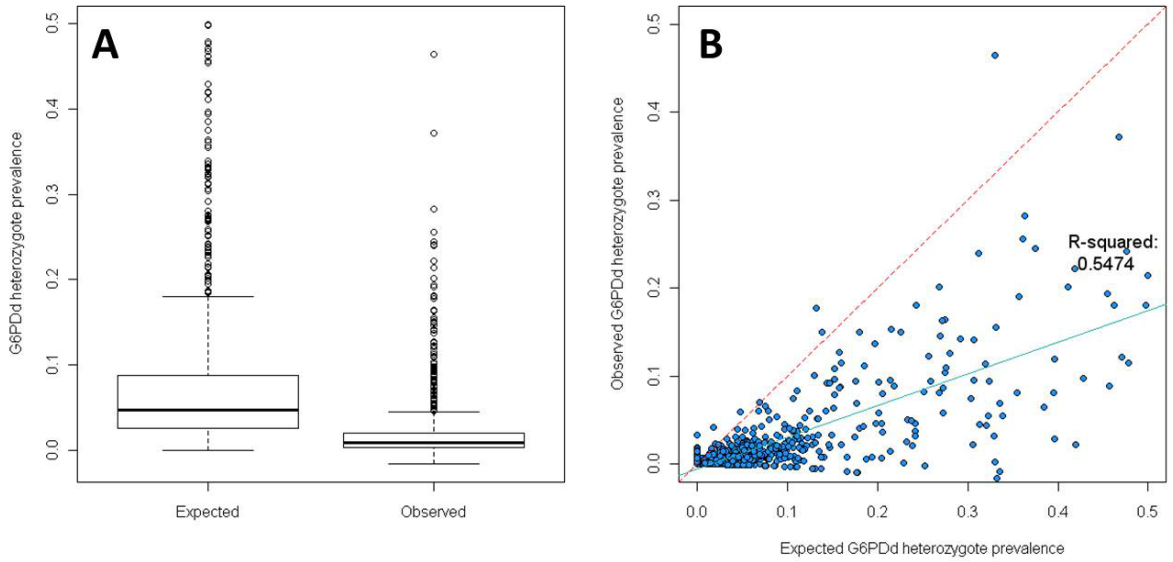


Figure S5.2. Comparison of expected and observed heterozygous rates. Expected heterozygous estimates are the calculated directly from the G6PDd frequency in males: $Expected = 2 \times q \times (1 - q)$. The observed heterozygous prevalence is estimated as the difference between the total observed G6PDd females and the expected proportion of homozygotes based on male G6PDd frequency: [$Observed = Female\ G6PDd\ prevalence - q^2$]. Only surveys sampling both sexes and with sample sizes ≥50 are included in these plots (n = 725). Panel B shows the 1:1 line (dashed red line) and the line of best fit (continuous blue line).

S5.4 Modelling phenotypic G6PDd prevalence in females.

When modelling the prevalence of G6PDd in females, the female population affected was considered as two distinct populations: homozygotes (whose prevalence was estimated directly from the population allele frequency (q^2), with the assumption that gene inheritance

is in Hardy-Weinberg equilibrium) and heterozygotes (a proportion of whom will be phenotypically deficient). The discordance between genetic heterozygosity and phenotypic deficiency complicates the modelling of heterozygous prevalence significantly [2,11]. In their most recent estimate of affected heterozygotes [5], the WHO imposed a fixed threshold of 10% of heterozygotes being phenotypically deficient, as did the more recent study by Nkhoma et al. [12]. However, the flexibility of the Bayesian model developed in this study meant that we could give the model the freedom to determine this threshold based directly on the input dataset. The expected variability in this threshold (Figure S5.3) indicated that imposing a single threshold was not the most appropriate method; instead, the observed variability supported the use of the spatially-variable deviance term, h, to moderate the number of heterozygotes likely to be deficient, as determined by the input dataset. A detailed description of these aspects of the model is given in Protocol S2.

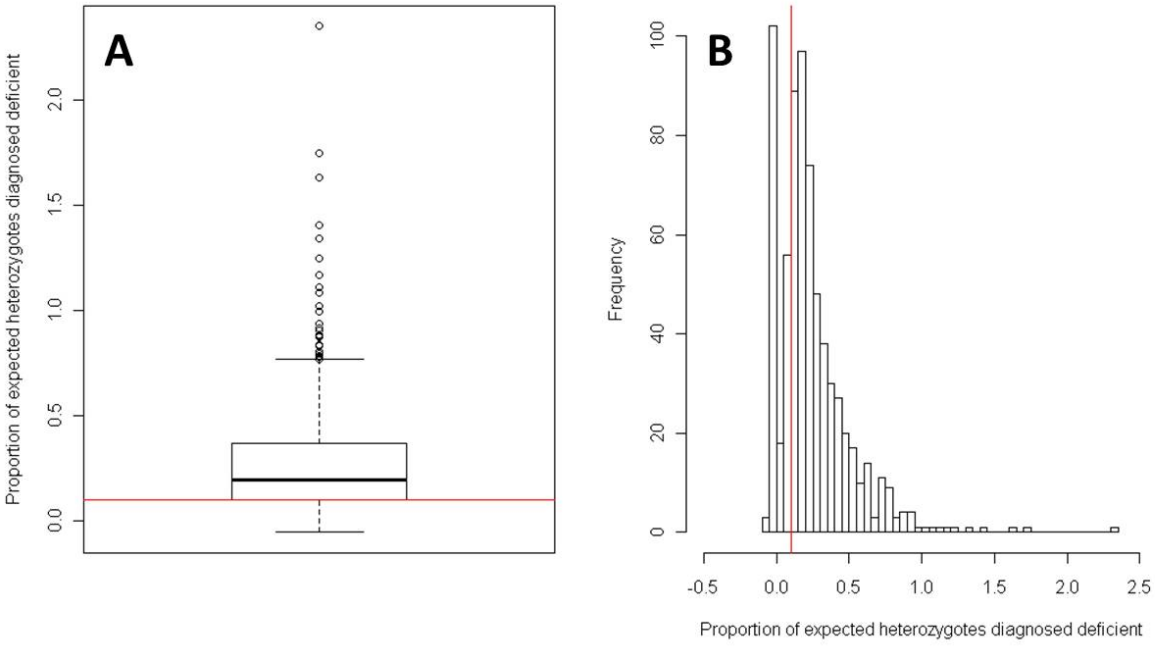


Figure S5.3. Proportion of expected heterozygote females diagnosed as phenotypically deficient. Expected heterozygotes are calculated from the corresponding prevalence of G6PDd males at each locality [q]. Plots show the proportion of genetic heterozygotes [$Expected = 2 \times q \times (1 - q)$] diagnosed deficient [$Observed = Female\ G6PDd\ prevalence - q^2$], in boxplot (Panel A) and histogram (Panel B) plots. The red lines show the fixed 10% threshold used by the WHO in their calculations of heterozygote diagnosis [5]. Summary statistics for the distribution shown: median 19.5% (IQR: 10.0% – 36.6%). Only surveys sampling both sexes and with sample sizes ≥50 are included in these plots (n = 725).

S5.5 Maps of G6PDd in females and population estimates

The spatial patterns of female G6PDd prevalence correspond to that for the allele frequency map (Figure 2B), and female population estimates, aggregated to national and regional areas, are given in Table 1 and Supplementary Table S1. The estimates of female homozygotes may be considered a highly conservative estimate of the overall number of females affected; in reality a proportion of heterozygotes will also share the G6PDd phenotype. Although G6PDd is frequently considered to be rare in females, it is clear from the assembled database and derived modelled population estimates that in many areas an important proportion of females will also be affected.

Figure S5.4 displays the estimated proportion of expected G6PDd heterozygotes (applying the Hardy-Weinberg equations to the median national summary estimates); while the WHO imposed a 10% threshold, the model developed here derived this threshold directly from the input data, with the flexibility for spatial variation. Across the predicted national predictions, a median proportion of 26.4% (IQR: 25.2-27.6) expected heterozygotes were predicted to be phenotypically deficient (Fig S5.4).

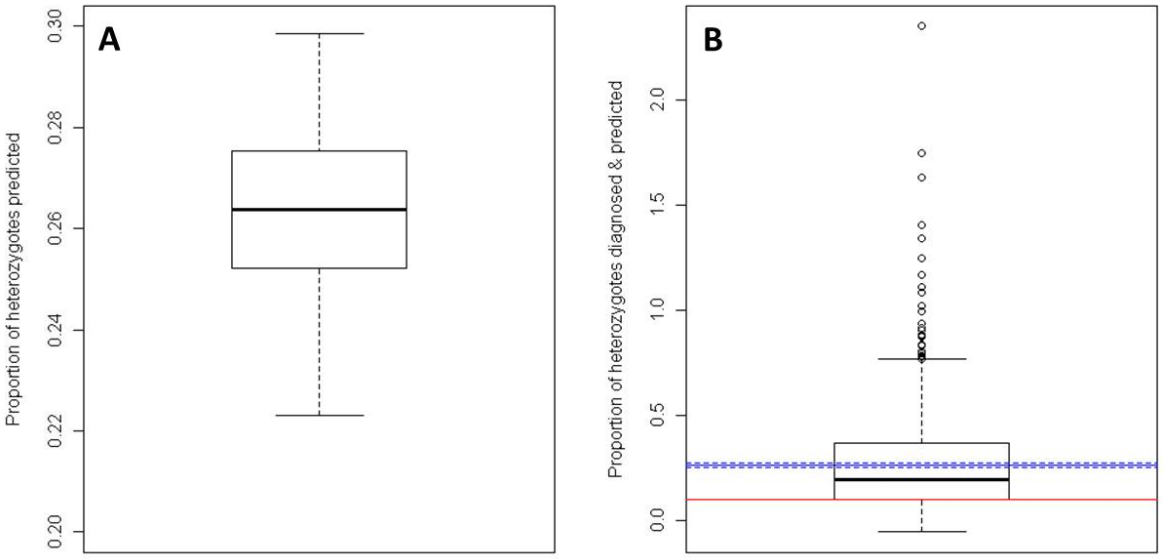


Figure S5.4. Proportion of expected heterozygote females predicted by the model national population estimates. Panel A plots the proportion of expected heterozygotes (derived from the model national estimates of allele frequency) predicted to be deficient; Panel B summarises the various heterozygote diagnostic cut-off limits: the boxplot data are of the input data points (n=725) [median value: 19.5% (IQR: 10.0-36.6)]; the red line represents the WHO [5] and Nkhoma et al. [12] 10% cut-off threshold, and the blue lines summarise the model's national heterozygote population predictions [Panel A: median value: 26.4% (solid line) (IQR: 25.2-27.6; dashed lines)]

S5.6 Improving the map of G6PDd in females

Although we present population estimates of affected females, these estimates are largely dependent upon the input dataset of surveys, and are therefore vulnerable to the same limitations as those original population survey diagnoses. While most methods are considered broadly equivalent for determining male hemizygote and female homozygote expression (Protocol S1), skewed X-allele inactivation [13] leaves the female heterozygous phenotype more ambiguous to diagnose and therefore harder to model [2,14]. This relationship is spatially variable, and ill-defined; presumably dependent upon the severity of the genetic variant, the Lyonization skew (which determines the relative proportions of deficient and normal genes in the mosaic of G6PD erythrocytic expression [6]), and, importantly, the sensitivity of the enzyme activity test used. It is well established that the proportion of heterozygotes identified from phenotypic tests is variable [15,16]. However, in terms of primaquine use, we make the assumption here that the proportion of deficient heterozygotes who are at risk of significant clinical side-effects will all be identified by the majority of tests; and that the border-line cases who may or may not be diagnosed, depending on the tests, will be at lower risk of severe reactions and thus not of main concern to policy-makers. However, this assumes that enzyme activity correlates with clinical severity, which has not yet been demonstrated [17].

The evidence-based, modelled estimates presented here for affected females are the first to be published, previous efforts having relied on a single pre-determined cut-off threshold [5,12]. However, these estimates constitute only a first attempt; standardising the diagnostic methods used, and developing better datasets and models for predicting phenotypic heterozygote expression from allele frequency data are necessary to improve the estimates for affected females.

Finally, further clinical data are required to understand the association between local enzyme activity level and risk of haemolysis, as well as relating this information to relative risk associated with different G6PDd allele frequencies. The model here relies on the diagnostic tests having appropriate cut-off limits, but the model could be refined if a better understanding of these relationships were available.

References

1. Mason PJ, Bautista JM, Gilsanz F (2007) G6PD deficiency: the genotype-phenotype association. Blood Reviews 21: 267-283.
2. Peters AL, Van Noorden CJ (2009) Glucose-6-phosphate dehydrogenase deficiency and malaria: cytochemical detection of heterozygous G6PD deficiency in women. J Histochem Cytochem 57: 1003-1011.
3. Beutler E (1994) G6PD deficiency. Blood 84: 3613-3636.
4. Luzzatto L (2010) Glucose-6-phosphate dehydrogenase (G6PD) deficiency. In: Warrell DA, Cox TM, Firth JD, editors. Oxford Textbook of Medicine. Oxford: Oxford University Press.
5. WHO Working Group (1989) Glucose-6-phosphate dehydrogenase deficiency. Bull World Health Organ 67: 601-611.
6. Beutler E, Yeh M, Fairbanks VF (1962) The normal human female as a mosaic of X-chromosome activity: studies using the gene for G-6-PD-deficiency as a marker. Proc Natl Acad Sci U S A 48: 9-16.
7. Lyon MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). Nature 190: 372-373.
8. Luzzatto L, Notaro R (2001) Malaria. Protecting against bad air. Science 293: 442-443.
9. Minucci A, Giardina B, Zuppi C, Capoluongo E (2009) Glucose-6-phosphate dehydrogenase laboratory assay: How, when, and why? IUBMB Life 61: 27-34.
10. Kuwahata M, Wijesinghe R, Ho MF, Pelecanos A, Bobogare A, et al. (2010) Population screening for glucose-6-phosphate dehydrogenase deficiencies in Isabel Province, Solomon Islands, using a modified enzyme assay on filter paper dried bloodspots. Malar J 9: 223.
11. Cappellini MD, Fiorelli G (2008) Glucose-6-phosphate dehydrogenase deficiency. Lancet 371: 64-74.
12. Nkhoma ET, Poole C, Vannappagari V, Hall SA, Beutler E (2009) The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. Blood Cells Mol Dis 42: 267-278.
13. Dobyns WB, Filauro A, Tomson BN, Chan AS, Ho AW, et al. (2004) Inheritance of most X-linked traits is not dominant or recessive, just X-linked. Am J Med Genet A 129A: 136-143.
14. Abdulrazzaq YM, Micallef R, Qureshi M, Dawodu A, Ahmed I, et al. (1999) Diversity in expression of glucose-6-phosphate dehydrogenase deficiency in females. Clin Genet 55: 13-19.
15. Ainoon O, Alawiyah A, Yu YH, Cheong SK, Hamidah NH, et al. (2003) Semiquantitative screening test for G6PD deficiency detects severe deficiency but misses a substantial proportion of partially-deficient females. Southeast Asian J Trop Med Public Health 34: 405-414.
16. Zaffanello M, Rugolotto S, Zamboni G, Gaudino R, Tato L (2004) Neonatal screening for glucose-6-phosphate dehydrogenase deficiency fails to detect heterozygote females. Eur J Epidemiol 19: 255-257.
17. Baird JK, Surjadjaja C (2011) Consideration of ethics in primaquine therapy against malaria transmission. Trends Parasitol 27: 11-16.

Supplementary Tables

Supplementary Table S1: National-level demographic metrics and G6PDd allele frequency and population estimates (in 1,000s)

Country	Total pop[1,2]	Sex-ratio[3]	Surveys	G6PDd allele freq (IQR)[4]	G6PDd males[1] (IQR)[4]	Homozygotes[1,5]	G6PDd females[1] (IQR)[4]
African MECs[6]							
Angola	18,994	98.1	2	15.3% (10.4 - 22.2)	1,435 (980 - 2,091)	223	947 (621 - 1,468)
Benin	9,219	97.3	0	23.0% (17.0 - 30.1)	1,044 (772 - 1,366)	246	706 (490 - 984)
Botswana†	1,977	101.7	2	3.6% (2.0 - 6.5)	36 (20 - 65)	1	19 (10 - 37)
Burkina Faso	16,250	98.5	0	9.4% (5.6 - 15.0)	757 (455 - 1,206)	72	452 (256 - 772)
Burundi	8,519	96.3	0	7.2% (3.3 - 15.2)	301 (137 - 635)	23	169 (73 - 394)
Cameroon	19,957	99.7	8	12.5% (9.9 - 15.5)	1,248 (990 - 1,543)	157	750 (568 - 967)
Cape Verde†	513	98.0	0	0.1% (0.0 - 0.5)	0 (0 - 1)	0	0 (0 - 1)
Central African Republic	4,506	97.1	0	9.2% (4.7 - 17.3)	203 (103 - 383)	19	126 (59 - 258)
Chad	11,509	98.9	0	13.4% (8.5 - 20.2)	767 (488 - 1,155)	104	496 (294 - 799)
Comoros	691	101.4	0	14.0% (5.8 - 30.4)	49 (20 - 106)	7	29 (10 - 71)
Congo	3,760	100.2	1	22.5% (17.3 - 29.6)	424 (326 - 557)	95	277 (205 - 387)
Cote d'Ivoire	21,571	103.9	0	15.0% (8.5 - 25.5)	1,654 (931 - 2,800)	240	991 (514 - 1,843)
Democratic Republic of the Congo	67,829	98.9	6	19.2% (14.7 - 25.1)	6,488 (4,974 - 8,459)	1,261	4,425 (3,270 - 6,066)
Djibouti	879	100.1	0	0.8% (0.3 - 2.7)	4 (1 - 12)	0	2 (1 - 6)
Equatorial Guinea	693	105.2	0	11.4% (6.1 - 20.1)	40 (22 - 72)	4	22 (11 - 43)
Eritrea	5,204	97.1	1	4.0% (2.7 - 6.1)	103 (68 - 157)	4	56 (36 - 88)
Ethiopia	84,996	99.1	6	1.0% (0.7 - 1.5)	422 (281 - 642)	4	218 (142 - 338)
Gabon	1,501	100.6	0	12.3% (6.0 - 23.4)	92 (45 - 176)	11	56 (25 - 117)
Ghana	24,339	103.6	2	19.6% (14.2 - 27)	2,429 (1,764 - 3,341)	460	1,498 (1,031 - 2,194)
Guinea	10,323	102.1	0	11.7% (7.4 - 18.8)	611 (385 - 982)	70	357 (214 - 621)
Guinea-Bissau	1,647	98.3	0	8.4% (4.4 - 15.3)	68 (36 - 124)	6	39 (19 - 76)
Kenya	40,847	99.8	45	11.3% (9.2 - 13.7)	2,310 (1,880 - 2,805)	262	1,377 (1,092 - 1,725)
Liberia	4,102	101.0	1	9.5% (5.2 - 16.9)	196 (107 - 348)	19	112 (58 - 216)
Madagascar	20,146	99.4	1	19.4% (11.5 - 30.3)	1,952 (1,154 - 3,046)	382	1,301 (711 - 2,212)
Malawi	15,690	100.1	0	20.8% (10.2 - 36.4)	1,629 (799 - 2,858)	338	1,067 (468 - 2,105)
Mali	13,362	99.8	3	12.2% (8.6 - 17.3)	813 (574 - 1,156)	99	499 (335 - 751)
Mauritania	3,359	101.0	0	9.6% (4.6 - 18.5)	162 (78 - 312)	15	95 (43 - 205)
Mayotte	199	99.5	0	12.4% (3.9 - 32.7)	12 (4 - 32)	2	7 (2 - 22)

Country	Total pop[1,2]	Sex-ratio[3]	Surveys	G6PDd allele freq (IQR)[4]	G6PDd males[1] (IQR)[4]	Homozygotes[1,5]	G6PDd females[1] (IQR)[4]
Mozambique	23,418	94.8	4	21.1% (14.7 - 29.8)	2,404 (1,670 - 3,394)	535	1,703 (1,112 - 2,587)
Namibia†	2,212	98.7	5	2.8% (1.8 - 4.6)	31 (20 - 51)	1	17 (11 - 29)
Niger	15,885	101.2	0	5.3% (2.6 - 10.3)	426 (211 - 819)	22	236 (111 - 497)
Nigeria	158,255	102.5	6	16.9% (14.1 - 20.2)	13,515 (11,317 - 16,185)	2,224	8,464 (6,898 - 10,477)
Rwanda	10,277	96.4	0	5.8% (3.3 - 10.1)	294 (169 - 509)	18	163 (90 - 298)
Sao Tome and Principe†	165	98.1	0	7.4% (2.3 - 20.8)	6 (2 - 17)	0	3 (1 - 11)
Senegal	12,866	98.4	1	15.1% (11.3 - 20.3)	966 (720 - 1,295)	148	598 (424 - 847)
Sierra Leone	5,837	95.5	0	7.9% (3.4 - 17.0)	226 (98 - 485)	19	132 (53 - 313)
Somalia	9,359	98.4	0	3.1% (1.2 - 7.7)	145 (56 - 356)	5	80 (29 - 215)
South Africa†	50,523	98.1	0	3.3% (1.8 - 6.2)	830 (438 - 1,563)	28	482 (242 - 960)
Sudan	43,204	101.5	8	15.3% (12.7 - 18.2)	3,322 (2,763 - 3,964)	499	2,118 (1,691 - 2,623)
Swaziland†	1,195	96.7	0	8.7% (4.6 - 15.5)	51 (27 - 91)	5	29 (15 - 57)
The Gambia	1,751	97.6	2	11.5% 8.1 - 15.9)	99 (70 - 138)	12	58 (39 - 85)
Togo	6,774	98.1	3	21.2% (16.7 - 26.6)	712 (560 - 893)	154	463 (346 - 616)
Uganda	33,798	99.9	12	14.5% (12.8 - 16.5)	2,457 (2,162 - 2,785)	358	1,468 (1,263 - 1,706)
United Republic of Tanzania	45,028	99.8	10	16.4% (11.9 - 22.3)	3,685 (2,671 - 5,019)	605	2,372 (1,643 - 3,431)
Zambia	13,254	100.5	1	21.0% (14.6 - 29.4)	1,393 (971 - 1,950)	291	923 (606 - 1,393)
Zimbabwe	12,645	97.2	2	14.8% (11.2 - 19.4)	924 (698 - 1,212)	141	586 (421 - 808)
American MECs[6]							
Argentina†	40,668	95.8	1	0.9% (0.5 - 1.6)	169 (98 - 313)	2	92 (51 - 181)
Belize†	313	97.3	0	2.2% (0.9 - 5.1)	3 (1 - 8)	0	2 (1 - 4)
Bolivia	9,995	99.5	0	0.2% (0.1 - 0.8)	11 (3 - 41)	0	6 (2 - 21)
Brazil	195,453	96.9	14	4.8% (3.6 - 6.5)	4,647 (3,501 - 6,213)	232	2,758 (1,994 - 3,897)
Colombia	46,305	96.8	2	4.9% (3.4 - 7.3)	1,118 (764 - 1,667)	57	638 (419 - 1,008)
Costa Rica†	4,640	103.1	3	0.4% (0.2 - 1.0)	9 (4 - 23)	0	4 (2 - 11)
Dominican Republic†	10,225	100.7	0	3.0% (0.9 - 10.0)	154 (44 - 511)	5	79 (22 - 288)
Ecuador	13,775	100.3	2	4.2% (2.4 - 7.5)	292 (166 - 519)	12	157 (87 - 294)
El Salvador†	6,201	90.5	1	3.3% (2.4 - 4.8)	98 (69 - 140)	4	56 (39 - 81)
French Guiana	231	100.3	0	0.7% (0.3 - 1.6)	1 (0 - 2)	0	0 (0 - 1)
Guatemala	14,378	95.1	0	2.7% (1.5 - 5.1)	189 (103 - 355)	5	102 (54 - 199)
Guyana	761	100.9	0	3.0% (1.4 - 6.4)	11 (5 - 25)	0	6 (3 - 13)
Haiti	10,188	98.4	0	5.2% (1.9 - 13.2)	261 (94 - 665)	14	141 (48 - 395)
Honduras	7,609	99.9	0	2.9% (1.5 - 5.8)	111 (55 - 219)	3	58 (28 - 118)

Country	Total pop[1,2]	Sex-ratio[3]	Surveys	G6PDd allele freq (IQR)[4]	G6PDd males[1] (IQR)[4]	Homozygotes[1,5]	G6PDd females[1] (IQR)[4]
Mexico†	110,568	97.3	25	1.0% (0.8 - 1.3)	555 (430 - 733)	6	291 (222 - 387)
Nicaragua†	5,822	97.9	0	1.5% (0.6 - 3.6)	43 (18 - 103)	1	22 (9 - 56)
Panama†	3,508	101.5	0	0.9% (0.4 - 2.5)	16 (6 - 44)	0	8 (3 - 22)
Paraguay†	6,462	101.8	0	3.2% (1.1 - 8.8)	105 (34 - 288)	3	54 (17 - 163)
Peru	29,493	100.4	3	0.2% (0.1 - 0.6)	33 (13 - 84)	0	17 (6 - 43)
Suriname	524	100.6	5	0.7% (0.4 - 1.3)	2 (1 - 3)	0	1 (1 - 2)
Venezuela	29,044	100.7	0	8.6% (4.0 - 18.0)	1,251 (583 - 2,617)	107	732 (316 - 1,701)
Eurasian MECs[6]							
Afghanistan	29,117	107.2	5	7.4% (5.6 - 9.8)	1,115 (845 - 1,470)	77	599 (436 - 833)
Azerbaijan†	8,932	97.8	17	10.2% (8.9 - 11.7)	452 (393 - 518)	47	267 (228 - 314)
Bangladesh	164,424	102.6	0	3.8% (2.4 - 5.9)	3,168 (2,002 - 4,942)	117	1,624 (1,007 - 2,636)
Bhutan†	723	112.5	0	5.9% (3.6 - 9.6)	23 (14 - 37)	1	11 (6 - 18)
Cambodia	15,056	95.8	11	14.3% (11.8 - 17.2)	1,055 (871 - 1,268)	158	653 (522 - 811)
China†	1,381,796	108.0	11	4.7% (3.5 - 6.8)	33,675 (25,014 - 48,717)	1,464	18,555 (13,427 - 27,859)
Dem People's Rep of Korea†	23,963	96.3	0	0.1% (0.0 - 0.4)	10 (2 - 42)	0	5 (1 - 22)
Georgia†	4,221	89.0	2	1.1% (0.7 - 1.7)	21 (14 - 33)	0	12 (8 - 19)
India	1,209,105	106.8	45	8.0% (6.9 - 9.3)	50,009 (43,246 - 57,985)	3,748	27,708 (23,452 - 32,947)
Indonesia	232,544	99.5	33	7.1% (5.3 - 9.4)	8,204 (6,180 - 10,901)	583	4,856 (3,484 - 6,756)
Iran (Islamic Republic of)†	75,084	103.0	12	11.8% (9.9 - 14.1)	4,510 (3,788 - 5,356)	518	2,661 (2,186 - 3,256)
Iraq†	31,443	100.6	1	10.6% (8.1 - 13.5)	1,669 (1,279 - 2,130)	176	970 (720 - 1,297)
Korea, Rep of†	48,517	99.4	0	0.2% (0.1 - 0.6)	50 (17 - 145)	0	24 (8 - 72)
Kyrgyzstan†	5,545	97.4	0	0.3% (0.1 - 1.2)	9 (3 - 33)	0	5 (1 - 17)
Lao People's Democratic Republic	6,434	99.6	1	15.6% (11.6 - 20.5)	500 (372 - 657)	78	315 (224 - 437)
Malaysia†	27,949	103.0	12	8.0% (6.6 - 9.6)	1,129 (942 - 1,359)	87	627 (506 - 782)
Myanmar	50,503	97.2	7	6.1% (4.1 - 9.3)	1,523 (1,028 - 2,309)	96	880 (577 - 1,384)
Nepal	29,950	98.4	0	5.3% (2.9 - 9.4)	786 (436 - 1,390)	42	434 (230 - 814)
Pakistan	189,875	103.4	8	15.0% (10.8 - 20.4)	14,495 (10,393 - 19,648)	2,106	8,900 (5,984 - 13,089)
Papua New Guinea	6,887	104.1	34	7.4% (6.0 - 9.3)	261 (212 - 325)	19	143 (114 - 185)
Philippines†	93,617	100.7	636	2.5% (2.4 - 2.5)	1,151 (1,117 - 1,187)	28	580 (556 - 604)
Saudi Arabia†	26,207	124.0	19	12.4% (10.4 - 14.9)	1,794 (1,511 - 2,161)	179	899 (741 - 1,113)
Solomon Islands†	536	107.1	41	22.3% (15.7 - 30.9)	62 (43 - 86)	13	38 (25 - 56)
Sri Lanka†	20,410	97.5	84	2.9% (2.6 - 3.3)	291 (258 - 331)	9	161 (140 - 188)
Tajikistan†	7,078	96.9	0	0.8% (0.4 - 1.9)	29 (12 - 66)	0	15 (6 - 34)

Country	Total pop[1,2]	Sex-ratio[3]	Surveys	G6PDd allele freq (IQR)[4]	G6PDd males[1] (IQR)[4]	Homozygotes[1,5]	G6PDd females[1] (IQR)[4]
Thailand†	68,141	96.7	20	13.6% (11.9 - 15.5)	4,544 (3,981 - 5,188)	638	2,830 (2,428 - 3,312)
Timor-Leste	1,168	104.0	0	5.0% (2.5 - 9.7)	29 (15 - 58)	1	15 (7 - 31)
Turkey†	75,699	99.5	79	3.8% (3.0 - 4.9)	1,437 (1,128 - 1,850)	55	783 (601 - 1,034)
Uzbekistan†	27,790	98.8	0	1.0% (0.4 - 2.4)	132 (53 - 330)	1	68 (27 - 179)
Vanuatu†	246	103.8	17	8.0% (6.9 - 9.3)	10 (9 - 12)	1	5 (4 - 6)
Viet Nam†	89,016	97.7	6	8.9% (6.0 - 13.9)	3,902 (2,625 - 6,102)	354	2,310 (1,488 - 3,922)
Yemen	24,324	101.3	0	4.6% (1.9 - 10.9)	565 (238 - 1,329)	26	308 (121 - 781)

[1] All population estimates are in 1,000s

[2] Source: GRUMP-adjusted UN population projected population estimates for 2010; World Population Prospects, the 2008 Revision.

[3] Number of males per 100 females in 2010. Source: United Nations Department of Economics and Social Affairs (2011) World Population Prospects, the 2010 Revision. http://esa.un.org/unpd/wpp/Excel-Data/population.htm.

[4] Interquartile range of the posterior predictive distribution for each G6PDd population estimate

[5] Homozygous female estimates are derived directly from the median allele frequency estimates, so are not given with modelled IQRs

[6] Malaria Endemic Countries (Protocol S1.6)

† Countries targetting malaria elimination

Supplementary Table S2: National areal prediction summary statistics and Monte Carlo standard errors (SE) for each model output

	Allele frequency (SE)				G6PDd male population (SE)				G6PDd female population (SE)			
Country	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**
African MECs												
Angola	17.4 (0.06) %	10.4 (0.10) %	15.3 (0.10) %	22.2 (0.10) %	1,635 (5.9)	980 (9.6)	1,435 (9.6)	2,091 (2.1)	1,143 (4.6)	621 (7.1)	947 (6.6)	1,468 (11.0)
Benin	24.0 (0.08) %	17.0 (0.11) %	23.0 (0.07) %	30.1 (0.13) %	1,091 (3.5)	772 (5.0)	1,044 (3.3)	1,366 (1.4)	769 (3.0)	490 (3.1)	706 (3.2)	984 (6.4)
Botswana	5.2 (0.04) %	2.0 (0.02) %	3.6 (0.04) %	6.5 (0.07) %	52 (0.3)	20 (0.2)	36 (0.4)	65 (0.1)	30 (0.2)	10 (0.1)	19 (0.2)	37 (0.4)
Burkina Faso	11.4 (0.08) %	5.6 (0.04) %	9.4 (0.06) %	15.0 (0.21) %	920 (6.7)	455 (3.3)	757 (4.9)	1,206 (1.2)	592 (5.0)	256 (2.4)	452 (3.1)	772 (13.0)
Burundi	11.3 (0.12) %	3.3 (0.07) %	7.2 (0.09) %	15.2 (0.22) %	473 (5.1)	137 (2.8)	301 (3.9)	635 (0.6)	314 (3.9)	73 (1.4)	169 (2.5)	394 (6.2)
Cameroon	13.0 (0.04) %	9.9 (0.05) %	12.5 (0.05) %	15.5 (0.06) %	1,297 (4.3)	990 (5.2)	1,248 (4.8)	1,543 (1.5)	798 (3.8)	568 (3.0)	750 (3.8)	967 (5.6)
Cape Verde	1.0 (0.03) %	0.0 (0.00) %	0.1 (0.00) %	0.5 (0.01) %	3 (0.1)	0 (0.0)	0 (0.0)	1 (0.0)	2 (0.1)	0 (0.0)	0 (0.0)	1 (0.0)
Central African Rep	12.5 (0.13) %	4.7 (0.09) %	9.2 (0.12) %	17.3 (0.22) %	277 (2.8)	103 (2.0)	203 (2.6)	383 (0.4)	192 (2.2)	59 (1.2)	126 (1.9)	258 (4.2)
Chad	15.4 (0.13) %	8.5 (0.09) %	13.4 (0.15) %	20.2 (0.25) %	882 (7.2)	488 (5.3)	767 (8.3)	1,155 (1.2)	605 (6.1)	294 (3.8)	496 (6.5)	799 (12.6)
Comoros	20.6 (0.15) %	5.8 (0.14) %	14.0 (0.17) %	30.4 (0.24) %	72 (0.5)	20 (0.5)	49 (0.6)	106 (0.1)	50 (0.4)	10 (0.2)	29 (0.4)	71 (0.8)
Congo	24.2 (0.12) %	17.3 (0.14) %	22.5 (0.12) %	29.6 (0.17) %	456 (2.2)	326 (2.6)	424 (2.3)	557 (0.6)	312 (1.7)	205 (1.7)	277 (1.9)	387 (2.7)
Cote d'Ivoire	18.6 (0.19) %	8.5 (0.10) %	15.0 (0.23) %	25.5 (0.31) %	2,045 (21.3)	931 (11.3)	1,654 (24.8)	2,800 (2.8)	1,358 (16.5)	514 (7.9)	991 (16.9)	1,843 (26.3)
Dem Rep of the Congo	20.5 (0.11) %	14.7 (0.12) %	19.2 (0.11) %	25.1 (0.15) %	6,928 (35.8)	4,974 (41.5)	6,488 (37.0)	8,459 (8.5)	4,891 (28.5)	3,270 (30.6)	4,425 (33.1)	6,066 (37.1)
Djibouti	2.9 (0.04) %	0.3 (0.01) %	0.8 (0.02) %	2.7 (0.06) %	13 (0.2)	1 (0.0)	4 (0.1)	12 (0.0)	7 (0.1)	1 (0.0)	2 (0.0)	6 (0.1)
Equatorial Guinea	14.8 (0.09) %	6.1 (0.09) %	11.4 (0.12) %	20.1 (0.09) %	53 (0.3)	22 (0.3)	40 (0.4)	72 (0.1)	33 (0.2)	11 (0.2)	22 (0.2)	43 (0.4)
Eritrea	4.8 (0.04) %	2.7 (0.03) %	4.0 (0.04) %	6.1 (0.07) %	124 (1.1)	68 (0.8)	103 (1.1)	157 (0.2)	70 (0.7)	36 (0.4)	56 (0.6)	88 (1.1)
Ethiopia	1.2 (0.01) %	0.7 (0.01) %	1.0 (0.01) %	1.5 (0.01) %	515 (4.9)	281 (2.4)	422 (3.0)	642 (0.6)	275 (3.0)	142 (1.3)	218 (1.6)	338 (3.2)
Gabon	16.8 (0.15) %	6.0 (0.06) %	12.3 (0.18) %	23.4 (0.31) %	126 (1.2)	45 (0.4)	92 (1.3)	176 (0.2)	87 (1.0)	25 (0.2)	56 (0.9)	117 (1.6)
Ghana	21.5 (0.08) %	14.2 (0.11) %	19.6 (0.09) %	27.0 (0.10) %	2,664 (10.4)	1,764 (13.4)	2,429 (11.0)	3,341 (3.3)	1,727 (8.1)	1,031 (7.5)	1,498 (7.0)	2,194 (9.3)
Guinea	14.4 (0.08) %	7.4 (0.06) %	11.7 (0.09) %	18.8 (0.15) %	749 (3.9)	385 (3.3)	611 (4.7)	982 (1.0)	479 (3.0)	214 (1.7)	357 (3.6)	621 (6.8)
Guinea-Bissau	11.4 (0.08) %	4.4 (0.06) %	8.4 (0.07) %	15.3 (0.21) %	93 (0.7)	36 (0.5)	68 (0.6)	124 (0.1)	59 (0.5)	19 (0.3)	39 (0.5)	76 (1.0)
Kenya	11.7 (0.05) %	9.2 (0.06) %	11.3 (0.05) %	13.7 (0.08) %	2,394 (10.0)	1,880 (12.3)	2,310 (9.6)	2,805 (2.8)	1,454 (6.4)	1,092 (7.7)	1,377 (6.1)	1,725 (9.7)
Liberia	12.7 (0.12) %	5.2 (0.05) %	9.5 (0.09) %	16.9 (0.20) %	261 (2.5)	107 (1.1)	196 (1.8)	348 (0.3)	166 (1.9)	58 (0.6)	112 (1.3)	216 (3.4)
Madagascar	22.3 (0.11) %	11.5 (0.12) %	19.4 (0.16) %	30.3 (0.20) %	2,241 (11.4)	1,154 (12.4)	1,952 (16.6)	3,046 (3.0)	1,626 (10.3)	711 (8.7)	1,301 (12.5)	2,212 (17.7)
Malawi	25.4 (0.17) %	10.2 (0.18) %	20.8 (0.31) %	36.4 (0.25) %	1,991 (13.0)	799 (14.3)	1,629 (24.1)	2,858 (2.9)	1,471 (11.1)	468 (9.7)	1,067 (17.7)	2,105 (16.9)
Mali	13.7 (0.08) %	8.6 (0.07) %	12.2 (0.07) %	17.3 (0.12) %	918 (5.0)	574 (4.9)	813 (4.8)	1,156 (1.2)	592 (3.8)	335 (2.6)	499 (3.6)	751 (6.6)
Mauritania	13.4 (0.09) %	4.6 (0.06) %	9.6 (0.08) %	18.5 (0.19) %	226 (1.5)	78 (1.1)	162 (1.3)	312 (0.3)	153 (1.4)	43 (0.7)	95 (1.0)	205 (2.6)
Mayotte	21.6 (0.21) %	3.9 (0.10) %	12.4 (0.24) %	32.7 (0.57) %	21 (0.2)	4 (0.1)	12 (0.2)	32 (0.0)	16 (0.2)	2 (0.1)	7 (0.2)	22 (0.5)
Mozambique	23.1 (0.09) %	14.7 (0.10) %	21.1 (0.15) %	29.8 (0.20) %	2,631 (9.7)	1,670 (11.8)	2,404 (16.9)	3,394 (3.4)	1,972 (9.0)	1,112 (9.3)	1,703 (15.4)	2,587 (20.6)
Namibia	4.0 (0.05) %	1.8 (0.03) %	2.8 (0.04) %	4.6 (0.07) %	44 (0.6)	20 (0.4)	31 (0.4)	51 (0.1)	26 (0.3)	11 (0.2)	17 (0.2)	29 (0.4)

	Allele frequency (SE)				G6PDd male population (SE)				G6PDd female population (SE)			
Country	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**
Niger	7.7 (0.08) %	2.6 (0.05) %	5.3 (0.06) %	10.3 (0.16) %	617 (6.0)	211 (3.7)	426 (4.7)	819 (0.8)	387 (4.7)	111 (2.1)	236 (2.7)	497 (8.5)
Nigeria	17.5 (0.07) %	14.1 (0.08) %	16.9 (0.07) %	20.2 (0.10) %	14,021 (52.3)	11,317 (63.9)	13,515 (54.2)	16,185 (16.2)	8,920 (36.9)	6,898 (44.0)	8,464 (39.2)	10,477 (51.0)
Rwanda	7.7 (0.04) %	3.3 (0.03) %	5.8 (0.06) %	10.1 (0.07) %	389 (2.1)	169 (1.5)	294 (2.8)	509 (0.5)	234 (1.6)	90 (0.6)	163 (1.7)	298 (2.8)
Sao Tome and Principe	15.0 (0.23) %	2.3 (0.05) %	7.4 (0.18) %	20.8 (0.57) %	12 (0.2)	2 (0.0)	6 (0.1)	17 (0.0)	9 (0.2)	1 (0.0)	3 (0.1)	11 (0.3)
Senegal	16.4 (0.05) %	11.3 (0.05) %	15.1 (0.06) %	20.3 (0.10) %	1,046 (3.4)	720 (3.0)	966 (4.1)	1,295 (1.3)	675 (2.7)	424 (1.6)	598 (2.3)	847 (5.2)
Sierra Leone	12.5 (0.13) %	3.4 (0.04) %	7.9 (0.07) %	17.0 (0.17) %	356 (3.6)	98 (1.1)	226 (2.0)	485 (0.5)	247 (3.3)	53 (0.7)	132 (1.6)	313 (4.8)
Somalia	6.1 (0.08) %	1.2 (0.03) %	3.1 (0.07) %	7.7 (0.16) %	282 (3.6)	56 (1.3)	145 (3.4)	356 (0.4)	186 (2.5)	29 (0.7)	80 (2.2)	215 (5.1)
South Africa	5.0 (0.04) %	1.8 (0.04) %	3.3 (0.04) %	6.2 (0.07) %	1,244 (9.4)	438 (9.0)	830 (10.8)	1,563 (1.6)	793 (6.7)	242 (4.9)	482 (5.6)	960 (11.9)
Sudan	15.6 (0.06) %	12.7 (0.06) %	15.3 (0.06) %	18.2 (0.08) %	3,402 (13.6)	2,763 (12.8)	3,322 (13.1)	3,964 (4.0)	2,202 (10.6)	1,691 (9.5)	2,118 (9.2)	2,623 (15.5)
Swaziland	11.6 (0.09) %	4.6 (0.05) %	8.7 (0.08) %	15.5 (0.12) %	68 (0.5)	27 (0.3)	51 (0.5)	91 (0.1)	44 (0.4)	15 (0.2)	29 (0.3)	57 (0.5)
The Gambia	12.6 (0.09) %	8.1 (0.06) %	11.5 (0.08) %	15.9 (0.16) %	109 (0.8)	70 (0.5)	99 (0.7)	138 (0.1)	67 (0.6)	39 (0.3)	58 (0.4)	85 (1.0)
Togo	22.1 (0.08) %	16.7 (0.09) %	21.2 (0.13) %	26.6 (0.12) %	741 (2.7)	560 (3.2)	712 (4.3)	893 (0.9)	499 (2.3)	346 (2.5)	463 (3.0)	616 (3.5)
Uganda	14.8 (0.02) %	12.8 (0.03) %	14.5 (0.03) %	16.5 (0.03) %	2,493 (3.5)	2,162 (5.1)	2,457 (4.8)	2,785 (2.8)	1,504 (2.2)	1,263 (2.8)	1,468 (2.8)	1,706 (3.5)
United Rep of Tanzania	17.9 (0.08) %	11.9 (0.10) %	16.4 (0.09) %	22.3 (0.16) %	4,027 (18.6)	2,671 (21.4)	3,685 (21.0)	5,019 (5.0)	2,716 (15.0)	1,643 (14.4)	2,372 (16.8)	3,431 (33.8)
Zambia	22.9 (0.13) %	14.6 (0.15) %	21.0 (0.11) %	29.4 (0.18) %	1,520 (8.8)	971 (9.7)	1,393 (7.1)	1,950 (2.0)	1,069 (7.6)	606 (6.2)	923 (6.5)	1,393 (10.2)
Zimbabwe	15.9 (0.05) %	11.2 (0.04) %	14.8 (0.07) %	19.4 (0.09) %	988 (3.2)	698 (2.4)	924 (4.3)	1,212 (1.2)	650 (2.6)	421 (1.5)	586 (3.6)	808 (4.3)
American MECs												
Argentina	1.3 (0.03) %	0.5 (0.01) %	0.9 (0.02) %	1.6 (0.04) %	261 (5.1)	98 (2.1)	169 (3.5)	313 (0.3)	159 (3.3)	51 (1.1)	92 (2.1)	181 (5.1)
Belize	4.1 (0.04) %	0.9 (0.01) %	2.2 (0.03) %	5.1 (0.07) %	6 (0.1)	1 (0.0)	3 (0.1)	8 (0.0)	4 (0.1)	1 (0.0)	2 (0.0)	4 (0.1)
Bolivia	1.0 (0.02) %	0.1 (0.00) %	0.2 (0.01) %	0.8 (0.02) %	51 (1.1)	3 (0.1)	11 (0.3)	41 (0.0)	30 (0.8)	2 (0.0)	6 (0.2)	21 (0.6)
Brazil	5.4 (0.04) %	3.6 (0.02) %	4.8 (0.04) %	6.5 (0.06) %	5,153 (38.8)	3,501 (21.3)	4,647 (40.4)	6,213 (6.2)	3,203 (28.3)	1,994 (16.6)	2,758 (25.1)	3,897 (48.0)
Colombia	5.8 (0.04) %	3.4 (0.02) %	4.9 (0.04) %	7.3 (0.06) %	1,311 (8.3)	764 (5.4)	1,118 (9.4)	1,667 (1.7)	798 (6.3)	419 (2.3)	638 (5.1)	1,008 (11.5)
Costa Rica	0.8 (0.01) %	0.2 (0.00) %	0.4 (0.00) %	1.0 (0.01) %	20 (0.2)	4 (0.1)	9 (0.1)	23 (0.0)	10 (0.1)	2 (0.0)	4 (0.0)	11 (0.1)
Dominican Republic	8.6 (0.14) %	0.9 (0.03) %	3.0 (0.08) %	10.0 (0.31) %	443 (7.1)	44 (1.5)	154 (4.2)	511 (0.5)	296 (5.3)	22 (0.7)	79 (2.3)	288 (10.2)
Ecuador	6.0 (0.03) %	2.4 (0.03) %	4.2 (0.05) %	7.5 (0.06) %	411 (2.1)	166 (2.1)	292 (3.2)	519 (0.5)	240 (1.5)	87 (1.1)	157 (1.9)	294 (2.0)
El Salvador	3.8 (0.01) %	2.4 (0.02) %	3.3 (0.02) %	4.8 (0.01) %	112 (0.3)	69 (0.5)	98 (0.6)	140 (0.1)	65 (0.2)	39 (0.3)	56 (0.3)	81 (0.2)
French Guiana	1.4 (0.02) %	0.3 (0.01) %	0.7 (0.01) %	1.6 (0.02) %	2 (0.0)	0 (0.0)	1 (0.0)	2 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	1 (0.0)
Guatemala	4.0 (0.04) %	1.5 (0.02) %	2.7 (0.03) %	5.1 (0.07) %	280 (2.8)	103 (1.1)	189 (2.4)	355 (0.4)	163 (1.9)	54 (0.6)	102 (1.4)	199 (2.8)
Guyana	5.3 (0.07) %	1.4 (0.02) %	3.0 (0.06) %	6.4 (0.12) %	20 (0.3)	5 (0.1)	11 (0.2)	25 (0.0)	12 (0.2)	3 (0.1)	6 (0.1)	13 (0.3)
Haiti	10.3 (0.14) %	1.9 (0.05) %	5.2 (0.12) %	13.2 (0.24) %	522 (6.9)	94 (2.4)	261 (6.1)	665 (0.7)	349 (5.5)	48 (1.2)	141 (3.4)	395 (9.6)
Honduras	4.6 (0.03) %	1.5 (0.02) %	2.9 (0.02) %	5.8 (0.07) %	176 (1.2)	55 (0.9)	111 (0.8)	219 (0.2)	101 (0.8)	28 (0.5)	58 (0.5)	118 (1.7)
Mexico	1.1 (0.01) %	0.8 (0.01) %	1.0 (0.01) %	1.3 (0.01) %	619 (5.7)	430 (5.0)	555 (6.1)	733 (0.7)	327 (3.2)	222 (2.5)	291 (3.3)	387 (4.4)
Nicaragua	3.0 (0.06) %	0.6 (0.01) %	1.5 (0.02) %	3.6 (0.07) %	88 (1.7)	18 (0.3)	43 (0.7)	103 (0.1)	51 (1.1)	9 (0.2)	22 (0.3)	56 (1.2)

	Allele frequency (SE)				G6PDd male population (SE)				G6PDd female population (SE)			
Country	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**
Panama	2.3 (0.05) %	0.4 (0.01) %	0.9 (0.02) %	2.5 (0.07) %	41 (0.9)	6 (0.1)	16 (0.3)	44 (0.0)	23 (0.6)	3 (0.1)	8 (0.2)	22 (0.7)
Paraguay	7.5 (0.09) %	1.1 (0.03) %	3.2 (0.06) %	8.8 (0.21) %	244 (2.8)	34 (0.8)	105 (1.9)	288 (0.3)	157 (2.2)	17 (0.4)	54 (0.9)	163 (3.5)
Peru	0.5 (0.01) %	0.1 (0.00) %	0.2 (0.00) %	0.6 (0.01) %	79 (1.2)	13 (0.3)	33 (0.5)	84 (0.1)	43 (0.7)	6 (0.1)	17 (0.4)	43 (0.8)
Suriname	1.1 (0.01) %	0.4 (0.01) %	0.7 (0.01) %	1.3 (0.02) %	3 (0.0)	1 (0.0)	2 (0.0)	3 (0.0)	1 (0.0)	1 (0.0)	1 (0.0)	2 (0.0)
Venezuela	13.3 (0.12) %	4.0 (0.06) %	8.6 (0.09) %	18.0 (0.32) %	1,936 (17.8)	583 (9.4)	1,251 (13.2)	2,617 (2.6)	1,325 (15.0)	316 (5.8)	732 (10.4)	1,701 (35.4)
Eurasian MECs												
Afghanistan	8.0 (0.06) %	5.6 (0.05) %	7.4 (0.05) %	9.8 (0.08) %	1,209 (9.5)	845 (7.2)	1,115 (8.2)	1,470 (1.5)	681 (6.0)	436 (3.6)	599 (5.2)	833 (7.8)
Argentina	1.3 (0.03) %	0.5 (0.01) %	0.9 (0.02) %	1.6 (0.04) %	261 (5.1)	98 (2.1)	169 (3.5)	313 (0.3)	159 (3.3)	51 (1.1)	92 (2.1)	181 (5.1)
Azerbaijan	10.4 (0.08) %	8.9 (0.07) %	10.2 (0.09) %	11.7 (0.09) %	461 (3.6)	393 (3.2)	452 (4.0)	518 (0.5)	275 (2.4)	228 (2.2)	267 (2.6)	314 (2.5)
Bangladesh	4.7 (0.03) %	2.4 (0.01) %	3.8 (0.02) %	5.9 (0.04) %	3,898 (25.1)	2,002 (12.1)	3,168 (20.3)	4,942 (4.9)	2,110 (15.5)	1,007 (7.0)	1,624 (12.3)	2,636 (19.4)
Bhutan	7.4 (0.05) %	3.6 (0.04) %	5.9 (0.06) %	9.6 (0.11) %	28 (0.2)	14 (0.1)	23 (0.2)	37 (0.0)	15 (0.1)	6 (0.1)	11 (0.1)	18 (0.2)
Cambodia	14.8 (0.03) %	11.8 (0.05) %	14.3 (0.04) %	17.2 (0.05) %	1,089 (2.1)	871 (3.4)	1,055 (3.3)	1,268 (1.3)	685 (1.6)	522 (2.2)	653 (1.7)	811 (3.2)
China	5.7 (0.04) %	3.5 (0.04) %	4.7 (0.05) %	6.8 (0.05) %	41,186 (314.1)	25,014 (273.3)	33,675 (350.2)	48,717 (48.7)	23,765 (192.0)	13,427 (164.0)	18,555 (204.4)	27,859 (222.7)
Dem People's Rep Korea	0.7 (0.02) %	0.0 (0.00) %	0.1 (0.00) %	0.4 (0.01) %	78 (2.8)	2 (0.0)	10 (0.2)	42 (0.0)	46 (2.0)	1 (0.0)	5 (0.1)	22 (0.4)
Georgia	1.3 (0.01) %	0.7 (0.01) %	1.1 (0.02) %	1.7 (0.02) %	26 (0.3)	14 (0.2)	21 (0.3)	33 (0.0)	15 (0.2)	8 (0.1)	12 (0.2)	19 (0.2)
India	8.2 (0.03) %	6.9 (0.03) %	8.0 (0.03) %	9.3 (0.04) %	51,368 (204.6)	43,246 (188.0)	50,009 (218.2)	57,985 (58.0)	28,756 (132.1)	23,452 (122.3)	27,708 (121.3)	32,947 (175.5)
Indonesia	7.7 (0.05) %	5.3 (0.05) %	7.1 (0.05) %	9.4 (0.06) %	8,948 (53.9)	6,180 (55.8)	8,204 (57.0)	10,901 (10.9)	5,494 (37.8)	3,484 (30.6)	4,856 (40.1)	6,756 (51.7)
Iran (Islamic Rep of)	12.3 (0.05) %	9.9 (0.04) %	11.8 (0.05) %	14.1 (0.06) %	4,672 (20.8)	3,788 (15.7)	4,510 (20.7)	5,356 (5.4)	2,803 (13.8)	2,186 (11.9)	2,661 (13.0)	3,256 (13.8)
Iraq	11.2 (0.03) %	8.1 (0.03) %	10.6 (0.04) %	13.5 (0.05) %	1,762 (5.1)	1,279 (5.3)	1,669 (6.0)	2,130 (2.1)	1,056 (3.5)	720 (4.0)	970 (4.1)	1,297 (6.1)
Korea, Rep of	0.8 (0.02) %	0.1 (0.00) %	0.2 (0.00) %	0.6 (0.01) %	182 (4.7)	17 (0.4)	50 (1.1)	145 (0.1)	100 (2.8)	8 (0.2)	24 (0.5)	72 (1.4)
Kyrgyzstan	1.6 (0.04) %	0.1 (0.00) %	0.3 (0.01) %	1.2 (0.04) %	43 (1.1)	3 (0.1)	9 (0.2)	33 (0.0)	26 (0.9)	1 (0.0)	5 (0.1)	17 (0.6)
Lao People's Dem Rep	16.6 (0.08) %	11.6 (0.09) %	15.6 (0.11) %	20.5 (0.11) %	533 (2.5)	372 (2.9)	500 (3.6)	657 (0.7)	350 (1.9)	224 (2.1)	315 (2.9)	437 (2.8)
Malaysia	8.3 (0.04) %	6.6 (0.04) %	8.0 (0.05) %	9.6 (0.06) %	1,177 (5.9)	942 (5.2)	1,129 (6.5)	1,359 (1.4)	669 (3.7)	506 (3.2)	627 (4.0)	782 (5.0)
Myanmar	7.4 (0.03) %	4.1 (0.03) %	6.1 (0.03) %	9.3 (0.04) %	1,833 (7.8)	1,028 (6.2)	1,523 (7.3)	2,309 (2.3)	1,115 (6.0)	577 (3.9)	880 (5.9)	1,384 (5.6)
Nepal	7.2 (0.06) %	2.9 (0.02) %	5.3 (0.03) %	9.4 (0.11) %	1,070 (8.8)	436 (3.4)	786 (4.2)	1,390 (1.4)	648 (6.5)	230 (2.0)	434 (3.2)	814 (10.2)
Pakistan	16.2 (0.09) %	10.8 (0.07) %	15.0 (0.11) %	20.4 (0.16) %	15,626 (90.9)	10,393 (63.4)	14,495 (108.4)	19,648 (19.6)	10,150 (75.4)	5,984 (50.3)	8,900 (71.1)	13,089 (130.8)
Papua New Guinea	7.9 (0.04) %	6.0 (0.03) %	7.4 (0.04) %	9.3 (0.05) %	277 (1.5)	212 (1.0)	261 (1.5)	325 (0.3)	157 (1.0)	114 (0.6)	143 (1.0)	185 (1.0)
Philippines	2.5 (0.01) %	2.4 (0.01) %	2.5 (0.01) %	2.5 (0.01) %	1,153 (6.1)	1,117 (5.9)	1,151 (6.0)	1,187 (1.2)	581 (3.2)	556 (3.0)	580 (3.2)	604 (3.6)
Saudi Arabia	12.9 (0.06) %	10.4 (0.08) %	12.4 (0.07) %	14.9 (0.05) %	1,877 (8.1)	1,511 (11.7)	1,794 (9.7)	2,161 (2.2)	957 (4.5)	741 (6.6)	899 (5.8)	1,113 (4.7)
Solomon Islands	24.0 (0.11) %	15.7 (0.11) %	22.3 (0.13) %	30.9 (0.17) %	66 (0.3)	43 (0.3)	62 (0.4)	86 (0.1)	43 (0.2)	25 (0.2)	38 (0.2)	56 (0.4)
Sri Lanka	3.0 (0.03) %	2.6 (0.02) %	2.9 (0.03) %	3.3 (0.03) %	299 (2.7)	258 (2.5)	291 (2.7)	331 (0.3)	168 (1.7)	140 (1.3)	161 (1.7)	188 (1.7)
Tajikistan	1.7 (0.03) %	0.4 (0.01) %	0.8 (0.02) %	1.9 (0.03) %	59 (1.0)	12 (0.2)	29 (0.6)	66 (0.1)	33 (0.7)	6 (0.1)	15 (0.3)	34 (0.5)
Thailand	13.8 (0.06) %	11.9 (0.06) %	13.6 (0.06) %	15.5 (0.05) %	4,636 (19.0)	3,981 (18.6)	4,544 (20.0)	5,188 (5.2)	2,923 (13.4)	2,428 (13.6)	2,830 (14.8)	3,312 (12.6)

	Allele frequency (SE)				G6PDd male population (SE)				G6PDd female population (SE)			
Country	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**	**Mean**	**Q25%**	**Median**	**Q75%**
Timor-Leste	7.5 (0.07) %	2.5 (0.03) %	5.0 (0.05) %	9.7 (0.11) %	44 (0.4)	15 (0.2)	29 (0.3)	58 (0.1)	25 (0.3)	7 (0.1)	15 (0.2)	31 (0.5)
Turkey	4.1 (0.04) %	3.0 (0.03) %	3.8 (0.04) %	4.9 (0.05) %	1,551 (14.8)	1,128 (12.1)	1,437 (14.8)	1,850 (1.9)	863 (8.8)	601 (6.9)	783 (8.3)	1,034 (12.8)
Uzbekistan	2.2 (0.04) %	0.4 (0.01) %	1.0 (0.01) %	2.4 (0.04) %	298 (5.4)	53 (0.9)	132 (1.4)	330 (0.3)	176 (3.8)	27 (0.5)	68 (0.9)	179 (3.6)
Vanuatu	8.2 (0.03) %	6.9 (0.02) %	8.0 (0.03) %	9.3 (0.04) %	10 (0.0)	9 (0.0)	10 (0.0)	12 (0.0)	6 (0.0)	4 (0.0)	5 (0.0)	6 (0.0)
Viet Nam	10.8 (0.11) %	6.0 (0.06) %	8.9 (0.10) %	13.9 (0.15) %	4,745 (46.5)	2,625 (24.5)	3,902 (43.1)	6,102 (6.1)	3,098 (36.6)	1,488 (15.5)	2,310 (33.0)	3,922 (54.1)
Yemen	8.4 (0.08) %	1.9 (0.04) %	4.6 (0.06) %	10.9 (0.15) %	1,031 (9.9)	238 (5.2)	565 (7.0)	1,329 (1.3)	656 (7.2)	121 (2.8)	308 (4.7)	781 (11.9)

Table S3. Reported observations of Class II and III G6PD variants from malaria endemic countries. Only MECs for which data were available are listed (n = 54 of 90 MECs).

The A- variant includes mostly variants carrying the 202 G→A and 376 A→G mutations; but also some diagnoses determined only from the 202 locus (in cases where only the 376A→G mutation was identified (G6PD A), the record was not included here as that mutation is from Class IV), as well as the 680 G→T/376 A→G and the 968 T→C/376 A→G variants.

Country	Number of occurrences	Class II variants	Class III variants
African MECs			
Angola	1		A- [1]
Benin	1		A- [2]
Burkina Faso	4		A- [3-5]
Cameroon	8		A- [6,7]
Cape Verde	1		A- [8]
Central African Republic	1		A- [9]
Comores	2	Mediterranean [10]	A- [10]
Congo	1		A- [11]
Côte d'Ivoire	4		A- [7,12,13]
DR Congo	1		A- [14]
Gabon	3		A- [7,15,16]
Ghana	7		A- [5,16-20]
Guinea	2		A- [21]
Kenya	6		A- [5,22-24]
Malawi	4		A- [25-27]
Mali	6		A- [5,7,28-30]
Mauritania	1		A- [7]
Mozambique	2		A- [31]
Namibia	6		A- [32,33]
Nigeria	21		A- [5,14,16,23,34-43] Ilesha [44]
Rwanda	1		A- [45]
Sao Tome and Principe	1		A- [46]
Senegal	7	Santamaria [47]	A- [7,47-51]
Sierra Leone	2		A- [52]
South Africa	2	Mediterranean [53]	A- [14,33]
Sudan	7	Mediterranean [54]	A- [54-59]
The Gambia	8	Santamaria [60]	A- [14,60-63]
Uganda	3		A- [64-66]
United Republic of Tanzania	6		A- [5,67-70]
American MECs			
Brazil	39	Amazonia [71] Ananindeau [71] Belem [71] Crispim [71] Chatham [72] Farroupilha [73]	A- [71-74,76-84] Bahia [78] Lages [73] Seattle [71,73,75,81] Seattle-like [76,85]

		Mediterranean [72-76] Santamaria [71]	
Costa Rica	3	Santamaria [86,87]	A- [86]
Ecuador	2		A- [88]
Guyana	1		A- [7]
Mexico	24	Santamaria [89] Union [89] Valladolid [90] Vanua Lava [89] Viangchan-Jammu [90]	A- [89,91-94] Mexico City [95] Seattle [89,94]
Panama	3	Mediterranean [96]	A- [96]

Eurasian MECs			
Cambodia	11	Canton [97] Kaiping [98] Valladolid [97] Viangchan-Jammu [97-99]	A- [100] Mahidol [97,101]
China	266	Canton [102-136] Chinese-1 [103,124,137] Coimbra [106,109,130,136-138] Fushan [109,121,139,140] Haikou [103] Hechi [109] Kaiping [103-110,112,113,115-119,121-127,129,132-136,138-144] Liuzhou [109] Miaoli [109,113,126,129,137] Nankang [103,124,126,145] Songklanagarind [109] Taipei [113,126,133,135,136] Taipei-Hakka [111] Union [106,124,126,127,132-134,136,146] Valladolid [122,124] Viangchan-Jammu [103,106,109,112,118,121-125,127,133,134] Gaohe/Kaiping [106,124]	A- [109,117] Chinese-5 [104-106,109,113,118,122-127,129,132-136] Gaohe [103-106,108-110,112,113,115,117,118,121-124,126,127,129,131-136,143,144] Guangzhou [106,117] Keelung [137] Mahidol [106,113,122,124,126,127,129,133,135,136,147] Mahidol-like [117] Nanning [109] Quing Yan [103,104,106,109,113,117,122-124,126,127,129,133-136,148] Ube Konan [144]
India	66	Chatham [149] Coimbra [150-153] Mediterranean [149,152-156] Namouru [150,151,153] Nilgiri [150,151]	Kalyan-Kerala [149-153,156-159]; Orissa [149,152-154,156,160,161]
Indonesia	76	Canton [162-167] Chatham [162,165,166,168-171] Coimbra [162,168,169,171] Kaiping [162,164-166,168-172] Mediterranean [163,171] Surabaya [162] Union [165,166] Vanua Lava [162,165,166,168-171,173] Viangchan-Jammu [165-169,171,174]	Bajo Maumere [168] Chinese-5 [168] Gaohe [162] Mahidol [163]
Iran	29	Canton [175,176] Chatham [175-182] Cosenza [177,178,181,182]	A- [175,176]

		Mediterranean [155,175-186]	
Iraq	11	Chatham [187,188] Mediterranean [187-191]	A- [188,189]
Lao PDR	1	Viangchan-Jammu [162]	
Malaysia	61	Andalus [192] Canton [192-200] Chatham [192,196,199] Coimbra [192,198-200] Kaiping [192,194,195,197-200] Mediterranean [192,198-200] Namouru [198] Nankang [196] Union [192,196] Vanua Lava [192,199] Viangchan-Jammu [192,194-196,198-200]	Chinese-5 [194-197] Gaohe [194-198,200] Mahidol [192,196,198-200] Orissa [192,199] Quing Yan [195,196]
Myanmar	27	Canton [162,201] Coimbra [201,202] Viangchan-Jammu [202] Kaiping [202] Mediterranean [202] Union [162,201] Valladolid [202]	Kerala-Kalyan [202] Mahidol [162,201-203]`
Nepal	1	Mediterranean [204]	
Pakistan	9	Chatham [205] Mediterranean [155,205-208]	Orissa [205]
Papua New Guinea	16	Viangchan-Jammu [209] Kaiping [210] Mediterranean-like [211,212] Union [210] Union-like [211,212] Vanua Lava [209]	
Philippines	2	Union [213,214]	
Saudi Arabia	81	Aures [215-218] Chatham [216-218] Kaiping [218] Mediterranean [191,215-230] Mediterranean-like [221-226] S. Antioco [216] Union [216] Viangchan [216]	A- [216,218,220-224,226-228,231] Kerala-Kalyan [218] Sibari [230]
Solomon Islands	1	Union [232]	
Thailand	45	Canton [233-237] Kaiping [233-239] Mediterranean [233] Songklanagarind [233] Union [233-237] Vanua Lava [240] Viangchan-Jammu [233,234,237-239]	Chinese-5 [234] Gaohe [233,237,238] Kerala-Kalyan [238] Mahidol [233-239,241,242] Quing Yan [233]
Turkey	7	Chatham [243,244] Mediterranean [243,245-247]	A- [243]
Vanuatu	4	Namouru [248] Naone [248] Union [248] Vanua Lava [248]	

Viet Nam	19	Bao Loc [249] Canton [249-251] Coimbra [252] Kaiping [249] Union [249] Viangchan-Jammu [249,252]	Chinese-5 [252] Gaohe [249,252] Mahidol [250] Quing Yan [249]

References

1. Nurse GT, Jenkins T, David JH, Steinberg AG (1979) The Njinga of Angola: a serogenetic study. Annals of Human Biology 6: 337-348.
2. Biondi G, Rickards O, Martinez-Labarga C, Taraborelli T, Ciminelli B, et al. (1996) Biodemography and genetics of the Berba of Benin. Am J Phys Anthropol 99: 519-535.
3. Meissner PE, Coulibaly B, Mandi G, Mansmann U, Witte S, et al. (2005) Diagnosis of red cell G6PD deficiency in rural Burkina Faso: comparison of a rapid fluorescent enzyme test on filter paper with polymerase chain reaction based genotyping. Br J Haematol 131: 395-399.
4. Modiano D, Luoni G, Sirima BS, Lanfrancotti A, Petrarca V, et al. (2001) The lower susceptibility to *Plasmodium falciparum* malaria of Fulani of Burkina Faso (West Africa) is associated with low frequencies of classic malaria-resistance genes. Trans R Soc Trop Med Hyg 95: 149-152.
5. Carter N, Pamba A, Duparc S, Waitumbi JN (2011) Frequency of glucose-6-phosphate dehydrogenase deficiency in malaria patients from six African countries enrolled in two randomized anti-malarial clinical trials. Malar J 10: 241.
6. Bernstein SC, Bowman JE, Kaptue Noche L (1980) Population studies in Cameroon: hemoglobin S, glucose-6-phosphate dehydrogenase deficiency and falciparum malaria. Hum Hered 30: 251-258.
7. Kahn A, Boivin P, Lagneau J (1973) [Phenotypes of erythrocytic glucose-6-phosphate dehydrogenase in black people. Examination of 301 black people living in France and description of 9 different variants. High incidence of deficiency of an enzyme of "B" mobility]. Humangenetik 18: 261-270.
8. Alves J, Machado P, Silva J, Goncalves N, Ribeiro L, et al. (2010) Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers. Blood Cells Mol Dis 44: 62-68.
9. Vergnes H, Sevin A, Sevin J, Jaeger G (1979) Population genetic studies of the Aka pygmies (Central Africa): a survey of red cell and serum enzymes. Hum Genet 48: 343-355.
10. Badens C, Martinez di Montemuros F, Thuret I, Michel G, Mattei JF, et al. (2000) Molecular basis of haemoglobinopathies and G6PD deficiency in the Comorian population. Hematol J 1: 264-268.
11. Bouanga JC, Mouele R, Prehu C, Wajcman H, Feingold J, et al. (1998) Glucose-6-phosphate dehydrogenase deficiency and homozygous sickle cell disease in Congo. Hum Hered 48: 192-197.

12. Coulibaly FH, Koffi G, Toure HA, Bouanga JC, Allangba O, et al. (2000) Molecular genetics of glucose-6-phosphate dehydrogenase deficiency in a population of newborns from Ivory Coast. Clin Biochem 33: 411-413.
13. Vergnes H, Cabannes R (1976) Polymorphism of erythrocyte and serum enzyme systems in the Gagu of the Ivory Coast. Ann Hum Biol 3: 423-429.
14. Saunders MA, Hammer MF, Nachman MW (2002) Nucleotide variability at G6pd and the signature of malarial selection in humans. Genetics 162: 1849-1861.
15. Mombo LE, Ntoumi F, Bisseye C, Ossari S, Lu CY, et al. (2003) Human genetic polymorphisms and asymptomatic Plasmodium falciparum malaria in Gabonese schoolchildren. Am J Trop Med Hyg 68: 186-190.
16. May J, Meyer CG (2003) A synonymous mutation of ancient origin in the glucose-6-phosphate dehydrogenase gene and assessment of haplotypes. Blood Cells Mol Dis 30: 144-145.
17. Acquaye CT, Oldham JH (1974) A preliminary report on survey and characterization of haemoglobins and glucose-6-phosphate dehydrogenase variants in Ghana. West Afr J Pharmacol Drug Res 1: 25P-28P.
18. Owusu SK, Opare-Mante A (1972) Electrophoretic characterisation of glucose-6-phosphate dehydrogenase in Ghana. Lancet 2: 44.
19. Timmann C, Evans JA, Konig IR, Kleensang A, Ruschendorf F, et al. (2007) Genome-wide linkage analysis of malaria infection intensity and mild disease. PLoS Genet 3: e48.
20. Armoo S, Wilson M, Boakye D, Quakyi I (2010) Studies on Abo Blood Groups, Haemoglobinopathies and G6pd Genotypes, and Plasmodium Falciparum Infection in Kpone-on-Sea, Ghana. American Journal of Tropical Medicine and Hygiene 83: 82-82.
21. Millimono TS, Loua KM, Rath SL, Relvas L, Bento C, et al. (2012) High prevalence of hemoglobin disorders and glucose-6-phosphate dehydrogenase (G6PD) deficiency in the Republic of Guinea (West Africa). Hemoglobin 36: 25-37.
22. Othieno-Obel A (1972) East African variant of glucose-6-phosphate dehydrogenase. East Afr Med J 49: 230-234.
23. Vulliamy TJ, Othman A, Town M, Nathwani A, Falusi Y, et al. (1991) Linkage disequilibrium of polymorphic sites in the G6PD gene in African populations and the origin of G6PD A. Gene Geogr 5: 13-21.
24. Shah SS, Macharia A, Uyoga S, Williams TN (2011) Modeling the relationship between genotype and biochemotype at the *G6PD* locus in Kenya. pp. Unpublished work.
25. Nkhoma ET, Mu JB, Krause MA, Diakite SA, Kalilani L, et al. (2009) Effect of the a-Form of G6pd Deficiency on Maternal Plasmodium Falciparum Parasitemia and Pregnancy Outcomes. American Journal of Tropical Medicine and Hygiene 81: 86-86.
26. Calis JC, Phiri KS, Faragher EB, Brabin BJ, Bates I, et al. (2008) Severe anemia in Malawian children. New England Journal of Medicine 358: 888-899.
27. Senga EL, Harper G, Koshy G, Kazembe PN, Brabin BJ (2011) Reduced risk for placental malaria in iron deficient women. Malaria journal 10: 47.
28. Duflo B, Diallo A, Toure K, Soula G (1979) [Glucose-6-phosphate dehydrogenase deficiency in Mali. Epidemiology and pathological aspects]. Bull Soc Pathol Exot Filiales 72: 258-264.
29. Junien C, Chaventre A, Fofana Y, Lapoumeroulie C, Floury B, et al. (1982) Glucose-6-phosphate dehydrogenase and hemoglobin variants in Kel Kummer Tuareg and related groups. Indirect evidence for alpha-thalassemia trait. Hum Hered 32: 318-328.
30. Crompton PD, Traore B, Kayentao K, Doumbo S, Ongoiba A, et al. (2008) Sickle cell trait is associated with a delayed onset of malaria: Implications for time-to-event analysis in clinical studies of malaria. Journal of Infectious Diseases 198: 1265-1275.
31. Reys L, Manso C, Stamatoyannopoulos G (1970) Genetic studies on southeastern Bantu of Mozambique. I. Variants of glucose-6-phosphate dehydrogenase. Am J Hum Genet 22: 203-215.

32. Nurse GT, Jenkins T (1977) Serogenetic studies on the Kavango peoples of South West Africa. Annals of Human Biology 4: 465-478.
33. Coetzee MJ, Bartleet SC, Ramsay M, Jenkins T (1992) Glucose-6-phosphate dehydrogenase (G6PD) electrophoretic variants and the PvuII polymorphism in southern African populations. Hum Genet 89: 111-113.
34. Ademowo OG, Falusi AG (2002) Molecular epidemiology and activity of erythrocyte G6PD variants in a homogeneous Nigerian population. East Afr Med J 79: 42-44.
35. Bienzle U, Guggenmoos-Holzmann I, Luzzatto L (1979) Malaria and erythrocyte glucose-6-phosphate dehydrogenase variants in West Africa. Am J Trop Med Hyg 28: 619-621.
36. Guggenmoos-Holzmann I, Bienzle U, Luzzatto L (1981) Plasmodium falciparum malaria and human red cells. II. Red cell genetic traits and resistance against malaria. Int J Epidemiol 10: 16-22.
37. Luzzatto L, Allan NC (1968) Relationship between the genes for glucose-6-phosphate dehydrogenase and for haemoglobin in a Nigerian population. Nature 219: 1041-1042.
38. Porter IH, Boyer SH, Watson-Williams EJ, Adam A, Szeinberg A, et al. (1964) Variation of Glucose-6-Phosphate Dehydrogenase in Different Populations. Lancet 283: 895-899.
39. Mockenhaupt FP, May J, Stark K, Falusi AG, Meyer CG, et al. (1999) Serum transferrin receptor levels are increased in asymptomatic and mild *Plasmodium falciparum*-infection. Haematologica 84: 869-873.
40. Bienzle U, Effiong CE, Aimaku VE, Luzzatto L (1976) Erythrocyte enzymes in neonatal jaundice. Acta Haematologica 55: 10-20.
41. Bienzle U, Ayeni O, Lucas AO, Luzzatto L (1972) Glucose-6-phosphate dehydrogenase and malaria. Greater resistance of females heterozygous for enzyme deficiency and of males with non-deficient variant. Lancet 1: 107-110.
42. May J, Meyer CG, Grossterlinden L, Ademowo OG, Mockenhaupt FP, et al. (2000) Red cell glucose-6-phosphate dehydrogenase status and pyruvate kinase activity in a Nigerian population. Trop Med Int Health 5: 119-123.
43. Porter IH, Boyer SH, Watson-Williams EJ, Adam A, Szeinberg A, et al. (1964) Variation of Glucose-6-Phosphate Dehydrogenase in Different Populations. Lancet 1: 895-899.
44. Vulliamy TJ, D'Urso M, Battistuzzi G, Estrada M, Foulkes NS, et al. (1988) Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. Proceedings of the National Academy of Sciences of the United States of America 85: 5171-5175.
45. Gahutu JB, Musemakweri A, Harms G, Mockenhaupt FP (2012) Prevalence of classic erythrocyte polymorphisms among 749 children in southern highland Rwanda. Transactions of the Royal Society of Tropical Medicine and Hygiene 106: 63-65.
46. Manco L, Botigue LR, Ribeiro ML, Abade A (2007) G6PD deficient alleles and haplotype analysis of human G6PD locus in Sao Tome e Principe (West Africa). Hum Biol 79: 679-686.
47. De Araujo C, Migot-Nabias F, Guitard J, Pelleau S, Vulliamy T, et al. (2006) The role of the G6PD AEth376G/968C allele in glucose-6-phosphate dehydrogenase deficiency in the seerer population of Senegal. Haematologica 91: 262-263.
48. Bouloux C, Gomila J, Langaney A (1972) Hemotypology of the Bedik. Hum Biol 44: 289-302.
49. Sarr JB, Pelleau S, Toly C, Guitard J, Konate L, et al. (2006) Impact of red blood cell polymorphisms on the antibody response to *Plasmodium falciparum* in Senegal. Microbes and Infection 8: 1260-1268.
50. Vergnes H, Gherardi M, Bouloux C (1975) Erythrocyte glucose-6-phosphate dehydrogenase in the Niokolonko (Malinke of the Niokolo) of Eastern Senegal. Identification of a slow variant with normal activity (Tacoma-like). Hum Hered 25: 80-87.
51. Courtin D, Milet J, Bertin G, Vafa M, Sarr JB, et al. (2011) G6PD A-variant influences the antibody responses to Plasmodium falciparum MSP2. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 11: 1287-1292.

52. Jalloh A, Jalloh M, Gamanga I, Baion D, Sahr F, et al. (2008) G6PD deficiency assessment in Freetown, Sierra Leone, reveals further insight into the molecular heterogeneity of G6PD A. J Hum Genet 53: 675-679.
53. Lane AB, Grant SC, Jenkins T (1983) G6pd Witwatersrand, a New Variant Found in a South-African Ashkenazi Jew and the Report of a 2nd Case of G6pd Panama. South African Journal of Science 79: 414-416.
54. Saha N, Samuel AP (1991) Characterization of glucose-6-phosphate dehydrogenase variants in the Sudan--including GdKhartoum, a hyperactive slow variant. Hum Hered 41: 17-21.
55. Bayoumi RA, Saha N (1987) Some blood genetic markers of the Nuba and Hawazma tribes of western Sudan. Am J Phys Anthropol 73: 379-388.
56. Saha N, el Seikh FS (1987) Some blood genetic characteristics of several Sudanese tribes. Am J Phys Anthropol 73: 397-406.
57. Saha N, Samuel AP, Omer A, Ahmed MA, Hussein AA, et al. (1978) A study of some genetic characteristics of the population of the Sudan. Ann Hum Biol 5: 569-575.
58. Saha N, Samuel AP, Omer A, Hoffbrand AV (1983) The inter- and intra-tribal distribution of red cell G6PD phenotypes in Sudan. Hum Hered 33: 39-43.
59. Samuel AP, Saha N, Omer A, Hoffbrand AV (1981) Quantitative expression of G6PD activity of different phenotypes of G6PD and haemoglobin in a Sudanese population. Hum Hered 31: 110-115.
60. Clark TG, Fry AE, Auburn S, Campino S, Diakite M, et al. (2009) Allelic heterogeneity of G6PD deficiency in West Africa and severe malaria susceptibility. Eur J Hum Genet 17: 1080-1085.
61. Atkinson SH, Rockett K, Sirugo G, Bejon PA, Fulford A, et al. (2006) Seasonal childhood anaemia in West Africa is associated with the haptoglobin 2-2 genotype. PLoS Med 3: e172.
62. Sirugo G, Schaefer EA, Mendy A, West B, Bailey R, et al. (2004) Is G6PD A- deficiency associated with recurrent stillbirths in The Gambia? Am J Med Genet A 128A: 104-105.
63. Welch SG, Lee J, McGregor IA, Williams K (1978) Red cell glucose 6 phosphate dehydrogenase genotypes of the population of two West African villages. Hum Genet 43: 315-320.
64. Johnson MK, Clark TD, Njama-Meya D, Rosenthal PJ, Parikh S (2009) Impact of the method of G6PD deficiency assessment on genetic association studies of malaria susceptibility. PLoS One 4: e7246.
65. Katrak S, Gasasira A, Arinaitwe E, Kakuru A, Wanzira H, et al. (2009) Safety and tolerability of artemether-lumefantrine versus dihydroartemisinin-piperaquine for malaria in young HIV-infected and uninfected children. Malar J 8: 272.
66. Parikh S, Johnson MK, Kamya MR, Dorsey G, Rosenthal PJ (2008) Glucose 6-Phosphate Dehydrogenase (G6pd) Deficiency Genotype-Phenotype Correlations in Malaria Association Studies. American Journal of Tropical Medicine and Hygiene 79: 243-243.
67. Enevold A, Lusingu JP, Mmbando B, Alifrangis M, Lemnge MM, et al. (2008) Reduced risk of uncomplicated malaria episodes in children with alpha+-thalassemia in northeastern Tanzania. Am J Trop Med Hyg 78: 714-720.
68. Shekalaghe S, Alifrangis M, Mwanziva C, Enevold A, Mwakalinga S, et al. (2009) Low density parasitaemia, red blood cell polymorphisms and Plasmodium falciparum specific immune responses in a low endemic area in northern Tanzania. BMC Infect Dis 9: 69.
69. Roberts DF, Papiha SS (1978) Les polymorphismes genetiques des Sukuma (Tanzania). L'Anthropologie 82: 565-574.
70. Shekalaghe SA, Ter Braak R, Daou M, Kavishe R, van den Bijllaardt W, et al. (2010) Haemolysis after a single dose of primaquine co-administered with an artemisinin is not restricted to glucose-6-phosphate dehydrogenase (G6PD A- variant) deficient individuals in Tanzania. Antimicrob Agents Chemother 54: 1762-1768.
71. Hamel AR, Cabral IR, Sales TS, Costa FF, Olalla Saad ST (2002) Molecular heterogeneity of G6PD deficiency in an Amazonian population and description of four new variants. Blood Cells Mol Dis 28: 399-406.

72. Saad ST, Salles TS, Carvalho MH, Costa FF (1997) Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in Brazil. Hum Hered 47: 17-21.
73. Weimer TA, Salzano FM, Westwood B, Beutler E (1998) G6PD variants in three South American ethnic groups: population distribution and description of two new mutations. Hum Hered 48: 92-96.
74. Azevedo ES, Alves AF, Da Silva MC, Souza MG, Muniz Dias Lima AM, et al. (1980) Distribution of abnormal hemoglobins and glucose-6-phosphate dehydrogenase variants in 1200 school children of Bahia, Brazil. Am J Phys Anthropol 53: 509-512.
75. Weimer TA, Salzano FM, Westwood B, Beutler E (1993) Molecular characterization of glucose-6-phosphate dehydrogenase variants from Brazil. Hum Biol 65: 41-47.
76. Weimer TA, Salzano FM, Hutz MH (1981) Erythrocyte isozymes and hemoglobin types in a southern Brazilian population. Journal of Human Evolution 10: 319-328.
77. Saad ST, Costa FF (1992) Glucose-6-phosphate dehydrogenase deficiency and sickle cell disease in Brazil. Hum Hered 42: 125-128.
78. Neto JPD, Dourado MV, dos Reis MG, Goncalves MS (2008) A novel c. 197T -> A variant among Brazilian neonates with glucose-6-phosphate dehydrogenase deficiency. Genetics and Molecular Biology 31: 33-35.
79. Santana MS, de Lacerda MV, Barbosa MG, Alecrim WD, Alecrim MG (2009) Glucose-6-phosphate dehydrogenase deficiency in an endemic area for malaria in Manaus: a cross-sectional survey in the Brazilian Amazon. PLoS One 4: e5259.
80. de Castro SM, Weber R, Matte U, Giugliani R (2007) Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in patients from the Southern Brazilian city of Porto Alegre, RS. Genetics and Molecular Biology 30: 10-13.
81. Oliveira RAG, Oshiro M, Hirata MH, Hirata RDC, Ribeiro GS, et al. (2009) A novel point mutation in a class IV glucose-6-phosphate dehydrogenase variant (G6PD Sao Paulo) and polymorphic G6PD variants in Sao Paulo State, Brazil. Genetics and Molecular Biology 32: 251-254.
82. Chan TY (1997) Co-trimoxazole-induced severe haemolysis: the experience of a large general hospital in Hong Kong. Pharmacoepidemiol Drug Saf 6: 89-92.
83. Mezzacappa MA, Facchini FP, Pinto AC, Cassone AE, Souza DS, et al. (2010) Clinical and genetic risk factors for moderate hyperbilirubinemia in Brazilian newborn infants. Journal of perinatology : official journal of the California Perinatal Association 30: 819-826.
84. Cardoso MA, Scopel KK, Muniz PT, Villamor E, Ferreira MU (2012) Underlying factors associated with anemia in amazonian children: a population-based, cross-sectional study. PLoS One 7: e36341.
85. Hutz MH, Yoshida A, Salzano FM (1977) Three rare G-6-PD variants from Porto Alegre, Brazil. Human Genetics 39: 191-197.
86. Beutler E, Kuhl W, Saenz GF, Rodriguez W (1991) Mutation analysis of glucose-6-phosphate dehydrogenase (G6PD) variants in Costa Rica. Hum Genet 87: 462-464.
87. Saenz GF, Chaves M, Berrantes A, Elizondo J, Montero AG, et al. (1984) A glucose-6-phosphate dehydrogenase variant, Gd(-) Santamaria found in Costa Rica. Acta Haematol 72: 37-40.
88. Martinez-Labarga C, Rickards O, Scacchi R, Corbo RM, Biondi G, et al. (1999) Genetic population structure of two African-Ecuadorian communities of Esmeraldas. Am J Phys Anthropol 109: 159-174.
89. Vaca G, Arambula E, Esparza A (2002) Molecular heterogeneity of glucose-6-phosphate dehydrogenase deficiency in Mexico: overall results of a 7-year project. Blood Cells Mol Dis 28: 436-444.
90. Vaca G, Arambula E, Monsalvo A, Medina C, Nunez C, et al. (2003) Glucose-6-phosphate dehydrogenase (G-6-PD) mutations in Mexico: four new G-6-PD variants. Blood Cells Mol Dis 31: 112-120.

91. Lisker R, Cordova MS, Graciela Zarate QB (1969) Studies on several genetic hematological traits of the Mexican population. XVI. Hemoglobin, S and glucose-6-phosphate dehydrogenase deficiency in the east coast. Am J Phys Anthropol 30: 349-354.
92. Beutler E, Kuhl W, Ramirez E, Lisker R (1991) Some Mexican glucose-6-phosphate dehydrogenase variants revisited. Hum Genet 86: 371-374.
93. Lisker R, Perez-Briceno R, Rave V, Yoshida A (1981) [Federal District glucose-6-phosphate dehydrogenase Gd(-). A new variant associated with moderate enzyme deficiency and occasional hemolytic anemia]. Rev Invest Clin 33: 209-211.
94. Medina MD, Vaca G, Lopez-Guido B, Westwood B, Beutler E (1997) Molecular genetics of glucose-6-phosphate dehydrogenase deficiency in Mexico. Blood Cells Mol Dis 23: 88-94.
95. Beutler E, Westwood B, Prchal JT, Vaca G, Bartsocas CS, et al. (1992) New glucose-6-phosphate dehydrogenase mutations from various ethnic groups. Blood 80: 255-256.
96. Cossio-Gurrola G, Arambula-Meraz E, Perea M, Garcia N, Correa AS, et al. (2010) Glucose-6-phosphate dehydrogenase (G6PD) molecular variant deficiency: identification in Panama pediatric population. Blood Cells Mol Dis 44: 115-116.
97. Kim S, Nguon C, Guillard B, Duong S, Chy S, et al. (2011) Performance of the CareStart G6PD deficiency screening test, a point-of-care diagnostic for primaquine therapy screening. PloS one 6: e28357.
98. Narazah MY, Devenish R, Rostenberghe HV, Nishiyama K, Shirakawa T, et al. (2004) Molecular basis of glucose-6-phosphate dehydrogenase (G6PD) deficiency in Cambodia. Faops 2004: Proceedings of the 13th Congress of the Federation of Asia and Oceania Perinatal Societies Faops 2004: 149-152.
99. Louicharoen C, Nuchprayoon I (2005) G6PD Viangchan (871G>A) is the most common G6PD-deficient variant in the Cambodian population. J Hum Genet 50: 448-452.
100. Viallard JL, Cottreau D, Kahn A, Dastugue B (1979) G6PD deficiency with Gd(-)A like variant in a Chinese family from Cambodia. Hum Genet 51: 213-215.
101. El-Hazmi MA (1987) Haemoglobinopathies, thalassaemias and enzymopathies in Saudi Arabia: the present status. Acta Haematol 78: 130-134.
102. Cai W, Filosa S, Martini G (1994) DNA haplotypes in the G6PD gene cluster studied in the Chinese Li population and their relationship to G6PDCanton. Hum Hered 44: 279-286.
103. Cai W, Filosa S, Martini G, Zhou Y, Zhou D, et al. (2001) [Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in the Han and Li nationalities in Hainan, China and identification of a new mutation in human G6PD gene]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 18: 105-109.
104. Chiu DTY, Zuo L, Chao L, Chen E, Louie E, et al. (1993) Molecular characterization of glucose-6-phosphate dehydrogenase (G6PD) deficiency in patients of Chinese descent and identification of new base substitutions in the human G6PD gene. Blood 81: 2150-2154.
105. Jiang W, Du C, Duan S, Ma L, Yang L, et al. (1999) [Molecular characterization of glucose-6-phosphate dehydrogenase variants in four ethnic groups in Yunnan province of China]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 16: 149-152.
106. Jiang W, Yu G, Liu P, Geng Q, Chen L, et al. (2006) Structure and function of glucose-6-phosphate dehydrogenase-deficient variants in Chinese population. Hum Genet 119: 463-478.
107. Wu CX, He Y, Shan KR, Li Y, Xiu J, et al. (2006) Study the mutations of glucose-6-phosphate dehydrogenase gene in Yao ethnic group in Guizhou Libo. Chinese Journal of Endemiology 25: 402-404.
108. Xiu J, Qi XL, Shan KR, Xie Y, He Y, et al. (2005) [G6PD Gene Mutations in Shui people in Sandu of Guizhou]. Zhongguo Shi Yan Xue Ye Xue Za Zhi 13: 147-150.
109. Yan T, Cai R, Mo O, Zhu D, Ouyang H, et al. (2006) Incidence and complete molecular characterization of glucose-6-phosphate dehydrogenase deficiency in the Guangxi Zhuang

autonomous region of southern China: description of four novel mutations. Haematologica 91: 1321-1328.
110. Zhang DT, Hu LH, Yang YZ (2005) Detection of three common G6PD gene mutations in Chinese individuals by probe melting curves. Clin Biochem 38: 390-394.
111. Chan TK (1983) Glucose 6 phosphate dehydrogenase (G6PD) [MD]: University of Hong Kong.
112. Au WY, Ma ES, Lam VM, Chan JL, Pang A, et al. (2004) Glucose 6-phosphate dehydrogenase (G6PD) deficiency in elderly Chinese women heterozygous for G6PD variants. Am J Med Genet A 129A: 208-211.
113. Bang-Ce Y, Hongqiong L, Zhensong L (2004) Rapid detection of common Chinese glucose-6-phosphate dehydrogenase (G6PD) mutations by microarray-based assay. Am J Hematol 76: 405-412.
114. Chan TK, Todd D, Lai MC (1972) Glucose 6-phosphate dehydrogenase: identity of erythrocyte and leukocyte enzyme with report of a new variant in Chinese. Biochem Genet 6: 119-124.
115. Deng C, Guo CB, Xu YH, Deng B, Yu JL (2007) Three mutations analysis of glucose-6-phosphate dehydrogenase deficiency in neonates in South-west China. Pediatr Int 49: 463-467.
116. Du CS, Ren X, Chen L, Jiang W, He Y, et al. (1999) Detection of the most common G6PD gene mutations in Chinese using amplification refractory mutation system. Hum Hered 49: 133-138.
117. Du CS, Xu YK, Hua XY, Wu QL, Liu LB (1988) Glucose-6-phosphate dehydrogenase variants and their frequency in Guangdong, China. Hum Genet 80: 385-388.
118. Li L, Zhou YQ, Xiao QZ, Yan TZ, Xu XM (2008) Development and evaluation of a reverse dot blot assay for the simultaneous detection of six common Chinese G6PD mutations and one polymorphism. Blood Cells Mol Dis 41: 17-21.
119. Li P, Thompson JN, Wang X, Song L (1998) Analysis of common mutations and associated haplotypes in Chinese patients with glucose-6-phosphate dehydrogenase deficiency. Biochem Mol Biol Int 46: 1135-1143.
120. McCurdy PR, Kirkman HN, Naiman JL, Jim RT, Pickard BM (1966) A Chinese variant of glucose-6-phosphate dehydrogenase. J Lab Clin Med 67: 374-385.
121. Xu W, Westwood B, Bartsocas CS, Malcorra-Azpiazu JJ, Indrak K, et al. (1995) Glucose-6 phosphate dehydrogenase mutations and haplotypes in various ethnic groups. Blood 85: 257-263.
122. Wang YF, Xia WQ, Ni PH, Hu YQ, Jiang XC (2010) Analysis of glucose-6-phosphate dehydrogenase gene mutations: A novel missense mutation. Journal of Shanghai Jiaotong University (Medical Science) 30: 698-702.
123. Jiang WY, Zhou BY, Yu GL, Liu H, Zeng JB, et al. (2012) G6PD genotype and its associated enzymatic activity in a Chinese population. Biochemical genetics 50: 34-44.
124. Yan JB, Xu HP, Xiong C, Ren ZR, Tian GL, et al. (2010) Rapid and reliable detection of glucose-6-phosphate dehydrogenase (G6PD) gene mutations in Han Chinese using high-resolution melting analysis. The Journal of molecular diagnostics : JMD 12: 305-311.
125. Huang YY, Huang CS, Yang SS, Lin MS, Huang MJ, et al. (2005) Effects of variant UDP-glucuronosyltransferase 1A1 gene, glucose-6-phosphate dehydrogenase deficiency and thalassemia on cholelithiasis. World J Gastroenterol 11: 5710-5713.
126. Tseng CP, Huang CL, Chong KY, Hung IJ, Chiu DT (2005) Rapid detection of glucose-6-phosphate dehydrogenase gene mutations by denaturing high-performance liquid chromatography. Clin Biochem 38: 973-980.
127. Tang TK, Liu TH, Tang CJ, Tam KB (1995) Glucose-6-phosphate dehydrogenase (G6PD) mutations associated with F8C/G6PD haplotypes in Chinese. Blood 85: 3767-3768.
128. Chan TK, Todd D (1972) Characteristics and distribution of glucose-6-phosphate dehydrogenase-deficient variants in South China. Am J Hum Genet 24: 475-484.

129. Chiang SH, Wu SJ, Wu KF, Hsiao KJ (1999) Neonatal screening for glucose-6-phosphate dehydrogenase deficiency in Taiwan. Southeast Asian J Trop Med Public Health 30 Suppl 2: 72-74.
130. Tang TK, Huang WY, Tang CJ, Hsu M, Cheng TA, et al. (1995) Molecular basis of glucose-6-phosphate dehydrogenase (G6PD) deficiency in three Taiwan aboriginal tribes. Hum Genet 95: 630-632.
131. Wan GH, Lin KK, Tsai SC, Chiu DT (2006) Decreased glucose-6-phosphate-dehydrogenase (G6PD) activity and risk of senile cataract in Taiwan. Ophthalmic Epidemiol 13: 109-114.
132. Wan GH, Tsai SC, Chiu DT (2002) Decreased blood activity of glucose-6-phosphate dehydrogenase associates with increased risk for diabetes mellitus. Endocrine 19: 191-195.
133. Huang CS, Hung KL, Huang MJ, Li YC, Liu TH, et al. (1996) Neonatal jaundice and molecular mutations in glucose-6-phosphate dehydrogenase deficient newborn infants. Am J Hematol 51: 19-25.
134. Ko CH, Yung E, Li K, Li CL, Ng PC, et al. (2006) Multiplex primer extension reaction screening and oxidative challenge of glucose-6-phosphate dehydrogenase mutants in hemizygous and heterozygous subjects. Blood Cells Mol Dis 37: 21-26.
135. Chang JG, Chiou SS, Perng LI, Chen TC, Liu TC, et al. (1992) Molecular characterization of glucose-6-phosphate dehydrogenase (G6PD) deficiency by natural and amplification created restriction sites: five mutations account for most G6PD deficiency cases in Taiwan. Blood 80: 1079-1082.
136. Tang TK, Yeh CH, Huang CS, Huang MJ (1994) Expression and biochemical characterization of human glucose-6-phosphate dehydrogenase in Escherichia coli: a system to analyze normal and mutant enzymes. Blood 83: 1436-1441.
137. Chen HL, Huang MJ, Huang CS, Tang TK (1997) Two novel glucose 6-phosphate dehydrogenase deficiency mutations and association of such mutations with F8C/G6PD haplotype in Chinese. J Formos Med Assoc 96: 948-954.
138. Qi XL, Shan K, Xie Y, Wu CX, Xiu J, et al. (2006) Study on the mutations of G6PD gene in Dong ethnic group in Guizhou Congjiang. Chinese Journal of Endemiology 25: 283-285.
139. Du C, He Y (1997) [A case of nt 1004C --> A G6PD gene mutation in Yunnan Han people]. Zhonghua Xue Ye Xue Za Zhi 18: 535-537.
140. Ren X, He Y, Du C, Jiang W, Chen L, et al. (2001) A novel mis-sense mutation (G1381A) in the G6PD gene identified in a Chinese man. Chin Med J (Engl) 114: 399-401.
141. Wu CX, Shan KR, He Y, Qi XL, Li Y, et al. (2007) Detection of glucose-6-phosphate dehydrogenase gene mutations of Tujia ehtnic in Jiangkou, Guizhou. Chinese Journal of Endemiology 26: 415-417.
142. Wang J, Matsuoka H, Hirai M, Mu L, Yang L, et al. (2010) The First Case of a Class I Glucose-6-phosphate Dehydrogenase Deficiency, G6PD Santiago de Cuba (1339 GA), in a Chinese Population as Found in a Survey for G6PD Deficiency in Northeastern and Central China. Acta Med Okayama 64: 49-54.
143. Xu W, Wang J, Hua X, Du C (1994) Detection of point mutations in exon 2 of the G6PD gene in Chinese G6PD variants. Chin Med Sci J 9: 20-23.
144. Chao LT, Du CS, Louie E, Zuo L, Chen E, et al. (1991) A to G substitution identified in exon 2 of the G6PD gene among G6PD deficient Chinese. Nucleic Acids Res 19: 6056.
145. Chen HL, Huang MJ, Huang CS, Tang TK (1996) G6PD NanKang (517 T-->C; 173 Phe-->Leu): a new Chinese G6PD variant associated with neonatal jaundice. Hum Hered 46: 201-204.
146. Tang TK, Chen HL, Huang CS, Liu TH (1995) Identification of a novel G6PD mutation (G6PD NanKang) and the association of F8C/G6PD haplotypes in Chinese. Blood 86: 525-525.
147. Yang Y, Zhu Y, Li D, Li Z, Lu H, et al. (2007) Characterization of glucose-6-phosphate dehydrogenase deficiency and identification of a novel haplotype 487G>A/IVS5-612(G>C) in the Achang population of Southwestern China. Sci China C Life Sci 50: 479-485.

148. Du CS, Liu LB, Liu B, Tokunaga K, Omoto K (1988) Glucose-6-phosphate dehydrogenase deficiency among three national minorities in Hainan Island, China. Gene Geogr 2: 71-74.
149. Sukumar S, Mukherjee MB, Colah RB, Mohanty D (2004) Molecular basis of G6PD deficiency in India. Blood Cells Mol Dis 33: 141-145.
150. Chalvam R, Mukherjee MB, Colah RB, Mohanty D, Ghosh K (2007) G6PD Namoru (208 T--> C) is the major polymorphic variant in the tribal populations in southern India. Br J Haematol 136: 512-513.
151. Chalvam R, Kedar PS, Colah RB, Ghosh K, Mukherjee MB (2008) A novel R198H mutation in the glucose-6-phosphate dehydrogenase gene in the tribal groups of the Nilgiris in Southern India. J Hum Genet 53: 181-184.
152. Chalvam R, Colah RB, Mohanty D, Ghosh K, Mukherjee MB (2009) Molecular heterogeneity of glucose-6-phosphate dehydrogenase deficiency among the tribals in Western India. Blood Cells Mol Dis 43: 156-157.
153. Sarkar S, Biswas NK, Dey B, Mukhopadhyay D, Majumder PP (2010) A large, systematic molecular-genetic study of G6PD in Indian populations identifies a new non-synonymous variant and supports recent positive selection. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 10: 1228-1236.
154. Nishank SS, Chhotray GP, Kar SK, Ranjit MR (2008) Molecular variants of G6PD deficiency among certain tribal communities of Orissa, India. Ann Hum Biol 35: 355-361.
155. Beutler E, Kuhl W (1990) The NT 1311 polymorphism of G6PD: G6PD Mediterranean mutation may have originated independently in Europe and Asia. Am J Hum Genet 47: 1008-1012.
156. Sukumar S, Colah R, Mohanty D (2002) G6PD gene mutations in India producing drug-induced haemolytic anaemia. Br J Haematol 116: 671-672.
157. Ishwad CS, Naik SN (1984) A new glucose-6-phosphate dehydrogenase variant (G-6-PD Kalyan) found in a Koli family. Hum Genet 66: 171-175.
158. Sayyed Z, Mukherjee MB, Mudera VC, Colah R, Gupte S (1992) Characterization of G6PD Rohini--a new class III Indian variant. Indian J Med Res 96: 96-100.
159. Sukumar S, Mukherjee MB, Colah RB, Mohanty D (2005) Two distinct Indian G6PD variants G6PD Jamnagar and G6PD Rohini caused by the same 949 G-->A mutation. Blood cells, molecules & diseases 35: 193-195.
160. Murhekar KM, Murhekar MV, Mukherjee MB, Gorakshakar AC, Surve R, et al. (2001) Red cell genetic abnormalities, beta-globin gene haplotypes, and APOB polymorphism in the Great Andamanese, a primitive Negrito tribe of Andaman and Nicobar Islands, India. Hum Biol 73: 739-744.
161. Kaeda JS, Chhotray GP, Ranjit MR, Bautista JM, Reddy PH, et al. (1995) A new glucose-6-phosphate dehydrogenase variant, G6PD Orissa (44 Ala-->Gly), is the major polymorphic variant in tribal populations in India. Am J Hum Genet 57: 1335-1341.
162. Iwai K, Hirono A, Matsuoka H, Kawamoto F, Horie T, et al. (2001) Distribution of glucose-6-phosphate dehydrogenase mutations in Southeast Asia. Hum Genet 108: 445-449.
163. Soemantri AG, Saha S, Saha N, Tay JS (1995) Molecular variants of red cell glucose-6-phosphate dehydrogenase deficiency in Central Java, Indonesia. Hum Hered 45: 346-350.
164. Yuniar I (2006) Prevalensi varian Canton dan Kaiping pada defisiensi enzim glukosa-6-fosfat dehidrogenase yang mengalami ikterus pada masa neonatus. Jakarta: Universitas Indonesia.
165. Suhartati (2007) Pola mutan gen glukosa 6 fosfat dehidrogenase di Surabaya. Majalah Ilmu Faal Indonesia: 130-139.
166. Suhartati (2007) Mutasi gen penyebab defisiensi glukosa 6 fosfat dehidrogenase (G6PD) di Surabaya dan Kepulauan Maluku Tenggara : Tinjauan biologi molekuler dan genetika populasi terhadap G6PD. Surabaya: Universitas Airlangga.
167. Francine MR (2000) Molecular analysis for the detection of glucose-6-phosphate dehydrogenase (G6PD) deficiency. Yogyakarta: Gadjah Mada University.

168. Kawamoto F, Matsuoka H, Kanbe T, Tantular IS, Pusarawati S, et al. (2006) Further investigations of glucose-6-phosphate dehydrogenase variants in Flores Island, eastern Indonesia. J Hum Genet 51: 952-957.
169. Matsuoka H, Arai M, Yoshida S, Tantular IS, Pusarawati S, et al. (2003) Five different glucose-6-phophate [correction phosphate]dehydrogenase (G6PD) variants found among 11 G6PD-deficient persons in Flores Island, Indonesia. J Hum Genet 48: 541-544.
170. Soewono S, Martini T, Shirakawa T, Nishiyama K (2000) Glucose-6-phosphate dehydrogenase (G6PD) deficiency variants in small isolated islands in eastern Indonesia. Jurnal Kedokteran Yarsi 8: 87-92.
171. Tantular IS, Matsuoka H, Kasahara Y, Pusarawati S, Kanbe T, et al. (2010) Incidence and mutation analysis of glucose-6-phosphate dehydrogenase deficiency in eastern Indonesian populations. Acta medica Okayama 64: 367-373.
172. Suhartati, Martini T, Soewono I, Shirakawa T, Nishiyama K (2002) Molecular study in G6PD deficiency, a pedigree analysis of a Javanese-Chinese family in Surabaya, Indonesia. Jurnal Kedokteran Yarsi 10: 30-34.
173. Kurniatun Y (2000) Deteksi defisiensi dehidrogenase glukosa-6-fosfat (G6PD) pada daerah endemik malaria dengan pendekatan molekuler. Yogyakarta: Universitas Gadjah Mada.
174. Francine MR, Sofro ASM, Artama WT (2001) Molecular analysis for the detection of glucose 6-phosphate dehydrogenase (G6PD) deficiency. Teknosains 14: 169-178.
175. Karimi M, Martinez di Montemuros F, Danielli MG, Farjadian S, Afrasiabi A, et al. (2003) Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in the Fars province of Iran. Haematologica 88: 346-347.
176. Karimi M, Yavarian M, Afrasiabi A, Dehbozorgian J, Rachmilewitz E (2008) Prevalence of beta-thalassemia trait and glucose-6-phosphate dehydrogenase deficiency in Iranian Jews. Arch Med Res 39: 212-214.
177. Rahimi Z, Vaisi-Raygani A, Nagel RL, Muniz A (2006) Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in the Kurdish population of Western Iran. Blood Cells Mol Dis 37: 91-94.
178. Mesbah-Namin SA, Sanati MH, Mowjoodi A, Mason PJ, Vulliamy TJ, et al. (2002) Three major glucose-6-phosphate dehydrogenase-deficient polymorphic variants identified in Mazandaran state of Iran. Br J Haematol 117: 763-764.
179. Noori-Daloii MR, Hajebrahimi Z, Najafi L, Mesbah-Namin SA, Mowjoodi A, et al. (2007) A comprehensive study on the major mutations in glucose-6-phosphate dehydrogenase-deficient polymorphic variants identified in the coastal provinces of Caspian Sea in the north of Iran. Clin Biochem 40: 699-704.
180. Noori-Daloii MR, Najafi L, Mohammad Ganji S, Hajebrahimi Z, Sanati MH (2004) Molecular identification of mutations in G6PD gene in patients with favism in Iran. J Physiol Biochem 60: 273-277.
181. Nezhad SRK, Fahmi F, Khatami SR, Musaviun M (2011) Molecular characterization of cosenza mutation among patients with glucose-6-phosphate dehydrogenase deficiency in Khuzestan province, southwest Iran. Iranian Journal of Medical Sciences 36: 40-44.
182. Gandomani MG, Khatami SR, Nezhad SR, Daneshmand S, Mashayekhi A (2011) Molecular identification of G6PD Chatham (G1003A) in Khuzestan province of Iran. Journal of genetics 90: 143-145.
183. Mortazavi YM, Soleimani MS, Lahijani ANA, Omidkhoda AO, Ghavamzadeh A (2008) The Frequency and Molecular Genetics of G6pd Deficiency in Northwest and Southeast of Iran. Haematologica-the Hematology Journal 93: 485-486.
184. Nezhad SRK, Mashayekhi A, Khatami SR, Daneshmand S, Fahmi F, et al. (2009) Prevalence and Molecular Identification of Mediterranean Glucose-6-Phosphate Dehydrogenase Deficiency in Khuzestan Province, Iran. Iranian Journal of Public Health 38: 127-131.

185. Mortazavi Y, Mirzamohammadi F, Ardestani MT, Mirimoghadam E, Vulliamy TJ (2010) Glucos 6-phosphate dehydrogenase deficiency in Tehran, Zanjan and sistan-balouchestan provinces: Prevalence and frequency of Mediterranean variant of G6PD. Iranian Journal of Biotechnology 8: 229-233.
186. Nakhaee A, Salimi S, Zadehvakili A, Dabiri S, Noora M, et al. (2012) The Prevalence of Mediterranean Mutation of Glucose-6-Phosphate Dehydrogenase (G6PD) in Zahedan. Zahedan Journal of Research in Medical Sciences 14: 39-43.
187. Al-Allawi N, Eissa AA, Jubrael JM, Jamal SA, Hamamy H (2010) Prevalence and molecular characterization of Glucose-6-Phosphate dehydrogenase deficient variants among the Kurdish population of Northern Iraq. BMC Blood Disord 10: 6.
188. Al-Musawi BM, Al-Allawi N, Abdul-Majeed BA, Eissa AA, Jubrael JM, et al. (2012) Molecular characterization of glucose-6-phosphate dehydrogenase deficient variants in Baghdad city - Iraq. BMC blood disorders 12: 4.
189. Hilmi FA, Al-Allawi NA, Rassam M, Al-Shamma G, Al-Hashimi A (2002) Red cell glucose-6-phosphate dehydrogenase phenotypes in Iraq. East Mediterr Health J 8: 42-48.
190. Benbassat J, Ben-Ishay D (1969) Hereditary hemolytic anemia associated with glucose-6-phosphate dehydrogenase deficiency (Mediterranean type). Isr J Med Sci 5: 1053-1059.
191. Kurdi-Haidar B, Mason PJ, Berrebi A, Ankra-Badu G, al-Ali A, et al. (1990) Origin and spread of the glucose-6-phosphate dehydrogenase variant (G6PD-Mediterranean) in the Middle East. Am J Hum Genet 47: 1013-1019.
192. Ainoon O, Yu YH, Amir Muhriz AL, Boo NY, Cheong SK, et al. (2002) Glucose-6-phosphate dehydrogenase (G6PD) variants in Malaysian Malays. Hum Mutat 21: 101.
193. Ainoon O, Joyce J, Boo NY, Cheong SK, Hamidah NH (1995) Nucleotide 1376 G-->T mutation in G6PD-deficient Chinese in Malaysia. Malays J Pathol 17: 61-65.
194. Othman A, Wong F, Boo N, Wang M, Nh H (2008) Rapid molecular screening of G6PD variants in Malaysian Chinese newborns using Taqman MGB SNP assay. International Journal of Laboratory Hematology. pp. (Suppl 1) 125-125.
195. Boo NY, Wong FL, Wang MK, Othman A (2009) Homozygous variant of UGT1A1 gene mutation and severe neonatal hyperbilirubinemia. Pediatr Int 51: 488-493.
196. Ainoon O, Boo NY, Yu YH, Cheong SK, Hamidah HN, et al. (2004) Complete molecular characterisation of glucose-6-phosphate dehydrogenase (G6PD) deficiency in a group of Malaysian Chinese neonates. Malays J Pathol 26: 89-98.
197. Ainoon O, Joyce J, Boo NY, Cheong SK, Zainal ZA, et al. (1999) Glucose-6-phosphate dehydrogenase (G6PD) variants in Malaysian Chinese. Hum Mutat 14: 352.
198. Wang J, Luo E, Hirai M, Arai M, Abdul-Manan E, et al. (2008) Nine different glucose-6-phosphate dehydrogenase (G6PD) variants in a Malaysian population with Malay, Chinese, Indian and Orang Asli (aboriginal Malaysian) backgrounds. Acta Med Okayama 62: 327-332.
199. Yusoff NM, Shirakawa T, Nishiyama K, Ghazali S, Ee CK, et al. (2002) Molecular heterogeneity of glucose-6-phosphate dehydrogenase deficiency in Malays in Malaysia. Int J Hematol 76: 149-152.
200. Nadarajan V, Shanmugam H, Sthaneshwar P, Jayaranee S, Sultan KS, et al. (2011) Modification to reporting of qualitative fluorescent spot test results improves detection of glucose-6-phosphate dehydrogenase (G6PD)-deficient heterozygote female newborns. International journal of laboratory hematology 33: 463-470.
201. Matsuoka H, Wang J, Hirai M, Arai M, Yoshida S, et al. (2004) Glucose-6-phosphate dehydrogenase (G6PD) mutations in Myanmar: G6PD Mahidol (487G>A) is the most common variant in the Myanmar population. J Hum Genet 49: 544-547.
202. Nuchprayoon I, Louicharoen C, Charoenvej W (2008) Glucose-6-phosphate dehydrogenase mutations in Mon and Burmese of southern Myanmar. J Hum Genet 53: 48-54.

203. Than AM, Harano T, Harano K, Myint AA, Ogino T, et al. (2005) High incidence of 3-thalassemia, hemoglobin E, and glucose-6-phosphate dehydrogenase deficiency in populations of malaria-endemic southern Shan State, Myanmar. Int J Hematol 82: 119-123.
204. Matsuoka H, Jichun W, Hirai M, Yoshida S, Arai M, et al. (2003) Two cases of glucose-6-phosphate dehydrogenase-deficient Nepalese belonging to the G6PD Mediterranean-type, not India-Pakistan sub-type but Mediterranean-Middle East sub-type. J Hum Genet 48: 275-277.
205. Moiz B, Nasir A, Moatter T, Naqvi ZA, Khurshid M (2011) Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in Pakistani population. International journal of laboratory hematology 33: 570-578.
206. Saha N, Ramzan M, Tay JS, Low PS, Basair JB, et al. (1994) Molecular characterisation of red cell glucose-6-phosphate dehydrogenase deficiency in north-west Pakistan. Hum Hered 44: 85-89.
207. Moiz B, Nasir A, Moatter T, Naqvi ZA, Khurshid M (2009) Population study of 1311 C/T polymorphism of Glucose 6 Phosphate Dehydrogenase gene in Pakistan - an analysis of 715 X-chromosomes. BMC Genet 10: 41.
208. Leslie T, Briceno M, Mayan I, Mohammed N, Klinkenberg E, et al. (2010) The impact of phenotypic and genotypic G6PD deficiency on risk of plasmodium vivax infection: a case-control study amongst Afghan refugees in Pakistan. PLoS medicine 7: e1000283.
209. Hung NM, Eto H, Mita T, Tsukahara T, Hombhanje FW, et al. (2008) Glucose - 6 - Phosphate Dehydrogenase (G6PD) variants in East Sepik Province of Papua New Guinea : G6PD Jammu, G6PD Vanua Lava, and a novel variant (G6PD Dagua). Tropical Medicine and Health 36: 163-169.
210. Wagner G, Bhatia K, Board P (1996) Glucose-6-phosphate dehydrogenase deficiency mutations in Papua New Guinea. Hum Biol 68: 383-394.
211. Yoshida A, Giblett ER, Malcolm LA (1973) Heterogeneous distribution of glucose-6-phosphate dehydrogenase variants with enzyme deficiency in the Markham Valley Area of New Guinea. Ann Hum Genet 37: 145-150.
212. Chockkalingam K, Board PG, Nurse GT (1982) Glucose-6-phosphate dehydrogenase deficiency in Papua New Guinea. The description of 13 new variants. Hum Genet 60: 189-192.
213. Silao CL, Shirakawa T, Nishiyama K, Padilla C, Matsuo M (1999) Molecular basis of glucose-6-phosphate dehydrogenase deficiency among Filipinos. Pediatr Int 41: 138-141.
214. Yoshida A, Baur EW, Moutlsky AG (1970) A Philippino glucose-6-phosphate dehydrogenase variant (G6PD Union) with enzyme deficiency and altered substrate specificity. Blood 35: 506-513.
215. Niazi GA, Adeyokunnu A, Westwood B, Beutler E (1996) Neonatal jaundice in Saudi newborns with G6PD Aures. Ann Trop Paediatr 16: 33-37.
216. Mohamed MM, El-Humiany AUR (2006) Molecular characterization of new variants of glucose-6-phosphate dehydrogenase deficiency gene isolated in Western province of Saudi Arabia causing hemolytic anemia. Pakistan Journal of Biological Sciences 9: 1605-1616.
217. Al-Jaouni SK, Jarullah J, Azhar E, Moradkhani K (2011) Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in Jeddah, Kingdom of Saudi Arabia. BMC research notes 4: 436.
218. Faiyaz-Ul-Haque M, Zaidi SH, Hasanato RM, Al-Abdullatif A, Cluntun A, et al. (2010) Genetics of glucose-6-phosphate dehydrogenase deficiency in Saudi patients. Clinical genetics 78: 98-100.
219. Al-Ali AK (1996) Common G6PD variant from Saudi population and its prevalence. Ann Saudi Med 16: 654-656.
220. El-Hazmi MA, Warsy AS (1990) Frequency of glucose-6-phosphate dehydrogenase variants and deficiency in Arabia. Gene Geogr 4: 15-19.

221. El-Hazmi MA, Warsy AS (1991) Glucose-6-phosphate dehydrogenase variants and sickle cell genes in Al-Qunfuda, Saudi Arabia. Trop Geogr Med 43: 174-179.
222. El-Hazmi MA, Warsy AS (1992) The frequency of glucose-6-phosphate dehydrogenase phenotypes and sickle cell gene in Al-Qassim. Ann Saudi Med 12: 463-467.
223. El-Hazmi MA, Warsy AS (1994) The frequency of glucose-6-phosphate dehydrogenase phenotypes and sickle cell genes in Al-Qatif oasis. Ann Saudi Med 14: 491-494.
224. El-Hazmi MA, Warsy AS, Bahakim HH, Al-Swailem A (1994) Glucose-6-phosphate dehydrogenase deficiency and the sickle cell gene in Makkah, Saudi Arabia. J Trop Pediatr 40: 12-16.
225. El-Hazmi MA, Warsy AS, Bahakim HM, Al-Swailem A (1993) Glucose-6-phosphate dehydrogenase deficiency and sickle cell genes in two regions of western Saudi Arabia. Ann Saudi Med 13: 250-254.
226. El-Hazmi MAF, Warsy AS (1986) Glucose-6-phosphate dehydrogenase polymorphism in the Saudi population. Hum Hered 36: 24-30.
227. El-Hazmi MAF, Warsy AS (1993) The Frequency of Hbs and Glucose-6-Phosphate-Dehydrogenase Phenotypes in Relation to Malaria in Western Saudi-Arabia. Saudi Medical Journal 14: 121-125.
228. Gelpi AP, King MC (1977) New data on glucose-6-phosphate dehydrogenase deficiency in Saudi Arabia. G6PD variants, and the association between enzyme deficiency and hemoglobins S. Hum Hered 27: 285-291.
229. Hellani A, Al-Akoum S, Abu-Amero KK (2009) G6PD Mediterranean S188F codon mutation is common among Saudi sickle cell patients and increases the risk of stroke. Genet Test Mol Biomarkers 13: 449-452.
230. Gari MA, Chaudhary AG, Al-Qahtani MH, Abuzenadah AM, Waseem A, et al. (2010) Frequency of Mediterranean mutation among a group of Saudi G6PD patients in Western region-Jeddah. International journal of laboratory hematology 32: 17-21.
231. Samuel AP, Saha N (1986) Distribution of red cell G6PD and 6PGD phenotypes in Saudi Arabia. Trop Geogr Med 38: 287-291.
232. Hirono A, Ishii A, Kere N, Fujii H, Hirono K, et al. (1995) Molecular analysis of glucose-6-phosphate dehydrogenase variants in the Solomon Islands. Am J Hum Genet 56: 1243-1245.
233. Laosombat V, Sattayasevana B, Janejindamai W, Viprakasit V, Shirakawa T, et al. (2005) Molecular heterogeneity of glucose-6-phosphate dehydrogenase (G6PD) variants in the south of Thailand and identification of a novel variant (G6PD Songklanagarind). Blood Cells Mol Dis 34: 191-196.
234. Nuchprayoon I, Sanpavat S, Nuchprayoon S (2002) Glucose-6-phosphate dehydrogenase (G6PD) mutations in Thailand: G6PD Viangchan (871G>A) is the most common deficiency variant in the Thai population. Hum Mutat 19: 185.
235. Panich V (1973) G6pd-Characterization in Thailand. Genetics 74: S208-S208.
236. Panich V, Sungnate T (1973) Characterization of glucose-6-phosphate dehydrogenase in Thailand. The occurrence of 6 variants among 50 G-6-PD deficient Thai. Humangenetik 18: 39-46.
237. Phompradit P, Kuesap J, Chaijaroenkul W, Rueangweerayut R, Hongkaew Y, et al. (2011) Prevalence and distribution of glucose-6-phosphate dehydrogenase (G6PD) variants in Thai and Burmese populations in malaria endemic areas of Thailand. Malar J 10: 368.
238. Ninokata A, Kimura R, Samakkarn U, Settheetham-Ishida W, Ishida T (2006) Coexistence of five G6PD variants indicates ethnic complexity of Phuket islanders, Southern Thailand. J Hum Genet 51: 424-428.
239. Laosombat V, Sattayasevana B, Chotsampancharoen T, Wongchanchailert M (2006) Glucose-6-phosphate dehydrogenase variants associated with favism in Thai children. Int J Hematol 83: 139-143.
240. Tanphaichitr VS, Hirono A, Pung-amritt P, Treesucon A, Wanachiwanawin W (2011) Chronic nonspherocytic hemolytic anemia due to glucose-6-phosphate dehydrogenase deficiency:

report of two families with novel mutations causing G6PD Bangkok and G6PD Bangkok Noi. Annals of hematology 90: 769-775.
241. Panich V, Sungnate T, Wasi P, Na-Nakorn S (1972) G-6-PD Mahidol. The most common glucose-6-phosphate dehydrogenase variant in Thailand. J Med Assoc Thai 55: 576-585.
242. Louicharoen C, Patin E, Paul R, Nuchprayoon I, Witoonpanich B, et al. (2009) Positively selected G6PD-Mahidol mutation reduces *Plasmodium vivax* density in Southeast Asians. Science 326: 1546-1549.
243. Oner R, Gumruk F, Acar C, Oner C, Gurgey A, et al. (2000) Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in Turkey. Haematologica 85: 320-321.
244. Cappellini MD, di Montemuros FM, Prandoni S, Tavazzi D, Iolascon A, et al. (2001) Molecular characterization of G6PD deficiency in subjects negative for common Mediterranean mutations. Blood 98: 13a-13a.
245. Canatan D, Bagci H, Gumuslu S, Bilmen S, Acikbas I, et al. (2006) The features of patients with favism in Turkey. HAEMA 9: 247-250.
246. Keskin N, Ozdes I, Keskin A, Acikbas I, Bagci H (2002) Incidence and molecular analysis of glucose-6-phosphate dehydrogenase deficiency in the province of Denizli, Turkey. Med Sci Monit 8: CR453-456.
247. Yildiz SM, Ariyurek SY, Aksoy K (2010) Detection of Mediterranean Mutation in the Glucose-6-Phosphate Dehydrogenase Gene with Microarray Technique. Turkish Journal of Biochemistry-Turk Biyokimya Dergisi 35: 63-66.
248. Ganczakowski M, Town M, Bowden DK, Vulliamy TJ, Kaneko A, et al. (1995) Multiple glucose 6-phosphate dehydrogenase-deficient variants correlate with malaria endemicity in the Vanuatu archipelago (southwestern Pacific). Am J Hum Genet 56: 294-301.
249. Matsuoka H, Thuan DT, van Thien H, Kanbe T, Jalloh A, et al. (2007) Seven different glucose-6-phosphate dehydrogenase variants including a new variant distributed in Lam Dong Province in southern Vietnam. Acta Med Okayama 61: 213-219.
250. Panich V, Bumrungtrakul P, Jitjai C, Kamolmatayakul S, Khoprasert B, et al. (1980) Glucose-6-phosphate dehydrogenase deficiency in South Vietnamese. Hum Hered 30: 361-364.
251. Toncheva D (1986) Variants of glucose-6-phosphate dehydrogenase in a Vietnamese population. Hum Hered 36: 348-351.
252. Hue NT, Charlieu JP, Chau TT, Day N, Farrar JJ, et al. (2009) Glucose-6-phosphate dehydrogenase (G6PD) mutations and haemoglobinuria syndrome in the Vietnamese population. Malar J 8: 152.

Dataset S1. Published and unpublished sources from which surveys were identified. Only sources reporting surveys included in the final model are listed. Population samples from these met the inclusion criteria for community representativeness, enzymatic diagnosis, sex-specificity and were geographically specific. References are listed alphabetically by first author name; n=261.

1. Abeyaratne KP, Premawansa S, Rajapakse L, Roberts DF, Pipiha SS (1976) A survey of glucose-6-phosphate-dehydrogenase deficiency in the North Central Province of Sri Lanka (formerly Ceylon). Am J Phys Anthropol 44: 135-138.
2. Abreu de Miani MS, Penalver JA (1983) [Incidence of beta-thalassemia carriers and those deficient in erythrocyte glucose-6-phosphate dehydrogenase in the greater Buenos Aires area]. Sangre (Barc) 28: 537-541.
3. Ademowo OG, Falusi AG (2002) Molecular epidemiology and activity of erythrocyte G6PD variants in a homogeneous Nigerian population. East Afr Med J 79: 42-44.
4. Ainoon O, Yu YH, Amir Muhriz AL, Boo NY, Cheong SK, et al. (2002) Glucose-6-phosphate dehydrogenase (G6PD) variants in Malaysian Malays. Hum Mutat 21: 101.
5. Akinkugbe FM (1980) Anaemia in a rural population in Nigeria (Ilora). Ann Trop Med Parasitol 74: 625-633.
6. Al Arrayed S (2005) Campaign to control genetic blood diseases in Bahrain. Community Genet 8: 52-55.
7. Ali N, Anwar M, Ayyub M, Bhatti FA, Nadeem M, et al. (2005) Frequency of glucose-6-phosphate dehydrogenase deficiency in some ethnic groups of Pakistan. J Coll Physicians Surg Pak 15: 137-141.
8. Allison AC (1960) Glucose-6-phosphate dehydrogenase deficiency in red blood cells of East Africans. Nature 186: 531-532.
9. Allison AC, Charles LJ, McGregor IA (1961) Erythrocyte glucose-6-phosphate dehydrogenase deficiency in West Africa. Nature 190: 1198-1199.
10. al-Nuaim L, Talib ZA, el-Hazmi MA, Warsy AS (1997) Sickle cell and G-6-PD deficiency gene in cord blood samples: experience at King Khalid University Hospital, Riyadh. J Trop Pediatr 43: 71-74.
11. al-Riyami A, Ebrahim GJ (2003) Genetic blood disorders survey in the Sultanate of Oman. J Trop Pediatr 49 Suppl 1: i1-20.
12. Amini F, Ismail E, Zilfalil BA (2011) Prevalence and molecular study of G6PD deficiency in Malaysian Orang Asli. Intern Med J 41: 351-353.
13. Amin-Zaki L, el-Din ST, Kubba K (1972) Glucose-6-phosphate dehydrogenase deficiency among ethnic groups in Iraq. Bull World Health Organ 47: 1-5.
14. Amoozegar H, Mirshekari M, Pishva N (2006) Does the history before blood transfusion identify donors who are glucose-6-phosphate dehydrogenase (G-6-PD) deficient? Turk J Hematol 23: 147-150.
15. Arambula E, Aguilar LJ, Vaca G (2000) Glucose-6-phosphate dehydrogenase mutations and haplotypes in Mexican Mestizos. Blood Cells Mol Dis 26: 387-394.
16. Ardati KO, Bajakian KM, Mohammad AM, Coe EL (1995) Glucose-6-phosphate-dehydrogenase phenotypes in Bahrain - quantitative-analysis and electrophoretic characterization. Saudi Med J 16: 102-104.
17. Askerova TA, Kichibekov BR, Movsum-zade KM (1992) [Hereditary glucose-6-phosphate dehydrogenase deficiency in newborn infants]. Pediatriia: 10-13.
18. Azevedo ES, Alves AF, Da Silva MC, Souza MG, Muniz Dias Lima AM, et al. (1980) Distribution of abnormal hemoglobins and glucose-6-phosphate dehydrogenase variants in 1200 school children of Bahia, Brazil. Am J Phys Anthropol 53: 509-512.
19. Azevedo ES, Costa Silva KM, Da Silva MCBO, Dias Lima AM, Mascaronhas Fortuna CM, et al. (1981) Genetic and anthropological studies in the island of Itaparica, Bahia, Brazil. Hum Hered 31: 353-357.

20. Azhar A (1998) Kajian genetika biokemis dehidrogenase glukosa-6-fosfat (G6PD) dan dehidrogenase 6-fosfoglukonat (6-PGD) pada tiga populasi Nusa Tenggara. Yogyakarta: Universitas Gadjah Mada.
21. Azhar A, Husin A (2001) Prevalence of glucose 6-phosphate dehydrogenase (G6PD) deficiency in two populations of Aceh province. Jurnal Kedokteran Yarsi 9: 93-95.
22. Azim AA, Kamel K, Gaballah MF, Sabry FH, Ibrahim W, et al. (1974) Genetic blood markers and anthropometry of the populations in Aswan Governorate, Egypt. Hum Hered 24: 12-23.
23. Azofeifa J, Barrantes R (1991) Genetic variation in the Bribri and Cabecar Amerindians from Talamanca, Costa Rica. Rev Biol Trop 39: 249-253.
24. Badens C, Leclaire M, Collomb J, Auquier P, Soyer P, et al. (2001) [Glucose-6-phosphate dehydrogenase et neonatal jaundice]. Presse Med 30: 524-526.
25. Baer A, Lie-Injo LE, Welch QB, Lewis AN (1976) Genetic factors and malaria in the Temuan. Am J Hum Genet 28: 179-188.
26. Balgir RS (2007) Genetic burden of red cell enzyme glucose-6-phosphate dehydrogenase deficiency in two major Scheduled Tribes of Sundargarh district, Northwestern Orissa, India. Current Science 92: 768-774.
27. Balgir RS, Sharma JC (1988) Genetic markers in the Hindu and Muslim Gujjars of Northwestern India. Am J Phys Anthropol 75: 391-403.
28. Banerjee B, Saha N, Daoud ZF, Khalaf FH, Qudah H (1981) A genetic study of the Jordanians. Hum Hered 31: 65-69.
29. Barretto OC, Nonoyama K (1978) [Malaria-dependent polymorphism related to erythrocyte glucose-6-phosphate dehydrogenase and glutathione among Brazilian Indians]. Rev Hosp Clin Fac Med Sao Paulo 33: 231-233.
30. Barretto OCO (1970) Erythrocyte glucose-6-phosphate dehydrogenase deficiency in Sao Paulo, Brazil. Rev Bras Pesqui Med Biol 3: 61-65.
31. Basu S, Jindal A, Kumar CS, Khan AS (1995) Genetic marker profile of primitive Kutia Kondh tribal population of Phulbani district (Orissa). Indian J Med Res 101: 36-38.
32. Bayoumi RA, Taha TS, Saha N (1985) A study of some genetic characteristics of the Fur and Baggara tribes of the Sudan. Am J Phys Anthropol 67: 363-370.
33. Beck HP, Felger I, Kabintik S, Tavul L, Genton B, et al. (1994) Assessment of the humoral and cell-mediated immunity against the *Plasmodium falciparum* vaccine candidates circumsporozoite protein and SPf66 in adults living in highly endemic malarious areas of Papua New Guinea. Am J Trop Med Hyg 51: 356-364.
34. Benabadji M, Benlatrache C, Merad F, Suaudeau C, Benmoussa M, et al. (1977) [Glucose-6-phosphate dehydrogenase deficiency in Algeria.]. Sem Hop 53: 899-904.
35. Bernstein RE (1965) Inborn errors of metabolism in Central Africa: red cell glucose-6-phosphate dehydrogenase deficiency and sickle haemoglobin. In: GJ S, editor. Science and Medicine in Central Africa. NY: Pergamon Press. pp. 739-747.
36. Bernstein SC, Bowman JE, Kaptue Noche L (1980) Population studies in Cameroon: hemoglobin S, glucose-6-phosphate dehydrogenase deficiency and falciparum malaria. Hum Hered 30: 251-258.
37. Bernstein SC, Bowman JE, Noche LK (1980) Interaction of sickle cell trait and glucose-6-phosphate dehydrogenase deficiency in Cameroon. Hum Hered 30: 7-11.
38. Best WR (1959) Absence of erythrocyte glucose-6-phosphate dehydrogenase deficiency in certain Peruvian Indians. J Lab Clin Med 54: 791.
39. Bienzle U, Ayeni O, Lucas AO, Luzzatto L (1972) Glucose-6-phosphate dehydrogenase and malaria. Greater resistance of females heterozygous for enzyme deficiency and of males with non-deficient variant. Lancet 1: 107-110.
40. Bienzle U, Okoye VC, Gogler H (1972) Haemoglobin and glucose-6-phosphate dehydrogenase variants: distribution in relation to malaria endemicity in a Togolese population. Z Tropenmed Parasitol 23: 56-62.
41. Blibech R, Gharbi Y, Mrad A, Zahra H, Mahjoub T, et al. (1989) Incidence of glucose-6-phosphate dehydrogenase (G6PD) deficiency in Tunisian populations. Nouv Rev Fr Hematol 31: 189-191.

42. Blinov MN, Rodriges N, Sanches Perovani Kh A (1973) [Incidence in the glucose-6-phosphate dehydrogenase deficiency of erythrocytes in the population of North Oriente province (Republic of Cuba)]. Probl Gematol Pereliv Krovi 18: 26-29.
43. Bloch M, Rivera H (1969) [Abnormal hemoglobins and glucose-6-phosphate dehydrogenase deficiency in El Salvador]. Sangre (Barc) 14: 121-124.
44. Bonne B, Godber M, Ashbel S, Mourant AE, Tills D (1971) South-Sinai Beduin. A preliminary report on their inherited blood factors. Am J Phys Anthropol 34: 397-408.
45. Bottini N, Meloni G, Porcu S, Gloria-Bottini F (2001) Cyclic seasonal variation of G-6-PD deficiency in newborn infants from Sardinia. Biol Rhythm Res 32: 413-421.
46. Bowman JE, Paul E. Carson, Henri Frischer, Robin D. Powell, Edward J. Colwell, et al. (1971) Hemoglobin and red cell enzyme variation in some populations of the republic of Vietnam with comments on the malaria hypothesis. Am J Phys Anthropol 34: 313-324.
47. Brabin L, Brabin BJ (1990) Malaria and glucose 6-phosphate dehydrogenase deficiency in populations with high and low spleen rates in Madang, Papua New Guinea. Hum Hered 40: 15-21.
48. Buchanan JG, Wilson FS, Nixon AD (1973) Survey for erythrocyte glucose-6-phosphate dehydrogenase deficiency in Fiji. Am J Hum Genet 25: 36-41.
49. Budtz-Olsen O, Kidson C (1961) Absence of red cell enzyme deficiency in Australian Aborigines. Nature 192: 765.
50. Cao A, Congiu R, Sollaino MC, Desogus MF, Demartis FR, et al. (2008) Thalassaemia and glucose-6-phosphate dehydrogenase screening in 13- to 14-year-old students of the Sardinian population: preliminary findings. Community Genet 11: 121-128.
51. Castaneda BF, Colwell EJ, Phintuyothin P, Hickman RJ (1972) Investigations of the fluorescent spot test for erythrocyte glucose-6-phosphate dehydrogenase deficiency in southeast Thailand. J Med Assoc Thai 55: 331-338.
52. Ceda-Flores RM, Arriaga-Rios G, Munoz-Campos J, Bautista-Pena VA, Angeles Rojas-Alvarado M, et al. (1990) [Frequency of color blindness and glucose-6-phosphate dehydrogenase enzyme deficiency in non-industrialized populations in the state of Nuevo Leon, Mexico]. Arch Invest Med (Mex) 21: 229-234.
53. Chan TK (1983) Glucose 6 phosphate dehydrogenase (G6PD) [MD Thesis]: University of Hong Kong.
54. Chan YK, Tay MT, Lim MK (1992) Xq28: epidemiology and sex-linkage between red-green colour blindness and G6PD deficiency. Ann Acad Med Singapore 21: 318-322.
55. Chien YH, Lee NC, Wu ST, Liou JJ, Chen HC, et al. (2008) Changes in incidence and sex ratio of glucose-6-phosphate dehydrogenase deficiency by population drift in Taiwan. Southeast Asian J Trop Med Public Health 39: 154-161 & additional data from authors.
56. Choremis C, Fessas P, Kattamis C, Stamatoyannopoulos G, Zannos-Mariolea L, et al. (1963) Three inherited red-cell abnormalities in a district of Greece. Thalassaemia, sickling, and glucose-6-phosphate-dehydrogenase deficiency. Lancet 1: 907-909.
57. Choremis C, Zannos-Mariolea L, Katamis MD (1962) Frequency of glucose-6-phosphate-dehydrogenase deficiency in certain highly malarious areas of Greece. Lancet 1: 17-18.
58. Chowdhury A (1976) Glucose-6-phosphate-dehydrogenase in a Bengalee sample of Calcutta. Man India 56: 263-268.
59. Cladera Serra A, Oliva Berini E, Torrent Quetglas M, Bartolozzi Castilla E (1997) [Prevalence of glucose-6-phosphate dehydrogenase deficiency in a student population on the island of Menorca]. Sangre (Barc) 42: 363-367.
60. Cocco P, Manca P, Dessi S (1987) Preliminary results of a geographic correlation study on G6PD deficiency and cancer. Toxicol Pathol 15: 106-108.
61. Compri MB, Saad ST, Ramalho AS (2000) [Genetico-epidemiological and molecular investigation of G-6-PD deficiency in a Brazilian community]. Cad Saude Publica 16: 335-342.
62. Daar S, Vulliamy TJ, Kaeda J, Mason PJ, Luzzatto L (1996) Molecular characterization of G6PD deficiency in Oman. Hum Hered 46: 172-176.

63. DaCosta H, Pattani J, Dandekar S, Kotnis U, Mehendale K, et al. (1967) Glucose-6-phosphate dehydrogenase (G-6-PD) defect in Maharashtrian children. Indian J Med Sci 21: 809-812.
64. David S, Trincao C (1963) [Drepanocytemia, erythrocytic glucose-6-phosphate dehydrogenase deficiency (G-6-Pd) and malaria in the Cuango Post (Lunda-Angola).]. An Inst Med Trop (Lisb) 20: 5-15.
65. Devi ST, Saran SK, Nair G (1993) Study of glucose-6-phosphate dehydrogenase (G6PD) in the Kissan tribals of Orissa and the Kannikar tribals of Kerala, India. Anthropol Anz 51: 179-181.
66. Diatewa M, Ganga-Zanzou SP, Gangoue N, Miehakanda J (1992) [Neonatal icterus and erythrocyte glucose-6-phosphate dehydrogenase deficiency in Congolese newborn infants in Brazzaville]. Arch Fr Pediatr 49: 939-940.
67. Doeblin TD, Ingall GB, Pinkerton PH, Donambaju KR, Bannerman RM (1968) Genetic studies of the Seneca Indians: haptoglobins, transferrins, G-6-PD Deficiency, hemoglobinopathy, color blindness, morphological traits and dermatoglyphics. Hum Hered 18: 251-260.
68. Doxiadis SA, Karaklis A, Valaes T, Stavrakakis D (1964) Risk of severe jaundice in glucose-6-phosphate-dehydrogenase deficiency of the newborn. Differences in population groups. Lancet 2: 1210-1212.
69. Dube RK, Dube B, Gupta YN (1976) Erythrocytic glucose-6-phosphate dehydrogenase deficiency at Varanasi. Indian J Pathol Microbiol 19: 245-251.
70. Duflo B, Diallo A, Toure K, Soula G (1979) [Glucose-6-phosphate dehydrogenase deficiency in Mali. Epidemiology and pathological aspects]. Bull Soc Pathol Exot Filiales 72: 258-264.
71. Duflo B, Ranque P, Quilici M, Balique H, Dembele O, et al. (1982) Glucose-6-phosphate-dehydrogenase deficiency and malaria in Mali. Nouvelle Presse Medicale 11: 2713-2713.
72. Egesie OJ, Joseph DE, Isiguzoro I, Egesie UG (2008) Glucose-6-phosphate dehydrogenase (G6PD) activity and deficiency in a population of Nigerian males resident in Jos. Niger J Physiol Sci 23: 9-11.
73. El-Hazmi MA, Warsy AS (1987) Interaction between glucose-6-phosphate dehydrogenase deficiency and sickle cell gene in Saudi Arabia. Trop Geogr Med 39: 32-35.
74. El-Hazmi MAF, Warsy AS (1997) Phenotypes of glucose-6-phosphate dehydrogenase in different regions of Saudi Arabia - A comparative assessment. Saudi Med J 18: 393-399.
75. El-Migdadi F, Al-Tellawi A, Al-Hussain S, Rawashdeh M (2008) Pyruvate kinase and glucose-6-phosphate dehydrogenase activities in children living above (Jordan Valley) and below (Amman and Irbid) sea level. Journal of Chinese Clinical Medicine 3: 627-632.
76. Eng LL, Giok PH (1964) Glucose-6-phosphate dehydrogenase deficiency in Indonesia. Nature 204: 88-89.
77. Eng LL, Ti TS (1964) Glucose-6-phosphate dehydrogenase deficiency in Malayans. Trans R Soc Trop Med Hyg 58: 500-502.
78. Estrada M, Gonzalez R (1983) [Neonatal jaundice and glucose-6-phosphate dehydrogenase deficiency in Havana]. Rev Invest Clin 35: 297-299.
79. Fernando WP, Ratnapala PR (1988) Report: A survey to ascertain the prevalence of G-6-PD enzyme deficiency in Sri Lanka, and its geographical and ethnical distribution. funded by ISTI.
80. Flatz G, Chakravartti MR, Das BM, Delbruck H (1972) Genetic survey in the population of Assam. I. ABO blood groups, glucose-6-phosphate dehydrogenase and haemoglobin type. Hum Hered 22: 323-330.
81. Flatz G, Sringam S (1963) Malaria and glucose-6-phosphate dehydrogenase deficiency in Thailand. Lancet 2: 1248-1250.
82. Flatz G, Sringam S, Premyothin C, Penbharkkul S, Ketusingh R, et al. (1963) Glucose-6-phosphate dehydrogenase deficiency and neonatal jaundice. Arch Dis

Childh 38: 566-570.
83. Fraser GR, Defaranas B, Kattamis CA, Race RR, Sanger R, et al. (1964) Glucose-6-phosphate dehydrogenase, colour vision and Xg blood groups in Greece: linkage and population data. Ann Hum Genet 27: 395-403.
84. Fraser GR, Grunwald P, Stamatoyannopoulos G (1966) Glucose-6-phosphate dehydrogenase (G6PD) deficiency, abnormal haemoglobins, and thalassaemia in Yugoslavia. J Med Genet 3: 35-41.
85. Fraser GR, Stamatoyannopoulos G, Kattamis C, Loukopoulos D, Defaranas B, et al. (1964) Thalassemias, abnormal hemoglobins and glucose-6-phosphate dehydrogenase deficiency in the Arta area of Greece: diagnostic and genetic aspects of complete village studies. Ann N Y Acad Sci 119: 415-435.
86. Fraser GR, Steinberg AG, Defaranas B, Mayo O, Stamatoyannopoulos G, et al. (1969) Gene frequencies at loci determining blood-group and serum-protein polymorphisms in two villages of northwestern Greece. Am J Hum Genet 21: 46-60.
87. Ganczakowski M, Town M, Bowden DK, Vulliamy TJ, Kaneko A, et al. (1995) Multiple glucose 6-phosphate dehydrogenase-deficient variants correlate with malaria endemicity in the Vanuatu archipelago (southwestern Pacific). Am J Hum Genet 56: 294-301.
88. Garcia SC, Moragon AC, Lopez-Fernandez ME (1979) Frequency of glutathione reductase, pyruvate kinase and glucose-6-phosphate dehydrogenase deficiency in a Spanish population. Hum Hered 29: 310-313.
89. Garg A, Bhatia BD, Chaturvedi P, Garg S (1984) G6PD deficiency in newborn infants. Indian J Pediatr 51: 29-33.
90. Garlipp CR, Ramalho AS (1988) [Clinical and laboratory aspects of glucose-6-phosphate-dehydrogenase (G-6-Pd) deficiency in Brazilian newborns]. Rev Bras Genet 11: 717-728.
91. Geerdink RA, Bartstra HA, Schillhorn Van Veen JM (1974) Serum proteins and red cell enzymes in Trio and Wajana Indians from Surinam. Am J Hum Genet 26: 581-587.
92. Geerdink RA, Okhura K, Li Fo Sjoe E, Schillhorn van Veen JM, Bartstra HA (1975) Serum factors and red cell enzymes in Carib and Arowak Indians from Surinam. Trop Geogr Med 27: 269-273.
93. Gelpi AP (1965) Glucose-6-phosphate dehydrogenase deficiency in Saudi Arabia: a survey. Blood 25: 486-493.
94. Gelpi AP (1967) Glucose-6-phosphate dehydrogenase deficiency, the sickling trait, and malaria in Saudi Arab children. J Pediatr 71: 138-146.
95. Ghosh K, Mukherjee MB, Shankar U, Kote SL, Nagtilak SB, et al. (2002) Clinical examination and hematological data in asymptomatic & apparently healthy school children in a boarding school in a tribal area. Indian J Public Health 46: 61-65.
96. Gibbs WN, Ottey F, Dyer H (1972) Distribution of glucose-6-phosphate dehydrogenase phenotypes in Jamaica. Am J Hum Genet 24: 18-23.
97. Giles E, Curtain CC, Baumgarten A (1967) Distribution of beta-thalassemia trait and erythrocyte glucose-6-phosphate dehydrogenase deficiency in the Markham River Valley of New Guinea. Am J Phys Anthropol 27: 83-88.
98. Gualandri V, Orsini GB, Porta E, Gerli GC (1983) [Glucose-6-phosphate dehydrogenase deficiency among the student population of Milan]. J Genet Hum 31: 201-209.
99. Gupta JC, Yagnik U, Seth P (1982) Incidence of glucose-6-phosphate dehydrogenase deficiency in Jabalpur area. Indian J Pathol Microbiol 25: 66-69.
100. Gupta S, Ghai OP, Chandra RK (1970) Glucose-6-phosphate dehydrogenase deficiency in the newborn and its relation to serum bilirubin. Indian J Pediatr 37: 169-176.
101. Gupte SC, Shaw AN, Shah KC (2005) Hematological findings and severity of G6PD deficiency in Vataliya Prajapati subjects. J Assoc Physicians India 53: 1027-1030.
102. Haghighi B, Suzangar M, Yazdani A, Mehnat M (1985) A genetic variant of human erythrocyte glucose 6-phosphate dehydrogenase. Biochem Biophys Res Commun 132: 1151-1159.
103. Harley JD, Agar NS, Turner TB (1976) Letter: sickle-cell anaemia and trait in Sydney.

Med J Aust 1: 894.
104. Hashmi JA, Farzana F, Ahmed M (1976) Abnormal hemoglobins, thalasemia trait & G6PD deficiency in young Pakistani males. J Pak Med Assoc 26: 2-4.
105. Hoan NKH (2010) Personal communication: unpublished data from Vietnam.
106. Hussein L, Yamamah G, Saleh A (1992) Glucose-6-phosphate dehydrogenase deficiency and sulfadimidin acetylation phenotypes in Egyptian oases. Biochem Genet 30: 113-121.
107. Ibrahim WN, Kamel K, Selim O, Azim A, Gaballah MF, et al. (1974) Hereditary blood factors and anthropometry of the inhabitants of the Egyptian Siwa Oasis. Hum Biol 46: 57-68.
108. Idel'son LI, Kotoian ER (1970) [Incidence of glucose-6-phosphate dehydrogenase deficiency in erythrocytes of the population in Armenia]. Probl Gematol Pereliv Krovi 15: 39-44.
109. Itskan SB, Saldanha PH (1975) [Erythrocyte glucose-6-phosphate dehydrogenase activity in the population of a malarial region in Sao Paulo (Iguape)]. Rev Inst Med Trop Sao Paulo 17: 83-91.
110. Jenkins T, Blecher SR, Smith AN, Anderson CG (1968) Some hereditary red-cell traits in Kalahari Bushmen and Bantu: hemoglobins, glucose-6-phosphate dehydrogenase deficiency, and blood groups. Am J Hum Genet 20: 299-309.
111. Jeremiah ZA, Uko EK, Usanga EA (2008) Relation of nutritional status, sickle cell trait, glucose-6-phosphate dehydrogenase deficiency, iron deficiency and asymptomatic malaria infection in the Niger Delta, Nigeria. J Med Sci 8: 269-274.
112. Jiang J, Ma X, Song C, Lin B, Cao W, et al. (2003) Using the fluorescence spot test for neonatal screening of G6PD deficiency. Southeast Asian J Trop Med Public Health 34 Suppl 3: 140-142.
113. Jiang W, Yu G, Liu P, Geng Q, Chen L, et al. (2006) Structure and function of glucose-6-phosphate dehydrogenase-deficient variants in Chinese population. Hum Genet 119: 463-478.
114. Johnson MK, Clark TD, Njama-Meya D, Rosenthal PJ, Parikh S (2009) Impact of the method of G6PD deficiency assessment on genetic association studies of malaria susceptibility. PLoS One 4: e7246.
115. Jolly JG, Sarup BM, Bhatnagar DP, Maini SC (1972) Glucose-6-phosphate dehydrogenase deficiency in India. J Indian Med Assoc 58: 196-200.
116. Kageoka T, Satoh C, Goriki K, Fujita M, Neriishi S, et al. (1985) Electrophoretic variants of blood proteins in Japanese. IV. Prevalence and enzymologic characteristics of glucose-6-phosphate dehydrogenase variants in Hiroshima and Nagasaki. Hum Genet 70: 101-108.
117. Kamal I, Gabr M, Mohyeldin O, Talaat M (1967) Frequency of glucose6-phosphate dehydrogenase deficiency in Egyptian children. Acta Genet Stat Med 17: 321-327.
118. Kamel K, Umar M, Ibrahim W, Mansour A, Gaballah F, et al. (1975) Anthropological studies among Libyans. Erythrocyte genetic factors, serum haptoglobin phenotypes and anthropometry. Am J Phys Anthropol 43: 103-111.
119. Kaneko A, Taleo G, Kalkoa M, Yaviong J, Reeve PA, et al. (1998) Malaria epidemiology, glucose 6-phosphate dehydrogenase deficiency and human settlement in the Vanuatu Archipelago. Acta Trop 70: 285-302.
120. Kaplanoglou LB, Triantaphyllidis CD (1982) Genetic polymorphisms in a North-Greek population. Hum Hered 32: 124-129.
121. Kate SL, Mukherjee BN, Malhotra KC, Phadke MA, Mutalik GS, et al. (1978) Red cell glucose-6-phosphate dehydrogenase deficiency and haemoglobin variants among ten endogamous groups of Maharshtra and West Bengal. Hum Genet 44: 339-343.
122. Kate SL, Phadke MA, Sainani GS, Mutalik GS (1976) Study of erythrocyte glucose-6-phosphate dehydrogenase and abnormal haemoglobins in an endogamous community--"Katkaris"--a survey. J Assoc Physicians India 24: 1-3.
123. Kattamis CA, Chaidas A, Chaidas S (1969) G6PD deficiency and favism in the island of Rhodes (Greece). J Med Genet 6: 286-291.

124. Kidson C (1961) Deficiency of glucose-6-phosphate dehydrogenase: some aspects of the trait in people of Papua-New Guinea. Med J Aust 48(2): 506-509.
125. Kigoni EP, Kujwalile JM, Nhonoli AM (1978) Frequency of glucose-6-phosphate dehydrogenase (G-6-PD) deficiency and its relationship to hepatitis B surface antigen (HBS-Ag) in normal Tanzanian males. East Afr Med J 55: 247-251.
126. Kirk RL, Keats B, Blake NM, McDermid EM, Ala F, et al. (1977) Genes and people in the Caspian Littoral: a population genetic study in Northern Iran. Am J Phys Anthropol 46: 377-390.
127. Kirkman HN, Walker DH (1982) Glucose-6-phosphate-dehydrogenase deficiency and Mediterranean fever in northern Sardinia - Reply. J Infect Dis 146: 302-302.
128. Knight RH, Robertson DH (1963) The prevalence of the erythrocyte glucose-6-phosphate dehydrogenase deficiency among Africans in Uganda. Trans R Soc Trop Med Hyg 57: 95-100.
129. Kotea R, Kaeda JS, Yan SL, Sem Fa N, Beesoon S, et al. (1999) Three major G6PD-deficient polymorphic variants identified among the Mauritian population. Br J Haematol 104: 849-854.
130. Krasnopol'skaia KD, Filippov IK, Sotnikova EN, Movsum-zade KM, Gadzhiev BO (1980) [Patterns in the distribution of GPD- alleles in Azerbaijan. I. Incidence and polymorphism of glucose-6-phosphate dehydrogenase deficiency in the Shekii region of the Azerbaijan SSR]. Genetika 16: 1685-1692.
131. Krasnopol'skaia KD, Iakovlev SA, Smirnova OA, Prytkov AN (1985) [Patterns of Gd- allele distribution in Azerbaijan. IV. The incidence and polymorphism of erythrocytic glucose-6-phosphate dehydrogenase deficiency in the settlement of Kobi, Apsheron District]. Genetika 21: 487-492.
132. Krasnopolskaya XD, Shatskaya TL (1987) Distribution of Gd- alleles in some ethnic groups of the USSR. Hum Genet 75: 258-263.
133. Krželj V, Markić J, Karaman K, Ćurin K, Unić I, et al. (2011) Personal communication: unpublished data from Croatia.
134. Kuhn VL, Lisboa V, de Cerqueira LP (1983) [Glucose-6-phosphate dehydrogenase deficiency in blood donors in a general hospital of Salvador, Bahia, Brazil]. Rev Paul Med 101: 175-177.
135. Kumakawa T, Suzuki S, Fujii H, Miwa S (1987) Frequency of glucose 6-phosphate dehydrogenase (G6PD) deficiency in Tokyo and a new variant: G6PD Musashino. Nippon Ketsueki Gakkai Zasshi 50: 25-28.
136. Kuwahata M, Wijesinghe R, Ho MF, Pelecanos A, Bobogare A, et al. (2010) Population screening for glucose-6-phosphate dehydrogenase deficiencies in Isabel Province, Solomon Islands, using a modified enzyme assay on filter paper dried bloodspots. Malar J 9: 223.
137. Lai HC, Lai MP, Leung KS (1968) Glucose-6-phosphate dehydrogenase deficiency in Chinese. J Clin Pathol 21: 44-47.
138. Le Xuan C, Le Si Q, Humbert C, Chu Quang G (1968) [Glucose-6-phosphate dehydrogenase deficiency in Viet-nam]. Nouv Rev Fr Hematol 8: 878-884.
139. Lefevre-Witier P, Vergnes H (1977) Enzyme polymorphisms of Ideles populations (Ahaggar, Algeria) and the Iwellemeden Kel Kummer Twaregs (Menaka, Mali). Hum Hered 27: 454-469.
140. Lehmann H, Ala F, Hedeyat S, Montazemi K, Nejad HK, et al. (1973) Biological studies of Yemenite and Kurdish Jews in Israel and other groups in south-west Asia. XI. The hereditary blood factors of the Kurds of Iran. Phil Trans R Soc London B 266: 195-205.
141. Lisker R, Cordova MS, Graciela Zarate QB (1969) Studies on several genetic hematological traits of the Mexican population. XVI. Hemoglobin, S and glucose-6-phosphate dehydrogenase deficiency in the east coast. Am J Phys Anthropol 30: 349-354.
142. Lisker R, Loria A, Cordova MS (1965) Studies on several genetic hematological traits of the Mexican population. VIII. Hemoglobin S, glucose-6-phosphate dehydrogenase deficiency, and other characteristics in a malarial region. Am J Hum Genet 17:

179-187.
143. Lisker R, Loria A, Gonzales Llaven J, Guttman S, Ruiz Reyes G (1962) [Preliminary note on the incidence of abnormal hemoglobulins and glucose-6-phosphate dehydrogenase deficiency in the Mexican population.]. Rev Fr Etud Clin Biol 7: 76-78.
144. Lothe F (1967) Erythrocyte glucose-6-phosphate dehydrogenase deficiency in Uganda. Nature 215: 299-300.
145. Lysenko A, Abrashkin-Zhuchkov RG, Alekseeva MI, Gorbunova Iu P, Krasil'nikov AA (1973) [Incidence of hereditary deficiency of glucose-6-phosphate dehydrogenase activity of erythrocytes in Azerbaijan SSR]. Probl Gematol Pereliv Krovi 18: 16-21.
146. Madanat F, Karadsheh N, Shamayleh A, Tarawneh M, Khraisha S, et al. (1986) Glucose-6-phosphate dehydrogenase deficiency in male newborns. Jordan Medical Journal 21: 205-212.
147. Markic J, Krzelj V, Markotic A, Marusic E, Stricevic L, et al. (2006) High incidence of glucose-6-phosphate dehydrogenase deficiency in Croatian island isolate: example from Vis island, Croatia. Croat Med J 47: 556-570.
148. Martinez-Labarga C, Rickards O, Scacchi R, Corbo RM, Biondi G, et al. (1999) Genetic population structure of two African-Ecuadorian communities of Esmeraldas. Am J Phys Anthropol 109: 159-174.
149. Martins MC, Olim G, Melo J, Magalhaes HA, Rodrigues MO (1993) Hereditary anaemias in Portugal: epidemiology, public health significance, and control. J Med Genet 30: 235-239.
150. Mathews ST, Kumaresan PR, Selvam R (1991) Glucose-6-phosphate dehydrogenase deficiency and malaria--a study on north Madras population. J Commun Dis 23: 178-181.
151. Matsuoka H, Arai M, Yoshida S, Tantular IS, Pusarawati S, et al. (2003) Five different glucose-6-phophate [correction phosphate]dehydrogenase (G6PD) variants found among 11 G6PD-deficient persons in Flores Island, Indonesia. J Hum Genet 48: 541-544.
152. Matsuoka H, Ishii A, Panjaitan W, Sudiranto R (1986) Malaria and glucose-6-phosphate dehydrogenase deficiency in North Sumatra, Indonesia. Southeast Asian J Trop Med Public Health 17: 530-536.
153. Matsuoka H, Nguon C, Kanbe T, Jalloh A, Sato H, et al. (2005) Glucose-6-phosphate dehydrogenase (G6PD) mutations in Cambodia: G6PD Viangchan (871G>A) is the most common variant in the Cambodian population. J Hum Genet 50: 468-472.
154. Mayer G, Mayoux A (1966) Recherches d'anomalies sanguines genetiques dans une population de la cote Est de Madagascar. Arch Inst Pasteur Madagascar 35: 209-211.
155. McGuinness R, Saunders RA (1967) Erythrocyte galactose-I-phosphate uridyl transferase and glucose-6-phosphate dehydrogenase activity in the population of the Rhondda Fach. Clin Chim Acta 16: 221-226.
156. Ménard D (2011) Personal communication: unpublished data from Cambodia.
157. Miall WE, Milner PF, Lovell HG, Standard KL (1967) Haematological investigations of population samples in Jamaica. Br J Prev Soc Med 21: 45-55.
158. Miguel A, Ramon M, Petitpierre E, Goos CM, Vermeesch-Markslag AM, et al. (1983) Population screening for glucose-6-phosphate dehydrogenase deficiency on the Baleares. Hum Genet 64: 176-179.
159. Milbauer B, Peled N, Svirsky S (1973) Neonatal hyperbilirubinemia and glucose-6-phosphate dehydrogenase deficiency. Isr J Med Sci 9: 1547-1552.
160. Mir NA, Fakhri M, Abdelaziz M, Kishan J, Elzouki A, et al. (1985) Erythrocyte glucose-6-phosphate dehydrogenase status of newborns and adults in eastern Libya. Ann Trop Paediatr 5: 211-213.
161. Mohammed N, Amanzai O, Rashid H, Jan S, Leslie T (2010) Report: Assessment of the prevalence of G6PD deficiency in Afghanistan. HealthNet TPO Malaria Control Programme, funded by a Global Fund for AIDS TB and Malaria Round 5 grant.
162. Monchy D, Babin FX, Srey CT, Ing PN, von Xylander S, et al. (2004) [Frequency of G6PD deficiency in a group of preschool-aged children in a centrally located area of

Cambodia]. Med Trop (Mars) 64: 355-358.
163. Mortazavi YM, Soleimani MS, Lahijani ANA, Omidkhoda AO, Ghavamzadeh A (2008) The Frequency and Molecular Genetics of G6pd Deficiency in Northwest and Southeast of Iran. Haematol-Hematol J 93: 485-486.
164. Moscarelli G, Ferraro G, Infantone MA, Scola S, Montaperto A, et al. (1999) Screening della glucosio 6 fosfato deidrogenasi in 500 gravide. Giornale Italiano di Ostetricia e Ginecologia 21: 27-28.
165. Motulsky AG, Vandepitte J, Fraser GR (1966) Population genetic studies in the Congo. I. Glucose-6-phosphate dehydrogenase deficiency, hemoglobin S, and malaria. Am J Hum Genet 18: 514-537.
166. Mourant AE, Kopec AC, Ikin EW, Lehmann H, Bowen-Simpkins P, et al. (1974) The blood groups and haemoglobins of the Kunama and Baria of Eritrea, Ethiopia. Ann Hum Biol 1: 383-392.
167. Murhekar KM, Murhekar MV, Mukherjee MB, Gorakshakar AC, Surve R, et al. (2001) Red cell genetic abnormalities, beta-globin gene haplotypes, and APOB polymorphism in the Great Andamanese, a primitive Negrito tribe of Andaman and Nicobar Islands, India. Hum Biol 73: 739-744.
168. Muzaffer MA (2005) Neonatal screening of glucose-6-phosphate dehydrogenase deficiency in Yanbu, Saudi Arabia. J Med Screen 12: 170-171.
169. Nasserullah Z, Al Jame A, Abu Srair H, Al Qatari G, Al Naim S, et al. (1998) Neonatal screening for sickle cell disease, glucose-6-phosphate dehydrogenase deficiency and a-thalassemia in Qatif and Al Hasa. Ann Saudi Med 18: 289-292.
170. Neto JPD, Dourado MV, dos Reis MG, Goncalves MS (2008) A novel c. 197T -> A variant among Brazilian neonates with glucose-6-phosphate dehydrogenase deficiency. Genet Mol Biol 31: 33-35.
171. Nezhad SRK, Mashayekhi A, Khatami SR, Daneshmand S, Fahmi F, et al. (2009) Prevalence and molecular identification of mediterranean glucose-6-phosphate dehydrogenase deficiency in Khuzestan Province, Iran. Iran J Public Health 38: 127-131.
172. Nicolielo DB, Ferreira RIP, Leite AA (2006) Activity of 6-phosphogluconate dehydrogenase in glucose-6-phosphate dehydrogenase deficiency. Rev Bras Hematol Hemoter 28: 135-138.
173. Nieuwenhuis F, Wolf B, Bomba A, De Graaf P (1986) Haematological study in Cabo Delgado province, Mozambique; sickle cell trait and G6PD deficiency. Trop Geogr Med 38: 183-187.
174. Ninokata A, Kimura R, Samakkarn U, Settheetham-Ishida W, Ishida T (2006) Coexistence of five G6PD variants indicates ethnic complexity of Phuket islanders, Southern Thailand. J Hum Genet 51: 424-428.
175. Nixon AD, Buchanan JG (1969) Survey for erythrocyte glucose-6-phosphate dehydrogenase deficiency in Polynesians. Am J Hum Genet 21: 305-309.
176. Nosten F, Bancone G (2011) Personal communication: unpublished data from Thailand.
177. Nuchprayoon I, Louicharoen C, Charoenvej W (2008) Glucose-6-phosphate dehydrogenase mutations in Mon and Burmese of southern Myanmar. J Hum Genet 53: 48-54.
178. Nuchprayoon I, Sanpavat S, Nuchprayoon S (2002) Glucose-6-phosphate dehydrogenase (G6PD) mutations in Thailand: G6PD Viangchan (871G>A) is the most common deficiency variant in the Thai population. Hum Mutat 19: 185.
179. Nurse GT, Jenkins T (1977) Serogenetic studies on the Kavango peoples of South West Africa. Ann Hum Biol 4: 465-478.
180. Nurse GT, Jenkins T, David JH, Steinberg AG (1979) The Njinga of Angola: a serogenetic study. Ann Hum Biol 6: 337-348.
181. Nwankwo MU, Bunker CH, Ukoli FA, Omene JA, Freeman DT, et al. (1990) Blood pressure and other cardiovascular disease risk factors in black adults with sickle cell trait or glucose-6-phosphate dehydrogenase deficiency. Genet Epidemiol 7: 211-218.
182. Ohkura K, Miyashita T, Nakajima H, Matsumoto H, Matsutomo K, et al. (1984)

Distribution of polymorphic traits in Mazandaranian and Guilanian in Iran. Hum Hered 34: 27-39.
183. Omer A, Ali M, Omer AH, Mustafa MD, Satir AA, et al. (1972) Incidence of G-6-PD deficiency and abnormal haemoglobins in the indigenous and immigrant tribes of the Sudan. Trop Geogr Med 24: 401-405.
184. Ondei LS, Silveira LM, Leite AA, Souza DR, Pinhel MA, et al. (2009) Lipid peroxidation and antioxidant capacity of G6PD-deficient patients with A-(202G>A) mutation. Genet Mol Res 8: 1345-1351.
185. Oudart JL, Tchernia G, Giscard R, Boal MR, Zucker JM, et al. (1971) [Erythrocyte glucose-6-phosphate dehydrogenase deficiency in African newborn infants in Dakar]. Afr J Med Sci 2: 87-100.
186. Padilla C (2011) Personal communication: unpublished data from the Philippines.
187. Palmarino R, Agostino R, Gloria F, Lucarelli P, Businco L, et al. (1975) Red cell acid phosphatase: another polymorphism correlated with Malaria? Am J Phys Anthropol 43: 177-186.
188. Pao M, Kulkarni A, Gupta V, Kaul S, Balan S (2005) Neonatal screening for glucose-6-phosphate dehydrogenase deficiency. Indian J Pediatr 72: 835-837.
189. Parsons IC, Ryan BPK (1962) Observations on Glucose-6-Phosphate Dehydrogenase Deficiency in Papuans. Med J Aust 2: 585-587.
190. Patel S (1977) Incidence of glucose-6-phosphate-dehydrogenase deficiency and correlation with some other laboratory findings among Tibetan refugees in Orissa. J Anthropol Soc Nip 85: 347-349.
191. Perine PL, Michael MT (1974) A preliminary survey for glucose-6-phosphate dehydrogenase deficiency and haemoglobin S in Ethiopia. Ethiop Med J 12: 179-184.
192. Plato CC, Cruz MT, Kurland LT (1964) Frequency of glucose-6-phosphate dehydrogenase deficiency red-green colour blindness and Xga blood-group among Chamorros. Nature 202: 728.
193. Plato CC, Rucknagel DL, Gershowitz H (1964) Studies on the distribution of glucose-6-phosphate dehydrogenase deficiency, thalassemia, and other genetic traits in the coastal and mountain villages of Cyprus. Am J Hum Genet 16: 267-283.
194. Prins HK, Loos JA, Meuwissen JH (1963) Glucose-6-phosphate dehydrogenase (G6pd) deficiency in West New Guinea. Trop Geogr Med 15: 361-370.
195. Ragab AH, el-Alfi OS, Abboud MA (1966) Incidence of glucose-6-phosphate dehydrogenase deficiency in Egypt. Am J Hum Genet 18: 21-25.
196. Rahimi Z, Raygani AV, Siabani S, Mozafari H, Nagel RL, et al. (2008) Prevalence of glucose-6-phosphate dehydrogenase deficiency among schoolboys in Kermanshah, Islamic Republic of Iran. East Mediterr Health J 14: 978-979.
197. Ramalho AS, Beiguelman B (1977) [Glucosephosphate dehydrogenase deficiency (G6-PD) in Brazilian blood donors]. AMB; Revista da Associação Médica Brasileira 23: 259-260.
198. Ratrisawadi V, Horpaopan S, Chotigeat U, Sangtawesin V, Kanjanapattanakul W, et al. (1999) Neonatal screening program in Rajavithi Hospital, Thailand. Southeast Asian J Trop Med Public Health 30 Suppl 2: 28-32.
199. Reclos GJ, Hatzidakis CJ, Schulpis KH (2000) Glucose-6-phosphate dehydrogenase deficiency neonatal screening: preliminary evidence that a high percentage of partially deficient female neonates are missed during routine screening. J Med Screen 7: 46-51.
200. Restrepo AM, Gutierrez E (1968) The frequency of glucose-6-phosphate dehydrogenase deficiency in Colombia. Am J Hum Genet 20: 82-85.
201. Reys L, Manso C, Stamatoyannopoulos G (1970) Genetic studies on southeastern Bantu of Mozambique. I. Variants of glucose-6-phosphate dehydrogenase. Am J Hum Genet 22: 203-215.
202. Richard F, Belhani M, Colonna P (1974) [G-6PD deficiency in newborns in Algiers (author's transl)]. Nouv Rev Fr Hematol 14: 453-459.
203. Ringelhann B, Dodu SRA, Konotey-Ahulu FID, Lehmann H (1968) A survey for

haemoglobin variants, thalassaemia and glucose-6-phosphate dehydrogenase deficiency in Northern Ghana. Ghana Med J 7: 120-124.
204. Roberts DF, Triger DR, Morgan RJ (1970) Glucose-6-phosphate dehydrogenase deficiency and haemoglobin level in Jamaican children. West Indian Med J 19: 204-211.
205. Saha N, Bhattacharyya SP, Mukhopadhyay B, Bhattacharyya SK, Gupta R, et al. (1987) A genetic study among the Lepchas of the Darjeeling area of eastern India. Hum Hered 37: 113-121.
206. Saha N, Hong SH, Wong HA, Tay JS (1990) Red cell glucose-6-phosphate dehydrogenase phenotypes in several Mongoloid populations of eastern India: existence of a non-deficient fast variant in two Australasian tribes. Ann Hum Biol 17: 529-532.
207. Saha N, Ramzan M, Tay JS, Low PS, Basair JB, et al. (1994) Molecular characterisation of red cell glucose-6-phosphate dehydrogenase deficiency in north-west Pakistan. Hum Hered 44: 85-89.
208. Saha N, Tay JS (1990) Genetic studies among the Nagas and Hmars of eastern India. Am J Phys Anthropol 82: 101-112.
209. Saleem TH, Mendis BS, Osanyintuyi SO (1991) Glucose-6-phosphate dehydrogenase deficiency in a rural Saudi population. J Trop Med Hyg 94: 327-328.
210. Samuel AP, Saha N, Omer A, Hoffbrand AV (1981) Quantitative expression of G6PD activity of different phenotypes of G6PD and haemoglobin in a Sudanese population. Hum Hered 31: 110-115.
211. Sanchez MC, Villegas VE, Fonseca D (2008) [Glucose-6-phosphate dehydrogenase deficiency: enzimatic and molecular analysis in a Bogota population]. Colombia Medica 39: 14-23.
212. Sans M, Alvarez I, Bentancor N, Abilleira D, Bengochea M, et al. (1995) Blood protein genetic-markers in a northeastern Uruguayan population. Rev Bras Genet 18: 317-320.
213. Santana MS, de Lacerda MV, Barbosa MG, Alecrim WD, Alecrim MG (2009) Glucose-6-phosphate dehydrogenase deficiency in an endemic area for malaria in Manaus: a cross-sectional survey in the Brazilian Amazon. PLoS One 4: e5259.
214. Sarma DK, Shukla R, Lodha A, Abdulla A, Pataridze L (2006) Neonatal screening for glucose-6-phosphate dehydrogenase (G6PD) deficiency: Experience in a private hospital. Emirates Med J 24: 211-214.
215. Say B, Ozand P, Berkel I, Cevik N (1965) Erythrocyte glucose-6-phosphate dehydrogenase deficiency in Turkey. Acta Paediatr Scand 54: 319-324.
216. Schuurkamp GJ, Bhatia KK, Kereu RK, Bulungol PK (1989) Glucose-6-phosphate dehydrogenase deficiency and hereditary ovalocytosis in the Ok Tedi impact region of Papua New Guinea. Hum Biol 61: 387-406.
217. Segeja MD, Mmbando BP, Kamugisha ML, Akida JA, Savaeli ZX, et al. (2008) Prevalence of glucose-6-phosphate dehydrogenase deficiency and haemoglobin S in high and moderate malaria transmission areas of Muheza, north-eastern Tanzania. Tanzan J Health Res 10: 9-13.
218. Seth PK, Seth S (1971) Biogenetical studies of Nagas: glucose-6-phosphate dehydrogenase deficiency in Angami Nagas. Hum Biol 43: 557-561.
219. Sethuraman M, Rao KV (1978) A survey of glucose-6-phosphate dehydrogenase deficiency & sickle-cell trait on a local population of Tirupati. Indian J Exp Biol 16: 1098-1099.
220. Shah SS, Macharia A, Uyoga S, Williams TN (2011) Personal communication: unpublished data from Kenya.
221. Shimizu H, Tamam M, Soemantri A, Ishida T (2005) Glucose-6-phosphate dehydrogenase deficiency and Southeast Asian ovalocytosis in asymptomatic *Plasmodium* carriers in Sumba island, Indonesia. J Hum Genet 50: 420-424.
222. Singh H (1986) Glucose-6-phosphate dehydrogenase deficiency: a preventable cause of mental retardation. Br Med J (Clin Res Ed) 292: 397-398.

223. Siniscalco M, Bernini L, Filippi G, Latte B, Meera Khan P, et al. (1966) Population genetics of haemoglobin variants, thalassaemia and glucose-6-phosphate dehydrogenase deficiency, with particular reference to the malaria hypothesis. Bull World Health Organ 34: 379-393.
224. Sonnet J, Michaux JL (1960) Glucose-6-phosphate dehydrogenase deficiency, haptoglobin groups, blood groups and sickle cell trait in the Bantus of west Belgian Congo. Nature 188: 504-505.
225. Sözüöz A, Çamber I (1998) G6PD deficiency in Turkish Cypriots. Turk J Med Sci 28: 673-676.
226. Stamatoyannopoulos G, Fessas P (1964) Thalassaemia, glucose-6-phosphate dehydrogenase deficiency, sickling, and malarial endemicity in Greece: a study of Five Areas. Br Med J 1: 875-879.
227. Stamatoyannopoulos G, Panayotopoulos A, Motulsky AG (1966) The distribution of glucose-6-phosphate dehydrogenase deficiency in Greece. Am J Hum Genet 18: 296-308.
228. Suradi R, Monitja HE, Munthe BG, Suparno (1979) Glucose-6-phosphate dehydrogenase deficiency in the Dr. Cipto Mangunkusumo General Hospital. Paediatr Indones 19: 30-40.
229. Suryantoro P (2003) Glucose-6-phosphate dehydrogenase (G6PD) deficiency in Yogyakarta and its surrounding areas. Southeast Asian J Trop Med Public Health 34 Suppl 3: 138-139.
230. Sutton RN (1963) Erythrocyte glucose-6-phosphate-dehydrogenase deficiency in Trinidad. Lancet 1: 855.
231. Syahyuni R (2003) Hubungan defisiensi glucose-6-phosphate dehydrogenase (G-6-PD) degan kepadatan parasit malaria pada anak usia sekolah di daerah endemis malaria [Author translation: Glucose-6-phosphate defficiency during school children in malaria endemic area]. Semarang: Universitas Diponegoro.
232. Szathmary EJE, Cox DW, Gershowitz H, Rucknagel DL, Schanfield MS (1974) The Northern and Southeastern Ojibwa: serum proteins and red cell enzyme systems. Am J Phys Anthropol 40: 49-65.
233. Tagarelli A, Bastone L, Cittadella R, Calabro V, Bria M, et al. (1991) Glucose-6-phosphate dehydrogenase (G6PD) deficiency in southern Italy: a study on the population of the Cosenza province. Gene Geogr 5: 141-150.
234. Tagarelli A, Cittadella R, Bria M, Brancati C (1992) Glucose-6-phosphate dehydrogenase (G6PD) deficiency in the Albanian ethnic minority of Cosenza province, Italy. Gene Geogr 6: 71-78.
235. Talafih K, Hunaiti AA, Gharaibeh N, Gharaibeh M, Jaradat S (1996) The prevalence of hemoglobin S and glucose-6-phosphate dehydrogenase deficiency in Jordanian newborn. J Obstet Gynaecol Res 22: 417-420.
236. Taleb N, Loiselet J, Guorra F, Sfeir H (1964) [on glucose-6-phosphate dehydrogenase deficiency in autochthonous populations of Lebanon.]. C R Hebd Seances Acad Sci 258: 5749-5751.
237. Tantular IS, Iwai K, Lin K, Basuki S, Horie T, et al. (1999) Field trials of a rapid test for G6PD deficiency in combination with a rapid diagnosis of malaria. Trop Med Int Health 4: 245-250.
238. Tartaglia M, Scano G, DeStefano GF (1996) An anthropogenetic study on the Oromo and Amhara of central Ethiopia. Am J Hum Biol 8: 505-516.
239. Thakur A, Verma IC (1992) Interaction of malarial infection and glucose-6-phosphate dehydrogenase deficiency in Muria gonds of district Bastar, central India. Trop Geogr Med 44: 201-205.
240. Tills D, Warlow A, Lord JM, Suter D, Kopec AC, et al. (1983) Genetic factors in the population of Plati, Greece. Am J Phys Anthropol 61: 145-156.
241. Tsoneva M, Proinova N, Mavrudieva M (1974) [Incidence of glucose-6-phosphate dehydrogenase deficiency in blood donors of Sofia]. Vutr Boles 13: 46-51.
242. Tuchinda S, Rucknagel DL, Na-Nakorn S, Wasi P (1968) The Thai variant and the

distribution of alleles of 6-phosphogluconate dehydrogenase and the distribution of glucose 6-phosphate dehydrogenase deficiency in Thailand. Biochem Genet 2: 253-264.
243. Tuda JSB, Kepel BJ, Nakatsu M, Matsuoka H (2007) Prevalensi defisiensi Glucose-6-Phosphate Dehydrogenase (G6PD) pada anak Sekolah Dasar yang tinggal di daerah endemis malaria di Sulawesi utara. Jurnal Kedokteran Yarsi 15: 59-63.
244. Turan Y (2006) Prevalence of erythrocyte glucose-6-phosphate dehydrogenase (G6PD) deficiency in the population of western Turkey. Arch Med Res 37: 880-882.
245. Usanga EA, Ameen R (2000) Glucose-6-phosphate dehydrogenase deficiency in Kuwait, Syria, Egypt, Iran, Jordan and Lebanon. Hum Hered 50: 158-161.
246. Voronov AA, Krasilnikov AA (1973) [Population study of glucose-6-phosphate deficiency in the Transcaucasus]. Probl Gematol Pereliv Krovi 18: 21-23.
247. Walter H, Neumann S, Nemeskeri J (1968) Investigations on the occurrence of glucose-6-phosphate-dehydrogenase deficiency in Hungary. Acta Genet Stat Med 18: 1-11.
248. Weimer TA, Salzano FM, Hutz MH (1981) Erythrocyte isozymes and hemoglobin types in a southern Brazilian population. J Hum Evol 10: 319-328.
249. Welch SG, Lee J, McGregor IA, Williams K (1978) Red cell glucose 6 phosphate dehydrogenase genotypes of the population of two West African villages. Hum Genet 43: 315-320.
250. White JM, Byrne M, Richards R, Buchanan T, Katsoulis E, et al. (1986) Red cell genetic abnormalities in Peninsular Arabs: sickle haemoglobin, G6PD deficiency, and alpha and beta thalassaemia. J Med Genet 23: 245-251.
251. Willcox M, Bjorkman A, Brohult J (1983) Falciparum malaria and beta-thalassaemia trait in northern Liberia. Ann Trop Med Parasitol 77: 335-347.
252. Woodfield DG, Scragg RFR, Blake NM (1974) Distribution of blood, serum protein and enzyme groups among the Fuyuge speakers of the Goilala sub district. Hum Hered 24: 507-519.
253. Wu CX, Shan KR, He Y, Qi XL, Li Y, et al. (2007) Detection of glucose-6-phosphate dehydrogenase gene mutations of Tujia ehtnic in Jiangkou, Guizhou. Chinese Journal of Endemiology 26: 415-417.
254. Xiu J, Qi XL, Shan KR, Xie Y, He Y, et al. (2005) [G6PD gene mutations in Shui people in Sandu of Guizhou]. Zhongguo Shi Yan Xue Ye Xue Za Zhi 13: 147-150.
255. Yamamoto T, Amano H, Sano A, Takahash.Y, Takahash.H (1974) Study about glucose-6-phosphate-dehydrogenase deficiency in Laos. Jpn J Hum Genet 19: 64-64.
256. Yenchitsomanus P, Summers KM, Chockkalingam C, Board PG (1986) Characterization of G6PD deficiency and thalassaemia in Papua New Guinea. P N G Med J 29: 53-58.
257. Young GP, Smith MB, Woodfield DG (1974) Glucose-6-phosphate dehydrogenase deficiency in Papua New Guinea using a simple methylene blue reduction test. Med J Aust 1: 876-878.
258. Yudhaputri FA, Baird JK, Nixon C (2011) Personal communication: unpublished data from Indonesia.
259. Yue PC, Strickland M (1965) Glucose-6-phosphate-dehydrogenase deficiency and neonatal jaundice in Chinese male infants in Hong Kong. Lancet 1: 350-351.
260. Zaidman JL, Leiba H, Scharf S, Steinman I (1976) Red cell glucose-6-phosphate dehydrogenase deficiency in ethnic groups in Israel. Clin Genet 9: 131-133.
261. Zannos-Mariolea L, Kattamis C (1961) Glucose-6-phosphate dehydrogenase deficiency in Greece. Blood 18: 34-47.

OPEN ACCESS Freely available online

PLoS NEGLECTED TROPICAL DISEASES

The International Limits and Population at Risk of *Plasmodium vivax* Transmission in 2009

Carlos A. Guerra[1]*, Rosalind E. Howes[1], Anand P. Patil[1], Peter W. Gething[1], Thomas P. Van Boeckel[1,2], William H. Temperley[1], Caroline W. Kabaria[3], Andrew J. Tatem[4,5], Bui H. Manh[6], Iqbal R. F. Elyazar[7], J. Kevin Baird[7,8], Robert W. Snow[3,9], Simon I. Hay[1]*

1 Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, United Kingdom, **2** Biological Control and Spatial Ecology, Université Libre de Bruxelles, CP160/12, Brussels, Belgium, **3** Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, KEMRI - University of Oxford - Wellcome Trust Collaborative Programme, Nairobi, Kenya, **4** Department of Geography, University of Florida, Gainesville, Florida, United States of America, **5** Emerging Pathogens Institute, University of Florida, Gainesville, Florida, United States of America, **6** Oxford University Clinical Research Unit, Bach Mai Hospital, National Institute of Infectious and Tropical Diseases, Ha Noi, Vietnam, **7** Eijkman-Oxford Clinical Research Unit, Jakarta, Indonesia, **8** Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, United Kingdom, **9** Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, CCVTM, Oxford, United Kingdom

Abstract

Background: A research priority for *Plasmodium vivax* malaria is to improve our understanding of the spatial distribution of risk and its relationship with the burden of *P. vivax* disease in human populations. The aim of the research outlined in this article is to provide a contemporary evidence-based map of the global spatial extent of *P. vivax* malaria, together with estimates of the human population at risk (PAR) of any level of transmission in 2009.

Methodology: The most recent *P. vivax* case-reporting data that could be obtained for all malaria endemic countries were used to classify risk into three classes: malaria free, unstable ($<$0.1 case per 1,000 people per annum (p.a.)) and stable ($\geq$0.1 case per 1,000 p.a.) *P. vivax* malaria transmission. Risk areas were further constrained using temperature and aridity data based upon their relationship with parasite and vector bionomics. Medical intelligence was used to refine the spatial extent of risk in specific areas where transmission was reported to be absent (e.g., large urban areas and malaria-free islands). The PAR under each level of transmission was then derived by combining the categorical risk map with a high resolution population surface adjusted to 2009. The exclusion of large Duffy negative populations in Africa from the PAR totals was achieved using independent modelling of the gene frequency of this genetic trait. It was estimated that 2.85 billion people were exposed to some risk of *P. vivax* transmission in 2009, with 57.1% of them living in areas of unstable transmission. The vast majority (2.59 billion, 91.0%) were located in Central and South East (CSE) Asia, whilst the remainder were located in America (0.16 billion, 5.5%) and in the Africa+ region (0.10 billion, 3.5%). Despite evidence of ubiquitous risk of *P. vivax* infection in Africa, the very high prevalence of Duffy negativity throughout Central and West Africa reduced the PAR estimates substantially.

Conclusions: After more than a century of development and control, *P. vivax* remains more widely distributed than *P. falciparum* and is a potential cause of morbidity and mortality amongst the 2.85 billion people living at risk of infection, the majority of whom are in the tropical belt of CSE Asia. The probability of infection is reduced massively across Africa by the frequency of the Duffy negative trait, but transmission does occur on the continent and is a concern for Duffy positive locals and travellers. The final map provides the spatial limits on which the endemicity of *P. vivax* transmission can be mapped to support future cartographic-based burden estimations.

Citation: Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The International Limits and Population at Risk of *Plasmodium vivax* Transmission in 2009. PLoS Negl Trop Dis 4(8): e774. doi:10.1371/journal.pntd.0000774

Editor: Jane M. Carlton, New York University School of Medicine, United States of America

Received March 17, 2010; **Accepted** June 24, 2010; **Published** August 3, 2010

Funding: SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#079091), which also supports CAG, PWG, and CWK. REH is funded by a Biomedical Resources Grant (#085406) from the Wellcome Trust to SIH. RWS is funded by a Wellcome Trust Principal Research Fellowship (#079080), which also supports APP and WHT. TPVB is funded by a grant from the Belgian Fond National pour la Recherche Scientifique and the Fondation Wiener-Anspach. AJT is supported by a grant from the Bill and Melinda Gates Foundation (#49446). BHM is funded by a grant from the University of Oxford - Li Ka Shing Foundation Global Health Programme. IRFE is funded by grants from the University of Oxford - Li Ka Shing Foundation Global Health Programme, the United States Navy, and the Oxford Tropical Network. JKB is funded by a grant from the Wellcome Trust (#B9RJIXO) and by the South East Asia Infectious Disease Research Network. This work forms part of the output of the Malaria Atlas Project (MAP, www.map.ox.ac.uk), principally funded by the Wellcome Trust, United Kingdom. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: carlos.guerraloaiza@zoo.ox.ac.uk (CAG); simon.hay@zoo.ox.ac.uk (SIH)

Author Summary

Growing evidence shows that *Plasmodium vivax* malaria is clinically less benign than has been commonly believed. In addition, it is the most widely distributed species of human malaria and is likely to cause more illness in certain regions than the more extensively studied *P. falciparum* malaria. Understanding where *P. vivax* transmission exists and measuring the number of people who live at risk of infection is a fundamental first step to estimating the global disease toll. The aim of this paper is to generate a reliable map of the worldwide distribution of this parasite and to provide an estimate of how many people are exposed to probable infection. A geographical information system was used to map data on the presence of *P. vivax* infection and spatial information on climatic conditions that impede transmission (low ambient temperature and extremely arid environments) in order to delineate areas where transmission was unlikely to take place. This map was combined with population distribution data to estimate how many people live in these areas and are, therefore, exposed to risk of infection by *P. vivax* malaria. The results show that 2.85 billion people were exposed to some level of risk of transmission in 2009.

Introduction

The bulk of the global burden of human malaria is caused by two parasites: *Plasmodium falciparum* and *P. vivax*. Existing research efforts have focussed largely on *P. falciparum* because of the mortality it causes in Africa [1,2]. This focus is increasingly regarded as untenable [3–6] because the following factors indicate that the public health importance of *P. vivax* may be more significant than traditionally thought: i) *P. vivax* has a wider geographical range, potentially exposing more people to risk of infection [7,8]; ii) it is less amenable to control [9,10]; and, most importantly, iii) infections with *P. vivax* can cause severe clinical syndromes [5,11–16].

A key research priority for *P. vivax* malaria is to improve the basic understanding of the geographical distribution of risk, which is needed for adequate burden estimation [6]. Recent work by the Malaria Atlas Project (MAP; www.map.ox.ac.uk) [17] has shown *P. falciparum* malaria mapping to be a fundamental step in understanding the epidemiology of the disease at the global scale [18,19], in appraising the equity of global financing for control [20] and in forming the basis for burden estimation [21,22]. The benefits of a detailed knowledge of the spatial distribution of *P. vivax* transmission, and its clinical burden within these limits, are identical to those articulated for *P. falciparum*: establishing a benchmark against which control targets may be set, budgeted and monitored. Such maps do not exist for *P. vivax*, making any strategic planning problematic. In addition, information about the global extent of *P. vivax* transmission and population at risk (PAR) is crucial for many nations that are re-evaluating their prospects for malaria elimination [23,24].

This paper documents the global spatial limits of *P. vivax* malaria using a combination of national case-reporting data from health management information systems (HMIS), biological rules of transmission exclusion and medical intelligence combined in a geographical information system. The output is an evidence-based map from which estimates of PAR are derived. The resulting map also provides the global template in which contemporary *P. vivax* endemicity can be estimated and it contributes to a cartographic basis for *P. vivax* disease burden estimation.

Methods

Analyses Outline

A schematic overview of the analyses is presented in Figure 1. Briefly, *P. vivax* malaria endemic countries (*Pv*MECs) were first identified and the following layers were progressively applied within a geographical information system to constrain risk areas and derive the final *P. vivax* spatial limits map: i) a *P. vivax* annual parasite incidence (*Pv*API) data layer; biological exclusion layers comprising of ii) temperature and iii) aridity data layers; iv) a medical intelligence exclusion layer; and v) a predicted Duffy negativity layer. A detailed description of these steps follows.

Identifying *Pv*MECs

Those countries that currently support *P. vivax* transmission were first identified. The primary sources for defining national risk were international travel and health guidelines [25,26] augmented with national survey information, pertinent published sources and personal communication with malariologists. Nations were grouped into three regions, as described elsewhere [19]: i) America; ii) Africa, Saudi Arabia and Yemen (Africa+); and iii) Central and South East (CSE) Asia. To further resolve PAR estimates, the CSE Asia region was sub-divided into West Asia, Central Asia and East Asia (Protocol S1).

Mapping case-reporting data

Methods described previously for mapping the global spatial limits of *P. falciparum* malaria [18] were used to constrain the area defined at risk within the *Pv*MECs using *Pv*API data (the number of confirmed *P. vivax* malaria cases reported per administrative unit per 1,000 people per annum (p.a.)). The *Pv*API data were obtained mostly through personal communication with individuals and institutions linked to malaria control in each country (Protocol S1). The format in which these data were available varied considerably between countries. Ideally, the data would be available by administrative unit and by year, with each record presenting the estimated population for the administrative unit and the number of confirmed autochthonous malaria cases by the two main parasite species (*P. falciparum* and *P. vivax*). This would allow an estimation of species-specific API. These requirements, however, were often not met. Population data by administrative unit were sometimes unavailable, in which cases these data were sourced separately or extrapolated from previous years. An additional problem was the lack of parasite species-specific case or API values. In such cases, a parasite species ratio was inferred from alternative sources and applied to provide an estimate of species-specific API. There was, thus, significant geographical variation in the ability to look at the relative frequency of these parasites between areas and this was not investigated further. Finally, although a differentiation between confirmed and suspected cases and between autochthonous and imported cases was often provided, whenever this was not available it was assumed that the cases in question referred to confirmed and autochthonous occurrences.

The aim was to collate data for the last four years of reporting (ideally up to 2009) at the highest spatial resolution available (ideally at the second administrative level (ADMIN2) or higher). A geo-database was constructed to archive this information and link it to digital administrative boundaries of the world available from the 2009 version of the Global Administrative Unit Layers (GAUL) data set, implemented by the Food and Agriculture Organization of the United Nations (FAO) within the EC FAO Food Security for Action Programme [27]. The *Pv*API data were averaged over the period available and were used to classify areas

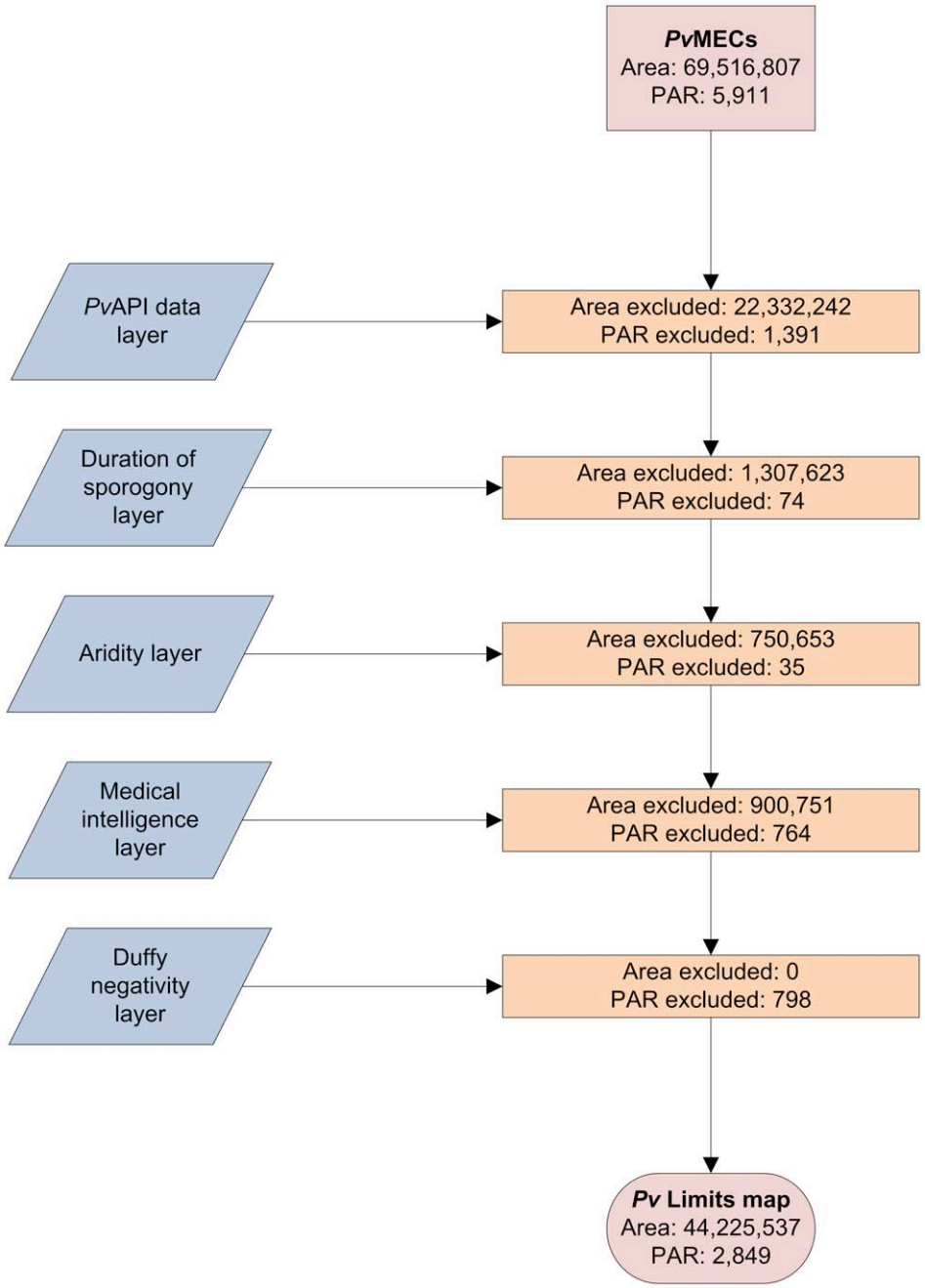


Figure 1. Flow chart of the various data and exclusion layers used to derive the final map. The pink rectangle denotes the surface area and populations of *Pv*MECs, whilst the pink ovoid represents the resulting trimmed surface area and PAR after the exclusion of risk by the various input layers, denoted by the blue rhomboids. Orange rectangles show area and PAR exclusions at each step to illustrate how these were reduced progressively. The sequence in which the exclusion layers are applied does not affect the final PAR estimates.
doi:10.1371/journal.pntd.0000774.g001

as malaria free, unstable (<0.1 case per 1,000 p.a.) or stable (≥0.1 case per 1,000 p.a.) transmission, based upon metrics advised during the Global Malaria Eradication Programme [28–30]. These data categories were then mapped using ArcMAP 9.2 (ESRI 2006).

Biological masks of exclusion of risk

To further constrain risk within national territories, two "masks" of biological exclusion were implemented (Protocol S2). First, risk was constrained according to the relationship between temperature and the duration of sporogony, based upon parameters specific to *P. vivax* [31]. Synoptic mean, maximum and minimum monthly temperature records were obtained from 30-arcsec (~1×1 km) spatial resolution climate surfaces [32]. For each pixel, these values were converted, using spline interpolation, to a continuous time series representing a mean temperature profile across an average year. Diurnal variation was represented by adding a sinusoidal component to the time series with a wavelength of 24 hours and the amplitude varying smoothly across the year determined by the difference between the monthly

minimum and maximum values. For *P. vivax* transmission to be biologically feasible, a cohort of anopheline vectors infected with *P. vivax* must survive long enough for sporogony to complete within their lifetime. Since the rate of parasite development within anophelines is strongly dependent on ambient temperature, the time required for sporogony varies continuously as temperatures fluctuate across a year [31]. For each pixel, the annual temperature profile was used to determine whether any periods existed in the year when vector lifespan would exceed the time required for sporogony, and hence when transmission was not precluded by temperature. This was achieved via numerical integration whereby, for cohorts of vectors born at each successive 2-hour interval across the year, sporogony rates varying continuously as a function of temperature were used to identify the earliest time at which sporogony could occur. If this time exceeded the maximum feasible vector lifespan, then the cohort was deemed unable to support transmission. If sporogony could not complete for any cohort across the year, then the pixel was classified as being at zero risk. Vector lifespan was defined as 31 days since estimates of the longevity of the main dominant vectors [33] indicate that 99% of anophelines die in less than a month and, therefore, would be unable to support parasite development in the required time. The exceptions were areas that support the longer-lived *Anopheles sergentii* and *An. superpictus*, where 62 days were considered more appropriate (Protocol S2) [18].

The second mask was based on the effect of arid conditions on anopheline development and survival [34]. Limited surface water reduces the availability of sites suitable for oviposition and reduces the survival of vectors at all stages of their development through the process of desiccation [35]. The ability of adult vectors to survive long enough to contribute to parasite transmission and of pre-adult stages to ensure minimum population abundance is, therefore, dependent on the levels of aridity and species-specific resilience to arid conditions. Extremely arid areas were identified using the global GlobCover Land Cover product (ESA/ESA GlobCover Project, led by MEDIAS-France/POSTEL) [36]. GlobCover products are derived from data provided by the Medium Resolution Imaging Spectrometer (MERIS), on board the European Space Agency's (ESA) ENVIronmental SATellite (ENVISAT), for the period between December 2004 and June 2006, and are available at a spatial resolution of 300 meters [36]. The layer was first resampled to a 1×1 km grid using a majority filter, and all pixels classified as "bare areas" by GlobCover were overlaid onto the *Pv*API surface. The aridity mask was treated differently from the temperature mask to allow for the possibility of the adaptation of human and vector populations to arid environments [37–39]. A more conservative approach was taken, which down-regulated risk by one class. In other words, GlobCover's bare areas defined originally as at stable risk by *Pv*API were stepped down to unstable risk and those classified initially as unstable to malaria free.

Medical intelligence modulation of risk

Medical intelligence contained in international travel and health guidelines [25,26] was used to inform risk exclusion and down-regulation in specific urban areas and sub-national territories, which are cited as being free of malaria transmission (Protocol S3). Additional medical intelligence and personal communication with malaria experts helped identify further sub-national areas classified as malaria free in Cambodia, Vanuatu and Yemen. Specified urban areas were geo-positioned and their urban extents were identified using the Global Rural Urban Mapping Project (GRUMP) urban extents layer [40]. Rules of risk modulation within these urban extents were as follows: i) risk within urban extents falling outside the range of the urban vector *An. stephensi* [41] (Protocol S3) was excluded; ii) risk within urban areas inhabited by *An. stephensi* was down-regulated by one level from stable to unstable and from unstable to free (Protocol S3). Specified sub-national territories were classified as malaria free if not already identified as such by the *Pv*API layer and the biological masks. These territories were mapped using the GAUL data set [27].

Duffy negativity phenotype

Since Duffy negativity provides protection against infection with *P. vivax* [42], a continuous map of the Duffy negativity phenotype was generated from a geostatistical model fully described elsewhere (Howes *et al.*, manuscript in preparation). The model was informed by a database of Duffy blood group surveys assembled from thorough searches of the published literature and supplemented with unpublished data by personal communication with relevant authors. Sources retrieved were added to existing Duffy blood group survey databases [43,44]. The earliest inclusion date for surveys was 1950, when the Duffy blood group was first described [45].

To model the Duffy system and derive a global prediction for the frequency of the homozygous Duffy negative phenotype ([Fy(a-b-)], which is encoded by the homozygous $FY{*}B^{ES}/{*}B^{ES}$ genotype), the spatially variable frequencies of the two polymorphic loci determining Duffy phenotypes were modelled: i) nucleotide −33 in the gene's promoter region, which defines positive/negative expression (T-33C); ii) the coding region locus (G125A) determining the antigen type expressed: Fy^a or Fy^b [46]. Due to the wide range of diagnostic methods used to describe Duffy blood types in recent decades, data were recorded in a variety of forms, each providing differing information about the frequency of variants at both loci. For example, some molecular studies sequenced only the gene's promoter region, and thus could not inform the frequency of the coding region variant; serological diagnoses only testing for the Fy^a antigen could not distinguish Fy^b from the Duffy negative phenotype. As part of the larger dataset, however, these incomplete data types can indirectly inform frequencies of negativity. Therefore, despite only requiring information about the promoter locus to model the negativity phenotype, variant frequencies at both polymorphic sites were modelled. This allowed the full range of information contained in the dataset to be used rather than just the subset specifically reporting Duffy negativity frequencies.

The model's general architecture and Bayesian framework will be described elsewhere (Howes *et al.*, manuscript in preparation). Briefly, the dataset of known values at fixed geographic locations was used to predict expression frequencies at each locus in all geographic sites where no data were available, thereby generating continuous global surfaces of the frequency of each variant. From the predicted frequency of the promoter region variant encoding null expression (-33C), a continuous frequency map of the Duffy negative population was derived.

Estimating the population at risk of *P. vivax* transmission

The GRUMP *beta* version provides gridded population counts and population density estimates for the years 1990, 1995, and 2000, both adjusted and unadjusted to the United Nations' national population estimates [40]. The adjusted population counts for the year 2000 were projected to 2009 by applying national, medium variant, urban and rural-specific growth rates by country [47]. These projections were undertaken using methods described previously [48], but refined with urban growth rates being applied solely to populations residing within the GRUMP

urban extents, while the rural growth rates were applied to the remaining population. This resulted in a 2009 population count surface of approximately 1×1 km spatial resolution, which was used to extract PAR figures. The PAR estimates in Africa were corrected for the presence of the Duffy negativity phenotype by multiplying the extracted population by [1 - frequency of Duffy negative individuals].

Results

Plasmodium vivax malaria endemic countries

A total of 109 potentially endemic countries and territories listed in international travel and health guidelines were identified [25,26]. Ten of these countries: Algeria, Armenia, Egypt, Jamaica (*P. falciparum* only), Mauritius, Morocco, Oman, Russian Federation, Syrian Arab Republic and Turkmenistan have either interrupted transmission or are extremely effective at dealing with minor local outbreaks. These nations were not classified as *Pv*MECs and are all considered to be in the elimination phase by the Global Malaria Action Plan [24]. Additionally, four malaria endemic territories report *P. falciparum* transmission only: Cape Verde [49], the Dominican Republic [50], Haiti [50,51] and Mayotte [52]. This resulted in a global total of 95 *Pv*MECs. Figure 1 summarises the various layers applied on the 95 *Pv*MECs in order to derive the limits of *P. vivax* transmission. The results of these different steps are described below.

Defining the spatial limits of *P. vivax* transmission at sub-national level

*Pv*API data were available for 51 countries. Data for four countries were available up to 2009. For 29 countries the last year of reporting was 2008, whilst 2007 and 2006 were the last years available for 11 and six countries, respectively. For Colombia the last reporting year was 2005. No HMIS data could be obtained for Kyrgyzstan and Uzbekistan, for which information contained in the most recent travel and health guidelines [25,26] was used to map risk. With the exception of Namibia, Saudi Arabia, South Africa and Swaziland, which were treated like all other nations, no HMIS data were solicited for countries in the Africa+ region, where stable risk of *P. vivax* transmission was assumed to be present throughout the country territories. In Botswana, stable risk was assumed in northern areas as specified by travel and health guidelines [25,26]. Amongst those countries for which HMIS data were available, 16 reported at ADMIN1 and 29 at ADMIN2 level. For Southern China, Myanmar, Nepal and Peru, data were available at ADMIN3 level. Data for Namibia and Venezuela were resolved at ADMIN1 and ADMIN2 levels. In total, 17,591 administrative units were populated with *Pv*API data. Protocol S1 describes these data in detail. Figure 2 shows the spatial extent of *P. vivax* transmission as defined by the *Pv*API data, with areas categorised as malaria free, unstable (*Pv*API<0.1 case per 1,000 p.a.) or stable (*Pv*API≥0.1 case per 1,000 p.a.) transmission [29].

Biological masks to refine the limits of transmission

Figure 3 shows the limits of *P. vivax* transmission after overlaying the temperature mask on the *Pv*API surface. The *P. vivax*-specific temperature mask was less exclusive of areas of risk than that derived for *P. falciparum* [18]. Exclusion of risk was mainly evident in the Andes, the southern fringes of the Himalayas, the eastern fringe of the Tibetan plateaux, the central mountain ridge of New Guinea and the East African, Malagasy and Afghan highlands. There was a remarkable correspondence between *Pv*API defined risk in the Andean and Himalayan regions and the temperature mask, which trimmed pixels of no risk at very high spatial resolution in these areas.

The aridity mask used here [36] was more contemporary and derived from higher spatial resolution imagery than the one used to define the limits of *P. falciparum* [18]. Figure 4 shows that the

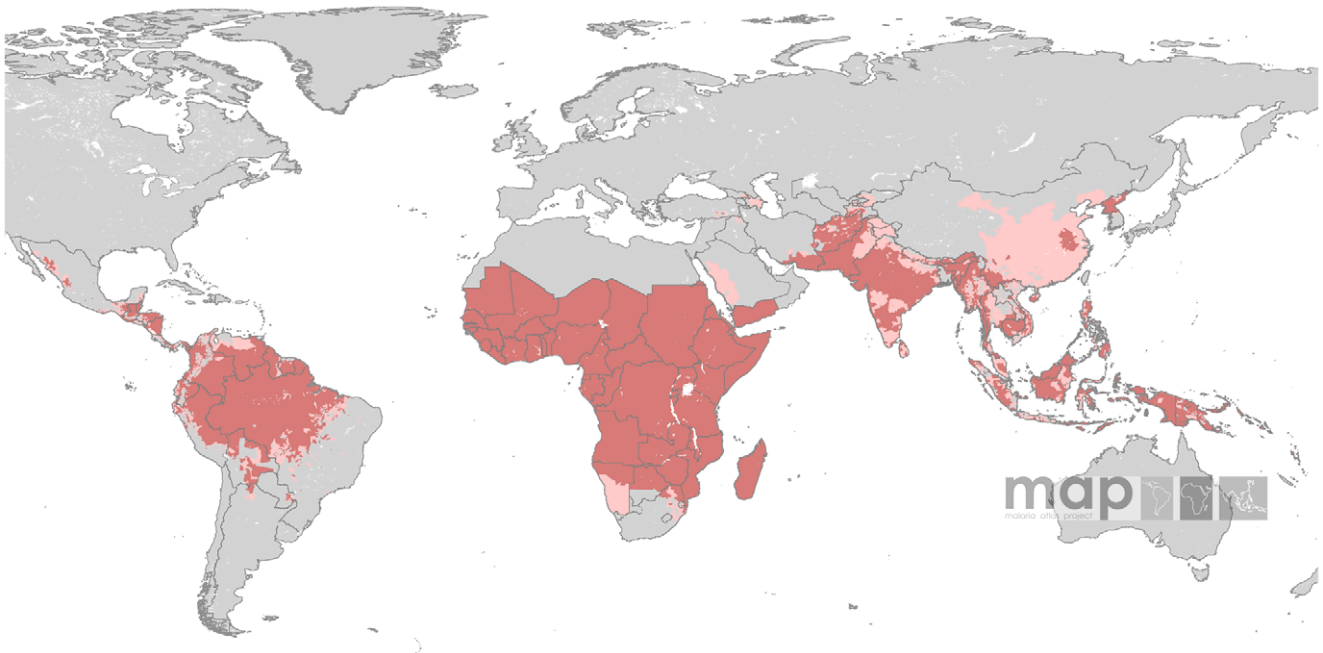


Figure 2. *Plasmodium vivax* malaria risk defined by *Pv*API data. Transmission was defined as stable (red areas, where *Pv*API≥0.1 per 1,000 people p.a.), unstable (pink areas, where *Pv*API<0.1 per 1,000 p.a.) or no risk (grey areas). The boundaries of the 95 countries defined as *P. vivax* endemic are shown.
doi:10.1371/journal.pntd.0000774.g002

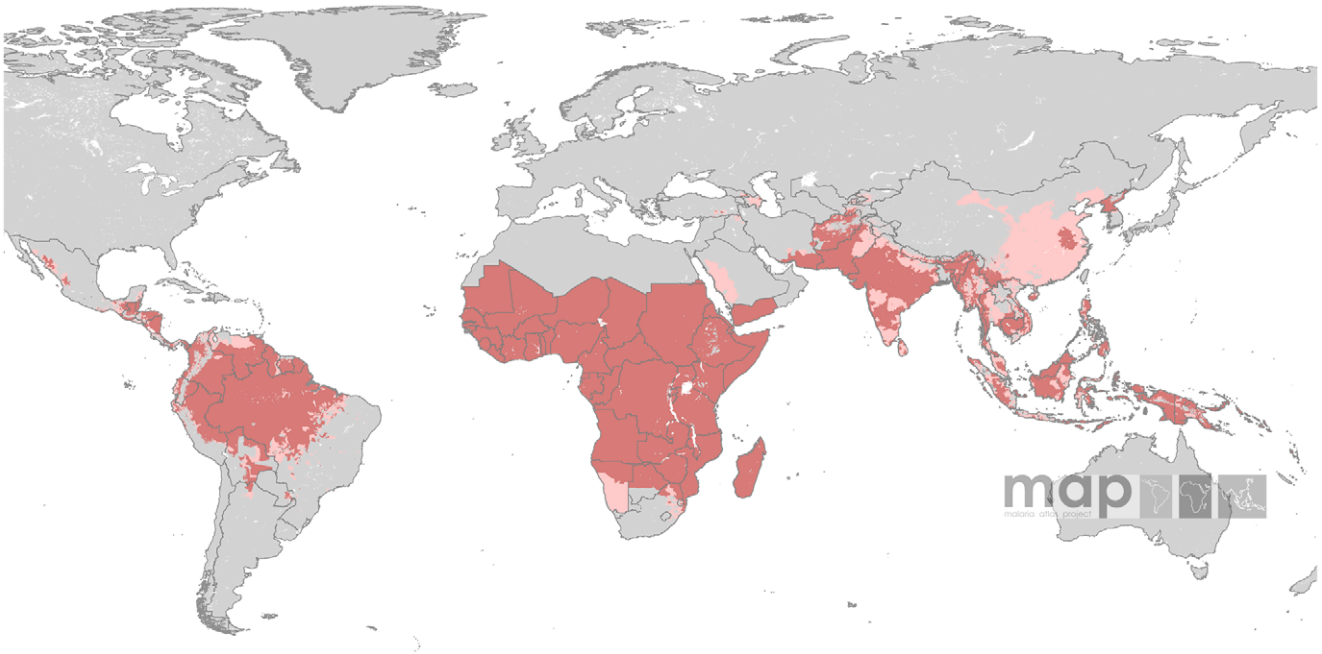


Figure 3. Further refinement of *Plasmodium vivax* transmission risk areas using the temperature layer of exclusion. Risk areas are defined as in Figure 2.
doi:10.1371/journal.pntd.0000774.g003

effects of the aridity mask were more evident in the Sahel and southern Saharan regions, as well as the Arabian Peninsula. In the western coast of Saudi Arabia, unstable risk defined by the *Pv*API layer was reduced to isolated foci of unstable risk by the aridity mask. In Yemen, stable risk was constrained to the west coast and to limited pockets along the southern coast. Similarly, endemic areas of stable risk defined by *Pv*API data in southern Afghanistan, southern Iran and throughout Pakistan were largely reduced to unstable risk by the aridity mask.

Medical intelligence used to refine risk

The two international travel and health guidelines consulted [25,26] cite 59 specific urban areas in 31 countries as being malaria free, in addition to urban areas in China, Indonesia (those

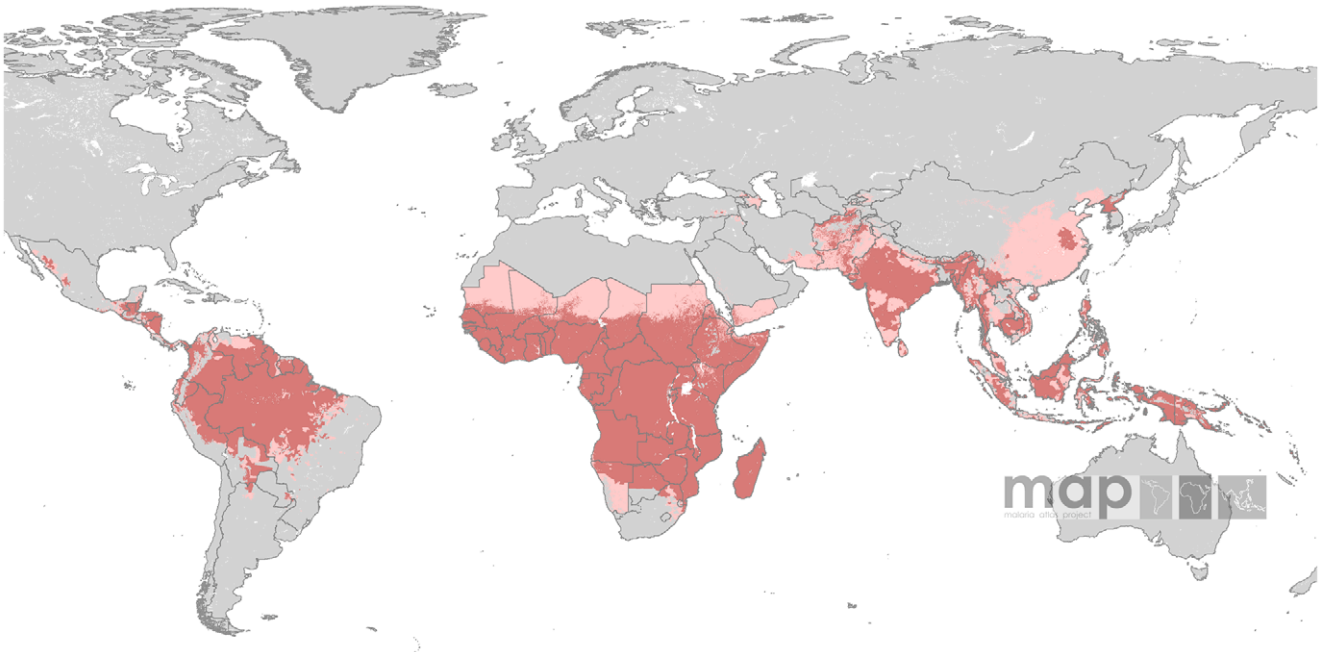


Figure 4. Aridity layer overlaid on the *Pv*API and temperature layers. Risk areas are defined as in Figure 2.
doi:10.1371/journal.pntd.0000774.g004

found in Sumatra, Kalimantan, Nusa Tenggara Barat and Sulawesi) and the Philippines (Protocol S3). A total of 42 of these cities fell within areas classified as malarious and amongst these, eight were found within the range of *An. stephensi*, as were some urban areas in south-western Yunnan, China. Risk in the latter was down-regulated from stable to unstable and from unstable to free due to the presence of this urban vector. In the remaining 34 cities and other urban areas in China, Indonesia and the Philippines, risk was excluded. In addition, 36 administrative units, including islands, are cited as being malaria free (Protocol S3). These territories were excluded as areas of risk, if not already classified as such by the *Pv*API surface and biological masks. In addition, the island of Aneityum, in Vanuatu [53], the area around Angkor Watt, in Cambodia, and the island of Socotra, in Yemen [54], were classified as malaria free following additional medical intelligence and personal communication with malaria experts from these countries.

Frequency of Duffy negativity

From the assembled library of references, 821 spatially unique Duffy blood type surveys were identified. Globally the data points were spatially representative, with 265 in America, 213 in Africa+ (167 sub-Saharan), 207 in CSE Asia and 136 in Europe. The total global sampled population was 131,187 individuals, with 24,816 (18.9%) in Africa+ and 33 African countries represented in the final database.

The modelled global map of Duffy negativity (Figure 5) indicates that the *P. vivax* resistant phenotype is rarely seen outside of Africa, and, when this is the case, it is mainly in localised New World migrant communities. Within Africa, the predicted prevalence was strikingly high south of the Sahara. Across this region, the silent Duffy allele was close to fixation in 31 countries with 95% or more of the population being Duffy negative. Frequencies fell sharply into southern Africa and into the Horn of Africa. For instance, the frequency of Duffy negativity in the South African population was 62.7%, increasing to 65.0% in Namibia and 73.5% across Madagascar. The situation was predicted to be highly heterogeneous across Ethiopia, with an estimated 50.0% of the overall population being Duffy negative.

Populations at risk of *P. vivax* transmission

The estimated *P. vivax* endemic areas and PAR for 2009 are presented in Table 1, stratified by unstable (*Pv*API$<$0.1 per 1,000 p.a.) and stable (*Pv*API$\geq$0.1 per 1,000 p.a.) risk of transmission, globally and by region and sub-region. It was estimated that there were 2.85 billion people at risk of *P. vivax* transmission worldwide in 2009, the vast majority (91.0%) inhabiting the CSE Asia region, 5.5% living in America and 3.4% living in Africa+, after accounting for Duffy negativity. An estimated 57.1% of the *P. vivax* PAR in 2009 lived in areas of unstable transmission, with a population of 1.63 billion.

Country level PAR estimates are provided in Protocol S4. The ten countries with the highest estimated PAR, in descending order, were: India, China, Indonesia, Pakistan, Viet Nam, Philippines, Brazil, Myanmar, Thailand and Ethiopia. PAR estimates in India accounted for 41.9% of the global PAR estimates, with 60.3% of the more than one billion PAR (1.19 billion) living in stable transmission areas. The situation in China was different as, according to the *Pv*API input data, areas of stable transmission were only found in the southern provinces of Yunnan and Hainan, and in the north-eastern province of Anhui, which reported an unusually high number of cases up to 2007. The latter is in accordance with a recent report documenting the resurgence of malaria in this province [55]. Transmission in the rest of China was largely negligible, with *Pv*API values well below 0.1 case per 1,000 people p.a. Given the reported cases, however, these were classified as unstable transmission areas and the total PAR estimated within them, after urban exclusions, was 583 million

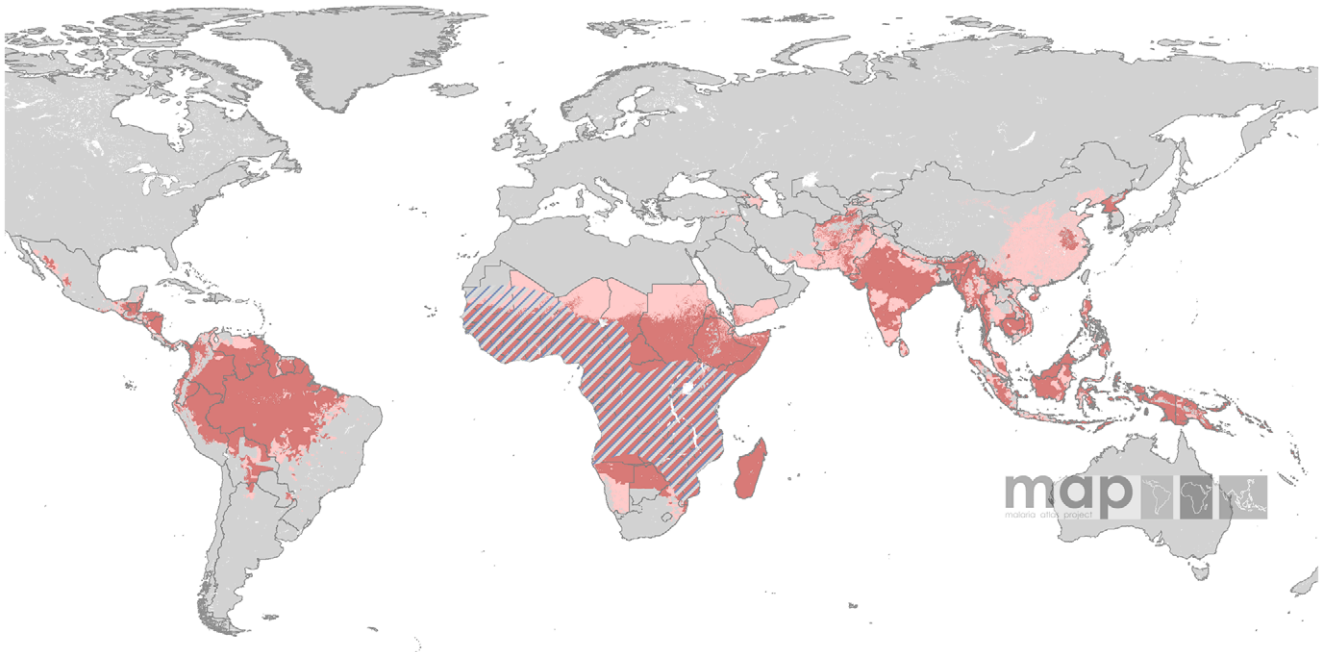


Figure 5. The global spatial limits of *Plasmodium vivax* malaria transmission in 2009. Risk areas are defined as in Figure 2. The medical intelligence and predicted Duffy negativity layers are overlaid on the *P. vivax* limits of transmission as defined by the *Pv*API data and biological mask layers. Areas where Duffy negativity prevalence was estimated as $\geq$90% are hatched, indicating where PAR estimates were modulated most significantly by the presence of this genetic trait.
doi:10.1371/journal.pntd.0000774.g005

Table 1. Regional and global areas and PAR of *Plasmodium vivax* malaria in 2009.

Region	Area (km^2)			PAR (millions)		
	Unstable	Stable	Any risk	Unstable	Stable	Any risk
Africa+	4,812,618	17,980,708	22,793,326	20.1	77.9	98.0
America	1,368,380	8,087,335	9,455,715	99.0	58.8	157.8
CSE Asia	5,848,939	6,127,549	11,976,488	1,509.0	1,084.2	2,593.2
West Asia	*2,007,247*	*2,800,612*	*4,807,859*	*653.9*	*845.2*	*1,499.2*
Central Asia	*3,156,574*	*1,277,219*	*4,433,793*	*694.3*	*129.2*	*823.4*
East Asia	*685,118*	*2,049,717*	*2,734,835*	*160.8*	*109.8*	*270.6*
World	12,029,937	32,195,600	44,225,537	1,628.1	1,220.9	2,849.0

doi:10.1371/journal.pntd.0000774.t001

people. All other countries reporting the highest PAR were in CSE Asia, with the exception of Brazil and Ethiopia.

Discussion

We present a contemporary evidence-based map of the global distribution of *P. vivax* transmission developed from a combination of mapped sub-national HMIS data, biological rules of transmission exclusion and medical intelligence. The methods used were developed from those implemented for *P. falciparum* malaria [18] and can be reproduced following the sequence of data layer assemblies and exclusions illustrated in Figure 1.

Plasmodium vivax is transmitted within 95 countries in tropical, sub-tropical and temperate regions, reaching approximately 43 degrees north in China and approximately 30 degrees south in Southern Africa. The fact that *P. vivax* has a wider range than *P. falciparum* [18] is facilitated by two aspects of the parasite's biology [56]: i) its development at lower temperatures during sporogony [31]; and ii) its ability to produce hypnozoites during its life cycle in the human host [57]. The sporogonic cycle of *P. vivax* is shorter (i.e. a lower number of degree days required for its completion) and the parasite's sexual stage is active at lower temperatures than other human malaria parasites (Protocol S2) [31]. Consequently, generation of sporozoites is possible at higher altitudes and more extreme latitudes. In the human host, hypnozoites of *P. vivax* temperate strains can relapse anywhere between months and years after the initial infection, often temporally coincident with optimal climatic conditions in a new transmission season [10,57].

The resulting maps produced an estimate of 2.85 billion people living at risk of *P. vivax* malaria transmission in 2009. The distribution of *P. vivax* PAR is very different from that of *P. falciparum* [18], due to the widespread distribution of *P. vivax* in Asia, up to northern China, and the high prevalence of the Duffy negativity phenotype in Africa. China accounts for 22.0% of the global estimated *P. vivax* PAR, although 93.1% of these people live in areas defined as unstable transmission (Protocol S4). An important caveat is that *Pv*API data from central and northern China could only be accessed at the lowest administrative level (ADMIN1) (Protocol S1). The very high population densities found in this country exacerbate the problem, inevitably biasing PAR estimates, despite urban areas in China being excluded from the calculations following information from the sources of medical intelligence that were consulted [25,26]. Malaria transmission in most of these unstable transmission areas in China is probably negligible given the very few cases reported between 2003 and 2007. It is important to stress the necessity to access *Pv*API data at a higher spatial resolution from China (i.e. at the county level) in order to refine these estimates and minimise biases.

Table 2. Published evidence of *Plasmodium vivax* malaria transmission in African countries.

Country	References*
Angola	[68–73]
Benin	[68,70,71,74]
Botswana	[72]
Burkina Faso	[68,71]
Burundi	[70–73]
Cameroon	[68,69,71–79]
Cen. African Rep.	[68]
Chad	[74]
Comoros	[68]
Congo	[68,70,71,73,74,76,77,80]
Côte d'Ivoire	[68–71,73,74,76,78]
Congo (DR)	[68,81]
Djibouti	[68,78]
Equatorial Guinea	[82]
Eritrea	[71,73,76,77,83,84]
Ethiopia	[68–74,76–79,85]
Gabon	[68,71,86]
Gambia	[71,72,76,78]
Ghana	[69–74,76–79]
Guinea	[68,69,71,76,77]
Kenya	[68–73,76–79]
Liberia	[68–73,76–79]
Madagascar	[68–73,76,78,87]
Malawi	[68,70,72,73]
Mali	[68,69,71]
Mauritania	[68,69,71,72,76,77,88,89]
Mozambique	[68–71,73,76,79,90]
Namibia	[70]
Niger	[68,69,71,76]
Nigeria	[69–74,76–79,91]
Rwanda	[68,71,72,78]
São Tomé and Príncipe	[68,92]
Senegal	[68,70,71,73,76,77]
Sierra Leone	[68,69,72–74,76,78]
Somalia	[69,70,78,79,93]
South Africa	[69–71,76–78]
Sudan	[68–74,76,77,79,94]
Togo	[70,71]
Uganda	[69–74,76–79,95]
Tanzania	[68–72,76,77,79]
Zambia	[69–72,78,96]
Zimbabwe	[68,69,71]

*The cited references mostly document imported cases from Africa. Evidence of transmission of *P. vivax* in Guinea Bissau and Swaziland could not be found in the published literature.

doi:10.1371/journal.pntd.0000774.t002

In Africa, the modelled prevalence of Duffy negativity shows that very high rates of this phenotype are present in large swaths of West and Central Africa (Figure 5). One of the functions of the Duffy antigen is being a receptor of *P. vivax* [46] and its absence has been shown to preclude infection with this parasite [58,59], although the extent of this has been questioned [60–63]. There is no doubt that the African continent has a climate highly conducive to *P. vivax* transmission (Protocol S2). Moreover, dominant African *Anopheles* have been shown to be competent vectors of this parasite [62,64,65]. In addition, there is a plethora of evidence of *P. vivax* transmission in Africa, mostly arising from travel-acquired *P. vivax* infections during visits to malaria endemic African countries (Table 2; Protocol S1). This evidence supports the hypothesis that *P. vivax* may have been often misdiagnosed as *P. ovale* in the region due to a combination of morphological similarity and the prevailing bio-geographical dogma driven by the high prevalence of Duffy negativity [60]. Despite the fact that the risk of *P. vivax* is cosmopolitan, PAR estimates in Africa were modulated according to the high limitations placed on infection by the occurrence of the Duffy negative trait. Consequently, the PAR in the Africa+ region accounts for only 3.5% of the global estimated *P. vivax* PAR. Although recent work has shown 42 *P. vivax* infections amongst 476 individuals genotyped as Duffy negative across eight sites in Madagascar [63], we have taken a conservative approach and consider it premature to relax the Duffy exclusion of PAR across continental Africa until this study has been replicated elsewhere.

Mapping the distribution of *P. vivax* malaria has presented a number of unique challenges compared to *P. falciparum*, some of which have been addressed by the methods used here. The influence of climate on parasite development has been allowed for by implementing a temperature mask parameterised specifically for the *P. vivax* life cycle. The question of Duffy negativity and *P. vivax* transmission has also been addressed by modelling the distribution of this phenotype and by allowing the predicted prevalence to modulate PAR. It is also worth noting that the accuracy of HMIS for *P. vivax* clinical cases, particularly in areas of coincidental *P. falciparum* risk, is notoriously poor [66], in part because microscopists are less likely to record the presence of a parasite assumed to be clinically less important. Here, HMIS data were averaged over a period of up to four years and used to differentiate malaria free areas from those that are malarious. Within the latter, a conservative threshold was applied to classify risk areas as being of unstable (*Pv*API$<$0.1 per 1,000 p.a.) or stable (*Pv*API$\geq$0.1 per 1,000 p.a.) transmission [29]. We believe that this conservative use of HMIS data balances, to some extent, anomalies introduced by *P. vivax* underreporting and the correspondence of the biological masks and *Pv*API data in many areas is reassuring.

The intensity of transmission within the defined stable limits of *P. vivax* risk will vary across this range and this will be modelled using geostatistical techniques similar to those developed recently for *P. falciparum* [19]. This modelling work will be cognisant of the unique epidemiology of *P. vivax*. First, in areas where *P. vivax* infection is coincidental with *P. falciparum*, prevalence of the former may be suppressed by cross-species immunity [67] or underestimated by poor diagnostics [66]. Second, there is the ability of *P. vivax* to generate hypnozoites that lead to relapses. These characteristics render the interpretation of prevalence measures more problematic [5]. Third, the prevalence of Duffy negativity provides protection against infection in large sections of the population in Africa [58,59]. An appropriate modelling framework is under development and will be the subject of a subsequent paper mapping *P. vivax* malaria endemicity using parasite prevalence data. These data are being collated in the MAP database, with nearly 9,000 *P. vivax* parasite rate records archived by 01 March 2010.

Supporting Information

Protocol S1 Defining risk of transmission of *Plasmodium vivax* using case reporting data. Document describing more extensively one of the layers used to create the final map.
Found at: doi:10.1371/journal.pntd.0000774.s001 (2.87 MB DOC)

Protocol S2 Defining the global biological limits of *Plasmodium vivax* transmission. Document describing more extensively two of the layers used to create the final map.
Found at: doi:10.1371/journal.pntd.0000774.s002 (0.42 MB DOC)

Protocol S3 Risk modulation based upon medical intelligence. Document describing more extensively one of the layers used to create the final map.
Found at: doi:10.1371/journal.pntd.0000774.s003 (0.36 MB DOC)

Protocol S4 Country level area and population at risk of *Plasmodium vivax* malaria in 2009. Country-level table of the estimated area and populations at risk of *P. vivax* malaria in 2009
Found at: doi:10.1371/journal.pntd.0000774.s004 (0.16 MB DOC)

Acknowledgments

We thank Anja Bibby for proof reading the manuscript. A large proportion of the *Pv*API data used to map risk in this paper could only be accessed with the help of people in the malaria research and control communities of each country and these individuals are listed on the MAP website (www.map.ac.uk/acknowledgements.html) and in Table 2 of Protocol S1. The authors also acknowledge the support of the Kenyan Medical Research Institute (KEMRI). This paper is published with the permission of the director of KEMRI.

Author Contributions

Conceived and designed the experiments: SIH. Performed the experiments: CAG REH APP PWG TPVB WHT. Analyzed the data: CAG REH APP PWG TPVB WHT CWK AJT BHM IRFE JKB RWS. Wrote the paper: CAG SIH.

References

1. Snow RW, Craig MH, Newton CRJC, Steketee RW (2003) The public health burden of *Plasmodium falciparum* malaria in Africa: deriving the numbers. Working Paper No. 11. Bethesda, Maryland, U.S.A.: Disease Control Priorities Project, Fogarty International Center, National Institutes of Health.
2. Hay SI, Guerra CA, Tatem A, Atkinson P, Snow RW (2005) Urbanization, malaria transmission and disease burden in Africa. Nat Rev Microbiol 3: 81–90.
3. Mendis K, Sina BJ, Marchesini P, Carter R (2001) The neglected burden of *Plasmodium vivax* malaria. Am J Trop Med Hyg 64: 97–106.
4. Baird JK (2007) Neglect of *Plasmodium vivax* malaria. Trends Parasitol 23: 533–539.
5. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, et al. (2007) Vivax malaria: neglected and not benign. Am J Trop Med Hyg 77: 79–87.
6. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, et al. (2009) Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. Lancet Infect Dis 9: 555–566.
7. Guerra CA, Snow RW, Hay SI (2006) Defining the global spatial limits of malaria transmission in 2005. Adv Parasitol 62: 157–179.
8. Guerra CA, Snow RW, Hay SI (2006) Mapping the global extent of malaria in 2005. Trends Parasitol 22: 353–358.

9. Sattabongkot J, Tsuboi T, Zollner GE, Sirichaisinthop J, Cui L (2004) *Plasmodium vivax* transmission: chances for control? Trends Parasitol 20: 192–198.
10. Baird JK (2009) Resistance to therapies for infection by *Plasmodium vivax*. Clin Microbiol Rev 22: 508–534.
11. Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, et al. (2008) *Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea. PLoS Med 5: e127.
12. Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, et al. (2008) Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. PLoS Med 5: e128.
13. Anstey NM, Russell B, Yeo TW, Price RN (2009) The pathophysiology of vivax malaria. Trends Parasitol 25: 220–227.
14. Kochar DK, Das A, Kochar SK, Saxena V, Sirohi P, et al. (2009) Severe *Plasmodium vivax* malaria: a report on serial cases from Bikaner in northwestern India. Am J Trop Med Hyg 80: 194–198.
15. Barcus MJ, Basri H, Picarima H, Manyakori C, Sekartuti, et al. (2007) Demographic risk factors for severe and fatal vivax and falciparum malaria among hospital admissions in northeastern Indonesian Papua. Am J Trop Med Hyg 77: 984–991.
16. Parakh A, Agarwal N, Aggarwal A, Aneja A (2009) *Plasmodium vivax* malaria in children: uncommon manifestations. Ann Trop Paediatr 29: 253–256.
17. Hay SI, Snow RW (2006) The Malaria Atlas Project: Developing global maps of malaria risk. PLoS Med 3: e473.
18. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, et al. (2008) The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. PLoS Med 5: e38.
19. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6: e48.
20. Snow RW, Guerra CA, Mutheu JJ, Hay SI (2008) International funding for malaria control in relation to populations at risk of stable *Plasmodium falciparum* transmission. PLoS Med 5: e142.
21. Gething PW, Patil AP, Hay SI (2010) Quantifying aggregated uncertainty in *Plasmodium falciparum* malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. PLoS Comput Biol 6: e1000724.
22. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, et al. (2010) Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. PLoS Med: in press.
23. Feachem R, Sabot O (2008) A new global malaria eradication strategy. Lancet 371: 1633–1635.
24. Roll Back Malaria Partnership (2008) The Global Malaria Action Plan: For a malaria-free world. .
25. Centers for Disease Control and Prevention (2009) CDC Health Information for International Travel 2010. Atlanta: U.S. Department of Health and Human Services, Public Health Service.
26. WHO (2010) International Travel and Health: Situation as on 1 January 2010. Geneva: World Health Organization.
27. FAO (2008) The Global Administrative Unit Layers (GAUL): Technical Aspects. Rome: Food and Agriculture Organization of the United Nations, EC-FAO Food Security Programme (ESTG).
28. Pampana E (1969) A textbook of malaria eradication. London: Oxford University Press.
29. Hay SI, Smith DL, Snow RW (2008) Measuring malaria endemicity from intense to interrupted transmission. Lancet Infect Dis 8: 369–378.
30. Yekutiel P (1980) III The Global Malaria Eradication Campaign. In: Klingberg MA, ed. Eradication of infectious diseases: a critical study. Basel, Switzerland: Karger. pp 34–88.
31. Nikolaev BP (1935) On the influence of temperature on the development of malaria *plasmodia* inside the mosquito. Leningrad Pasteur Institute of Epidemiology and Bacteriology.
32. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25: 1965–1978.
33. Kiszewski A, Mellinger A, Spielman A, Malaney P, Sachs SE, et al. (2004) A global index representing the stability of malaria transmission. Am J Trop Med Hyg 70: 486–498.
34. Shililu JI, Grueber WB, Mbogo CM, Githure JI, Riddiford LM, et al. (2004) Development and survival of *Anopheles gambiae* eggs in drying soil: influence of the rate of drying, egg age, and soil type. J Am Mosq Control Assoc 20: 243–247.
35. Gray EM, Bradley TJ (2005) Physiology of desiccation resistance in *Anopheles gambiae* and *Anopheles arabiensis*. Am J Trop Med Hyg 73: 553–559.
36. Bicheron P, Defourny P, Brockmann C, Schouten L, Vancutsem C, et al. (2008) GLOBCOVER: Products Description and Validation Report. Tolouse: MEDIAS-France.
37. Omer SM, Cloudsley-Thompson JL (1970) Survival of female *Anopheles gambiae* Giles through a 9-month dry season in Sudan. Bull World Health Organ 42: 319–330.
38. Omer SM, Cloudsley-Thomson JL (1968) Dry season biology of *Anopheles gambiae* Giles in the Sudan. Nature 217: 879–880.
39. Bouma MJ, Parvez SD, Nesbit R, Winkler AM (1996) Malaria control using permethrin applied to tents of nomadic Afghan refugees in northern Pakistan. Bull World Health Organ 74: 413–421.
40. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, et al. (2006) Determining global population distribution: methods, applications and data. Adv Parasitol 62: 119–156.
41. Hay SI, Sinka ME, Okara RM, Kabaria CW, Mbithi PM, et al. (2010) Developing maps of the dominant *Anopheles* vectors of human malaria. PLoS Med 7: e1000209.
42. Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. N Engl J Med 295: 302–304.
43. Mourant AE, Kopec AC, Domaniewska-Sobczak K (1976) The Distribution of the Human Blood Groups and other Polymorphisms. London: Oxford University Press.
44. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. Princeton, New Jersey: Princeton University Press.
45. Cutbush M, Mollison PL (1950) The Duffy blood group system. Heredity 4: 383–389.
46. Langhi DM, Jr., Bordin JO (2006) Duffy blood group and malaria. Hematology 11: 389–398.
47. UNPD (2008) World Urbanization Prospects: The 2007 Revision Population Database. http://esa.un.org/unpp/. December 2009.
48. Hay SI, Noor AM, Nelson A, Tatem AJ (2005) The accuracy of human population maps for public health application. Trop Med Int Health 10: 1073–1086.
49. Ministério da Saúde de Cabo Verde, Direcção Geral de Saúde, Programa Nacional de Luta Contra o Paludismo (2009) Plano Estratégico de Pré-Eliminação do Paludismo 2009–2013.
50. PAHO (2006) Regional Strategic Plan for Malaria in the Americas 2006–2010. Washington DC: Pan American Health Organization.
51. Lindo JF, Bryce JH, Ducasse MB, Howitt C, Barrett DM, et al. (2007) *Plasmodium malariae* in Haitian refugees, Jamaica. Emerg Infect Dis 13: 931–933.
52. Tchen J, Ouledi A, Lepere JF, Ferrandiz D, Yvin JL (2006) Epidemiologie et prevention du paludisme dans les iles du sud-ouest de l'Ocean Indien. Med Trop (Mars) 66: 295–301.
53. Kaneko A, Taleo G, Kalkoa M, Yamar S, Kobayakawa T, et al. (2000) Malaria eradication on islands. Lancet 356: 1560–1564.
54. EMRO (2008) Technical discussion on malaria elimination in the Eastern Mediterranean Region: vision, requirements and strategic outline. World Health Organization Regional Office for the Eastern Mediterranean.
55. Zhang W, Wang L, Fang L, Ma J, Xu Y, et al. (2008) Spatial analysis of malaria in Anhui province, China. Malar J 7: 206.
56. Coatney GR, Collins WE, Warren M, Contacos PG (2003) The primate malarias.[CD-ROM; original book published 1971]. AtlantaGeorgia, USA: Centers for Disease Control and Prevention.
57. Garnham PCC (1988) Malaria parasites of man: life-cycles and morphology (excluding ultrastructure). In: Wernsdorfer WH, McGregor I, eds. Malaria: principles and practice of malariology. Edinburgh: Churchill Livingstone. pp 61–96.
58. Welch SG, McGregor IA, Williams K (1977) The Duffy blood group and malaria prevalence in Gambian West Africans. Trans R Soc Trop Med Hyg 71: 295–296.
59. Mathews HM, Armstrong JC (1981) Duffy blood types and vivax malaria in Ethiopia. Am J Trop Med Hyg 30: 299–303.
60. Rosenberg R (2007) *Plasmodium vivax* in Africa: hidden in plain sight? Trends Parasitol 23: 193–196.
61. Cavasini CE, Mattos LC, Couto AA, Bonini-Domingos CR, Valencia SH, et al. (2007) *Plasmodium vivax* infection among Duffy antigen-negative individuals from the Brazilian Amazon region: an exception? Trans R Soc Trop Med Hyg 101: 1042–1044.
62. Ryan JR, Stoute JA, Amon J, Dunton RF, Mtalib R, et al. (2006) Evidence for transmission of *Plasmodium vivax* among a duffy antigen negative population in Western Kenya. Am J Trop Med Hyg 75: 575–581.
63. Menard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR, et al. (2010) *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. Proc Natl Acad Sci U S A 107: 5967–5971.
64. Collins WE, Roberts JM (1991) *Anopheles gambiae* as a host for geographic isolates of *Plasmodium vivax*. J Am Mosq Control Assoc 7: 569–573.
65. Taye A, Hadis M, Adugna N, Tilahun D, Wirtz RA (2006) Biting behavior and *Plasmodium* infection rates of *Anopheles arabiensis* from Sille, Ethiopia. Acta Trop 97: 50–54.
66. Mayxay M, Pukrittayakamee S, Newton PN, White NJ (2004) Mixed-species malaria infections in humans. Trends Parasitol 20: 233–240.
67. Maitland K, Williams TN, Newbold CI (1997) *Plasmodium vivax* and *P. falciparum*: Biological interactions and the possibility of cross-species immunity. Parasitol Today 13: 227–231.
68. Gautret P, Legros F, Koulmann P, Rodier MH, Jacquemin JL (2001) Imported *Plasmodium vivax* malaria in France: geographical origin and report of an atypical case acquired in Central or Western Africa. Acta Trop 78: 177–181.
69. Holtz TH, Kachur SP, MacArthur JR, Roberts JM, Barber AM, et al. (2001) Malaria surveillance–United States, 1998. MMWR CDC Surveill Summ 50: 1–20.
70. Newman RD, Barber AM, Roberts J, Holtz T, Steketee RW, et al. (2002) Malaria surveillance–United States, 1999. MMWR Surveill Summ 51: 15–28.
71. Causer LM, Newman RD, Barber AM, Roberts JM, Stennies G, et al. (2002) Malaria surveillance–United States, 2000. MMWR Surveill Summ 52: 9–21.
72. Filler S, Causer LM, Newman RD, Barber AM, Roberts JM, et al. (2003) Malaria surveillance–United States, 2001. MMWR Surveill Summ 52: 1–14.

73. Thwing J, Skarbinski J, Newman RD, Barber AM, Mali S, et al. (2007) Malaria surveillance–United States, 2005. MMWR Surveill Summ 56: 23–40.
74. Mali S, Steele S, Slutsker L, Arguin PM (2008) Malaria surveillance–United States, 2006. MMWR Surveill Summ 57: 24–39.
75. Durante Mangoni E, Severini C, Menegon M, Romi R, Ruggiero G, et al. (2003) Case report: An unusual late relapse of *Plasmodium vivax* malaria. Am J Trop Med Hyg 68: 159–160.
76. Shah S, Filler S, Causer LM, Rowe AK, Bloland PB, et al. (2004) Malaria surveillance–United States, 2002. MMWR Surveill Summ 53: 21–34.
77. Eliades MJ, Shah S, Nguyen-Dinh P, Newman RD, Barber AM, et al. (2005) Malaria surveillance–United States, 2003. MMWR Surveill Summ 54: 25–40.
78. Skarbinski J, James EM, Causer LM, Barber AM, Mali S, et al. (2006) Malaria surveillance–United States, 2004. MMWR Surveill Summ 55: 23–37.
79. Mali S, Steele S, Slutsker L, Arguin PM (2009) Malaria surveillance–United States, 2007. MMWR Surveill Summ 58: 1–16.
80. Culleton R, Ndounga M, Zeyrek FY, Coban C, Casimiro PN, et al. (2009) Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. J Infect Dis 200: 1465–1469.
81. Comellini L, Tozzola A, Baldi F, Brusa S, Serra L, et al. (1998) *Plasmodium vivax* congenital malaria in a newborn of a Zairian immigrant. Ann Trop Paediatr 18: 41–43.
82. Rubio JM, Benito A, Roche J, Berzosa PJ, Garcia ML, et al. (1999) Semi-nested, multiplex polymerase chain reaction for detection of human malaria parasites and evidence of *Plasmodium vivax* infection in Equatorial Guinea. Am J Trop Med Hyg 60: 183–187.
83. Peruzzi S, Gorrini C, Piccolo G, Calderaro A, Dettori G, et al. (2007) Prevalence of imported malaria in Parma during 2005–2006. Acta Biomed 78: 170–175.
84. Sintasath DM, Ghebremeskel T, Lynch M, Kleinau E, Bretas G, et al. (2005) Malaria prevalence and associated risk factors in Eritrea. Am J Trop Med Hyg 72: 682–687.
85. Teka H, Petros B, Yamuah L, Tesfaye G, Elhassan I, et al. (2008) Chloroquine-resistant *Plasmodium vivax* malaria in Debre Zeit, Ethiopia. Malar J 7: 220.
86. Poirriez J, Landau I, Verhaeghe A, Savage A, Dei-Cas E (1991) [Atypical forms of *Plasmodium vivax*. Apropos of a case]. Ann Parasitol Hum Comp 66: 149–154.
87. Rabarijaona LP, Randrianarivelojosia M, Raharimalala LA, Ratsimbasoa A, Randriamanantena A, et al. (2009) Longitudinal survey of malaria morbidity over 10 years in Saharevo (Madagascar): further lessons for strengthening malaria control. Malar J 8: 190.
88. Cortes H, Morillas-Marquez F, Valero A (2003) Malaria in Mauritania: the first cases of malaria endemic to Nouakchott. Trop Med Int Health 8: 297–300.
89. Lekweiry KM, Abdallahi MO, Ba H, Arnathau C, Durand P, et al. (2009) Preliminary study of malaria incidence in Nouakchott, Mauritania. Malar J 8: 92.
90. Wejda BU, Huchzermeyer H, Dormann AJ (2002) Hotel malaria in Greece: Mozambique origin, American vector, German victims. J Travel Med 9: 277.
91. Erhabor O, Babatunde S, Uko KE (2006) Some haematological parameters in plasmodial parasitized HIV-infected Nigerians. Niger J Med 15: 52–55.
92. Snounou G, Pinheiro L, Antunes AM, Ferreira C, do Rosario VE (1998) Non-immune patients in the Democratic Republic of Sao Tome e Principe reveal a high level of transmission of *P. ovale* and *P. vivax* despite low frequency in immune patients. Acta Trop 70: 197–203.
93. Peragallo MS, Sabatinelli G, Majori G, Cali G, Sarnicola G (1997) Prevention and morbidity of malaria in non-immune subjects; a case-control study among Italian troops in Somalia and Mozambique, 1992–1994. Trans R Soc Trop Med Hyg 91: 343–346.
94. Himeidan YE, Elbashir MI, El-Rayah el A, Adam I (2005) Epidemiology of malaria in New Halfa, an irrigated area in eastern Sudan. East Mediterr Health J 11: 499–504.
95. Illamperuma C, Allen BL (2007) Pulmonary edema due to *Plasmodium vivax* malaria in an American missionary. Infection 35: 374–376.
96. Blossom DB, King CH, Armitage KB (2005) Occult *Plasmodium vivax* infection diagnosed by a polymerase chain reaction-based detection system: a case report. Am J Trop Med Hyg 73: 188–190.

OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

A Long Neglected World Malaria Map: *Plasmodium vivax* Endemicity in 2010

Peter W. Gething[1]*, Iqbal R. F. Elyazar[2], Catherine L. Moyes[1], David L. Smith[3,4], Katherine E. Battle[1], Carlos A. Guerra[1], Anand P. Patil[1], Andrew J. Tatem[4,5], Rosalind E. Howes[1], Monica F. Myers[1], Dylan B. George[4], Peter Horby[6,7], Heiman F. L. Wertheim[6,7], Ric N. Price[7,8,9], Ivo Müeller[10], J. Kevin Baird[2,7], Simon I. Hay[1,4]*

1 Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, United Kingdom, **2** Eijkman-Oxford Clinical Research Unit, Jakarta, Indonesia, **3** Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **4** Fogarty International Center, National Institutes of Health, Bethesda, Maryland, United States of America, **5** Department of Geography and Emerging Pathogens Institute, University of Florida, Gainesville, Florida, United States of America, **6** Oxford University Clinical Research Unit - Wellcome Trust Major Overseas Programme, Ho Chi Minh City, Vietnam, **7** Nuffield Department of Medicine, Centre for Tropical Medicine, University of Oxford, Oxford, United Kingdom, **8** Global Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia, **9** Division of Medicine, Royal Darwin Hospital, Darwin, Northern Territory, Australia, **10** Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea

Abstract

Background: Current understanding of the spatial epidemiology and geographical distribution of *Plasmodium vivax* is far less developed than that for *P. falciparum*, representing a barrier to rational strategies for control and elimination. Here we present the first systematic effort to map the global endemicity of this hitherto neglected parasite.

Methodology and Findings: We first updated to the year 2010 our earlier estimate of the geographical limits of *P. vivax* transmission. Within areas of stable transmission, an assembly of 9,970 geopositioned *P. vivax* parasite rate (*Pv*PR) surveys collected from 1985 to 2010 were used with a spatiotemporal Bayesian model-based geostatistical approach to estimate endemicity age-standardised to the 1–99 year age range ($PvPR_{1-99}$) within every 5×5 km resolution grid square. The model incorporated data on Duffy negative phenotype frequency to suppress endemicity predictions, particularly in Africa. Endemicity was predicted within a relatively narrow range throughout the endemic world, with the point estimate rarely exceeding 7% $PvPR_{1-99}$. The Americas contributed 22% of the global area at risk of *P. vivax* transmission, but high endemic areas were generally sparsely populated and the region contributed only 6% of the 2.5 billion people at risk (PAR) globally. In Africa, Duffy negativity meant stable transmission was constrained to Madagascar and parts of the Horn, contributing 3.5% of global PAR. Central Asia was home to 82% of global PAR with important high endemic areas coinciding with dense populations particularly in India and Myanmar. South East Asia contained areas of the highest endemicity in Indonesia and Papua New Guinea and contributed 9% of global PAR.

Conclusions and Significance: This detailed depiction of spatially varying endemicity is intended to contribute to a much-needed paradigm shift towards geographically stratified and evidence-based planning for *P. vivax* control and elimination.

Citation: Gething PW, Elyazar IRF, Moyes CL, Smith DL, Battle KE, et al. (2012) A Long Neglected World Malaria Map: *Plasmodium vivax* Endemicity in 2010. PLoS Negl Trop Dis 6(9): e1814. doi:10.1371/journal.pntd.0001814

Editor: Jane M. Carlton, New York University, United States of America

Received April 24, 2012; **Accepted** July 29, 2012; **Published** September 6, 2012

Funding: SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066), which also supports PWG, CAG, and KEB. CLM and APP are funded by a Biomedical Resources Grant from the Wellcome Trust (#091835). REH is funded by a Biomedical Resources Grant from the Wellcome Trust (#085406). IRFE is funded by grants from the University of Oxford—Li Ka Shing Foundation Global Health Program and the Oxford Tropical Network. DLS and AJT are supported by grants from the Bill and Melinda Gates Foundation (#49446, #1032350) (http://www.gatesfoundation.org). PH is supported by Wellcome Trust grants 089276/Z/09/Z and the Li Ka Shing Foundation. RNP is a Wellcome Trust Senior Fellow in Clinical Science (#091625). JKB is supported by a Wellcome Trust grant (#B9RJIXO). PWG, APP, DLS, AJT, DBG, and SIH also acknowledge support from the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health (http://www.fic.nih.gov). This work forms part of the output of the Malaria Atlas Project (MAP, http://www.map.ox.ac.uk), principally funded by the Wellcome Trust, UK (http://www.wellcome.ac.uk). MAP also acknowledges the support of the Global Fund to Fight AIDS, Tuberculosis, and Malaria (http://www.theglobalfund.org). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: peter.gething@zoo.ox.ac.uk (PWG); simon.i.hay@gmail.com (SIH)

Introduction

The international agenda shaping malaria control financing, research, and implementation is increasingly defined around the goal of regional elimination [1–6]. This ambition ostensibly extends to all human malarias, but whilst recent years have seen a surge in research attention for *Plasmodium falciparum*, the knowledge-base for the other major human malaria, *Plasmodium vivax*, is far less developed in almost every aspect [7–11]. During 2006–2009 just 3.1% of expenditures on malaria research and development were committed to *P. vivax* [12]. The notion that control approaches developed primarily for *P. falciparum* in

Author Summary

Plasmodium vivax is one of five parasites causing malaria in humans. Whilst it is found across a larger swathe of the globe and potentially affects a larger number of people than its more notorious cousin, *Plasmodium falciparum*, it receives a tiny fraction of the research attention and financing: around 3%. This neglect, coupled with the inherently more complex nature of vivax biology, means important knowledge gaps remain that limit our current ability to control the disease effectively. This patchy knowledge is becoming recognised as a cause for concern, in particular as the global community embraces the challenge of malaria elimination which, by definition, includes *P. vivax* and the other less common Plasmodium species as well as *P. falciparum*. Particularly conspicuous is the absence of an evidence-based map describing the intensity of *P.* vivax endemicity in different parts of the world. Such maps have proved important for other infectious diseases in supporting international policy formulation and regional disease control planning, implementation, and monitoring. In this study we present the first systematic effort to map the global endemicity of *P. vivax*. We assembled nearly 10,000 surveys worldwide in which communities had been tested for the prevalence of *P. vivax* infections. Using a spatial statistical model and additional data on environmental characteristics and Duffy negativity, a blood disorder that protects against *P. vivax*, we estimated the level of infection prevalence in every 5×5 km grid square across areas at risk. The resulting maps provide new insight into the geographical patterns of the disease, highlighting areas of the highest endemicity in South East Asia and small pockets of Amazonia, with very low endemic setting predominating in Africa. This new level of detailed mapping can contribute to a wider shift in our understanding of the spatial epidemiology of this important parasite.

holoendemic Africa can be transferred successfully to *P. vivax* is, however, increasingly acknowledged as inadequate [13–17]. Previous eradication campaigns have demonstrated that *P. vivax* frequently remains entrenched long after *P. falciparum* has been eliminated [18]. The prominence of *P. vivax* on the global health agenda has risen further as evidence accumulates of its capacity in some settings to cause severe disease and death [19–25], and of the very large numbers of people living at risk [26].

Amongst the many information gaps preventing rational strategies for *P. vivax* control and elimination, the absence of robust geographical assessments of risk has been identified as particularly conspicuous [9,27]. The endemic level of the disease determines its burden on children, adults, and pregnant women; the likely impact of different control measures; and the relative difficulty of elimination goals. Despite the conspicuous importance of these issues, there has been no systematic global assessment of endemicity. The Malaria Atlas Project was initiated in 2005 with an initial focus on *P. falciparum* that has led to global maps [28–30] for this parasite being integrated into policy planning at regional to international levels [4,31–36]. Here we present the outcome of an equivalent project to generate a comprehensive evidence-base on *P. vivax* infections worldwide, and to generate global risk maps for this hitherto neglected disease. We build on earlier work [26] defining the global range of the disease and broad classifications of populations at risk to now assess the levels of endemicity under which these several billion people live. This detailed depiction of geographically varying risk is intended to contribute to a much-needed paradigm shift towards geographically stratified and evidence-based planning for *P. vivax* control and elimination.

Numerous biological and epidemiological characteristics of *P. vivax* present unique challenges to defining and mapping metrics of risk. Unlike *P. falciparum*, infections include a dormant hypnozoite liver stage that can cause clinical relapse episodes [37,38]. These periodic events manifest as a blood-stage infection clinically indistinguishable from a primary infection and constitute a substantial, but geographically varying, proportion of total patent infection prevalence and disease burden within different populations [37,39–41]. The parasitemia of *P. vivax* typically occurs at much lower densities compared to those of falciparum malaria, and successful detection by any given means of survey is much less likely. Another major driver of the global *P. vivax* landscape is the influence of the Duffy negativity phenotype [42]. This inherited blood condition confers a high degree of protection against *P. vivax* infection and is present at very high frequencies in the majority of African populations, although is rare elsewhere [43]. These factors, amongst others, mean that the methodological framework for mapping *P. vivax* endemicity, and the interpretation of the resulting maps, are distinct from those already established for *P. falciparum* [28,29]. The effort described here strives to accommodate these important distinctions in developing a global distribution of endemic vivax malaria.

Methods

The modelling framework is displayed schematically in Figure 1. In brief, this involved (i) updating of the geographical limits of stable *P. vivax* transmission based on routine reporting data and biological masks; (ii) assembly of all available *P. vivax* parasite rate data globally; (iii) development of a Bayesian model-based geostatistical model to map *P. vivax* endemicity within the limits of stable transmission; and (iv) a model validation procedure. Details on each of these stages are provided below with more extensive descriptions included as Protocols S1, S2, S3, and S4.

Updating Estimates of the Geographical Limits of Endemic *Plasmodium vivax* in 2010

The first effort to systematically estimate the global extent of *P. vivax* transmission and define populations at risk was completed in 2009 [26]. As a first step in the current study, we have updated this work with a new round of data collection for the year 2010. The updated data assemblies and methods are described in full in Protocol S1. In brief, this work first involved the identification of 95 countries as endemic for *P. vivax* in 2010. From these, *P. vivax* annual parasite incidence (*Pv*API) routine case reports were assembled from 17,893 administrative units [44]. These *Pv*API and other medical intelligence data were combined with remote sensing surfaces and biological models [45] that identified areas where extreme aridity or temperature regimes would limit or preclude transmission (see Protocol S1). These components were combined to classify the world into areas likely to experience zero, unstable (*Pv*API $<0.1‰$ per annum), or stable (*Pv*API $\geq0.1‰$ per annum) *P. vivax* transmission. Despite the very high population frequencies of Duffy negativity across much of Africa, the presence of autochthonous transmission of *P. vivax* has been confirmed by a systematic literature review for 42 African countries [26]. We therefore treated Africa in the same way as elsewhere in this initial stage: regions were deemed to have stable *P. vivax* transmission unless the biological mask layers or *Pv*API data suggested otherwise.

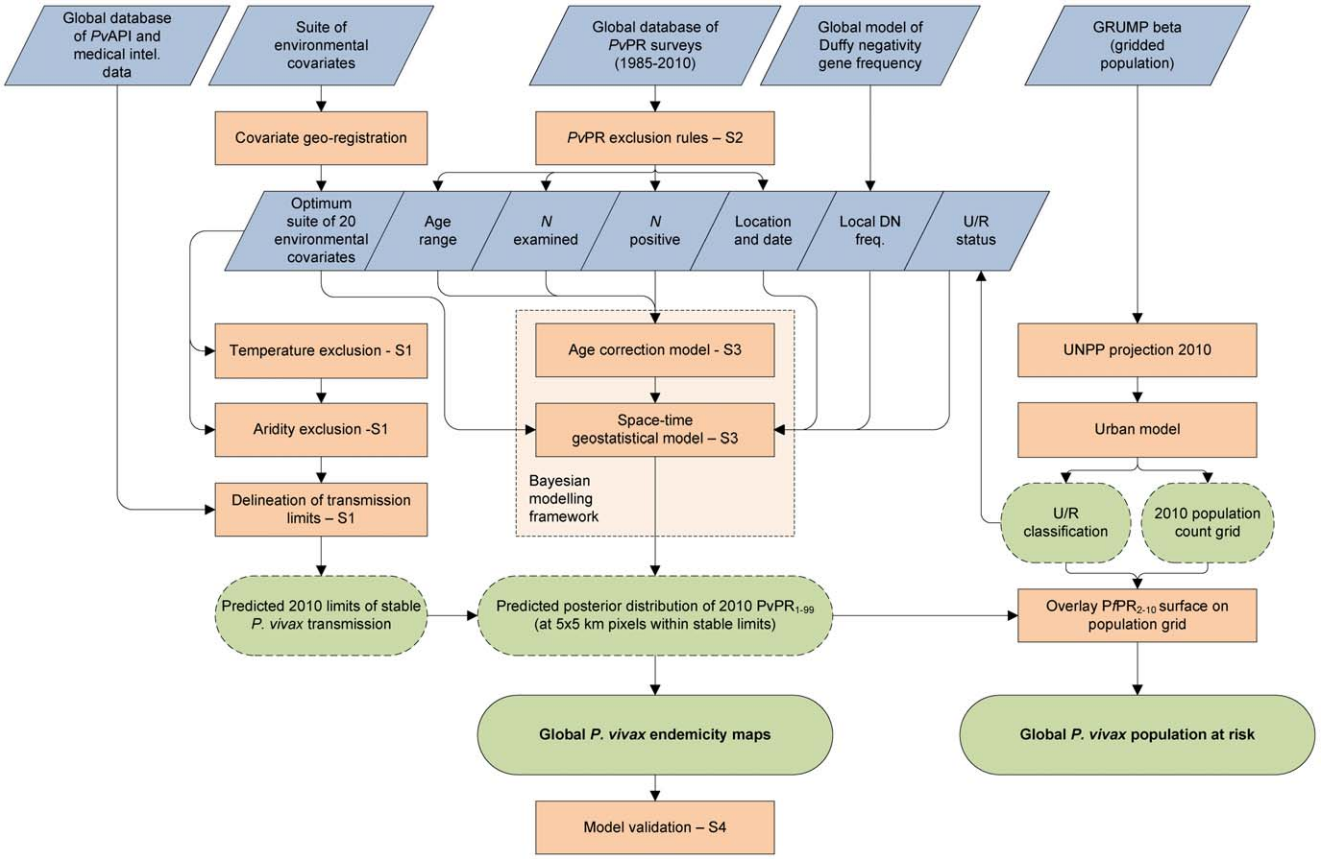


Figure 1. Schematic overview of the mapping procedures and methods for *Plasmodium vivax* endemicity. Blue boxes describe input data. Orange boxes denote models and experimental procedures; green boxes indicate output data (dashed lines represent intermediate outputs and solid lines final outputs). U/R = urban/rural; UNPP = United Nations Population Prospects. Labels S1-4 denote supplementrary information in Protocols S1, S2, S3, and S4.
doi:10.1371/journal.pntd.0001814.g001

Creating a Database of Georeferenced *Pv*PR Data

As with *P. falciparum*, the most globally ubiquitous and consistently measured metric of *P. vivax* endemicity is the parasite rate (*Pv*PR), defined as the proportion of randomly sampled individuals in a surveyed population with patent parasitemia in their peripheral blood as detected *via*, generally, microscopy or rapid diagnostic test (RDT). Whilst RDTs can provide lower sensitivity and specificity than conventional blood smear microscopy, and neither technique provides accuracy comparable to molecular diagnostics (such as polymerase chain reaction, PCR), the inclusion of both microscopically and RDT confirmed parasite rate data was considered important to maximise data availability and coverage across the endemic world.

To map endemicity within the boundaries of stable transmission, we first carried out an exhaustive search and assembly of georeferenced *Pv*PR survey data from formal and informal literature sources and direct communications with data generating organisations [46]. Full details of the data search strategy, abstraction and inclusion criteria, geopositioning and fidelity checking procedure are included in Protocol S2. The final database, completed on 25th November 2011, consisted of 9,970 quality-checked and spatiotemporally unique data points, spanning the period 1985–2010. Figure 2A maps the spatial distribution of these data and further summaries by survey origin, georeferencing source, time period, age group, sample size, and type of diagnostic used are provided in Protocol S2.

Modelling *Plasmodium vivax* Endemicity within Regions of Stable Transmission

We adopt model-based geostatistics (MBG) [47,48] as a robust and flexible modelling framework for generating continuous surfaces of malaria endemicity based on retrospectively assembled parasite rate survey data [28,29,49]. MBG models are a special class of generalised linear mixed models, with endemicity values at each target pixel predicted as a function of a geographically-varying mean and a weighted average of proximal data points. The mean can be defined as a multivariate function of environmental correlates of disease risk. A covariance function is used to characterise the spatial or space-time heterogeneity in the observed data, which in turn is used to define appropriate weights assigned to each data point when predicting at each pixel. This framework allows the uncertainty in predicted endemicity values to vary between pixels, depending on the observed variation, density and sample size of surveys in different locations and the predictive utility of the covariate suite. Parts of the map where survey data are dense, recent, and relatively homogenous will be predicted with least uncertainty, whilst regions with sparse or mainly old surveys, or where measured parasite rates are extremely variable, will have greater uncertainty. When MBG models are fitted using Bayesian inference and a Markov chain Monte Carlo (MCMC) algorithm, uncertainty in the final predictions as well as all model parameters can be represented in the form of predictive posterior distributions [50].

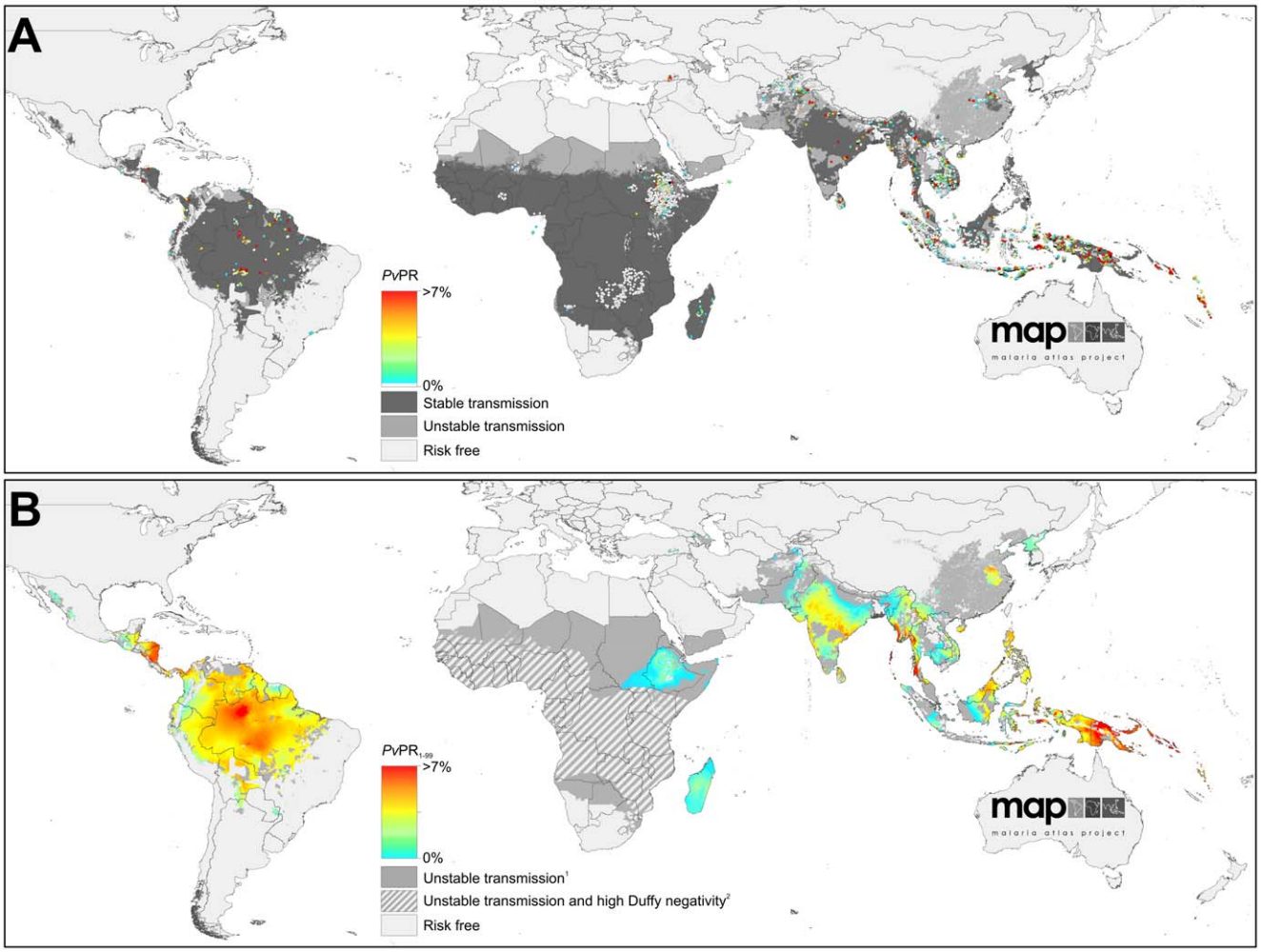


Figure 2. The spatial distribution of *Plasmodium vivax* malaria endemicity in 2010. Panel A shows the 2010 spatial limits of *P. vivax* malaria risk defined by *Pv*API with further medical intelligence, temperature and aridity masks. Areas were defined as stable (dark grey areas, where *Pv*API $\geq$0.1 per 1,000 pa), unstable (medium grey areas, where *Pv*API $<$0.1 per 1,000 pa) or no risk (light grey, where *Pv*API = 0 per 1,000 pa). The community surveys of *P. vivax* prevalence conducted between January 1985 and June 2010 are plotted. The survey data are presented as a continuum of light green to red (see map legend), with zero-valued surveys shown in white. Panel B shows the MBG point estimates of the annual mean $PvPR_{1-99}$ for 2010 within the spatial limits of stable *P. vivax* malaria transmission, displayed on the same colour scale. Areas within the stable limits in (A) that were predicted with high certainty ($>$0.9) to have a $PvPR_{1-99}$ less than 1% were classed as unstable. Areas in which Duffy negativity gene frequency is predicted to exceed 90% [43] are shown in hatching for additional context.
doi:10.1371/journal.pntd.0001814.g002

We developed for this study a modified version of the MBG framework used previously to model *P. falciparum* endemicity [28,29], with some core aspects of the model structure remaining unchanged and others altered to capture unique aspects of *P. vivax* biology and epidemiology. The model is presented in full in Protocol S3. As in earlier work [28,29,49], we adopt a space-time approach to allow surveys from a wide time period to inform predictions of contemporary risk. This includes the use of a spatiotemporal covariance function which is parameterised to downweight older data appropriately. We also retain a seasonal component in the covariance function, although we note that seasonality in transmission is often only weakly represented in *Pv*PR in part because of the confounding effect of relapses occurring outside peak transmission seasons [51]. A minimal set of covariates were included to inform prediction of the mean function, based on *a priori* expectations of the major environmental factors modulating endemicity. These were (i) an indicator variable defining areas as urban or rural based on the Global Rural Urban Mapping Project (GRUMP) urban extent product [52,53]; (ii) a long-term average vegetation index product as an indicator of overall moisture availability for vector oviposition and survival [54,55]; and (iii) a *P. vivax* specific index of temperature suitability derived from the same model used to delineate suitable areas on the basis of vector survival and sporogony [45].

Age Standardisation

Our assembly of *Pv*PR surveys was collected across a variety of age ranges and, since *P. vivax* infection status can vary systematically in different age groups within a defined community, it was necessary to standardise for this source of variability to allow all surveys to be used in the same model. We adopted the same model form as has been described [56] and used previously for *P. falciparum* [28,29], whereby population infection prevalence is expected to rise rapidly in early infancy and plateau during childhood before declining in early adolescence and adulthood. The timing and relative magnitude of these age profile features are

likely distinct between the two parasites in different endemic settings [51,57], and so the model was parameterised using an assembly of 67 finely age-stratified *Pv*PR surveys (Protocol S2), with estimation carried out in a Bayesian model using MCMC. The parameterised model was then used to convert all observed survey prevalences to a standardised age-independent value for use in modelling, and then further allowed the output prevalence predictions to be generated for any arbitrary age range. We chose to generate maps of all-age infection prevalence, defined as individuals of age one to 99 years (thus $PvPR_{1–99}$). We excluded infection in those less than one year of age from the standardisation because of the confounding effect of maternal antibodies, and because parasite rate surveys very rarely sample young infants. We deviated from the two-to-ten age range used for mapping *P. falciparum* [28,29] because the relatively lower prevalences has meant that surveys are far more commonly carried out across all age ranges.

Incorporating Duffy Negativity

Since Duffy negative individuals are largely refractory to *P. vivax* infection [58], high population frequencies of this phenotype have a dramatic suppressing effect on endemicity, even where conditions are otherwise well suited for transmission [26]. The predominance of Duffy negativity in Africa has led to a historical perception that *P. vivax* is absent from much of the continent, and a dearth of surveys or routine diagnoses testing for the parasite have served to entrench this mantra [59]. However, evidence exists of autochthonous *P. vivax* transmission across the continent [26], and therefore we did not preclude any areas at risk *a priori*. Instead, we used a recent map of estimated Duffy negativity phenotypic frequency [43] and incorporated the potential influence of this blood group directly in the MBG modelling framework. The mapped Duffy-negative population fraction at each location was excluded from the denominator in *Pv*PR survey data, such that any *P. vivax* positive individuals were considered to have arisen from the Duffy positive population subset. Thus in a location with 90% Duffy negativity, five positive individuals in a survey of 100 would give an assumed prevalence of 50% amongst Duffy positives. Correspondingly, prediction of *Pv*PR was then restricted to the Duffy positive proportion at each pixel, with the final prevalence estimate re-converted to relate to the total population. This approach has two key advantages. First, predicted *Pv*PR at each location could never exceed the Duffy positive proportion, therefore ensuring biological consistency between the *P. vivax* and Duffy negativity maps. Second, where *Pv*PR survey data were sparse across much of Africa, the predictions could effectively borrow strength from the Duffy negativity map because predictions of *Pv*PR were restricted to a much narrower range of possible values.

Model Implementation and Map Generation

The *P. vivax* endemic world was divided into four contiguous regions with broadly distinct biogeographical, entomological and epidemiological characteristics: the Americas and Africa formed separate regions, whilst Asia was subdivided into Central and South East sub-regions with a boundary at the Thailand-Malaysia border (see Protocol S2). This regionalisation was implemented in part to retain computational feasibility given the large number of data points, but also to allow model parameterisations to vary and better capture regional endemicity characteristics. Within each region, a separate MBG model was fitted using a bespoke MCMC algorithm [60] to generate predictions of $PvPR_{1–99}$ for every 5×5 km pixel within the limits of stable transmission. The prediction year was set to 2010 and model outputs represent an annualised average across the 12 months of that year. Model output consisted of a predicted posterior distribution of $PvPR_{1–99}$ for every pixel. A continuous endemicity map was generated using the mean of each posterior distribution as a point estimate. The uncertainty associated with predictions was summarised by maps showing the ratio of the posterior distribution inter-quartile range (IQR) to its mean. The IQR is a simple measure of the precision with which each *Pv*PR value was predicted, and standardisation by the mean produced an uncertainty index less affected by underlying prevalence levels and more illustrative of relative model performance driven by data densities in different locations. This index was then also weighted by the underlying population density to produce a second map indicative of those areas where uncertainty is likely to be most operationally important.

Refining Limits Definition and Population at Risk Estimates

In some regions within the estimated limits of stable transmission, $PvPR_{1–99}$ was predicted to be extremely low, either because of a dense abundance of survey data reporting zero infections or, in Africa, because of very high coincident Duffy negativity phenotype frequencies. Such areas are not appropriately described as being at risk of stable transmission and so we defined a decision rule whereby pixels predicted with high certainty (probability >0.9) of being less than 1% $PvPR_{1–99}$ were assigned to the unstable class, thereby modifying the original transmission limits. These augmented mapped limits were combined with a 2010 population surface derived from the GRUMP *beta* version [52,53] to estimate the number of people living at unstable or stable risk within each country and region. The fraction of the population estimated to be Duffy negative [43] within each pixel was considered at no risk and therefore excluded from these totals.

Model Validation

A model validation procedure was implemented whereby 10% of the survey points in each model region were selected using a spatially declustered random sampling procedure. These subsets were held out and the model re-fitted in full using the remaining 90%. Model predictions were then compared to the hold-out data points and a number of different aspects of model performance were assessed using validation statistics described previously [28,29]. The validation procedure is detailed in full in Protocol S4.

Results

Model Validation

Full validation results are presented in Protocol S4. In brief, examination of the mean error in the generation of the *P. vivax* malaria endemicity point-estimate surface revealed minimal overall bias in predicted *Pv*PR with a global mean error of -0.41 (Americas -1.38, Africa 0.03, Central Asia -0.43, South East Asia -0.43), with values in units of *Pv*PR on a percentage scale (see Protocol S4). The global value thus represents an overall tendency to underestimate prevalence by just under half of one percent. The mean absolute error, which measures the average magnitude of prediction errors, was 2.48 (Americas 5.05, Africa 0.53, Central Asia 1.52, South East Asia 3.37), again in units of *Pv*PR (see Protocol S4).

Global *Plasmodium vivax* Endemicity and Populations at Risk in 2010

The limits of stable and unstable *P. vivax* transmission, as defined using *Pv*API, biological exclusion masks and medical intelligence

data are shown in Figure 2A. The continuous surface of *P. vivax* endemicity predicted within those limits is shown in Figure 2B. The uncertainty map (posterior IQR:mean ratio) is shown in Figure 3A and the population-weighted version in Figure 3B.

We estimate that *P. vivax* was endemic across some 44 million square kilometres, approximately a third of the Earth's land surface. Around half of this area was located in Africa (51%) and a quarter each in the Americas (22%) and Asia (27%) (Table 1). However, the uneven distribution of global populations, coupled with the protective influence of Duffy negativity in Africa, meant that the distribution of populations at risk was very different. An estimated 2.48 billion people lived at any risk of *P. vivax* in 2010 (Table 1), of which a large majority lived in Central Asia (82%) with much smaller fractions in South East Asia (9%), the Americas (6%), and Africa (3%). Of these, 1.52 billion lived in areas of unstable transmission where risk is very low and case incidence is unlikely to exceed one per 10,000 per annum. The remaining 964 million people at risk lived in areas of stable transmission, representing a wide diversity of endemic levels. The global distribution of populations in each risk class was similar to the total at risk, such that over 80% of people in both classes lived in Central Asia (Table 1).

Plasmodium vivax Endemicity in the Americas

Areas endemic for *P. vivax* in the Americas extended to some 9.5 million square kilometres, of which the largest proportion was in the Amazonian region of Brazil (Figure 2B). Interestingly, only a relatively small fraction of these areas (15%) experienced unstable rather than stable transmission, suggesting a polarisation between areas at stable risk and those where the disease is absent altogether (Table 1). The regions of highest endemicity were found in Amazonia and in Central America – primarily Nicaragua and Honduras – with predicted mean $Pv\mathrm{PR}_{1-99}$ exceeding 7% in all three locations. An important feature of *P. vivax* throughout the Americas is that its distribution is approximately inverse to that of the population. This is particularly true of the two most populous endemic countries of the region, Brazil and Mexico, and it means that, whilst the Americas contributed 53% of the land area experiencing stable transmission worldwide, they housed only 5% of the global population at that level of risk.

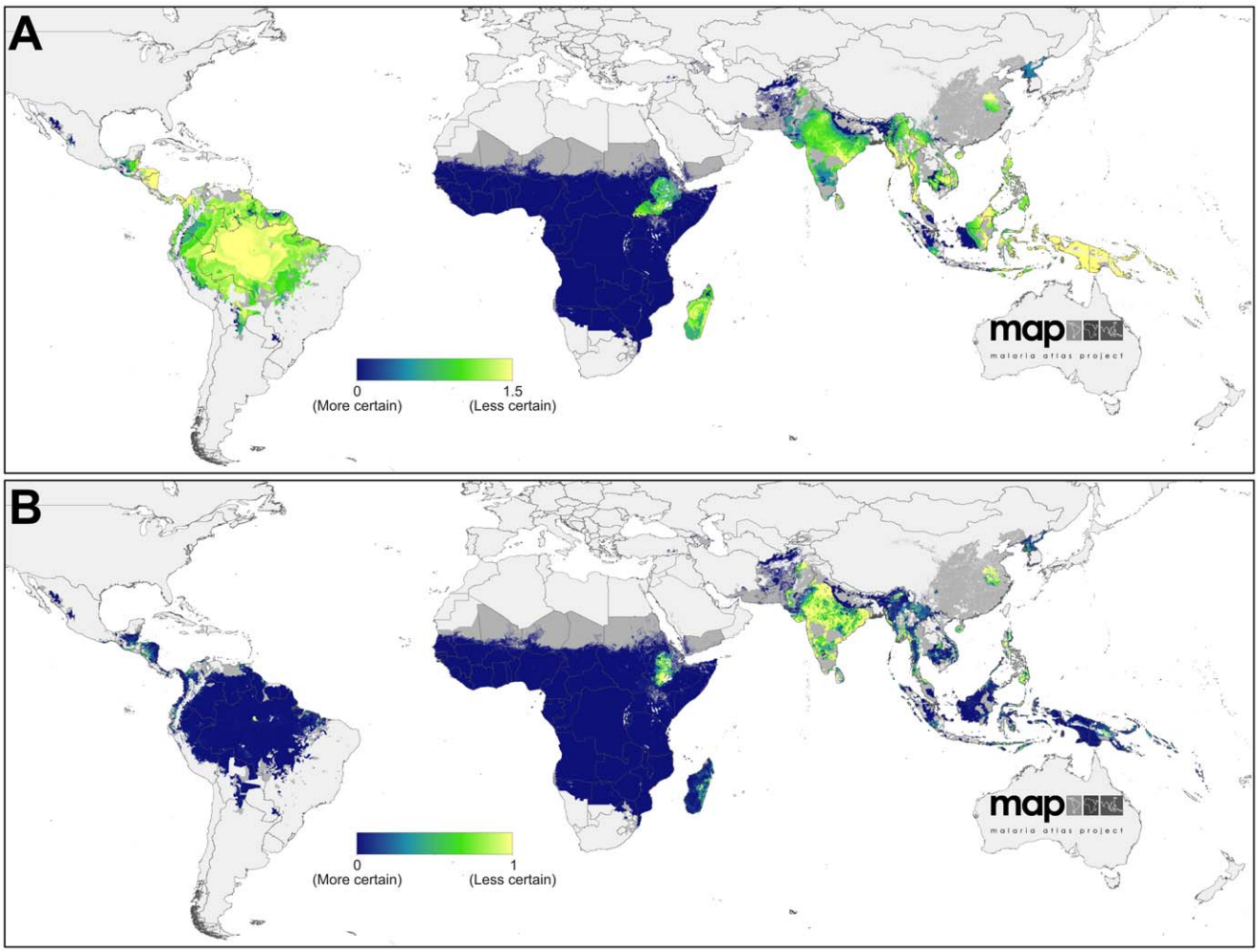


Figure 3. Uncertainty associated with predictions of *Plasmodium vivax* endemicity. Panel A shows the ratio of the posterior inter-quartile range to the posterior mean prediction at each pixel. Large values indicate greater uncertainty: the model predicts a relatively wide range of $Pv\mathrm{PR}_{1-99}$ as being equally plausible given the surrounding data. Conversely, smaller values indicate a tighter range of values have been predicted and, thus, a higher degree of certainty in the prediction. Panel B shows the same index multiplied by the underlying population density and rescaled to 0–1 to correspond to Panel A. Higher values indicate areas with high uncertainty and large populations.
doi:10.1371/journal.pntd.0001814.g003

Table 1. Area and populations at risk of *Plasmodium vivax* malaria in 2010.

Region	Area (million km^2)			Population (millions)		
	Unstable	Stable	Any risk	Unstable	Stable	Any risk
America	1.38	8.08	9.46	87.66	49.79	137.45
Africa+	20.60	1.86	22.46	48.72	37.66	86.38
C Asia	5.60	3.63	9.24	1,236.92	812.55	2,049.47
SE Asia	0.96	1.78	2.74	150.17	64.90	215.07
World	28.55	15.35	43.90	1,523.47	964.90	2,488.37

Risk is stratified into unstable risk (*Pv*API<0.1 per 1,000 people pa) and stable risk (*Pv*API≥0.1 per 1,000 people pa).
doi:10.1371/journal.pntd.0001814.t001

Uncertainty in predicted $PvPR_{1-99}$ was relatively high throughout much of the Americas (Figure 3B). This reflects the heterogeneous landscape of endemicity coupled with the generally scarce availability of parasite rate surveys in the region (see Figure 2A). However, when this uncertainty is weighted by the underlying population density (Figure 3B), its significance on a global scale is placed in context: because most areas at stable risk are sparsely populated, the population-weighted uncertainty was very low compared to parts of Africa and much of Asia.

Plasmodium vivax Endemicity in Africa, Yemen and Saudi Arabia (Africa+)

Our decision to assume stable transmission of *P. vivax* in Africa unless robust *Pv*API or biological mask data confirmed otherwise meant that much of the continent south of the Sahara was initially classified as being at stable risk (Figure 2A). However, by implementing the MBG predictions of $PvPR_{1-99}$ throughout this range and reclassifying *a posteriori* those areas likely to fall below an endemicity threshold of 1% $PvPR_{1-99}$, the majority of stable risk areas were downgraded to unstable (Figure 2B). Thus, in the final maps, 92% of endemic Africa was at unstable risk, with the majority of Madagascar and Ethiopia, and parts of South Sudan and Somalia making up most of the remaining area at stable risk. Even in these areas, endemicity was uniformly low, with predicted endemicity values rarely exceeding a point estimate of 2% $PvPR_{1-99}$. We augmented the final map with an additional overlay mask delineating areas where Duffy negativity phenotype prevalence has been predicted to exceed 90% (Figure 2B). The influence of this blood group on the estimated populations at risk is profound: of the 840 million Africans living in areas within which transmission is predicted to occur, only 86 million were considered at risk, contributing just 3% to the global total (Table 1).

Uncertainty in predicted $PvPR_{1-99}$ followed a similar pattern to the magnitude of the predictions themselves (Figure 3B). Certainty around the very low predicted endemicity values covering most of the continent was extremely high – reflecting the increased precision gained by incorporating the Duffy negativity information that compensated for the paucity of *P. vivax* parasite rate surveys on the continent. The pockets of higher endemicity in Madagascar and northern East Africa were predicted with far less certainty. In the population-weighted uncertainty map (Figure 3B), the lower population densities of Madagascar reduced the index on that island whereas the densely populated Ethiopian highlands remained high.

Plasmodium vivax Endemicity in Central and South East Asia

Large swathes of high endemicity, very large population densities and a negligible presence of Duffy negativity combine to make the central and south-eastern regions of Asia by far the most globally significant for *P. vivax*. We estimate that India alone contributed nearly half (46%) of the global population at risk, and two thirds (67%) of those at stable risk. China is another major contributor with 19% of the global populations at risk, primarily in unstable transmission regions, whilst Indonesia and Pakistan together contributed a further 12%. Within regions of stable transmission, endemicity is predicted to be extremely heterogeneous (Figure 2B). Areas where the point estimate of $PvPR_{1-99}$ exceeded 7% were found in small pockets of India, Myanmar, Indonesia, and the Solomon Islands, with the largest such region located in Papua New Guinea.

The uncertainty map (Figure 3A) reveals how the most precise predictions were associated with areas of uniformly low endemicity and abundant surveys, such as Afghanistan and parts of Sumatra and Kalimantan in Indonesia. Conversely, areas with higher or more heterogeneous endemicity, such as throughout the island of New Guinea, were the most uncertain. The population-weighted uncertainty map (Figure 3B) differs substantively, indicating how the populous areas of Indonesia, for example, were relatively precisely predicted whereas India, China, and the Philippines had the largest per-capita uncertainty.

Discussion

The status of *P.* vivax as a major public health threat affecting the world's most populous regions is becoming increasingly well documented. The mantra of vivax malaria being a very rarely threatening and relatively benign disease [7,10] has been challenged with evidence suggesting that it can contribute a significant proportion of severe malaria disease and death attributable to malaria in some settings [61]. Some reports have pointed especially to very young children being a major source of morbidity [20,62] and some hospital-based studies have reported comparable mortality rates between patients classified with severe *P. vivax* and severe *P. falciparum* [21,24,63]. The recognition of a lethal threat by this parasite comes with evidence of failing chemotherapeutics against the acute attack [64] and overdue acknowledgement of the practical inadequacy of the only available therapy against relapse [65]. As the international community defines increasingly ambitious targets to minimise malaria illness and death [66–68], and to progressively eliminate the disease from endemic areas [1–6], further sustained neglect of *P. vivax* becomes increasingly untenable.

Here we have presented the first systematic attempt to map the global distribution of *P. vivax* endemicity using a defined evidence base, transparent methodologies, and with measured uncertainty. These new maps aim to contribute to a more rational international appraisal of the importance of *P. vivax* in the broad context of

malaria control and elimination policies, as well as providing a practical tool to support control planning at national and sub-national levels.

Interpreting *P. vivax* Endemicity in 2010

In 2010, areas endemic for *P. vivax* covered a huge geographical range spanning three major continental zones and extending into temperate climates. In the Americas, whilst important pockets of high endemicity are present, the majority of areas of stable transmission coincide with lower population densities, diminishing the contribution of this continent to global populations at risk. In Africa the protection conferred by Duffy negativity to most of the population means the large swathes of the continent in which transmission may occur contain only small populations at biological risk. Thus it is primarily in Asia where very large populations coincide with extensive high endemic regions, and as a result nine out of every ten people at risk of *P. vivax* globally live on that continent.

A number of important contrasts arise when comparing this map with the equivalent 2010 iteration for *P. falciparum* [28]. Perhaps most obvious are the lower levels of observed endemicity at which *P. vivax* tends to exist within populations experiencing stable transmission. We used a cartographic scale between 0% and 7% to differentiate global variation in *P. vivax* endemicity, although point estimates exceeded that upper threshold in localised areas. For *P. falciparum* the equivalent scale spanned 0% to 70% [28], suggesting an approximate order-of-magnitude difference in prevalence of patent parasitemia. In part, this difference reflects the decision to standardise our predictions across the 1–99 age range, and values would have been higher if we had opted for the peak 2–10 age range used for *P. falciparum*. This difference might be accentuated by the likely more rapid acquisition of immunity to *P. vivax* than *P. falciparum* in the most highly endemic areas [57]. A number of other biological and epidemiological differences between the two species also mean these lower apparent levels of endemicity must be interpreted differently. One factor is the lower sensitivities of microscopy and RDT diagnoses for a given level of *P. vivax* infection prevalence, because infections tend to be associated with much lower parasite densities which increase the likelihood of false negative diagnoses [9]. A number of studies in both high and low endemic settings have found microscopy to underestimate prevalence by a factor of up to three when compared with molecular diagnosis [57,69–72]. The decreasing cost and time implications of molecular diagnosis may mean that these gold standard diagnostic techniques become the standard for parasite rate surveys in the future. A global map of PCR-positive parasitemia rates would almost certainly reveal a larger underlying reservoir of infections and, possibly, reveal systematic differences in patterns of endemicity than we are able to resolve currently with less sensitive diagnostic methods.

The lower parasite loads must be interpreted in the context of implications for progression to clinical disease. For example, *Plasmodium vivax* is known to induce fevers at comparatively lower parasite densities than *P. falciparum*, a feature likely linked to overall inflammatory responses of greater magnitude [16]. *P. vivax* is also comparable to *P. falciparum* in its potential to cause anaemia regardless of lower parasite densities, due to a combination of dyserythropoesis and repeated bouts of haemolysis [22]. A recent hospital-based study at a site in eastern Indonesia of hypo- to meso-endemic transmission of both species showed far lower frequencies of parasitemia >6,000/uL among inpatients classified as having not serious, serious, and fatal illness with a diagnosis of *P. vivax* compared to *P. falciparum* [24]. Further, the majority of case reports describing severe and fatal illness with a diagnosis of vivax malaria typically show parasitemia >5,000/uL. In contrast, the World Health Organization threshold for severe illness attributable to hyperparasitemia with *P. falciparum* is >200,000/uL [73]. In brief, the relationship between prevalence and risk of disease and transmission for *P. vivax* is distinct from that for *P. falciparum*, and it is weighted more heavily towards substantial risks at much lower parasite densities and levels of prevalence of microscopically patent parasitemia.

The capacity of *P. vivax* hypnozoites to induce relapsing infections has a number of important implications. First, because dormant liver stage infections are not detectable in routine parasite rate surveys, our maps do not capture the potentially very large reservoir of asymptomatic infections sequestered in each population. Evidence is emerging that this hidden reservoir may be substantially larger than previously thought, with long-latency *P. vivax* phenotypes both prevalent and geographically widespread [37]. Whilst not contributing to clinical disease until activated, these dormant hypnozoites ultimately play a vital role in sustaining transmission since they are refractory to blood-stage antimalarial chemotherapy and interventions to reduce transmission. Hypnozoites also ensure an ability of *P. vivax* to survive in climatic conditions that cannot sustain *P. falciparum* transmission. Second, the *P. vivax* parasite rates observed in population surveys detect both new and relapsing infections, although the two are almost never distinguishable. This confounds the relationship between observed infection prevalence and measures of transmission intensity such as force of infection or the entomological inoculation rate. This, in turn, has implications for the use of transmission models seeking to evaluate or optimise control options for *P. vivax* [2,9,27,74]. The current unavailability of any diagnostic method for detecting hypnozoites [75] and our resulting ignorance about the size and geographic distribution of this reservoir therefore remain critical knowledge gaps limiting the feasibility of regional elimination [9]. It is also worth noting that conventional parasite rate data do not measure multiplicity of infection which is an additional potential confounding effect between observed infection prevalence and transmission intensity.

P. vivax in Africa and Duffy Polymorphism

Our map of *P. vivax* endemicity and estimates of populations at risk in Africa are heavily influenced by a single assumption: that the fraction of the population estimated to be negative for the Duffy antigen [43] is refractory to infection with *P. vivax*. A body of empirical evidence is growing, however that *P. vivax* can infect and cause disease in Duffy negative individuals, as reported in Madagascar [76] and mainland sub-Saharan Africa [77–80] as well as outside Africa [81,82]. Whether the invasion of erythrocytes *via* Duffy antigen-independent pathways is a newly evolved mechanism, or whether this capacity has been overlooked by the misdiagnosis of *P. vivax* in Africa as *P. ovale* remains unresolved [9,42,59]. Whilst this accumulated evidence stands contrary to our simplifying assumption of complete protection in Duffy negative individuals, there is currently no evidence to suggest that such infections are anything but rare and thus are unlikely to have any substantive influence on the epidemiology or infection prevalence of *P. vivax* at the population scale throughout most of Africa. We also make no provision in our model for a protective effect in Duffy-negative heterozygotes, although such protection has been observed in some settings [83–86]. The movement and mixing within Africa of human populations from diverse ethnographic backgrounds complicates contemporary patterns of Duffy negativity and, in principle, could yield local populations with substantially reduced protection from *P. vivax* infection in the future. Indeed, the implications for our map of population

movement go beyond the effect of Duffy negativity: the carriage of parasites from high to low endemic regions, for example by migratory workers, may play an important role in sustaining transmission in some regions and further research is required to investigate such processes.

Mapping to Guide Control

There exists for *P. falciparum* a history of control strategies linked explicitly to defined strata of endemicity, starting with the first Global Malaria Eradication Programme [18,87,88] and undergoing a series of refinements that now feature in contemporary control and elimination efforts. Most recently, stratification has been supported by insights gained from mathematical models linking endemic levels to optimum intervention suites, control options, and timelines for elimination planning [2,89–95]. In stark contrast, control options for *P. vivax* are rarely differentiated by endemicity, and there is little consensus around how this may be done. In part, the absence of agreed control-oriented strata of *P. vivax* endemicity stems from the biological complexities and knowledge gaps that prevent direct interpretation of infection prevalence as a metric for guiding control. It is also to some extent inevitable that the dogma of unstratified control becomes self-propagating: risk maps are not created because control is not differentiated by endemicity, but that differentiation cannot proceed without reliable maps.

As well as providing a basis for stratified control and treatment, the endemicity maps presented here have a number of potential applications in combination with other related maps. First, there is an urgent need to better identify regions where high *P. vivax* endemicity is coincident with significant population prevalence of glucose-6-phosphate dehydrogenase deficiency (G6PDd). This inherited blood disorder plays a key role in chemotherapy policy for *P. vivax* because primaquine, the only registered drug active against the hypnozoite liver stage is contra-indicated in G6PDd individuals in whom it can cause severe and potentially fatal haemolytic reactions [96,97]. A new global map of G6PDd prevalence is now available (Howes et al, submitted) which can be combined with the endemicity maps presented here to provide a rational basis for estimating adverse outcomes and setting appropriate testing and treatment protocols. Moreover, in practice most clinical infections are managed without differentiating the causative parasite species: combining the endemicity maps for *P. vivax* and *P. falciparum* may therefore inform unified strategies for malaria control programs and policy [28]. It has been proposed, for example, that artemesinin-based combination therapy (ACT) be adopted for all presumptively diagnosed malaria in areas coendemic for both species, as opposed to a separate ACT/chloroquine treatment strategy [98]. Further, in some regions more than 50% of patients diagnosed with falciparum malaria go on to experience an attack of vivax malaria in the absence of risk of reinfection [99]. This high prevalence of hypnozoites may also justify presumptive therapy with primaquine against relapse with any diagnosis of malaria where the two species occur at relatively high frequencies. Such geographically specific cross-parasite treatment considerations hinge on robust risk maps for both species.

Future Challenges in *P. vivax* Cartography

Numerous research and operational challenges remain unaddressed that would provide vital insights into the geographical distribution of *P. vivax* and its impacts on populations. Perhaps the highest priority is to improve understanding of the link between infection prevalence and clinical burden in both *P. vivax* mono-endemic settings and where it is coendemic with *P. falciparum*. Official estimates of national and regional disease burdens for *P. vivax* remain reliant on routine case reporting of unknown fidelity and are only crudely distinguished from *P. falciparum* [100]. It is illuminating that only 53 of the 95 *P. vivax* endemic countries were able to provide vivax-specific routine case reporting data, and there is a clear mandate for strengthening the routine diagnosis and reporting of *P. vivax* cases. Cartographic approaches to estimating *P. vivax* burden can therefore play a crucial role in triangulating with these estimates to provide insight into the distribution of the disease independent of health system surveillance and its attendant biases [27,101–105]. There is also a particular need to define burden and clinical outcomes associated with *P. vivax* in pregnancy [9,106] and other clinically vulnerable groups, most notably young children. Linking infection prevalence to clinical burden implies the need to better understand the contribution of relapsing infections to disease. Whilst the magnitude of this contribution is known to be highly heterogeneous, its geographical pattern is poorly measured and causal factors only partially understood [39,41].

Further challenges lie in understanding how *P. falciparum* and *P. vivax* interact within human hosts and how these interactions manifest at population levels. Comparison of the maps for each species reveals a complete spectrum from areas endemic for only one parasite through to others where both species are present at broadly equal levels. Whilst identifying these patterns of coendemicity is an important first step, the implications in terms of risks of coinfection and clinical outcomes, antagonistic mechanisms leading to elevated severe disease risk, or cross-protective mechanisms of acquired immunity remain disputed [20,107–109].

Conclusions

To meet international targets for reduced malaria illness and death, and to progress the cause of regional elimination, the malaria research and control communities can no longer afford to neglect the impact of *P. vivax*. Its unique biology and global ubiquity present challenges to its elimination that greatly surpass those of its higher-profile cousin, *P. falciparum*. Making serious gains against the disease will require substantive strengthening of the evidence base on almost every aspect of its biology, epidemiology, control and treatment. The maps presented here are intended to contribute to this effort. They are all made freely available from the MAP website [110] along with regional and individual maps for every malaria-endemic country. Users can access individual map images or download the global surfaces for use in a geographical information system, allowing them to integrate this work within their own analyses or produce bespoke data overlays and displays. We will also make available, where permissions have been obtained, all underlying *P. vivax* parasite rate surveys used in this work.

Supporting Information

Protocol S1 Updating the global spatial limits of *Plasmodium vivax* malaria transmission for 2010. S1.1 Overview. S1.2 Identifying Countries Considered *P. vivax* Malaria Endemic. S1.3 Updating National Risk Extents with *P. vivax* Annual Parasite Incidence Data. S1.4 Biological Masks of Transmission Exclusion. S1.5 Risk Modulation Based on Medical Intelligence. S1.6 Assembling the *P. vivax* Spatial Limits Map. S1.7 Refining Regions of Unstable Transmission after MBG Modelling. S1.8 Predicting Populations at Risk of *P. vivax* in 2010.
(DOC)

Protocol S2 The Malaria Atlas Project *Plasmodium vivax* parasite prevalence database. S2.1 Assembling the

*Pv*PR Data. S2.2 Database Fidelity Checks. S2.3 Data Exclusions. S2.4 The *Pv*PR Input Data Set. S2.5 Age-Standardisation. S2.6 Regionalisation.
(DOC)

Protocol S3 Bayesian model-based geostatistical framework for predicting $Pv\mathrm{PR}_{1-99}$. S3.1 Bayesian Inference. S3.2 Model Overview. S3.3 Formal Presentation of Model.
(DOC)

Protocol S4 Model validation procedures and additional results. S4.1 Creation of Validation Sets. S4.2 Procedures for Testing Model Performance. S4.3 Validation Results.
(DOC)

Acknowledgments

The large global assembly of parasite prevalence data was critically dependent on the generous contributions of data made by a large number of people in the malaria research and control communities and these individuals are listed on the MAP website (http://www.map.ac.uk/acknowledgements). We thank Professor David Rogers for providing the Fourier-processed remote sensing data. We are grateful for the comments of three anonymous referees that have helped strengthen the manuscript.

Author Contributions

Conceived and designed the experiments: PWG SIH. Performed the experiments: PWG APP DLS. Analyzed the data: PWG APP DLS IRFE CAG KEB. Contributed reagents/materials/analysis tools: IRFE CLM CAG MFM KEB APP AJT REH DBG PH HFLW RNP IM JKB. Wrote the paper: PWG.

References

1. MalERA (2011) Introduction: a research agenda to underpin malaria eradication. PLoS Med 8: e1000406.
2. Chitnis N, Schapira A, Smith DL, Smith T, Hay SI, et al. (2010) Mathematical modeling to support malaria control and elimination. Roll Back Malaria Progress and Impact Series, number 5. Geneva, Switzerland: Roll Back Malaria. 48 p.
3. Tanner M, Hommel M (2010) Towards malaria elimination–a new thematic series. Malaria J 9: 24.
4. Feachem RGA, Phillips AA, Targett GA, on behalf of the Malaria Elimination Group, editors (2009) Shrinking the Malaria Map: a Prospectus on Malaria Elimination. San Francisco, U.S.A.: The Global Health Group, University of California - Santa Cruz Global Health Sciences. 187 p.
5. Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, et al. (2010) Operational strategies to achieve and maintain malaria elimination. Lancet 376: 1592–1603.
6. Tatem A, Smith D, Gething P, Kabaria C, Snow R, et al. (2010) Ranking elimination feasibility among malaria endemic countries. Lancet 376: 1579–1591.
7. Baird JK (2007) Neglect of *Plasmodium vivax* malaria. Trends Parasitol 23: 533–539.
8. Mendis K, Sina BJ, Marchesini P, Carter R (2001) The neglected burden of *Plasmodium vivax* malaria. Am J Trop Med Hyg 64: 97–106.
9. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, et al. (2009) Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. Lancet Infect Dis 9: 555–566.
10. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, et al. (2007) *Vivax* malaria: neglected and not benign. Am J Trop Med Hyg 77: 79–87.
11. Carlton JM, Sina BJ, Adams JH (2011) Why Is *Plasmodium vivax* a neglected tropical disease? PLoS Neglect Trop D 5: e1160.
12. PATH (2011) Staying the Course? Malaria Research and Development in a Time of Economic Uncertainty. Seattle: PATH.
13. Bockarie MJ, Dagoro H (2006) Are insecticide-treated bednets more protective against *Plasmodium falciparum* than *Plasmodium vivax*-infected mosquitoes? Malaria J 5: 15.
14. Luxemburger C, Perea WA, Delmas G, Pruja C, Pecoul B, et al. (1994) Permethrin-impregnated bed nets for the prevention of malaria in schoolchildren on the Thai-Burmese border. T Roy Soc Trop Med H 88: 155–159.
15. Bousema T, Drakeley C (2011) Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. Clin Microbiol Rev 24: 377–410.
16. Andrade B, Reis-Filho A, Souza-Neto S, Clarencio J, Camargo L, et al. (2010) Severe *Plasmodium vivax* malaria exhibits marked inflammatory imbalance. Malaria J 9: 13.
17. Baird JK (2010) Eliminating malaria - all of them. Lancet 376: 1883–1885.
18. Yekutiel P (1980) III The Global Malaria Eradication Campaign. In: Klingberg MA, editor. Eradication of infectious diseases: a critical study. Basel, Switzerland: Karger. pp. 34–88.
19. Barcus MJ, Basri H, Picarima H, Manyakori C, Sekartuti, et al. (2007) Demographic risk factors for severe and fatal *vivax* and *falciparum* malaria among hospital admissions in northeastern Indonesian Papua. Am J Trop Med Hyg 77: 984–991.
20. Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, et al. (2008) *Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea. PLoS Med 5: e127.
21. Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, et al. (2008) Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. PLoS Med 5: e128.
22. Anstey NM, Russell B, Yeo TW, Price RN (2009) The pathophysiology of *vivax* malaria. Trends Parasitol 25: 220–227.
23. Kochar DK, Das A, Kochar SK, Saxena V, Sirohi P, et al. (2009) Severe *Plasmodium vivax* malaria: a report on serial cases from Bikaner in northwestern India. Am J Trop Med Hyg 80: 194–198.
24. Nurleila S, Syafruddin D, Elyazar IRF, Baird JK (2011) Morbidity and mortality caused by *falciparum* and *vivax* malaria in hypo- to meso-endemic West Sumba, Indonesia: a two year retrospective hospital-based study. Am J Trop Med Hyg 85 (Suppl 451-A-516): 471.
25. Kochar DK, Saxena V, Singh N, Kochar SK, Kumar SV, et al. (2005) *Plasmodium vivax* malaria. Emerg Infect Dis 11: 132–134.
26. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. PLoS Neglect Trop D 4: e774.
27. MalERA (2011) A Research Agenda for Malaria Eradication: monitoring, evaluation, and surveillance. PLoS Med 8: e1000400.
28. Gething P, Patil A, Smith D, Guerra C, Elyazar I, et al. (2011) A new world malaria map: *Plasmodium falciparum* endemicity in 2010. Malaria J 10: 378.
29. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6: e1000048.
30. Hay SI, Snow RW (2006) The Malaria Atlas Project: developing global maps of malaria risk. PLoS Med 3: e473.
31. Global Partnership to Roll Back Malaria, Johansson EW, Cibulskis RE, Steketee RW (2010) Malaria funding and resource utilization: the first decade of Roll Back Malaria. Geneva, Switzerland: World Health Organization on behalf of the Roll Back Malaria Partnership Secretariat. 95 p.
32. McLaughlin C, Levy J, Noonan K, Rosqueta K (2009) Lifting the burden of malaria: an investment guide for impact-driven philanthropy. Philadelphia: The Center for High Impact Philanthropy. 1–79 p.
33. Zanzibar Malaria Control Program (2009) Malaria elimination in Zanzibar: a feasibility assessment. Zanzibar, Tanzania: Zanzibar Ministry of Health and Social Welfare.
34. Anonymous (2009) Kenya Malaria Monitoring and Evaluation Plan 2009–2017. Nairobi, Kenya: Division of Malaria Control, Ministry of Public Health and Sanitation, Government of Kenya.
35. World Bank (2009) World Development Report 2009: Reshaping Economic Geography. Washington, DC. 383 p.
36. DFID (2010) Breaking the Cycle: Saving Lives and Protecting the Future. The UK's Framework for Results for Malaria in the Developing World. London, UK. 65 p.
37. White NJ (2011) Determinants of relapse periodicity in *Plasmodium vivax* malaria. Malaria J 10: 297.
38. Krotoski WA, Collins WE, Bray RS, Garnham PC, Cogswell FB, et al. (1982) Demonstration of hypnozoites in sporozoite-transmitted *Plasmodium vivax* infection. Am J Trop Med Hyg 31: 1291–1293.
39. Betuela I, Rosanas A, Kiniboro B, Stanisic D, Samol L, et al. (2011) Relapses are contributing significantly to risk of *P. vivax* infection and disease in Papua New Guinean children 1–5 years of age. Trop Med Int Health 16: 60–61.
40. Orjuela-Sanchez P, da Silva NS, da Silva-Nunes M, Ferreira MU (2009) Recurrent parasitemias and population dynamics of *Plasmodium vivax* polymorphisms in rural Amazonia. Am J Trop Med Hyg 81: 961–968.
41. Battle KE, Van Boeckel T, Gething PW, Baird JK, Hay SI (2011) A review of the geographical variations in *Plasmodium vivax* relapse rate. Am J Trop Med Hyg 85 (Suppl 451-A-516): 470.
42. Mercereau-Puijalon O, Menard D (2010) *Plasmodium vivax* and the Duffy antigen: a paradigm revisited. Transfus Clin Biol: 176–183.
43. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, et al. (2011) The global distribution of the Duffy blood group. Nat Commun 2: 266.
44. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, et al. (2008) The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. PLoS Med 5: e38.
45. Gething PW, Van Boeckel T, Smith DL, Guerra CA, Patil AP, et al. (2011) Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. Parasite Vector 4: 92.

46. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, et al. (2007) Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. Malaria J 6: 17.
47. Diggle PJ, Tawn JA, Moyeed RA (1998) Model-based geostatistics. J Roy Stat Soc C-App 47: 299–326.
48. Diggle PJ, Ribeiro PJ (2007) Model-based geostatistics; Bickel P, Diggle P, Fienberg S, Gather U, Olkin I et al., editors. New York: Springer. 228 p.
49. Gething PW, Patil AP, Hay SI (2010) Quantifying aggregated uncertainty in *Plasmodium falciparum* malaria prevalence and populations at risk *via* efficient space-time geostatistical joint simulation. PLoS Comput Biol 6: e1000724.
50. Patil AP, Gething PW, Piel FB, Hay SI (2011) Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol 27: 245–252.
51. Lin E, Kiniboro B, Gray L, Dobbie S, Robinson L, et al. (2010) Differential patterns of infection and disease with *P. falciparum* and *P. vivax* in young Papua New Guinean children. PLoS One 5: e9047.
52. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, et al. (2006) Determining global population distribution: methods, applications and data. Adv Parasitol 62: 119–156.
53. CIESIN/IFPRI/WB/CIAT (2007) Global Rural Urban Mapping Project (GRUMP) alpha: Gridded Population of the World, version 2, with urban reallocation (GPW-UR). Available: http://sedac.ciesin.columbia.edu/gpw. Palisades, New York, USA: Center for International Earth Science Information Network, Columbia University/International Food Policy Research Institute/ The World Bank/and Centro Internacional de Agricultura Tropical.
54. Scharlemann JP, Benz D, Hay SI, Purse BV, Tatem AJ, et al. (2008) Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing. MODIS data. 3: e1408.
55. Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ (2006) Global environmental data for mapping infectious disease distribution. Adv Parasitol 62: 37–77.
56. Smith DL, Guerra CA, Snow RW, Hay SI (2007) Standardizing estimates of the *Plasmodium falciparum* parasite rate. Malaria J 6: 131.
57. Mueller I, Widmer S, Michel D, Maraga S, McNamara DT, et al. (2009) High sensitivity detection of Plasmodium species reveals positive correlations between infections of different species, shifts in age distribution and reduced local variation in Papua New Guinea. Malar J 8: 41.
58. Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) Resistance factor to *Plasmodium vivax* in blacks Duffy-blood-group genotype, FyFy. New Engl J Med 295: 302–304.
59. Rosenberg R (2007) *Plasmodium vivax* in Africa: hidden in plain sight? Trends Parasitol 23: 193–196.
60. Patil A, Huard D, Fonnesbeck CJ (2010) PyMC: Bayesian stochastic modelling in Python. J Stat Softw 35: e1000301.
61. Price RN, Douglas NM, Anstey NM (2009) New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. Curr Opin Infect Dis 22: 430–435.
62. Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Hasanuddin A, et al. (2009) *Vivax* malaria: a major cause of morbidity in early infancy. Clin Infect Dis 48: 1704–1712.
63. Rogerson SJ, Carter R (2008) Severe *vivax* malaria: newly recognised or rediscovered. PLoS Med 5: e136.
64. Baird JK (2009) Resistance to therapies for infection by *Plasmodium vivax*. Clin Microbiol Rev 22: 508–534.
65. Baird JK, Surjadjaja C (2011) Consideration of ethics in primaquine therapy against malaria transmission. Trends Parasitol 27: 11–16.
66. Anonymous The Abuja Declaration and the Plan of Action: An extract from The African Summit on Roll Back Malaria, Abuja, 25 April 2000 (WHO/ CDS/RBM/2000.17) Roll Back Malaria/World Health Organization. 1–11 p.
67. WHO (2005) Global strategic plan. Roll Back Malaria. 2005–2015. Geneva: World Health Organization. 44 p.
68. RBMP (2008) The global malaria action plan for a malaria free world. Geneva, Switzerland: Roll Back Malaria Partnership (R.B.M.P), World Health Organization.
69. Harris I, Sharrock WW, Bain LM, Gray KA, Bobogare A, et al. (2010) A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. 9: 254.
70. Katsuragawa TH, Soares Gil LH, Tada MS, de Almeida e Silva A, Neves Costa JDA, et al. (2010) The dynamics of transmission and spatial distribution of malaria in riverside areas of Porto Velho, Rondonia, in the Amazon region of Brazil. PLoS One 5.
71. da Silva NS, da Silva-Nunes M, Malafronte RS, Menezes MJ, D'Arcadia RR, et al. (2010) Epidemiology and control of frontier malaria in Brazil: lessons from community-based studies in rural Amazonia. Trans Roy Soc Trop Med Hyg 104: 343–350.
72. Steenkeste N, Rogers WO, Okell L, Jeanne I, Incardona S, et al. (2010) Sub-microscopic malaria cases and mixed malaria infection in a remote area of high malaria endemicity in Rattanakiri province, Cambodia: implication for malaria elimination. 9: 108.
73. WHO (2010) Guidelines for the treatment of malaria. Second edition. Geneva, Switzerland: World Health Organization. 194 p.
74. MalERA (2011) A Research Agenda for Malaria Eradication: modeling. PLoS Med 8: e1000403.
75. MalERA (2011) A Research Agenda for Malaria Eradication: diagnoses and diagnostics. PLoS Med 8: 1–10.
76. Menard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR, et al. (2010) *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. P Natl Acad Sci USA 107: 5967–5971.
77. Mendes C, Dias F, Figueiredo J, Mora VG, Cano J, et al. (2011) Duffy negative antigen is no longer a barrier to *Plasmodium vivax* – molecular evidences from the African west coast (Angola and Equatorial Guinea). PLoS Neglect Trop D 5: e1192.
78. Wurtz N, Lekweiry KM, Bogreau H, Pradines B, Rogier C, et al. (2011) *Vivax* malaria in Mauritania includes infection of a Duffy-negative individual. Malaria J 10: 336.
79. Ryan JR, Stoute JA, Amon J, Dunton RF, Mtalib R, et al. (2006) Evidence for transmission of *Plasmodium vivax* among a Duffy antigen negative population in western Kenya. Am J Trop Med Hyg 75: 575–581.
80. Koita OA, Sangaré L, Sango HA, Dao S, Keita N, et al. (2012) Effect of Seasonality and Ecological Factors on the Prevalence of the Four Malaria Parasite Species in Northern Mali. J Trop Med 2012.
81. Pasvol G (2007) Eroding the resistance of Duffy negativity to invasion by *Plasmodium vivax*? T Roy Soc Trop Med H 101: 953–954.
82. Cavasini CE, Mattos LCd, Couto AA, Bonini-Domingos CR, Valencia SH, et al. (2007) *Plasmodium vivax* infection among Duffy antigen-negative individuals from the Brazilian Amazon region: an exception? T Roy Soc Trop Med H 101: 1042–1044.
83. Kasehagen LJ, Mueller I, Kiniboro B, Bockarie MJ, Reeder JC, et al. (2007) Reduced *Plasmodium vivax* Erythrocyte Infection in PNG Duffy-Negative Heterozygotes. PLoS One 2.
84. Sousa TN, Sanchez BAM, Ceravolo IP, Carvalho LH, Brito CFA (2007) Real-time multiplex allele-specific polymerase chain reaction for genotyping of the Duffy antigen, the *Plasmodium vivax* invasion receptor. Vox Sang 92: 373–380.
85. Cavasini CE, de Mattos LC, D'Almeida Couto AAR, D'Almeida Couto VSC, Gollino Y, et al. (2007) Duffy blood group gene polymorphisms among malaria vivax patients in four areas of the Brazilian Amazon region. Malaria J 6: 167.
86. Albuquerque SRL, Cavalcante FD, Sanguino EC, Tezza L, Chacon F, et al. (2010) FY polymorphisms and vivax malaria in inhabitants of Amazonas State, Brazil. Parasitol Res 106: 1049–1053.
87. Pampana E (1969) A textbook of malaria eradication. London: Oxford University Press.
88. Macdonald G, Göeckel GW (1964) The malaria parasite rate and interruption of transmission. Bull World Health Organ 31: 365–377.
89. Smith DL, Hay SI (2009) Endemicity response timelines for *Plasmodium falciparum* elimination. Malaria J 8: 87.
90. Smith T, Maire N, Ross A, Penny M, Chitnis N, et al. (2008) Towards a comprehensive simulation model of malaria epidemiology and control. Parasitology 135: 1507–1516.
91. Chitnis N, Schapira A, Smith T, Steketee R (2010) Comparing the effectiveness of malaria vector-control interventions through a mathematical model. Am J Trop Med Hyg 83: 230–240.
92. Smith T, Killeen GF, Maire N, Ross A, Molineaux L, et al. (2006) Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of *Plasmodium falciparum* malaria: Overview. Am J Trop Med Hyg 75: 1–10.
93. Ross A, Maire N, Sicuri E, Smith T, Conteh L (2011) Determinants of the cost-effectiveness of intermittent preventive treatment for malaria in infants and children. PLoS One 6: e18391.
94. Okell LC, Drakeley CJ, Bousema T, Whitty CJ, Ghani AC (2008) Modelling the impact of artemisinin combination therapy and long-acting treatments on malaria transmission intensity. PLoS Med 5: e226.
95. Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, et al. (2010) Reducing *Plasmodium falciparum* malaria transmission in Africa: a model-based evaluation of intervention strategies. PLoS Med 7: e1000324.
96. Ruwende C, Hill A (1998) Glucose-6-phosphate dehydrogenase deficiency and malaria. J Mol Med 76: 581–588.
97. Cappellini MD, Fiorelli G (2008) Glucose-6-phosphate dehydrogenase deficiency. Lancet 371: 64–74.
98. Douglas NM, Anstey NM, Angus BJ, Nosten F, Price RN (2010) Artemisinin combination therapy for *vivax* malaria. Lancet Infect Dis 10: 405–416.
99. Douglas NM, Nosten F, Ashley EA, Phaiphun L, van Vugt M, et al. (2011) *Plasmodium vivax* recurrence following falciparum and mixed species malaria: risk factors and effect of antimalarial kinetics. Clin Infec Dis 52: 612–620.
100. WHO (2011) World malaria report 2011. Geneva: World Health Organization. 246 p.
101. Rowe AK, Kachur SP, Yoon SS, Lynch M, Slutsker L, et al. (2009) Caution is required when using health facility-based data to evaluate the health impact of malaria control efforts in Africa. Malaria J 8: 209.
102. Gupta S, Gunter J, Novak R, Regens J (2009) Patterns of *Plasmodium vivax* and *Plasmodium falciparum* malaria underscore importance of data collection from private health care facilities in India. Malaria J 8: 227.
103. Hay SI, Gething PW, Snow RW (2010) India's invisible malaria burden. Lancet 376: 1716–1717.
104. Mueller I, Slutsker L, Tanner M (2011) Estimating the burden of malaria: The need for improved surveillance. PLoS Med 8: e1001144.

105. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, et al. (2010) Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. PLoS Med 7: e1000290.
106. Nosten F, McGready R, Simpson JA, Thwai KL, Balkan S, et al. (1999) Effects of *Plasmodium vivax* malaria in pregnancy. Lancet 354: 546–549.
107. Snounou G, White NJ (2004) The co-existence of *Plasmodium*: sidelights from *falciparum* and *vivax* malaria in Thailand. Trends Parasitol 20: 333–339.
108. Maitland K, Williams TN, Bennett S, Newbold CI, Peto TEA, et al. (1996) The interaction between *Plasmodium falciparum* and *P. vivax* in children on Espiritu Santo island, Vanuatu. T Roy Soc Trop Med H 90: 614–620.
109. Maitland K, Williams TN, Newbold CI (1997) *Plasmodium vivax* and *P. falciparum*: biological interactions and the possibility of cross-species immunity. Parasitol Today 13: 227–231.
110. The Malaria Atlas Project website (2012) Available: www.map.ox.ac.uk. Accessed 2012 Aug 12.

Fy^a/Fy^b antigen polymorphism in human erythrocyte Duffy antigen affects susceptibility to *Plasmodium vivax* malaria

Christopher L. King[a,b,1], John H. Adams[c], Jia Xianli[a], Brian T. Grimberg[a], Amy M. McHenry[c], Lior J. Greenberg[a], Asim Siddiqui[a], Rosalind E. Howes[d], Monica da Silva-Nunes[e], Marcelo U. Ferreira[e], and Peter A. Zimmerman[a]

[a]Center for Global Health and Diseases, Case Western Reserve University, Cleveland, OH 44106; [b]Veterans Affairs Medical Center, Cleveland, OH 44106; [c]College of Public Health, University of South Florida, Tampa, FL 33612; [d]Department of Zoology, University of Oxford, Oxford OX1 2JD, United Kingdom; and [e]Departamento de Parasitologia, Instituto de Ciencias Biomédicas, Universidade de São Paulo, 05508-900, São Paulo, Brazil

Edited by Louis H. Miller, National Institutes of Health, Rockville, MD, and approved November 4, 2011 (received for review June 18, 2011)

Plasmodium vivax **(Pv) is a major cause of human malaria and is increasing in public health importance compared with falciparum malaria. Pv is unique among human malarias in that invasion of erythrocytes is almost solely dependent on the red cell's surface receptor, known as the Duffy blood-group antigen (Fy). Fy is an important minor blood-group antigen that has two immunologically distinct alleles, referred to as Fy^a or Fy^b, resulting from a single-point mutation. This mutation occurs within the binding domain of the parasite's red cell invasion ligand. Whether this polymorphism affects susceptibility to clinical vivax malaria is unknown. Here we show that Fy^a, compared with Fy^b, significantly diminishes binding of Pv Duffy binding protein (PvDBP) at the erythrocyte surface, and is associated with a reduced risk of clinical Pv in humans. Erythrocytes expressing Fy^a had 41–50% lower binding compared with Fy^b cells and showed an increased ability of naturally occurring or artificially induced antibodies to block binding of PvDBP to their surface. Individuals with the Fy^{a+b-} phenotype demonstrated a 30–80% reduced risk of clinical vivax, but not falciparum malaria in a prospective cohort study in the Brazilian Amazon. The Fy^{a+b-} phenotype, predominant in Southeast Asian and many American populations, would confer a selective advantage against vivax malaria. Our results also suggest that efficacy of a PvDBP-based vaccine may differ among populations with different Fy phenotypes.**

Duffy binding protein | resistance

The parasite *Plasmodium vivax* plays a major role in the overall burden of malaria, causing severe morbidity and death (1). At least 80 million individuals worldwide suffer from vivax malaria; indeed, it is the most widely distributed malarial species outside of sub-Saharan Africa (2). Global efforts to eliminate malaria, largely based on reducing transmission, have been considerably less effective with *P. vivax* than with *Plasmodium falciparum* (3, 4), in part because of the former's efficient transmission in diverse ecological settings and its ability to reinitiate blood-stage infection from a dormant liver hypnozoite phase (5). Thus, success at *P. vivax* elimination may depend more on developing vaccines to prevent infection and suppress re-emergent blood-stage parasites.

P. falciparum demonstrates capacity to invade erythrocytes through multiple receptor pathways (6). In contrast, *P. vivax* red cell invasion appears to be primarily dependent on the Duffy antigen (Fy) (7). Although Duffy-independent *P. vivax* infection and disease can occur (8), alternative invasion pathways are not understood. As detailed understanding of host and parasite genetic polymorphisms and immune response inhibition of receptor-ligand interaction is of critical importance for vaccine development, here we have investigated the relevance of the $Fy^a \rightarrow Fy^b$ antigen polymorphism on susceptibility to clinical *P. vivax* malaria.

The gene that encodes the Duffy antigen has two major polymorphisms. A Asp→Gly amino acid substitution (codon 42) in the N-terminal region is associated with the Fy^b and Fy^a blood-group antigens, respectively (Fig. 1*A*). The second polymorphism T→C transition at nucleotide −33 in the Duffy gene promoter ablates Duffy expression on erythrocytes (ES; erythrocyte silent). The blood group and expression phenotypes associated with these polymorphisms have been well characterized; nomenclature and biological properties of Duffy have been summarized previously (8, 9) (Table S1).

Because of the critical role played by the Duffy antigen in *P. vivax* erythrocyte invasion, the corresponding parasite ligand, the Duffy binding protein (PvDBP), which is expressed at the parasite's cellular surface upon invasion, is a major vaccine candidate (10). The binding domain of PvDBP to Fy has been indentified in a 330-aa cysteine-rich region referred to as region II, designated PvDBPII (11, 12). Naturally acquired and artificially induced antibodies to PvDBPII inhibit parasite invasion in vitro (13) and protect against clinical malaria in children (14), supporting PvDBPII as a leading vaccine candidate. The critical residues of Fy, to which PvDBPII binds, map to N-terminal region amino acids 8–42 (Fig. 1*A*) (15, 16). Given studies on non-human primates indicating that Fy^b is the ancestral allele (17, 18), we hypothesized that Fy^a decreased the efficiency of PvDBPII binding, thereby reducing susceptibility to *P. vivax* malaria. Indeed, cross sectional association studies performed in the Brazilian Amazon region suggested that individuals expressing the Fy^b compared with Fy^a antigen may be more susceptible to *P. vivax* infection (19). Additionally, prior studies showed that an orthologous protein expressed by the simian malaria parasite, *Plasmodium knowlesi*, which infects human erythrocytes in a Duffy-dependent manner, preferentially bound Fy^b- compared with Fy^a-expressing erythrocytes, both in vivo and in vitro (20).

Results

Fy^a/Fy^b Polymorphism Affects Binding of PvDBPII to Erythrocytes. To examine whether human erythrocytes expressing Fy^a showed differential binding of recombinant PvDBPII compared with those expressing Fy^b, we screened blood samples from a group of healthy North American volunteers. Using a PCR assay, individual Fy genotypes were established (Table S1). Samples were assayed for degree of recombinant PvDBPII binding to red cells using flow cytometry. We found 40–50% lower binding of

Author contributions: C.L.K., J.H.A., B.T.G., M.U.F., and P.A.Z. designed research; C.L.K., J.X., B.T.G., A.M.M., L.J.G., A.S., R.E.H., M.d.S.-N., and M.U.F. performed research; C.L.K. contributed new reagents/analytic tools; C.L.K., M.U.F., and P.A.Z. analyzed data; and C.L.K., J.H.A., M.U.F., and P.A.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: cxk21@case.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1109621108/-/DCSupplemental.

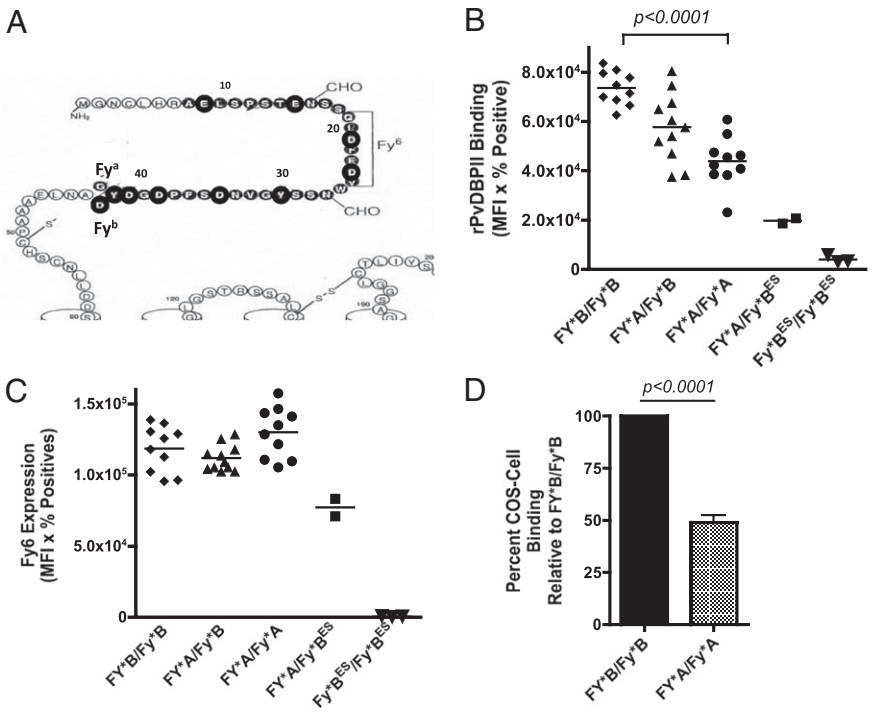


Fig. 1. Relationship of *FY* genotype on binding to PvDBP. (*A*) N-terminal binding domain (black residues) of PvDBP to Fy. Fy6 is a mAb and the corresponding epitope. $Fy^a \rightarrow Fy^b$ is the only polymorphism in the N-terminal region. Red highlights indicate negatively charged amino acids; the overall pI is 3.4 for the N-terminal region. In contrast, the binding domain of PvDBPII [consisting of 330 aa (32)] has a pI of 9.6. (*B*) Flow cytometry assessment of recombinant PvDBPII binding levels to erythrocytes from people differing by Duffy genotype. (*C*) Level of Duffy expression using mAb Fy6. (*D*) Relative binding of PvDBPII expressing COS-cells to erythrocytes of *FY*A/FY*A* ($n = 12$) vs. *FY*B/FY*B* ($n = 12$, $P < 0.0001$) blood donors; combines results of three separate experiments. For example, the mean number of rosettes per 30 high-powered field was 83 ± 11 for *FY*B/FY*B* erythrocytes compared with 46 ± 5 for *FY*A/FY*A* ($P = 0.007$) for one experiment (Fig. S1).

PvDBPII (0.2 μg per 10^6 red cells) (Fig. S2) to erythrocytes from *FY*A/FY*A* (i.e., phenotypically Fy^{a+b-}) compared with *FY*B/FY*B* (i.e., phenotypically Fy^{a-b+}) blood donors (Fig. 1*B*) ($P < 0.0001$). Erythrocytes from *FY*A/FY*B* donors displayed intermediate binding (Fy^{a+b+}). Our observed differences in PvDBPII binding could not be attributed to levels of Fy expression, which were similar for *FY*B/FY*B*, *FY*A/FY*A*, and *FY*A/FY*B* genotypes (Fig. 1*C*). *FY*A/FY*B*^{ES} cells expressed approximately half the levels of Fy compared with *FY*A/FY*A* cells; as expected, their binding was significantly reduced compared with cells from corresponding *FY*A* homozygotes. Duffy-negative erythrocytes (*FY*B*^{ES}/*FY*B*^{ES}) failed to express Fy and did not bind PvDBPII. We also performed erythrocyte rosetting assays, where PvDBPII was surface-expressed on COS cells. In these studies, cells from *FY*A/FY*A* donors bound COS cells at a 50% lower level compared with erythrocytes from *FY*B/FY*B* donors (Fig. 2*D* and Fig. S2).

To investigate mechanisms responsible for the differential binding, we looked at differences in electrostatic charge, as well as tyrosine (Tyr) sulfation between Fy^a and Fy^b (Fig. 1*A*). The less-efficient parasite binding of Fy^a may be a result of the charge neutrality of Gly42 (replacing Asp42), because the N-terminal region of Fy is negatively charged but PvDBPII is positively charged.

Prior studies have shown that the degree of sulfation of Tyr41 markedly affected binding of PvDBPII to Fy (21). Although Choe et al. observed no substantial difference in sulfation of their Fy^a vs. Fy^b constructs [60 codons Duffy amino terminus joined to human IgG1 Fc domain (20)], no data were provided to compare PvDBPII interaction with their constructs corresponding to Fy^a vs. Fy^b or native Fy^a vs. Fy^b antigens on the red cell surface. To make these comparisons, we treated erythrocytes with the enzyme arylsulfatase, which selectively and partially removes sulfate groups from Tyr (Fig. S2) (22). Interestingly, enzymatic treatment of *FY*A/FY*A* erythrocytes reduced PvDBPII binding by 42% (Fig. 2*A*) ($P = 0.0004$), but had no effect on PvDBPII binding to *FY*B/FY*B* erythrocytes (Fig. 2*C*). Enzymatic treatment did not affect quantitative expression of Fy on erythrocytes (Fig. 2 *C* and *D*, and Fig. S2*B*). Arylsulfatase concentrations that would result in complete removal of sulfate groups could not be used because it caused erythrocyte lysis. These results suggest that Fy^a may be more susceptible to loss of sulfate groups from tyrosines compared with Fy^b.

Binding Inhibitory Antibodies Show Greater Blocking of PvDBPII to Fy^a- Compared with Fy^b-Expressing Erythrocytes. Previous studies have shown that PvDBPII-specific antibodies inhibit *P. vivax* erythrocyte invasion in vitro (13) and correlate with protection against blood-stage infection in vivo (14). To determine whether antibodies that inhibit *P. vivax* invasion would bind differentially to Fy^a- vs. Fy^b-expressing cells, a binding-competition assay was performed using antibodies directed against PvDBPII. We found that similar concentrations of either naturally acquired (Fig. 3 *A* and *B*) or artificially induced inhibitory antibodies (Fig. 3 *C* and *D*) effected 200–300% greater inhibition of PvDBPII binding to erythrocytes from *FY*A/FY*A* compared with *FY*B/FY*B* donors. For these experiments, PvDBPII-specific antibodies were affinity-purified from human and rabbit sera. Therefore, the concentration of these PvDBPII antibodies would be higher than in circulating blood. Of note, antibody preparations were affinity-purified to enrich for antibodies directed to PvDBPII, and thus antibody concentrations used are unlikely to represent circulating antibody levels in individuals. Overall, these results suggest that binding inhibitory antibodies directed against PvDBPII may

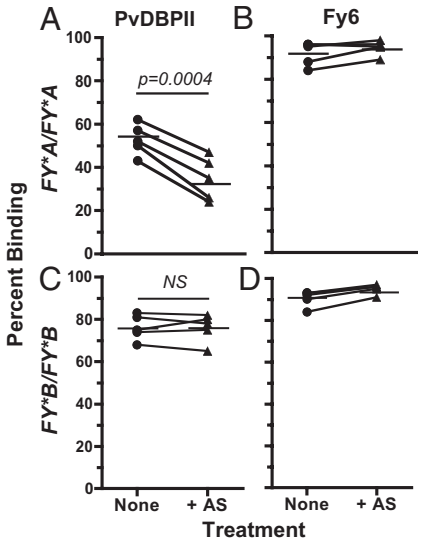


Fig. 2. Effect of arysulfatase treatment on binding of PvDBPII to erythrocytes from blood donors with different *FY* genotypes. (*A*) Treatment of erythrocytes from *FY*A/FY*A* donors selectively and partially removes sulfonate groups from tyrosine residues (Fig. S3) and reduced binding by PvDBPII. (*C*) Identical treatment of erythrocytes from *FY*B/FY*B* donors did not reduce PvDBPII binding. (*B* and *D*) Enzyme treatment of erythrocytes does not affect Fy expression on erythrocytes. Symbols (circle to triangle) paired by lines compare red blood cells from each individual before and after treatment with 500 million units of arylsulfatase.

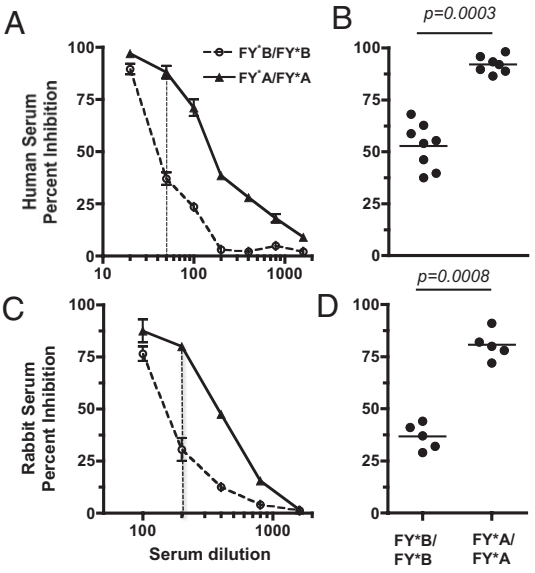


Fig. 3. Fy^a/Fy^b polymorphism affects binding inhibitory antibody blocking activity of recombinant PvDBPII to erythrocytes. (*A*) Inhibition of PvDBPII by different serum dilutions of pooled affinity-purified human binding inhibitory Abs of PvDBPII binding to erythrocytes from an *FY*A/FY*A* compared with an *FY*B/FY*B* blood donor. (*B*) Binding inhibitory Abs from human serum (1:50 dilution) show consistently greater blocking of PvDBPII binding to erythrocytes from *FY*A/FY*A* (n = 7) compared with *FY*B/FY*B* (n = 8) blood donors. (*C* and *D*) Affinity-purified rabbit serum generated against PvDBPII also shows a greater capacity to inhibit binding of PvDBPII to erythrocytes from *FY*A/FY*A* (n = 5) compared with the *FY*B/FY*B* (n = 5) individuals at a serum dilution of 1:200. Each datapoint (*A* and *C*) represents mean (± SD) percent inhibition for individuals with Fy^a vs. Fy^b. Each dot (*B* and *D*) represents means of duplicate or triplicate binding assays for one donor. Statistical comparison uses a Student t test.

inhibit more efficiently when *P. vivax* is accessing the red blood cell through contact with Fy^a compared with Fy^b.

Fy^a/Fy^b Polymorphism Is Associated with Reduced Risk to Clinical *P. vivax*. We then sought to determine if our in vitro findings associated with *FY* genotype correlated with in vivo susceptibility to uncomplicated clinical *P. vivax* malaria. For this study, we analyzed data from 400 individuals (5–74 y of age) living in a *P. vivax*-endemic region of the Brazilian Amazon studied previously (Table S2). All individuals were actively followed for clinical malaria over 14 mo, as determined by blood-smear microscopy and PCR-confirmation during a time when there was a surge in malaria infection (23). Overall, 124 cases of *P. vivax* and 66 cases of *P. falciparum* malaria were diagnosed, with annual incidence rates of 0.31 and 0.17, respectively. Mixed infection was found in 31 cases.

Prior analysis of the complete dataset suggested that duration of residence in the community and distance from the Iquiri River influenced risk for malaria (23). Here we observed a slight reduction in risk for *P. vivax* malaria, with longer duration of residence in the transmission area [risk ratio 0.94, 95% confidence interval (CI) 0.91–0.98, P = 0.01, negative binomial analysis], consistent with development of acquired immunity. Individuals who lived in the high-transmission area close to the river were at increased risk of *P. vivax* malaria (risk ratio 3.24, 95% CI 1.60–6.05, P = 0.001). Similar risk ratios for development of *P. falciparum* malaria were associated with time and location (risk ratio 2.89, 95% CI 1.72–5.67, P < 0.001). There was no significant association for *FY* genotype and prevalence of individuals living in high vs. low transmission areas, nor for *FY* genotype and duration of residence (Table S2) (Pearson's χ^2 test, P = 0.12).

Individuals with *FY*A/FY*B*^{ES} and *FY*A/FY*A* genotypes showed the lowest incidence of clinical *P. vivax* (Fig. 4). Negative binominal analysis adjusting for duration of local residence and transmission areas showed respectively, 80% and 29% reduced risk of clinical vivax malaria for individuals with *FY*A/FY*B*^{ES} and *FY*A/FY*A* compared with the *FY*A/FY*B* genotype (Table 1). In contrast, individuals with *FY*B/FY*B*^{ES} and *FY*B/FY*B* genotypes had 220–270% greater risk of vivax malaria compared with the *FY*A/FY*B* genotype (Fig. 3*B*). Interestingly, we observed no significant difference in the percent of subjects with antibody responses against PvDBPII in association with *FY* genotype (see Table 1 for sample sizes, *FY*A/FY*B*^{ES} 16.7%, *FY*A/FY*A* 11.9%, *FY*A/FY*B* 24.4%, *FY*B/FY*B*^{ES} 21.4%, *FY*B/FY*B* 11.3%). Results show a consistent decreased susceptibility

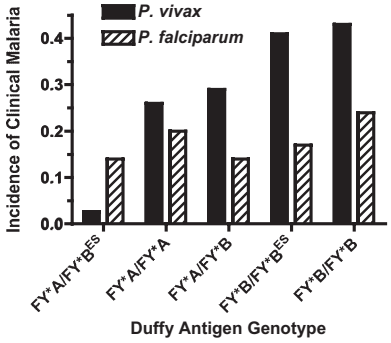


Fig. 4. Effect of *FY* genotype on the unadjusted annual incidence of clinical *P. vivax* (black bars) and *P. falciparum* malaria (striped bars). Incidence is expressed as mean number of clinical episodes per person-years of follow-up. Using a negative binomial analysis, the overall effect was highly significant, P < 0.001. The adjusted effect of specific genotype on *P. vivax* risk is shown in Table 1.

Table 1. Effect of Different *FY* genotype on risk of clinical vivax malaria

Genotype	*n*	Risk ratios (95% CI)	*P*
*FY*A/FY*B^{ES}*	35	0.204 (0.09–0.87)	**0.005**
*FY*A/FY*A*	52	0.715 (0.31–1.21)	0.06
*FY*A/FY*B*	140	Comparator	
*FY*B/FY*B^{ES}*	76	2.17 (0.91–4.77)	0.09
*FY*B/FY*B*	87	2.70 (1.36–5.49)	**0.002**

Rate ratios are adjusted for location and duration of residence in endemic area using *FY*A/FY*B* as the comparator groups based on a negative binomial analysis. Boldface indicates statistical significance.

to vivax malaria associated with Fy^{a} and increased susceptibility to vivax malaria with increased expression of Fy^{b}. There was no association between *FY* genotype and risk for *P. falciparum* in the multivariate analysis (overall risk ratio 1.08, 95% CI 0.87–2.38, $P = 0.42$). The analysis was repeated excluding individuals with mixed *P. vivax* and *P. falciparum* infections; *FY*A/FY*B^{ES}* and *FY*A/FY*A* had risk ratios of 0.197 (CI 0.07–0.98), $P = 0.03$ and 0.684 (CI 0.28–1.45), $P = 0.09$ compared with the *FY*A/FY*B* genotype. These results imply that differences in binding of PvDBPII to Fy^{a} compared with Fy^{b} translate to decreased *P. vivax* erythrocyte invasion efficiency, and lower parasitemias in Fy^{a+b-} compared with Fy^{a-b+} individuals. It should be noted that parasitemia levels were not specifically determined in this study.

Geographical Distribution of the Duffy Allelle Frequencies. The global distribution of the three major FY alleles, *FY*A*, *FY*B*, and *FY*B^{ES}*, are shown in Fig. 5. This map was derived from a suite of allele-frequency maps assembled from a database of Duffy blood-typing surveys from 1950 to 2010 (24) (Table S3). The *FY*A* allele appears to be advancing to fixation in many Asian populations. In contrast, *FY*B^{ES}* has followed a different pattern and has achieved fixation in Africa. The *FY*B* allele predominates in European populations. Admixture of *FY*A* and *FY*B* usually occurs in populations from relatively recent European migrations into North Africa and the Americas following the much early migration of populations from Asia.

Discussion

Our study demonstrates that PvDBPII binding is significantly lower to Fy^{a} than the ancestral Fy^{b} antigen (Fig. 1*B*). Concordantly, we found that Fy^{a} was associated with protection, but Fy^{b} was associated with increased infection and disease (Table 1). Although *P. vivax* parasitemia has been shown to correlate with risk of clinical vivax malaria (25), we did not evaluate the relationship of the *FY* genotype with parasitemia, as this feature of infection was not recorded in the Brazilian longitudinal study. Despite our recent findings that *P. vivax* is able to infect human red cells through a Duffy-independent mechanism (8), it is well known that Duffy-negativity (*FY*B^{ES}/FY*B^{ES}*) is responsible for high-level resistance to *P. vivax* blood-stage infection (7), suggesting that this parasite's invasion mechanism is heavily reliant on access to Fy. Given the burden of illness and death associated with vivax malaria (1), and proposals that *P. vivax* originated in Asia following lateral transfers of simian parasites from Old World monkeys (26), the emergence of a major vivax-resistance allele in African populations alone (*FY*B^{ES}*) is puzzling. Therefore, we hypothesize that *FY*A* has also been positively selected to reduce efficiency of *P. vivax* red cell invasion to improve human fitness to *P. vivax* malaria. The observation that *FY*A* has advanced to fixation in many Asian and American populations where vivax malaria is most highly endemic (Fig. 5 and *SI Methods*) supports this conclusion. Although historical record of vivax malaria in Africa is scanty, *FY*B^{ES}* predominates in most African ethnicities and holds vivax malaria at very low prevalence.

Prior studies investigating Fy and other red cell surface proteins as receptors for *P. knowlesi*/*P. vivax* red cell invasion have reported results that both support and vary with our present findings. Although evidence is limited, in vitro infection studies by Miller et al. suggested that *P. knowlesi* displayed lower efficiency in infecting human Fy^{a+b-} compared with Fy^{a-b+} red cells (27). Additional studies did not compare parasite invasion between Fy^{a+b-} and Fy^{a-b+} red cells directly (20, 28, 29) but did

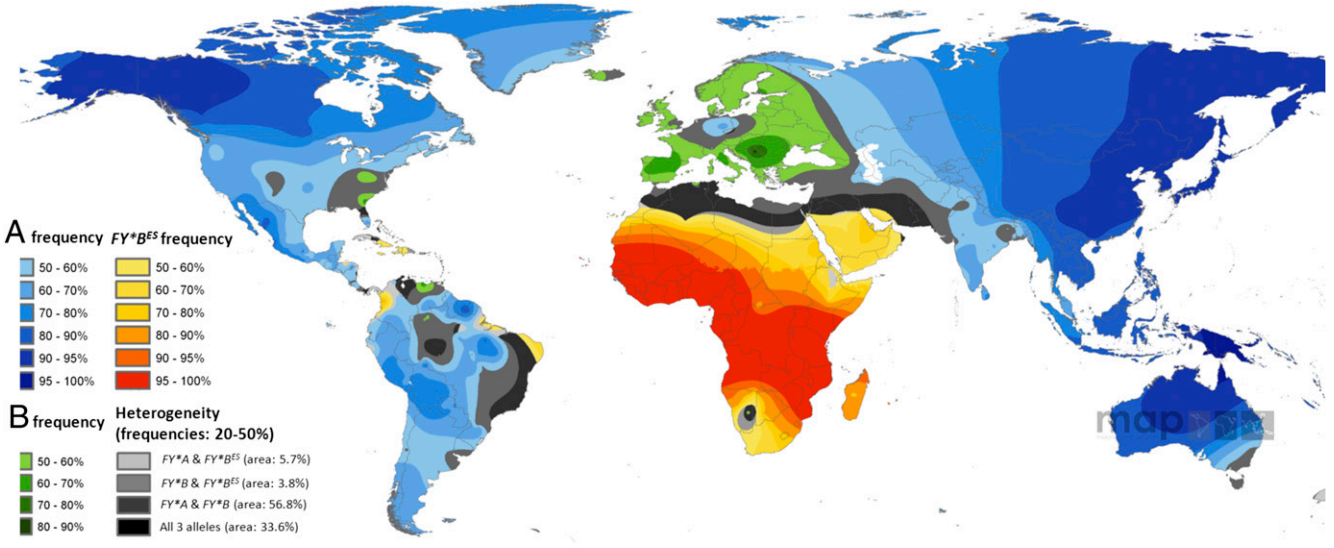


Fig. 5. Global frequencies of the *FY* alleles. Areas predominated by a single allele (frequency ≥ 50%) are represented by a color gradient (blue, *FY*A*; green, *FY*B*; red/yellow, *FY*B^{ES}*). Areas of allelic heterogeneity where no single allele predominates, but two or more alleles each have frequencies ≥ 20%, are shown in gray-scale: palest for heterogeneity between the silent *FY*B^{ES}* allele and either *FY*A* or *FY*B* (when coinherited, these do not generate new phenotypes), and darkest being co-occurrence of all three alleles (and correspondingly the greatest genotypic and phenotypic diversity). Overall percentage surface area of each class is listed in the legend. Refer to *SI Methods* for a methodological summary and further detail about the map surface.

compare binding of parasite proteins to human Fy^{a+b-} and Fy^{a-b+} red cells. Haynes et al. showed that a 135-kDa protein from *P. knowlesi* culture supernatants bound to Fy^{a-b+} erythrocytes much better than to Fy^{a+b-} cells (20). Similar studies found that a 135–140 "native" protein corresponding to the full-length Duffy binding protein from *P. vivax* culture supernatants showed similar binding to Fy^{a+b-} vs. Fy^{a-b+} red cells (30). A subsequent study using the same recombinant PvDBPII construct as the current study and a similar flow cytometry erythrocyte-binding assay showed no significant difference in PvDBPII binding to Fy^{a+b-} vs. Fy^{a-b+} erythrocytes (31). Experiments in this latter study used PvDBPII at concentrations 50-fold higher that those shown here that reveal preferential binding to Fy^{a-b+} vs. Fy^{a+b-} erythrocytes. High-concentration PvDBPII obscure differential binding to Fy^{a-b+} vs. Fy^{a+b-} erythrocytes (Fig. S1). Overall differences and similarities suggest that outcomes of in vitro studies may be sensitive to variation in parasite species/strains and their parasitemias, as well as differences in antigen polymorphism and concentration.

Results from our in vitro studies were consistent with our in vivo observations, suggesting a relationship between Fy^a compared with Fy^b expressed on erythrocytes with the amounts of PvDBPII erythrocyte binding in vitro and further susceptibility to clinical vivax malaria in vivo. Individuals who were Fy^{a+b-} (particularly the *FY*A/FY*B*^{ES} genotype) had the lowest binding of PvDBPII to their erythrocytes in vitro and the greatest resistance to vivax malaria. Those who were Fy^{a-b+} displayed the highest binding to PvDBPII and the greatest sensitivity to clinical vivax malaria. Interestingly, even though erythrocytes from *FY*B/FY*B*^{ES} donors express approximately half the amount of Duffy antigen (Table S1) compared with *FY*A/FY*A* and *FY*A/FY*B*, they were more susceptible to *P. vivax* malaria than the *FY*A/FY*A* and *FY*A/FY*B* genotypes (Table 1). The reasons for the relationships between these in vitro and in vivo results are unclear, and may be related to cohort sample size. However, our in vivo findings are consistent with a significant protective effect of the *FY*A* allele.

The mechanism by which Fy^a-expressing red cells show reduced binding to PvDBPII and reduced susceptibility to vivax malaria needs to be fully elucidated. As the N-terminal region of Fy is negatively charged and PvDBPII is positive, less efficient parasite binding to Fy^a may be because of the electrostatic neutrality of Gly42 (pI = 6) vs. negatively charged Asp42 (pI = 3.1). Additionally, because sulfation of Fy appears to influence PvDBPII binding, observed increased lability of PvDBPII binding to Fy^{a+b-} vs. Fy^{a-b+} following arylsulfatase treatment of donor cells suggests that *P. vivax* red cell invasion efficiency may be susceptible to differences in Fy sulfation.

In conclusion, our observations related to *P. vivax* interaction with Fy^a vs. Fy^b and subsequent development of naturally acquired immunity has important implications for vaccine trials using PvDBPII. In vitro studies demonstrate that both naturally acquired and artificially induced antibodies block erythrocyte binding of recombinant PvDBPII to Fy^a- better than Fy^b-expressing erythrocytes. Although we observed that the FY genotype is not associated with any significant differences in PvDBPII-specific antibody responses, our results suggest that naturally acquired immunity to *P. vivax* infection and disease may be more effective in populations where the *FY*A* allele predominates. Additionally, our findings indicate that it will be important to test PvDBPII-based vaccine in populations that carry combinations of both *FY*A* and *FY*B* alleles (Fig. 5).

Methods

Detailed information is provided in *SI Methods*.

Participants. Erythrocytes for the binding experiments were obtained from malaria uninfected volunteers at Case Western Reserve University. Malaria-exposed subjects were recruited as part of longitudinal cohort study performed in the Brazilian Amazon in 2004–2005, as previously described (23). All work with human samples was performed in accordance with approved Institutional Review Board protocols of the Veterans Affairs Medical Center, Cleveland, OH, University Hospitals, Cleveland, OH, and Ethical Review Board of the Institute of Biomedical Sciences of the University of São Paulo, Brazil.

Binding Experiments. Binding experiments with recombinant Pv Duffy Binding protein and Fy6 mAb that recognizes N-terminal region of Duffy antigen used fresh human erythrocytes from subjects previously genotyped for Duffy and were performed as previously described (13). Binding inhibition levels by anti-PvDBPII were assessed as previously described (13).

Statistical Analysis. A negative binomial regression analysis (SAS version 9.2; SAS Institute) was used because of the overdispersion of the data. Sample size was 400, input risk variables were Duffy genotype, location, and duration of residence in the study area with the primary output being risk of clinical *P. vivax*. There was little interaction between Duffy genotype and location and duration of residence (Table S2).

Mapping Duffy Genotypes. The probability distribution based on a Bayesian model is summarized as a single statistic: in this case, the median value, as this corresponds best to the input dataset, as previously described (24). Median values of the predictions were generated for each allele frequency at a 10 × 10-km resolution on a global grid with GIS software (ArcMap 9.3; ESRI).

ACKNOWLEDGMENTS. We thank Chetan Chitnis for providing the nDARCIg; Kevin Moore for providing the monoclonal antibody that recognizes sulfonated tyrosine; Jennifer Cole-Tobian for help with statistical analysis; Christine J. Julian for help in preparing the manuscript; and Menachem Shoham and Martin Stone for advice in experimental design and insights into the interaction between Duffy antigen and *Plasmodium vivax* Duffy binding protein. This research was supported in part by Veterans Affairs Research Service, and US Public Health Service Grant R01 AI064478; field work was supported by the Brazilian agencies Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Grants 2003/09719-6 and 2005/51988-0) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant 470067/2004-7); M.d.S.-N. received a PhD scholarship from FAPESP and M.U.F. is supported by a research scholarship from CNPq.

1. Price RN, Douglas NM, Anstey NM (2009) New developments in *Plasmodium vivax* malaria: Severe disease and the rise of chloroquine resistance. *Curr Opin Infect Dis* 22: 430–435.
2. Guerra CA, et al. (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl Trop Dis* 4:e774.
3. Baird JK (2007) Neglect of *Plasmodium vivax* malaria. *Trends Parasitol* 23:533–539.
4. Tatem AJ, et al. (2010) Ranking of elimination feasibility between malaria-endemic countries. *Lancet* 376:1579–1591.
5. Krotoski WA (1989) The hypnozoite and malarial relapse. *Prog Clin Parasitol* 1: 1–19.
6. Persson KE, et al. (2008) Variation in use of erythrocyte invasion pathways by *Plasmodium falciparum* mediates evasion of human inhibitory antibodies. *J Clin Invest* 118:342–351.
7. Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med* 295: 302–304.
8. Ménard D, et al. (2010) *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc Natl Acad Sci USA* 107:5967–5971.
9. Murphy PM, et al. (2000) International union of pharmacology. XXII. Nomenclature for chemokine receptors. *Pharmacol Rev* 52:145–176.
10. Arévalo-Herrera M, Chitnis C, Herrera S (2010) Current status of *Plasmodium vivax* vaccine. *Hum Vaccin* 6:124–132.
11. Chitnis CE, Miller LH (1994) Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. *J Exp Med* 180:497–506.
12. Adams JH, et al. (1992) A family of erythrocyte binding proteins of malaria parasites. *Proc Natl Acad Sci USA* 89:7085–7089.
13. Grimberg BT, et al. (2007) Antibodies directed *Plasmodium vivax* Duffy binding protein region inhibit *P. vivax* invasion of human erythrocytes. *PLoS Med* 4:e337.
14. King CL, et al. (2008) Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage *Plasmodium vivax* infection. *Proc Natl Acad Sci USA* 105:8363–8368.
15. Pogo AO, Chaudhuri A (2000) The Duffy protein: A malarial and chemokine receptor. *Semin Hematol* 37:122–129.
16. Horuk R, Wang ZX, Peiper SC, Hesselgesser J (1994) Identification and characterization of a promiscuous chemokine-binding protein in a human erythroleukemic cell line. *J Biol Chem* 269:17730–17733.

17. Li J, Iwamoto S, Sugimoto N, Okuda H, Kajii E (1997) Dinucleotide repeat in the 3′ flanking region provides a clue to the molecular evolution of the Duffy gene. *Hum Genet* 99:573–577.
18. Tournamille C, et al. (2004) Sequence, evolution and ligand binding properties of mammalian Duffy antigen/receptor for chemokines. *Immunogenetics* 55:682–694.
19. Cavasini CE, et al. (2007) Duffy blood group gene polymorphisms among malaria vivax patients in four areas of the Brazilian Amazon region. *Malar J* 6:167.
20. Haynes JD, et al. (1988) Receptor-like specificity of a *Plasmodium knowlesi* malarial protein that binds to Duffy antigen ligands on erythrocytes. *J Exp Med* 167: 1873–1881.
21. Choe H, et al. (2005) Sulphated tyrosines mediate association of chemokines and *Plasmodium vivax* Duffy binding protein with the Duffy antigen/receptor for chemokines (DARC). *Mol Microbiol* 55:1413–1422.
22. Wilkins PP, Moore KL, McEver RP, Cummings RD (1995) Tyrosine sulfation of P-selectin glycoprotein ligand-1 is required for high affinity binding to P-selectin. *J Biol Chem* 270:22677–22680.
23. da Silva-Nunes M, et al. (2008) Malaria on the Amazonian frontier: Transmission dynamics, risk factors, spatial distribution, and prospects for control. *Am J Trop Med Hyg* 79:624–635.
24. Howes RE, et al. (2011) The global distribution of the Duffy blood group. *Nat Commun* 2:266.
25. Müller I, et al. (2009) Three different *Plasmodium* species show similar patterns of clinical tolerance of malaria infection. *Malar J* 8:158.
26. Carter R (2003) Speculations on the origins of *Plasmodium vivax* malaria. *Trends Parasitol* 19:214–219.
27. Miller LH, Mason SJ, Dvorak JA, McGinniss MH, Rothman IK (1975) Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* 189:561–563.
28. Mason SJ, Miller LH, Shiroishi T, Dvorak JA, McGinniss MH (1977) The Duffy blood group determinants: Their role in the susceptibility of human and animal erythrocytes to *Plasmodium knowlesi* malaria. *Br J Haematol* 36:327–335.
29. Barnwell JW, Nichols ME, Rubinstein P (1989) In vitro evaluation of the role of the Duffy blood group in erythrocyte invasion by *Plasmodium vivax*. *J Exp Med* 169: 1795–1802.
30. Wertheimer SP, Barnwell JW (1989) *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: Identification of a parasite receptor-like protein. *Exp Parasitol* 69:340–350.
31. Tran TM, et al. (2005) Detection of a *Plasmodium vivax* erythrocyte binding protein by flow cytometry. *Cytometry A* 63:59–66.
32. Chitnis CE, Chaudhuri A, Horuk R, Pogo AO, Miller LH (1996) The domain on the Duffy blood group antigen for binding *Plasmodium vivax* and *P. knowlesi* malarial parasites to erythrocytes. *J Exp Med* 184:1531–1536.

Supporting Information

King et al. 10.1073/pnas.1109621108

SI Methods

Erythrocyte Samples. Healthy volunteers from the local Cleveland area donated finger-prick samples of blood and were screened for their Duffy genotypes using PCR and a Bioplex fluorescent conjugated ligase chain detection (1). Institutional Review Boards at the Veterans Affairs Medical Center, Cleveland, OH and University Hospitals, Cleveland, OH approved the study. Ethical Review Board of the Institute of Biomedical Sciences of the University of São Paulo, Brazil gave approval for the studies performed in Brazil. Written informed consent was obtained from each adult participant and from the parent or legal guardian of every minor.

Erythrocyte Binding Assays and Duffy Expression. PvDBPII (the binding domain of *Plasmodium vivax* Duffy binding protein to Fy, indentified in a 330-aa cysteine-rich region referred to as region II) binding to RBC was assessed using a flow cytometry based-assay (2, 3). Finger-prick blood (~100 μL) was collected directly into 0.5 mL of PBS and washed twice in PBS using a microcentrifuge at 8,000 × *g* for 1 min. Following the final spin, 10 μL of packed RBC were removed from the pellet and resuspended in 40 μL of PBS (1:5 dilution). One microliter of this RBC was further diluted into 100 μL PBS plus 1% BSA to yield a total of ~10^6 erythrocytes. Recombinant PvDBPII was added (0.2 μg total) to the 100-μL erythrocyte suspension and incubated for 2 h at room temperature or overnight at 4 °C. Production of recombinant PvDBPII and the different variants was performed as previously described (2, 4). Following binding to PvDBPII, each sample was washed three times with PBS/1% BSA and incubated (1 h in the dark at 4 °C) with rabbit anti-PvDBPII (1:8,000), washed, then followed by the secondary antibody phycoerythrin-conjugated goat–anti-rabbit antibody (Invitrogen). The amount of antibody was titered to obtain optimal signal with each lot of antibody (ranging from 1:5–1:50 dilution). LSRII-based (Becton-Dickinson) flow cytometry evaluated 50,000 erythrocytes (using a Blue 488 laser).

PvDBPII binding to RBC varied with the amount of recombinant protein used (Fig. S1). Maximal binding of recombinant PvDBPII (rPvDBPII) occurred at 1 μg/10^6 RBC for erythrocytes with the *FY*B*/*FY*B* genotype, whereas maximal binding for *FY*A*/*FY*A* occurred at 5–10 μg/10^6 RBC. The greatest difference in PvDBPII binding between *FY*B*/*FY*B* and *FY*A*/*FY*A* occurred between concentrations of 0.2 and 0.5 μg/10^6 RBC. Thus, 0.2 μg/10^6 was routinely used to examine differences in binding.

To evaluate Duffy expression, erythrocytes were prepared as described above, except rather than adding PvDBPII, the mAb Fy6 (kindly provided by John Barnwell (Atlanta, GA); recognizes Fy N-terminal amino acids 19–25) (Fig. 1*A*) was added at a final dilution of 1:25,000 for 15 min at 37 °C. Cells were then washed, a 1:6 dilution of phycoerythrin-conjugated goat anti-mouse Abs (Sigma) was added and then cells were evaluated by flow cytometry.

Erythrocyte binding to PvDBPII expressed on COS cells was also assessed, as recently described (5). COS7 (green monkey kidney epithelial) cells were maintained in DMEM (Sigma) containing 10% fetal bovine sera (FBS). Only cells between the passage numbers of 5 and 20 were used for binding assays. COS7 cells were plated in 24-well plates at a density of 35,000 cells per well and were transiently transfected with endotoxin-free pEGFP-PvDBPII DNA using Lipofectamine or Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Forty-two hours posttransfection, the transfected COS7 cells were incubated with Fy antigen-positive human erythrocytes for 2 h (2.5×10^7 cells per well, previously washed three times with incomplete DMEM). Wells were washed three times with PBS to remove nonadherent erythrocytes and binding was scored by counting the number of rosettes per 30 fields of view at 200× magnification. EGFP fluorescence was used to confirm surface expression of the PvDBPII construct.

Binding Inhibition Using Antibodies to Recombinant PvDBPII. To evaluate the ligand–receptor inhibitory ability, affinity-purified human or rabbit anti-PvDBPII Abs were incubated with rPvDBPII at the specified dilutions (1 h at 37 °C) before combining with donor erythrocytes. Percent binding was evaluated by assessing the percentage of erythrocytes with bound rPvDBPII following exposure to test serum, divided by the percentage of erythrocytes with bound rPvDBPII following exposure to prebled rabbit serum and multiplied by 100.

Removal of Sulfated Tyrosine Residues with Arylsulfatase. Erythrocytes were prepared as described above. An additional binding assay was performed, using a chimeric molecule consisting of the 60-aa N-terminal region Duffy antigen linked to the human FcgR binding region of human IgG1 heavy chain (nDARCIg), as previously described (2, 6), in which 0.1 μg/mL of the protein was coated on Immulon 4 plates. Both RBC and chimera-treated plates with the Fyb construct were incubated with 100–1,000 mu (million units) of arylsulfatase (Sigma) overnight at 4 °C and then washed with PBS/1%PBS or ELISA buffer before addition of rPvDBPII. Maximal reduction of binding occurred in a range of 500–1,000 mu/mL (Fig. S2*A*) that corresponded to loss of recognition of mAb, called PSG2 (Fig. S3*B*) that binds with high affinity and exquisite specificity to sulfotyrosine residues on proteins (7) (kindly provided by Kevin L. Moore (Oklahoma City, OK), University of Oklahoma Health Sciences Center) (Fig. S3*B*). Human RBC (10^6 in 100 μL PBS/1% BSA) were treated with maximum amount of arylsulfatase 500 mu/mL overnight at 4 °C to achieve maximal activity. Arylsulfatase concentrations >500 mu/mL resulted in partial erythrocyte lysis and, thus, could not be reliably used. Following treatment with arylsulfatase, RBC were washed twice with PBS/1%BSA and a flow-based binding assay was performed as described above.

Study Area, Design, and Population. The study population, design and incidence of symptomatic *P. vivax* and *Plasmodium falciparum* have been previously described in detail (8, 9). Briefly, the study population was from a settlement area in Acre, northwestern Brazil, with an equatorial humid climate. This area began settlement around 1982 with immigrants from Southeast and South Brazil. A population-based open cohort study was started in March 2004 and ended in May 2005, where active case detection (5 d/wk) as well as passive surveillance was performed for malaria. A total of 123 households with 509 inhabitants were studied over this period. Blood samples were obtained on 425 subjects between ages of 5 and 90 y, of which 400 were successfully genotyped and phenotyped for Duffy positivity and Fy^a and Fy^b using the molecular methods described above and microtyping method (DiaMed-ID Microtyping System, DiaMed AG), respectively. The diagnosis of clinical malaria was based on blood smear with any parasite present and confirmed by a semiquantitative PCR analysis, accompanied by an oral temperature of >38 °C on examination or reporting of fever within the past 2 d accompanied by a headache or respiratory distress, myalgias, and other symptoms consistent with malaria and absence of

other obvious causes for the symptoms. All Giemsa-stained thick blood smears had at least 100 fields examined for malaria parasites under 700× magnification by two experienced microscopists. A reference microscopist in Rio Branco, the capital of Acre, reviewed all positive slides and 10% of the negative ones. Levels of parasitemia were not recorded. Mixed *P. vivax* and *P. falciparum* infections by blood smear occurred in 12% of clinical cases. In this situation the individual was considered to have *P. vivax*-attributed illness if the *P. vivax* parasite density was higher by at least a factor of 2 compared with *P. falciparum*. A minimal interval of 28 d between two or more consecutive sample examinations was required to count the last positive slide as a new malaria episode. When different species were detected in samples obtained less than 28 d apart, the subject was considered to have a single episode of mixed-species infection. Factors that could affect malaria exposure were recorded, such as bednet use, location, and age; however, because all were immigrants and the majority from malaria-free areas (60.1%), age did not correlate with malaria risk during follow-up (9). Cumulative exposure was therefore estimated as length of residence in malaria-endemic areas.

Statistical Analysis. Binding levels were compared using Student *t* test in GraphPad Prizm. The risk of malaria infection was calculated by using a negative binominal regression (SAS Inc.), with *FY* genotypes as the independent variable, and with outcomes being mean annual incidence of clinical *P. vivax* or *P. falciparum* infections per person-years of follow-up, adjusted by location and number of years of residence in the area. Age was found not to be associated with risk of malaria (9) because of the low transmission rates and immigration of most subjects from areas in Brazil not endemic for malaria.

Mapping the Duffy Allele Frequencies. Fig. 5 was derived from the suite of allele-frequency maps described in detail by Howes et al. (10) and in Table S3. A total of 821 spatially unique surveys conformed to the inclusion criteria of geographic specificity and community representation. This geopositioned evidence base informed a bespoke multilocus Bayesian geostatistical model adapted for both serological and molecular diagnostic types.

The output of the Bayesian model is a probability distribution of plausible allele frequencies based on the input dataset at each prediction location. To present this output as a single summary map, the probability distribution is summarized as a single statistic, in this case the median value, as this corresponds best to the input dataset. Median values of the predictions were generated for each allele frequency at a 10 × 10-km resolution on a global grid. GIS software (ArcMap 9.3; ESRI) was used to selectively represent areas of allelic predominance (displayed in color) and allelic heterogeneity (shown in grayscale) in a single synthesized map (Fig. 5).

Estimating Population Numbers of Different Duffy Phenotypes. To generate fully additive population counts for the various Duffy phenotypes, the Bayesian outputs of the model previously described (10) were summarized as mean predicted values. Allele frequencies were assumed to be in Hardy–Weinberg equilibrium and genotype frequencies calculated accordingly to generate the six allelic combinations listed below, by multiplying the 10 × 10-km gridded map surfaces in GIS software (ArcMap 9.3; ESRI):

$1 = FY^*A^2 + FY^*B^2 + FY^*B^{(ES)2} + 2(FY^*A \times FY^*B) + 2(FY^*A \times FY^*B^{ES}) + 2(FY^*B \times FY^*B^{ES})$.

Phenotype surfaces were then made by adding surfaces where required: $Fy^{(a+b-)}$ are *FY*A/FY*A or FY*A/FY*B*^{ES}; $Fy^{(a-b+)}$ are *FY*B/FY*B or FY*B/FY*B*^{ES}; $Fy^{(a+b+)}$ are *FY*A/FY*B*; and $Fy^{(a-b-)}$ are *FY*B*^{ES}*/FY*B*^{ES}.

To derive population estimates for each phenotype, the Duffy surfaces were transposed to 1 × 1-km grids, for compatibility with the high-resolution population surface used. As previously described (11), the dataset was the β-version of the Global Rural Urban Mapping Project gridded population database (http://sedac.ciesin.columbia.edu/gpw/), which was projected from 2000 to 2010 using separate urban and rural growth rates estimated by the 2007 United Nations World Urbanization Prospects (http://esa.un.org/unup/). The national population totals were then adjusted to match those reported in the 2008 United Nations World Population Prospects report (http://esa.un.org/unpp/). This population surface was overlaid on each of the Duffy phenotype frequency maps, as previously described (11, 12), to derive gridded population counts for each phenotype, then summarized by country to estimate the relative proportions of each phenotype (Table S3).

1. Ménard D, et al. (2010) *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc Natl Acad Sci USA* 107:5967–5971.
2. Grimberg BT, et al. (2007) *Plasmodium vivax* invasion of human erythrocytes inhibited by antibodies directed against the Duffy binding protein. *PLoS Med* 4:e337.
3. Tran TM, et al. (2005) Detection of a *Plasmodium vivax* erythrocyte binding protein by flow cytometry. *Cytometry A* 63:59–66.
4. Singh S, et al. (2001) Biochemical, biophysical, and functional characterization of bacterially expressed and refolded receptor binding domain of *Plasmodium vivax* duffy-binding protein. *J Biol Chem* 276:17111–17116.
5. McHenry AM, Barnwell JW, Adams JH (2010) *Plasmodium vivax* DBP binding to *Aotus nancymaae* erythrocytes is Duffy antigen dependent. *J Parasitol* 96:225–227.
6. Choe H, et al. (2005) Sulphated tyrosines mediate association of chemokines and *Plasmodium vivax* Duffy binding protein with the Duffy antigen/receptor for chemokines (DARC). *Mol Microbiol* 55:1413–1422.
7. Hoffhines AJ, Damoc E, Bridges KG, Leary JA, Moore KL (2006) Detection and purification of tyrosine-sulfated proteins using a novel anti-sulfotyrosine monoclonal antibody. *J Biol Chem* 281:37877–37887.
8. da Silva NS, et al. (2010) Epidemiology and control of frontier malaria in Brazil: Lessons from community-based studies in rural Amazonia. *Trans R Soc Trop Med Hyg* 104:343–350.
9. da Silva-Nunes M, et al. (2008) Malaria on the Amazonian frontier: Transmission dynamics, risk factors, spatial distribution, and prospects for control. *Am J Trop Med Hyg* 79:624–635.
10. Howes RE, et al. (2011) The global distribution of the Duffy blood group. *Nat Commun* 2:266.
11. Hay SI, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med* 6:e1000048.
12. Guerra CA, et al. (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl Trop Dis* 4:e774.

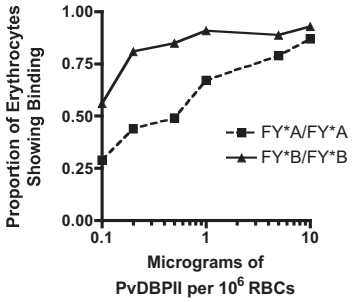


Fig. S1. The proportion of erythrocytes that bound recombinant PvDBPII varies with protein concentration and Duffy genotype. This is one of five experiments showing the impact of PvDBPII concentration on erythrocyte binding from individuals with *FY*A/FY*A* or *FY*B/FY*B* genotypes. Shown are mean percent binding of duplicates assays for each PvDBPII concentration for two individuals, one with the *FY*A/FY*A* genotype and the other with *FY*B/FY*B* genotype. Variation in percent binding in duplicate cultures is less than 10%.

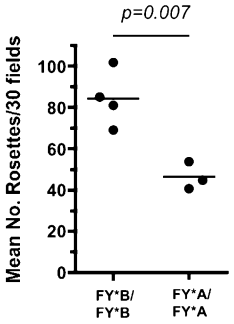


Fig. S2. The level of PvDBPII binding to erythrocytes is influenced by *FY* genotype as shown in an erythrocyte binding assay in which COS7 cells were transfected with plasmid expressing the gene encoding PvDBPII and then incubated with erythrocytes expressing either *FY*A/FY*A* or *FY*B/FY*B*, resulting in formation of rosettes as described in the *SI Methods*. Assays were performed in triplicate and each dot represents mean of three assays from one individual. There were seven different individuals examined, *FY*B/FY*B* (n = 4) and *FY*A/FY*A* (n = 3). The assays were performed twice with similar results with the combined results shown in Fig. 1*D*.

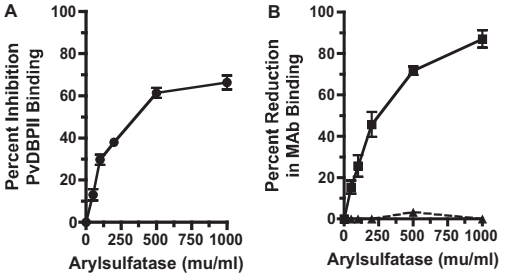


Fig. S3. Treatment of N-terminal region of Fy with arylsulfatase removes sulfonated tyrosines and reduces PvDBPII binding. (*A*) Treatment of a chimeric molecule nDARCIg [60-aa N-terminal region of the Fy antigen fused to the FcgR1 binding region of the IgG1 heavy chain (2, 6)] with varying concentrations of arylsulfatase reduces binding with PvDBPII. (*B*) Treatment with arylsulfatase results in reduced binding of mAb PGS2 that specifically recognizes sulfated tyrosines (■), demonstrating the effectiveness of the arylsulfatase treatment, but did not affect binding by mAb Fy6 (▲) that recognizes the N-terminal region of the Duffy antigen/receptor for chemokines (Fig. 1*A*). Percent reduction in mAb binding was determined by optical density (OD), with the specific mAbs following treatment with different concentrations arylsulfatase divided by OD in the absence of arylsulfatase.

Table S1. Duffy blood group nomenclature relevant to the current study

			Phenotype	
Allele	Antigen	Genotype	Serological	Expression*
*FY*A*	Fy^a	*FY*A/FY*A*	$Fy^{a+/b-}$	2× Fy^a, 0× Fy^b
		*FY*A/FY*B^{ES}*	$Fy^{a+/b-}$	1× Fy^a, 0× Fy^b
*FY*B*	Fy^b	*FY*B/FY*B*	$Fy^{a-/b+}$	0× Fy^a, 2× Fy^b
		*FY*B/FY*B^{ES}*	$Fy^{a-/b+}$	0× Fy^a, 1× Fy^b
		*FY*A/FY*B*	$Fy^{a+/b+}$	1× Fy^a, 1× Fy^b
*FY*B^{ES}*	No antigen	*FY*B^{ES}/FY*B^{ES}*	$Fy^{a-/b-}$	0× Fy^a, 0× Fy^b

Alleles correspond with antigens. Genotypes (allele combinations) correspond with phenotypes. An alternative gene name for the Duffy blood group is the Duffy antigen/receptor for chemokines (DARC), which is consistent with the blood-group mutations database at the National Center for Biotechnology. The official nomenclature is yet to be determined. Genotypes are designated using standard nomenclature whereby genotype is signified by uppercase letters and italicized fonts for both alleles, e.g., *FY*B/FY*B*. Lowercase designations indicate phenotypes. For purposes of this study, ES (erythrocyte silent) pertains only to the B allele.

*Expression phenotypes based on composite flow cytometry with Fy^b mAb and chemokine binding based on results in the article and the references provided.

Table S2. Study population in Brazil

Genotype	*FY*A/FY*B^{ES}*	*FY*A/FY*A*	*FY*A/FY*B*	*FY*B/FY*B^{ES}*	*FY*B/FY*B*
n = 400 (%)	35 (8.8)	52 (13.0)	140 (35)	76 (19.0)	87 (21.8)
Median age (range)	28 (5–52)	23 (5–74)	25 (5–71)	21 (4–64)	25 (5–67)
Sex (M/F)	13/21	21/31	78/68	39/33	51/34
Median years in area (range)	11 (2–22)	8 (0.1–22)	8 (0.1–24)	10 (0.2–23)	10 (0.2–24)
High-transmission area	42.8[†]	59.6	42.1	50	50.5
Intermediate-transmission area	40	23.1	28.6	27.6	20.7
Low-transmission area	17.2	19.2	30	22.4	28.7

[†]Percentage of individuals living in the different transmission zones as a function of distance from Iquiri River, as determined previously (9).

Table S3. Proportions of population numbers for Duffy phenotypes in *P. vivax*-endemic countries

Region/country	$Fy^{(a+b+)}$ (%)	$Fy^{(a+b-)}$ (%)	$Fy^{(a-b+)}$ (%)	$Fy^{(a-b-)}$ (%)	2010 United Nations estimated country population (in thousands)
Africa					
Angola	0.08	2.85	3.34	93.73	18,993
Benin	0.00	0.43	2.10	97.46	9,212
Botswana	2.16	12.22	22.23	63.39	1,978
Burkina Faso	0.00	0.22	0.23	99.55	16,287
Burundi	0.02	1.99	1.75	96.24	8,519
Cameroon	0.03	1.12	3.56	95.29	19,958
Central African Republic	0.22	3.05	7.74	89.00	4,506
Chad	0.61	3.86	17.64	77.89	11,506
Comoros	0.15	9.99	2.64	87.22	691
Congo	0.02	1.48	2.27	96.23	3,759
Congo, DR	0.09	2.86	4.26	92.80	67,827
Côte d'Ivoire	0.00	0.30	0.19	99.50	21,571
Djibouti	12.06	40.13	23.22	24.59	879
Equatorial Guinea	0.01	0.83	1.58	97.58	693
Eritrea	5.61	16.86	37.99	39.54	5,224
Ethiopia	4.79	14.67	35.29	45.24	84,976
Gabon	0.01	1.02	2.29	96.68	1,501
Gambia	0.00	0.17	0.10	99.73	1,751
Ghana	0.00	0.29	0.38	99.33	24,333
Guinea	0.00	0.28	0.33	99.39	10,324
Guinea-Bissau	0.00	0.21	0.20	99.59	1,647
Kenya	0.03	2.01	1.90	96.06	40,863
Liberia	0.00	0.38	0.41	99.20	4,102
Madagascar	1.36	22.84	8.30	67.51	20,146
Malawi	0.00	0.90	0.51	98.58	15,692
Mali	0.01	0.67	0.60	98.72	13,323
Mauritania	0.06	3.33	1.67	94.94	3,366
Mozambique	0.08	2.19	4.19	93.54	23,406
Namibia	4.66	15.45	23.28	56.61	2,212
Niger	0.13	2.76	5.47	91.64	15,891
Nigeria	0.03	1.10	3.98	94.89	158,259
Rwanda	0.03	2.32	2.33	95.32	10,277
São Tomé and Príncipe	0.01	0.85	3.09	96.05	165
Senegal	0.00	0.29	0.32	99.38	12,861
Sierra Leone	0.00	0.36	0.46	99.18	5,836
Somalia	2.28	10.11	18.68	68.93	9,359
South Africa	3.25	10.89	30.12	55.75	50,492
Sudan	3.32	11.78	33.75	51.16	43,192
Swaziland	0.70	5.40	20.58	73.31	1,202
Togo	0.00	0.35	1.11	98.54	6,780
Uganda	0.08	3.16	4.20	92.56	33,796
United Republic of Tanzania	0.00	1.29	0.57	98.14	45,040
Zambia	0.09	2.65	4.73	92.53	13,257
Zimbabwe	0.13	2.77	7.07	90.03	12,644
Americas					
Argentina	35.99	42.32	20.23	1.46	40,666
Belize	7.34	76.26	4.59	11.81	313
Bolivia	25.44	68.59	5.38	0.60	10,031
Brazil	21.77	36.54	28.51	13.18	195,423
Colombia	8.73	58.28	9.52	23.48	46,300
Costa Rica	36.17	43.86	18.53	1.45	4,640
Ecuador	8.00	70.63	6.32	15.05	13,775
El Salvador	13.70	69.46	8.17	8.66	6,194
French Guiana	14.26	74.64	5.92	5.18	231
Guatemala	16.26	67.11	9.35	7.28	14,377
Guyana	33.29	54.49	11.37	0.86	761
Honduras	8.79	70.44	6.45	14.31	7,616
Mexico	26.10	60.37	10.86	2.67	110,645
Nicaragua	18.99	61.45	12.12	7.44	5,822
Panama	18.30	50.45	18.62	12.63	3,508
Paraguay	28.06	59.54	10.57	1.82	6,460

Table S3. Cont.

Region/country	$Fy^{(a+b+)}$ (%)	$Fy^{(a+b-)}$ (%)	$Fy^{(a-b+)}$ (%)	$Fy^{(a-b-)}$ (%)	2010 United Nations estimated country population (in thousands)
Peru	7.92	79.07	4.13	8.88	29,496
Suriname	19.67	72.65	5.68	2.00	524
Venezuela	23.51	23.87	44.34	8.29	29,044
West Asia					
Afghanistan	29.44	49.62	17.38	3.56	29,117
Azerbaijan	40.86	34.92	23.38	0.84	8,934
Bangladesh	29.68	51.85	15.72	2.75	164,425
Bhutan	21.15	72.23	5.43	1.19	708
Georgia	47.78	23.71	28.45	0.05	4,219
India	28.28	52.40	15.89	3.43	1,214,464
Iran (Islamic Republic of)	27.35	39.53	25.23	7.89	75,078
Iraq	24.43	41.41	25.08	9.08	31,467
Kyrgyzstan	29.01	58.02	11.20	1.77	5,550
Nepal	27.60	61.57	9.35	1.48	29,853
Pakistan	29.43	45.16	20.88	4.53	184,753
Saudi Arabia	8.48	27.69	29.09	34.73	26,246
Sri Lanka	16.77	72.26	6.73	4.23	20,410
Tajikistan	30.04	53.18	14.37	2.41	7,075
Turkey	42.71	26.00	30.68	0.60	75,705
Uzbekistan	30.64	52.50	14.62	2.25	27,794
Yemen	10.63	28.49	32.08	28.80	24,256
Central Asia					
Cambodia	20.87	76.20	2.73	0.20	15,053
China	12.72	86.39	0.83	0.07	1,354,146
Korea, DPR	10.68	88.92	0.40	0.00	23,991
Lao People's Democratic Republic	16.00	80.22	2.87	0.91	6,436
Myanmar	23.65	69.34	5.94	1.07	50,496
Republic of Korea	10.01	89.62	0.36	0.00	48,501
Thailand	24.36	71.11	4.15	0.38	68,139
Viet Nam	12.17	84.92	1.86	1.05	89,029
East Asia					
Indonesia	11.72	83.89	2.66	1.73	232,517
Malaysia	24.37	66.08	8.03	1.51	27,914
Papua New Guinea	1.93	97.86	0.15	0.06	6,888
Philippines	15.24	80.94	2.81	1.01	93,617
Solomon Islands	8.35	87.63	2.36	1.66	536
Timor-Leste	8.01	89.66	1.22	1.11	1,171
Vanuatu	2.13	97.58	0.11	0.18	246

Phenotype frequencies were estimated from the allele frequency maps presented by Howes et al. (10), summarized in Fig. 5. National population estimates were derived from the high resolution Duffy phenotype and 2010 United Nations population maps (adapted from http://esa.un.org/unpp/). *P. vivax* endemic countries and their regional categorizations are based on those recently determined by Guerra et al. (12).