

DYNAMICS UNDERLYING
PLASMODIUM VIVAX MITOCHONDRIAL
GENOME DIVERSITY ACROSS THE
EASTERN HEMISPHERE

A PAST HUMAN MIGRATORY PERSPECTIVE



Leonie Renée Jacoba Raijmakers

Merton College, Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Archaeological Science, University of Oxford
Hilary Term, 2018

ABSTRACT

Plasmodium vivax (*Pv*) malaria is a human infecting blood parasite distributed widely across both tropical and temperate regions. In order to increase the understanding of past dynamics influencing its current distribution, this thesis explores its origins, spread and evolutionary past through diversity and evolution of the mitochondrial genome. Exploring several different factors that would have affected its dispersal, including mosquito vector species and geographic distance, the main focus of *Chapter 2* is on understanding when and with which past human migrations it spread across the continents of the Eastern Hemisphere. A special emphasis on the Melanesian region is included in *Chapter 3*, which shows considerable diversity in human populations and cultures, and has high incidence of all four species of human infecting malarias (*P. vivax*, *P. falciparum*, *P. malariae*, *P. ovale*). Although previous publications indicated an especially high level of diversity in *Pv* mitochondrial genomes in Melanesia; in this study it is shown to be a sampling artefact due to denser sampling. In both chapters a novel cross-disciplinary data comparison is undertaken, matching *Pv* mitochondrial genome phylogeny and population genetics with modern human mitochondrial genome data, human and hominid archaeological data, archaeological data from human commensal species and phylogenetic data from human associated diseases. Results indicate that not only the current Melanesian *Pv* but also the *Pv* strains found across the Asian continent to the east of India were likely introduced by the first wave of anatomically modern humans (*Homo sapiens*) to leave Africa. The strong patterning seen across this eastern region has ostensibly been shaped not only by relatively stable human populations for the last several thousand years, but is also associated with a strong regional heterogeneity of mosquito vector species and clades. In contrast, the present study confirms previously observed homology in *Pv* mitochondrial genetics from India to the west. Presumably the homology is due to increased human population movement and contact between the western regions, as well as greater overlap in mosquito vector species across the region, as shown in this study. Even so, with the addition of data from new sites across the western half of the Eastern Hemisphere, including samples from central and western Asia, there is a detection of low levels of population diversity. Lastly, *Chapter 4* gives an overview of the applications of different genetic markers used in malaria research over time, reviewing the continued value of using mitochondrial DNA, on its own and in combination with other available genetic data – in an age of whole genome sequencing.

INDEX CHAPTERS

CHAPTER I

<u>1. 1. INTRODUCTION</u>	<u>2</u>
<u>1. 2. DNA RESEARCH IN ARCHAEOLOGY</u>	<u>9</u>
<u>1. 3. MALARIA, ORIGIN & SPREAD</u>	<u>12</u>
1. 3.1 PLASMODIUM VIVAX & PLASMODIUM FALCIPARUM: AGE & EVOLUTION	12
1. 3.2 POPULATION DYNAMICS & DISEASE EVOLUTION	16
<u>1. 4. LARGE-SCALE DIVERSITY PATTERNS</u>	<u>22</u>
1. 4.1 DISEASE GENETICS AND HUMAN MIGRATIONS	22
<u>1. 5. SMALL-SCALE DIVERSITY PATTERNS</u>	<u>26</u>
1. 5.1 POPULATIONS AND DISEASE	26
1. 5.2 “PARASITE-DRIVEN WEDGE”	26
1. 5.3 HUMAN ACTIVITIES AND VECTOR-BORNE DISEASES	30
1. 5.4 HUMAN INVASION, DISEASE AND IMMUNITY	32
<u>1. 6. CONCLUSION</u>	<u>36</u>

CHAPTER II

<u>2. 1. INTRODUCTION</u>	<u>39</u>
<u>2. 2. GENETIC STRUCTURE</u>	<u>41</u>
2. 2.1 INTRODUCTION	41
2. 2.2 METHODS	42
2. 2.3 RESULTS & DISCUSSION	45
<u>2. 3. HUMAN MIGRATION</u>	<u>53</u>
2. 3.1 INTRODUCTION	53
2. 3.2 METHODS	57
2. 3.3 RESULTS & DISCUSSION	59
<u>2. 4. GEOGRAPHY</u>	<u>63</u>
2. 4.1 INTRODUCTION	63
2. 4.2 METHODS	64
2. 4.3 RESULTS & DISCUSSION	65

<u>2. 5. MOSQUITO VECTOR</u>	<u>67</u>
2. 5.1 INTRODUCTION	67
2. 5.2 METHODS	71
2. 5.3 RESULTS & DISCUSSION	73
<u>2. 6. OVERALL RESULTS & DISCUSSION</u>	<u>75</u>
<u>2. 7. OVERALL CONCLUSION</u>	<u>80</u>

CHAPTER III

<u>3. 1. INTRODUCTION</u>	<u>82</u>
<u>3. 2. METHODS</u>	<u>87</u>
<u>3. 3. RESULTS</u>	<u>92</u>
<u>3. 4. DISCUSSION</u>	<u>99</u>

CHAPTER IV

<u>4. 1. INTRODUCTION</u>	<u>106</u>
<u>4. 2. MALARIA PHYLOGENETICS SEQUENCING APPLICATIONS OVER TIME</u>	<u>109</u>
<u>4. 3. CURRENT APPLICATION OF MTDNA IN PHYLOGENETICS</u>	<u>113</u>
<u>4. 4. MTDNA AND MS DATA USAGE & AGGREGATION</u>	<u>115</u>
<u>4. 5. DISCUSSION & CONCLUSION</u>	<u>125</u>

CHAPTER V

<u>5. 1. CONCLUSION</u>	<u>127</u>
-------------------------	------------

APPENDICES

<u>APPENDIX</u>	<u>I</u>
<u>CHAPTER II</u>	<u>II</u>
APPENDIX S2.1	II
APPENDIX S2.2	III
APPENDIX S2.3	IV
APPENDIX S2.4	V
APPENDIX S2.5	VI
APPENDIX S2.6	VII
APPENDIX S2.7	VIII
APPENDIX S2.8	XI
APPENDIX S2.9	XII
APPENDIX S2.10	XIII
APPENDIX S2.11	XIV
APPENDIX S2.12	XV
APPENDIX S2.13	XVI
APPENDIX S2.14	XVII
APPENDIX S2.15	XVIII
APPENDIX S2.16	XXII
<u>CHAPTER III</u>	<u>XXIII</u>
APPENDIX S3.1	XXIII
APPENDIX S3.2	XXIV
APPENDIX S3.3	XXV
APPENDIX S3.4	XXVI
APPENDIX S3.5	XXVII
APPENDIX S3.6	XXVIII
APPENDIX S3.7	XXIX
APPENDIX S3.8	XXX
APPENDIX S3.9	XXXI
<u>CHAPTER IV</u>	<u>XXXII</u>
APPENDIX S4.1	XXXII
APPENDIX S4.2	XXXIII

ACKNOWLEDGEMENTS

REFERENCES

INDEX CHAPTER I

<u>1. 1. INTRODUCTION</u>	<u>2</u>
<u>1. 2. DNA RESEARCH IN ARCHAEOLOGY</u>	<u>9</u>
<u>1. 3. MALARIA, ORIGIN AND SPREAD</u>	<u>12</u>
1. 3.1 PLASMODIUM VIVAX & PLASMODIUM FALCIPARUM: AGE & EVOLUTION	12
1. 3.2 POPULATION DYNAMICS & DISEASE EVOLUTION	16
<u>1. 4. LARGE-SCALE DIVERSITY PATTERNS</u>	<u>22</u>
1. 4.1 DISEASE GENETICS AND HUMAN MIGRATIONS	22
<u>1. 5. SMALL-SCALE DIVERSITY PATTERNS</u>	<u>26</u>
1. 5.1 POPULATIONS AND DISEASE	26
1. 5.2 “PARASITE-DRIVEN WEDGE”	26
1. 5.3 HUMAN ACTIVITIES AND VECTOR-BORNE DISEASES	30
1. 5.4 HUMAN INVASION, DISEASE AND IMMUNITY	32
<u>1. 6. CONCLUSION</u>	<u>36</u>

CHAPTER I

INTRODUCTION

A PARASITES' PAST:

PLASMODIUM VIVAX MALARIA AND HUMAN MIGRATIONS

*"I recognised you as soon as I saw you
As if we called each other
The DNA in my blood is telling me
That it's you I've been looking for." – BTS, DNA*

1. 1. INTRODUCTION

With increased ease and access to genome sequencing, there has been growing interest in understanding past human migrations through the relatively novel method of studying the genetics of human diseases (Curtin 1968, Brockmann *et al.* 2006, Tatem *et al.* 2006, Bryant *et al.* 2007, Moodley *et al.* 2009, Culleton *et al.* 2011, Bos *et al.* 2011, Miao *et al.* 2011, Comas *et al.* 2013, Taylor *et al.* 2013). Many major infectious diseases are thought to have originated around the time of the emergence of agriculture, some 10,000 years ago (ya), and to have spread with subsequent human population expansion and migration (Wolfe *et al.* 2007). Some diseases, however, are more ancient. Certain strains of tuberculosis (*Mycobacterium tuberculosis*) were present in humans before their exit out of Africa and spread with the earliest known human migrations, some 62,000-75,000 ya, as well as with the subsequent large migration around 25,000-38,000 ya (Comas *et al.* 2013). Malaria is another disease that has been strongly affiliated with humans - and probably hominids before that - for many hundreds of thousands of years (Garnham 1966 in Groube 1993). It too is now believed to have travelled with the earliest migrations out of Africa into Asia (Miao *et al.* 2011), especially with recent evidence for Africa as its place of origin (Rayner *et al.* 2011, Culleton *et al.* 2011, Prugnolle *et al.* 2013). This thesis seeks to uncover which, if any, of the major human or even hominid migrations was associated with the spread of *Plasmodium vivax* (Pv) malaria, by studying the phylogenetics of the Pv mitochondrial genome and relating it to what is known of the human past through different disciplines, including archaeology.

At the present time, malaria is topical across a broad range of disciplines. The interest surrounding it has been sparked by a marked decrease in the effectiveness of eradication efforts. For *Plasmodium falciparum* (Pf), the most prolific and malignant human malaria species, this is partially due to the

parasite's increasing resistance to the current drugs available (White 2004, Strode *et al.* 2014, Blasco *et al.* 2017, Imwong *et al.* 2017, Ochomo *et al.* 2017). For *Pv*, chloroquine resistance is also increasing but vivax malaria remains highly susceptible to other antimalarial treatments such as artemisinin combination therapy, primaquine and tafenoquine. Effectiveness of eradication efforts is caused by a combination of resistance, lack of investments that mean lack of support for poor health systems that cannot deliver the interventions effectively, and the fact that much of the remaining malaria (at least in Asia and Americas) is in geographically or demographically difficult to reach populations – such as on unstable border regions and mobile migrant populations.

In addition, certain anophelene mosquito vectors - key to malaria transmission - are developing resistance to the chemicals used in bed nets (Sokhna *et al.* 2013, Miles *et al.* 2017). This, and resistance of parasites to treatment, forms issues for regions currently infected by malaria, but also poses a threat for regions which have already eradicated the disease but remain receptive to malaria transmission. The threat to malaria-free status comes in several forms for certain regions, depending on their current status. For example, Taiwan, Hong Kong and Okinawa no longer have malaria parasites circulating in their population. However, this elimination has taken place without removing the resident transmission vectors (Yip 2009). Reintroduction of the disease from surrounding areas still occurs on low levels, but these events can as yet be easily controlled with medication. An example of new outbreaks occurring in previously malaria-free regions include a small *Pv* outbreak in the Southern Peloponnese, Greece, in 2009 (Andriopoulos 2013), and, since then, an increase in *Pv* patients across several other regions in Greece (Dimopoulou 2016), where effective *Anopheles* mosquito vectors are still prevalent. Removing effective medication from the equation has the consequence of reversing elimination as these populations become re-infected. To counteract the adaptation of the mosquito vectors and malaria parasites to the existing medication and pesticides, there has been a marked reinvigoration of the global malaria eradication programme. A global eradication programme was run by the World Health Organisation (WHO) in the 1950s and 1960s, but was abandoned in 1969 due to slow progress and the development of resistance in the anophelene mosquitoes to the pesticide DDT (Yip 2009). With support from the Global Funds for AIDS, TB, Malaria – and other bilateral donors – many countries have re-intensified their malaria control programs in the early 2000s, leading to dramatic reductions in malaria burden in numerous parts of the world (World Health Organization 2008, Global Partnership to Roll Back Malaria *et al.* 2009). This has prompted all countries in the Asia-Pacific and Americas to re-declare malaria elimination as a formal goal. For reasons stated above, unfortunately, since 2016 the progress towards malaria control and elimination has stalled and is at risk of going backwards (World Health Organization 2018).

At such a critical time, it is of great importance to understand malaria and the dynamics behind its spread and diversity. Medical research generally focuses on understanding the current situation on local scales, molecular levels and curing or preventing the disease. However, this thesis aims to help understand the long-term evolution and impact of the disease, approaching the subject from a broader perspective and much longer time scale than most. This chapter has the function of setting out where the knowledge of the past of malaria and humans in the Eastern Hemisphere currently stands: to provide a framework for further in-depth study of different forces acting on and between malaria and humans that could have influenced the genetic diversity of both species. The approach in later *Chapters 2 & 3*, using geographical mapping of malaria phylogenetics, is designed to assist in understanding the routes and timing of the spread of the disease. In addition, once the historic connection between humans and malaria has been established, the connection can possibly be used to clarify past global human migration routes in a similar fashion to previous studies of disease phylogeny (Curtin 1968, Brockmann *et al.* 2006, Tatem *et al.* 2006, Bryant *et al.* 2007, Bos *et al.* 2011, Miao *et al.* 2012, Comas *et al.* 2013).

Linking the presence of malaria to certain migrant populations also provides an opportunity to confirm the minimum extent of time over which the migrants have been exposed to the disease. This will aid in understanding the process of human genetic adaptation to a disease such as malaria, by giving an indication of the varying exposure times of the different migrant populations to the disease. An understanding of the period of exposure will, in turn, clarify to what extent malaria played a role in shaping the human cultural landscape in the region. As such, combining the datasets on humans and malaria is necessary in order to elucidate the role of disease in human evolutionary and cultural processes, and to gain insight into the role of human past activities in the spread and diversity of this disease. Before this can be done, however, it is important to first understand the timing and dynamics behind the spread and persistence of the disease throughout the globe.

Current research into *Pv* phylogenetics is starting to shed light on these interconnected pasts of humans and malaria on a global scale. Past studies with *Plasmodium falciparum* (*Pf*) have already indicated an introduction of the disease to the Americas with the African slave trade (Yalcindag *et al.* 2012). For *Pv*, the African-American connection in the context of human migration was recently published by Rodrigues *et al.* (2018). A possible European origin for American *Pv* was discussed in Gelabert *et al.* (2016), with the sequencing of a historic sample of *Pv* from Spain. Even so the origin of South American *Pv* and how it connects to *Pv* in the rest of the world remains unclear, possibly due to the lack of European samples. The South American connections are studied extensively by others (Taylor *et al.* 2012, Rodrigues *et al.* 2018) and the complex positioning of American data in the *Pv* mtDNA phylogeny as observed in other studies (Mu *et al.* 2005, Culleton *et al.* 2011, Miao *et al.* 2012) may obscure some of the finer details found in other

regions. This thesis concentrates exclusively on *Pv* in the Eastern Hemisphere: the African, Eurasian and Oceanian continents, from the Atlantic to the Pacific.

Possibly more substantial evidence for a connection between human migration and *Pv* distribution is the presence of a divide in Central, Eastern and South East Asian *Pv* versus an Indo-Malay, Australasian branch of *Pv* (Miao *et al.* 2012). The divide could be the result of two main human migration events out of Africa and into Asia following different routes at different times, as has been found with tuberculosis (Comas *et al.* 2013) and indicated by human whole-genome variation (Gignoux *et al.* 2011, Stewart & Chinnery 2015) and archaeology (Bae *et al.* 2017).

A recent paper published by Skoglund & Mathieson (2018) reviews the evidence for human migratory movements in the context of currently available publications on ancient DNA in hominids. The review reiterates the origin of modern humans from Africa, with all modern humans outside of Africa falling within one genetic clade that branches off a diverse African phylogeny. Though on a single branch off the African tree, the population movement and diversification of the Eurasian (and American) clade has been varied and complex. It has included an early split that has shaped the variety of populations on the very eastern edges of the Eurasian and Oceanian continents. This split saw two major population movements: one taking a southern route along the Indian Ocean coast towards and into Australasia, also known as haplogroup M; the other took a more northerly route, probably through the Levant and possibly up through central Asia and along the northern side of the Himalayas, also known as human haplogroup N (Stewart & Chinnery 2015, Bae *et al.* 2017). A third haplogroup R emerged from haplogroup N and gives rise to several lineages in Europe and eastern Asia whilst it takes a similar northern route to haplogroup N (Stewart & Chinnery 2015).

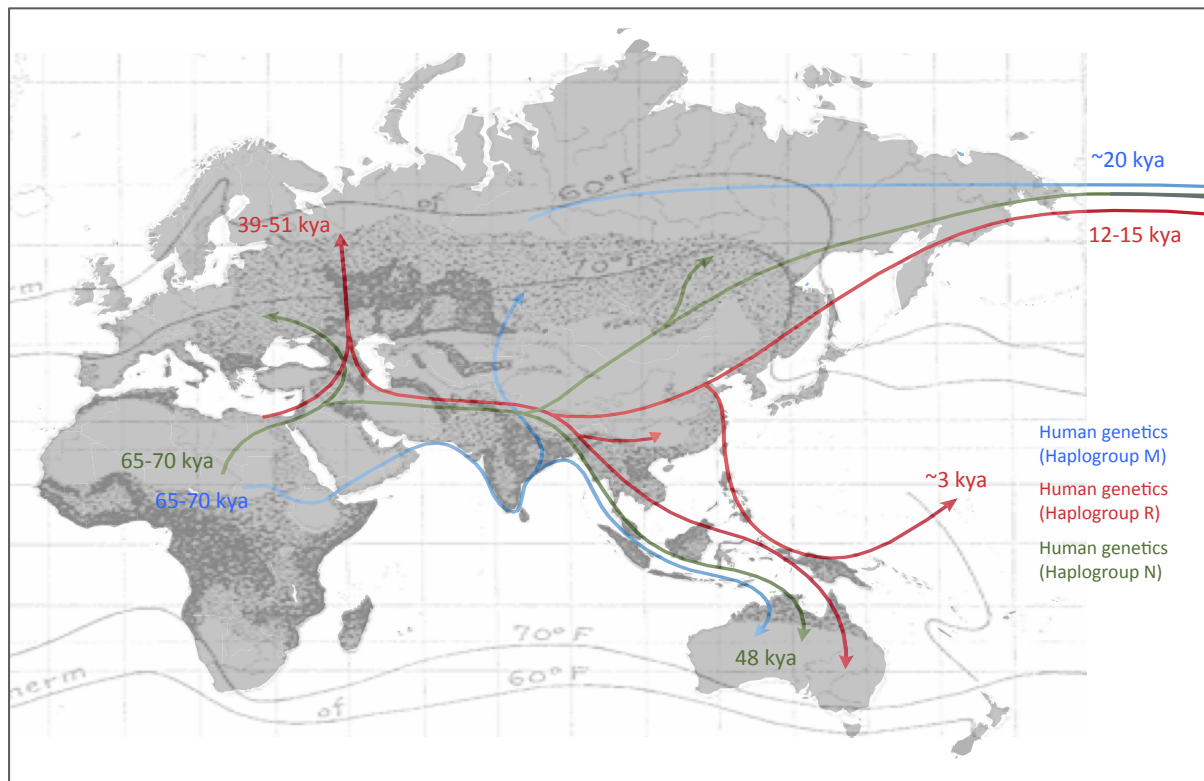


Figure 1: Broad indication of the migratory routes and timings of the earliest *Homo sapiens* to exit Africa, based on Bae *et al.* 2017. Routes as found for the different human mitochondrial genome haplogroups are given in light blue (haplogroup M), red (haplogroup R), and green (haplogroup N). Dates are estimated ages of time to the most recent common ancestor of populations in the region marked. Overlain with the malaria distribution map by Boyd 1930, indicating the presence of both human infecting malaria parasites as well as anophelene vectors. Shaded/dotted regions denote areas with malaria infections: increased density indicates higher number of infections across the human population in the region.

Assuming that malaria vectors were present along both routes, as they were until recently (Boyd 1930, MacDonald 1957, Sinka *et al.* 2012), the expectation is that if *Pv* came out of Africa with all migrants, it had the opportunity to travel via both migration routes. This can be seen in *Figure 1*, where most estimated migration routes pass through regions that are, or until recently were, malarious. Climate throughout the past 100 ky was changeable (Timmerman & Friedrich 2016, Stewart *et al.* 2017), so consistency of vector availability is unpredictable. Yet at times more suitable habitat for transmission existed in the tropical and subtropical regions; there is much evidence for deserts having been less extensive than they are today, e.g. at 95 - 107 and 75 - 90 kya (Timmermann & Friedrich 2016) and during the Neolithic, 12 - 6 kya (Platt *et al.* 2017). Deserts appear to be a huge barrier for malaria survival, as seen in *Figure 1*, with absence of malaria in the Sahara in Northern Africa, the central areas of the Arabian Peninsula and the Gobi desert, in Southern Mongolia and Northeastern China. The spread of *Pv* across the Eastern Hemisphere is explored in greater depth in *Chapter 2*, which looks at the global distribution of *Pv* according to its mitochondrial genome as influenced by several factors, including human migration.

From the two separate global human migration events, the first wave of *Homo sapiens* out of Africa connects to the first migration into Australasia by anatomically modern humans. Few traces have

remained in intermediate geographic regions through a high density of subsequent migrations. If vivax malaria distribution patterns mirror that of *H. sapiens*, the closest genetic relations of *Pv* from the Australasian region are expected to be found at its original source, in Africa. Both malaria and human populations across the region display especially high diversity around Melanesia, the region in the Pacific encompassing the islands of Papua, the Solomon Islands, Vanuatu and Fiji. Much research into human diversity has focused on the varying origins of the human populations in Melanesia (Bellwood *et al.* 1995, Spriggs 1997, Terrell *et al.* 2001, Pawley *et al.* 2005, Kirch 2010, Donohue & Denham 2010, Gignoux *et al.* 2011, Spriggs 2011, Xu *et al.* 2012). These varying origins represent the main determinant of the large-scale diversity in the Indo-Pacific region, both in cultural and in genetic terms.

Many areas of research have looked into the migratory movements and timings of people into the Indo-Pacific and Melanesia, as there is much focus and ongoing interest in the timing of migrations ‘Out of Africa’. Archaeological, geological, palaeoecological, climatological, linguistic and genetic studies including modern human, ancient, commensal/agricultural and, recently, disease genetics, have all contributed to what is currently known about past human migratory movements (Jin & Su 2000, Larson *et al.* 2007, Atkinson *et al.* 2008, Dobney *et al.* 2008, Petraglia *et al.* 2010, Bellwood 2011, Bellwood *et al.* 2011, Gignoux *et al.* 2011, Rasmussen *et al.* 2011, Henn *et al.* 2012, Jinam *et al.* 2012, Moodley *et al.* 2012, Paraskevis *et al.* 2012, Soares *et al.* 2012, Stewart & Stringer 2012, Comas *et al.* 2013, Malaspinas *et al.* 2016, Skoglund *et al.* 2016, Soares *et al.* 2016, Gomes *et al.* 2017). It has been recognised that at least two main migratory movements of anatomically modern humans occurred into Oceania (e.g. Chappell 1993, Spriggs 1997, Kayser *et al.* 2006, Hudjashov *et al.* 2007, Duggan *et al.* 2014, Gomes *et al.* 2015). ‘Oceania’ is used here in a broad sense, referring to all the tropical island groups between Asia and the Americas, yet pertaining especially to the Melanesian region.

These two migratory waves into Oceania connect back to the two different global migration routes, and the migrations are represented by a duality in linguistics and human genetics. The two-way split is shown in the Australoid/Papuan versus the Austronesian populations. The Australoids first entered the region 75,000 to 40,000 ya (Kirch 2010, Summerhayes *et al.* 2010, Gomes *et al.* 2015, Clarkson *et al.* 2017, Larruga *et al.* 2017, Nagle *et al.* 2017) with possible subsequent migrant waves between 15,000 and 5,000 ya, represented by - intraregional movement of - Papuans (Soares *et al.* 2008, Gomes *et al.* 2015) who did not reach Australia as sea-levels between Australia and Papua New Guinea had heightened. The much later Austronesian migration came to the region from 4,000 ya, moving far out into the Pacific and probably towards Madagascar across the Indian Ocean (Serjeantson *et al.* 1982, 1983, 1992, Serjeantson 1984, Kirch 2010). The Austronesian migration into Australasia and specifically Melanesia at 4,000 ya is believed to have strong connections with the *H. sapiens* that took the northern route and settled in East Asia, though its exact origins are under debate (Spriggs 1993, Bellwood *et al.* 1995, Terrell *et al.* 2001, Kirch 2010, Donohue & Denham 2011, Spriggs 2011, Xu *et al.* 2012, Stewart & Chinnery 2015, Bae *et al.* 2017).

Apart from this larger scale variation, however, there is a conspicuous representation of smaller scale differentiation between population groups in Melanesia. This differentiation is present in languages, cultural practices and genetics (Serjeantson *et al.* 1992, Spriggs *et al.* 1993, Spriggs 1997, Pawley *et al.* 2005). In this region, however, besides general genetic variations, populations display many different genetic adaptations to coping with malaria (Serjeantson *et al.* 1992, Howes *et al.* 2013). Malaria has played a significant selective role in the evolution of human genetic blood defects, which act as an adaptive mechanism against the more adverse consequences of malarial infection (Flint *et al.* 1986, Flint *et al.* 1998, Kwiatkowski 2005, Williams 2006). The recent discovery of a genetic polymorphism found only in Austronesian people that is associated with a decrease in the risk of contracting vivax malaria (Rosanas-Urgell *et al.* 2012) suggests that the past relationship between human populations and malaria across Austronesia merits further study. As such, *Chapter 3* of this thesis will focus on discovering when and with which human migrations *Pv* entered the Melanesian region.

So far, it has been shown that establishing when malaria entered regions across the globe is important to developing an understanding of its role in shaping human populations in those regions. Doing so can help explain some of the larger-scale diversity found in malaria in the region, as well as creating an opportunity to gain insight into long-term interactions between human populations and disease. As has been mentioned previously, malaria is a disease that significantly influences human populations and exerts strong selective pressure on the human genome to adapt and survive the consequences of the disease. In addition, it would be expected that malaria could select favourably for certain behavioural adaptations. It follows that the earlier malaria was present in human populations, the longer humans would have had to adapt to it, including both the rise of genetic mutations which lessen the severity of the disease, and behavioural changes which disrupt transmission. Viewing society from the perspective of long-term malarial selection could help us understand not only human -cultural- evolution, but also give insight into how societies might operate at their most efficient whilst dealing with the burden of malarial disease.

1. 2. DNA RESEARCH IN ARCHAEOLOGY

As Jobling (2012) explains, there are two methods of using information from genomes to interpret the past: (1) by studying ancient DNA directly; or (2) by mathematically reconstructing the patterns revealed in modern DNA. In archaeology, any material studied is usually at least a few centuries old. The first way to read information in DNA is thus relatively simple to understand in an archaeological context, as it concerns studying ancient biological material itself. It is easy to see how studying ancient DNA (aDNA) from bones, plants or other organic matter found in archaeological digs fits into the discipline of studying ancient materials to further an understanding of the past of human civilizations. DNA from hominid remains can reveal the relations between the hominid and modern humans.

Ancient genomes from anatomically modern humans found in several sites across Asia confirm some human history as deduced from modern genomes and archaeological finds, but disagree with other previously constructed dynamics. For example, the sequenced draft genome for Neanderthals (Green *et al.* 2010) provided relatively clear data indicating that around 1 to 4% of modern human Eurasian genes are likely to have come from Neanderthals, revealing possible interbreeding of Neanderthals with our *Homo sapiens* ancestors at some point in the past. This was unexpected; the current findings are that the genetic contribution of Neanderthals to human lineages outside of Africa is around 2%, and there is a 3% contribution of the still mysterious Denisovan lineage(s) in some of Asia's human populations, specifically in Oceania (Skoglund & Mathieson 2018).

What makes human ancient DNA (aDNA) especially useful is that it gives both a place and a time point to refer to in regards human movement and evolution. With aDNA, the estimated mutation rate over time for the genome gets calibration points from both time and number of mutations, making the molecular clock much more precise. In reference to information about the past obtained from human aDNA, ancient mitochondrial genomes confirm that the most recent common ancestors (TMRCA) of African and non-African lineages can be dated back to 62-95 thousand years ago (kya), and that non-African sub-lineages diversify around 45-55 kya. These dates correspond with earlier estimates of these timings. However, some archaeological findings of anatomically modern human remains in Australia and Indonesia push back these dates considerably (Clarkson *et al.* 2017, Westaway *et al.* 2017): the archaeological finds in the Arnhemland escarpment (Australia) and Lida Ajer Cave (Sumatra, Indonesia) are dated to 65 (59.3-70.7) kya and 68 kya respectively, considerably predating the sub lineage diversification estimate above. It should be kept in mind, however, that biological processes--and especially those that deal with population dynamics, such as population genetics--are not straightforward. Genetic relationships over time are influenced by multiple variables. This usually leads

to a genetic time overestimation of a physical geographic split. This is due to a lack of representation of all data at all time points, such that the odds of finding the exact ancestor of two lineages are rather low; the representative ancestor will be found slightly farther back in time.

Similarly, when studying human disease in the past, the most straightforward method is to use ancient DNA of the disease, found either in direct association with certain hominid sites or with certain events in the past. Examples of such research include the studies tracing the origins of the Black Death, *Yersinia pestis*, by Bos *et al.* (2011) and tracing human migratory and trade movements through bowel parasites, including *Caenorhabditis elegans*, preserved in archaeological materials (Patrik Flammer, *pers. comm.* 2012). Such an approach might be possible through analysing preserved soft tissue of human remains for malaria sequences, with varying origins. Alternatively, old samples of European vivax could aid in understanding the origins of South American vivax, which could have been introduced from Spain or Portugal instead of Africa during the age of exploration and slavery (Gelabert *et al.* 2016, Rodrigues *et al.* 2018).

Ancient or old samples are rare, but do exist. However, when such samples are available, there are further complications; when working with ancient DNA, there are often only very few samples, and the DNA tends to be severely fragmented. The information retrieved is thus both scattered and scant. Though ancient DNA helps increase the precision of time and route estimates, aiding significantly to increasing evolutionary insight in a species, it is important not to forget that even these reconstructions rely on models and assumptions about population dynamics and are not infallible. This is where the second method of reading DNA comes into play, which relies more heavily on reconstructive models but has alternative strengths. Instead of using aDNA, insight is gained through reading several samples of modern DNA and studying the patterns that are revealed. In a way, DNA is an archaeological record on its own; it accumulates mutations over time. As chances of the same mutation occurring more than once are negligible if not under selection, it is possible to trace relatedness by comparing levels of similarity between different individuals. The problem in this method lies not in the number of samples – though this can be an extra complication due to restrictions in time and funding – but in having to infer through statistics how samples are related and what this tells us about the past. These ‘phylogenetic’ methods of reconstructing genetic relationships based on the DNA code – whether ancient or modern – provide an estimation of the linear associations between genetic samples across time. In both cases time is estimated: for aDNA through archaeological methods of estimating a sample’s age; for modern DNA based on estimations of mutation rates (how much time passes between the accumulation of one mutation and the next).

To understand processes in the past which have led to the patterns seen across geographic space, extra dimensions can be added that help clarify time outside the information contained in the DNA. ‘Phylogeography’ overlays a geographical map with phylogenetic data to demonstrate the genetic relatedness between samples from their sites. Other dimensions come from data provided by archaeology, linguistics, palaeo-anthropology and the earth sciences. As Oppenheimer (2012) states: ‘*A combined approach should give increased confidence by triangulation*’. The combination of different classes of data help fill gaps, and provide checks through expected overlap. Data from other disciplines provide much of the broader understanding of human activities within the observed timeframe. One must, however, remain wary that information provided by one area in an integrated data-set does not form the basis for conclusions of another, as this will lead to false positives through data reiteration. Another hazardous strategy is to randomly pick the best fitting story about the past to explain the patterns in the data. It is advisable to remain aware of alternative explanations, which could be much more complex, and keep a certain amount of objectivity. Approaching a genetic dataset from a combined evolutionary and archaeological perspective, bringing an awareness of alternative explanations for genetic variation on at least two different levels prior to looking at the data, is arguably a great advantage.

1. 3. MALARIA, ORIGIN AND SPREAD

Malaria is a disease caused through infection by one or more specific species of single celled protozoan organisms, of the genus *Plasmodium*. Infections are passed to humans through anophelene mosquito vectors, *Anopheles* spp.. Both humans and mosquitoes form crucial parts to the parasites' life cycle. Sexual replication takes place in the mosquito vector. 10-18 days after infection of the mosquito, the parasites are found in 'sporozoite' form, in its salivary glands. The mosquito passes on these sporozoites to humans when feeding on them for a blood meal. In the human host, the parasites first invade and grow in the liver cells. In *Pv*, this phase in the liver can be prolonged by the parasites going into a dormant 'hypnozoite' form, for several months up to four years. From the liver cells, the parasites are released into the blood stream, where they invade and multiply in the red blood cells. Another difference between *Pv* and *Pf* is found at this stage: whereas *Pv* invades only young red blood cells (i.e. erythroplasts and reticulocytes), *Pf* can invade blood cells of different ages (Miller *et al.* 2002). After multiplying, the parasites burst out of the blood cell walls and release 'merozoite' daughter cells into the bloodstream, which reinvade and multiply in red blood cells in a repetitive cycle. This stage is what causes the symptoms associated with malaria. Another form of blood stage parasite – the 'gametocyte' – infects the mosquito when it feeds off a malaria infected person, and starts the whole cycle over again.

Malaria and human ancestors have been associated with one another from before the emergence of the hominids (Garnham 1966 in Groube 1993, Silva *et al.* 2011). Garnham (1966) contributes that the current complex equilibrium in the relationship between malaria and its definitive and intermediate hosts, respectively, has taken a long time to establish, indicating a very long period of evolution. What is currently known of malaria's past depends on the species. For many years, research has mainly focused on the most malignant malaria parasite – *Pf*, which accounts for most infections deaths worldwide and can be cultured *in-vitro* – providing ample data to understand its past. The recent focus of research on *Pv* should aid in rapidly enlightening its origins. A brief overview of what is known for both is given below.

1. 3.1 *PLASMODIUM VIVAX* & *PLASMODIUM FALCIPARUM*: AGE & EVOLUTION

Previous age estimates calculate the origins of *Pf* to stem from before the human-chimpanzee split, infecting human ancestors in Africa for the first time between 112,000 and 1,036,000 years ago (Escalante & Ayala 1994, Joy *et al.* 2003, 2006, Silva *et al.* 2011, Baron *et al.* 2011). A species of *Plasmodium* infecting gorillas has been found to be the immediate precursor, or at least closest relative, of *Pf* (Liu *et al.* 2010, Rayner *et al.* 2011), re-igniting the discussion on apes being a recurring source of re-infection for humans by plasmodia. There have been studies concluding that *Pf* originated from primates relatively recently –

about 10,000 ya. However, the date estimates from these studies have later been attributed to the genetic effects of a relatively small, bottle-necked – but already established – *Pf* population, suddenly expanding during the onset of agriculture around 10,000 ya (*for discussions see*: Tishkoff *et al.* 2001, Volkman *et al.* 2001, Coluzzi *et al.* 2002, Barry *et al.* 2003, Joy *et al.* 2003, Hartl 2004, Baron *et al.* 2011).

Recent publications have brought to light the possibility that *Pf* travelled along with at least one of the several human migration waves into Oceania in the distant past, possibly even dating back to the earliest known migrations out of Africa between 130,000 and 40,000 ya (Joy *et al.* 2006). Conflicting ideas about when *Pf* branched out of Africa still remain. The earlier mentioned increase of *Pf* due to agricultural development around 10,000 ya, in conjunction with increased human migration at the time due to population increase, is a strong argument for a later introduction of *Pf* (Livingstone 1958, Tishkoff *et al.* 2001, Coluzzi *et al.* 2002, Hartl *et al.* 2002, Hartl 2004, Joy *et al.* 2003, Rich *et al.* 2009). Vedic texts from India (4,000 – 3,500 ya) indicate that *Pv* and *P. malariae* reached India before *Pf* (Carter & Mendis 2002). The latter is mentioned in texts around 3,000 years old.

Unlike *Pf*, the place of origin for *Pv* has been subject to much debate. Supporting evidence was present for both Asian and African heritages. High genetic diversity and the presence of closely related malaria species in primate/ape species, namely macaques, are both strong arguments to indicate an area of origin. As such, previously, evidence was slightly stronger for an Asian origin (Escalante *et al.* 2005, Mu *et al.* 2005), until the discovery of multiple forms of malaria – closely related to vivax – infecting apes in Africa (Culleton *et al.* 2011, Prugnolle *et al.* 2013, Liu *et al.* 2014). The high diversity of these forms of malaria indicates a strong possibility for an extensive African origin for *Pv*. Besides these multiple vivax-related malaria species in Africa, there is evidence for a protracted presence of vivax malaria in Africa through the presence of human genetic adaptations which specifically provide protection against vivax malaria – indicating vivax origins are expected to lie in Africa. It is also estimated to be older than first thought, largely pre-dating the first known human migrations out of Africa (Mu *et al.* 2005, Silva *et al.* 2011).

In addition to the discovery of the new vivax-like primate malaria species, the advent of molecular diagnostics has found *Pv* to be more common in western and central Africa than previously hypothesised (Culleton & Ferreira 2012, Howes *et al.* 2015, Poirier *et al.* 2016). The occurrence of *Pv* across these regions was doubted for several decades due to the presence of Duffy negativity – the specific genetic adaptation against vivax malaria, which modifies part of surface of red blood cells, making it harder for the parasites to enter – in most human populations in the area. Evidence has accumulated for *Pv* infecting people who are Duffy antigen negative (personal observation & Niangaly *et al.* (unpublished, 2017), Ryan *et al.* 2006, Menard *et al.* 2010, Mendes *et al.* 2011, Wurtz *et al.* 2011), which is a genetic adaptation in mainly central and western African human populations that protects against *Pv* infection. Together with the evidence

provided by the new vivax-like malaria and the widespread presence of Duffy negativity in human populations, the evidence mounts for the root of *Pv* to lie in Africa.

Besides the uncertainty of an African or Asian origin, some arguments were presented for *P. vivax* originating in the Americas. The arguments were mainly based on the genetic proximity of a *Pv*-like *Plasmodium* species, found in monkeys in the Americas. Until the recent publications by Culleton *et al.* (2011) and Prugnolle *et al.* (2013), on African vivax-type parasites in wild apes, the closest relatives of *Pv* were believed to occur in Old World macaques (*P. cynomolgi*, *fragile* and *knowlesi*) (Waters *et al.* 1991, 1993, Escalante & Ayala 1994, Mu *et al.* 2005), and South American monkeys (*P. simium*) (Escalante *et al.* 2005, Mu *et al.* 2005). The monkey-infecting species *P. simium* in the New World is so closely related to *Pv* that it has likely descended from the human strain, but the timing of this evolution from human to monkey is unknown. It has been implied that the introduction of *Pv* that led to the evolution of *P. simium* was pre-Columbian, and possibly connected to the Austronesian expansion through Polynesia (Escalante *et al.* 2005, Cormier 2011). However, these findings are still questionable and evidence is mounting for the introduction of *Pv* into the Americas through the European colonists (Rodrigues *et al.* 2018, Gelabert *et al.* 2016).

Over the past few years, there has been much discussion about the different ages for *Pv* and *Pf* (Ricklefs & Outlaw 2010, Silva *et al.* 2011). One argument that has been used in the past to argue a greater age for *Pv* than *Pf* is its reduced malignancy (Brooks & McLennan 1992). When a disease is adapting to a new host, it can be unusually malignant, or virulent, as it still needs to adapt to the host and the host to the disease. Killing off a host quickly, before infecting new hosts, is a strategy that hampers survival of the disease and thus generally uncommon in diseases that have evolved in conjunction with their host for a long period of time (Wolfe *et al.* 2007). However, estimated ages of both *Pv* and *Pf* date back to hominids, which puts both parasites well beyond an initial adaptive period. It is unlikely that the level of malignancy for either reflects their respective ages, instead pointing towards a difference in evolutionary strategy.

Reduction in virulence – how harmful a disease is to the host – is only one part of a suite of evolutionary strategies that usually caters towards transmission or survival of the disease. Disease transmission strategies tend to trade-off between virulence and host survival. Fast and high transmission rates are often associated with impacting the host system & behaviour in ways that increase transmission rates, often at high cost to the hosts – also known as increased virulence of the disease. Such strategies include increasing fluid or airborne transmission, with physical effects to the host that significantly increase such transmission rates including the forced release of bodily fluids (e.g. snot, vomiting, sweating, diarrhoea, or even opening up of skin and release of body fluids through open wounds/pustules), or the release of infections into the air (e.g. through coughing and sneezing).

High transmission strategies through increased virulence impacts the host's body significantly, often leading to reduced survival rates. In addition, a high transmission strategy means available hosts in the vicinity will be rapidly exposed to the disease. The combination of these has the effect of rapidly depleting the available hosts in a host population, removing infectible hosts either through survival – with assumed developed immunity – or through death. A modern complication for the survival of a disease that shows severe symptoms on infection, is the possibility for immediate medical attention. Patients can get treated, isolated or quarantined early on, reducing infection opportunities. Such isolation strategies work for *Pf*, however, *Pv* can transmit before clinical symptoms appear in infected patients, and the *Pv* cycle in humans includes a possible latent stage in the form of the hypnozoite. The hypnozoite hides in liver cells for periods of time that can last up to several years and cause continuous resurgence of the disease in a patient. There is currently only one drug that can eliminate these cells, primaquine – also known as tafenoquine (Campo et al. 2015). The issue is that many populations across infectious sites in Southeast Asia and the Pacific carry a gene – glucose-6-phosphate dehydrogenase (G-6-PD), ironically an anti-malarial genetic adaptation – that makes them more susceptible to oxidative damage to haemoglobin, and erythrocyte (young red blood cell) destruction. Leading to illness and possible death.

The alternative strategy to high transmission strategies, at the other end of the spectrum, is to focus on long-term persistence within a host. This strategy focuses on high host survival rates and long infection duration – increasing the chances of infection over time, often utilising low detection rates for an increased opportunity for exposure to a much larger host population, as hosts with low to no symptoms have relatively high mobility and could be in contact with a large selection of different populations over time. Trade-offs include the possibility of an infected host not infecting others before recovering, receiving treatment or death, forming a dead-end to the disease.

Countering the argument for a younger age due to higher virulence, these theories of evolutionary adaptation form an alternative explanation: instead of the higher virulence being a maladaptation of a young human disease still adapting to its new hosts, it is much more likely *Pf* evolved this strategy. Characteristics exhibited by *Pf* that classify it as such include its relatively short and intense virulence/pathology, indicating dependence on fast transmission and the availability of ample numbers of both hosts and vectors (Wolfe et al. 2007). In doing so, it is characteristic of a crowd disease which would struggle to survive in small migratory human populations. The assumption is that, with an increasing human population, *Pf* has evolved towards being an opportunistic crowd-disease, the traits of which it has managed to exploit after the onset of agriculture and the growth of human populations. This high virulence, crowd disease type strategy is also suspected to have been adopted by more modern

strains of tuberculosis (Brooks & McLennan 1992, Wolfe *et al.* 2007, Joy *et al.* 2012, Comas *et al.* 2013), contrasting with the lower transmission strategies used by older strains. Possibly indicating that some of these ancient diseases continue to adapt to novel population dynamics. *Pf* differs from the usual crowd-diseases, and can be sustained by relatively smaller populations, for two main reasons: 1) immunity to malaria requires many exposures and is only retained with constant re-infection; and 2) the mosquito vector prolongs its persistence as it complicates control over, and predictability of, infections.

The two host system in malaria complicates the already complex system of host-parasite or host-disease strategies. In malaria, the necessity for two hosts overcomes some of the issues of both high and low virulence strategies. For example, in both *Pv* and *Pf* the mosquito vector aids in sustaining transmission even if patients are immobilised. However, having to function and survive in two vastly different host species, also increases issues of having to adapt to, survive in, and control infection pathways through two very different types of species. Though the life stage forms of the malaria parasites presenting themselves inside the two hosts are different – providing opportunities for the parasite to specialise and adapt to each host in generally well-separated systems – it is important to keep in mind that some effects of malaria we see in humans might be adaptations to evolutionary pathways in mosquitoes, and vice versa.

1. 3.2 ORIGIN & THE DUFFY NEGATIVITY CONUNDRUM

As established, there is increasing evidence for Africa as the place of *Pv*'s origin. Classical population genetic theory first supported an Asian origin, because generally, a species' place of origin is likely the one with the highest genetic diversity within the species itself. It is important to address how there are certain exceptions to this general, population genetics-based, assumption. Usually only a limited number of migrants from the original founding population settle new areas, causing genetic variation in migrant populations to be generally lower (Novak 2007). This decrease in genetic variation through selective migration – termed as founder effects or population bottlenecks – across a population can be mitigated in several ways. One is when multiple introductions of varying strains occur, especially if migrants from different populations accumulate in one area, or an uncommonly variable selection of the founding population is included amongst the migrants (Novak 2007, Lavergne & Molofsky 2007). Other processes influencing diversity are: 1) increased selection pressures in different directions (disruptive selection) or, 2) isolation of many smaller populations (genetic drift, and allopatric speciation through fragmentation of populations). These mechanisms could have contributed to the diversity found in *Pv* today.

There are two main identifiable factors influencing disruptive selection. Firstly, and generally the most obvious one – as it influences selection in many species – is the environment. The many different environments in which malaria is found across the globe would have forced malaria to adapt (Vellend *et al.* 2007). Not only because malaria was adapting to the environment itself, but malaria would have also been adapting to changes in behaviour, biology and availability of its vectors.

Another, more uncommon, form of disruptive selection acting on both *Pv* and *Pf* malaria species is the huge variation in species found in its mosquito hosts – as present across the globe. Disruptive selection could have been influenced by the many different species of mosquito present in different regions of the world that can act as vectors for malaria, where malaria strains adapt more towards one vector than another (Joy *et al.* 2008). The different environments found in the island Southeast Asian (ISEA) and Melanesian areas would have had a similar effect, impacting host migration as well as vector distribution (Vellend *et al.* 2007). Different mosquito species and the environment could also both have been drivers of allopatric speciation, i.e. isolating populations through distribution across different mosquito species or island geographies, mountain ranges and varying environments that limit human or vector migration.

If vivax malaria was introduced during the first human migrations, there would have been time for Asian *Pv* mtDNA to diversify beyond the original sequences introduced by the colonising population, especially when combined with the processes mentioned above. Besides the influence of the previously discussed processes of adaptation and diversification of the parasite, on the distribution patterns of *Pv* mtDNA seen today, it is possible that dynamics in the vectors contributed to its distribution too. The African *Pv* founder population could have been burdened with negative selection pressures and have declined to smaller numbers in the past. The strongest suggested cause of this negative selection pressure is that humans in the area developed Duffy negativity. Duffy negativity is a genetic mutation in a critical receptor for *Pv* invasion, and conferring very strong – albeit only partial – protection against vivax malaria. The occurrence of a high prevalence of Duffy negativity across especially Western and Central Africa is taken as evidence for an extended joint past of vivax malaria and humans in the region (Livingstone 1984), with Duffy negativity a human adaptation to high levels of exposure to, and selection from, *Pv*.

As Duffy negativity prevents the malaria parasites from effectively invading the hosts' red blood cells, the absence of Duffy basically bars the parasite from replicating. This not only prevents the host from getting ill, but it also prevents the parasite from invading mosquitoes and moving on to a new host, thus preventing the spread of malaria. The strong protection against vivax malaria that Duffy negativity confers to its human carriers means that a growing Duffy negative population of human hosts complicates the spread and survival of vivax malaria. An increasing number of Duffy negative human

individuals within a population confers what is known as ‘herd immunity’, protecting the population from infection. Chances of an increasingly rare Duffy positive host getting infected by a random, infected mosquito, become significantly reduced. Even were this infection to occur, the chances of infection occurring in sequence – which is required for the survival of the disease in a population – would also be very much reduced.

Even if some (low) levels of infection do occur in the Duffy negative human host, as has been increasingly observed (personal observation, Ryan *et al.* 2006, Menard *et al.* 2010, Mendes *et al.* 2011, Wurtz *et al.* 2011, Albsheer *et al.* 2019) to transfer the malaria parasite to mosquitoes a certain density of parasites is required. High parasite densities are also required further downstream, for mosquitoes to pass on the infection effectively back to humans. As such, a significant decrease in parasite density within an infected host complicates the spread of the parasite, and thus malarial disease; besides protecting the infected host itself from the negative effects of malarial disease. Duffy negativity, therefore, is believed to have all but disappeared – and at the very least gone through a severe population bottle-neck, which lowered genetic diversity – in Western and Central African populations, at some point in the past.

Since the discovery of Duffy negativity in the 1970s (Miller *et al.* 1976), it has been taken for granted that vivax malaria was hardly present in those areas of West Africa – and many other parts of Africa – with high levels of Duffy negativity. The underlying issue was the assumption that *Pv* would be unable to sustain populations in areas containing high levels of resistance in humans. However, recent publications have shown that *Pv* can still infect Duffy negative individuals (personal observation, Ryan *et al.* 2006, Menard *et al.* 2010, Mendes *et al.* 2011, Wurtz *et al.* 2011), and that apes might have acted as possible reservoirs for the disease in the distant past (Prugnolle *et al.* 2013) – indicating the possibility of *Pv* presence even in Duffy-negative areas of Africa being sustained through time. Whether the reduced, but still present, invasive properties of *Pv* within Duffy negative individuals is a new adaptation or one which has been around for a long time, is unclear. However, Boyd (1930) shows that vivax malaria was still present in large areas of Africa, including Duffy negative regions, around the start of the 20th century. It is, however, notable that at that time, confusion of *Pv* with *P. ovale* was rife. Even so, there is value in the reflection that observations by Boyd and earlier malarialogists were (dis)missed in favour of novel popular theories, demonstrating that certain scientific theories can find traction beyond the observed, even with the plausible presence of counter-evidence. Even so, extensive observations, going back several decades, exist of Duffy-positive travellers from Europe acquiring *P. vivax* infections during their visits to West or Central Africa (e.g. Baranova *et al.* 2019), indicating that *P. vivax* has been persisting at low levels throughout Africa for a prolonged period of time. Irrespective of mode of persistence of *P. vivax* in Africa, the pressure of high rates of Duffy negativity in the human host population would have reduced the population size of *P. vivax* in Africa, resulting in potential genetic bottlenecks and

the associated reduced diversity. Possibly causing the lowered diversity seen across this continent, especially when compared to the Asian population. The caveat to this is that in all current studies, there are very few genome sequences of *P. vivax* from West and Central Africa – one each from Nigeria and Sao Tomé (Culleton *et al.* 2011), and Mauritania (Mu *et al.* 2005). As such, the continent-wide diversity may be larger than currently assumed.

1. 3.3 CURRENT UNDERSTANDING OF GLOBAL DISTRIBUTION PATTERNS *Pv* AND *Pf*

Having introduced some of the disease dynamics, and their differences in *Pf* and *Pv*, as well as what is known about their possible ages, this section shortly discusses the differences in their distributions, and what is known of the history and other dynamics behind them.

We already know *Pf* is prevalent across much of Africa, with reduced *Pv* presence due to the Duffy negative adaptation in many African populations. Another note on their general differences in distribution pattern is that *Pf* is restricted to tropical regions. As it develops more slowly during the developmental stages in the mosquito, it needs high temperatures to successfully transmit. In addition, it misses the hypnozoite life stage, which allows *Pv* to bridge between seasons lacking mosquito vectors or seasons that become too cold for the parasite to survive in the mosquito host. *Pv* is present, and has been present much more extensively, across both tropical and temperate zones. Temperate zones historically included coastal areas in the UK, the Netherlands and along parts of Scandinavia. The most northern European latitudes for *Pf*, in the recent past and excluding single introductions from travellers, reach into Spain and Italy. Not much is known about species diversity of either parasite in Europe, especially since both were largely eradicated, and any current ones are likely to have been more recent reintroductions (Rodrigues *et al.* 2017).

In Asia, *Pv* is still present in parts of North and South Korea, Japan, Mongolia and Northern China, including Yunnan (Miao *et al.* 2012). It has distinct genetic diversity, present even within the slowly mutating mitochondrial genome, across sub-continental regions from the Pacific islands of Vanuatu in the Southeast, to Korea and Japan in the Northeast, and Bangladesh in the west (Miao *et al.* 2012, Culleton *et al.* 2011). This diversity will be discussed further later on in the thesis (*see Chapter 2*), and has been discussed in general previously, including its possible underlying driving mechanisms (*see Chapter 1.3.2*). In contrast, Asian diversity in *Pf* is both much less diversified and less differentiated, especially in those genetic regions pertaining to historic dynamics (e.g. Schmedes *et al.* 2019, Tanabe *et al.* 2010, Conway 2003). This lack of variation, which is not only seen in Asia but across the globe, is thought to be linked with a later

(re)introduction of *Pf* into the region, due to a more generally delayed spread across the globe of *Pf* compared to *Pv*, as discussed earlier.

Diversity in African *Pv* is very low, with hardly any showing in previous studies of the mitochondrial genome (Culleton *et al.* 2011). Whether this remains true with an increase in samples from the African continent – without Madagascar, which is the main contributor to the African datasets in current publications – is the question. Due to the low diversity found across its current samples, estimating the age of *Pv* presence in Africa is complex. *Pf* diversity in Africa also has a low diversity, though much higher compared to *Pv*, and generally high compared to the genetic diversity of the species overall (e.g. Schmedes *et al.* 2019, Tanabe *et al.* 2010, Conway 2003). The relatively high diversity of *Pf* in Africa, and the presence of ape-associated malaria lineages in Africa that are genetically most closely related to *Pf*, have made placing the origins of *Pf* in Africa relatively easy from early on (Livingstone 1958, Tishkoff *et al.* 2001, Coluzzi *et al.* 2002, Hartl *et al.* 2002, Hartl 2004, Joy *et al.* 2003, Rich *et al.* 2009), confirmed later by the discovery of *Pf*-associated malaria species in African gorillas (Liu *et al.* 2010, Rayner *et al.* 2011). Even so, it must be noted that there is low to no differentiation of the *Pf* mitochondrial genome at a more localised geographic level, across the continent.

For South America, the introduction route and timing for *Pf* have always been clearly linked with the introduction of slaves from Western Africa around the 1600s (Yalcindag *et al.* 2012, Rodrigues *et al.* 2017). Only recently, similar routes and timing for the introduction of *Pv* have been suggested, this time from Southern European origins, coming with the European explorers and traders instead of the West Africans (Rodrigues *et al.* 2017). Though evidence for this theory is mounting, it still remains questionable as it is mainly theoretical, with some support from limited genetics. As a consequence, the route of *Pv* introduction into the Americas remains a little unclear.

The strongest differentiation found in *Pf* mtDNA by the study by Schmedes *et al.* (2019), is the clustering of samples from the Philippines, Southeast Asia, Papua New Guinea and Vanuatu. As with *Pv*, this region appears both more strongly differentiated and more strongly diversified than most other regions. It could be an indication that there are strong selective factors at play across this region, and factors that impact both, regardless of age of introduction – be it climate, geography, human populations or vector species.

The focus of this project is on *Pv* as it is widespread across the Indo-Pacific region and believed to have been present there for a considerably long period of time (Groube 1993, Mu *et al.* 2005), possibly from the first known human migrations, estimated to be over 40k years ago (kya). It is historically the most prolific malaria species, with widespread human genetic adaptations, such as several versions of the

glucose-6-phosphate dehydrogenase (G-6-PD) deficiency gene spread across the globe (Mehta *et al.* 2000, Howes *et al.* 2013). The presence of another malaria-associated mutation, which produces Southeast Asian Ovalocytosis (SAO), found across the Austronesian population, indicates exposure and adaptation to malaria before migration, and the possibility of these human populations forming a source of proliferation (Rosanas-Urgell *et al.* 2012). Though there are many arguments against the ability of *Pv* to cross the Pacific, given time constraints and an absence of *Anopheles* spp. vectors for its further propagation, it is not an unsuitable migrant. Its ability to remain latently present in humans for several years, as so-called hypnozoites, concealed in the liver until circumstances trigger it to re-assert itself in the bloodstream, make *Pv* an opportune migratory disease and give it the ability to survive even in short-term absence of mosquitoes. As immunity to *Pv* is built up relatively fast (Mueller *et al.* 2009) it is generally less debilitating than *Pf*, and those infected with an active form can usually still travel and survive.

The environment across most of ISEA and Melanesia would have been suitable for a pre-Austronesian introduction of malaria. The Papua New Guinea highlands have been estimated to have developed forms of gardening from around 10,000 ya (Denham *et al.* 2003, 2008). Though the highlands themselves would have likely been unsuitable for malaria, such practices present in the coastal areas would have provided additional facilitation for mosquitoes and the spread of malaria well before the Austronesian expansion. The main malaria vector species in Melanesia, which includes members of the *An. punctulatus* complex, are thought to have evolved well before the arrival, and evolution, of the first humans (Logue *et al.* 2013, 2015). The tropical environment and prior presence of *Anopheles* mosquito species would likely have been sufficient for malaria to establish itself. Unlike *Pf*, *Pv* has been indicated to survive amongst small, isolated, migratory indigenous populations in Southern America (Ivo Mueller *pers. comm.* 2013), confirming its suitability as an early migratory disease.

1. 4. LARGE-SCALE DIVERSITY PATTERNS

1. 4.1 DISEASE GENETICS AND HUMAN MIGRATIONS

In recent years, there has been a global focus on the spread of disease through human trade, travel and other migratory activities. Humans represent an extremely widespread and well-connected species, with many migrants moving between populations. Numerous studies have recognised that these vast connection networks have had and could have serious implications for the spread of unwanted species, including diseases (Brockmann *et al.* 2006, Tatem *et al.* 2006, Tatem & Hay 2007, Colizza & Vespignani 2010, Bos *et al.* 2011, Chen & Vasilakis 2011). Species propagated through human movement, besides agricultural species for trade and subsistence, and those such as pets that are deliberately exchanged, include: commensals such as rats, mice and house geckoes; and parasites, diseases and disease vectors such as mosquitoes, ticks, yellow fever and hepatitis C (Moritz *et al.* 1993, Austin 1999, Duplantier *et al.* 2002, Matisoo-Smith & Robins 2004, Bryant *et al.* 2007, Searle *et al.* 2009, Hardouin *et al.* 2010, Delatte 2011, Short & Petren 2011, Tonione *et al.* 2011, Markov *et al.* 2012). The spread of some diseases that have migrated with(in) human beings goes back to the very origins of modern humans, such as the bacterium *Helicobacter pylori*, found in the stomachs of around 50% of the current human population (Moodley *et al.* 2012). Furthermore, studying the past of such stowaway diseases also gives an indication of the associated processes of invasion, adaptation and evolution of the disease and occasionally of humans (Lambrinos 2004, Facon *et al.* 2008). This knowledge contributes not only toward understanding the current dynamics behind the diseases studied, but also in predicting the spread patterns of introduced diseases in the future (McNeill 1998, Bryant *et al.* 2007, Markov *et al.* 2012, Moodley *et al.* 2012).

Most recent publications using disease genetics to trace human migration patterns originate from groups focusing on the genetic variation of diseases. A few of them focus on the human past that is likely associated with the observed patterns. One such study of great interest is the previously mentioned study by Moodley *et al.* (2009), which focuses on the ISEA migrations and their place in global human migrations through the DNA of the stomach bacterium *Helicobacter pylori*. Its results are a good indicator for what we expect to find for the phylogeny of *Plasmodium vivax* in a global setting, focusing on the Indo-Pacific migrations. The main discovery was a dichotomy in the genetics of *H. pylori* around ISEA, which is strongly linked to the two different populations found there. The first, associated with Papuans/Melanesians/Non-Austronesians, is linked to the African strains but with a split from other Asian *H. pylori* dating back to around 31,000-37,000 ya, indicating the different Asian strains diverged after humans first emerged from Africa, yet well before the time of the Austronesians migrating into ISEA. They show the presence of different strains across the Asia-Pacific region.

Figure 2 shows the historic movement connecting populations of *H. pylori*, including around the Asia-Pacific area. It shows splits in populations between Asia, East Asia, Papua/Melanesia and Austronesia. Due to the time of infection being restricted to the first weeks of life through mother to child transmission, the genetics of *H. pylori* is a strong proxy for human genetics. This can be seen by the correspondence with human migration estimates from human mitochondrial DNA, as published by Bae *et al.* (2017). Besides demonstrating past origins in the Melanesian area, the study by Moodley *et al.* (2009) goes further, showing broad migratory movements across the globe. The results from Bae *et al.* (2017) also show not all movements have been unidirectional, in contrast with the simplified version as given in Figure 2, with influence from populations in eastern Asia diffusing back through to Africa. The further from Africa, the less backflow there appears to be, possibly with oceans blocking/slowing much of the traffic for the Asia-Pacific and American zones.

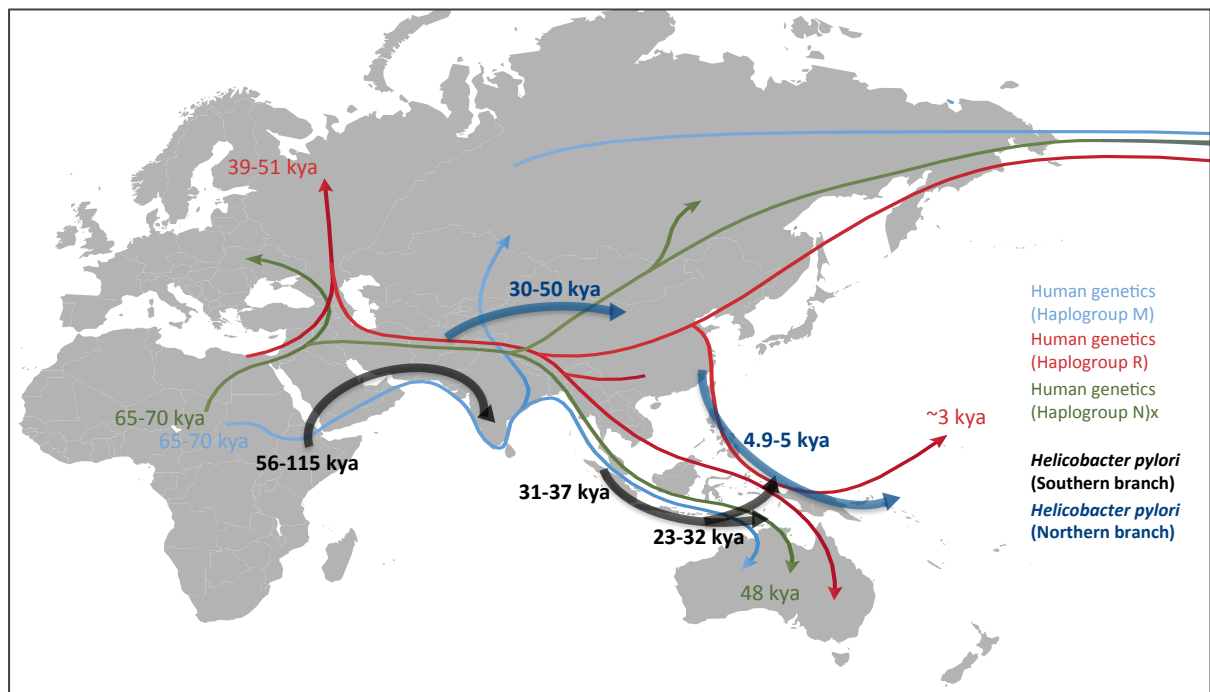


Figure 2: The broader global context of the phylogenetics of *H. pylori* and human migrations based on Moodley *et al.* 2009 and Bae *et al.* 2017. Migration directions of humans according to calculations by isolation with migration models for the genetics of eight different *H. pylori* population pools are given in dark blue and black, denoting the northern and southern routes respectively. Routes as found for the different human mitochondrial genome haplogroups are given in light blue (haplogroup M), red (haplogroup R), and green (haplogroup N). For *H. pylori* dates given are estimated ages of time to the most recent common ancestor of populations across all the regions migrated towards, following an eastward movement. It denotes the estimated earliest time of migrants settling in the region in question.

For *Pv*, a similar framework would be of great interest for understanding human Indo-Pacific migrations, and the resulting genetic variations within a global context. This project aims to understand the patterns of *Pv* distribution and explore the possibilities for using *Pv* in a similar fashion with an extended dataset of *Pv* mitochondrial genomes. An earlier study by Miao *et al.* (2012), involving the

mtDNA of *Pv* and human movements through Asia, indicates a comparable pattern present in *Pv* as seen in *H. pylori*. The paper concentrated on mtDNA diversity, but did not look into the driving forces causing it or estimated the times of introduction. In addition the dataset was relatively small with no or little representation from large geographic regions, including mainland Africa and Asia outside of East Asia and India. These regions are crucial to understand the past dynamics of *Pv* as they cover routes connecting Africa, the supposed region of origin, to East Asia and Melanesia, sites that have high diversity and are well represented in the *Pv* mtDNA dataset.

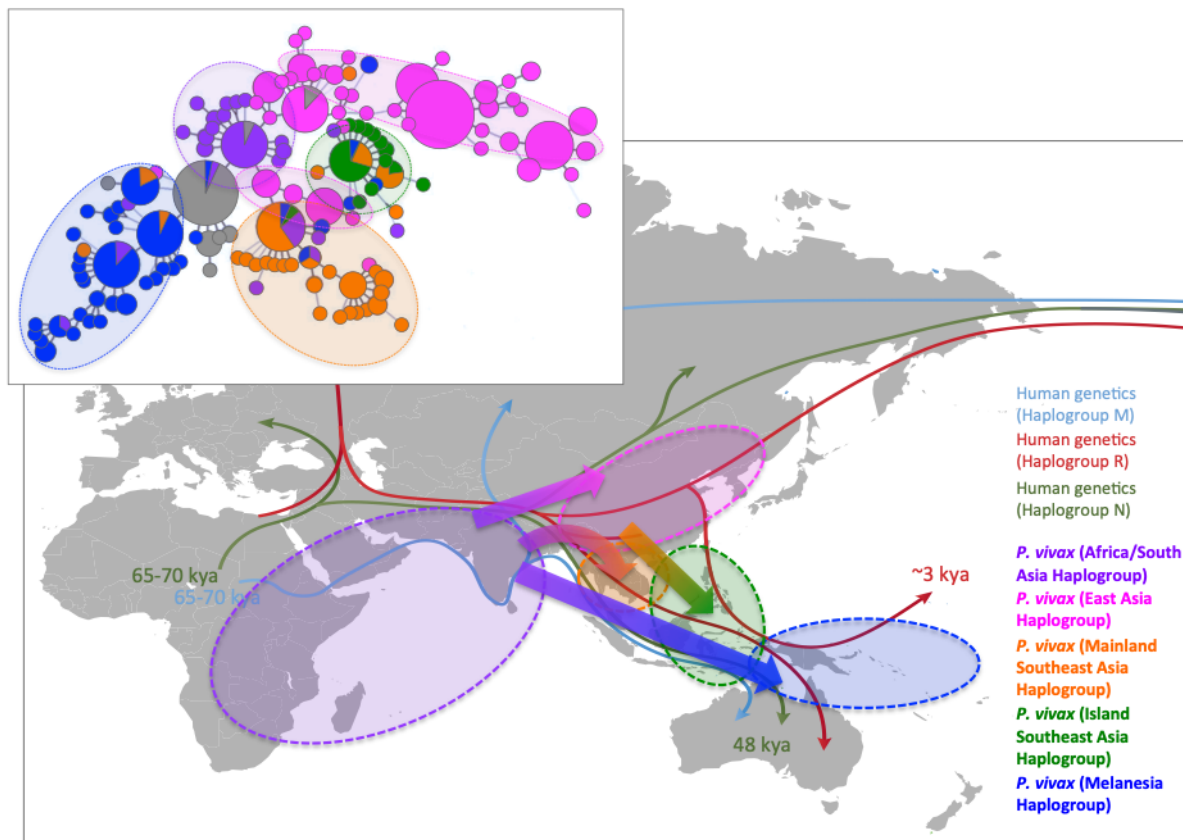


Figure 3: Map representing the main geographic regions for the main haplogroups as given in Miao *et al.* 2012: Fig. 2, and their connections, overlain on a map showing the main human migrations based on data from human mitochondrial genomes, based on that published by Bae *et al.* 2017. Insert: the haplotype network of *Pv* mt genome sequences by Miao *et al.* 2012, recoloured to match the map. Each circle represents a different haplotype, each colour the region from which a sample originates. The size of the circles indicates the number of samples with that haplotype, the smallest circles indicating a single sample. Each line is a mutational step.

Figure 3 shows a rough geographical representation of the results from the study of mitochondrial DNA (mtDNA) diversity of *Plasmodium vivax* across Asia, and its global connections, published by Miao *et al.* (2012). The figure demonstrates a high diversity in the Melanesian region (dark blue), a connection between the Americas and Melanesia (grey and dark blue haplotypes respectively), and connections between African and South Asian samples (purple) that are placed as intermediaries between East Asia, Southeast Asia and Melanesia. The latter connection is explained through migrations not only in the

deep past, but also more recent history, including Indian Ocean trade and later connections through Western colonisation.

Some of these findings overlap with our predictions as set out in the introduction. The differences are that the focus of the paper by Miao *et al.* (2012) is mainly on the genetic diversity found in China and is less concerned with the implications that the migratory dynamics behind the diversity might have for understanding the human past. Even so it remains an extremely useful pilot study for our purposes. Two other papers that provide key insight into the past of *Pv* through mtDNA are by Mu *et al.* (2005) and Culleton *et al.* (2011). Both these papers also show strong association between Melanesian and South/Central American samples, which is unsurprising as both the papers (Culleton *et al.* 2011, Miao *et al.* 2012) incorporate the data from Mu *et al.* (2005) into their study. Two more recent papers focused explicitly on the origin of American *Pv* (Taylor *et al.* 2012, Rodrigues *et al.* 2018). In combination with another paper containing data on an old Spanish sample (Gelabert *et al.* 2016) these studies have still not satisfactorily resolved the origins of American *Pv*, though there is increasingly strong evidence for an African/European origin, similar to that of *Pf*.

As with *H. pylori* it is possible that the two major human migrations to ISEA and Melanesia could have introduced distinct strains of *Pv*. The connection between *Pv* and humans is not as strong as that of *H. pylori*, making it more complicated to use as a direct proxy for human migrations. However, its greater malignancy means that the implications for its influence on exposed human populations and human genetic evolution are significantly larger, and therefore even a basic understanding of the past between humans and *Pv* is of great interest.

1. 5. SMALL-SCALE DIVERSITY PATTERNS

1. 5.1 POPULATIONS AND DISEASE

Besides the global large-scale patterns that have been discussed in the previous section, there is a plethora of smaller scale variation of human genetics and culture found in Melanesia. Current theories on drivers behind population-level diversity contribute towards understanding past interactions between humans and disease, both within the region and in a more general sense. The main focus of this section will be on the influence of disease, malaria in particular, on a population and vice versa. This includes activities by populations that influence disease levels (e.g. through changing their direct environment), as well as the influence of disease on human population dynamics.

Since Groube's (1993) publication on the influence of malaria on human population diversity in Melanesia, there appears to have been only sporadic mention of the theories he puts forward. Though studying the effects of disease on host populations has been an established area of research in the biological sciences for a while (Hamilton *et al.* 1982), Groube's theories have mainly resided within the spheres of archaeology and anthropology (Fincher & Thornhill 2008). Fincher & Thornhill (2008) discuss how infectious disease can act as a wedge between populations, stimulating isolation and thus diversification of populations over time. This has much in common with Groube's theories concerning the population diversification in the Pacific and the role of malaria. Though the theory is interesting enough to be considered, there are several reasons for it failing to have generated wider attention, the reasons for which are laid out in the following paragraphs.

1. 5.2 "PARASITE-DRIVEN WEDGE"

Groube's (1993) explanation for much of the human population structure found across the Melanesian region is the burden of malaria. As malaria is a disease that is relatively host density-dependent, small, separate human populations would have been better able to survive malaria, especially as large populations are subject to epidemics (Kermack & McKendrick 1927, 1932, 1933, Bailey 1957, Aaron & May 1982, Bailey & Norman 1982; all in Groube 1993 pg 170, also in Fincher & Thornhill 2008). Groube's theory follows the thought that whilst existing in relative isolation, these populations would build up a balance of immunity and infection with their local malaria strains. Each population would adapt to deal with slightly varying strains of malaria. Any migrants would run the risk of being exposed to a strain of malaria for which they had no immunity. The implication from this would be that populations gained an advantage in isolated existence, beyond contact with specific, close, communities (Groube 1993, Fincher & Thornhill 2008).

Groube calls this a “parasite driven-wedge”, where isolation between populations is increased over time through exposure and adaptation to local parasite strains. If there was movement between populations, this could have happened mainly at the time of marriage. Groube (1993) suggests that as pregnant women would especially run the risk of malaria becoming fatal during pregnancy, it is likely that any marital associated movements would involve residential mobility by the male partner, creating a matrilocal marriage system. Two areas of highest malaria infection in the world, West Africa and Island Melanesia, have a conspicuous predominance of matrilocal marriage in their local cultures, suggesting some possibility of truth in these conjectures.

Fincher & Thornhill (2008) predict that the areas with the highest levels of infections will have the highest levels of population diversity as well as smaller population contact networks, due to the host-parasite co-evolution of highly infected populations inducing isolation, similar to Groube's theories. Their measure of diversity is not genetics but language, a system that is likely to show diversification at an earlier stage, since it is more plastic than genetics. With these predictions in mind, it is interesting that the areas mentioned by Groube (1993) as having high malarial load, are also regions with extremely high cultural diversity. Island Melanesia is an especially strong case for this argument.

From the above it follows that some of the cultural diversity of the Melanesian region could be explained through disease limiting population densities. However, it is wise to approach these theories with caution. Research into strain specific immunity, and the relationship between pregnancy and malaria infection, actually shows a multi-faceted picture that is not as straightforward as what is assumed to come to the conclusions made by the researchers above.

Although there is evidence of local immunity to *Pf* in human populations in Melanesia (Alyssa Barry *pers. comm.* 2018) the argument needs to be approached with care. Besides strain specific immunity to *Pv* there is a build-up of strain transcending immunity (Doolan *et al.* 2009, Mueller *et al.* 2013). In lab tests, mainly for treating syphilis, patients were exposed to only a single strain of *Pv*. Subsequent exposure to different strains would make the patient very ill, though with a slight reduction of especially medical complications (Mueller *et al.* 2013, Snounou & Pérignon 2013, Doolan *et al.* 2009, Langhorne *et al.* 2008). These studies probably led to the early impression that immunity was strongly strain specific. However, once a patient has been exposed to several different strains, as is the case generally in natural populations, the strain transcending immunity becomes more pronounced, significantly reducing illness from exposure to most – if not all – strains within a species (Doolan *et al.* 2009). For *Pv*, the reduction in severe illness in humans built up from strain-transcending immunity can last up to 3.5 years (Mueller *et al.* 2013) and protection against clinical disease up to 30 years (Langhorne *et al.* 2008).

In regions with continuously high malarial infection, there would therefore be little extra risk in migrating between populations. Even so, the isolation of human populations in Melanesia whilst exposed to high levels of malaria has probably led to the marked diversity in genetic adaptations found in human populations, including a marked variety in G-6-PD (Howes *et al.* 2013) and human leucocyte antigens (HLA, Vina *et al.* 2012, Skoglund *et al.* 2016), and varying prevalence of SAO, Duffy negativity and Gerbich negativity across different Melanesian regions (Zimmerman *et al.* 1999, Patel *et al.* 2004, Rosanas-Urgell *et al.* 2012). For these adaptations to become prevalent would have taken several human generations (Hedrick 2011), and could have influenced genetic structuring in malaria populations too. However, this genetic structuring is expected to be present in other regions of the malaria genome and not the mitochondria. With its slow mutation rate, extremely long-term isolation would be necessary for mutations on the mtDNA to become saturated in a population, besides the presence of only three genes on the mtDNA that could be targets of selection pressure. Some patterns of connection and isolation of human populations on a shorter time-scale have been detected with microsatellites (Koepli *et al.* 2015, Fola *et al.* 2018).

The argument for matrilocality due to malaria also has little strength. In populations that are continuously exposed to malaria, the main adults at risk of dying from the disease are women in their first or second pregnancy. Research has moved on since the past assumptions that it was due to the natural immuno-suppression occurring in pregnant women, or that the placenta provides non-immunised shelter for the parasites (Fried *et al.* 1998). Amongst others, to find that though the placenta provides an extra organ for parasite sequestration in *Pf* infections (Beeson & Duffy 2005, Cowman *et al.* 2016), it is not necessarily non-immunised. The placenta itself increases sequestration opportunities for *Pf* infected mature cells, thereby facilitating increased disease. In addition, pregnancy also selects for parasites with a different binding mechanism for invading red blood cells, which also have different antigens expressed on the surface of infected erythrocytes (Rogerson *et al.* 2007, Cowman *et al.* 2016). Due to this, parasites and parasitised cells are generally unrecognised by the immune system, even if the patients have built up immunity to the normal *Pf* parasites from regular exposure (Beeson & Duffy 2005, Rogerson *et al.* 2007, Doolan *et al.* 2009, Cowman *et al.* 2016). Over subsequent pregnancies, the immune system appears to build up recognition of these parasites. According to this theory, all women – local or migrant – are at risk of increased illness during pregnancy from *Pf*. This means there is no perceivable extra benefit to a matrilocality-based society for areas with *Pf* infection.

The risk for increased illness seems mostly absent in *Pv* (Desai *et al.* 2007), or at the very least much reduced compared to *Pf* (Nosten *et al.* 1999) – possibly due to the lack of organ sequestration behaviour. In contrast, increased levels of maternal anaemia and lower the birth weight in the new-born remain present, for both *Pv* and *Pf* (Desai *et al.* 2007, McGready *et al.* 2012, Doolan *et al.* 2009). *Pf* can also lead to

stillbirths and spontaneous abortions, but this feature is argued over in *Pv*. Desai *et al.* (2007) claim no influence, however the data provided and studied by McGready *et al.* (2012) indicates similar rates of abortion for *Pf* and *Pv*. The difference here may be epidemiology. In women who were exposed to *Pv* from early childhood, vivax malaria does not seem to cause much morbidity in pregnancy. However, in non-immune women experiencing their first vivax malaria episode, the morbidity generally is higher in pregnant compared to non-pregnant women. McGready *et al.*'s population is more of the latter, non-immune, type. Even so, according to both *Pv* retains the negative influence on maternal anaemia and birthweight in multigravid women, more so than *Pf*. In short, though there is morbidity associated with infection, for *Pv* it does not appear to be much more significant for non-pregnant women. For *Pf*, there is increased risk, but, as far as is known, this risk is not dependant on strain, and disease presents even in immune women during their 1st and 2nd pregnancies. Leading to the conclusion that neither disease is likely to play a significant role in separating populations based on the influence it has on pregnant women.

From this, one may conclude that small-scale population separation caused by malaria only is unlikely. Even so, if the effects of malaria were combined with other factors, such as environmental influences, malaria itself could exacerbate already existing barriers between populations. This is believed to be the case in Papua New Guinea (PNG) between the highland and coastal human populations, where the strong separation is probably at least partially due to disease in general and malaria specifically. The area between an altitude of 500 and 1200m is plagued by unstable outbreaks of epidemic malaria. Other environmental factors contribute to the region being seen as a 'population sink' that cannot sustain human life without migrants from other areas (Michael Bangs *pers. comm.* 2012, Ivo Mueller *pers. comm.* 2013). Environmental factors complicating life in the area include high rainfall and steep terrain, complicating agricultural activities, but also disease pressure from both the previously mentioned epidemic malaria and the highland associated pneumonia. Thus this area forms a large and consistent barrier between the highland and coastal areas. It is not unlikely that this could have acted as a wedge, with populations on both sides adapting separately to widely different environments.

Purely environmental influences, such as the island geography of ISEA and Melanesia will have contributed to the diversity in the area, be it on a smaller scale. Geographic separation of populations can result in diversification through lack of continued contact. In biology this is termed the process of allopatric speciation, if it were to result in the populations diversifying into separate species. In the case of humans and culture, a term such as 'allopatric diversification' might be appropriate. The separate development of human cultures through geographic separation by oceanic barriers, however, does not completely explain the patterning of diversity found. The populations are known to have had the ability to cross oceanic barriers, and did so with purpose. It would be unlikely that they settled the Australasian

region if this was not the case, and genetic diversity would have been much lower if settlement had been through accidental crossings (Chappell 1993, Kirch 2000). The earliest Australasian settlers are estimated to have arrived in ISEA from Africa through a Southern Asian route from 60,000 ya to at least 40,000 ya. Even with lowered sea-levels at that time the first human settlers would have had to cross broad, deep sea-channels from Sunda, the Afro-Eurasian continent, to reach the enlarged continent of Sahul, which included a connected New Guinea and Australia (Chappell 1993, Kirch 2000).

Even if discounting the oceanic barriers, the ecology of the area is expected to have played a significant role in population diversification. On the large island of New Guinea, linguistic diversity is one of the most prolific in the world (Spriggs 1997, Kirch 2000, Pawley *et al.* 2005). The geography is varied, including barriers in the form of high mountains, large rivers and mangroves, and both geography as well as ecological variation have been shown to influence cultural and genetic diversity in the region (Stoneking *et al.* 1989, Friedlaender *et al.* 2008, Romaine 2013). Population variation is also increased by the Papuan/Melanesian versus Austronesian descent of the populations, especially as the populations had the opportunity to mix their cultures at different ratios depending on the input of each. With all these factors influencing population diversity, it is possible that the significant presence of disease and its role in selection can explain some of the gaps in understanding the high diversity of the region, but it is hard to tease apart from the multiple dynamics that contributed to it.

1. 5.3 HUMAN ACTIVITIES AND VECTOR-BORNE DISEASES

According to O'Brien and Laland (2012), humans impact their own evolution through influencing their surroundings. This includes the consequences of human activities affecting disease dynamics. In Boyd's book on malaria parasitology (1930), there is the observation that long-distance dispersal of malaria likely occurs through human migrations. In addition, human activities can influence malaria infection rates, and malaria transmission, in several ways. This includes the movement of new susceptible people into an area, increasing infectivity and parasitaemia transmission levels, which in turn exceeds the usual threshold of the local population and creates an epidemic. Alternatively the proliferation of vectors can occur through activities such as land development that increases the number of (stagnant) pools or sunlit pools, for example by obstructing natural drainage or cutting down trees.

One of the early migrations out of Africa is thought to have been connected to early agricultural land clearing (Salamini *et al.* 2002). The development of early agriculture is connected to large-scale population increase, which ultimately led to mass migrations (Cormier 2011). Clearings made in African forests through agricultural practices by humans led to a higher incidence of the mosquito *Anopheles gambiae*,

thus increasing the transmission of *Pf* malaria around 6,000 ya (Livingstone 1958, Tishkoff *et al.* 2001, Coluzzi *et al.* 2002, Hartl *et al.* 2002, Hartl 2004, Joy *et al.* 2003, Rich *et al.* 2009). According to Coluzzi *et al.* (2002) *A. gambiae* also became increasingly anthropophilic, whereas it was previously predominantly a livestock/wildlife feeder. Wilkinson (1994) argues that swidden agriculture, which creates open spaces and allows people to move into forested areas, may have been the main cause of the lowland Mayan civilisation's downfall through an increase of mosquitoes and yellow fever.

Other aspects of agricultural practices, such as stagnant water in cleared areas or flooded agricultural land for crops such as taro and rice, could enlarge the breeding grounds for mosquitoes (Livingstone 1958, Klinkenberg *et al.* 2003, Ohba *et al.* 2012). Alternatively, it has been found that in Southern Asia the more rigorous removal of forests in favour of plantations decreases the occurrence of the malaria-transmitting mosquito *Anopheles dirus* (Obsomer *et al.* 2007), with more dramatic decline of mosquito populations for urban areas. Such contrasting effects of human environmental modification on the increase or decrease of malaria exposure are due to distinct characteristics of particular mosquito species. Some, such as *A. dirus* are typical forest dwellers, whereas *A. sinensis* is commonly found in rice-paddies. An indirect effect of agricultural activities on human disease is that it facilitates for larger populations. The increased density of human populations provides a sustained breeding ground for infectious diseases (McNeill 1998, Oppenheimer 2012). The increase in human vectors within a relatively small area would have been a contributing factor to increasing malaria levels when early land-clearing practices increased mosquito vectors.

Not just agriculture, but other variations in lifestyle among human populations can have a significant effect on the disease presence and manifestation in different populations. This has been shown for Chagas' disease in South America (Coimbra 1988). The disease infects mainly humans in the highlands, but in the lowlands is a zoonotic disease. The difference is likely due to variations in lifestyle influencing host preference of the blood-sucking 'kissing bugs' that are the insect vector of this disease. Such variations include the type of housing, how much the population moves and the presence of domestic animals.

Overall the relationship between human activities, vectors and disease is intricate. ISEA and Melanesia form a good study area to study these relationship due to its environmental diversity and varied cultures, both now and in the past. Though the main *Anopheles* vectors in Melanesia are not strongly anthropophilic, and will readily bite domestic and wild animals, densities of domestic animals are rather low and make humans an accessible target (Ivo Mueller *pers. comm.* 2013).

1. 5.4 HUMAN INVASION, DISEASE AND IMMUNITY

Processes whereby human activities influence mosquito densities and thus disease transmission could have had significant repercussions on populations during the initial stages of such activities, comparable to introductions of novel diseases into previously unexposed populations (McNeill 1998). Newly infected populations often lack the necessary immunity to a well-developed disease, as well as missing any possible cultural adaptations. Impacts of such infections have been demonstrated by the introduction of small-pox on the Native Indian populations of the Americas during European colonisation (Snow & Lanphear 1988, McNeill 1998). In reverse, populations with established diseases could be hard to invade by novel, non-adapted, migrants (McNeill 1998). The latter scenario protects initial settlers, with the disease only facilitating for migrants that are in some way already adapted. This dynamic has been referenced in some cases as the main difficulty in invading and settling malarious regions in Africa and South Asia by Europeans.

Studies on many other species have shown that for species in general it is harder to invade an already established population, especially one that is occupying the available resources, if other things are equal (Stachowicz *et al.* 2002, Hardouin *et al.* 2010). Few studies have looked at what the effect would be if - instead of the invading population - the original population had a natural weapon in the form of disease, in addition to having the advantage of being the initial settlers. The lack of attention for such a 'native over settler' point of view could be due to the more recent experience of Western history being involved with large-scale colonisations. Western colonisations still form a demonstration of this theory, as in contrast to replacing the indigenous populations in Northern America and Australia, they did not manage to settle in significant numbers in the tropical, malarious, regions of Africa and Southern Asia.

The reverse scenario of advantaged early settlers could well be the situation for the Australoid (Papuan) populations colonising Sahul, particularly if they brought malaria with them. The barrier of disease, besides the geographical barrier of the ocean, might provide an explanation for why migrants that populated Asia failed to settle the Melanesian areas to a noticeable degree until the Austronesian influx. A barrier formed by infectious disease could have remained in place until a sea-faring population adapted to the disease arrived and settled the fringe areas. Adapted migrants, especially those bringing in a new strain or species of malaria affecting the earlier settlers, would have less trouble settling in the infected regions (for example *P. vivax* infected and adapted Austronesians, bringing genetic mutations such as SAO with them). As novel exposure to the disease would still affect the settlers, uncolonised areas are likely to have had a preference as migratory destinations.

Groube (1993) discusses an alternative scenario besides the ones mentioned above. As the dynamics of malaria are very much dependent on the population structure and immunity of the host populations, an

imbalance caused by immigrants could affect the colonised populations to such a degree that they are at least as disadvantaged as the migrant populations. The dynamic is as follows: unimmunised migrants enter the area, they get infected and increase the density or prevalence of infection. With a constant presence of migrants, levels of parasitaemia are raised to above the threshold levels of what the local populations are used to, triggering an epidemic that affects migrants as well as locals. This is also known as the ‘spillback’ effect, as opposed to the ‘spillover’ effect of new migrants bringing novel diseases with them (Kelly *et al.* 2009). Constant influx of migrants will bring non-infected migrants into the area, whilst most of the local population will be suffering from malaria. There is good reasoning behind this process described by Groube, but it is difficult to see how the migrants would have the advantage. Any new migrants moving into an area exposed to holoendemic malaria, let alone a raging epidemic, are unlikely to avoid being infected, which removes any advantage. At best the playing field would be level. Advantage is given to those with a more rapid development of immunity to the disease, i.e. those that already have an existing level of immunity, or those that have experience of how to deal with the disease. Behavioural adaptation to prior exposure could include knowledge of local cures, if not an already adapted way of life. An example of how rapidly immunity is acquired by non-exposed migrants moving into a holoendemic malarious region is given in the study by Baird *et al.* (2003). Javanese migrants moved into West Papua, suffering severe malarial infections at first exposure, comparable to that of indigenous children. These migrants rapidly gained immunity, but around 7.5% of the infected migrants had to be evacuated due to the severity of the disease, indicating that non-exposed migrants would have a severe disadvantage when entering a malarious region.

1. 5.5 MALARIA AND HUMAN GENETIC ADAPTATION

Besides high diversity in general human genetics, the region of focus in this study is also rich in genetic mutations related to malaria adaptations (Serjeantson *et al.* 1992, Howes *et al.* 2013). Several adaptations are exclusive to the Melanesian area, or local even within certain areas of Melanesia, suggesting high levels of population isolation. However, there is evidence of enough contact between populations to spread the most effective genetic adaptations against the disease (Serjeantson *et al.* 1992). The ages of several human genetic adaptations to malaria have been estimated by different studies; an overview of this is given in *Table 1* (Hedrick 2011). These date estimates vary depending on the study, due to differences in assumptions made regarding the basic values used in the calculations, such as the chances of survival in malarious areas of individuals with adaptive mutations versus those without (the selection coefficient) or reproductive rates of those that survive (Hedrick 2011). As demonstrated in *Table 1*, small variations in these assumptions can give substantially different results. Additionally there is substantial variation in the standard deviation estimates within studies. More rigorous methods would be useful and currently anything based too heavily on precise dates taken from these calibrations would seem questionable.

However, even with the variability of the estimates of times of origin it remains clear from most data in the table that populations in ISEA have long been exposed to malaria infections.

There is evidence for a link between malaria and Austronesian migrants through the presence of the G-6-PD deficiency gene in Vanuatu populations (Kaneko *et al.* 1998). The frequency of the G-6-PD deficiency variant (allele) in populations across the different islands and regions of Vanuatu corresponds strongly to the level of malarial infection present in each area. The areas with a high malarial load contain human populations with a higher frequency of the G-6-PD deficiency gene than those areas with low or no malaria presence. All areas, including those that have had little to no gene flow between them, contain the same G-6-PD deficiency mutation. This suggests the humans came into the area carrying the variant, and those populations still carrying it have been continuously exposed to malaria (Kaneko *et al.* 1998).

Table 1: Overview of the estimated age of alleles in generations and years (for a 25-year generation length) that confer resistance to malaria. *Table after Hedrick 2011, pg. 289: Table 2.* Selection coefficients with an * are assumed values, not estimated, and — indicates no selection (neutrality) is assumed.

Gene allele	Age		Selection coefficient (s)	Reference
	Generations	Years		
<i>b-globin</i>	S	45–70	1125–1750	0.152*
		10–28	250–700	—
	C	75–150	1875–3750	0.04–0.09
		38–120	950–3000	—
	E	100.3 (62–222)	2508 (1550–5550)	0.079 (0.035–0.099)
<i>G6PD</i>	A-	254.3	6357 (3840–11 760)	0.044
	A-	93	2325 (1200–3862)	—
	A-	1800	45 000 (25 000–65 000)	—
	A-	40	1000	0.25
	Med	133.2	3330 (1600–6640)	0.034
	Med	400	10 000 (0–35 000)	—
	Mahidol	63	1575	0.23
<i>Duffy</i>	ES (null)	1323 310 490	33 075 (6500–97 200)	—
	ES (null)	310	7750 (3625–13 125)	—
	ES (null)	490	12 250 (4250–26 500)	—
<i>ABO</i>	O01	—	1.15 million	—
	O02	—	2.5 million	—
<i>HLA-B</i>	B53	100	2500	0.028*
	B53	86	2150	0.041*

Other evidence for a link between malaria and a specific group of migrants in the Asia-Pacific region is the occurrence of a *Pv* resistance variant associated with Southeast Asian Ovalocytosis (SAO), in certain ethnic groups (Mgone *et al.* 1996, Patel *et al.* 2004). The mutation mainly protects against *Pv* infections, yet in addition it appears to help reduce the severity of *Pf* induced cerebral malaria (Genton *et al.* 1995, Rosanas-Urgell 2012). The SAO mutation is lethal in homozygous individuals (Genton *et al.* 1995), which argues that to remain present in the population there must have been a high incidence of *Pv* malaria. Through having a long-term selective advantage in the heterozygous individual, such mutations with negative consequences will be maintained in the population (Livingstone 1958, 1964).

A possible consequence of contact between two populations carrying different adaptive variants with adverse health effects is that the mutations can accumulate in offspring with mixed ancestry. Instead of conferring an advantage, the decreased fitness of mixed offspring could create a barrier to population admixture, besides any cultural barriers, as mixed progeny would have lower chances of survival to an extent that outweighs any relative advantage gained over malaria due to accumulation of deleterious mutations. Where separated populations adapted to malaria in different ways such a risk would be an added barrier.

1. 6. CONCLUSION

This chapter has demonstrated that the human cultural and genetic diversity found across the Eastern Hemisphere, and especially within ISEA and Melanesia, is complex and includes a diverse array of contributing factors. Whilst substantial research has been conducted on the topic of *Pv* mitochondrial diversity, there is still much left unexplained and there have been no comprehensive studies looking into several influencing factors at the same time. Much of the past dynamics behind the distribution of *Pv* and its connections with the human past has been left unexplored. Though the complexity and scale of the area tackled in this thesis are considerable, it is precisely this diversity that will make for interesting comparisons and help gain a more thorough understanding of factors impacting the interactions between humans and malarial disease. In addition it will aid in identifying avenues for future research into parasite and human evolution combined.

To such an end this project aims to augment a thorough comprehension of the dynamics behind the diversity of *Pv* in the Eastern Hemisphere. There are several factors that have been shown to influence genetic variation on smaller geographic scales or are expected to have a role in the distribution of *Pv*. *Chapter 2* explores the distribution and history of, and interaction between, the two vectors on a global scale. Studying each vector separately without accounting for the influence of the other could overlook important interactions or lead to misinterpretation of data patterns. The expectation is that *Pv* distribution follows human migration and settlement patterns, possibly dating back to the emergence of *Homo sapiens* from Africa, though the emergence could have occurred earlier, with other primates, or later. As strong influence of mosquito vector species on the *Pv* genome has been found in recent publications, the expectation is that patterns in human population distribution could be influenced by alternative underlying patterns in anophelene species.

Chapter 3 takes a closer look at the influence of past human migration dynamics on the mtDNA variety of *Pv* within a smaller regional context. It looks at the origin of *Pv* in Melanesia, set within the context of the human past both locally and in the surrounding regions. As previous studies show a high diversity of *Pv* in this region there are two alternative, yet compatible, hypotheses. Firstly, the introduction of *Pv* into Melanesia occurred very early on in time, for example with the early *H. sapiens* Papuan populations, and as such it has had ample time to diversify. Alternatively, or additionally, there were several introductions of *Pv*, possibly from different origins such as from both Papuan and Austronesian populations. Attempts at studying smaller scale dynamics in this region failed due to the low mutation rate of mtDNA being unable to pick up on more recent historical dynamics.

The functionality and remaining use for the mitochondrial genome for *Pv*, and malaria in general, with the introduction of whole genome sequencing is the focus for *Chapter 4*. The chapter reviews the use of different genetic sequences for varying research purposes in the past, with a strong emphasis on research in malaria phylogenetics and population structure. As whole genome sequencing is still relatively slow and costly - both resource and time wise - the expectation is that the mitochondrial genome, and its large publicly available database, will still remain a useful tool. In addition, the knowledge garnered on the genome throughout the past years, including an understanding of its relatively consistent and stable mutation rate across the genome, is anticipated to retain its value especially when looking into the past of *Pv*. The chapter also explores the possibility of combining mtDNA data with data from microsatellites as combining data from mtDNA with that of other parts of the genome could increase resolution in phylogenetics and population dynamics, leading to greater insight over slightly smaller time periods than that covered by the mitochondrial genome.

Overall this thesis attempts to answer how *Pv* and its diversity and distribution are connected to the human past. It assesses the value of using the mitochondrial genome as a tool to answer this question and explores what and how other dynamics influence the connections seen. Besides providing an increased understanding of the past of *Pv*, comprehending the influence of these dynamics could provide helpful insights for future research into the past of humans, their migration dynamics and their adaptation to the disease.

INDEX CHAPTER II

<u>2. 1. INTRODUCTION</u>	<u>39</u>
<u>2. 2. GENETIC STRUCTURE</u>	<u>41</u>
2. 2.1 INTRODUCTION	41
2. 2.2 METHODS	42
2. 2.3 RESULTS & DISCUSSION	45
<u>2. 3. HUMAN MIGRATION</u>	<u>53</u>
2. 3.1 INTRODUCTION	53
2. 3.2 METHODS	57
2. 3.3 RESULTS & DISCUSSION	59
<u>2. 4. GEOGRAPHY</u>	<u>63</u>
2. 4.1 INTRODUCTION	63
2. 4.2 METHODS	64
2. 4.3 RESULTS & DISCUSSION	65
<u>2. 5. MOSQUITO VECTOR</u>	<u>67</u>
2. 5.1 INTRODUCTION	67
2. 5.2 METHODS	71
2. 5.3 RESULTS & DISCUSSION	73
<u>2. 6. OVERALL RESULTS & DISCUSSION</u>	<u>75</u>
<u>2. 7. OVERALL CONCLUSION</u>	<u>80</u>

CHAPTER II

DISTRIBUTION OF *PLASMODIUM VIVAX* MALARIA MITOCHONDRIAL GENETICS AND ITS DRIVERS ACROSS THE EASTERN HEMISPHERE

*“Our encounter is a mathematical formula,
[...] the way of the universe
The evidence of fate given to me” – BTS, DNA*

2. 1. INTRODUCTION

Malaria is a globally widespread disease, with a strong presence in tropical regions. One of the five human infecting species, *Plasmodium vivax* is geographically most widely spread, occurring in both tropical and temperate regions. Until the early 20th century it infected populations across Europe, including the Netherlands and the Mediterranean, and is still present in more northern latitudes such as Japan and the Korean Peninsula (Deeks 1925, Boyd 1930, Gething *et al.* 2012, Verhave 2013). *P. vivax* is currently not only the most widely spread species of human infecting malaria, except for *Plasmodium falciparum* it is also the most prevalent. This is expected to change, however, as *P. vivax* is proving exceptionally hard to eradicate, still maintaining high levels of diversity in areas of intensive control efforts (Gunawardena *et al.* 2014, Delgado-Ratto *et al.* 2016, Waltmann *et al.* 2017). The increased recognition of the persistence, prevalence and societal impact of *P. vivax* has augmented recent interest in studying its history, biology and epidemiology. Significant progress in these areas of research has been aided by the discovery and sequencing of vivax-like parasites in African apes in 2011 (Rayner *et al.* 2011, Culleton *et al.* 2011, Prugnolle *et al.* 2013). The discovery of these African parasites has pushed back the estimated age of origin for *P. vivax*, and put its likely place of origin in Africa. However, the origins, spread and evolution of this parasite currently remain less well understood than *P. falciparum*.

The wide distribution and ancient origins of *P. vivax* enhance the interest in its evolutionary dynamics and epidemiology over the long-term, especially as these dynamics are vital for estimating the impact of the disease on its human hosts. This chapter elucidates the patterns of genetic diversity in the *P. vivax* mitochondrial genome across the Afro-Eurasian continent and Pacific with the help of an increased mitochondrial genomic data-set for *P. vivax*, which fills important geographic gaps in the currently available data including the addition of substantial sample numbers from mainland Africa and Melanesia, and new samples for western and central Asia. It also explores three possible driving factors behind these patterns: geographical distance, mosquito vector species, and past and present human migratory movements. Understanding how

each factor has helped shape the distribution and diversity of *P. vivax* gives insight into the diversification of the disease over time.

Though factors influencing the global distribution pattern and diversity of *Plasmodium vivax* malaria (*Pv*) are starting to be better understood, research usually focuses only on one specific factor (e.g. drug resistance, geographic separation, vector species differences, migration) and/or region (e. g. East Africa, India, China, South America, Thailand). Studying only a small part of the bigger picture can lead to data misinterpretation or neglect of interactions that enhance or dampen patterns observed. For example, previous studies have shown a strong geographic patterning in *Pv* mtDNA, up to the point of using this feature to deduce origin of infection in migrant/traveller patients (Rodrigues *et al.* 2015). The geographic structuring has also been used by many studies to deduce the origin and phylogeny of *Pv* in certain regions, e.g. in the Americas (Taylor *et al.* 2012, Rodrigues *et al.* 2017) and China (Miao *et al.* 2012, Mu *et al.* 2005). However, most of these studies look at these regions with a bias towards relatively recent human activity (<1000 years ago), and do not take into account, or only in a minor way, other contributing factors to the genetic structures seen. Due to the regional focus of most of these publications, the general distribution of *Pv* mtDNA across large sections of the world has received mainly summary attention. In contrast, in the past two years much progress has been made in attempting to understand the distribution of *Pv* microsatellite (MS) data on both global (Koepfli *et al.* 2015) and regional (Delgado-Ratto *et al.* 2016, Fola *et al.* 2017, Waltmann *et al.* 2017) levels, as well as in understanding patterning in whole genome sequence (WGS) data (Pearson *et al.* 2016, Hupalo *et al.* 2016). A comparison across these different forms of genetic data will be discussed in depth in *Chapter 4*.

So far, however, large-scale in-depth studies integrating the three main factors that affect *P. vivax* distribution - vector distribution, human migration and geography - have not been done for the mitochondrial genome itself. This chapter explores how much the influence of each can be observed in the mtDNA distribution and how the three could have influenced each other. With the increased size of the mitochondrial dataset and the addition of samples from previously unrepresented, or under-sampled, regions, the opportunity has been created to look at the distribution of *Pv* across a much larger geographic scale. This analysis focuses on the distribution of *P. vivax* in the Eastern hemisphere. Previously available data have been augmented with samples from Sudan, Armenia, Azerbaijan and Uzbekistan, which fill important geographic gaps in the data. An attempt was made to add data from West Africa, but the low endemicity of *Pv* across this region and across the samples received led to the extraction of only half a mitochondrial genome. Even so, the new data add significantly to what was available to the west of Pakistan. The genetic variation and history of *P. vivax* in the Americas has been discussed in other publications (Taylor *et al.* 2012, Rodrigues *et al.* 2017), and since the complex but likely recent patterning in the American data may mask the patterns seen across the African, Eurasian and Pacific regions, samples from the Americas have not been included in this study.

2. 2. GENETIC STRUCTURE

2. 2.1 INTRODUCTION

In the past decade an increasing number of papers have been published studying the genetic structure and distribution of *P. vivax* populations (e.g. Mu *et al.* 2005, Culleton *et al.* 2011, Miao *et al.* 2012, Taylor *et al.* 2012, Koepfli *et al.* 2015, Waltmann *et al.* 2017, Fola *et al.* 2017). Regional structuring has been found in nuclear genome microsatellites (MS) (Koepfli *et al.* 2015), the mitochondrial genome (mtDNA) (Mu *et al.* 2005, Culleton *et al.* 2011, Miao *et al.* 2012, Taylor *et al.* 2012) and across the full parasite genome in nuclear single nucleotide polymorphisms (SNP) (Pearson *et al.* 2016, Hupalo *et al.* 2016). Hupalo *et al.* (2016) identified the parasites' adaptation to human and mosquito, in addition to more recent selection pressures from antimalarial drugs, as driving causes behind the observed SNP variation.

Previous studies indicate that though little variation is present on the *Pv* mtDNA, the existing variation is strongly regional (Mu *et al.* 2005, Culleton *et al.* 2011, Miao *et al.* 2012, Taylor *et al.* 2012). Regional fingerprints are strong enough to broadly identify the original region of infection of a patient with *Pv* from the mtDNA on a continental scale (Rodrigues *et al.* 2015). An exception to this exists for the area spanning across Africa to South Asia, where *Pv* has remarkably low genetic variation (Taylor *et al.* 2012, Hupalo *et al.* 2016). The homogenous distribution of *P. vivax* strains across this vast region has been observed in the mitochondrial genome for a decade or more, and its occurrence has been confirmed in the nuclear genome SNPs by Hupalo *et al.* (2016). The lack of variation in African *Pv*, amongst other factors, contributed to the view, held until recently, that the geographic origin of *Pv* was situated in Southeast Asia (Perkins & Schall 2002, Escalante *et al.* 2005, Mu *et al.* 2005).

With the identification of Africa as the origin of *Pv* instead (Rayner *et al.* 2011, Culleton *et al.* 2011, Prugnolle *et al.* 2013, Liu *et al.* 2014), interest in the dynamics behind the observed genetic structure has increased. As a start, this section introduces and examines the differences in *Pv* mitochondrial genetics across the eastern hemisphere utilising several different methods. With the addition of samples from new regions situated between Asia and Africa and increased sample numbers for mainland Southeast Asia and Melanesia, this section aims to confirm and add more depth and understanding to previously observed distribution patterns.

2. 2.2 METHODS

P. vivax isolates were sequenced from archival samples of infected human blood, sampled across sites in Papua New Guinea (PNG, 126), the Solomon Islands (SOI, 17), Vanuatu (VAN, 3), Sudan (SUD, 33), Armenia (ARM, 4), Uzbekistan (UZB, 9) and Azerbaijan (AZB, 1). Originally these samples were taken for epidemiological and population genetic studies, as published in Arnott *et al.* (2014), Koepfli *et al.* (2012, 2013, 2015), Waltmann *et al.* (2015), Brega *et al.* (2004), Leclerc *et al.* (2004) and Menegon *et al.* (2010, 2014). Human research ethics clearance to conduct research on these archival samples for this study specifically was obtained from the Walter & Eliza Hall Institute's (WEHI) human research ethics committee (HREC) in Melbourne, Australia, under WEHI-HREC project number 14/14. For all samples the subject of this study fell within the consent obtained from patients, as well as within the parameters of previous, local, human ethics permissions. Around 1400 infected blood samples from Mali were received from collaborators; however, only two of these showed *Pv* infection, one with a parasitaemia too low for successful amplification. From the second only half a sequence was obtained. Despite these efforts to increase the dataset for African *Pv* mtDNA samples, only samples from Sudan produced usable sequences.

DNA from the samples was purified across several different labs with varying methods, mainly including extractions with QIAamp DNA Blood Mini Kits (Qiagen, Valencia, CA). Genotyping of highly variable nuclear size polymorphic markers (including: MSP2, MS8, MS16, MS20, Pv3.27) and merozoite surface protein 1 (*msp1F3*) identified single-clone infections. From these, the Melanesian mitochondrial genomes were sequenced (n=150, from PNG, VAN & SOI), according to a modified two-fragment PCR protocol based on that by Culleton *et al.* (2011). The PCR amplifications of 3 and 3.5kb were accomplished with Roche Fast Start High Fidelity enzyme (Roche Diagnostics AG, Switzerland), according to the provided Roche protocol for long amplifications: 2 minutes at 94° C, 10 cycles of 30 sec at 94° C, 1 minute at an optimised annealing temperature of 61° C, 3.5 min at 68° C, and then 30 cycles adding 20 seconds to the extension time per cycle. Final concentrations of the reaction included 1x buffer with MgCl as provided by Roche, 0.25mM dNTPs, 2.0U enzyme, and 0.5uM of each primer. Primer sequences and a protocol overview can be found in *Appendix S2.1*.

Sequencing of the samples from Sudan, Azerbaijan, Armenia and Uzbekistan (n = 47) used a variation on the above method, as the two-fragment protocol was less effective. Failure was possibly due to increased fragmentation of the substrate DNA. Instead, a ten-fragment amplification protocol was used. For seven out of ten fragments the protocol utilised the same mastermix formula as the two-fragment amplification. The thermocycle was adapted for shorter fragments to 40 normal cycles without added time, the elongation step altered to 72°C for 80 seconds, and the final elongation at 72°C for 7 minutes. The three remaining fragments used Solis Biodyne FirePol Pink enzyme (0.75U) and 1.5-2.5mM MgCl, besides the concentrations of buffer, dNTPs and primers from the two-fragment protocol. The thermocycle ran at 94°C for 15 minutes, 30 cycles of 30 sec

After alignment of all sequences, around 90bp was removed from the end (bp 5847-5991, *see Appendix S3.9*). The removed sequence was missing in several previously published sequences, and did not show any variation in the other sequences used. All gaps due to varying string lengths of single nucleotides were removed. This decision was made due to the high likelihood of both sequencing and DNA replication errors in such areas, making them unsuitable for phylogenetic inference. In addition, methods incorporating insertions and deletions (indels) into substitution models are still a point of contention with the current version of the *BEAST2* (Bouckaert *et al.* 2014) software employed.

The expanded dataset of all obtained *P. vivax* mitochondrial sequences from sample sites across the Pacific, Asia and Africa were grouped according to country and several different combinations for geographical regions of varying sizes, as shown in *Appendix S2.2*. These groups were used for analysing population dynamics and phylogenetic relationships. For population diversity and differentiation estimates, such as H_s and Tajima's D , country level categories were used (*Appendix S2.2*, column 'Country 2B S#>1'). Some samples were removed from the dataset, either because their geographic location was too broad to fall within one location grouping (South Asia (n=1), Africa (n=1)), or because the sample size was too small (Afghanistan (n=2), Iran (n=2), Namibia (n=1), Philippines (1), Sri Lanka (1)).

Polymorphism data across the entire mitochondrial genome for all samples was calculated with *DnaSP* and single nucleotide polymorphism (SNP) position data was visualised with the R package *adegenet* (Jombart *et al.* 2010). The same package was used to calculate within population expected heterozygosity (H_s) for each population. Results are displayed with population category names and sample size number in *Table 2*. Interpopulation differentiation was calculated from sequence data with the R package *hierfstat* (Goudet 2005). The value for population genetic variation chosen were the 'chord distance' values, D_{CH} , taken from Cavalli-Sforza & Edwards (1967), and found to be the best formula for calculating the relationship between populations by Takezaki & Nei (1996). The values can be found in the table in *Appendix S2.5*. These values were used to generate a heatmap and dendrogram, illuminating the relatedness between the different populations, in *Appendix 2.14*.

To better understand and visualise the differences in geographic variation between the different genomes, a principal component analyses (PCA) was generated with R (*ade4*), shown in *Figure 5*. A certain amount of 'jitter' was added to the data points (factor = 0.5, amount = 0.4) as many of the sequences were exactly the same and overlapped on the PCA figure. As Reich *et al.* (2008) argue in their review of the uses of PCA, it is useful for detecting finer structures in populations, working around confounding factors in epidemiological studies and inferring species' history especially concerning large migration events. An alternative method to calculate principal components, called the Discriminant Analysis of Principal Components (DAPC, in R package *adegenet*), was also used (Jombart *et al.* 2010). This method identifies and describes clusters of genetically related individuals,

and generally outperforms the much-used STRUCTURE analyses (Jombart *et al.* 2010) in showing relatedness, as well as being faster than Bayesian calculations.

Phylogenetic relationships were estimated using a Bayesian Markov Chain Monte Carlo framework in *BEASTv2.1.3* and *Tracer v1.6* (Rambaut *et al.* 2014; see *Appendix S2.3*). For the main model settings, a Coalescent Bayesian Skyline study model, HKY (Hasegawa *et al.* 1985) substitution model and strict clock branch rate model were used, with no rate of base pair invariance (I) and variance across sites (gamma) between 4 and 12. Even at 12 the *gammaShape mean* value remained close to zero, indicating a high gamma estimate, indicating variance across sites is high. Initial runs were based on jModelTest (Darriba *et al.* 2012) suggestions, but these were later simplified to conform more closely with the type of data and analyses used, taking into account that jModelTest is not the optimal tool for finding the correct model for Bayesian formats. See *Appendix S2.3* for further details.

Clock rate was set at ~0.00398 substitutions per site per million years (sub/site/Ma), the average median rate estimate from previous publications (see Comas *et al.* 2013: 0.00189-0.00600 & 0.00211-0.00601, Cornejo & Escalante 2006: 0.0032–0.0043, Mu *et al.* 2005: 0.0045–0.0056, Jongwutiwes *et al.* 2005: 0.0026–0.0036 substitutions per site per million years). Both Comas *et al.* (2013) and Mu *et al.* (2005) calculated their rates by calibrating a divergence time of 6-14.2 Ma (Pacheco *et al.* 2012) between a clade of *P. gonderi* and *Plasmodium sp.* from mandrills, with *P. vivax* and its nine closest Asian *Plasmodium sp.* relatives. Jongwutiwes *et al.* (2005) calibrated rates to *P. falciparum* and *P. reichenowi* host divergence dates, human and chimpanzee respectively, with *P. knowlesi* and *P. vivax* as outgroups, and using strong evidence for constant clock rates. Comas *et al.* (2013) used a more rigorous method to calculate rates, independent of assumptions about *P. vivax* host relationship with that of hosts of closely related strains, using one mitochondrial genome from each of the eleven species of primate-infecting *Plasmodium sp.* in a birth-death model in BEAST. As the other rates were very similar and the methods not seen as inherently wrong, the mutation rate estimate for this chapter was based on all four publications. These are currently still the standards for calculating the *Plasmodium* mitochondrial genome clock rate, and so the decision was made to use a similar rate in this paper as the starting estimate for the molecular clock, without the need for repeating the same calibration methods. Using this rate also meant the generally problematic addition of an outgroup could be avoided. Flexibility of the rate was set to vary within 95% of the highest and lowest rates as calculated in the literature (0.00189 – 0.00601 sub/site/Ma respectively), or a factor 10 faster or slower than the set rate.

2. 2.3 RESULTS & DISCUSSION

Across the entire mitochondrial genome there is relatively low variation, as is seen in *Figure 2* and in the polymorphism data in *Table 1*, showing only 309 variable sites and 326 mutations over more than 5500 nucleotides in 818 samples. *Figure 2* shows that polymorphism density is generally

similar across the genome, with the exception of one region of heightened SNP density, and one region of very low SNP density. These both fall outside the protein-coding regions, but much of the mtDNA genome outside the protein coding regions is comprised of codes for ribosomal RNA and DNA (rRNA and rDNA respectively), including fragmented codes that assemble after transcription, as found in *P. falciparum* (e.g. Feagin *et al.* 1997).

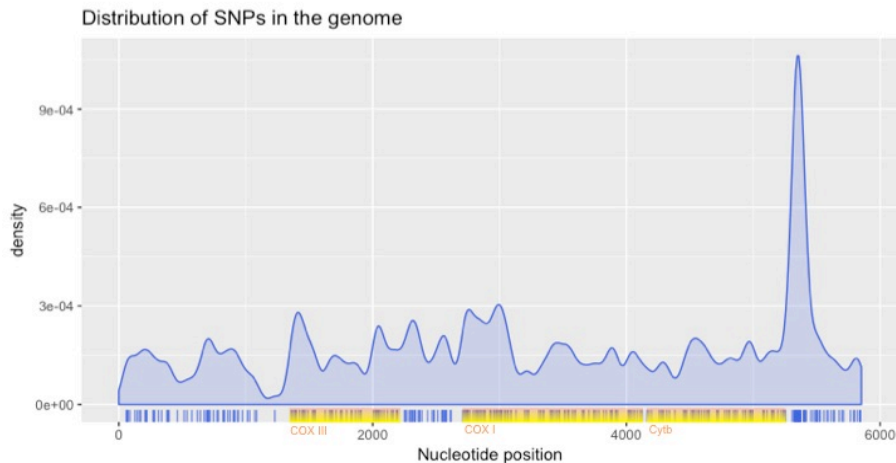


Figure 2. *P. vivax* mitochondrial genome distribution of single nucleotide polymorphisms
Overview of locations of polymorphisms on the mt genome (x-axis, dark blue) and the density of SNPs for each region (y-axis). The three gene coding sequences located on the mt genome are coloured (yellow, COXIII, COXI and Cytb from left).

Sample number	N	818
Number of variable sites	S	309
Total number of mutations	Eta	326
Number of haplotypes	h	352
Haplotype diversity	Hd	0.977
Hd variance	Hd variance	0.000
Hd standard deviation	Hd std dev	0.000
GC content noncoding sites	G+Cn	0.374
GC content coding sites	G+Cc	0.254
Total G+C	G+C	0.304

Table 1. *P. vivax* polymorphism data
Polymorphism data for the mitochondrial genome data set.

Most rRNA and rDNA codes on the mtDNA are much more conserved than the protein-coding regions, but a few others much less so. The spike in polymorphisms, seen towards 5300bp in *Figure 2*, is caused by a few different factors: 1) the three samples from Borneo miss 42bp from 5336-5377. Some extra mutations occur along the same stretch in other samples – mainly single mutations in one or at most two individuals; 2) little prior to that section, a heightened number of mutations occur within a span of three base pairs. One of these mutations is geographically specific to a sub-group of Cambodian samples (variant frequency: 0.5% overall). Another is present in a limited set of samples, mostly from Myanmar, East Asia (China and Korea), and a limited number of Melanesian samples (13.2% variant frequency). The third mutation is present at 3.5% variant frequency, and occurs on its own or in conjunction with the previous. This mutation too is mainly restricted to samples from the Eastern half of the eastern hemisphere.

As such a spike in variation on the basis of such little polymorphism emphasises, that the density of SNPs remains very low across the mitochondrial genome. Overall, this confirms that applying a single substitution rate for the entire genome is largely appropriate for the *Pv* mitochondrial genome.

With such low density of SNPs, which is reiterated by the low values for polymorphic sites (S) per population in *Table 2*, it is unsurprising that heterozygosity (H_s) within populations are generally low, remaining under $H_s = 0.2$ overall. Even so haplotype diversity (H_d) is generally high at >0.6 for all except Vanuatu, and the average number of differences within populations (K) above 1.0 for all regions except Bangladesh. This shows that though there are relatively few mutations across the genome many of the mutations are unique, as is confirmed by the number of haplotypes (h) being $>5\%$ higher than the number of mutations (Eta), and the number of mutations in turn being considerably ($\sim 7\%$) higher than the number of variable sites (S), see *Table 1*. In all 352 different varieties of haplotype are represented in the dataset of 818 samples.

Population	Full Name	N	S	h	H_s	H_d	K	Pi	StDv Pi
ETH	Ethiopia	9	12	5	0.014	0.806	2.833	0.00048	0.00022
SUD	Sudan	34	24	19	0.017	0.898	2.832	0.00030	0.00004
MAD	Madagascar	66	52	35	0.022	0.821	2.509	0.00043	0.00006
WAS	West Asia & Turkey	31	36	21	0.060	0.953	4.151	0.00071	0.00008
UZB	Uzbekistan	9	6	6	0.033	0.833	1.778	0.00030	0.00010
PAK	Pakistan	54	56	46	0.045	0.990	3.669	0.00063	0.00005
IND	India	32	22	13	0.051	0.688	1.948	0.00033	0.00011
BAN	Bangladesh	8	3	4	0.018	0.643	0.750	0.00013	0.00005
MMR	Myanmar	32	19	16	0.056	0.845	2.177	0.00036	0.00007
THA	Thailand	44	37	25	0.100	0.958	3.632	0.00057	0.00006
CAB	Cambodia	62	45	33	0.139	0.953	4.887	0.00083	0.00007
VIE	Vietnam	17	17	12	0.108	0.941	3.779	0.00065	0.00009
BOR	Borneo	4	56	4	0.130	1.000	28.333	0.00126	0.00031
IDO	Indonesia	47	30	26	0.038	0.892	1.998	0.00034	0.00004
PNG	Papua New Guinea	190	101	88	0.066	0.887	2.776	0.00046	0.00004
SOI	Solomon Islands	21	19	13	0.108	0.943	3.967	0.00063	0.00008
VAN	Vanuatu	11	8	3	0.041	0.345	1.455	0.00025	0.00015
CHSTH	China South	34	15	11	0.136	0.875	3.774	0.00064	0.00007
CHGU	China Guizhou	26	7	5	0.102	0.683	2.548	0.00044	0.00008
CHAN	China Anhui	37	14	10	0.026	0.607	1.231	0.00021	0.00006
KOR	Korea	35	24	18	0.198	0.924	5.640	0.00097	0.00006
		N	S	h		Hd	Kt	Pi	StDev Pi
TOTAL*		803	360*	359*		0.980	5.756		
TOTAL ALL		818	309	352		0.977	5.248	0.00091	0.00002

Table 2. *P. vivax* genetic diversity and heterozygosity (H_s)

Genetic diversity measures for each population with $N > 5$, including Borneo. Sample size: N, polymorphic sites (including indels and missing sites): S, haplotype number: h, heterozygosity: H_s , haplotype diversity: H_d , average number of differences within the population: K, nucleotide diversity: Pi, standard deviation of Pi: StDv Pi.

Rows are grouped alternatingly by broad geographic regions, from East to West, then North.

In terms of global differences in diversity, *Table 2* shows that all populations west of Bangladesh retain $H_S < 0.06$ (*Table 2* all populations above BAN), whereas H_S values including Bangladesh and to the east are more variable and have more diversity with $H_S > 0.1$ for 8 out of 14 values (*Table 2* all populations from MMR down). Western Asia (WAS), a combination of samples from Armenia, Azerbaijan and Turkey, has the highest H_S in the west. The DAPC in *Figure 3* shows that the Armenia/Azerbaijan population diverges from the Turkish population as well as most other western populations, shown in red. This indicates that Armenia/Azerbaijan is introducing a higher level of heterozygosity when the populations are combined.

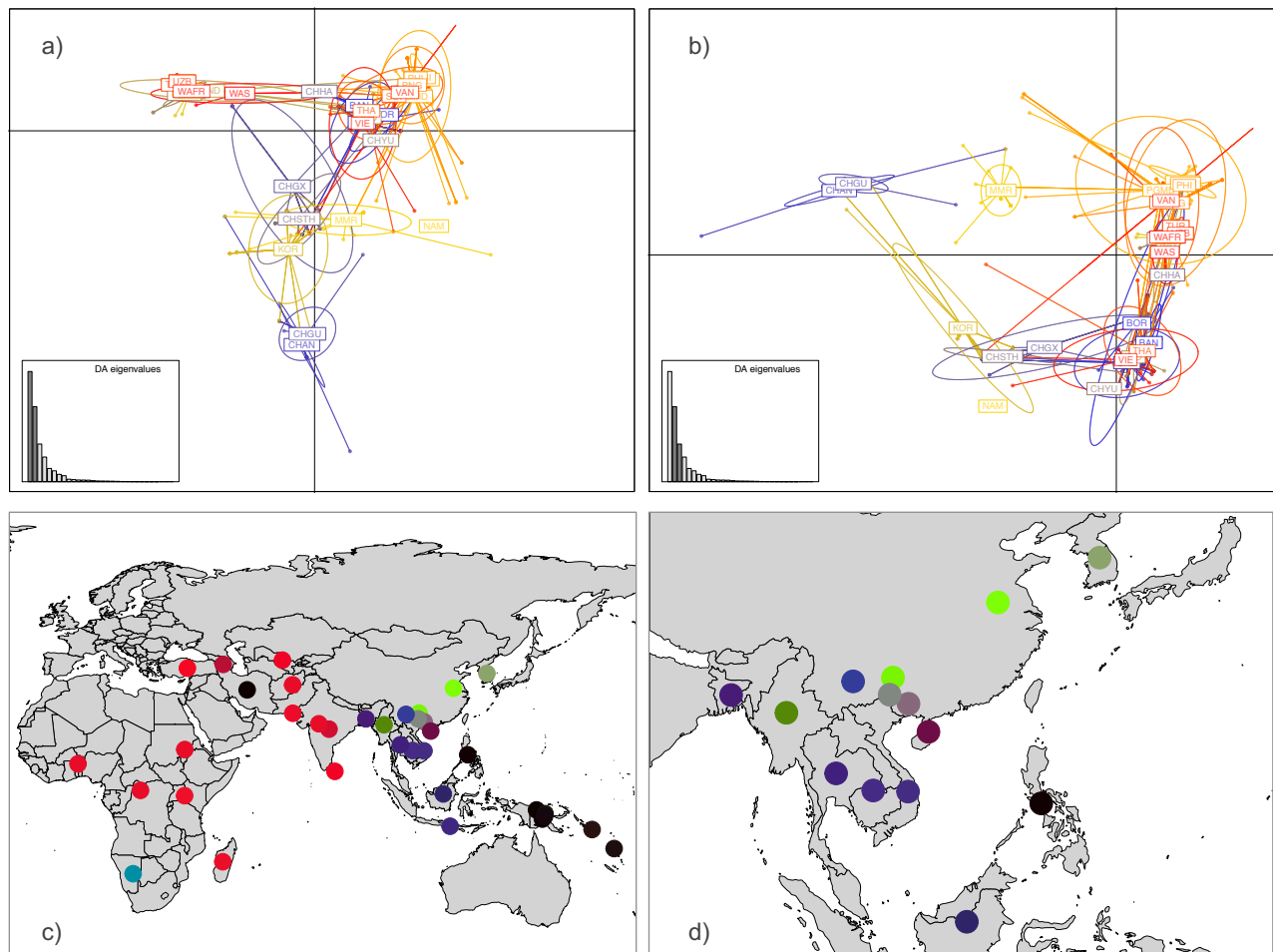


Figure 3. *P. vivax* DAPC 1

Discriminant analyses of principal components (DAPC) for the *P. vivax* populations as given in *Appendix Table S2.2* column "Sample Map", according to DA eigenvalues a) 1 & 2, and b) 2 & 3. c) Shows the relationship between populations as coloured along the rgb spectrum for eigenvalues 1, 2 & 3 according to DAPC analyses type 1, d) A close up of the populations in East Asia.

DAPC calculations were done with the package *adeigenet* in R.

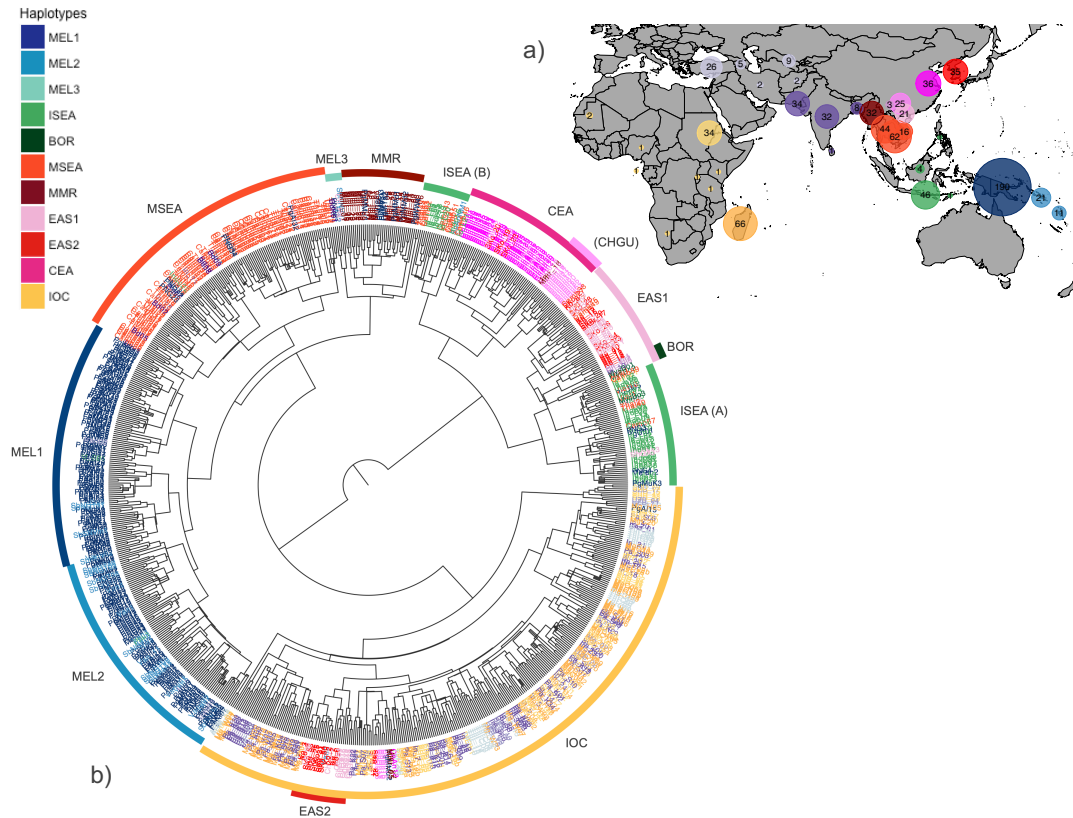


Figure 4. *P. vivax* mitochondrial genome phylogenetics

a) Map of *Pv* sample locations, coloured by geographic grouping “Regional” as given in the index in *Appendix S2.2*.
 b) Phylogenetic tree, run in BEAST2 with settings as given in *Appendix S2.3*. The colour-rings on the outside indicate the different haplogroups, the sample colours indicate their geographic sampling origin, and corresponds to the regions in 6a.

The phylogenetic tree in *Figure 4b* shows that this is through a sample that is of probable Melanesian origin. Other such foreign introductions likely include the samples from Iran, as seen in the DAPC results in *Figures 3 & 6* (DAPC *Pv*) and the phylogenetic tree results in *Figure 4b*, which strongly suggest a Melanesian origin; and the same is true for the sample from the Philippines. The Pilipino and Iranian sequences come from the same publication as some Melanesian samples, so it is unclear if contamination could have occurred during lab-work. However, there is a modern migrant connection between PNG and the Philippines. In addition, the samples from Melanesia and Armenia from this current study were sequenced entirely separately. A different study using aliquots from the same set of Central Asian samples also found PNG-type individuals in Azerbaijan/Armenia using microsatellite data (Koepfli et al. 2015). As such, the probability of the sequence data being from the original source, without contamination, is high.

The origin of the single sample from Namibia in southern Africa remains unclear, as it appears difficult to relate to any other haplotype groups. The argument to exclude such small populations from further analyses is strong, as there is a high chance that small sample numbers are comprised of rare patients coming in from foreign destinations and their sample being used for medical studies. For any small population, such an imported sample will strongly influence the data. Larger sample

sizes are more likely to originate from local population studies, and in addition any imports will have a reduced impact on population genetic analyses.

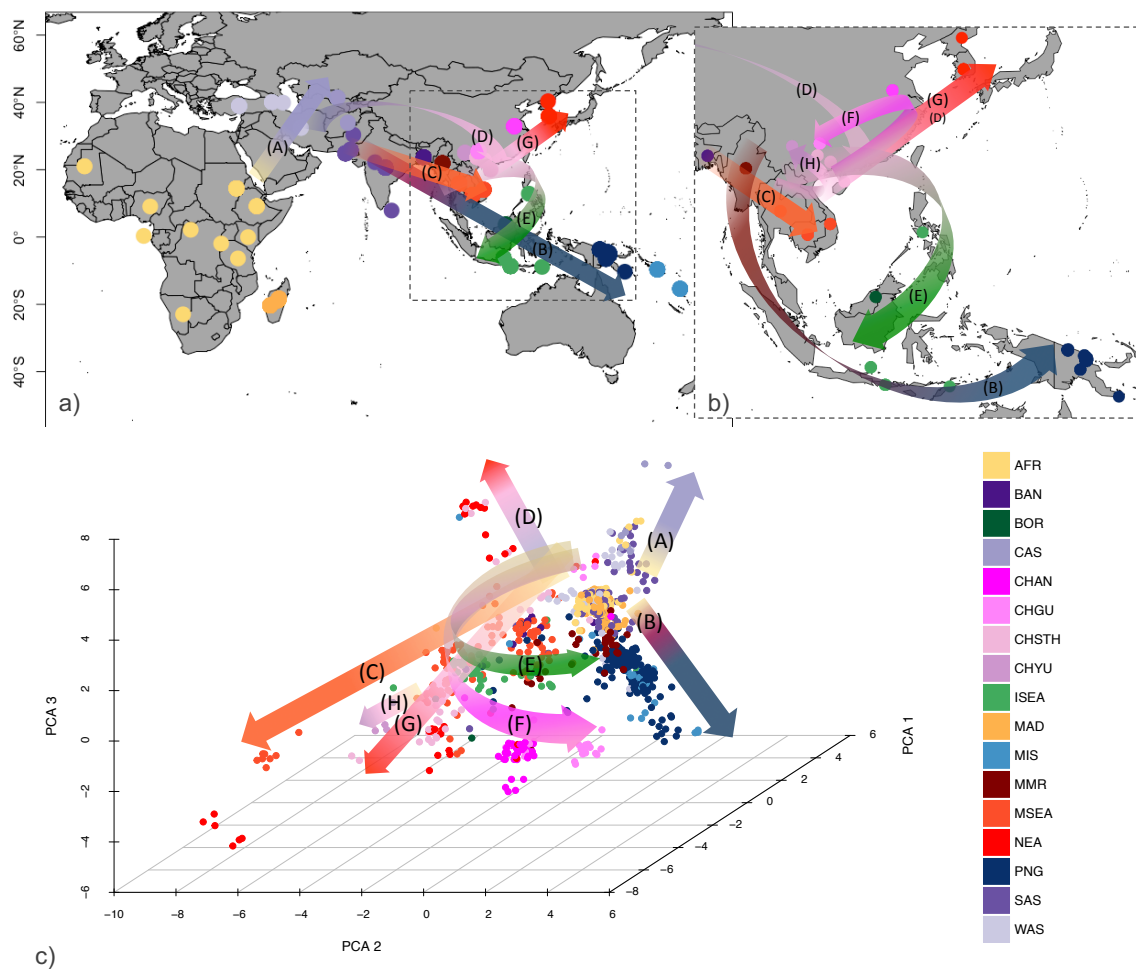


Figure 5. *Plasmodium vivax* mitochondrial genome migration reconstruction from PCA

a. & b. Arrows indicate geographic trajectories of genetic clines derived from the clines in PCA data, as given in c., demonstrating rough migration directions of *P. vivax*: (A) from the Middle-East/Africa towards central Asia (Uzbekistan), (B) from Southern Asia/Africa, through Bangladesh and Myanmar to Melanesia, (C) from Southern Asia/Africa to Southeast Asia, (D) from Central Asia/Africa to Southern China and on to Northeast Asia, (E) from Southeast Asia and Southern China to Borneo and Island Southeast Asia, (F) from Southern China towards Central Eastern China (Anhui) and into inland Southern China (Guizhou), (G) Southeast Asia to Southern China and northwards.

c. Principal component analyses (PCA) of *P. vivax* mtDNA. Arrows indicate genetic clines found across sample populations. Coloured according to geographical region.

AFR: Africa, BAN: Bangladesh, BOR: Borneo, CAS: Central Asia, CHAN: China Anhui, CHGU: China Guizhou, CHSTH: China South, CHYU: China Yunnan, ISEA: Island Southeast Asia, MAD: Madagascar, MIS: Island Melanesia, MMR: Myanmar, MSEA: Mainland Southeast Asia, NEA: Northeast Asia (Korea), PNG: Papua New Guinea, SAS: South Asia, WAS: West Asia.

Overall the results show strong stratification of populations, especially to the east of India. This stratification is found not only in the phylogenetic tree in *Figure 4b*, but also in the PCA in *Figure 5*, and throughout the DAPC1 analyses in *Figures 3 & 6*. Population groupings are relatively similar across all methods – phylogenetic tree, PCA, DAPC1, DAPC2, population differentiation heatmap/dendrogram – showing that the genetic background for the variation is stable and consistent. In the PCA this structuring at first glance seems clear only along the Myanmar-Melanesia axis, which is the most thoroughly sampled region.

East Asian samples from Central and North China show strong local signatures that are distinct from surrounding populations, as confirmed by the phylogenetic tree. Southern Chinese samples and samples from Korea diversify along two different axes, which is also picked up in the Pv phylogeny. The samples from Yunnan, southern China, branch off from the ‘Southern’ branch of the Southern China/Korea phylogeny, and so does the island Southeast Asian population. A variation on the structure seen in the phylogenetic tree is that mainland Southeast Asia here is grouped more strongly with the East Asian clades than with Myanmar/Melanesia. This could explain why occasionally the phylogeny will group mainland Southeast Asia as an earlier branch from the tree than both Myanmar and Melanesia. In *Figure 5* the associations between samples from different regions seen in the PCA have been interpreted on the geographic scale, which shows a complexity of movement in Pv through eastern Asia.

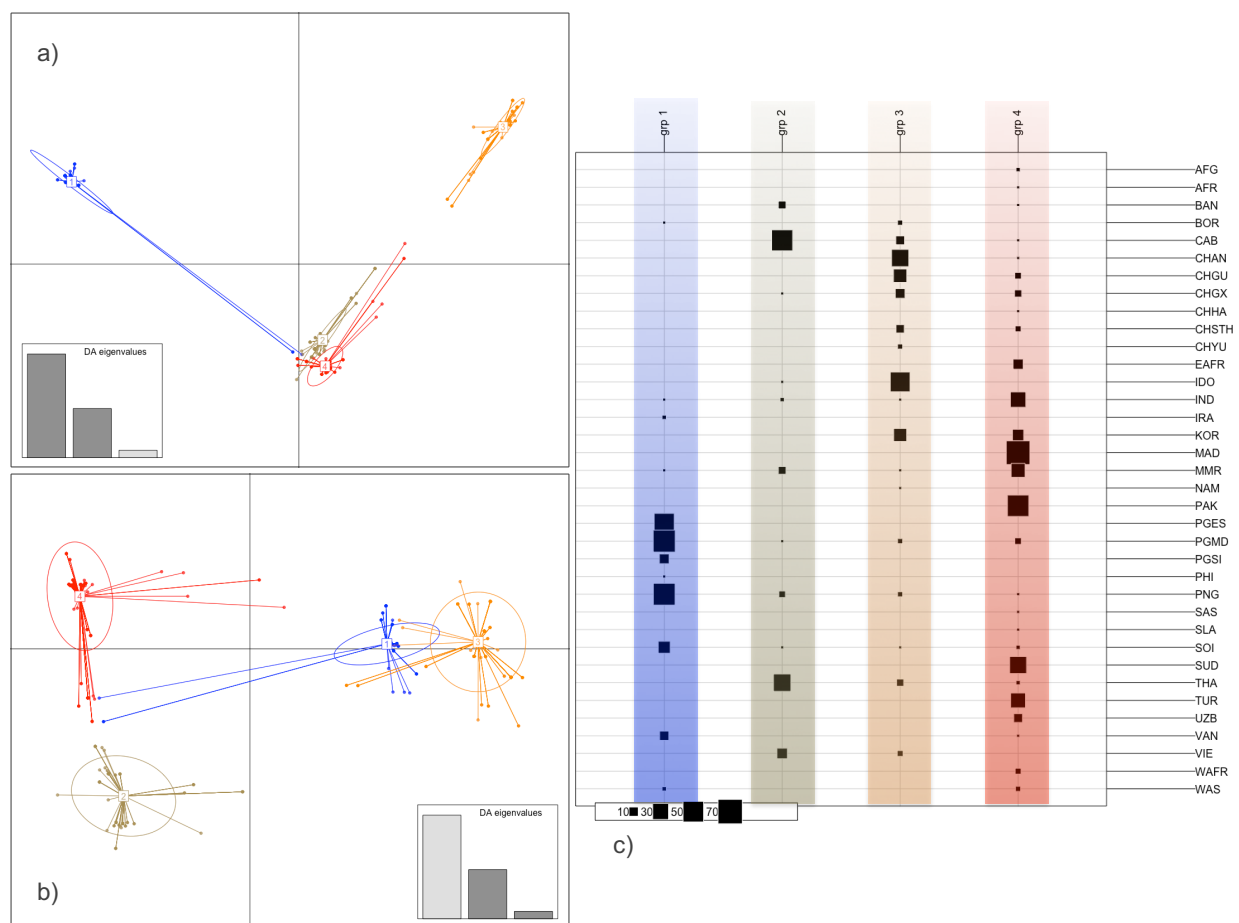


Figure 6. DAPC clusters

- a) Spread of samples and cluster allocation (1 = blue, 2 = beige, 3 = orange, 4 = red) determined by DAPC (*adeigenet* package in R, Jombart *et al.* 2010), shown for eigenvalues 1 (x axis) & 2 (y axis). The samples are allocated to clusters based on similarity across a set number of eigenvalues (n.pca) calculated from the data (n.pca = 1000).
- b) Spread of samples and cluster allocation (1 = blue, 2 = beige, 3 = orange, 4 = red) determined by DAPC, shown for eigenvalues 2 (x axis) & 3 (y axis).
- c) Overview of sample sizes from each region that fall into each DAPC determined cluster (grp). Populations are given on a country level, as is shown in *Appendix Table S2.2* under the ‘Country’ column. Sample numbers are given by the size of the square for each region per group (grp).

The groupings according to geography are most clearly displayed in the DAPC cluster data, *Figure 6*, which finds the best representative number of groupings and shows the number of samples from each region that make up the groups. All methods find a strong single grouping for the populations west of, and including, India. The region spans from South Asia, across Central and West Asia to Africa and Madagascar (Group 4, red in the DAPC cluster in *Figure 4*). Papua New Guinea and island Melanesia (Solomon Islands and Vanuatu) make up a second group (blue Cluster 1 in DAPC, *Figure 4*). The third includes mainland Southeast Asia (Cambodia, Vietnam, Thailand), Bangladesh (Cluster 2, olive in *Figure 4*) and partially Myanmar. Myanmar branches off from the Melanesian group in the phylogenetic tree in *Figure 4b*, and the PCA in *Figure 5* shows that Myanmar sits between the mainland Southeast Asian and Melanesian branches, indicating it as a central hub in its relationship with surrounding regions. Indonesia, Borneo, Southern China and Northeast Asia (Korea), form a fourth group (DAPC Cluster 3, orange, *Figure 4*), as shown in the dendrogram & heatmap (*Appendix S2.14*) and phylogenetic tree, but are more loosely associated with each other, as demonstrated by the DAPC1 population scatterplots in *Fig. 5 a & b*.

In *Figure 6 c & d* the DAPC1 results show some variation among these populations in East Asia. Some of this variation is caused by the small sample size in two of the southern China populations (Hainan = 1 and Yunnan = 3). As mentioned earlier, other points on this map also demonstrate the importance of having a larger sample size, as almost all points represented by populations of $n=1$ or 2 are very different from what is expected from their surroundings (Namibia, Philippines, Iran, Afghanistan), some of which could be imported by international travellers and others that could possibly give new insights but are too small in number to do so. As a precautionary measure, all these populations were either removed (Namibia, Philippines, Iran, Afghanistan), or combined with nearby populations (“China South”: Hainan & Yunnan, with Southern China and Guizhou) from the data looking at interpopulation differentiation in the next sections.

Even though there are some populations with small sample sizes that seem to consist of imported infections, overall imported infections make up only a minor portion of the samples, as can be observed through the generally strong geographical clustering of the samples throughout the results.

2. 3. HUMAN MIGRATION

2. 3.1 INTRODUCTION

As two main hosts exist for human malaria parasites - malaria mosquitoes and humans - it has two possible vehicles for dispersion. Dispersal distances for mosquitoes are generally limited; for example for the main African vector *An. gambiae* most individuals travel under 3 km with the occasional exception (Thomas *et al.* 2013), and *An. quadrimaculatus* was observed to generally disperse within 2 km of origin (Le Prince & Griffith 1917). In contrast, human travel distances are much larger, including significant global connections in modern society. As discussed earlier, mosquitoes may play a significant role in genetic diversification of malaria parasites due to the differences in mosquito host species selecting for genetic adaptations, forming barriers through preferential selection. These dynamics have been further discussed in *Section 2.4* on vectors. However, before the distribution currently seen as influenced by mosquito vectors the parasite most likely spread from its assumed African origins with its human hosts, and has been spread throughout its current range with human migrants. This section looks at traces left of past human migrations in the geographic distribution patterns of the *Pv* mitochondrial genome, aiming to understand the history and distribution of *Pv* in association with its human host.

Numerous diseases spread with the migratory activity of their host. Evidence of past human migration spreading the tuberculosis bacterium *Mycobacterium tuberculosis* (Comas *et al.* 2013), the hepatitis C virus (Markov *et al.* 2012) and the strong relationship between the genetic variation in the stomach gut bacteria *Helicobacter pylori* (Moodley *et al.* 2009) with the distribution of the human genome show that it is possible to use the modern genome to trace past migration patterns of the disease. Other phylogenetic studies reconstructing the historic spread of diseases have used ancient genomes (aDNA) from archaeological samples to help understand past genetic distribution and migration dynamics (e.g. black death *Yersinia pestis* (Bos *et al.* 2011, Achtman 2012) and tuberculosis (Fletcher *et al.* 2003)); however this is not always possible. Currently for *Pv* the only non-contentious ‘ancient’ sample that has been sequenced is around 80 years old, from a medical blood slide and originating from Spain (Gelabert *et al.* 2016). *Pv* genome has been hard to sequence, even from modern samples, so using a modern dataset to understand its past makes sense and is possible through improved bioinformatics methodologies and an enlarged data-set of the mitochondrial genome (>800 samples currently available).

To understand the possible connection between *Pv* mitochondrial genome differentiation and human migration, we have to recognise the major past human movement patterns. The mitochondrial genome evolves relatively slowly, on average 1 mutation per site per ~2500k years, with around 8000 sites. For that reason it is understood to be a good target for studying dynamics that have occurred over several thousands of years. As a start, this section covers the spread of the

with the arrival of *H. sapiens*, with evidence of genetic mixing taking place (Bae *et al.* 2017, Skoglund & Mathieson 2018).

Modern human genetics supports the above movement, in the form of haplogroup M (light blue in *Figure 7*), but estimates the time of emergence to around 65-70 kya (Stewart & Chinnery 2015). Human genetics also indicates this out-of-Africa movement coincides with a second emergence of *H. sapiens* from Africa, taking a northern route through Egypt and the Levant. Part of this migratory group moves further to the north and together with the emergence of a new haplogroup R populates Europe. Haplogroup N also heads across western and central Asia, eastwards across Asia, north of the Himalayas. This migration split again east of the Himalayas, with some populations heading south towards Southeast Asia, and others heading further eastwards and to the North.

Ancient DNA from anatomically modern human skeletons has helped calibrate dates and routes, through giving improved genetic mutation rates, showing ‘missing links’ between human populations and the place and time these links were present (Skoglund & Mathieson 2018). The improved rates support a split between African and non-African populations around 75 kya, and a divergence between non-African (east and west Eurasian) branches from around 47-55 kya (Y chromosome), 47 or 50 kya (autosomal, clean split model), or 50-60 kya (autosomal, general model) (Skoglund & Mathieson 2018). A quick note to these early dispersal routes and their dates is that especially the dates of emergence out of Africa are continuously being pushed back. The most recent example of this is the publication in January 2018 of the discovery of a skeleton with early *H. sapiens* features found in Israel and dated to 194-154 kya (Hershkovitz *et al.* 2018). This discovery pushes back the dates for the oldest findings of *H. sapiens* outside of Africa by over 70 ky, indicating that certain movements could have occurred much earlier.

Some extra definition to the second of the two main routes is given in *Figure 7*, as seen in the human mitochondrial genetic haplogroup N (green). Haplogroup R (red) splits off haplogroup N after exiting Africa, and takes a similar route to haplogroup N, with some variation in western and eastern Asia. The latter two migrations also head through both mainland and island Southeast Asia towards Australia. A much later migration into the Pacific region comes from haplogroup R, which comes from Eastern coastal China through Taiwan and the Philippines. These so-called Austronesians arrived in the Pacific from around 4.5 kya, and are thought to be connected to the archaeological dispersal of Lapita pottery (Vilar *et al.* 2008, Sheppard 2011). They brought a rich cultural landscape with them and with their maritime expertise not only reached far into the Pacific, to Hawaii and New Zealand, but also crossed the Indian Ocean to Madagascar (Bellwood *et al.* 1995, Tryon 1995, not shown in *Figure 7*).

A last significant movement seen in *Figure 7* is that archaeological evidence points towards the Japanese islands being colonised from both the Korean peninsula and southeastern Russia, around

40 kya (Bae *et al.* 2017). Besides the main routes of importance as mentioned above, *Figure 7* shows further levels of detail for both archaeological finds (Bae *et al.* 2017) and human genetics (Stewart & Chinnery 2015). These details are further supplemented by the recent publication by Skoglund & Mathieson (2018) that reviews all major recent findings by aDNA from anatomically modern human remains found across the world. For the eastern hemisphere as discussed in this chapter, some of the major dynamics include the observation of significant population admixture from Northern Africa, through the Levant to both Europe and stretching through Western and Central Asia to the Altai mountains. The admixture includes ancestral European Mesolithic populations from around 8500 ya being introgressed by early farming populations from the Levant and Iran, up till around 4500 ya. The flow between Europe (reached by 5000 BC) and the steppe populations stretching to the Altai Mountains in the East (by 6000 BC) was extensive and bi-directional. Though there is no current aDNA available from South Asia, there are many indications that the populations migrating across the Levant and Central Asia moved down into Southern Asia as well (Slatkin & Racimo 2016, Skoglund & Mathieson 2018).

In contrast, there are few human mtDNA samples from Eastern Asia, 1x China (Tianyuan, ~40 kya), 8x Tibet (1-3 kya), 2x Korea (Devil's Gate, 7.7 kya (Siska *et al.* 2017)) and 2x Japan (Final Jomon period, 3kya) (Slatkin & Racimo 2016, Skoglund & Mathieson 2018). The individuals from Tibet and Korea are strongly related to extant populations, indicating strong population stratification and consistency. Japanese populations show a mixture of both Jomon and Yayoi ancestry, indicating a two-pronged colonisation of early Jomon hunter-gatherers and later Southeast Asia related Yayoi rice agriculturalists. Southeast Asian archaeology and anthropology also claim relatively stable in-situ populations with no extensive overhaul of languages or genetics by populations from outside the region (Bellwood & Glover 2004). Even though much influence was received from the East, including India, and the North, China, this seems to have been mainly cultural permeation and trading, rather than actual colonisation.

2. 3.2 METHODS

Human mitochondrial genome data were obtained from the eHGDp database in R (*adegenet* package, Jombart *et al.* 2010), using sequences published by Rosenberg *et al.* (2002), Cann *et al.* (2002), Ramachandran *et al.* (2005) and Wang *et al.* (2007). Data from the New World was removed. The data were used for principal component analyses and genetic correlation analyses between sample populations, as was done for *Pv* (see methods section for *Genetic Structure*). PCA values and DAPC were calculated to show the underlying variation and geographic connections across the entire dataset. Distance between original sample populations was calculated and the D_{ch} values were converted into a heatmap and dendrogram. These were all calculated for the original populations to enable indirect comparisons of the larger datasets of *Pv* and humans, as many populations in both datasets did not overlap and the next step of using more direct comparisons might lose depth due to loss of data.

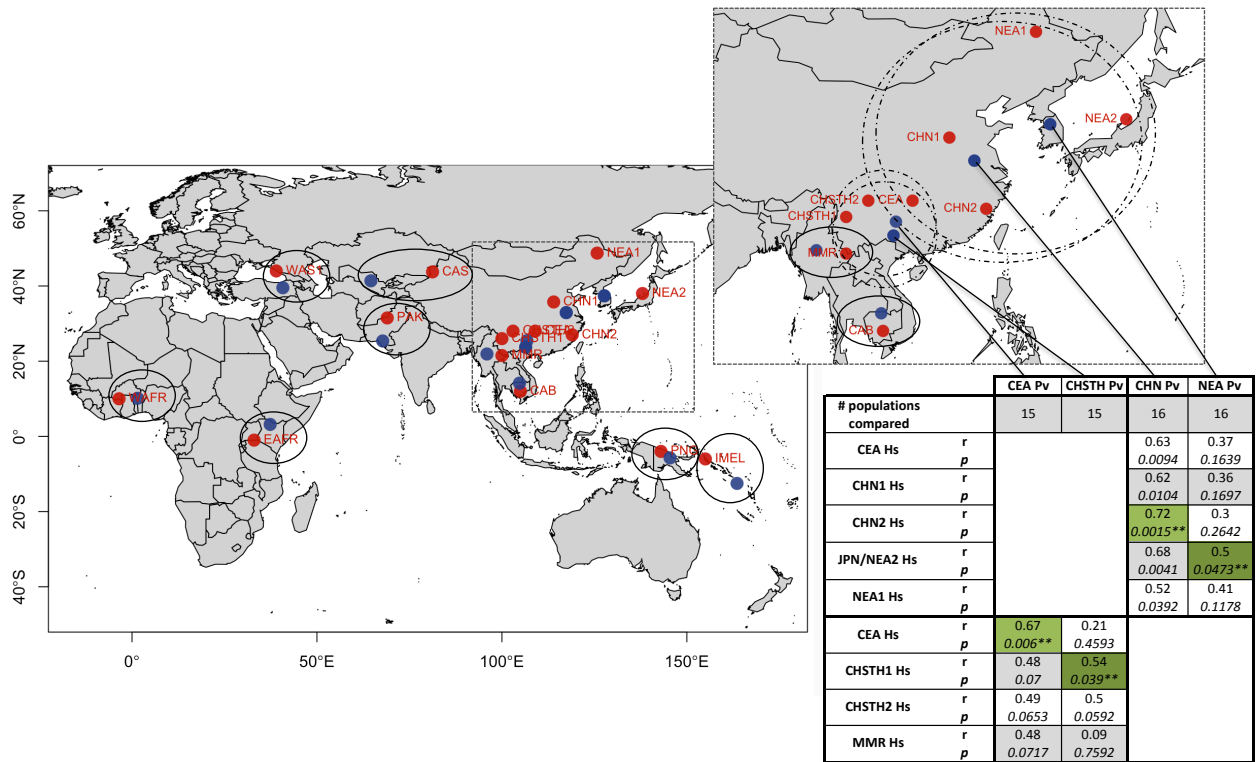


Figure 8 & Table 3. Comparison pairs of *Pv* and human sample sites

Figure 8: Corresponding sample locations of *Pv* mtDNA (blue) and human genome (red). A close up of the East Asian region is given, with

Table 3: a table of Spearman correlations testing the closest link between surrounding human populations and certain *Pv* populations (*Pv* populations: CEA, CHSTH, CHN & NEA).

CAB: Cambodia, CAS: Central Asia, CEA: Central East Asia, CHN: China North, CHSTH: China South, EAfr: East Africa, IMEL: Island Melanesia, ISEA: Island Southeast Asia, MMR: Myanmar, NEA: Northeast Asia, PNG: Papua New Guinea, SAS: South Asia, WAFR: West Africa, WAS: West Asia. Appendix S2.4 shows the regions used for each population from the human data, and Appendix S2.2 (*Hs Comparison*) the *Pv* populations.

Populations were then grouped into regions to correspond with similar regions from *Pv*. Some *Pv* populations were re-assigned to correspond with similar geographic structure as the human data, and genetic variation between populations was re-calculated for these groupings in both humans

and *Pv*, as can be seen in *Table 3* and *Figure 8*. New population locations and structuring for *Pv* can be found in *Appendix S2.2*, under the column “Hs”, and seen in *Figure 8*. *Figure 8* shows that several human populations geographically surround one or two *Pv* populations across East Asia. Pearson correlations were tested between *Pv* interpopulation genetic values (D_{ch}) and the varying human interpopulation genetic values (D_{ch}) to see which of the surrounding human populations was the best fit with the *Pv* population given over all *Pv* & *Hs* comparison. Results are given in *Table 3*. See *Appendix S2.2* for an overview of the *Pv* population division, *Appendix S2.4* for that of the human populations, and *Appendix S2.13* for an overview of the populations where cross-comparison occurs.

With this grouping of populations, correlations between human and *Pv* population genetic differentiation were computed to check for pattern overlap. Linear model correlation values, including F_{ST} , correlation coefficients and ANOVA, were calculated for *Pv* D_{ch} values with *Hs*, Geo and Vec. Calculations were approached in the same manner as the sections on geography and vectors, using the R packages *stat* and *Hmisc*. Results are shown in tables in *Appendices S2.9 - S2.11*. Pearson correlation was checked (r_s : 0.19, p : 0.15), and a plot of *Hs* vs *Pv* genetic differentiation values is shown in *Appendix S2.12*.

For a broader comparison of human migratory data with the distribution and timing of *Pv*, data from previous publications was gathered to give an overview of times of divergence across the eastern hemisphere. The data includes divergence estimates for modern human genome, dates of archaeological finds of human remains or activities (split into identified *Homo sapiens* and general unidentified *Homo* spp.), some dates on remains of introduced commensals (e.g. pig and dog) and dates of the ancestry of diseases (*Helicobacter pylori*, *Mycobacterium tuberculosis*, and Hepatitis B virus). The references for these dates, as displayed in *Figure 10*, are given in *Appendix S2.15*. For some regions, there was more than one set of date estimates for one data-type. These have been aggregated into a single date, taking the minimum 95% interval and the maximum 95% interval across all data and averaging the mean. If 95% intervals were unavailable from the data, for aggregate data the mean was used as maximum and minimum value, in single data the mean was decreased by 1 kya for the minimum and increased by 1 kya for the maximum. This interval was used as it gives a date range of relatively high precision compared to the rest of the data and makes the data visible in the figure. For all other data, the given mean and 95% intervals were taken from the reference.

2. 3.3 RESULTS & DISCUSSION

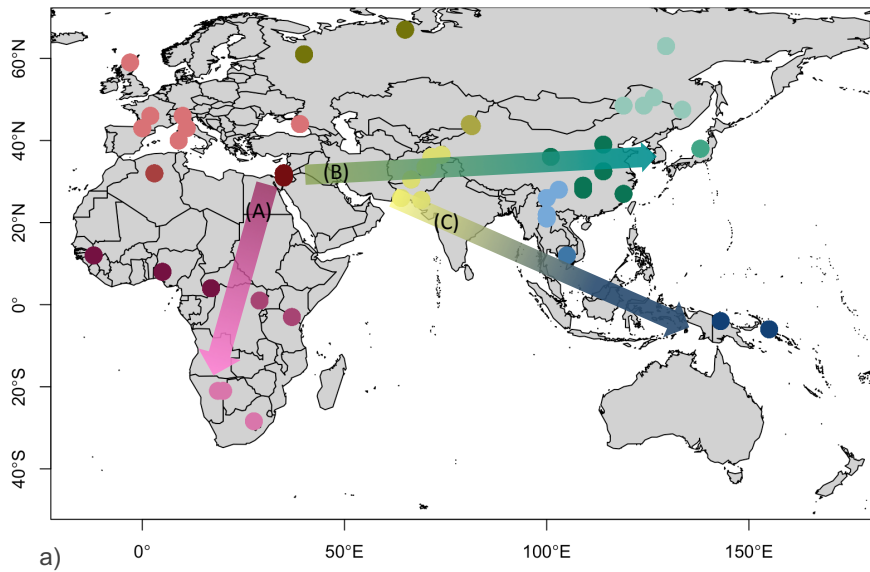
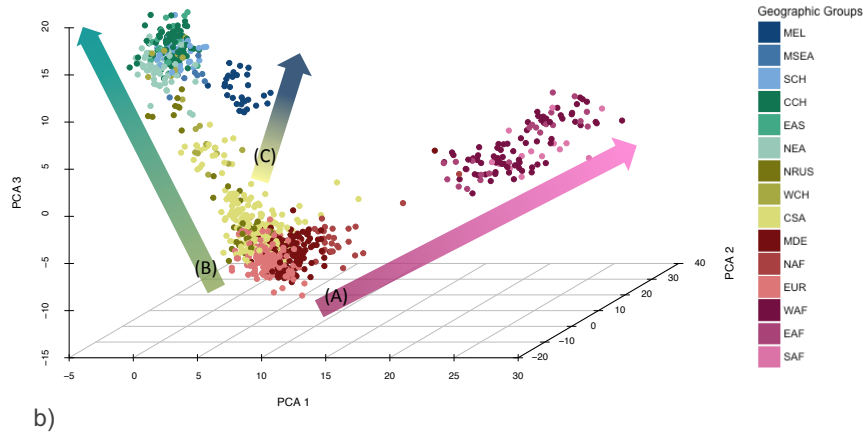


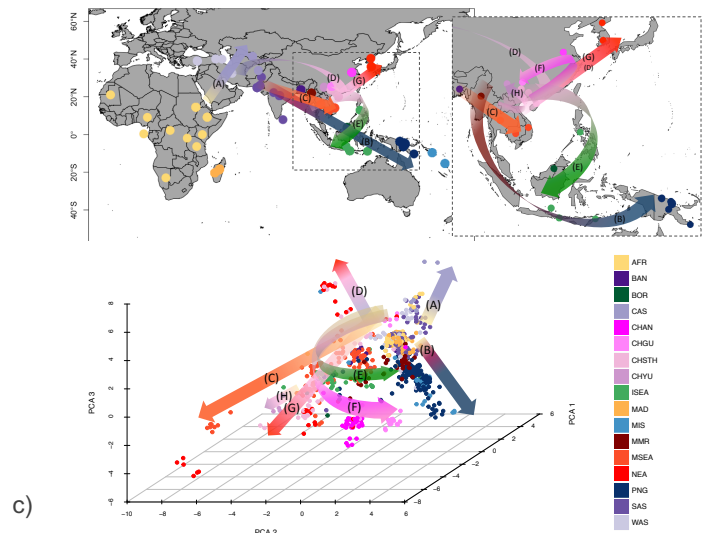
Figure 9. PCA of human genome SNP data diversity

a. Map of human sample locations, coloured by geographic grouping as given in the index. Arrows indicate geographic trajectories of genetic clines found by the PCA in b., demonstrating rough migration directions of anatomically modern human populations: (A) from the Middle-East towards Southern and Western Africa, (B) from the Middle-East/Europe to Southern Asia, Central China, East and Southeast Asia, (C) from Southern Asia to Melanesia.



b. Principal component analyses (PCA) of human SNP data from the eHGP database. Arrows indicate genetic clines found in the human genomes across sample populations. Coloured according to geographical region. MEL: Melanesia, MSEA: Mainland Southeast Asia, SCH: Southern China, CCH: Central China, EAS: East Asia, NEA: Northeast Asia, NRUS: Northern Russia, WCH: Western China, CSA: Central South Asia, MDE: Middle East, NAF: North Africa, EUR: Europe, WAF: West Africa, EAF: East Africa, SAF: Southern Africa.

c. For reference purposes, Figure 3 has been included to demonstrate the complexity of the patterns seen in Pv.

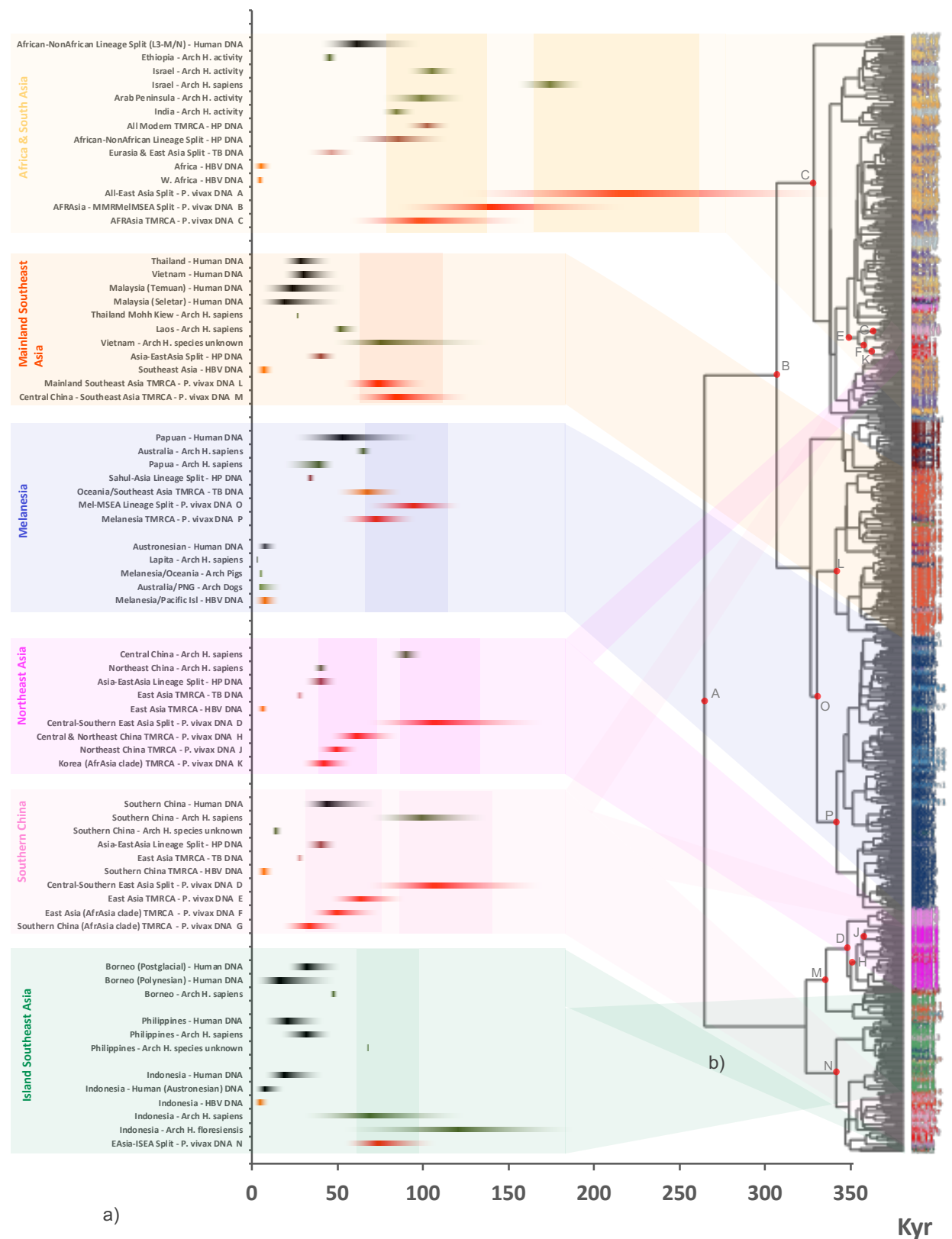


The principal components analysis for the human dataset shows strong structuring in 3D space, as seen in *Figure 9*. The distribution is along three main axes geographical axes: African (pink), Asian (yellow, green) and Melanesian (blue) branching off the Asian line (*Figure 9b*). The directions of

the PCA results have been shown in their geographic setting, which shows that the axes run along almost linear geographic clines (*Figure 9a*).

The similarity between the two datasets is underlined by the results from the Pearson and linear model correlation tests. The Pearson correlation is significant across the entire dataset (lowest values in *Table 8*, $t: 0.72$, $p: 0.0015$). More telling is that for the ANOVA F_{ST} values are not significant taken across the entire Hs versus Pv dataset, but become significant once comparisons between populations within the western region are removed. As Pv has a flat structure for the genetic differentiation across this region, it makes sense to remove this from the comparison. This does imply that human genetic differentiation is correlated to geographic distance, and so it could be argued that it is hard to tease these factors apart. However, ANOVA tests that combine for the influence of combined geography and human genetic variation find no significant correlation with geography, whether across all data or the specific E-W dataset. This could be due to the more limited combined dataset of Pv and humans leading to odd structuring of the data in geographic space, for example including 13 populations with only five in the west that in addition give very little information for Pv.

The time comparisons in *Figure 10* show interesting patterns concerning the age estimates for Pv lineages and how they correspond with other datasets. The earliest population split in Pv occurs well before the split between African and non-African humans. However, the estimated time for the most recent common ancestor for the African-Asian clade is estimated to occur around the same time as the most recent common ancestor for modern *Helicobacter pylori*, a stomach gut bacterium often seen as a proxy for human genetic diversification, and the earliest human activities seen in the archaeological record around Northeast Africa and the Middle East. This pattern is subsequently seen almost consistently, where Pv estimated dates for the most recent common ancestor will fall a little earlier than the estimated dates for humans, but will overlap with dates from archaeological sites. The sites with corresponding ages to Pv diversification are mainly of modern humans, *Homo sapiens*, except for mainland Southeast Asia where the dates correspond to hominid archaeology for which the species is unknown. For mainland and island Southeast Asia the human ancestry dates are relatively recent and do not relate to the first settlers, as the data comes from populations with an East Asian heritage. This is comparable to the Austronesians in Papua New Guinea that arrived much later (from around 8 kya) than the early Papuan populations (from around 65 kya). Papuan and Austronesian heritage is further discussed in *Chapter 3*, which focuses on the Melanesian region and the relationship between Pv and these two migration events.



Lastly, the recent review looking at aDNA and movements of anatomically modern humans across history (Skoglund & Mathieson 2018) suggests that much movement and admixture took place between human populations in India and further west, at least from 5,000 ya. The genetic remains of humans in Eastern Asia, on the other hand, suggest that long-term population stratification took place in these regions from >4,000 ya. These findings strongly correspond with the patterns seen in the *Pv* mtDNA.

2. 4. GEOGRAPHY

2. 4.1 INTRODUCTION

Geography is seen as one of the major factors involved in speciation and genetic diversification across many species. Genetic similarity between populations generally declines over geographic distance, as contact between populations, and so the exchange and spread of genomes, decreases. Isolation by distance (IBD), as it is often called, can be interpreted relatively easily as the genetic variation between populations is expected to correlate with the geographic distance between them. Naturally this is a simplified interpretation that is much more complex, as it does not take into account for example geographic barriers such as deserts, mountains and oceans. These geographic features may form barriers to movement and contact between populations, and so tend to increase the differentiation between populations over smaller geographic distances. As is shown for humans in Pagani *et al.* (2016), such barriers can stratify the IBD correlation into several different layers.

Another geographic influence on *Pv* genetic variation is the direct and indirect effect of climate on the parasite. Though more tolerant than *Pf* to colder temperatures (Gething *et al.* 2011a), there are still limits to the climates *Pv* can survive in. This is not only due to the presence/absence of the mosquito vector host, but also depends on the biological ability of *Pv* to tolerate colder temperatures. The ability to survive in colder temperatures also depends on the strain of *Pv*. Strains found in more northern latitudes show local adaptation to the long winter season and lack of mosquito vectors by extending the dormancy period as hypnozoites in their life cycles (White 2011, White *et al.* 2016). This is assumed to have underlying genetic mechanisms, though the specific genetic regions involved are still unknown.

Not only do geography and climate condition genetic exchange in *Pv*, forming drivers of allopatric speciation, but within similar geographic and climatic ranges mosquito vector and human dispersal or behaviour also drive local adaptation/selection. These factors driving sympatric speciation in *Pv* malaria complicate direct geographic correlations, as both hosts to the disease come with their own set of complexities that influence distribution and genetic variation in malaria. The expectation is that IBD is present in *P. vivax* but is less straightforward due to the large number of variables playing a role in the genetic variation. Influences of the human and mosquito hosts are discussed in separate sections. The focus for this section is exploring the role of geographic distance on the genetic variation found in the *P. vivax* mitochondrial genome.

2. 4.2 METHODS

Geographic distances were calculated through the Haversine formula¹, using the latitude and longitude of sites. Samples were divided into populations as shown in the table in *Appendix S2.2* (column 'Country 4') with all populations <4 samples removed. The table also includes coordinates and the resulting distances in kilometres between populations.

To largely follow land-routes, for distances between locations crossing the Middle East (such as from Africa to the Pacific), the geographic location of Iran (IRA) was used as a mid-point between populations to the east and to the west. Turkey (TUR), Armenia & Azerbaijan (CAS) and western Asia (WAS) were exempted from this, as these populations are already centrally located.

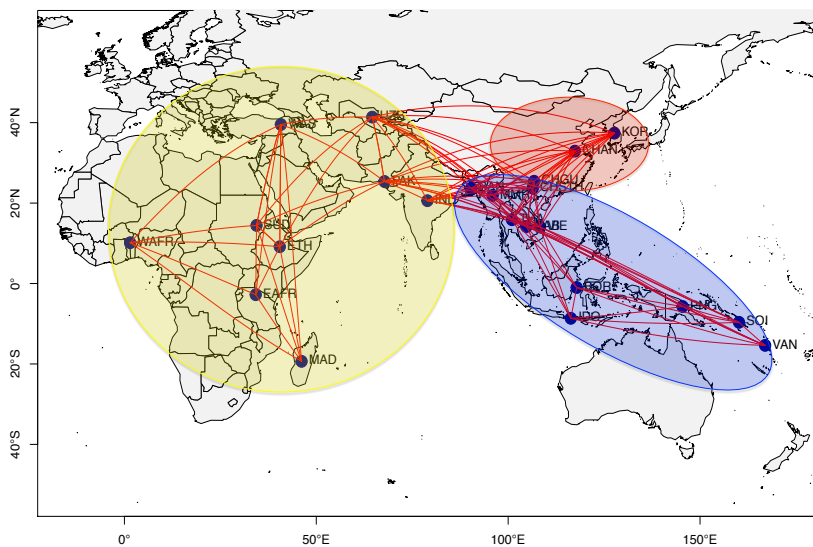


Figure 11. Geographic connections between sample sites

The figure shows a simplified network of connections between sample sites, as used for calculating geographic distances. Each regional division (West, East, North) has been indicated by a colour panel (Yellow, Blue, Red, respectively). Also given in Table 2. All connections between east and west cross through Pakistan (PAK), Uzbekistan (UZB) or western Asia (WAS).

Table 4. Population names, sample sizes (N), assigned geographic group (Geo) and colour (GeoCol) for visualisation in further analyses.

Population	N	Geo	GeoCol
BAN	8	E	Blue
BOR	4	E	Blue
CAB	62	E	Blue
CHAN	36	N	Red
CHGU	26	N	Red
CHSTH	34	E	Blue
ETH	9	W	Yellow
IDO	47	E	Blue
IND	32	W	Yellow
KOR	35	N	Red
MAD	66	W	Yellow
MMR	32	E	Blue
PAK	54	W	Yellow
PNG	190	E	Blue
SOI	21	E	Blue
SUD	34	W	Yellow
THA	44	E	Blue
UZB	9	W	Yellow
VAN	11	E	Blue
VIE	17	E	Blue
WAS	31	W	Yellow

Genetic structure in *Pv*, as demonstrated previously, shows strong differences in population genetics between populations to the west and to the east of India. In addition, populations from central and northern China tend to be surprisingly different from nearby populations. As seen in *Figure 11* and *Table 4*, the populations have been divided into these three categories - E, W, N - as a way to track differences in inter- and intraregional variation in IBD values, as the expectation is that the genetic differences between these regions will influence the correlation between genetic and geographic distance. The category for each population can be found in *Appendix S2.5*. D_{ch} values, calculated in the same manner as in the *Genetic Structure* section, can be found in the lower half of the matrix in *S2.5*. Correlations between D_{ch} and geographic distance were calculated with the Pearson correlation coefficient using the *Hmisc* package in R (rs: 0.19, p: 0.15).

¹ Haversine formula two step for Excel used from: <https://webtrough.wordpress.com/2011/09/19/latlon-distance-formula-in-excel-haversine-and-spherical-law-of-cosines/>

2. 4.3 RESULTS & DISCUSSION

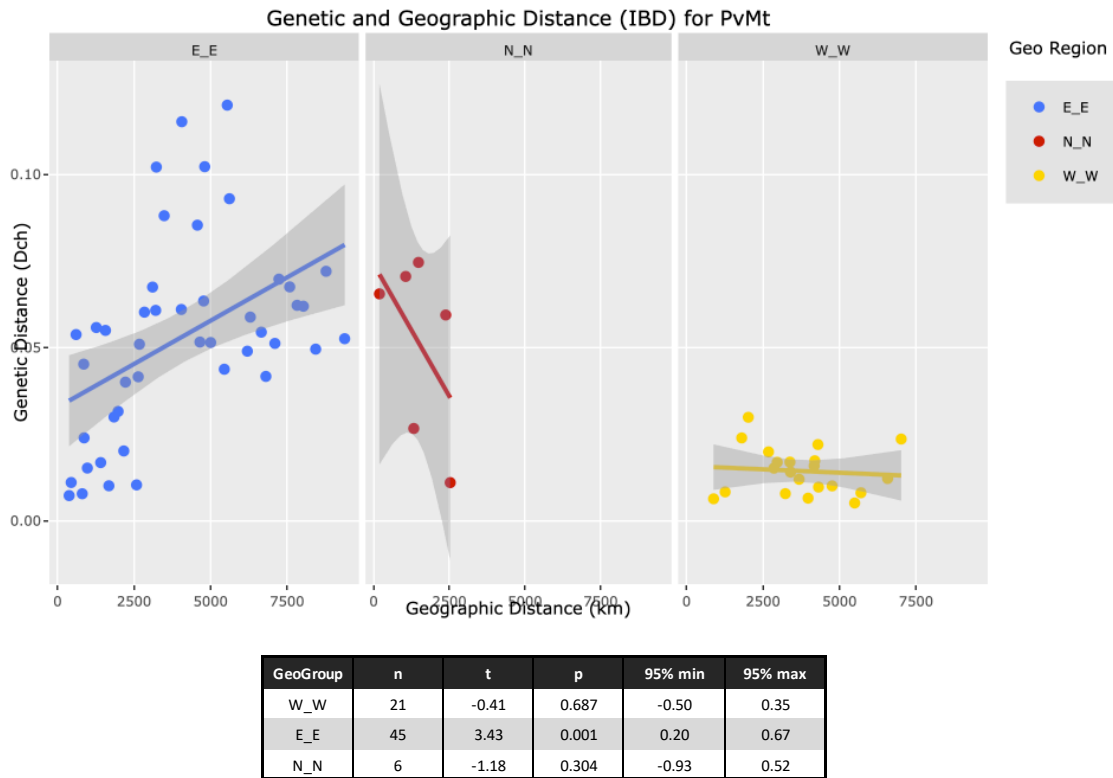


Figure 12 & Table 5. Geographic distance versus genetic differentiation (D_{CH})

Figure 12: Correlations between geographic distance and population differentiation, divided into different cross-geographic groupings depending on the geographic location of the sampled population, as shown in *Figure 11*. 'East' denotes anything to the east of and including Bangladesh, except for the four 'North' populations, to the north of and including southern China. 'West' denotes anything to the west from India, inclusive. Correlation slopes (lm) for each group are included.

Table 5: Group size (n), Pearson correlation coefficients (t) and p-value (p) comparing linear correlations between geographic distance and genetic differentiation within each geographical group (GeoGroup), as shown in *Figure 9*.

Correlations between distance and genetic differentiation, as present between populations, are shown in *Figure 12* and *Table 5*, divided into subsets of geographic groups (East, North, West). The correlation is strongly dependant on what groups are being compared: within the eastern region there is a significant positive correlation (blue, $t = 3.43$, $p = 0.001$), whereas there is a non-significant, almost flat correlation found within the western region (yellow, $t = -0.41$, $p = 0.687$), showing that no matter the distance travelled in the western half of the eastern hemisphere, the genetic similarity between regions remains similar. The northern region shows a strong but insignificant negative correlation (red, $n = 6$, $t = -1.18$, $p = 0.304$). It is hard to say anything significant about this region due to its small sample size, however the negative correlation is likely due to the samples from those regions furthest apart, Korea and Southern China, are the most closely related. This is clearly seen in the DPCAs and phylogenetic trees in *Figures 3, 6 & 4* respectively. Taken overall, the isolation by distance theory holds true ($r = 0.17$, $p = 0.013$), with a significant, weakly positive correlation between distance and genetic relatedness.

The differences between the patterns of IBD seen across the three regions is discussed more in-depth in section 2.6 of this chapter. Broadly, the reason is likely related to a combination of the other two factors discussed in this chapter: 1) Human migration patterns in the past – with significant movement between African and Asian populations, up to Southern Asia; a stronger stratification of populations to the east; and possible coastal movement of humans that link Southern China with Korea, breaking up the possible stratification of the Northern region. Lower diversity in the West could also be caused by historically higher transmission of *Pv* malaria across this region. 2) A strong suggestion is found for the influence of vector species distribution influencing the regional genetic variation in *Pv* – further discussed in section 2.5 of this chapter. In addition, an influence from climate might play a role too.

Further evidence for the linear relationship between geographic and genetic distance is found in the F_{ST} for a linear model. Correlations between D_{CH} and geographic distance were found to be significant across all populations with >5 samples, including Borneo (F_{ST} : 12.75, DF: 207, p : 0.00*), and for all samples east of India, incorporating the northern populations (F_{ST} : 11.49, DF: 88, p : 0.00*). For all samples west of and including India, the correlation decreases (F_{ST} : 2.753, DF: 18, p : 0.09) and is no longer significant, due to the loss of variation in D_{CH} across this geographic region.

2. 5. MOSQUITO VECTOR

2. 5.1 INTRODUCTION

When considering malaria genetics, it is important to realise that besides geographic barriers that limit the presence of the disease or the movement of the human or mosquito hosts, adaptation to the host or vector can be a limiting factor to its spread. Genetically, humans are very similar and all belong to one species, but for *Anopheles* mosquito vectors this is not the case. Many different species, and sub-species, of *Anopheles* mosquitoes host malaria (Sinka *et al.* 2012), and these species and most of their diversity are likely much older than the earliest humans; with for example previous studies indicating the presence of the *An. farauti* complex in Melanesia before anatomically modern human occupation (Hasan *et al.* 2008). For *P. falciparum* it has been shown that regional vectors are strongly related to the effective transmission of regional parasite strains (Molina-Cruz & Barillas-Mury 2014, Molina-Cruz *et al.* 2015), to the extent that transmission of non-local strains can be blocked entirely. This host-parasite compatibility has a genetic background and experiments have shown that parasites with the regional variant of the circumsporozoite (*csp*) gene have increased compatibility with the regional host (Joy *et al.* 2008).

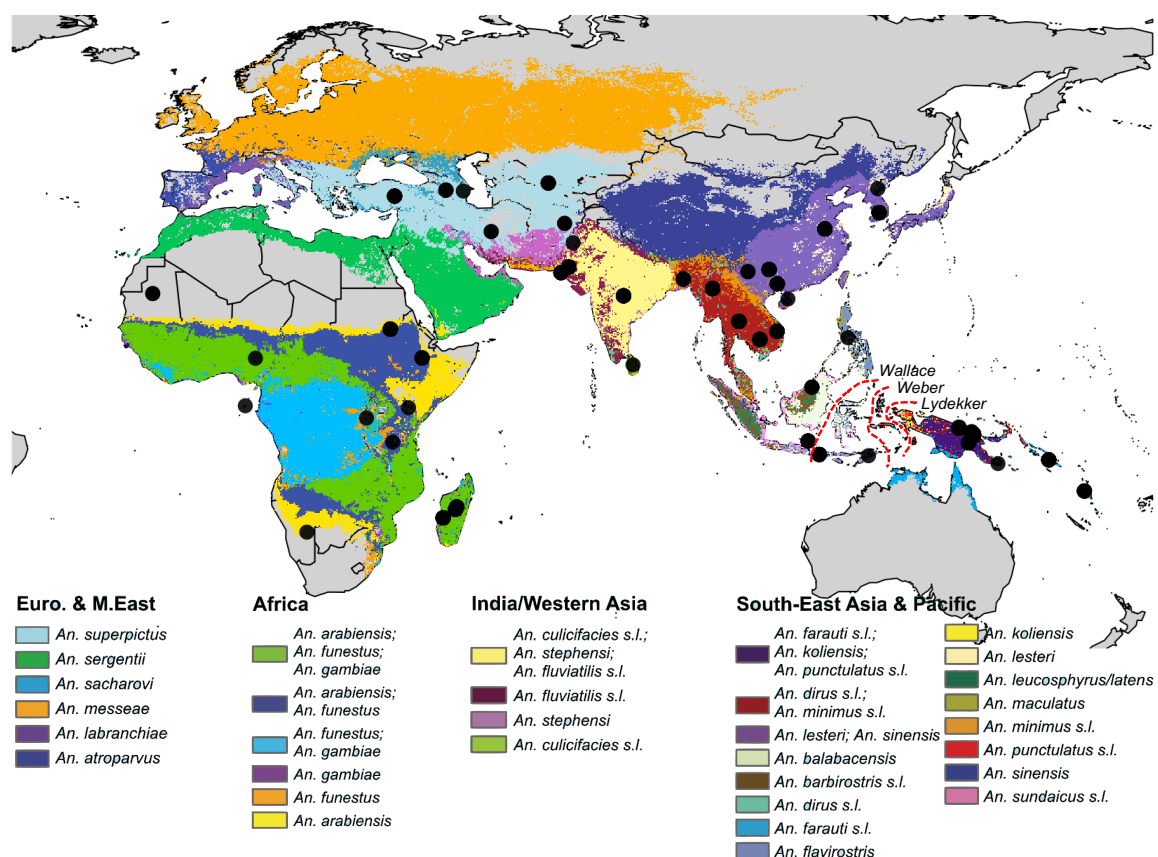


Figure 13. Main malaria *Anopheles* vector species per region

Map with the main *Anopheles* vector species, data, legend and colours as given in Sinka *et al.* 2012, and sample locations of all used *Plasmodium vivax* mitochondrial genomes. The location of the geographical divides between the Oriental and Australian regions (Wallace, Weber and Lydekker lines), according to Gaylord 1977, are marked in red.

Foreign strains may have an advantage over local populations regarding transmitting within human populations, which will lack specific immunity to strains they have not been exposed to.

Alternatively, foreign strains might be hindered by host genetic resistance mechanisms if local parasites are better adapted, and the barrier to transmission within vectors could explain why strong structuring remains within the mitochondrial genome, if imported strains have disadvantages not only in numbers, but also in host compatibility and subsequent transmission and infection. The Assam province in Northeast India is assumed to act as such a barrier. Chloroquine resistance in *Pf* has been present to the east of this region since the 1970s, yet showed delayed spread to the west (Goswami *et al.* 2013). This geographic boundary is associated with a change in major mosquito vector species from *A. dirus/minimus* in Southeast Asia to *A. fluviatilis/stephensi/culicifacies* in South Asia (Sinka *et al.* 2012), see *Figure 13*. Similarly, while chloroquine resistance has been present in Colombia since the 1960s, *P. falciparum* in Central America remains chloroquine sensitive (Sá *et al.* 2009).

Another example concerns *Pv*. In southern Mexico, coastal and foothill vector species, *An. albimanus* and *An. pseudopunctipennis* respectively, meet in a hybrid zone near the town of Tapachula (Joy *et al.* 2008). Within these populations there is a strong dichotomy in both the circumsporozoite (*csp*) gene of *Pv* and 24 other microsatellite (MS) markers, which vary according to mosquito vector species that hosts the parasite. Each malaria parasite was shown to reproduce more successfully in their sympatric mosquito species. Within the mosquito hybrid zone overlap and interchange of the genetic variation is present across *Pv* parasites, with some vectors carrying parasites of both types. Between coastal populations and foothill populations separately the variation between malaria parasites was absent. Joy *et al.* (2008) show that mosquito vector species can form a selective pressure on malaria parasites and form a barrier to genetic introgression.

It can be concluded that there is a definite genetic impact from mosquito vector species on malaria parasites. When looking at genetic correlations between *Pv* parasite population structure and mosquito vector, it makes sense to first check whether genes involved in adaptation process to the mosquito vector are associated with mosquito vector distribution. *Pv* genes that are suspected to be involved in processes of malaria transmission in the mosquito vector – such as Pvs47 and Pvs48/45 – have shown distinct geographical differentiation, even with a limited number of polymorphisms (Woo *et al.* 2013, Tachibana *et al.* 2015). Pvs47 shows a slightly similar patterning to *Pv* mtDNA. Sequences from Korea (40) and Indonesia (3) show close genetic proximity, with Thailand (10) forming a separate cluster closely related to Indonesia. Vanuatu (9) forms a separate cluster that is distinct. However, with only five countries sampled in the eastern hemisphere, all to the east of India (namely: India, Thailand, Korea, Indonesia and Vanuatu), and only three samples from India and Indonesia, it is hard to make further intra- or even interregional comparisons from the currently available data. There are no overlapping mosquito vector species distributions for the countries sampled.

Hupalo *et al.* (2016) did observe the impact of adaptation to mosquito vectors across multiple sections of the *P. vivax* genome, where high amounts of diversity was detected in genes involved in mosquito vector interactions when comparing between geographic areas, in this case Old and New World. The larger dataset of *Pv* genetics provided by Hupalo *et al.* (2016) also includes data on Pvs47. Combining the two datasets – Tachibana *et al.* 2015, Hupalo *et al.* 2016 – gives an increased coverage of samples from new locations across the eastern hemisphere, namely Madagascar, Cambodia, Myanmar, and Papua New Guinea. An overview of the mutations in Pvs47 found across samples from populations across the eastern hemisphere, from these datasets, is given in *Supplementary Table S2.7*.

Table S2.7 shows a few geographical trends in mutations across the Pvs47 gene: 1) between India and Myanmar, there is a loss of the mutation in position 57 for all populations to the east, and an increase in occurrence in a mutation in position 373; 2) a mutation is occasionally found in position 273, restricted to Myanmar and the Mainland Southeast Asian populations (Thailand, Cambodia); 3) in populations across East Asia when moving eastwards, the mutation in position 230 appears to become more fixed; 4) the biggest switch is visible in Melanesian populations (PNG & Vanuatu), with the loss of mutations in positions in 230 and 233 – commonly found in all the other regions – and the novel, frequent, occurrence of a mutation in position 240. These observations show that there is variation present in the Pvs47 gene, which broadly follows some of the regional mosquito vector variation seen by Sinka *et al.* (2012). However, none of the regionally typical mutations (Ile273Val & Ser62Asn for Mainland Southeast Asia, Glu240Asp for Melanesia) are fixed within the populations. Only the absence of the mutations Ser57Thr (throughout the region east of India), and Val230Ile & Met233Ile (in Melanesia) are relatively comprehensive. In all, in its current state the Pvs47 database does not appear to show any clear association between its genetic diversity and the main mosquito vectors present in the region.

An interest in comparing and correlating vector species with the genetic differentiation in malaria parasites remains strong, even for the mitochondrial genome that does not contain genes involved in specifically vector-related processes. The paper by Joy *et al.* (2008) demonstrates and emphasises that the presence of a transmission barrier created by vector species has consequences for regions on the genome that are not involved in processes regarding vector adaptation: vector specificity is enough for populations to diverge in isolation and differentiate in other regions on the genome as the barrier prevents hybridisation. In the analyses of SNPs across the full genome of 195 isolates, geographical structuring for the eastern hemisphere has some overlapping similarities with the *Pv* mtDNA dataset, and some notable differences. Similarities include the genetic differentiation between India and Myanmar, and the strong phylogenetic separation of the population in Papua New Guinea. Differences include that Myanmar and North Korea cluster with Thailand, and Cambodia and Vietnam show some differentiation from the Myanmar/NKorean/Thai cluster. A quick comparison of the genetic diversity found in *Pv*, as detected by Hupalo *et al.* (2016), with the

map of Anopheline distributions in *Figure 13*, shows that the geographic distribution of *Pv* only partially corresponds with the geographic distribution of Anopheles vector species: with correspondence in the India/Myanmar switch and the PNG differentiation, but non-correspondence of *Pv* genetic structuring with vector populations in East Asia.

As the mitochondrial genome evolves only very slowly, it may be the best part of the genome to compare with Anopheline species structuring, as the age and global distribution of Anopheline species has been estimated to ages well beyond the ~160kya age estimate of *Pv*'s global most recent common ancestor (Hasan *et al.* 2008). The mitochondrial genome might not undergo directly selective changes from adaptation to the vector but the species barriers could help maintain and evolve local signatures in the genome once the first adapted parasites are established in a population. As a consequence, newly imported strains would have a disadvantage unless there is compatibility between vectors they were adapted and are adapting to. Another reason to explore this connection with mtDNA is the current size and geographical distribution of the mtDNA database, which is comprehensive enough to explore inter- and intraregional levels of variation and possible correlations with mosquito vector distributions.

2. 5.2 METHODS

As there is a significant differentiation between mosquito species, and malaria is known to adapt to that level, it was decided to compare genetic variation in *Pv* to variation at the mosquito species level.

A mathematical model was set up to calculate vector differentiation between different regions. To do so, the major vector species were taken from Sinka *et al.* (2012) for each region (region x vs. region y), and the number of species shared by two regions ($m_{x=y}$) were divided by the total number of different species present in the two regions ($m_{x\neq y} + m_{y\neq x} + m_{x=y}$). The vector overlap index (V_{xy}) is calculated as follows:

$$V_{xy} = \frac{(m_{x=y})}{(m_{x\neq y} + m_{y\neq x} + m_{x=y})}$$

Values for V_{xy} can be found in the appendix in *Table S2.5*. Geographic distances between sample sites were calculated in the same way as that for the section on geography. Genetic variation (D_{ch}) is also taken from the sections above. For the comparison between vectors and geographic distances no samples sites were removed ($n=816$), besides the non-specified African ($n=1$) and South Asian ($n=1$) samples – as sample size does not influence either value. Correlations were calculated with the package *Hmisc* in R (Harrell *et al.* 2018), using the Spearman rank test because V_{xy} values behaved like discrete variables more suited to rank tests.

Table 6 gives an overview of the major vectors, the phylogenetic clade and series (Sinka *et al.* 2012, Harbach & Kitching 2016, Harbach 2013, *vectorbase.org*) of the major vectors and the dominant geographic haplotype of the *Pv* samples collected for the regions as categorised in *Table S2.2*: ‘Country 2B’. The column giving the major vector species in *Table 6* is colour-coded according to the colour in which the respective sample region lies in *Figure 11*.

Using the information on the distribution of mosquito species from *Table 6*, and the DPCA groupings and regional frequency of the *Pv* mtDNA from *Chapter 2.2*, a heatmap-style data overview was made (*Figure 14*). This data overview enables visual comparison of the frequency each species and each DPCA group was present within every geographical region. It was made using *ggplot2* in R, with each bar colour representative of the frequency at which the category was present.

Population	Full Name	N	Main PvmH Haplotypes	DAPC Group	Major Vector Species*			Vector Clade**	Vector Series**
MAD	Madagascar	66	IOC1 (& IOC2)	4	AB FT GB			Cellia, Cellia, Cellia	Pyrethophorus, Myzomyia, Pyrethophorus
EAFR	East Africa	3	IOC1 & IOC2	4	AB FT GB			Cellia, Cellia, Cellia	Pyrethophorus, Myzomyia, Pyrethophorus
WAFR	West Africa	4	IOC1	4	AB FT GB			Cellia, Cellia, Cellia	Pyrethophorus, Myzomyia, Pyrethophorus
ETH	Ethiopia	9	IOC1	4	AB FT			Cellia, Cellia	Pyrethophorus, Myzomyia
SUD	Sudan	34	IOC1 & IOC2	4	AB FT			Cellia, Cellia	Pyrethophorus, Myzomyia
TUR	Turkey	26	IOC1 & IOC2	4	SP	SC		Cellia, Anopheles	Neocellia, Anopheles
WAS	West Asia	5	IOC1 & IOC2	4	SP			Cellia	Neocellia
UZB	Uzbekistan	9	IOC1	4	SP			Cellia	Neocellia
AFG	Afghanistan	2		4	SP	ST		Cellia, Cellia	Neocellia, Neocellia,
IRA	Iran	2		1	SP	ST	FV	Cellia, Cellia, Cellia	Neocellia, Neocellia, Myzomyia
PAK	Pakistan	54	IOC1 & IOC2	4	CF ST FV			Cellia, Cellia, Cellia	Myzomyia, Neocellia, Myzomyia
IND	India	32	IOC1 & IOC2	4	CF ST FV			Cellia, Cellia, Cellia	Myzomyia, Neocellia, Myzomyia
BAN	Bangladesh	8		2	DR MN			Cellia, Cellia	Neomyzomyia, Myzomyia
MMR	Myanmar	32	MMR	4	DR MN			Cellia, Cellia, (Anopheles)	Neomyzomyia, Myzomyia
THA	Thailand	44	MSEA	2	DR MN			Cellia, Cellia	Neomyzomyia, Myzomyia
CAB	Cambodia	62	MSEA	2	DR MN			Cellia, Cellia	Neomyzomyia, Myzomyia
VIE	Vietnam	17	MSEA	2	DR MN	SN		Cellia, Cellia, Anopheles	Neomyzomyia, Myzomyia, Myzorrhynchus
CHSTH	China South	34	EAS1 & EAS2	3 (4)	LT SN			Anopheles, Anopheles	Myzorrhynchus, Myzorrhynchus
CHGU	China Guizhou	26	CEA	3	LT SN			Anopheles, Anopheles	Myzorrhynchus, Myzorrhynchus
CHAN	China Anhui	36	CEA	3	LT SN			Anopheles, Anopheles	Myzorrhynchus, Myzorrhynchus
KOR	Korea	35	EAS1 & EAS2	3 (4)	LT SN			Anopheles, Anopheles	Myzorrhynchus, Myzorrhynchus
BOR	Borneo	4	BOR	3	BB	BR	LS	Cellia, Anopheles, Cellia, Cellia	Anopheles, Myzorrhynchus, Neomyzomyia, Pyrethophorus
IDO	Indonesia	47	ISEA	3	BB	BR	FL	Cellia, Anopheles, Cellia, Cellia, Cellia	Anopheles, Myzorrhynchus, Myzomyia, Neomyzomyia, Pyrethophorus
PNG	Papua New Guinea	190	MEL1 & MEL2	1	FR KL PL		SD	Cellia, Cellia, Cellia	Neomyzomyia, Neomyzomyia, Neomyzomyia
SOI	Solomon Islands	21	MEL2 (& MEL1)	1	FR KL PL			Cellia, Cellia, Cellia	Neomyzomyia, Neomyzomyia, Neomyzomyia
VAN	Vanuatu	11	MEL2	1	FR			Cellia	Neomyzomyia

* Sinka et al. 2012

** Harbach & Kitching 2016, Harbach 2013, vectorbase.org

Table 6.

Overview of information gathered on sampling populations: main haplotypes, DAPC group, main mosquito vectors (species, clade and series. Species coloured according to the dominant vectors for the sampling region in Figure 10) and climate zone (coloured according to main climate of the sampling location/country as given in Figure 8 and Peel et al. 2007). Populations have been grouped into main geographic regions, from west to east (Africa, West and Central Asia, South Asia, Southeast Asia including Bangladesh, Northeast Asia, Island Southeast Asia, And Melanesia).

2. 5.3 RESULTS & DISCUSSION

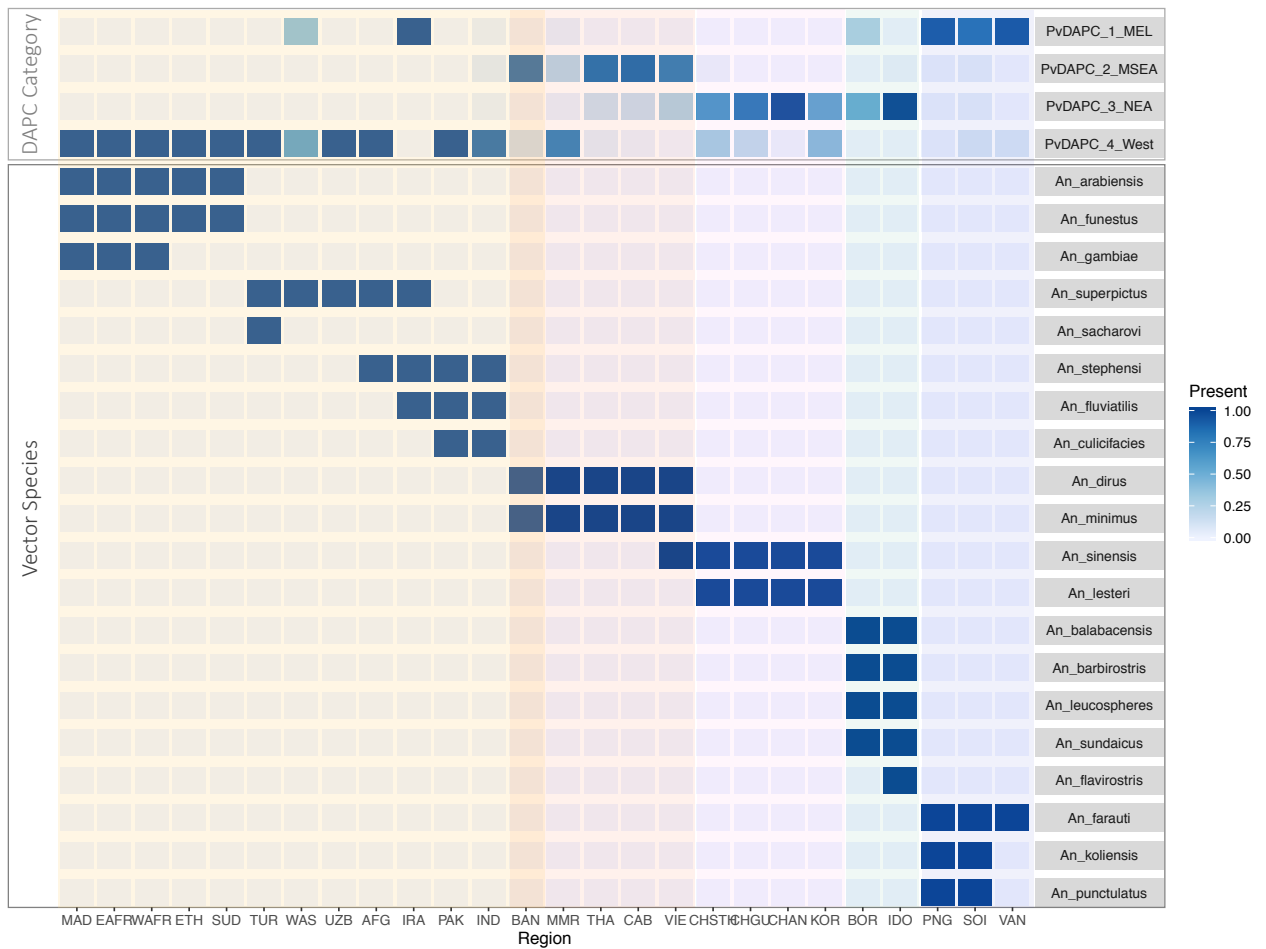


Figure 14. Regional Mosquito Vector Species Distribution and Pv mtDNA DAPC

Overview of frequency with which each DAPC subgroup is present in the Pv mtDNA samples from the different geographic regions (out of total sample number from each region; PvDAPC_1_MEL – PvDAPC_4_West), as well as the recorded presence or absence (1 or 0 respectively) of *Anopheles* mosquito vector species as a major vector across each geographic region (*An_arabiensis* – *An_punctulatus*). Geographic regions are named according to the abbreviations used in Table S2.2.

The correlation between vector species overlap with geographic and genetic distance is skewed, as there are many data-points with no major vector overlap between regions ($V_{xy} = 0$). Working with these values showed the data was not well suited to this format of calculating correlations, for several reasons including this high number of regions without any degree of vector correspondence. The results were however still significant, in favour of negative correlations between vector and geographic distance (Spearman signed rank correlation coefficient: $r_s = -0.54$, $p = 0.00$), and vector and *Pv* D_{ch} (Spearman signed rank correlation coefficient: $r_s = -0.54$, $p = 0.35$). For this reason the results have been added in the appendix, *Appendix S2.8*. In fact the correlation between vector species and *Pv* is much more clearly visible when plotting the vector species presence together with the *Pv* DPCA for all the regions, as shown in *Figure 14*.

A clear connection between the major vector species and the differentiation of *Pv* mtDNA is shown in *Figure 14*. Island Southeast Asia versus Melanesia shows an interesting divide, where vector species and malaria phylogeny vary but climate across this region is generally similar (Peel *et al.*

2007, Sinka *et al.* 2012). A difference in species is found across many different clades and species, such as plants and animals, and is usually observed as a general divide in species to the east and west of what is called the Wallace line (Mayr 1944, Camerini 1993, Barber *et al.* 2000), which is geographically noted by the deep-sea trough formed by the Lombok straight to the east of Java and Borneo and denotes the edge of the old Sunda continent. Nevertheless, as seen in *Figure 13*, the mosquito vector species composition in this region changes across another geographic divide: the region between the Lydekker and Weber lines. These lie slightly further to the east and denote the edges of the old Sahul continent, which combined Australia and Papua New Guinea (Mayr 1944, Gaylord 1977). All three lines show divisions transversing several zoological taxa and are established geographic divides between the Oriental and Austral regions, even though climate across them is continuous (Peel *et al.* 2007, Sinka *et al.* 2012).

Another differentiation between vector species and climate is seen between Southern China and the rest of more northern East Asia. The major vector species for this region are the same as for the remainder of China and Northeast Asia, yet climate varies and is grouped with the northern parts of mainland Southeast Asia. The Southern China phylogeny clusters only with Northeast Asia, suggesting the connection between these regions is neither geographic nor climatic.

A point of interest is the major phylogenetic split in *Pv* mtDNA between East & Island Southeast Asia and the rest of the samples. Though looking at the mosquito vectors at a species level does not show clearly that there is a vector-related dynamic behind this split, a strong differentiation between these two phylogenetic groupings becomes clear at the mosquito clade level, as seen in *Table 6*. The *Pv* mtDNA phylogenetic split is echoed in the major vector clades for these regions: most regions are entirely dominated by mosquito vectors from the Cella, whereas both East Asia and Island Southeast Asia share the characteristic of having major *Anopheles* representation. This is of special interest as for both Anopheline mosquitoes and for *Pv* mtDNA, the major split in their phylogenies aligns across the same geographic regions.

2. 6. OVERALL RESULTS & DISCUSSION

There are strong indications that all three of the factors explored in this chapter, human migration, geographic distance and mosquito vector species, have had a significant role in shaping the genetic diversity of *Pv* mtDNA. It is obviously hard to separate out the interaction between these factors, and they are expected to be closely connected to each other. Human migration over time includes a link to geographic distance over time, and mosquito species will vary over distance, just as general geographic climate parameters will generally change over distance too. However, for human migration there are levels in which human historical movement is present in *Pv* genetics, such as seen in the PCA (*Figure 5*) and Bayesian Phylogenetic Tree (*Figure 4b*), which are not directly associated with geographic distance.

For instance, in *Figure 5* the PCAs show population movement and isolation of small populations in China (*Figure 17 F & H*) that are not a direct correlation to geographical distance in relation to its supposed African source. In fact there are many interesting patterns across this East Asian region for which connections with both recent and more remote human history would be an interesting avenue of research. The patterns seen here appear to confirm the suggestion by Miao *et al.* (2012) that *Pv* across East Asia indicate an ancient presence. The mtDNA across China and Korea is stratified into an interior haplogroup found in Anhui and Guizhou, and a more coastal haplotype found in Southern China and Korea. *Pv* samples linked to Korea are the ones on or near the southern Chinese coast. As these two regions are the furthest apart on a continental route, the link between the two regions is more likely to be coastal – possibly an old marine trading link. Unlike the links across and around the Indian Ocean, the East Asian routes do not appear to reach far inland in either location.

Though currently still speculative, the pattern in human movement found in aDNA data as reviewed by Skoglund & Mathieson (2018) suggests strong mixing of human populations to the west of India throughout much of human history. The current findings in East Asia on the other hand indicate strong stratification and possible long-term isolation/continuance of populations. These patterns would agree with the current distribution seen in the *Pv* mitochondrial genome. However, the human data is based on a limited number of samples (2-3 in East Asia), and this observation comes with limited historic knowledge of human movements in the region. As such, caution is required in making general conclusions based on such limited data, and an in-depth study of the dynamics in East Asia specifically would be of great interest, especially with future additions to the archaeological dataset.

Another indicator of a connection with human movement in the past that does not have a direct correspondence to distance is the relationship of ISEA *Pv* with that of EAsia, as seen in *Figure 4*,

and other results across section 2.2. Geographically the expectation is for ISEA to take up a position as an intermediate between Melanesian and either Myanmar or MSEA *Pv*. Instead this link indicates the movement of the disease from Eastern Asia to ISEA, considerably far removed in the past to be associated with the first anatomically modern humans - or earlier - making this southern movement. Possibly the vectors from source and sink populations were more compatible, as it is assumed that at least some *Pv* strains from another introduction coming through mainland southeast Asia and ending in Melanesia would have been present before this introduction.

The difference in these two geographically close but genetically distant lineages found in ISEA and MSEA/Melanesia correspond to a broad lineage split found in both *Pv* and humans. The human mitochondrial lineages M and N, as seen in *Figure 7*, correspond well in geographic distribution with the two main lineages seen in *Pv*, as shown in the phylogenetic tree in *Figure 4b*. This distribution indicates links between human and *Pv* lineages and provides a strong basis for the argument that the *Pv* currently found across the Eastern Hemisphere spread to its current locations mainly with the earliest known anatomically modern human beings to migrate out of Africa, suggesting its' spread neither post- or predates this migratory movement. However, the dating of this split in *Pv* lies well before that of the human lineages (*Figure 10*). There is some room for arguing an older date might be appropriate for humans, as archaeological findings in Australia and Indonesia are constantly pushing back dates of anatomically modern human presence in the region (Clarkson *et al.* 2017, Westaway *et al.* 2017). These dates however would not go as far back as the current *Pv* lineage age. An improvement of *Pv* dates, for example with an enlarged data-set, improved mathematical model, or better understanding of the mutation rate of the mitochondrial genome, could make a much larger difference in this date estimate, especially as the error margin increases the further back in time/lineage the calculations go.

The correspondence of the introduction of *Pv* to certain regions with archaeological evidence of *H. sapiens*, as seen in *Figure 10*, also underlines this connection between *Pv* and *Hs*. The estimated age of *Pv*'s most recent common ancestors predating that of modern humans underlines its long-established presence. Though the dates, as well as the presence of *Pv* type and other malarial infections found across all species of ape, indicate a long association with hominids even before the rise of modern humans, it appears it was *H. sapiens* that spread the disease on a global scale. At the very least the form of *Pv* found today.

Lastly, the flat distribution of *Pv* mtDNA variation from Africa to India indicates the lack of the usually isolating influence from increased geographic distance across this region. Instead the lack of variation indicates consistent contact across the entire region, over extended periods of time. This corresponds with the very lengthy history of trading around the Indian Ocean, from India to Madagascar. However, the fact that this connection reaches considerably inland across the entire region, including Mauritania, Niger, Turkey and Armenia, which are expected to have been more

isolated from this contact, indicates there is more contributing to the pattern in *Pv* genetics than this one simplified historic movement of humans. From the PCA in *Figure 5* some distinction is found in the samples from Uzbekistan (UZB/CAS), denoting that some isolation of this region from the main Africa-South Asia axis was present.

In contrast the historic sample of *Pv* from Spain is supposedly more different from the Indian Ocean subtype (Gelabert *et al.* 2016), which suggests an isolation of Europe from this widespread bi-continental interaction. Though North African samples are missing from this data set, the European fingerprint as suggested by the century-old Spanish sample indicates isolation from Africa (including north-western Africa as denoted by Mauritanian samples) and western Asia (including the geographically continent-bridging Turkey). Historically southern Spain, where the sample originates, has strong human connections with these regions or at least areas close to them. In this light it would be of interest to obtain sequences from *Pv* samples from North Africa as well as further (historic) samples from Europe, to gain confirmation for the findings of the published study and an increased understanding of the genetic landscape across the region. This is especially important for underlining claims of a possible European import of *Pv* to South America (Taylor *et al.* 2015, Gelabert *et al.* 2016, Rodrigues *et al.* 2018).

The interaction of geographic distance and mosquito vector as a combination on the genetic diversity of *Pv* can be partially discounted as it was tested negative (or non-positive) in the ANOVA tests of variance. Separately they do influence *Pv* mitochondrial diversity, with IBD present east of India and a generally strong correlation between local mosquito vector and mitochondrial genome variation. This is the first study that looks at the correspondence of vectors and *Pv* mitochondrial genetics at this scale. It gives a good indication that the presumed connection between vectors and *Pv* genetics, as shown previously in smaller local studies (Joy *et al.* 2008), is extendable to a global level even in non-vector associated mtDNA regions. This agrees with results across the entire genome as found by Hupalo *et al.* (2016), but contrasts in that it is found in non-vector associated parts of the genome.

Issues with making correlations with vector related data are: the great diversity in *Anopheles* species, their relatedness and their distribution. In addition, vector species identification is not always reliable, which complicates already patchy distribution measures (Sinka *et al.* 2012). To complicate matters further, even when identified, the role of major and minor vectors in regions is not well understood. Minor vectors could be the major distributors of anthropolimnetic disease, whereas major vectors could be more drawn to other animal hosts (Ubalee *et al.* 2016). An additional complication is that even subpopulations within species have been shown to have diversified and adapted up to a genetic level to climate, variation in host preference, resting habits, infection rates and differentiation in human environment, including changes in human behaviour (Beebe & Cooper 1999, Govella & Ferguson 2012, Killeen *et al.* 2017, Thomsen *et al.* 2017).

In sum, which vectors are present, which vectors are the major carriers and distributors of anthropomorphic disease, and how vector species are related to each other complicate the understanding of differences in species association between vectors from different regions. Besides the implications this has for complicating the understanding of the epidemiology and transmission of malaria, this makes creating a measure for vector association across regions a very complex task. However, the section on vectors in this chapter makes a start with a very simplified attempt that compares the number of shared major vector species between sampling regions, and finds strong patterning even at this level. In *Figure 14* it is shown to be a good fit with *Pv* clades, and overall a much better fit than geographic distance (IBD).

Most notably, this simple start of comparing major vectors with parasite genetics already suggests vector species have a major influence on mitochondrial genome distribution of *Plasmodium vivax*. More so than what is currently detected in genes involved in processes within the mosquito host, such as *Pvs47*. Though the mitochondrial genome lacks genes directly involved in processes to do with the mosquito vector, the compatibility present within similar vector host species has probably greatly influenced the genomic patterns seen. The close association of dominant clade types, mainly *Cellia* versus *Anopheles* (see *Table 6*), is echoed in the main split in the phylogenetic tree (see *Figure 4b*). *Pv* mtDNA lineages from East Asia and Island Southeast Asia cluster together, in opposition to samples from the rest of the eastern hemisphere. As human movement from Mainland Southeast Asia to Melanesia would have passed through Island Southeast Asia, it is expected that it would form a medium between the two. The pattern seen does not correspond to this expectation. Instead the connection with East Asia puts up an expectation that the current *Pv* in Island Southeast Asia is a later introduction by humans coming from China through Taiwan and the Philippines. The connection is emphasised by the samples from Borneo taking up a bridging position between the two.

It remains questionable how and when the Island Southeast Asian samples were introduced. Assumedly before its introduction *Pv* forms were present that were related to Mainland Southeast Asia and Melanesia. The current form coming from East Asia might have had a closer affinity to the *Anopheles* clade, which is well represented in both regions, and infiltrate to dominant levels from that platform. Human migrants have had a similar dynamic. However, the timing of this movement for humans and for *Pv* does not correspond. The estimated age for the most recent common ancestor (TMRCA) for ISEA and East Asia *Pv* falls a factor 10 earlier than that of the estimated arrival time of humans from Eastern Asia. In fact, the 73 kya (53 - 108) of ISEA *Pv* fits well around the estimated TMRCA for Mainland Southeast Asia (74 kya, 52 - 104) and Melanesia (72 kya, 51 - 97). Even so it is improbable that the current form of *Pv* in ISEA is originally the intermediate form from MSEA and Melanesia that has adapted/evolved in a similar direction as East Asian *Pv* due to vector selection dynamics. With the mtDNA lacking genes involved in

mosquito vector interactions, direct selection on mitochondrial genes for the vector species is unlikely.

The discord between date and route connections with surrounding regions is something to be looked into further in the future, and could point towards issues with either the underlying mathematical model or simply a lack of data. The increase of sample numbers has previously improved date estimates for many of the regions, including Mainland Southeast Asia and Melanesia. Alternatively, the *Pv* mitochondrial genome might be indicating human migratory movements between regions at a much earlier date than other data sources have detected so far. This trend is already shown for the Australasian region with the most recent finds of anatomically modern humans in Australia and Indonesia pre-dating previous finds, and simultaneously corresponding better with the *Pv* age estimates (Clarkson *et al.* 2017, Westaway *et al.* 2017). As this region is of great archaeological and malaria epidemiological interest, *Chapter 3* will give an in-depth analyses of the human migration dynamics associated with the spread of *Pv* in the Melanesian region.

From the distribution of the genetic diversity in *Pv*, ancient human migration seems to have played a major role in its initial distribution. Subsequently, over time, the interaction between human migration and mosquito vector species, both influenced through geographic parameters such as climate and distance, appears to have either maintained and enhanced the difference or lessened it. The two different dynamics are seen across the two spaces to the east and west of India respectively. There is much room to explore the further influence of geographic barriers and connectors, such as mountainous regions, deserts, rivers and oceans, on the current distribution of *Pv*. The influence of several of these on the distribution of human populations has been shown by Pagani *et al.* (2016), and is expected to have a similar impact on that of the genetic distribution of *Pv*.

One factor left out of this overview of genetic variation in *Pv* mtDNA is selection for resistance by anti-malarial drugs. So far mutations on only one gene, cytochrome b (*cytb*), in the *Pf* mitochondrial genome have been found to play a role in resistance to the anti-malarial drug atovaquone (Fisher *et al.* 2012). For the *Pv* dataset used in this study, 817 samples had the same nucleotide sequence at this locus. Only one had a single synonymous substitution on the very last nucleotide for the sequence encoding the quinol oxidation site. As such it seems that the mitochondrial genome is relatively free from drug resistance selection mechanisms, at least in direct association.

2. 7. OVERALL CONCLUSION

Results indicate that besides each independent factor influencing patterns of distribution, interactions are of clear additional interest. There is evidence of the earliest anatomically modern human migrations, emerging from Africa over 100.000 years ago, introducing malaria into different geographic regions across the eastern hemisphere. Vector species differences and geography either maintain and enhance this pattern, or reduce it. Specifically in the east, varying mosquito vector species between geographic regions (e.g. Myanmar, China, Melanesia) appear to form a barrier to colonisation by new Pv strains. Combined with relatively low levels of human migration, this interaction has probably maintained differences in the mtDNA as introduced by ancient human migrations.

Further in-depth study is necessary for several of the factors on their own, including the development of a stronger framework for vector species representation, as well as the interactions of the factors, especially within certain geographic regions such as East Asia. Even so, this chapter shows that studying interactions between factors influencing evolution and distribution helps explain genetic patterns that cannot be understood in terms of individual causal variables.

INDEX CHAPTER III

<u>3. 1. INTRODUCTION</u>	<u>82</u>
<u>3. 2. METHODS</u>	<u>87</u>
<u>3. 3. RESULTS</u>	<u>92</u>
<u>3. 4. DISCUSSION</u>	<u>99</u>

CHAPTER III

A PARASITE'S PAPUAN PAST: AGE AND ORIGIN OF MELANESIAN *PLASMODIUM VIVAX* MALARIA RECONSTRUCTED FROM MITOCHONDRIAL GENETIC PHYLOGENY

*"From the day the universe was formed
Past infinite worlds
Probably in our past life and in our next lives
We'll be together forever" – BTS, DNA*

3. 1. INTRODUCTION

The Melanesian region is well known for the remarkable cultural and genetic diversity of its human populations. Globally, the area is also one of the most complex in diversity and transmission of malaria. In numerous past elimination efforts across the world, *Plasmodium vivax* malaria has been an especially persistent survivor. *P. vivax* (*Pv*) is geographically the most widely distributed human malaria species, and exhibits high global genetic diversity. The disease still has a profound impact on human populations, ranging from death to stunted opportunity for economic growth, which likely extends throughout the human past. However, the origins and drivers of the disease's diversity and distribution are still largely unknown. This lack of knowledge affects estimates of its past and present impact, and our understanding of its epidemiology and evolution. It is assumed that much of the current distribution and diversity can be attributed to its long evolutionary history and association with past human migrations. To explore these assumptions within a smaller but diverse geographical setting, this chapter aims to understand the origin and distribution of *Pv* in and around the complex Melanesian region. Accordingly, phylogenetic and population genetic methods were applied to a large dataset of global *Pv* mitochondrial genomes. The results from the genomic analyses presented demonstrate that much of the unique mitochondrial genetic diversity of *Pv* currently found across the East Asian, Southeast Asian and Pacific regions is due to its early spread, predating 65,000 years ago. The patterns seen in the mitochondrial genome of the parasites across these regions correspond with those broadly observed in human mitochondrial and whole genome data, concluding that the parasite likely migrated to Melanesia with the earliest anatomically modern humans to move out of Africa along a southern coastal route across the Asian continent.

The Melanesian region, reaching from island New Guinea in the west to Fiji in the east, constitutes a frontier for the distribution of human malarias. Western Melanesia (New Guinea, Solomon Island and Vanuatu) is notorious for the high frequency and pronounced diversity of human malaria

parasites and mosquito vectors, whereas Fiji and New Caledonia lack appropriate malaria vectors (Belkin 1962) and have thus always been malaria free. Previous studies looking at global diversity of the *Plasmodium vivax* malaria mitochondrial genome (mtDNA) have shown geographic specificity and high mtDNA diversity in New Guinean *Pv* populations (Mu *et al.* 2005, Jongwutiwes *et al.* 2005, Miao *et al.* 2012, Taylor *et al.* 2013). Such a pattern indicates that *Pv* may have had a long, isolated history in Papua New Guinea (PNG). This extensive Papuan past for *Pv* combined with a diverse and well-studied human ethnography makes the Melanesian region a strategic choice to investigate *Pv*'s connection with human history. This paper focuses on determining when, and from where, *Pv* was introduced to the Melanesian area.

Current interest in studying the spread of diseases in conjunction with human migratory movements in present-day, historic and prehistoric times has yielded a variety of new insights into the past of both humans and the disease agents. Amongst other things, such studies have indicated ages of certain diseases to be older than first thought (*M. tuberculosis*: Donoghue 2009, Gagneux 2012, Comas *et al.* 2013; *H. pylori*: Moodley *et al.* 2012), discovered variations in patterns of genetic adaptation in the disease agent (such as variance in virulence; see Pepperell *et al.* 2013, Comas *et al.* 2013) and provided new, or increasingly fine-scale, insight into human migration patterns (Bos *et al.* 2011, Moodley *et al.* 2009, Moodley *et al.* 2012).

Together with its distant relative *Plasmodium falciparum*, the malaria parasite *Pv* is still one of the main causes of death in low-income countries (2013, WHO). In both the first campaign to eliminate malaria in the 1960s and 70s, and the current one, *Plasmodium vivax* transmission has proven to be particularly hard to interrupt (Sattabongkot *et al.* 2004, Gunawardena *et al.* 2013, Waltmann *et al.* 2015, Kittichai *et al.* 2017). This includes elimination efforts across Melanesia. Previous studies looking at genetic diversity in malaria parasites have helped monitor the success of recent control and inform the strategy of elimination efforts (Abdullah *et al.* 2013, Noviyanti *et al.* 2015, Fola *et al.* 2018). The focus of this paper differs in that it seeks to understand the long-term evolutionary dynamics behind the distribution of the disease. Knowing the chronology and distribution patterns of *Pv* creates the opportunity to understand the role of past human migratory dynamics in parasite diversity and propagation. In addition, insight into these connections can increase understanding of human population dynamics, long-term human-parasite-vector interactions and subsequent genetic adaptation of each to the other.

One of the main reasons for the failure to eliminate *Pv* is its ability to create a dormant form infecting liver cells. Dormant hypnozoites can remain hidden for weeks to months, up to four years, until they are released into the bloodstream and a re-infection occurs (Boyd 1949, White 2011, White *et al.* 2016). The period of dormancy provides the parasite with the opportunity to bridge periods of time, whether seasonal or geographic, without its main *Anopheles* mosquito vector. The occurrence of *Pv* throughout temperate regions such as Finland and northern Russia (Huldén *et al.* 2005), and other parts

of coastal northern Europe up until the 1960s (Boyd 1949, Verhave 2013), was aided by this specific adaptation. In short, this trait facilitates both the parasite's spread and survival, and gives *Pv* the ability to have spread more readily and to sustain itself within smaller host populations than *P. falciparum* (*Pf*). Studies comparing *Pf* to *Pv* at similar transmission, suggest that *Pv* uses this strategy to maintain population fitness (e.g. Jennison *et al.* 2015, Noviyanti *et al.* 2015). Though a tougher parasite to eliminate and more widespread across the globe, the origin, evolutionary history and biological processes of *Pv* are less well understood than those of *Pf*.

At this time, studying connections between early human migrations and the spread of malaria is facilitated by the recent evidence for *Pv* African origins dating to before 200,000 years ago (Rayner *et al.* 2011, Culleton *et al.* 2011, Prugnolle *et al.* 2013, Liu *et al.* 2014), increasingly better methods for studying phylogeny, and a growing global database of genome sequences. Even so, few in-depth studies have been conducted to properly identify sources and timings of *Pv* introduction into different geographical regions, with the exception of a recent study by Rodrigues *et al.* (2018), which explores the origins of South American *Pv*. By contextualising the origin of Melanesian *Pv* within a broader regional setting, this chapter also sets up the opportunity to start exploring the dynamics behind its Asian diversity. The evidence for an African origin of *Pv* is of special interest, as Asian and Southeast Asian regions have long been seen as more likely places of origin, partially due to the large genetic diversity presented across the region (Carter & Mendis 2002, Escalante *et al.* 2005, Mu *et al.* 2005, Cornejo & Escalante 2006, Tachibana *et al.* 2012) and the presence of the closely related non-human relapsing malaria species *P. cynomolgi* and *P. simiovale* in Southeast Asian monkeys (Escalante 2005, Jongwutiwes *et al.* 2005, Mu *et al.* 2005).

Melanesia's high human genetic diversity relates to migrations from two very different human populations: the Papuans and Austronesians (e.g. Soares *et al.* 2016, Montenegro *et al.* 2016, Skoglund *et al.* 2016). Both human populations display genetic adaptations to malaria (Serjeantson *et al.* 1992, Howes *et al.* 2013), indicating long-term exposure, but have very different geographic and historic origins. With the Indigenous peoples in Australia, the Papuans are believed to be representatives of the earliest waves of anatomically modern human migrants that moved out of Africa. They reached New Guinea from around ~65,000 – 45,000 years ago (ya) (Lum 2007, Denham *et al.* 2009, Gosden 2009, Kirch 2010, Gosden 2010, Summerhayes *et al.* 2010, Pagani *et al.* 2017), following on the expansion from Africa through a southern Asian route between at least 75,000 and 62,000 ya (Rasmussen *et al.* 2011).

During the Palaeolithic, 15,000 – 30,000 ya, contact between Indonesia and continental SE Asia was present west of the Wallace line. Human Y-chromosomal genetics indicate a significant split in human population structure thought to relate to limited Palaeolithic human movements between Bali and Flores, two Indonesian islands divided by the deep ocean trough that makes up the Wallace line. Melanesian lineages remain dominant east of Flores while western Y-chromosomes relate more to mainland Asia. More recent Neolithic human migrants such as the Austronesians appear to have had

a limited impact on the paternal gene pool in West and Central Indonesia, despite large cultural impacts (Karafet *et al.* 2010). Around 9,000 ya, a sea-level rise increased distances between the Eurasian Sahel and the Australo-Melanesian Sunda continents (Chappell 1993). This effectively isolated the Papuan and Aboriginal Australian populations from the rest of the world.

The second great movement of modern humans in the area, which reached Melanesia, was the arrival of the Austronesians from ~6,000 – 3,500 ya (Skoglund *et al.* 2016, Montenegro *et al.* 2016). This migration event is generally called the Austronesian movement, and the Austronesian settlers reached all the way to New Zealand and Hawaii in the Pacific, and Madagascar across the Indian Ocean. Though Austronesian genetic and linguistic origins remain a much-debated topic, they are often linked to ancient tribes in Taiwan (Friedlander *et al.* 2008, Moodley *et al.* 2009, Rasmussen *et al.* 2011, Skoglund *et al.* 2016, Montenegro *et al.* 2016). Archaeologists, anthropologists and geneticists have studied the movements of the Austronesian population around the Indo-Pacific extensively through disciplines as varied as linguistics (Tryon 1995, Kirch 2010), human genetics (Friedlaender *et al.* 2008, Vilar *et al.* 2008, Soares *et al.* 2016, Skoglund *et al.* 2016), commensal species genetics (overviews: Matisoo-Smith 2009; pigs: Larson *et al.* 2007, Dobney *et al.* 2008; chickens: Storey *et al.* 2007, 2012; rats: Matisoo-Smith & Robins 2004; rice: Donohue & Denham 2010, Blench 2011), human gut bacteria genetics (Moodley *et al.* 2009), pottery (see chapters in: Sheppard *et al.* 2009) and stone tools (Groube *et al.* 1986, Evans & Mountain 2005, Specht 2005). Conclusions tend to vary depending on the reference data used and the group interpreting it, but the general consensus is that the large-scale patterns of distribution and past human population connectivity broadly follow the two migrations as presented above.

The precise origin, movement and time of emergence for the Austronesian people has been under debate for decades. Opinions are divided between local emergence in the Indonesian region versus movement from the Taiwanese area through the Philippines; and a slow movement model with admixing migrant and local populations versus a fast model without much initial contact; and any set of different combinations of the models. Four recent publications (Lipson *et al.* 2014, Soares *et al.* 2016, Montenegro *et al.* 2016, Skoglund *et al.* 2016) confirm the generally more accepted northern Taiwanese origin of Austronesians in the Melanesian region. Skoglund *et al.* and Montenegro *et al.* prescribe a rapid colonisation of Melanesia and Micronesia, at least up to the Bismarcks (Montenegro *et al.* 2016), without much initial interaction with the earlier settlers¹. Lipson *et al.* and Soares *et al.* support an early Austronesian entry (~10 kya) to ISEA laying the basis of Austronesian genetics, and a later (4 kya) re-entry that spread much wider aided by the earlier population. All four papers support the

¹ Montenegro and colleagues model routes of dispersal based on climatic and oceanographic data, Skoglund *et al.* find evidence for the widely accepted “Express Train to Polynesia” model (e.g. Kirch 2010, and related option of the ‘leapfrog model’ presented by Sheppard, 2011) in ancient genomes from Vanuatu and Tonga. Skoglund suggests that though Austronesians were the first to settle the Pacific east of New Ireland, there was substantial movement later by both Austronesian and Papuan populations that led to the presence of Papuan populations, especially men, further into the Remote Pacific region. Soares *et al.* show genetic evidence for a much earlier, 6-10 kya, expansion of first migrants out of Taiwan, with smaller migrations later on bringing the Austronesian languages and cultures. Montenegro shows that oceanographic and climatic models accommodate theories of fast movement over long distances.

mixing of culture and movement of Austronesians and Papuans in the Melanesian region after initial settlement.

By the time the first Europeans arrived in the 16th century, malaria was already prevalent. It took its toll, amongst others, on the Spanish explorer Mendaña, the first European to reach the Solomon Islands. As malaria is most likely to travel within infected humans and its presence in Melanesia appears to go back at least several thousands of years, it is probable that the Papuans, the Austronesians, or both, brought the parasite with them. Both populations show a wide range of genetic adaptations to malaria that indicate long-term exposure. Type and occurrence of these vary to great extent across the region. These adaptations specifically include Duffy heterozygosity (Zimmerman *et al.* 1999) and Gerbich-negative blood group (Serjeantson *et al.* 1992, Bellwood *et al.* 1995) in Papuan populations, and Southeast Asian Ovalocytosis (SAO) originating from Austronesian populations (Tsukahara *et al.* 2006). More generally the genetic adaptations present across Melanesia cover a large range of unique glucose-6-phosphate dehydrogenase (G6PD) mutations (Howes *et al.* 2013) and alpha- & beta-thalassaemia modifications (Serjeantson *et al.* 1992, Kaneko *et al.* 1998, Müller *et al.* 2003). Interestingly, several of these adaptations seem to provide protection predominantly against *Pv* malaria.

With the phylogenetic analyses of the *Pv* mitochondrial genome, this chapter explores the origins and the age of *Pv* in Melanesia and how it connects to *Pv* in surrounding parts of Asia. It confirms previous indications that the parasites brought to Melanesia are relatively old (Jongwutiwes *et al.* 2005, Mu *et al.* 2005, Taylor *et al.* 2013), and finds that *Pv* was an early introduction from Papuan migrants coming through Southeast Asia. In addition, the possible origins of the very different *Pv* strains found in neighbouring Island Southeast Asia are addressed, for the first time. Unlike the Melanesian *Pv*, Island Southeast Asian *Pv* appears to have originated from an early East Asian, possibly Austronesian, migration into the region. However, the origin and age of these parasites do not readily correspond with any of the known human histories of the region, confirming there are still many unknowns to the complex past of humans and disease in this region.

3. 2. METHODS

Samples were sequenced from previously obtained infected human blood, sampled for epidemiological and population genetic studies, across sites in Papua New Guinea (PNG), the Solomon Islands (SOI) and Vanuatu (VAN), as published in Arnott *et al.* (2014), Koepfli *et al.* (2012, 2013, 2015) and Waltmann *et al.* (2015). These samples intersected both culturally and genetically with Papuan and Austronesian populations. An overview of samples and associated ethnographical data for all samples with known geographic locations across Melanesia can be found in *Appendix S3.2*. The following projects provided infected blood samples: PNG: cross-sectional surveys in 2005/06 (Senn *et al.* 2012, Arnott *et al.* 2013, Jennison *et al.* 2015), PNG & Solomon Is: in-vivo drug efficacy studies 2003/04 (Koepfli *et al.* 2012, 2013), Solomon Is: cross-sectional survey in 2012 (Waltmann *et al.* 2015), Vanuatu: in-vivo drug trial 2013 (unpublished). Each project was carried out with local governmental and ethical permissions, from The National Health Research & Ethics Committee (NHREC) of the Government of Papua New Guinea Medical Research Advisory Committee, National Department of Health, Papua New Guinea; Ministry of Health & Medical Services, Solomon Islands; the Directorate of Public Health, Vanuatu; WPRO-ERC (WHO); and institute Human Ethics Committees at WEHI, Basel TPH and Menzies. Human ethics permission for the research presented here was obtained from the Walter & Eliza Hall Institute's (WEHI) human research ethics committee (HREC) in Melbourne, Australia, under WEHI-HREC project number 14/14. Ethics approval from the University of Oxford's Central University Research Ethics Committee (CUREC) will be applied for in the near future.

DNA from these samples was purified across several different labs with varying methods, but included extractions with QIAamp DNA Blood Mini Kits (Qiagen, Valencia, CA), processed according to the manufacturer's instructions. As it is unknown what kits were used for all the different extractions, the process should not have influenced the results of the PCR, apart from failure to amplify if DNA was low quality. Genotyping of highly variable nuclear size polymorphic markers (including: MSP2, MS8, MS16, MS20, Pv3.27) and merozoite surface protein 1 (*m*sp1F3) identified single-clone infections. From these, mitochondrial genomes were sequenced (n=150, from PNG, VAN & SOI), according to a modified two-fragment PCR protocol based on that by Culleton *et al.* (2011). The PCR amplifications of 3 and 3.5kb were accomplished with Roche High Fidelity Fast Start enzyme (Roche Diagnostics AG, Switzerland), according to the provided Roche protocol for long amplifications: 2 min at 94° C, 10 cycles of 30 sec at 94° C, 1 min at an optimised annealing temperature of 61° C, 3.5 min at 68° C, and then 30 cycles adding 20 sec to the extension time per cycle. Final concentrations of the reaction included 1x buffer with MgCl, as provided by Roche, 0.25mM dNTPs, 2.0U enzyme, and 0.5uM of each primer. See *Appendices S3.1.1 & S3.1.2* for precise protocol and primer details. PCR products were then Sanger sequenced at a commercial facility using an Applied Biosystems 3730xl platform. Primers used for PCR and sequencing (*Appendix S3.1.3*) were designed in-house, many based extensively on Culleton *et al.* (2011).

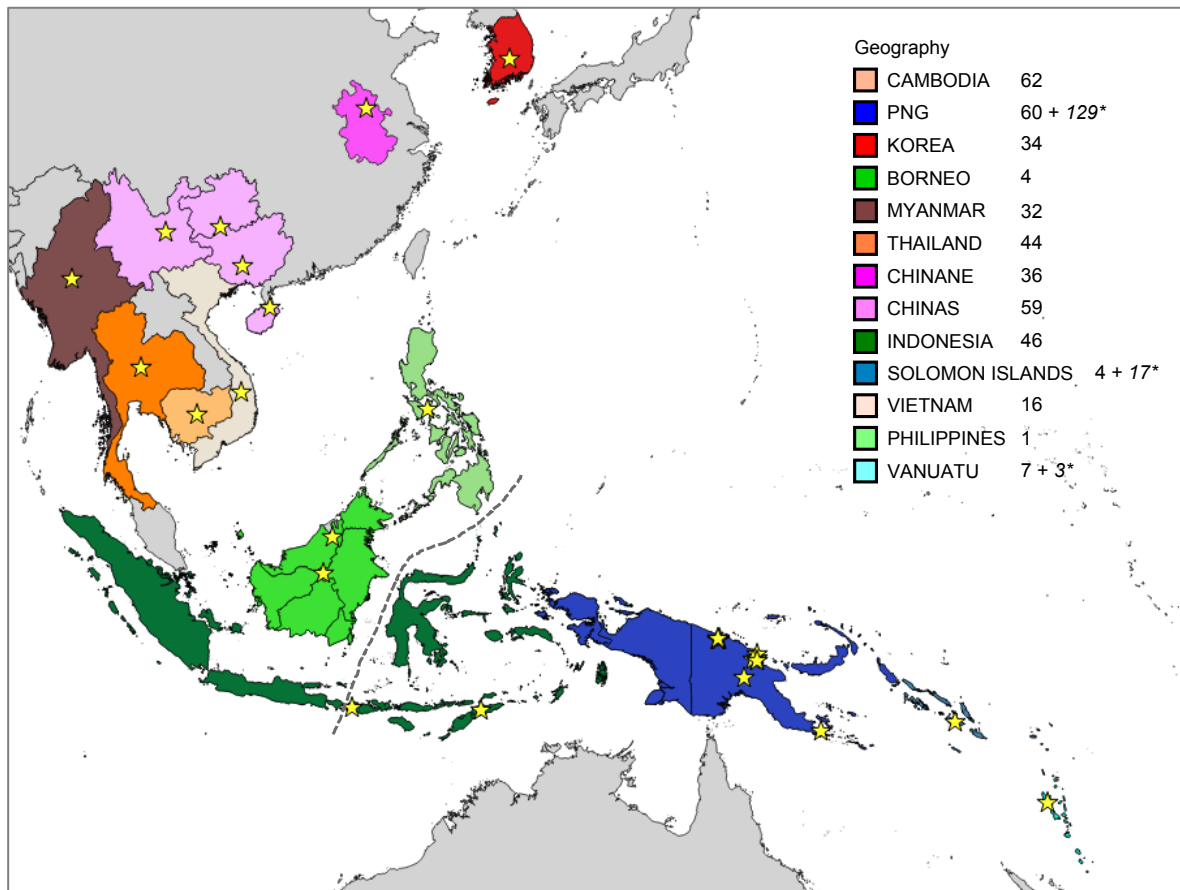


Figure 1

Sample sites across East Asia, Southeast Asia and Melanesia. Yellow stars indicate sample sites (approximated for those references with country or province indications); country of sample origin coloured according to the key. Sample numbers from each country given behind the key. The dotted line indicates the location of the Wallace line.

Note that PNG denotes all of island New Guinea, the Bismarck's, New Ireland, New Britain and surrounding smaller islands. Samples from these regions would be grouped together geographically due to their connected human pasts.

*New samples from this publication.

Sanger sequencing results were aligned in *Geneious v8.0.4* to a consensus sequence generated from all *Pv* whole mitochondrial genome sequences available on *Genbank* (see *Appendix S3.8*). Though the *Sal1* reference strain is usually used for alignment, the generated reference sequence gives a better indication of the direction of haplotype diversity from the start. Checking for erroneous calls using the reference sequence as a guide meant all base pairs were compared to those most commonly found globally, instead of against a single reference strain. The sequenced genomes were manually checked and conservatively edited, in accordance with the original sequencing chromatograms. A few ambiguous calls were noted (1-2 in <10% of sequences), where two nucleotides were equally or close to equally present in repeated sequencing efforts. These were then called towards the consensus of similar sequences, or towards the consensus sequence if similar sequence options were not available. As a result true diversity was possibly underestimated. Generated data (n= 149) were further increased and aligned with samples from around Melanesia, East- and Southeast Asia (n=358, all regions in *Table 1*) from previously published papers, as made accessible on *Genbank*. See *Table 1* for references, and *Figure 1* (map generated in *QGIS*) for sample site locations.

After alignment of all 507 sequences, around 90bp were removed from the end (bp 5847-5991, *see* S3.8). The removed sequence was missing in several previously published sequences, and did not show any variation in any of the other sequences used. In addition, all gaps due to varying string lengths of single nucleotides were removed. This decision was made due to the high likelihood of both sequencing and DNA replication errors in such areas, making them unsuitable for phylogenetic inference. In addition, incorporating insertions and deletions (indels) into substitution models is problematic with the current version of the *BEAST2* (Bouckaert *et al.* 2014) software employed.

Table 1

Sample site locations and sample sizes of *P. vivax* across East & Southeast Asia and Melanesia. Italic & starred samples are new data from this thesis; others were sourced on Genbank from referenced publications.

Region	Country	n	Geogroup	Published	Genbank Accession Numbers
Melanesia	Subtotal	222			
	Papua New Guinea	60	PNG	Mu <i>et al.</i> 2005, Jungwutiwes <i>et al.</i> 2005	AY791631-89; AY598119
	<i>Papua New Guinea*</i>	129	<i>PNG</i>	<i>This paper</i>	<i>N/A</i>
	Solomon Islands	4	MIS	Mu <i>et al.</i> 2005, Jungwutiwes <i>et al.</i> 2005, Miao <i>et al.</i> 2012	AY791564; AY598116-8; XA_SOI1
	<i>Solomon Islands*</i>	17	<i>MIS</i>	<i>This paper</i>	<i>N/A</i>
	Vanuatu	7	MIS	Mu <i>et al.</i> 2005	AY791561, AY791565-6, AY791568-70, AY791574
	<i>Vanuatu*</i>	3	<i>MIS</i>	<i>This paper</i>	<i>N/A</i>
	Melanesia Unk.	2	PNG	Mu <i>et al.</i> 2005	AY791690, AY791692
Island SE Asia	Subtotal	51			
	Indonesia (non-Papua)	45	ISEA	Mu <i>et al.</i> 2005, Jungwutiwes <i>et al.</i> 2005, Taylor <i>et al.</i> 2013	AY791562-3, AY791567, AY791572, AY791576-80, AY791582; AY791552; AY598098-108; KC330533-57
	Borneo	4	ISEA	Mu <i>et al.</i> 2005, Cox-Singh <i>et al.</i> 2006	AY791555; DQ396547-49
	Philippines	1	ISEA	Mu <i>et al.</i> 2005	AY791571
	Timor	1	ISEA	Mu <i>et al.</i> 2005	AY791573
Mainland SE Asia	Subtotal	122			
	Thailand	44	MSEA	Mu <i>et al.</i> 2005, Jungwutiwes <i>et al.</i> 2005	AY791556, AY791558-60, AY791583; AY598035-73
	Cambodia	19	MSEA	Taylor <i>et al.</i> 2013	KC330514-32
	Cambodia	43	MSEA	Pearson <i>et al.</i> 2016	Personal
	Vietnam	9	MSEA	Mu <i>et al.</i> 2005, Jungwutiwes <i>et al.</i> 2005	AY791581, AY791584; AY598121-7
	Vietnam	7	MSEA	Pearson <i>et al.</i> 2016	Personal
South China	Subtotal	59			
	China (South)	12	CHSTH	Jungwutiwes <i>et al.</i> 2005	AY598128-39
	China (Guizhou)	25	CHIN	Miao <i>et al.</i> 2012	JQ240370-95
	China (Guangxi)	18	CHSTH	Mu <i>et al.</i> 2005	AY791588-605
	China (Yunnan)	3	CHSTH	Mu <i>et al.</i> 2005	AY791585-7
	China (Hainan)	1	CHSTH	Mu <i>et al.</i> 2005	AY791606
NE Asia	Subtotal	70			
	Korea	34	NEA	Mu <i>et al.</i> 2005, Iwagami <i>et al.</i> 2010, Miao <i>et al.</i> 2012, Taylor <i>et al.</i> 2013	AY791554; AB550270-80; JQ240367-9, JQ240397-8; KC330497-513
	China (Anhui)	36	CHIN	Miao <i>et al.</i> 2012	JQ240331-66
Myanmar	Subtotal	32			
	Myanmar	32	MMR	Miao <i>et al.</i> 2012	JQ240396-429
Madagascar	Subtotal	66			
		66	MAD	Culleton <i>et al.</i> 2011, Taylor <i>et al.</i> 2013	JN788738-62; KC330465-80; KC30440-64
TOTAL		507			

To test for human heritage related differences, mitochondrial sequences from known sites across Melanesia were separated into Papuan (Pap: 93), Austronesian (Aus: 42) or dual (Both: 22) heritage, based on the linguistic heritage of the sample sites, as seen in S2. To determine the degree of differentiation and contact, K_{ST}^* and N_M values were calculated with *dnasp5* (Rozas & Rozas 1995, Rozas *et al.* 2003).

The expanded dataset of all obtained *Pv* mitochondrial sequences from sample sites across the Pacific, East and Southeast Asia were grouped according to geographical regions, as shown in *Table 1*. These groups were used for analysing population dynamics and phylogenetic relationships. Diversity values (intraregional haplotype diversity: H_d , intraregional neutrality test: *Tajima's D*, interregional population differentiation: K_{ST}^*) were calculated for each region in *dnasp5* (Rozas & Rozas 1995, Rozas *et al.* 2003); the results are shown in *Appendices S3.5.1 & S3.5.2*.

To better understand and visualise the geographic distribution between the different mitochondrial genomes, R (*ade4*) was used to generate a principle component analyses (PCA), as shown in *Figure 3b*. For the PCA extra data from Madagascar was included. This was done to include a central focal point representing the African-South Asian haplotypes, to which the Malagasy *Pv* belongs (Taylor *et al.* 2013, Culleton *et al.* 2011, also see *Chapter 2*), and to affirm that Malagasy *Pv* is indeed not part of an Austronesian introduction from East Asia. *Network 4.6.1.3* (Fluxus-Engineering 2015) was used to create a median joining haplotype network using the neighbour joining method (Bandelt *et al.* 1999) from all 506 *Pv* mitochondrial genome sequences across East Asian & Pacific region, without star contractions.

Phylogenetic relationships were estimated using a Bayesian Markov Chain Monte Carlo framework in *BEASTv2.1.3* and *Tracer v1.6* (Rambaut *et al.* 2014; see *S3*). For the main model settings, a Coalescent Bayesian Skyline study model, HKY (Hasegawa *et al.* 1985) substitution model and strict clock branch rate model were used, with no variance across sites (gamma) and high estimated rate of base pair invariance (I). See *Table S3* for further details. The Skyline plot estimating population size over time was produced from the Bayesian run, using *BEASTv2.1.3* (see *Appendix S3.4*).

Clock rate was first test-run with a relaxed clock log-normal setting, starting at 0.004 substitutions per million years, the average rate estimate from previous publications (see Comas *et al.* 2013: 0.00189-0.00600 & 0.00211-0.00601, Cornejo & Escalante 2006: 0.0032–0.0043, Mu *et al.* 2005: 0.0045–0.0056, Jongwutiwes *et al.* 2005: 0.0026–0.0036 substitutions per site per million years). Both Comas *et al.* (2013) and Mu *et al.* (2005) calculated their rates by calibrating a divergence time of 6-14.2 Ma (Pacheco *et al.* 2012) between a clade of *P. gonderi* and *Plasmodium sp.* from mandrills, with *Pv* and its nine closest Asian *Plasmodium sp.* relatives. In addition, Comas *et al.* calculated rates using one mitochondrial genome from each of the eleven species in a birth-death model in BEAST. Jongwutiwes *et al.* (2005) calibrated rates to *Pf* and *P. reichenowi* host divergence dates, human and chimpanzee respectively, with *P. knowlesi*

and *Pv* as outgroups, and using strong evidence for constant clock rates. These are currently still the standards for calculating the *Plasmodium* mitochondrial genome clock rate, and so the decision was made to use a similar rate in this paper as the starting estimate for the molecular clock, without the need for repeating the same calibration methods. All test runs with a relaxed clock lognormal model ended up with a median speed of ~0.0044 sub/site/Ma, with a 95% interval ranging from around 0.0041-0.0047sub/site/Ma. The rate of 0.0044 sub/site/Ma was consequently used as the strict clock rate in *BEAST2*.

Another model was run to estimate a broad indication of mutation rates necessary for Melanesian *Pv* given a Papuan (clade age range estimate set at: 44.8 kya, 95% HPD: 29.7–59.2 kya) vs. Austronesian (clade age range estimate set at: 6.0 kya, 95% HPD: 3.3–9.0 kya) origin for the Melanesian phylogenetic branches as determined in earlier runs. This was done to double check the interval for sensible mutation rates that might overlap with either of these migrations. See *Appendix S3.4* for run settings.

Two other analyses were conducted to study the variation and relationships between the genetic samples. Both a general principle component analyses (PCA – using *R* packages: *genind*, *ade4*, *ade4genet* (Jombart et al. 2010)), and the discriminant analyses of principle components (DAPC, – using *R* packages: *genind*, *ade4genet* (Jombart et al. 2010)).

3. 3. RESULTS

Along the ~5900bp length of the *Pv* mtDNA, 182 mutations over 173 sites were identified, most of these (117) being singletons. 116 mutations were found in coding regions, of which 68 were non-synonymous. This compares similarly to the variation seen on the *Pf* mt genome (Preston *et al.* 2014).

For the Melanesian region, there is no clear differentiation between samples from Austronesian and Papuan sites, divided according to human linguistic and genetic data, as found in *Table S2*. Low K_{ST}^* values (*Aus-Pap*: 0.023, *Aus-Both*: 0.014, *Pap-Both*: 0.001). Additionally, high migration estimates, N_M (*Aus-Pap*: 22.88, *Aus-Both*: 11.03, *Pap-Both*: 77.18) were measured between them. In the phylogenetic tree and PCA for the Eastern Asian & Pacific, all the Melanesian haplotypes are closely related, as shown in *Figure 3*, indicating a single origin. Time related approximations using *BEAST* for the timing of an Austronesian versus a Papuan introduction resulted in median mutation rate estimates of 4.3972E-8 (95% HPD: 2.3278E-8 – 7.6053E-8) and 5.993E-9 (95% HPD: 3.7696E-9 – 9.1065E-9) respectively. This indicates a Papuan introduction fits within current mutation rate estimates. Further implications of these results are discussed later in this section.

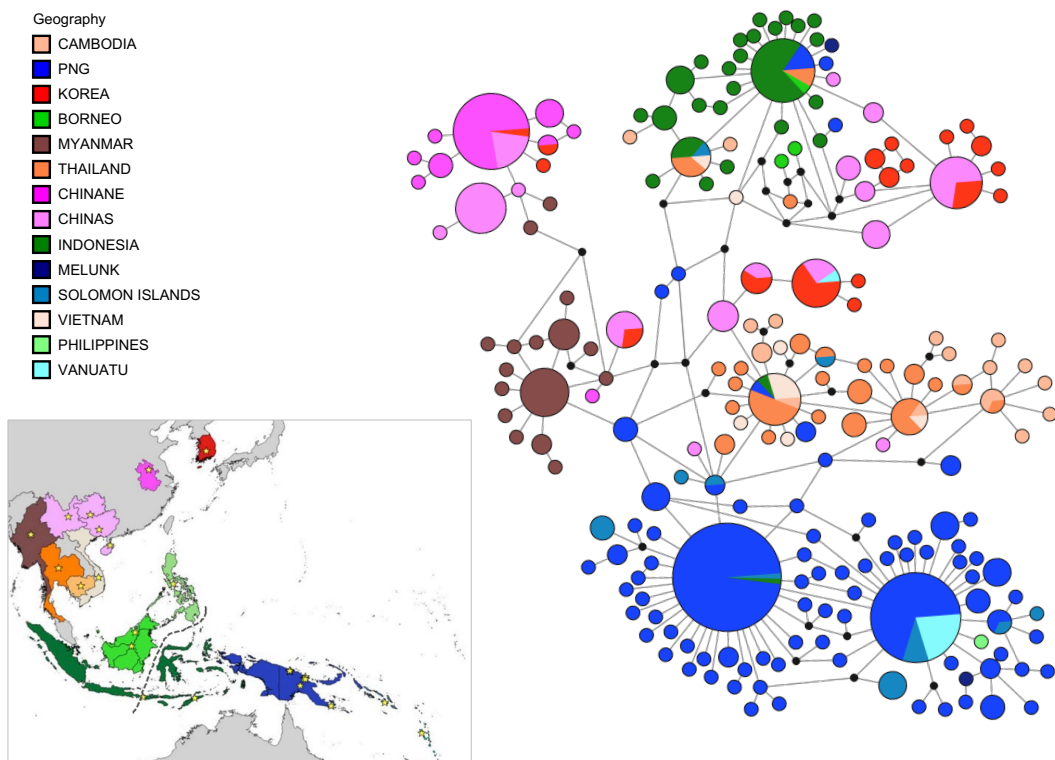


Figure 2

Median joining haplotype network of East Asian, Southeast Asian and Melanesian *P. vivax* mitochondrial genome sequences. Each circle represents a different haplotype, each colour the region from which a sample originates. Sizes of the nodes indicate the number of samples per haplotype. Each connection represents a mutational step. Produced in *Network 4.6.1.3*. Network of 506 samples, all indels removed from sequences. No star contractions.

Colouring according to legend, by region: PNG: dark blue; SOI: grey blue; VAN: light blue; Melanesia Unknown (Melunk): blue black; Thailand: bright orange; Cambodia: light orange; Vietnam: pale cream; Indonesia: dark green; Borneo: bright green; Philippines: light green; Southern China: Light Pink; NE China: Dark Pink; Korea: Red; Myanmar: Brown.

Across the six broader geographic regions (as seen in Table 1) the haplotype network in Figure 2, and the haplotype diversity estimates (H_d , MEL: 0.87; ISEA: 0.91; MSEA: 0.96; MMR: 0.89; SCh: 0.90; NEA: 0.87) show that all countries and regions have high haplotype diversity. These results are similar to those previously observed by Taylor *et al.* (2013) (see Appendix S3.5.1). Almost all geographic regions show a radial formation of numerous single haplotypes around a main haplotype. This constellation-like structure is a sign of increased diversity due to population growth. Population growth is also supported by the negative Tajima's D values indicating an excess of rare single nucleotide polymorphisms (SNPs) (Appendix S3.5.1), and by the Bayesian Skyline Plot (Appendix S3.4) that shows population growth across the entire region over time.

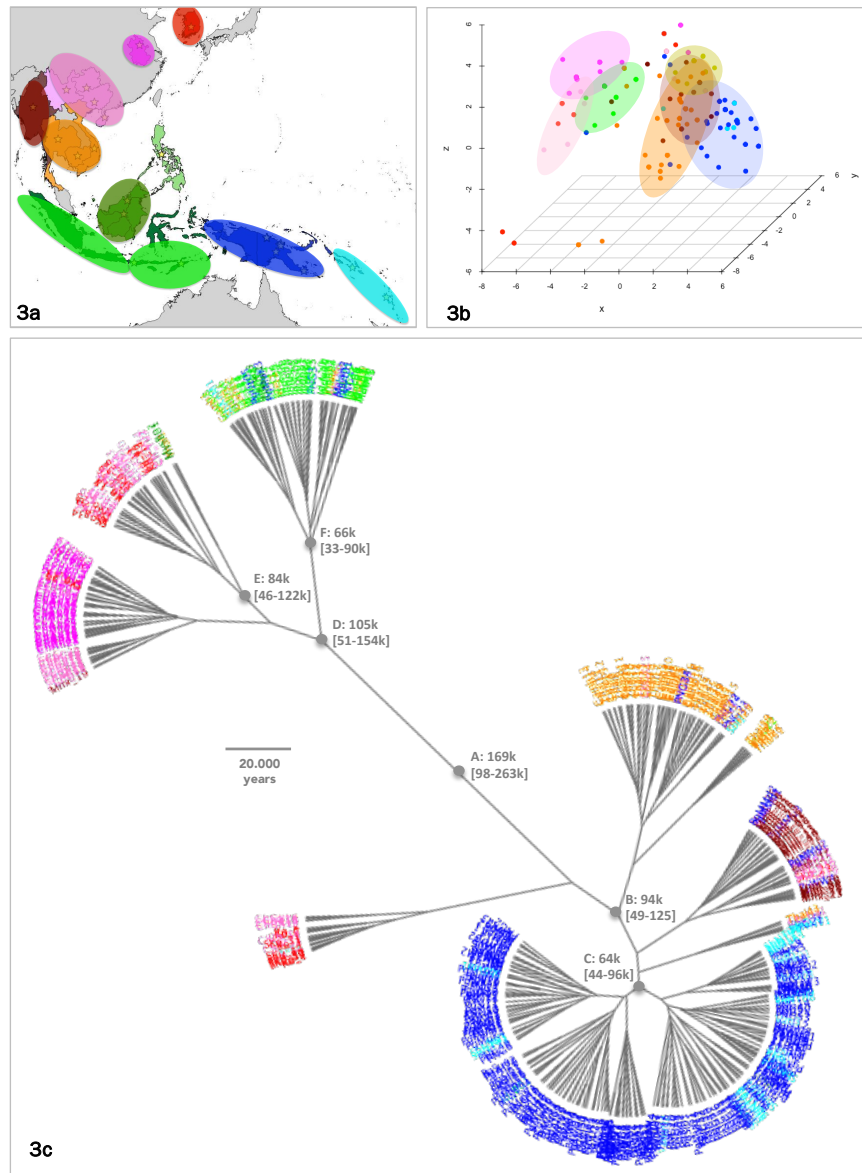


Figure 3

(3a) Regional division and colour coding of samples by geographic origin.

(3b) Principle components analyses of the whole mitochondrial genome of the *Plasmodium vivax* samples across three axes (x, y, z are PCA1, PCA2, PCA3 respectively). Coloured according to geographic origin as seen in 3a, with the addition of samples originating from Madagascar (mustard yellow), and all Island Southeast Asian samples are coloured light green. Clear geographical clusters have been circled in approximation of their limits within the graph.

(3c) Bayesian MCC tree estimated from 506 mitochondrial sequences, unrooted, no outgroup, times from GTR + I sites model with strict clock at 0.0044 substitutions per million years (S3.3 1a). 2x 750 000 000 runs, sampled every 10 000, maximum sum of clade credibility consensus tree with common ancestor node heights. Simplified model used for tree layout was S3.3 2b. For a linearised tree with posterior probability values, see Appendix S3.6.1.

Colour according to region: Melanesia: Blue (Van and SOI light blue); Mainland SE Asia: Orange; Island SE Asia: Green (Indonesia light green, Borneo dark green); South China: Light Pink; NE China: Dark Pink; Korea: Red; Myanmar: Brown; Madagascar (PCA only): Mustard Yellow.

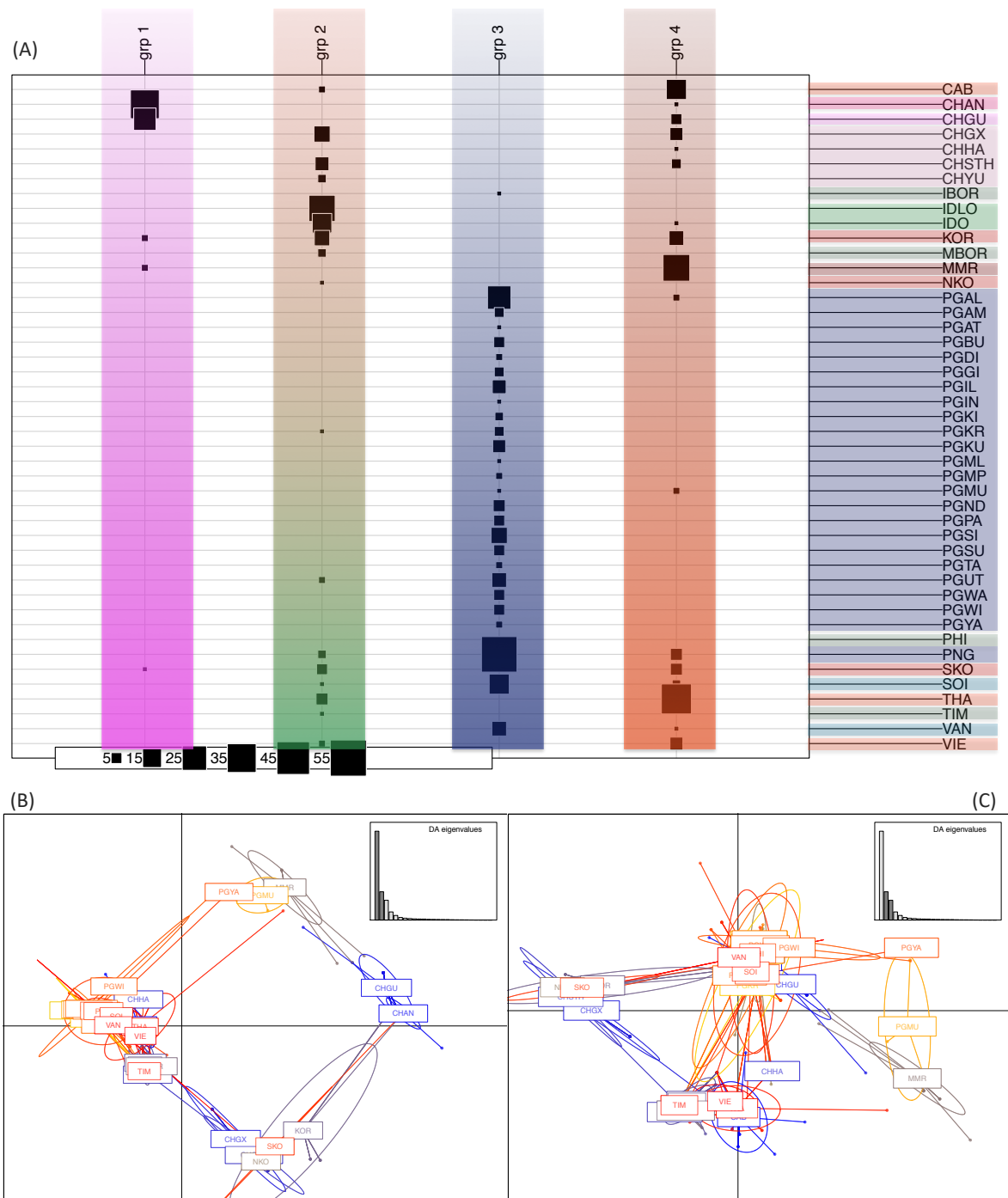


Figure 4

DAPC clusters for the samples across the East Asian & Pacific region.

(4a): The four main clusters recognised by the DAPC analyses. Groups have been coloured according to the main geographical regions of origin for the samples represented. Group 1 contains mainly Northern Chinese samples; Group 2 contains mainly samples from Island SE Asia, Southern China and Korea; Group 3 contains almost all Melanesian samples; Group 4 contains the samples from Myanmar, Mainland SE Asia, and the remaining samples from across East Asia (China & Korea).

(4b), (4c): Principle components analyses of the whole mitochondrial genome of the *Plasmodium vivax* samples for the three main eigenvalues. 4b: axes 1 & 2. 4c: axes 1 & 3. Samples have been coloured and clustered according to the region they were originally sampled in.

Colour according to region: Melanesia: Blue (Van and SOI light blue); Mainland SE Asia: Orange; Island SE Asia: Green (Indonesia light green, Borneo dark green); South China: Light Pink; NE China: Dark Pink; Korea: Red; Myanmar: Brown.

Strong regional patterning, found in the haplotype network, PCA, Bayesian phylogenetic tree and DAPC (Figure 2, Figure 3b & 3c, Figure 4 respectively), indicates evolution in isolation for many pockets of *Pv* across the Pacific, East and Southeast Asia. The length of this isolation depends on the

time taken to evolve, and with a simplification of current estimates of mt genome evolution rates (~ 0.0044 sub/mya/site) the time of isolation can range from 0 – 44,000 years for each mutation on the genome. As most regions have a genetic signature that varies one or two mutations from their closest related region, the time of evolution between nearest neighbours lies an estimated 44 - 88 ky apart. It follows that the distinct Melanesian lineages, and their diversity, indicate a long and isolated past of *Pv* in Melanesia. This is confirmed by the phylogenetic data, where the estimated time to the most recent common ancestor of the Melanesian lineage falls around 65 kya (see *Figure 3*; 32-121 kya 95% HPD), which corresponds with the earliest time that anatomically modern humans, *Homo sapiens*, are thought to have reached New Guinea. In addition, *Pv* diversification intensifies from just under 30 kya, indicating a growing population. Atkinson *et al.* (2008) observed human population growth in New Guinea around this time, based on similar methods using the human genome. Atkinson links agricultural development to the increase in human population size, and the increase in human host numbers could have instigated significant growth in the *Pv* population.

The interpopulation K_{ST}^* values reported in *S3.5.2*, the PCA in *Figure 3b* and the phylogenetic tree in *Figure 3c*, show that Melanesian *Pv* is most similar to those of Mainland Southeast Asia (MSE Asia) and Myanmar, much more so than to its geographic neighbour, Island Southeast Asia (ISE Asia). Both the PCA and the phylogenetic tree show a diversification of two broad groups in different directions from a central axis. The two groups fall broadly into: East Asia/Island Southeast Asia, and Myanmar/Mainland Southeast Asia/Melanesia. The PCA central axis is made up of South Asian – African haplotypes, denoted by Madagascar here (for relationship between Madagascar, South Asia and Africa see *Chapter 2*, and papers by Jongwutiwes *et al.* 2005, Miao *et al.* 2012, Taylor *et al.* 2013, Rodrigues *et al.* 2014). From the central axis there is a clear movement into Myanmar, and consecutive movement into both mainland Southeast Asia and Melanesia, along separate axes. In the second group, direction of migration is less clear with the northeast China (Anhui) cluster seeming closest to the central Africa-South Asia axis, but the South China and Korea distributions are not very strong and spread across the group. Also, in the PCA in *Figure 3b*, island Southeast Asia does not appear to clearly branch off another region, but appears to be closest to northeast China, to which it runs parallel. However, according to the DAPC in *Figure 4*, island Southeast Asia clusters with several samples from southern China and mainland Southeast Asia, showing a regional connection with all neighbours except for Melanesia.

The *Pv* mt genomes from the three most southern provinces in South China (Guangxi, Yunnan & Hainan), with Korea, fall into at least three sets of varying regional haplotype groups that fit into the broader network in different locations. This can be clearly seen in the DAPC results in *Figure 4*, where a significant number of samples from these regions are contained in groups 2 and 4. The pattern indicates that this region has seen greater population movement involving *Pv* than most of the other, more clearly defined, East Asian regions. As many of the haplotypes have their own distinct diversification patterns and are unique to the region, the population movements involved in their

distribution could be of an older age. There is also evidence for more recent movement of *Pv* strains to Southern China by modern human migration importing single haplotypes from MSE Asia, ISE Asia and Myanmar. The Chinese provinces of Guizhou (most Northern province in light pink) and Anhui (dark pink), found further inland, each have their own, neighbouring, distinct haplotypes, as seen in the top-left of *Figure 3b & c*. These patterns, and the ones discussed below, can be seen more clearly in *Figure 5*, which shows the haplotype group that each sample belongs to. The map and pie charts once again show the strong geographical stratification of the *Pv* mitochondrial genome, with only Southern China and northeast Asia (Korea), displaying a mixed set of regional haplotypes.

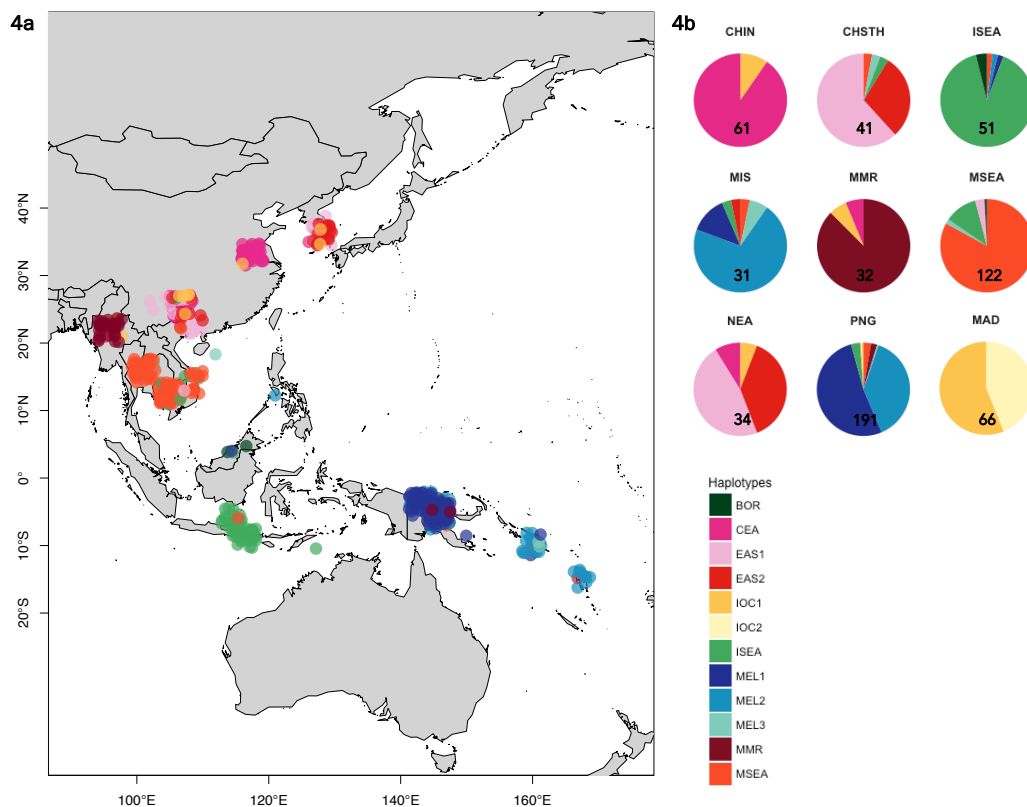


Figure 5

Sample locations and haplogroup association (4a). Representation of *P. vivax* mitochondrial haplotypes as found on the tree in figure 3, coloured according to the legend and grouped per geographical region in the pie charts. Sample numbers per region are given in the pie charts (4b). Geographical regions are as given in Table 1:

Colouring according to legend, by haplogroup association: BOR: Borneo; CEA: Central East Asia; EAS1, 2: East Asia 1, 2; IOC1, 2: Indian Ocean 1, 2; ISEA: Island Southeast Asia; MEL1, 2, 3: Melanesia 1, 2, 3; MMR: Myanmar; MSEA: Mainland Southeast Asia.

Only two Melanesian-like haplotypes have been sampled in other geographic regions, namely one from Indonesia (non-New Guinea) and one from the Philippines. These are likely to have been exports from PNG, or Indonesian provinces on New Guinea; the numbers of exported Melanesian strains are low and movement of people in and out of PNG has only become common in the last 200 years. Timber labourers from Indonesia and Filipino professionals are commonly found in PNG, indicating the presence and possible infection of these nationalities. In addition, a handful of SE Asian, both Mainland and Island, haplotypes have been introduced to the Melanesian region. These are likely recent introductions from modern-day labourers and migrants, as the haplotypes are

restricted in number and show little diversity. Most of these isolates have haplotypes similar to those from Indonesia, Thailand and Myanmar. Again, timber labourers were brought from Indonesia to PNG in the last two decades and are known to have been infected heavily with malaria whilst there (Baird *et al.* 2003). Going further back, German colonists brought Malay plantation labourers to PNG in the early 19th century. As observed by Koch during his expeditions to this area in the late 1800s, many of these labourers as well as Chinese immigrants were susceptible to malaria whilst working on the plantations (Koch 1900 referenced in Stanisic *et al.* 2010). These migrants could have been sources of parasite migration to and from PNG.

Within the Melanesian region it is notable that there is no variation found in *Pv* strains in Vanuatu, besides one sample that appears to originate from China. Sampling numbers are low, but large enough to observe variation; if diversity were as high as that observed across all other geographic locations. This could indicate that Vanuatu was more recently infected from a single location, or limited number of locations, in PNG or SOI. Sampling was extensive enough to pick up an introduced strain from China. However, its low variation could be influenced by its placement at the end of the range of malaria transmission, with a relatively smaller number of infected migration events occurring from a limited number of neighbouring sources, and an already low transmission causing loss of diversity. The Solomon Islands, on the other hand, appear to have a set of unique local *Pv* variants, indicating that these islands might have been infected early enough to develop a regional signature, or at least have been exposed to several introductions.

A final observation from the PCA and phylogenetic tree in *Figure 3* is that ISE Asia and MSE Asia each cluster strongly on their own, and at opposite ends to each other. Mainland SE Asia is more closely related to Melanesia, whilst Melanesia's neighbour Island SE Asia is more closely related to East Asia. This pattern strongly suggests that there were two routes of *Pv* introduction into the western Pacific. There is one small branch off the East Asian/Korean clade that contains two distinct Malaysian Borneo samples, a region that may be of interest for exploring the routes that connect ISE Asia and East Asia.

The strong geographical stratification of the mtDNA sequences aids in illuminating past connections between populations and estimating the broad time frame of isolation. To compare the age of *Pv* in Melanesia with that of modern humans and their migration patterns and activities, a time-line was constructed (*Figure 6*), which includes genetic data from lineage ages of humans, early diseases (*Helicobacter pylorii*, *Mycobacterium tuberculosis*) and a *Pv* age estimate for the region from Taylor *et al.* (2013). Data from archaeological sites identifying human presence around Melanesia are also included. The graph shows that there is a clear correspondence in the estimated time of *Pv* emergence in Melanesia with the presence of Papuans in Melanesia. The estimated median mutation rate of 5.993E-9 (95% HPD: 3.7696E-9 – 9.1065E-9) for a Papuan time of introduction, as shown in the blue normal distribution in *Figure 6*, is much closer to the one calculated through lineage tracing than the

one calculated for an Austronesian time of introduction, which is a factor ten off at 4.3972E-8 (95% HPD: 2.3278E-8 – 7.6053E-8, red normal distribution). Currently, all evidence points towards *Pv* malaria having spread to Melanesia, and across Asia, with the earliest known modern human migrants.

3. 4. DISCUSSION

Studies involving whole genome sequencing have the potential to provide a comparison of movement and evolution across all time periods, if mitochondrial, apicoplast and nuclear genomes are used together to highlight the underlying past behind the diversity of *Pv*. Such work has as yet to be done, with the result that the dataset for such an analysis is currently much smaller than the wealth of information available for each genome independently from much larger datasets.

Previous papers have shown high genetic diversity in *Pv* microsatellite markers (MS), with little or no structure at the local scale in Papua New Guinea (Koepfli *et al.* 2013, Jennison *et al.* 2015). However, on a larger geographic scale, MS show strong global structuring including regional fingerprints in Melanesia (Koepfli *et al.* 2015, Fola *et al.* 2018, Waltmann *et al.* 2018). Recent genome-wide SNP data also show strong regional structuring on a global scale (Pearson *et al.* 2016, Hupalo *et al.* 2016).

The nuclear genome itself is unsuitable for studying phylogenetics over long periods of time as it has unusually high recombination rates (Hartl *et al.* 2002, Hartl 2004). Both Hartl (2004) and Koepfli *et al.* (2015) argue that the use of mitochondrial genome might give enhanced resolution for regional signatures, where that of the nucleotide genome is too variable or fragmented. Strong regional signatures for *Pv* mitochondrial DNA have previously been indicated by Rodrigues *et al.* (2014). The decision to work with mtDNA was based on the strength of using mtDNA to inform long-term *Pv* phylogeny, as well as the availability of an extensive global database of mtDNA sequences which would add to the breadth of understanding global movement of *Pv*. A further comparison and discussion of the different uses of different parts and sizes of the genome is given in *Chapter 4*.

The data presented here confirm that mtDNA displays strong regional fingerprints. For regional patterns to show strongly in mtDNA, isolation of populations is necessary for long periods of time. Sustaining these patterns for thousands of years requires no or limited introductions of new strains of *Pv*, as is generally observed in the data. There are some strains present in Papua New Guinea that clearly have a SE Asian origin, yet most of these match known modern (<200 years) migration patterns.

Broadly, the phylogenies created by the BEAST tree correspond in some degree with previous phylogenies and age estimates for *Pv* (Culleton *et al.* 2011, Miao *et al.* 2012, Taylor *et al.* 2013). The comparison with human presence, disease and archaeological dates shows that the timing of *Pv* entry into the region corresponds with the earliest known humans entering the area (*see Figure 6*). This means there is strong support for *Pv* migrating along with the earliest modern human migrants, and thus being a Papuan introduction.

A Papuan introduction is also confirmed by the observation that Melanesian *Pv* is more closely related to *Pv* from Mainland Southeast Asia and Myanmar, corresponding to the southerly Asian route of first human migration into Melanesia. In order to be introduced by the Austronesians, the closest relatives would come from Eastern Asia, as this is the generally recognised place of origin for the Austronesians (Lipson *et al.* 2014, Soares *et al.* 2016). Moreover, the mutation rate would have to be more than ten times faster ($\sim 4.4\text{E-}9$ versus $\sim 4.4\text{E-}8$) to correspond with an Austronesian introduction. Such a rate does not fall within any current estimation for the *Plasmodium* mitochondrial genome mutation rate over the long term (Ricklefs & Outlaw 2012, Bensch *et al.* 2013), and is especially high when including the assumption of front-loading, where mutation rate estimates in the present are much higher than when they are extrapolated over time. Dynamics such as strong purifying selection on the mtDNA (Ho & Shapiro 2011) decrease variation over the long term. Consequently, mutation rates estimated over the long term should be lower than those observed to occur over a short period of time.

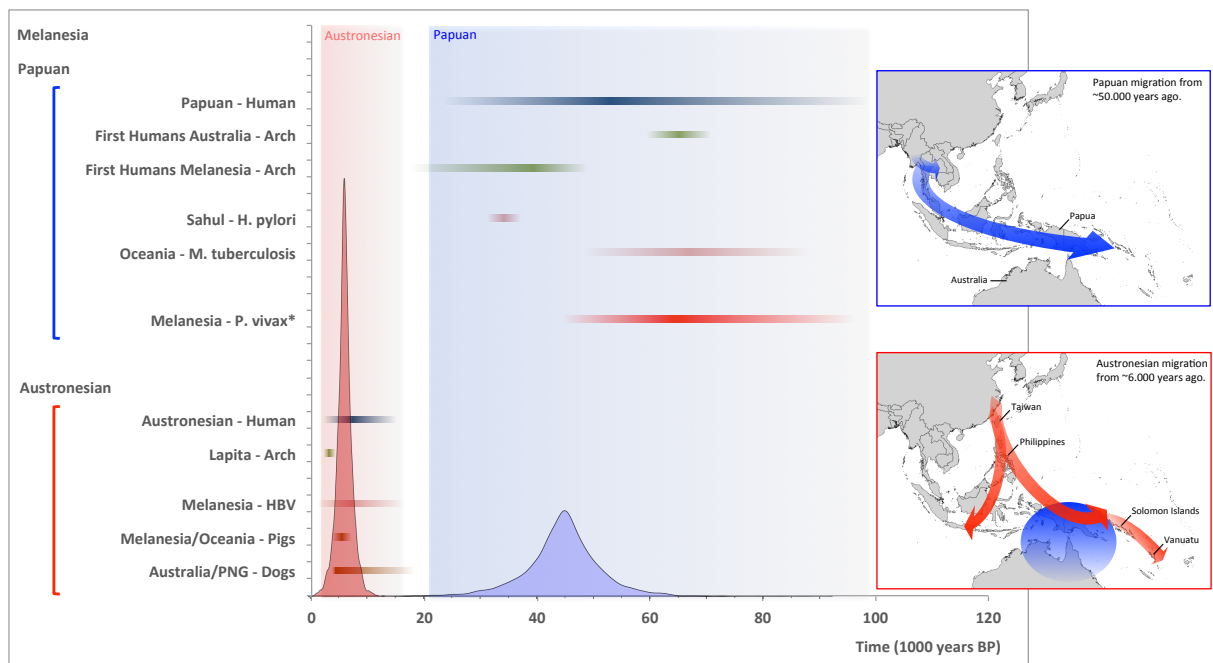


Figure 6

Archaeological evidence of human activity in Melanesia (green), 95% intervals for Melanesian lineage age estimates of the most recent common ancestor (MRCA) of Papuan and Austronesian motifs in human genomes (dark blue), other disease genomes (pink), domesticates (orange) and *P. vivax* mtDNA (red). Two maps show the general routes of migration of the Papuans (top) and Austronesians (bottom), based on Skoglund *et al.* 2016.

Two normal distribution graphs show the Melanesian *P. vivax* branch age estimate settings given a Papuan (blue: 45 kya, 95% HPD: 30–60 kya) and an Austronesian (red: 6035 ya, 95% HPD: 3350–9050 ya) emergence. The Papuan age estimate results in a median mutation rate of $5.993\text{E-}9$ (median, 95% HPD: $3.7696\text{E-}9$ – $9.1065\text{E-}9$), whereas the Austronesian age estimate leads to a median mutation rate of $4.3972\text{E-}8$ (95% HPD: $2.3278\text{E-}8$ – $7.6053\text{E-}8$).

The period of early Papuan settlement (~ 80 - 20 kya) is shaded blue, the estimated period of Austronesian settlement (15 - 2 kya) is shaded red. The reference data used to calculate the lineage estimates can be found in the table in *Appendix S3.7*.

The human Austronesian populations across ISEA, Melanesia and even in Madagascar display genetic adaptations to malaria in the form of Southeast Asian Ovalocytosis (SAO, Kimura *et al.* 2003, Wilder *et al.* 2009, Paquette *et al.* 2015), which provides strong protection against *Pv* malaria (Genton *et al.* 1995, Rosanas-Urgell *et al.* 2012). SAO is present in both malarious and non-malarious regions of Vanuatu,

indicating that the original Austronesian settlers were likely exposed to malaria for a substantial amount of time before migrating (Tsukuhara *et al.* 2006). This is in line with suggestions that the global distribution of SAO demonstrates the trait was acquired by Austronesian populations prior to their dispersal to the Pacific (Kimura *et al.* 2003) and Madagascar (Rabe *et al.* 2002). However, there is currently no evidence of these Austronesian migrants bringing particular Austronesian-associated strains of *Pv* with them to Melanesia, whether this is due to none being introduced or introductions having been unsuccessful in the long term. The mtDNA of Malagasy *Pv* indicates a common Indian Ocean origin (see *Global Chapter*, Culleton *et al.* 2012, Miao *et al.* 2012, Taylor *et al.* 2012, Rodrigues *et al.* 2017). Based on the mtDNA sequences of current Vanuatu and SOI isolates it is much more likely that Austronesian migrants in Melanesia picked up *Pv* strains from Papua New Guinea and deposited them on the SOI and Vanuatu. The chronology of the introductions of these strains is hard to estimate due to their low levels of diversification.

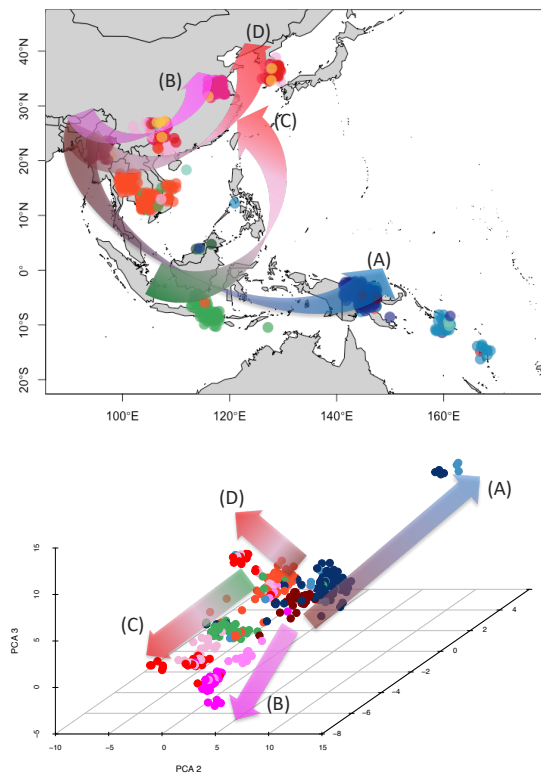


Figure 7

Reconstruction of possible geographical movements of *P. vivax* populations in the past, according to the PCA associations and clines seen in (7b).

(A) denotes the movement from Myanmar (brown) to Papua New Guinea and the rest of Melanesia (blue)

(B) shows the movement from Myanmar (on the central axis), through central China (medium pink) to northern China (dark pink).

(C) shows a movement from Myanmar through island SE Asia (green), which appears to be either a branch off, or feeding into a coastal trajectory from southern China (light pink) to northeastern Asia (red) – a similar trajectory is seen in

(D) which goes through mainland SE Asia instead, but then appears to hug the coast, moving up from southern China to northeastern Asia.

Although major evidence for the presence of an Austronesian *Pv* lineage in Melanesia is lacking, the lineages from neighbouring ISE Asia raise some intriguing questions. Geographically this is the closest region to Melanesia, yet phylogenetically it is one of the furthest removed, in the present dataset. Its close relation to East Asian *Pv* indicates an origin from this region, which is in line with human data. However, current estimation of the ISEA clade age shows it to be as old or older than the Melanesian clade, which does not correspond with a much later human Austronesian introduction. Whether its entry into the current geographic location corresponds relatively closely with the estimated age, or it was introduced later, is uncertain. If the former, *Pv* in West and Central Indonesia would not have been carried in by the Austronesians. Visualisation of the possibly

geographical movement as denoted by the PCA analyses, as seen in *Figure 7*, adds to the interest in the complexity of patterns seen around this region.

There are some possible alternative drivers behind the diversity in ISEA, which pushes back time estimates artificially. Firstly, there is a possibility that the ISE Asian *Pv* clade diversified in isolation around Taiwan, and Austronesian migrants introduced a wide variety of different Taiwanese and/or East Asian strains before 4 ky ago. Some theories push the first Austronesian expansions to 10 kya (e.g. Soares *et al.* 2016), which would support the argument for an Austronesian introduction of ISEA *Pv*. In addition, diversification across the region could have accelerated with increasing human numbers and a patchy geographic setting. This remains an unlikely scenario. Besides high haplotype diversity starting around 30 kya, a strong regional signature and early distinction from the East Asian clades at over 51 kya, there are traces of a Bornean stepping-stone, which accentuates a linear timeline. All these make it more likely that *Pv* in West and Central Indonesia diversified within the region before the movement of the Austronesian people. The Palaeolithic link between Western Indonesia and mainland Asia dated at 30-14 kya based on Y-chromosomal genetics (Karafet *et al.* 2010) make it possible chronologically that the ISE Asian *Pv* clade may have been introduced from mainland Asia by pre-Austronesian but post-Papuan migrants. The clines seen in the results from the PCA in *Figure 7* indicate a possible movement of *P. vivax* through ISEA first, before moving back northwards, onto the mainland, and from southern China to north-eastern Asia along the coast – instead of ISEA branching off of the southern China samples. In that case an unidentified route of early humans coming southwards to ISEA and moving back up would have had to occur, possibly through an island route using the Philippines and Borneo, instead of the other way around.

Alternatively, vector species compatibility and adaptation could have driven the diversity patterns seen across the region. The *Anopheles farauti* complex found in Melanesia is unique, and much older than human populations (Logue *et al.* 2013, 2015). Its compatibility with the local Melanesian strains of *Pv* could exclude the intrusion of other strains, including the ones found in ISEA. Adaptation to and selection by local vectors (Joy *et al.* 2012) could drive some of the population structure and diversity seen in the different strains of *Pv*. In contrast, as discussed in Chapter 3, ISEA has a vector species composition that relates to that of Eastern Asia. This could have facilitated the introduction of later strains from the East Asian region into ISEA.

Large parts of Indonesia would have been initially settled by Papuan migrants moving through ISEA to Melanesia, around ~60 kya. The geographical assumption of this movement is underlined by archaeological evidence of modern human remains on Sumatra dating back to 63-73 kya (Westaway *et al.* 2017). In addition, remnants of early Papuan settlers in ISEA are found in the presence of Papuan languages in parts of Central Indonesia, such as Timor (Mona *et al.* 2009). If the Papuans brought parasites to Melanesia, such parasites should also have been introduced to ISEA. However, these parasites are no longer found, at least not in significant numbers. The single sample from East Timor,

which is one of the few geographic regions outside Melanesia with Papuan human populations, cannot provide significant insight into the situation. It does fall within the ISEA clade with other Indonesian samples. The reason for the strong separation of ISEA and Melanesian samples is unknown, but one factor that is significantly different between these regions is the species of *Anopheles* mosquito vector (see Chapter 2.4 for an in-depth discussion of the role of mosquito vectors).

In other words, when and with whom ISEA *Pv* was introduced is unclear, and so is the reason why there are no traces of Papuan *Pv* in Indonesia whilst most other populations have remained in place for many tens of thousands of years. A Papuan introduction fits the time frame of ISEA *Pv* introduction, but not the phylogeny. A Palaeolithic introduction could fit the time frame but not the source. An Austronesian introduction fits the source but not the time frame. This corresponds with similar observations on the diversification of the three Malaysian Borneo samples, which fall well within an Austronesian route, but not an Austronesian timing. Expanding sample numbers would increase the accuracy of phylogenetic timing, as demonstrated by the narrowing of the Melanesian age estimate in this chapter as compared to the data from Taylor *et al.* (2013), see Appendix S3.7. Further sampling and genotyping of *Pv* strains from Taiwan, the Philippines and along the Austronesian migration routes in ISE Asia will be required to clarify the origins of *Pv* isolates currently circulating in the Indo-Malayan Region.

A side observation to the clear distinction between Melanesian and ISEA *Pv* samples is that even to this day the two have remained relatively strongly separated. The reasons for this are likely manifold: though population movement has become more frequent recently, levels of movement in the past are estimated to have been low. In addition, the different mosquito vectors present in each region might form a barrier to intrusion, on top of the already high prevalence of local strains. Thus there is a double mechanism of selection against introduced strains on each side. Even so, Melanesian type *Pv* strains were found in ISEA, and low levels of ISEA type *Pv* strains were found in parts of Indonesian Papua by Hupalo *et al.* (2016). Over time, with increased movement by modern day populations, these numbers are expected to increase, in relative numbers if not in absolute.

While the present study clarifies the main origin of *Pv* in Melanesia, it does not have sufficient resolution to entirely resolve the history of *Pv* in Southeast and East Asia. The immigrant haplotypes observed in this study highlight the importance of constructing the global *Pv* mt genome profile as soon as possible, before increased admixture confuses the historic patterns. In addition, the study once more brings into focus that modern migration is a complicating factor for *Pv* elimination, underlining the importance of understanding its origin and migratory patterns, both current and past. For this purpose, mitochondrial DNA is demonstrably useful (Rodrigues *et al.* 2014).

For a more complete understanding of *Pv* regional phylogenetic connections and past population dynamics it is crucial to heighten coverage in surrounding areas, in this case specifically around

certain regions in Island SE Asia. For example, Borneo looks like a probable East Asia-ISEA intermediate; East Timor merits further study as an outpost of Papuan populations, the Philippines needs further exploration as an Austronesian travel route, and different parts of Indonesia require further scrutiny for local classification that might be related to past human movements and ethnicities. The addition of these samples is of specific interest to better understand the migration of *Pv*, and possibly of human movements, in the East Asian and Pacific regions. Other drivers of genetic diversification of *Pv* strains have not been discussed here, as they have been discussed in a broader setting in *Chapter 2*.

INDEX CHAPTER IV

<u>4. 1. INTRODUCTION</u>	<u>106</u>
<u>4. 2. MALARIA PHYLOGENETICS SEQUENCING APPLICATIONS OVER TIME</u>	<u>109</u>
<u>4. 3. CURRENT APPLICATION OF MTDNA IN PHYLOGENETICS</u>	<u>113</u>
<u>4. 4. MTDNA AND MS DATA USAGE & AGGREGATION</u>	<u>115</u>
<u>4. 5. DISCUSSION & CONCLUSION</u>	<u>125</u>

CHAPTER IV

DOES SIZE MATTER?

PLASMODIUM VIVAX PHYLOGENETICS, AND THE APPLICATION & AGGREGATION OF DIFFERENT GENOME SEQUENCING DATA

*“All of this is not a coincidence
We are completely different [...]*
Because we have found our destiny” – BTS, DNA

4. 1. INTRODUCTION

Since the discovery of the function of the nucleic acid sequences which make up the genome, their sequential code has been used for many scientific purposes: from practical applications such as determining relationships between individuals, forensic evidence, disease discovery and drug design; to gaining a broader understanding of the development, adaptation and evolutionary history of humans – and all other organisms – through studying the similarities and differences across the genetic codes, as well as the functions of the encoded enzymes. Crucially, molecular techniques for understanding evolution and adaptation have been used to comprehend the history, functioning and epidemiology of many infectious diseases. Studying the genome of infectious diseases has often helped to clarify many aspects of their history, including the time and region of origin, the route and pace of spread, their adaptation – and with that knowledge, an understanding of their virility and impact on human populations patterns (e.g. for tuberculosis: Fletcher *et al* 2003, Donoghue 2009, Gagneux 2012, Comas *et al.* 2013; *Helicobacter pylori*: Moodley *et al.* 2012; black death: Bos *et al.* 2011; hepatitis C: Markov *et al.* 2012; hepatitis B: Dupinay *et al.* 2010).

Phylogenetics and phylogeography have, through the use of genetic sequences, become popular methods to reconstruct a species’ spread over space and time. With technological progress, the genetic sequences employed have evolved from small, targeted areas such as microsatellites (MS), small subunit ribosomal ribonucleic acids (SSU rRNA) and single genes, to the mitochondrial genome (mtDNA), chromosomes, and now the entire genomes of organisms (see *Section 4.2* for references). The current competency to extract and process large amounts of data, combined with the general attitude that more data is inherently better, has led to an increased concentration of current research and publication demands on whole genome data. However, whole genomes of many organisms are generally large, as many nucleic acid sequences easily run over several hundreds of

thousands to millions of base pairs. This means that often demands on laboratory work, data-processing and analyses are high, both in terms of cost and time.

Demands for using whole genomic datasets can be especially high in research on infectious diseases, as these frequently form an exception to the generalisation of organisms having very large single genomes. Smaller sized genomes are found in numerous viruses and bacteria. The smaller sizes, and the prolific use of whole genome sequencing in the medical world, have led to an inherent expectancy that new research on medical phylogenetics uses whole genomes. However, this is not always the most practical or relevant method for all genomic research applications, especially for those pathogenic organisms, such as malaria, with a larger genome; just as it would be inappropriate to sequence the entire malaria genome simply to diagnose a patient, certain parts of the genome are more useful for some phylogenetic applications than others.

The part of a sequence that is especially useful for a certain avenue of research often varies, and sequencing the entire genome remains expensive and cumbersome in terms of both laboratory work and computational processing. The limits of these sequencing methods are pushed further and faster by the ever-growing demands on them, which are increasing at a high rate. As such, it should be considered when high volumes of data are necessary and useful, and in which instances a more constrained approach is of greater or equal value. In a similar way to patient diagnosis, there is a strong argument to saving time and money in all sectors of research by using a targeted approach. This is not only useful to streamline research; it is also essential for the operation of smaller labs or those with limited resources. The approach of using shorter sequences has a twofold advantage in practical application: (1) the acquisition of new data is relatively straightforward; and (2) there are extended data sets available on them, which have been built up over extended periods of time. Despite catching up at a rapid rate, whole-genome data currently remain much more rare, and more complex to process.

What is lacking is an overview of the different genomes or genes used so far, and how their differences can, and have been, exploited in an appropriately targeted manner. Targeted uses of sequences to discover genetic trends are applied to phylogenetics, population genetics, genetic adaptation and genomic transcription. Weighing up the heavy costs of generating whole genome sequences (WGS), i.e. time and resources, against its actual usefulness in each instance, facilitates efficiency in labs' research approach. As this thesis is on the phylogenetics of *Plasmodium vivax* (*Pv*), this chapter will evaluate the history of phylogenetic approaches – and methods for DNA research within this field – in *Section 4.2*. In *Section 4.3* this chapter will also discuss the current pros and cons of using the mitochondrial genome for phylogenetics, in contrast to MS and WGS – reviewing what information on the history and population structure of *Pv* malaria can be gained from

both the mitochondrial genome and nuclear MS markers, compared to results seen using WGS. In addition, *Section 4.4* demonstrates the “do’s and don’ts” of aggregating marker datasets with different evolutionary histories from the same individuals and populations, which can happen thoughtlessly in WGS analyses. For this purpose, I analyse global mtDNA and MS data sets separately, as well as combined. The aim of this analysis is to test whether additional MS data allows for detection of more fine-scale differentiation on top of the broader evolutionary diversity already found in the mtDNA, as seen in *Chapters 2 & 3*.

The global datasets of *Pv* for both mtDNA (>1000 global samples) and microsatellites (>1000 global samples each, for at least 10 MS) are large, and partially overlapping. This allows for the production of an overview of which types of information can be gleaned from each dataset separately, from the combined data, and how these compare with, for example, phylogenetic whole genome data (Hupaló *et al.* 2016, Prugnolle *et al.* 2016). The expectation is that WGS will show an added depth of information, yet that the mtDNA remains suitable – and possibly the most efficient single candidate – for teasing apart the distribution of global *Pv* populations, especially when looking at the past beyond 500 years ago. Some population structure is expected to be found in nuclear genome MS (nDNA) data, as published in Koepfli *et al.* (2015). To understand the continued use of mtDNA in an age of increased whole genome sequences, this chapter will also look at what current publications of *Pv* whole genome sequencing studies have found, and how this differs from those purely based on mtDNA.

4. 2. MALARIA PHYLOGENETICS SEQUENCING APPLICATIONS OVER TIME

Before genetic sequencing, most phylogenies of *Plasmodium* spp., as with many other species, were based on comparative physiology of the organisms (e.g. Barta 1989). Genetic applications to study *Plasmodium* phylogeny started to be used in the 1980s (McCutchen 1984, Barta 1989). The 1990s saw a proliferation of genetic sequencing, with target sequences diversifying from mainly small ribosomal RNAs to genes, including the circumsporozoite protein gene (*csp*) and towards the end of the decade the first usages of the cytochrome b (*cytb*) gene on the mitochondrial genome (Waters *et al.* 1991, Brooks & McLennan 1992, Siddall & Barta 1992, Snounou *et al.* 1993, Arai *et al.* 1994, Escalante & Ayala 1994, Escalante *et al.* 1995, McCutchen *et al.* 1996, Qari *et al.* 1996, Kimura *et al.* 1997, Escalante *et al.* 1998). Genetic sequencing methods were soon used as a method of patient diagnosis, such that laboratory methods were developed for molecular diagnostics, including the ability to tell apart the different species infecting patients (Snounou *et al.* 1993, Arai *et al.* 1994, Kimura *et al.* 1997). Such developments went hand in hand with researching the broader *Plasmodium* species phylogenies and their position within the broader Apicomplexan phylum and its Haemosporidian subgroup.

With increased sensitivity of the sequencing regions used – including merozoite surface proteins (*Msp*) and microsatellites (MS) – and an increase in the number of regions sequenced per study (Barta 2001, Rathore *et al.* 2001), phylogenetic research began to take an interest in the origin and distribution of each separate species of *Plasmodium* from the late 1990s and early 2000s (Lim *et al.* 2000, Gardner *et al.* 2002, Leclerc *et al.* 2004). This interest was especially focused on the more virulent *P. falciparum* (*Pf*), which was entirely sequenced by 2002 mainly for the purpose of finding regions of interest for the development of antimalarial drugs (Carlton *et al.* 2002).

During this time, and across the following decade, there was a shift from using more widely known gene regions that were well preserved across different phyla, to finding regions with varying mutation speeds applicable to the question at hand. Regions with high mutability, such as microsatellites (Barta *et al.* 2001), though used in phylogenetics studies at first, were more applicable for studying population structures in real time and for informing multiplicity of infection, depending on the set of markers used (Barta *et al.* 2001, Leclerc *et al.* 2004, Imwong *et al.* 2006, Miao *et al.* 2012, Sutton 2013, Koepfli *et al.* 2015). The same goes for the glutamate rich protein (*Glurp*), *Msp1* and *Msp2* (Lim *et al.* 2000, Birkenmeier *et al.* 2010, Mwingira *et al.* 2011, Bowman *et al.* 2013, Congpuong *et al.* 2014).

By contrast, regions with low mutability, such as the mitochondrial genes *cytb* and cytochrome c oxidase subunit 1 (*cox1*), and in later studies the entire mitochondrial genome, were used and have retained the use to reconstruct phylogenies not only across species but also within them (Escalante *et al.* 1998, Barta 2001, Rathore *et al.* 2001, Carlton *et al.* 2002, Perkins & Schall 2002, Joy *et al.* 2003, Jongwutiwes *et al.*

2005, Mu *et al.* 2005, Liu *et al.* 2014, Prugnolle *et al.* 2015, Taylor *et al.* 2015, Rodrigues *et al.* 2018). Other genes or parts of genomes used for reconstructing phylogenies have come and gone, often in varying combinations (Escalante *et al.* 2005, Liu *et al.* 2014).

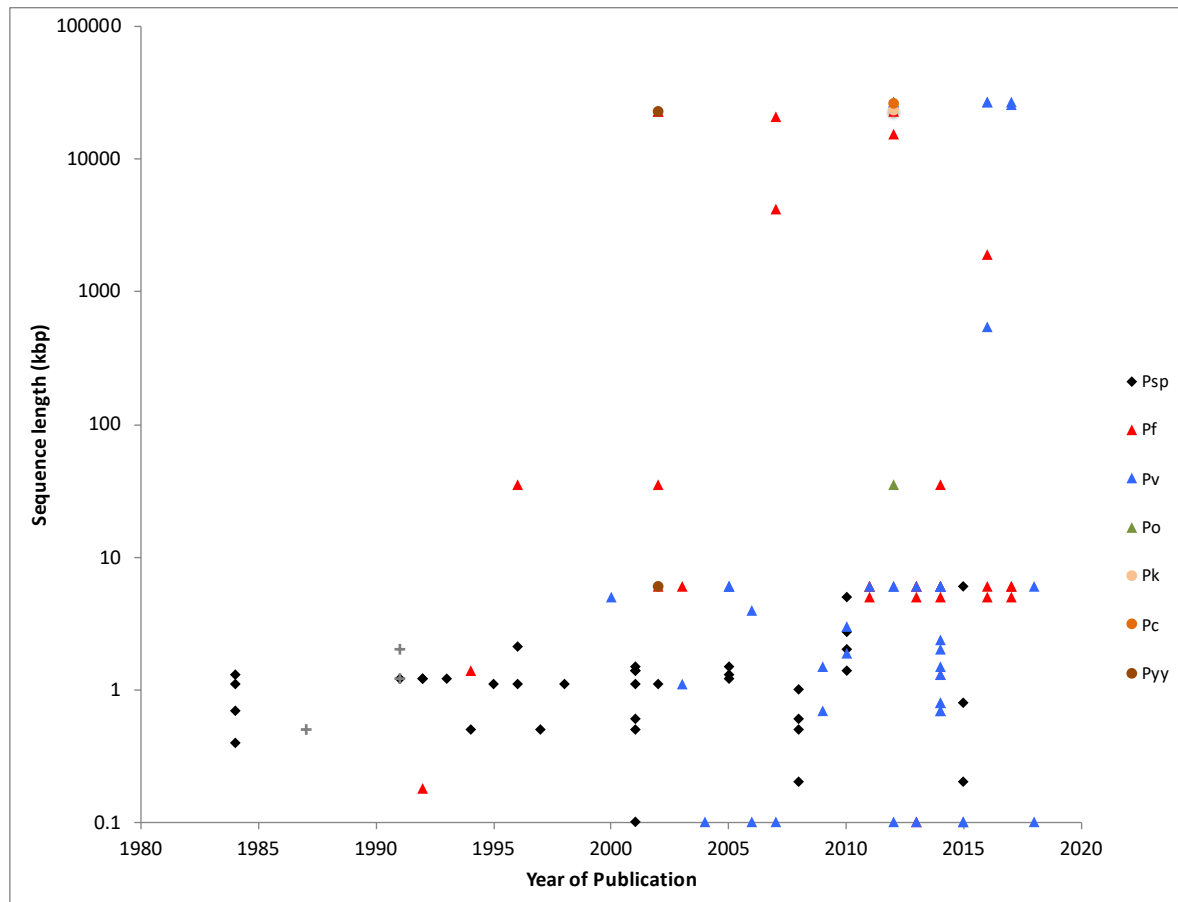


Figure 1

Overview of genetic sequence lengths used in *Plasmodium* spp. research, as published per year, coloured by species. Psp: Several *Plasmodium* spp., Pf: *Plasmodium falciparum*, Pv: *Plasmodium vivax*, Po: *Plasmodium ovale*, Pk: *Plasmodium knowlesi*, Pc: *Plasmodium cynomolgi*, Pyy: *Plasmodium y. yoelii*. Grey crosses denote publications on broader phyla or clades containing *Plasmodium* spp. Publications used can be found in *Appendix 4.1*.

Figure 1 gives an overview of the sequence lengths used in 73 publications studying genetic differentiation in or between *Plasmodium* spp., over time (year of publication) since the early days of gene sequencing, starting in 1984. An overview of the publications used can be found in *Appendix 4.1*. With regards to topic, the publications are focused on phylogenetics, population structure analyses and other topics comparing genetic variation across different populations or geographic regions – with the exception of publications where a large genome was newly sequenced for each species. The main focus is on *Pv*, but as progress on research in *P.* spp. was initially more concentrated on *Pf* the species was included to give a comprehensive overview. *Appendix 4.1* includes a summary of the genes researched, the main topic of research and the focal species, phylum or clade focused on.

In *Figure 1* a trend towards the use of larger sequences is visible, as described earlier. It also confirms the continuation of the use of smaller sequences for research, and illustrates how *Pf* research has remained continuously ahead of that on *Pv*, occasionally by a decade, such as with whole genome sequencing. Though this review is not a comprehensive overview, the trend of *Pf* research far outpacing *Pv* before 2010 remains to this day. This is likely due to *Pf* generally being much more virulent than *Pv*, attracting more attention. In fact, *Pv* has been recognised as a neglected tropical disease, whereas *Pf* does not qualify as such (Carlton *et al.* 2008, Carlton *et al.* 2011, Gething *et al.* 2012). *Pf* has had large amounts of research interest and investment, and thus had earlier opportunity to access costly genetic sequencing platforms. A cautionary note here is that besides attracting monetary investment at a higher rate, research for *Pf* is strongly facilitated by the ability to culture it in labs. This is currently still not possible for *Pv*. Regardless, the attention for *Pv* has increased dramatically in recent years. The cause for this is two-fold: 1) an increased interest in malaria in general, due to its resurgence and rapid development of drug-resistance, and 2) the realisation that eradicating *Pv* will likely be a much more complex task compared to *Pf* due to its dormant hypnozoite stage as well as occasional non-symptomatic carriers. *Figure 1* also confirms the continuous use and popularity of the (whole) mitochondrial genome from the start of its use up to the current date, for both *Pf* and *Pv*.

More modern methods used to look at phylogeny and population history, are through the application of whole genome sequencing (WGS) methods. WGS has been helpful in understanding the long-term evolution of some diseases. For example: WGS increased the precision of mutation rate estimates for both TB (Roetzer *et al.* 2013) and HIV (Cuevas *et al.* 2015, Zanini *et al.* 2017), increasing the understanding of their phylogenies over time. In both cases, however, the time span studied was relatively small compared to that of *Pv* in this thesis. In those studies, data from samples spanning much of the time line studied were available. Such is not currently the case; when looking into long-term phylogeny of malaria, where the time span from samples is at most 100 years but the lineage times span several tens of thousands. Arguably even 100-year old genomes, such as used in the study by Gelabert *et al.* (2016), would give very valuable insights. Such archival samples can provide original data from regions where malaria has now been eradicated. In addition, archival or archaeological samples could provide information regarding mutation rates of the *Pv* genome, even if this data comes from smaller time spans. Such samples of pre- or early drug treatment times can provide insight into which genes changed under selection pressure. Lastly, some strains had certain reputations, such as the Dutch *Pv*, which was known for its relatively mild symptoms and was used profusely in the experimental treatment of syphilis patients (Swellengrebel 1938, Verhave 2013). Sequencing the genomes of such parasites could provide functional insights into genes involved in pathogenesis processes.

For samples that are degraded due to their age, certain Next Generation Sequencing (NGS) methods, frequently used for WGS, are often seen as an appropriate platform. As long as these methods are combined with laboratory preparation methods that work around the low volume, and highly degraded length and quality, of the DNA targeted. As such, there is an increase in WGS used for ancient DNA studies (Stoneking & Krause 2011, Carpenter *et al.* 2013 (methodology), Hofreiter *et al.* 2014 (review), Linderholm 2015 (review), Fu *et al.* 2016, Gelabert *et al.* 2016). Sanger sequencing, NGS and 3rd Generation Sequencing (e.g. MinION and PacBio) methods, targeting long sequence reads - often used for modern genome sequencing - are not appropriate for ancient DNA. The latter two are not applicable, as the ancient DNA of the organism has fragmented over time into pieces too small to work with these methods. The former, as it is inefficient, especially as there will be a significant amount of short reads that need to be targeted. Naturally, when it is known which parts of the genome hold the most valuable information to understand the phylogeny over the timeline envisioned, Sanger sequencing methods can again be applicable.

Several NGS methods used for WGS are aimed at processing small fragments and so are useful for ancient DNA (e.g. Illumina, Roche 454 and Ion Torrent). Another useful aspect of using these methods, is that given the low volumes of DNA present in ancient samples, it is of great value to extract as much of the sequence as possible – a specific aim when using WGS. With aDNA the expectation is that not all genomic material can be retrieved and sequenced, assuming the presence of gaps in the data from the start. This is compatible with the issues around WGS. Even though the mitochondrial genome is a strong tool for understanding past phylogenies of modern DNA, when it comes to ancient DNA, NGS applications that target whole genomes might be more useful. Having additional data with many gaps instead of little data with many gaps would appear to be preferable. Even so, the decision whether to use WGS, more targeted sequencing of smaller markers, or both, would likely depend on: the amount of sample extracted & how much of it is used for each application, what is already known about the genome and its variation, and the research questions posed. There is significant importance in understanding what to sequence to answer which questions, especially with the increased interest in studying the past of infections, including malaria, from aDNA records such as medical archives or human remains – such as the studies of old malaria samples from Spain (Rodrigues *et al.* 2018) or even detecting infections in Egyptian mummies (Lalremruata *et al.* 2013).

4. 3. CURRENT APPLICATION OF MTDNA IN PHYLOGENETICS

Most of the history currently known for *Plasmodium* spp. – and other species including humans – is largely based on the information derived from mitochondrial genome sequences. There are many reasons why the mitochondrial genome has been used so prevalently in phylogenetic studies, despite having many opponents to its use, especially at a time when whole genome sequencing methods are becoming increasingly faster and less costly. The strongest argument against using only the mitochondrial genome is that it represents only a single genetic history for a species, and only for one sex (Ho & Shapiro 2011). It is inherited in a linear fashion, only through the maternal line, and thus does not have any recombination. The lack of recombination and linear inheritance are, however, the very reason the mitochondrial genome has been useful for phylogenetic studies: it is relatively stable and mathematically less complex to work with initially, especially when studying the phylogenetics of an organism deep into its past, due to its stability and continuity. For this reason it is an appropriate tool for investigating past migrations, across very broad geographies and/or long time scales. Some issues arise in the application of models reconstructing timelines with mtDNA, and these should be kept in mind. The main one is the strong weighting of mutations towards recent times – i.e. a relative overrepresentation of more recently accumulated mutations – due to strong purifying selection on the mtDNA, which removes erroneous mutations from the gene pool over longer periods of time (Ho & Shapiro 2011). Ho & Shapiro (2011) suggest these errors are avoided by sampling across different loci, and making sure data quality is of a high standard, as minor errors in the sequence data can lead to major errors in the results. This is why some WGS methods – certain ones of which trade out precision for size, missing out on parts of the DNA in favour of large coverage generally – can be less valuable for phylogenetic reconstructions than a more precise and targeted method of sequencing.

As Sutton (2013) explains in his review of microsatellite markers, for population genetics and phylogenetic research it is even more important to use a genomic marker that best reflects the level of variation necessary for the region and topic in question. It is sensible to use a marker that has the right amount of diversity for the study, rather than one that contains too much or too little variation across the populations studied. The amount of genetic variation is often influenced by the rate and type of evolution leading to particular patterns of diversity. Too little variation, which can be caused by either a low rate or similar selection pressures on the same part of a genome, will detect no population structure, as populations are excessively homogenous. On the other hand, an excess of diversity can compound population structure, as populations will display too much heterozygosity across all populations, diluting any structure, if present. Thus, the appropriate molecular markers will give a decipherable population structure. The same goes for time depth: for phylogenies estimating

time of divergence, it is important to use a genomic region that has a relevant mutation rate for the period of time and geographical region studied.

So far, the mitochondrial genome has suited the purpose for most phylogenetic studies looking into the past migration of *Plasmodium* parasites, including research into its origin and spread (Jongwutiwes *et al.* 2005, Mu *et al.* 2005, Imwong *et al.* 2006, Culleton *et al.* 2011, Manske *et al.* 2012, Taylor *et al.* 2013, Rodrigues *et al.* 2018). The resolution of mtDNA has been suitable enough for looking at the distribution of *Plasmodium* spp. on a global scale, especially for most regions infected with *Plasmodium vivax*. In Eastern Asia *Plasmodium vivax* populations can be discerned almost to a level between countries, as is seen in *Chapter 2* of this thesis and numerous other studies (Jongwutiwes *et al.* 2005, Mu *et al.* 2005, Imwong *et al.* 2006, Culleton *et al.* 2011, Manske *et al.* 2012, Taylor *et al.* 2013, Rodrigues *et al.* 2018). Based on this several studies claim it is a useful tool for diagnosing the origin of infections in recent migrants (Rodrigues *et al.* 2014, Ferreira & Rodrigues 2014). However, the resolution is not always so distinct. This is because any marker used for geographic positioning is only as good as its resolution. On the intercontinental scale mtDNA is the correct tool, but to understand smaller scale dynamics requires markers that have different fixed alleles or unique combinations of alleles presented across the populations in question. For mtDNA, the distinctions that are so clear on the eastern side of the Asian continent are minimal when looking at *Pv* from Africa across Eurasia up to India (see *Chapter 2*, Mu *et al.* 2005, Taylor *et al.* 2013). For this western region, the addition of other markers to aid strain identification would be useful.

Microsatellite (MS) data (Koepfli *et al.* 2015) show similar results to those seen in the mtDNA, as does whole genome sequencing (WGS) (Hupalo *et al.* 2016). However, both the MS and WGS datasets lack strong geographical representation across Africa, western and southern Asia. The MS data is largely represented by Madagascar and Sudan, with some samples from Uzbekistan, Azerbaijan and India. WGS data from Hupalo *et al.* contain several samples from India, three from Madagascar and one from Mauritania. With increased sample sizes from sites across this region it is possible that either of these genetic templates could find similar, or even greater, resolution than the mitochondrial genome, though currently it looks like the potential is low or the differentiation will be marginal. For MS data, the resolution for differentiation across the entire dataset is quite low, but strong enough for a general measure of divergence, as seen in Koepfli *et al.* (2015). Data from another WGS study by Pearson *et al.* (2016) is so greatly skewed towards samples from Southeast Asia and Papua - both Papua New Guinea and Indonesian Papua - that inferences about global patterns are not possible based on their results.

4. 4. MTDNA AND MS DATA USAGE & AGGREGATION

To illustrate the differences between mitochondrial and MS genomic data, a direct comparison between the two types of genetic data is shown in *Figure 2*, which compares the MS data from Koepfli *et al.* (2015) with the mtDNA data from *Chapter 2*. The two datasets overlap in some of their data; the same samples from several study sites were sequenced for both mtDNA and MS (mtDNA: $n = 813$, MS: $n = 624$, mtDNA and MS overlap: $n = 132$). South American data have been removed from both datasets, as this thesis concentrates on the spread of *Pv* across the eastern hemisphere. R (packages: *ggmap*, *ggalt*, *RcolorBrewer*, *maptools*, *sp*) was used to generate sample location maps for both datasets (*Figure 2a & 2b*). In addition, R (*ade4*) was used to generate a principle component analysis (PCA) for each dataset, as shown in *Figure 2c & 2d*. As Reich *et al.* (2008) argue in their review of the uses of PCA, it is useful for detecting finer structures in populations; it corrects for stratification in disease studies and can infer species history, especially concerning large migration events. This makes it a useful method for visualising differentiation between populations.

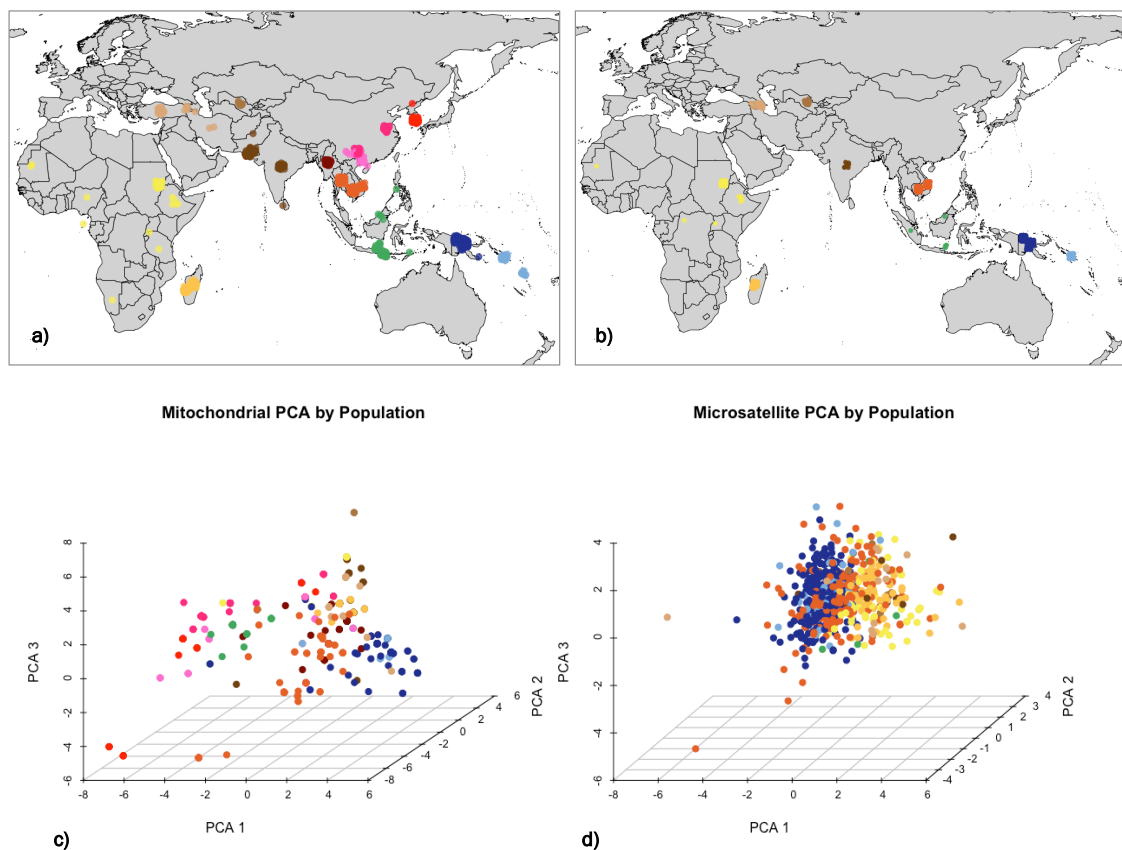


Figure 2

Sample site locations of a) mitochondrial genome samples, and b) nuclear genome microsatellite samples, of *Pv* across East & Southeast Asia and Melanesia.

Principle Component Analyses of c) mitochondrial genome samples, and d) nuclear genome microsatellite samples, of *Pv*. Colours are according to geographic origin as seen in a) and b).

The PCA results for the mtDNA (*Figure 2c*) show a strong degree of population variation across the first three PCA axes, indicating that the mitochondrial genome is picking up on variation between populations large enough to tell geographic structures apart relatively easily. The 11 MS used by Koepfli *et al.* (2015) all fall within the panel Sutton (2013) described as being well-balanced between endemism and diversity. As Koepfli *et al.* discovered, a panel of three of the microsatellites used in their study is enough to tell most populations apart at a global regional scale. The PCA results shown in *Figure 2d* show a degree of stratification across the microsatellite panel, but the separation between the populations is, although seemingly more linear, not as strong as in the mitochondrial genome. The STRUCTURE data analysis approach in Koepfli *et al.* shows a clearer regional stratification, which corresponds with that seen in the phylogenetic tree in *Chapter 2 (Figure 6)*. However, the mtDNA shows good resolution, independent of data analyses methods used, whereas the strength of the results using MS data vary.

Overall, when looking at dynamics on the continent-to-country-wide scale, the mitochondrial genome appears most appropriate on a time scale of >10,000 years. Yet, as mentioned previously, using only the mitochondrial genome comes with its own issues. The mitochondrial genome is maternally inherited, which means the dynamics of about half the population (the other sex) are ignored. Beebe & Cooper (2000) formulate it well when describing the differences in phylogenetic use for mitochondrial and nuclear regions to study the mosquito vector *Anopheles punctulatus* complex: “Mitochondrial coding regions have been extensively used for displaying phylogenetic relationships as they are single copy, clonally inherited, non-recombining and easy to manipulate. Nuclear genes have the potential of recombination events removing and inserting sequences of a different evolutionary history and thus removing linearity from the phylogenetic picture. However, the selection of possible markers is large compared with the mitochondria. Inferring phylogenies from a single gene can be misleading, as the phylogeny of the gene may not match the phylogeny of the organism; thus, the more markers that can be used, the higher the possibility of an accurate phylogenetic tree.”

In short, when using only mtDNA, other dynamics may be missed. For example, mtDNA lacks resolution on a smaller time scale, due to the slow rate of mutation of approximately 1 mutation per 10 - 20,000 years (*see Chapters 2 & 3*). Though the mtDNA in many species has a higher mutation rate than the nDNA (Ballard & Dean 2001), the mutation rate is still low enough to trace back ancestries several tens of thousands of years relatively easily (Jongwutiwes *et al.* 2005, Mu *et al.* 2005, Culleton *et al.* 2011, Manske *et al.* 2012, Taylor *et al.* 2013, Rodrigues *et al.* 2018). Specifically for *Plasmodium* spp., the mitochondrial genome is highly conserved, and a normalisation of mutation rates across both coding and non-coding regions has been deemed appropriate for phylogenetic calculations (Jongwutiwes *et al.* 2005, Mu *et al.* 2005, Preston *et al.* 2014, Taylor *et al.* 2013).

By contrast, nuclear genomes in *Pv* have vastly varying speeds of mutation, depending on the region of the genome, complexity, and its coding status, amongst other factors (Carlton *et al.* 2008). Nuclear genomes (nDNA) have generally been used to understand more recent population dynamics and structure across many species, specifically certain microsatellites (MS) and single nucleotide polymorphisms (SNPs) on the WGS of nDNA (Mu *et al.* 2007, Volkman *et al.* 2007, Neafsey *et al.* 2012, Hupalo *et al.* 2016, Prugnolle *et al.* 2016). This is because the recombination process in nDNA complicates the reconstruction of relationships between genetic sequences, increasing in complexity over time. The nuclear recombination rates in Plasmodia are especially high (Putaporntip *et al.* 2002, Carlton *et al.* 2013), which complicates calculations that depend on reconstructing genetic relationships from sequences.

Some of the polymorphic markers with an observed high rate of mutation include the microsatellites, often used in population genetic studies (Sutton 2013). These sites have been selected for population and infection studies because of their high mutation rates, which can inform on population diversity and change on recent timescales compared to mtDNA (Leclerc *et al.* 2004, Imwong *et al.* 2006, Ferreira *et al.* 2007, Miao *et al.* 2012, Sutton 2013, Fola *et al.* 2018). However, there are some mathematical issues regarding the use of microsatellites even for understanding population dynamics (Putman & Carbone 2014). Because of their fast mutation rate – 10^{-4} - 10^{-2} mutations per locus per generation (Lai & Sun 2003) – they are unsuitable for phylogenetic studies looking at time spans beyond 100-10,000 years before the present, depending on the rate of mutation saturation (Jongwutiwes *et al.* 2005). Instead, the mitochondrial genome is used for phylogenetics looking further into the past of a species.

However, there is a notion that adding MS to mtDNA data (*Figure 3*) for a higher resolution phylogeny might aid in understanding more recent population dynamics – as an increased amount of data, with a much shallower time-depth, is added to explore genetic relationships between samples for which information on relatedness over a larger time scale is already available. The theory is that adding a preselected marker panel that shows similar geographic resolution as seen in the mtDNA, such as the one suggested by Sutton (2013) and used by Koepfli *et al.* (2015), to the mitochondrial data set, might increase resolution. The obvious danger with this is the potential for over-parameterisation, and the weight of slowing down with too much data, such as with whole genome sequences.

This is why in the following trial, looking at how the datasets interact when combined, a PCA and a discriminant analysis of principal components (DAPC) were used for comparing the mitochondrial and microsatellite combined data. Both the PCA and DAPC are relatively quick methods to find data stratification. The data used were those samples for which both MS and mtDNA data was available ($n = 132$). Results for the initial PCA analyses with R (*ade4*, *ade4net*, *ggplot*, *plotly*) are shown in the Supplementary materials, *Figure S4.2*. It displayed a handful of outliers, with the main data

clustered closely together around the origin of the first three PCAs. Principle component 1 in the original analyses stood out as controlling most of the direction. The main outliers were found along this trajectory, with the remaining outliers appearing to dictate the other two main PCA directions. Therefore, the five most dominant outliers were removed, and the data re-analysed to see if the PCA could detect diversity within the main cluster more sensitively. Concentration of the initial principal components on the outliers, instead of separating the central cluster, could indicate too much ‘noise’ detracting from the factors that are important. The bar graph for PCA eigenvalues inserted in *Figure 3a* shows there are many different components having an almost equally strong influence on the structure of the data. In this case there is still some variation in the levels of differentiation detected, but the pattern points towards the danger of losing transparency in combined datasets, through the data becoming over-parameterised, and the PCA being unable to distinguish axes of importance. Here, added value is still present in most dimensions, and especially detectable in the first three principle components.

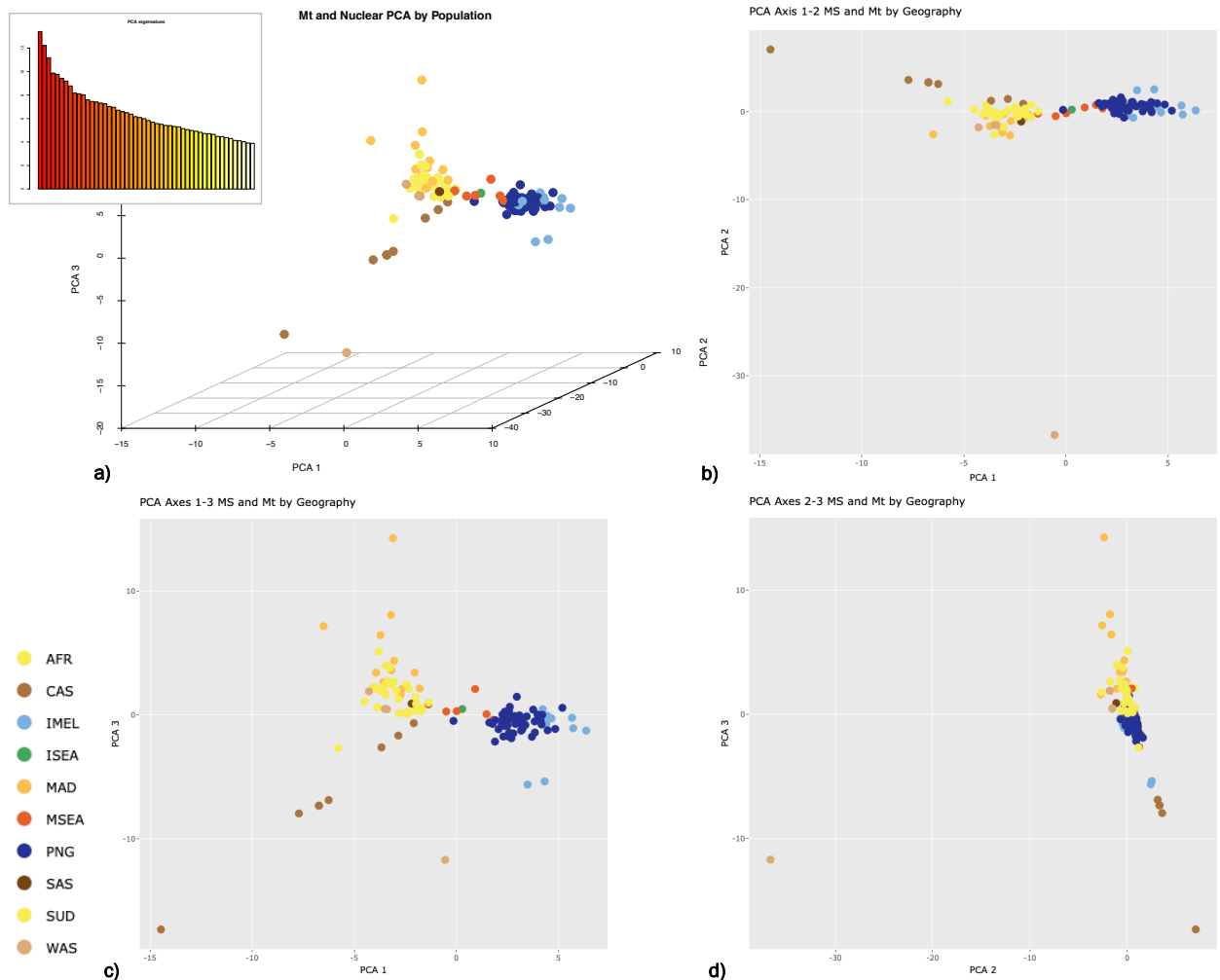


Figure 3

Principle Component Analyses of a) combined mitochondrial genome and nuclear microsatellite data – with initial outliers (samples: Mg_H02, SUD_60, SUD_47, SUD_28, PgAI01) removed – according to the first three PCA dimensions. The inserted bar graph indicates the levels of resolution provided by each principle component, up to 50 dimensions. b), c) & d) show the PCA for the combined data according to two axes for each graph: PCA1 & 2 in b), axes PCA1 & 3 in c), PCA2 & 3 in d). Colours are according to geographic origin as seen in *Figure 2*.

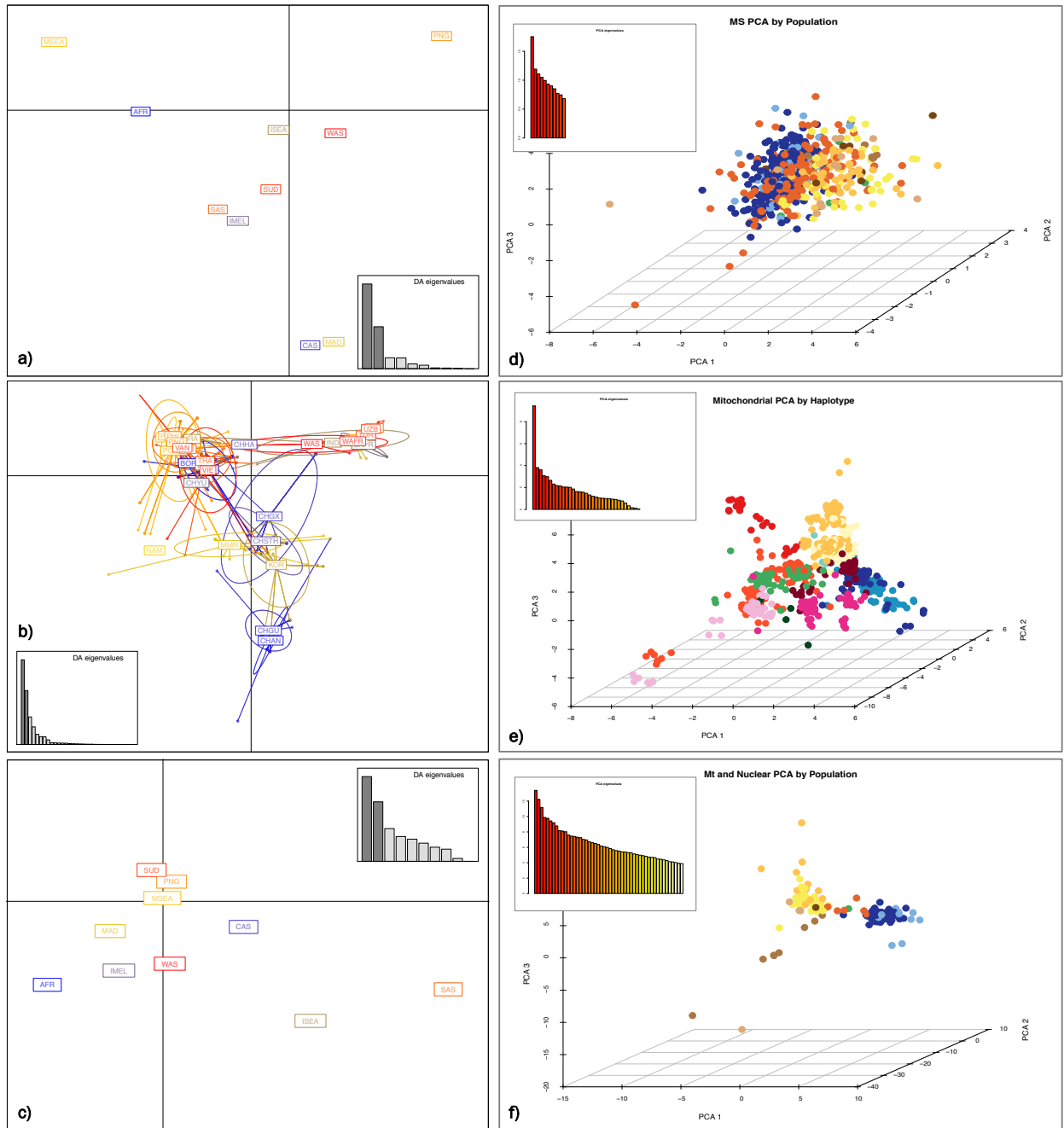


Figure 4

Figures a) microsatellite, b) mitochondrial, and c) combined mitochondrial and nuclear microsatellite genomic data, show the DAPC distribution for each dataset according to the two main principle components extracted.

d), e) and f) show the first three PCA dimensions for the microsatellite, mitochondrial and combined genomic datasets, respectively. The inserted bar graph indicates the levels of resolution provided by each principle component, up to 50 dimensions.

Colours are according to geographic origin as seen in Figure 2.

The results of these analyses, as seen in Figure 3, show clear stratification patterns per geographic region in the data. The three PCAs also appear to detect disparate dimensions present in the data. Compared to the data from the mitochondrial genome and the MS on their own, the combined data picks up on slightly different patterns, shown in Figure 4f; in this case clearly separating out the Central Asian data along an axis different from either the African (AFR, SUD, MAD), or East Asian/Melanesian (ISEA, MSEA/PNG, IMEL) data. In addition, both the African and East

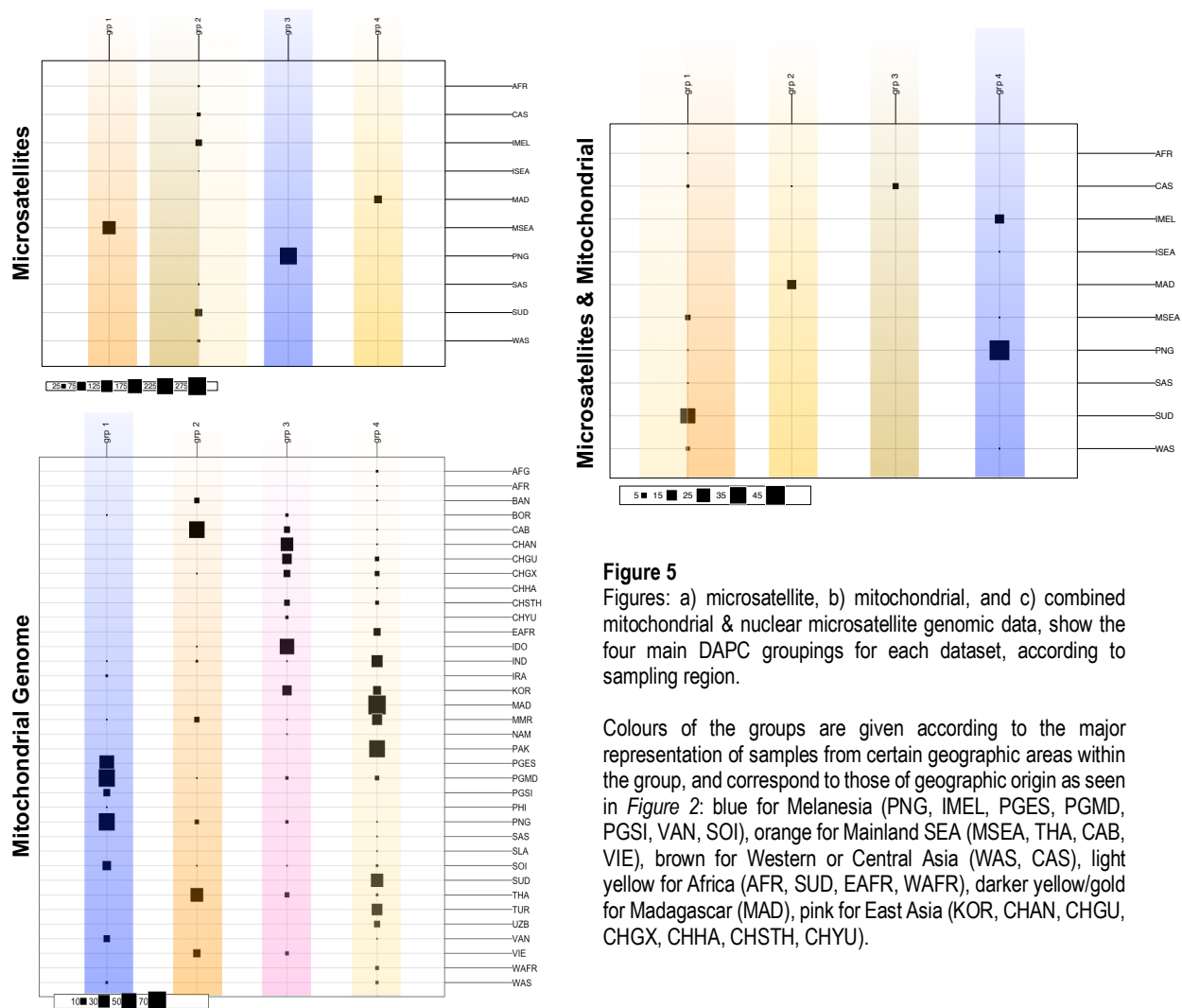
Asian/Melanesian axes display a geographically related trend of separation: samples from Madagascar are generally further removed from the centre than those from Africa/Sudan; and samples from Island Melanesia (IMEL: Vanuatu or Solomon Islands) are further from the centre point than those from PNG. Generally, these PCA trends in the combined dataset run along similar lines to the geographic distance from the origin of modern humans, somewhere in Eastern Africa – also hypothesised to be the probable starting point for the spread of modern *P. vivax*. In contrast, the distribution seen in the DAPC analyses in *Figure 4c* is much less clearly segregated, and is much more comparable to the results for the MS dataset (*Figure 4a*), where there are no clear geographic patterns or clumping of neighbouring regions, as is the case in the mtDNA dataset (*Figure 4b*).

The samples mainly originate from two mitochondrial haplotype groups as found in *Chapter 2*; samples are predominantly from Melanesia and Africa/Central Asia/South Asia, with minimal representation from other areas. This means that there are only two main groups to tease apart, which are clearly visible in *Figures 3a, b & c*. Another separation is found across the Central & Western Asian samples, which are seen slightly separated out in the mtDNA dataset – this separation could have been increased by the MS data.

There are several possible reasons for the structuring seen in the combined PCA data. Some of the main structures – such as the separation of Africa and Melanesia, with Mainland Southeast Asia on a more central axis between the two – are also seen in the mitochondrial genome. Central and western Asian samples also appear to branch out in the mitochondrial genome data, though this pattern is not as clear as in the combined data. The difference could be due to the smaller number of regions sampled; the combined dataset lacks samples from Far East Asia and Southern Asia, and has a limited number of samples from Mainland and Island Southeast Asia. As such, there is room for the principle components detected within the combined dataset to vary from those found in mtDNA, as it allows other patterns to come forward – even if the mtDNA data are the main provider of variation. Another reason for the variation in patterns is that the number of samples in the combined dataset is much smaller than either of the original datasets on their own. This could lead to patterns, though present on a smaller scale, not yet being fully detectable.

Due to the limited number of samples in the combined dataset, the concern was that some depth in diversity would be missed. There has been much discussion about the minimum number of samples necessary for a stable dataset (Preacher & MacCallum 2002, Mundfrom *et al.* 2005, Jung & Marron 2009, de Winter *et al.* 2009). Although previously suggested, following a set minimum number of samples is not entirely the appropriate approach. The number of samples necessary for stable data is influenced by several components, mainly the number of factors and the number of variables, as well as in some cases a level of loading (de Winter *et al.* 2009). The combined dataset used here has number of samples:

n = 132 (-5); to factors f: MS = 11, mtDNA SNPs = 406; to variables: p, nucleotides = max 4 for every mtDNA SNP factor, repeat units = many for MS. According to de Winter *et al.* (2009), lower sample numbers can be used with lower numbers of factors and higher numbers of variables. This would be the case for the MS dataset on its own. However, better resolution was present in the mtDNA dataset on its own, possibly due to the strength of the data used and the large number of samples. The combined dataset is a combination of two extremes on the data scale – which surprisingly seems to work in this particular PCA analyses – possibly due to the mtDNA data overpowering most of the MS data, but the MS data still getting enough influence to increase and extend levels of separation. The increased depth of variation in the data is not only visible in the seemingly enhanced separation of data in the 3D plot, but also visible in the increase in the number of principle components, as seen in the bar graph in *Figure 4f*. This indicates a scope for combining datasets to understand relationships between different geographic populations – in this case likely focused on the past, but with some input of modern-day dynamics to help increase resolution. It is important to understand how the different parts of the data are contributing, and whether combining the datasets is valid for its purpose.



In this case, how the MS and mtDNA data combine in the PCA and categorised DAPC (*Figure 5*) appears to suit the purposes of our study, and in a format that makes sense according to the structuring of the two combined datasets. The aim of this approach is to detect more fine-scale differentiation on top of the broader evolutionary diversity found in the mtDNA. Though the data comes from different types of markers, which lie on varying parts of the genome and have distinct evolutionary histories, this example appears to detect levels of diversity in the combined dataset similar to those detected in the mtDNA and MS datasets separately, even without appropriate measures to account for the specific types of evidence provided by the mtDNA and MS data. The databases are heavily influenced by ‘identity by state’ (IBS, for MS) versus ‘identity by descent’ (IBD, for mtDNA). MtDNA is better for inferring identity by descent because mutations are rarer and not shuffled by recombination. The rate of evolution, being slow, gives the time-frames necessary for understanding long-term phylogenies. In contrast, MS tends to give insight into the last 10-50 years of connections between more localised populations – such as within countries – and can be useful for questions on more recent time scales and local population genetics where there is a significant chance that parasites are related if they carry the same alleles.

Combining the datasets appears to give broad strokes of similarity with the patterns found in the mtDNA, enhanced by data found in the MS, making it possible to detect more subtle differences between regions. Though the differences in type of data for each of these datasets (such as IBD and IBS), the risk of over-parameterisation, and the separate levels at which each dataset measures genetic variation, are all strong reasons not to combine these datasets – especially if the aim is to make conclusions about phylogeny and descent – to study genetic differentiation between regions, both these datasets are appropriate. In addition, the structuring of the principle components in the PCA makes sense for our purposes; the dominant contributor to the principle components appears to be the mtDNA, and secondary patterns are added by the MS. Geographically and historically, this structuring makes the most sense; mtDNA provides most of the variation between regions, while MS contributes to the diversity within regions. As such, it appears that most of the results of combining the dataset are in line with our reasons for attempting the fusion. The exception is the DAPC dimensions per population as seen in *Figure 4c*, the pattern of which might indicate that the MS data played a larger role for the dominant dimensions detected there. Once a MS level is chosen as the primary dominant, mtDNA input could possibly worsen the situation. As observed, there is the risk of over-parameterisation, which means there are heightened chances of choosing an erroneous dominant parameter.

Another point is that even with the positive results observed, the current combined data are too limited to show the more enhanced diversity patterns present in the other datasets, and are of interest

to a better understanding of both regional and global diversity, and possibly history, of *Pv*. The lack of data is likely also the reason why certain – less important – patterns of division are detected more sensitively in the DAPC groupings in *Figure 5c*, compared to the ones that are much clearer in *Figures 5a & b*. The main patterns seen in *Figure 5a & b* which are missing in *Figure 5c* – i.e. East Asia, and Mainland SE Asia – are the ones that lack data entirely or are much reduced in sample numbers.

The process of combining genetic data with vastly differing backgrounds happens, often inadvertently, with WGS data, and should be approached with caution. Currently, the focus of WGS data is purely on SNPs, but each SNP comes with its own evolutionary history, and it is important to keep this in mind. Regarding increasing resolution for time depth with an increase in data: using MS data phylogeny for reconstructing historical connections is generally intractable – due to the uncertainty in speeds of mutation accumulation, and the high variability possible between different microsatellite regions. According to some, the issue of calibration makes it almost impossible to compare or combine results across studies or regions, making results difficult to interpret or generalise. This makes MS data mainly useful for applications that necessitate highly variable markers, over a smaller time frame. Even then, there are highly polymorphic genes that can serve a similar purpose. The expectation is that increasing data with SNPs from WGS data is a better approach, though they too have similar negatives to using MS data, including: 1) speeds of mutation of most of these SNPs are still unknown, extremely variable across the entire genome and under the influence of recombination (*Pv*: Pearson *et al.* 2016, *Pf*: Su *et al.* 1999, Volkman *et al.* 2006, Volkman *et al.* 2007, Bopp *et al.* 2013); and 2) insights are often complicated by patients and mosquitoes combining multiple infections (Pearson *et al.* 2016), which increases the complexity of the data gathered, especially when it is unclear which mutations are coming from the same strain. Due to the much higher variation present on the nuclear genome, the separate genomes of multiple infections are harder to identify.

It is argued by some that the strength of WGS lies in the vast number of markers which can be used to answer a question of interest. More accurate might be the statement that the strength of WGS lies in the vast number of markers that can be used to answer questions of interest. Currently, the phylogenetic tree for WGS data, as seen in Hupalo *et al.* (2016), has no more, and possibly less, resolution than the one from mtDNA, as seen in *Chapter 2*. This could be due to less data, and/or a less targeted focus on the geographic region of interest, which could change the parameterisation of data targeted for separation. In contrast, the MS data combined from different samples, and studies, across the globe by Koepfli *et al.* (2015), shows a distinct differentiation between the separate geographical regions sampled, demonstrating the possibility to use MS data for some comparisons between regions and studies.

In the end, it is not so much about the size of the genome data used that is of importance to historical investigations and the understanding of past dynamics behind the spread and distribution of a species. What is critical are the specific characteristics of the markers used. As such, the expectation is that the mtDNA template will remain a stable tool for phylogenetics research investigating the broad geographical relationships of samples and populations over large amounts of time, including for example the provision of an extra dimension for understanding historic population dynamics and connections in the human host. As the understanding of nuclear genome evolution grows, with the aid of WGS applications, knowledge will hopefully increase of supplementary genetic datasets that can add depth and resolution on numerous different geographic and temporary scales, besides the deep time of mtDNA and shallow time of MS.

4. 5. DISCUSSION & CONCLUSION

A clear result from this overview is that there is still great value in using smaller genome sequences, such as mtDNA and genetic markers, in the lab. Not only are they often cheaper and faster to work with, there is value in their precision to answer targeted research questions. There is also value in the already available knowledge about the behaviour of the markers regarding, for example, mutation sites and rate; and the large amount of data that has been built up over years of research. Accumulating data, for example through WGS, can give higher resolution for both spatio-temporal information and for mutation rates, but the possibility remains that valuable data get lost in the general noise, through over-parameterisation and/or as a result of aggregating data with varying evolutionary histories. Knowing what data to use and how to apply them to your research question in a focused manner is of substantial importance. As such, the continued use of well-understood smaller genomic markers, including the mtDNA, holds significant value. This can be extended to the use of WGS data in ancient DNA research settings, where some markers, and/or the aggregated dataset, will not be appropriate for all questions asked.

As shown in the data example in this chapter, aggregate data can be used, but with great care as to what questions are being asked. Not only should these be appropriate for both data types separately and makes sense for the combined dataset, but there should be a thoughtful approach to which hierarchical structuring is applicable for the results of the combined data. In our case it made most sense, both geographically and temporally, to have the main patterns from the mtDNA dominate the principle components before those from MS. Though in this case testing of the data to see which individual dataset contributed most to which principle components was left out, the suggestion is to take this approach in any future use of data combining strategies – to test whether the stratification of the principle components follows the expectations and needs of the study.

A further issue with WGS lies not so much in the amount of data generated, but in the issues arising from the process, such as gaps in the data. Many laboratories using WGS methods will follow up their sequencing with numerous PCRs and different sequencing methods to fill some of the gaps left during the initial process (Hupalo *et al.* 2016, Pearson *et al.* 2016, Benavente *et al.* 2017). Though this is not uncommon for traditional sequencing methods, it does increase both cost and time of an already costly and time-intensive method. Filling these gaps in WGS is not a standard practice, which means much data can be lost especially in genomes with low variation, as is the case with mtDNA (Gardner *et al.* 2002, Kidgell *et al.* 2006, Mu *et al.* 2007, Volkman *et al.* 2007). Whether this is of importance depends on the research aim of the sequencing, but when it comes to understanding phylogenetics of a genome that has highly informative but low numbers of mutations, any loss of data can make a huge difference in the inferences made.

For example, although for aDNA applications the benefits and negatives of WGS overlap with its aims and expectations, for sequences with low variation that are highly informative in phylogenies – such as the mitochondrial genome – missing data can be crucial in how one sample is related to others. This sensitivity means that the increased inaccuracy of many of the Next Generation Sequencing methods can be an Achilles heel in phylogenetics and other precision dependent analyses, especially in highly conserved sequences that depend on a small number of specific mutations for differentiation. At times, this means WGS data can only be compared to other WGS data as smaller comparisons lose value. The consequence is that use of WGS data carries a high risk of losing out on both incorporating, and being incorporated in, studies that cover smaller specific sequences. Naturally with the development of novel sequencing methods and platforms, and the reduction in cost and time, these issues are slowly being resolved.

As with TB and HIV, the main use of WGS in malaria research has been to understand the functioning of the organism through discovering genes and their functions (Carlton *et al.* 2002, Gardner *et al.* 2002, Neafsey *et al.* 2012, Tachibana *et al.* 2012), with only perfunctory attention given to phylogenetics. Following from that, focus has been to understand which regions are involved in the development of anti-malarial resistance within the parasite and on setting new steps towards targeting regions for drug development (Carlton *et al.* 2002, Mu *et al.* 2007, Volkman *et al.* 2007, Hupalo *et al.* 2016, Pearson *et al.* 2016, Benavente *et al.* 2017). It is expected that in the future this avenue of research will increase significantly, as drug resistance to all currently available drugs on the market has become a big issue, especially within Southeast Asia. Progress on the eradication of malaria has recently ground to a sudden halt due to these new developments, and medical research funding has globally flagged malaria as a priority subject.

The increase in research on drug resistance and anti-malarial drug development utilising WGS will likely feed into phylogenetic research, either through access to data or through technical and theoretical developments. Though much of the mtDNA and MS data available for *Pv* have been explored to a great extent, and despite the fact that the addition of WGS data from the same regions does not currently show significant supplementary resolution, avenues for improvement remain. This enhancement lies mainly in additional global coverage. As discussed in the previous chapters of this thesis there are several geographic regions of interest that are not yet present in the current datasets, or are significantly underrepresented. Adding these regions to the current mtDNA dataset is still of great value to understanding the past of *Pv*. In addition there is much room to explore the past of other *Plasmodium* spp. besides *Pv* and *Pf*.

CHAPTER V

CONCLUSION

THE DISTRIBUTION OF *PLASMODIUM VIVAX* ACROSS THE EASTERN HEMISPHERE: OBSERVATIONS OF HISTORY THROUGH THE MITOCHONDRIAL GENOME

“Because we have found our destiny.

DNA.”

– BTS, DNA

5.1 CONCLUSION

The previous chapters have shown that reconstructing the past dynamics behind the current distribution of *Plasmodium vivax* (*Pv*), as found in its mitochondrial genome, is not a straightforward task. To understand its relationship with human history there needs to be an understanding not only of human history itself, but also of the other factors influencing the genetic patterns as seen across the geographic landscape. *Chapter 2* indicated the strong influence of anophelene mosquito vector species distribution on the possible preservation, or homogenisation, of historic introduction patterns. In other species, including humans, genetic patterns have been strongly influenced by geography, including distances, climate and geographic barriers. In the present study, the influence of these factors was detected in the genetics of *Pv*, but there is still much room for further research and improvement on this topic. It would be interesting to see if there are any other geographic barriers besides the usual - oceanic, desert, mountain - that influence the distribution of *Pv*. As climate strongly controls vector distribution, the expectation is that more unexpected climatic barriers could be found that are unusual for the human host species, but correspond with vector species, such as the India/Bangladesh and Mexican forest/coastal distributions discussed in *Chapter 2*.

Besides a strong influence from vector species and some indication as to the influence of geographic parameters, *Chapter 2* showed a strong connection between the distribution of *Pv* and the past of anatomically modern human (AMH) migrations. Date estimates for the time of the most recent common ancestors (TMRCA) of *Pv* corresponded with archaeological finds of *Homo sapiens* (*Hs*) in the same region. In addition, patterns of age distribution and phylogenetic connections of *Pv* overlapped geographically with those of *Hs*. This was confirmed in *Chapter 3*, where a closer look at the age of introduction of *Pv* indicated an early arrival with Papuan migrants, around 65 thousand years ago (kya). The introduction of *Pv* in the Melanesian region is not connected with the much later arrival of the Austronesian human populations.

A complicating factor to the patterns in the dataset is that the Indonesian, or Island Southeast Asian (ISEA), samples are a difficult fit with surrounding regions. Though they fit an origin from East Asia, corresponding with the human movements of the Austronesian peoples, the TMRCA of *Pv* for this region is much earlier than the arrival of the Austronesian peoples. This could mean the parasite was introduced from East Asia far earlier than the movement of the Austronesians or arrived in the region from another geographic source that gave rise later to both the East Asian and the ISEA strains. Alternatively, the introduction could have occurred with the Austronesians but the adaptation to the local vectors and diverse landscape, and possibly competition with earlier subtypes of *Pv* already present in the area, drove an increased rate of genetic diversification. To increase our understanding of the past dynamics in this region additional sampling across the Indonesian islands and subsequent research would be of great interest. Further sampling around Borneo/Kalimantan would be of value as the current three samples appear to differ from both East Asia and from the other ISEA samples, taking up an intermediate position that indicates a possible route from East Asia to ISEA. If archival samples from Taiwan were available these might help to elucidate the connections within the region even further, as Taiwan is seen as one of the main regions contributing to the Austronesian migration. *Pv* genetic material could possibly confirm or exclude a connection between ISEA vivax malaria and the Austronesians, in similar fashion to the recently sequenced archival Spanish sample from the early 1900s (Gelabert *et al.* 2017) aiding in the understanding of the origins of American *Pv*.

Another region that would be of significant interest to study in greater detail is East Asia itself. The patterning and diversity found across China and neighbouring regions indicate a diverse past. This includes possible coastal links between Korea and South-eastern China, and the isolation of interior regions such as Guizhou, which is separated from mainland Southeast Asia (MSEA) by a river delta system but has a similar climate to that region. Inland connections between Anhui and Guangxi also tentatively exist between their historic human populations, the Han. Due to the regional distinctions found in *Pv* mtDNA across East Asia, modern imports can be relatively easily identified. For example, claims that refugees and migrants from Myanmar introduced *Pv* to Southern China could be checked relatively easily through mitochondrial DNA, as the fingerprint for each region is quite distinct. According to the dataset used in *Chapter 2*, dating to before 2005, there was little to no indication of this being the case at the time.

These future avenues of research for understanding the past of *Pv* through its current genetic distribution can include genomic data from other sources besides the mitochondrial genome. Currently the diversity of single nucleotide polymorphisms (SNPs) from sequencing across the entire *Pv* genome (WGS) has the potential to add extra depth to identifying regional genetic structures. However, so far WGS has not found increased resolution for this application, but equally it has not been specifically applied to do so. Instead the WGS studies have mainly focused on diversity in drug

resistance adaptation genes. There is still much scope for WGS data to be used for other applications including increasing the understanding of diversity and which parts of the genome contribute to it. Hupalo *et al.* (2016) made a good start on this, recognising the influence of mosquito vector species on the genetic diversity in *Pv*. In addition, the nuclear and apicoplast genomes could be explored for regions that can increase the accuracy of the molecular clock, such as has been done in tuberculosis, including the conceivable use of different regions for looking at varying time depths. However, caution is advised for increasing data at any cost. Not all data are appropriate for all applications, and more data increases both costs in the lab and costs of bioinformatic processing. Combining data from different sources for analyses can introduce complications, though many of these can be avoided by working with only one type of genetic data, such as SNPs. Though there is a distinct possibility for combining data with different evolutionary histories and research applications, much care should be taken in understanding whether the application makes sense for both datasets, how to combine these datasets in a sensible way, and whether the results are reasonable given the aims. Currently it remains unclear if combining data from microsatellites (MS) to mitochondrial SNPs would be useful for detecting smaller scale differentiation patterns, within the more global distribution found across populations with purely mtDNA data. There are indications of success, but the dataset used to test this in *Chapter 4* is limited in sample numbers and especially in regions of origin.

Even with the advance and suggested increase of the use of WGS, *Chapter 4* emphasises that mitochondrial DNA remains an appropriate and useful template for the exploration of phylogenetics, especially when looking into the past over several thousand years ago. The relative consistency of mutation rates across the genome and lack of recombination have made it a solid approach for a molecular clock. Equally it is a useful template for understanding recent introductions by modern migrants as the mitochondrial genome has a strong regional signature on a continental to countrywide scale. Another recent application of the mitochondrial genome is for detecting and identifying infections in patients, where two different sequences from the mtDNA appear better tools for diagnoses than the one currently used: SSU18 rRNA.

In all, this thesis has shown strong connections between the large-scale early *H. sapiens* migratory movements and the spread of *Plasmodium vivax* across the Eastern Hemisphere. In addition, it provides more resolution around regions in East Asia that are, and could be, connected to more recent past migrations. Lastly, the resolution and strong regional fingerprint of *Pv* detects recent introductions of *Pv* between global regions, such as introductions from Melanesia to the Philippines, Central Asia and Iran seen in *Chapter 2*. Though resolution between mutational steps is around 10,000 years, the assumption is that population dynamics would have caused older introductions to be either removed or assimilated. Single or minimal presence of certain haplotypes is therefore assumed to be an indication of relatively recent introduction. This variety in levels of distinction and

the understanding of mutational rates across the genome make mtDNA a good template for studying the connection between *Pv* and human migration movements.

Lastly, the extended exposure of human populations to malaria and its strong selective pressure on human genetics provides another avenue for further research into their connected pasts. Genetic adaptations to malaria provide a time-frame estimate for exposure to the disease. In addition, the prolific variation found in these adaptations can indicate historic connections between populations. Examples of this include the presence of the Southeast Asian Ovalocytosis antigen in both Pacific and certain East Asian populations, and the presence of both sickle-cell anaemia and the same Duffy negative antigen in Africa and Afro-American populations.

In all, this thesis has provided a broad framework for further research. It has confirmed historical and evolutionary influence on the geographical variation patterns seen in the *Pv* mitochondrial genome, and identified at least two major contributors: early *Homo sapiens* dispersal patterns, and the distribution of anopheline vector species. In addition it has identified regions of high interest regarding genetic patterning in *Pv*, such as East Asia, Island Southeast Asia and the India/Bangladesh divide. It is hoped that the results from this thesis will contribute to a broad variety of further inquiry into malaria genetics and human population dynamics in the future.

APPENDIX

INDEX APPENDIX

<u>APPENDIX</u>	<u>I</u>
<u>CHAPTER II</u>	<u>II</u>
APPENDIX S2.1	II
APPENDIX S2.2	III
APPENDIX S2.3	IV
APPENDIX S2.4	V
APPENDIX S2.5	VI
APPENDIX S2.6	VII
APPENDIX S2.7	VIII
APPENDIX S2.8	XI
APPENDIX S2.9	XII
APPENDIX S2.10	XIII
APPENDIX S2.11	XIV
APPENDIX S2.12	XV
APPENDIX S2.13	XVI
APPENDIX S2.14	XVII
APPENDIX S2.15	XVIII
APPENDIX S2.16	XXII
<u>CHAPTER III</u>	<u>XXIII</u>
APPENDIX S3.1	XXIII
APPENDIX S3.2	XXIV
APPENDIX S3.3	XXV
APPENDIX S3.4	XXVI
APPENDIX S3.5	XXVII
APPENDIX S3.6	XXVIII
APPENDIX S3.7	XXIX
APPENDIX S3.8	XXX
APPENDIX S3.9	XXXI
<u>CHAPTER IV</u>	<u>XXXII</u>
APPENDIX S4.1	XXXII
APPENDIX S4.2	XXXIII

CHAPTER II

APPENDIX S2.1

Table S2.1

Mastermix and thermocycle are the same for the amplification of both halves of the mitochondrial genome.

S2.1.1. Mastermix reagents, concentrations and volumes for the PCR amplification of the *P. vivax* mitochondrial genome.

S2.1.2. Thermocycle protocol for the PCR amplification of the *P. vivax* mitochondrial genome.

MASTERMIX 1 RFS			
Total Volume for 1x:	30	ul	
	[I]	[F]	x1
ddH ₂ O		(total vol. - (MM+DNA))	20.95
Buffer (10x + MgCl ₂)	10x	1x	3
dNTPs	10mM	0.25mM	0.75
Taq/Enzyme (Roche FSHF)	5U/ul	0.05U/ul (1.5U)	0.3
Primer Forward	10uM	0.5uM	1.5
Primer Reverse	10uM	0.5uM	1.5
SubTotal			28
DNA			2
TOTAL			30

MASTERMIX 2 MEntConc			
Total Volume for 1x:	30	ul	
	[I]	[F]	x1
Sterile ddH ₂ O		(total vol. - (MM+DNA))	22
Buffer B1	10x	1x	3
MgCl ₂	25mM	1.5-2.5mM (6-10uM/100ul)	2
dNTPs	10mM	0.25mM	0.35
Taq/Enzyme (FirePol Pink)	5U/ul	0.025U/ul (0.75U)	0.35
Primer Forward	10uM	0.5uM	0.75
Primer Reverse	10uM	0.5uM	0.75
SubTotal			29
DNA			1
TOTAL			30

2PP Seq Set	Long Thermo Prot	
PP1A	PvMt_0323F	PvMt_3193R
631	72	
PP2A	PvMt_2946F	PvMt_0426R
843		74

10 PP Seq Set		
PP1B	PvMtQ_5795F	PvMt_0426R
631	68	74
PP2B	PvMt_0323F	PvMtP_1166R
843	72	72
PP3B	PvMyO_1022F	PvMt_1699R
677	70	60
PP4B	PvMtP_1584F	PvMtQ_2402R
818	70	70
PP5B	PvMtP_2260F	PvMt_3193R
933	72	72
PP6B	PvMtQ_2957F	PvMtP_3890R
933	70	64
PP7B	PvMtQ_3790F	PvMtR4433
643	70	70
PP8B	PvMtP_4117F	PvMtP_4927R
810	66	70
PP9B	PvMtE_4685F	PvMt_5409R
724	62	60
PP10B	PvMtP_4978F	PvMtP_5966R
988	68	66

LONG THERMOCYCLE RFS			
	Temp.	Time	
	94 C.	2 min.	
10x	94 C	30 sec.	Den
	61 C	1 min.	An
	72C	3 min. 30 sec.	El
30x	94 C	30 sec.	Den
	61 C	1 min.	An
	72C	min. 30 sec. +20 sec/cyc	El
	72C	7 min.	
	4 C	%	forever

SHORT THERMOCYCLE FPP			
	Temp.	Time	
	94 C	15 min.	In. Den
40x	94 C	30 sec.	Den
	61 C	1 min.	An
	72 C	1 min. 10 sec.**	El
	72C	10 min.	final el.
	4 C	---	forever

**1 min. 10 sec. used for all reactions here

SHORT THERMOCYCLE RFS			
	Temp.	Time	
	94 C	2 min.	
40x	94 C	30 sec.	Den
	61 C	1 min.	An
	72C	1 min. 20 sec.	El
	72C	7 min.	
	4 C	%	forever

PrimerPair:			FragLength (bp)	**Elongation
PP1B	PvMtQ_5795F	PvMt_0426R	631	1 min.
PP2B	PvMt_0323F	PvMtP_1166R	843	1 min.
PP3B	PvMyO_1022F	PvMt_1699R	677	1 min.
PP4B	PvMtP_1584F	PvMtQ_2402R	818	1 min. 20 sec
PP5B	PvMtP_2260F	PvMt_3193R	933	1 min.
PP6B	PvMtQ_2957F	PvMtP_3890R	933	1 min. 20 sec
PP7B	PvMtQ_3790F	PvMtR4433	643	1 min.
PP8B	PvMtP_4117F	PvMtP_4927R	810	1 min.
PP9B	PvMtE_4685F	PvMt_5409R	724	1 min.
PP10B	PvMtP_4978F	PvMtP_5966R	988	1 min. 20 sec

PrimerPair	PCRPrimerName	Sequence 5' - 3'	SequencingPrimerName	Sequence 5' - 3'
PP1B	PvMtQ_5795F	ATGGATTGGATGTCAGTTACAC	PvMtN_0424R	ATAGACCGAACCTTGGACTC
	PvMt_0426R	GTCCAAGGTTTCGGTCTATAATTTTCC		
PP2B	PvMt_0323F	ACGCTGACTTCTCGGCTAACTTC	PvMtR_1165R	TCTGGAAGCGTCTGTAAGG
	PvMtP_1166R	AGTTATATTCGGAAGCGTCTGTAAG		
PP3B	PvMyO_1022F	GCAGGACGGAGATTACCCGAC	PvMtL_1693R	ACTAGCGGTGGAATACATTG
	PvMtL_1699R	ATCCACCGCTAGTGTTC		
PP4B	PvMtP_1584F	AGGGAACAACTGCCTCAAGACG	PvMtL_1585F	CAAACCTGCCTCAAGACGTTTC
	PvMtQ_2402R	TAGCAAGTGCATCATGTATGACAGC		
PP5B	PvMtP_2260F	ATCCTAAATGTAAGTATTCAGTTGTTTG	PvMt_2268F	AGTATTCAGTTGTTTGAAGTGATG
	PvMt_3193R	GAACGCTTTTAACGCCTGACATGG		
PP6B	PvMtQ_2957F	TATTACTAAGATAAAGAACTCCAGGC	PvMtL_2974F	GGCGTTAACTGTAGAGTTTG
	PvMtP_3890R	ATGTTAATAATAGCATTACAGATGTA		
PP7B	PvMtQ_3790F	CCTGTTGCAGTAGATGTTATCATTG	PvMtR_3794F	CAGTAGATGTTATCATTGTTGG
	PvMtR_4433	GCTGCATTACCTAATATTACTCCTG		
PP8B	PvMtP_4117F	ACCAGCATTGGTGTATTAGTCATG	PvMtL_4925R	AGATATTTCTGGAGTATAACGAC
	PvMtP_4927R	GCATAAGATATTTCTGGAGTATAACG		
PP9B	PvMtE_4685F	GACGTATTCCTGATTATCCAGA	PvMtL_5405R	GACTTAATAGATTGGATAGAAG
	PvMtL_5409R	GACTTAATAGATTGGATAGAAG		
PP10B	PvMtP_4978F	AAGAGAATTATGGAGTGGATGGTG	PvMtS_4979F	AGAGAATTATGGAGTGGATGGTGT
	PvMtP_5966R	TTAACATAATTATAACCTTACGGTCT		

also used for 2 fragment protocol

APPENDIX S2.2

Table S2.2

Sample origins and sample sizes of *P. vivax* across the Eastern Hemisphere. Italic & starred samples are new data from this thesis; others were sourced on Genbank from referenced publications. Geographic groupings for different sections of data analyses are given too.

Region	Country	n	Country	Country 2B S&P>1	Country 4 S&P>4	Regional	Hs Comparison	Published	Genbank Accession Numbers
Melanesia	Subtotal	222							
	Papua New Guinea	64	PNG	PNG	PNG	PNG	PNG	Mu <i>et al.</i> 2005, Jungwuthies <i>et al.</i> 2005	AY791631-89; AY598119
	Papua New Guinea* (EastSepik)	48	PNG	PNG	PNG	PNG	PNG	*This publication	N/A
	Papua New Guinea* (Madang)	66	PNG	PNG	PNG	PNG	PNG	*This publication	N/A
	Papua New Guinea* (Sugomaru)	11	PNG	PNG	PNG	PNG	PNG	*This publication	N/A
	Solomon Islands	4	SOI	SOI	SOI	IMEI	IMEI	Mu <i>et al.</i> 2005, Jungwuthies <i>et al.</i> 2005, Miao <i>et al.</i> 2012	AY791564; AY598116-8; XA_SOI1
	Solomon Islands*	17	SOI	SOI	SOI	IMEI	IMEI	*This publication	N/A
	Vanuatu	7	VAN	VAN	VAN	IMEI	IMEI	Mu <i>et al.</i> 2005	AY791561, AY791565-6, AY791568-70, AY791574
	Vanuatu*	4	VAN	VAN	VAN	IMEI	IMEI	*This publication	N/A
	Melanesia Unk.	1	PNG	PNG	PNG	PNG	PNG	Mu <i>et al.</i> 2005	AY791690, AY791692
Island SE Asia	Subtotal	52							
	Indonesia (non-Papua)	46	IDO	IDO	IDO	ISEA	NA	Mu <i>et al.</i> 2005, Jungwuthies <i>et al.</i> 2005, Taylor <i>et al.</i> 2013	AY791562-3, AY791567, AY791572, AY791576-80, AY791582; AY791552; AY598098-108; KC330533-57
	Borneo	4	BOR	BOR	BOR	ISEA	NA	Mu <i>et al.</i> 2005, Cox-Singh <i>et al.</i> 2006	AY791555; DQ396547-49
	Philippines	1	PHI	NA	NA	ISEA	NA	Mu <i>et al.</i> 2005	AY791571
	Timor	1	TIM	NA	NA	ISEA	NA	Mu <i>et al.</i> 2005	AY791573
Mainland SE Asia	Subtotal	122							
	Thailand	44	THAI	THAI	THAI	MSEA	NA	Mu <i>et al.</i> 2005, Jungwuthies <i>et al.</i> 2005	AY791556, AY791558-60, AY791583; AY598035-73
	Cambodia	19	CAB	CAB	CAB	MSEA	CAB	Taylor <i>et al.</i> 2013	KC330514-32
	Cambodia	43	CAB	CAB	CAB	MSEA	CAB	Pearson <i>et al.</i> 2016	Personal
	Vietnam	9	VIE	VIE	VIE	MSEA	NA	Mu <i>et al.</i> 2005, Jungwuthies <i>et al.</i> 2005	AY791581, AY791584; AY598121-7
	Vietnam	7	VIE	VIE	VIE	MSEA	NA	Pearson <i>et al.</i> 2016	Personal
South China	Subtotal	34							
	China (South)	12	CHSTH	CHSTH	CHSTH	CHSTH	CHSTH	Jungwuthies <i>et al.</i> 2005	AY598128-39
	China (Guangxi)	18	CHGX	CHSTH	CHSTH	CHSTH	CHSTH	Mu <i>et al.</i> 2005	AY791588-605
	China (Yunnan)	3	CHYU	CHSTH	CHSTH	CHSTH	CHSTH	Mu <i>et al.</i> 2005	AY791585-7
	China (Hainan)	1	CHHA	CHSTH	CHSTH	CHSTH	CHSTH	Mu <i>et al.</i> 2005	AY791606
Central China	Subtotal	61							
	China (Guizhou)	25	CHGU	CHGU	CHGU	CHGU	CEA	Miao <i>et al.</i> 2012	JQ240370-95
	China (Anhui)	36	CHAN	CHAN	CHAN	CHAN	CHN	Miao <i>et al.</i> 2012	JQ240331-66
NE Asia	Subtotal	35							
	Korea	17	KOR	KOR	KOR	NEA	NEA	Taylor <i>et al.</i> 2013	KC330497-513
	North Korea	1	KOR	KOR	KOR	NEA	NEA	Mu <i>et al.</i> 2005	AY791554
	South Korea	17	KOR	KOR	KOR	NEA	NEA	Wagami <i>et al.</i> 2010, Miao <i>et al.</i> 2012, Rodrigues <i>et al.</i> 2014	AB550270-80; JQ240367-9, JQ240397-8; KF668404
Myanmar	Subtotal	32							
	Myanmar	32	MMR	MMR	MMR	MMR	MMR	Miao <i>et al.</i> 2012	JQ240396-429
South Asia	Subtotal	96							
	Bangladesh	8	BAN	BAN	BAN	BAN	NA	Jungwuthies <i>et al.</i> 2005, Mu <i>et al.</i> 2005	AY598109-15; AY791691
	India	32	IND	IND	IND	SAS	NA	Jungwuthies <i>et al.</i> 2005, Mu <i>et al.</i> 2005, Rodrigues <i>et al.</i> 2014	AY598120; AY791607-22, AY791624, AY791628, AY791630; KF668367, KF668370, KF668379, KF668382, KF668407, KF668412-13, KF668421, KF668424-25, KF668427-29
	Pakistan (Karachi)	34	PAK	PAK	PAK	SAS	PAK	Taylor <i>et al.</i> 2013	KC330406-439
	Pakistan (Sindh)	20	PAK	PAK	PAK	SAS	PAK	Taylor <i>et al.</i> 2013	KC330386-405
	Sri Lanka	1	SLA	NA	NA	SAS	NA	Mu <i>et al.</i> 2005	AY791557
	South Asia Unknown	1	NA	NA	NA	SAS	NA	Rodrigues <i>et al.</i> 2014	KF668361, KF668426
Central Asia	Subtotal	11							
	Uzbekistan*	9	UZB	UZB	UZB	CAS	CAS	*This publication	
	Afghanistan	2	AFG	AFG	NA	CAS	NA	Rodrigues <i>et al.</i> 2014	KF668381, KF668399
Western Asia	Subtotal	33							
	Armenia*	4	ARM	WAS	WAS	WAS	WAS	*This publication	
	Azerbaijan*	1	AZB	WAS	WAS	WAS	WAS	*This publication	
	Iran	2	IRA	IRA	NA	WAS	NA	Mu <i>et al.</i> 2005	AY791625-26
	Turkey	26	TUR	WAS	WAS	WAS	WAS	Culleton <i>et al.</i> 2011, Taylor <i>et al.</i> 2013	JN788767-76; KC330481-96
Africa	Subtotal	52							
	Africa Unknown	1	NA	NA	NA	AFR	NA	Rodrigues <i>et al.</i> 2014	KF668380
	Ethiopia	9	ETH	ETH	ETH	AFR	EAFR	Mu <i>et al.</i> 2005, Rodrigues <i>et al.</i> 2014	AY791521-25; KF668392, KF668401, KF668411, KF668423
	Kenya	1	KEN	EAFR	NA	AFR	EAFR	Rodrigues <i>et al.</i> 2014	KF668389
	Mauritania	2	MAU	WAFR	NA	AFR	WAFR	Mu <i>et al.</i> 2005	AY791526-27
	Namibia	1	NAM	NA	NA	AFR	NA	Mu <i>et al.</i> 2005	AY791528
	Nigeria	1	NIG	WAFR	NA	AFR	WAFR	Culleton <i>et al.</i> 2011	JN788763
	Rwanda	1	RWA	EAFR	NA	AFR	EAFR	Culleton <i>et al.</i> 2011	JN788764
	Sao Tomé	1	STO	WAFR	NA	AFR	WAFR	Culleton <i>et al.</i> 2011	JN788765
	Sudan	1	SUD	SUD	SUD	AFR	NA	Rodrigues <i>et al.</i> 2014	KF668362
	Sudan*	33	SUD	SUD	SUD	AFR	NA	*This publication	
	Tanzania	1	TAN	EAFR	NA	AFR	EAFR	Culleton <i>et al.</i> 2011	JN788766
Madagascar	Subtotal	66							
	Madagascar	66	MAD	MAD	MAD	MAD	MAD	Culleton <i>et al.</i> 2011, Taylor <i>et al.</i> 2013	JN788738-62; KC330465-80, KC330440-64
	TOTAL	816							

APPENDIX S2.3

Table S2.3

BEAST2 run information for phylogenetic tree.

Panel	Variable	Subvariable	Value/Type
Iterations			1 920 600 000
Runs Combined (LogCombiner)			3
Burn In			10%
Tip Dates			None
Site Model	SubRate		1 (est)
	GammaCatCount		12
	Shape (alpha)		1.0 (est)
	Proplnv		0
	Kappa		2 (est)
ClockModel	Model		Strict Clock
	Clock.Rate		4.40E-09
	Clock.Rate	LowerLimit	4.40E-11
	Clock.Rate	UpperLimit	4.40E-07
	Clock.Rate	Estimate	Yes
Initialization	bPopSizes	Upper	1.00E+12
	bPop&bGroup	Dimensions	9
Priors	Tree.t		CBS
	Tree.t	StartTree	RandomTree
	Tree.t	ConditionalRoot	no
	Tree.t	Type	unscaled
	MarkovChainPopSizes.t	Distribution	Jeffreys
	GammaShape		LogNormal
	GammaShape	Mean	5
	GammaShape	S	1
	GammaShape	Offset	0
MCMC	ChainLength		750000000
	StoreEvery		100000
	PreBurnin		100000
Delta	SampleFromPrior		Yes
RandomSeedNr			Yes

APPENDIX S2.4

Table S2.4

Overview of human population origins and sites, and divisions for data processing.

Nationality	Geo	Sample#	GeoGroup	GeoNr	Latitude	Longitude	Pv Pop Assigned
Orcadian	EUROPE	15	EUR	1	59	-3	OTH
Adygei	EUROPE	17	EUR	1	44	39	WAS1
Russian	EUROPE	25	NRUS	2	61	40	OTH
Basque	EUROPE	24	EUR	1	43	0	OTH
French	EUROPE	28	EUR	1	46	2	OTH
Italian	EUROPE	13	EUR	1	46	10	OTH
Sardinian	EUROPE	28	EUR	1	40	9	OTH
Tuscan	EUROPE	8	EUR	1	43	11	OTH
Mozabite	MIDDLE_EAST	29	NAF	3	32	3	OTH
Bedouin	MIDDLE_EAST	47	MDE	4	31	35	WAS2
Druze	MIDDLE_EAST	44	MDE	4	32	35	WAS2
Palestinian	MIDDLE_EAST	50	MDE	4	32	35	WAS2
Balochi	CENTRAL_SOUTH_ASIA	24	CSA	5	30.5	66.5	PAK
Brahui	CENTRAL_SOUTH_ASIA	25	CSA	5	30.5	66.5	PAK
Burusho	CENTRAL_SOUTH_ASIA	25	CSA	5	36.5	74	PAK
Hazara	CENTRAL_SOUTH_ASIA	23	CSA	5	33.5	70	PAK
Kalash	CENTRAL_SOUTH_ASIA	24	CSA	5	36	71.5	PAK
Makrani	CENTRAL_SOUTH_ASIA	25	CSA	5	26	64	PAK
Pathan	CENTRAL_SOUTH_ASIA	24	CSA	5	33.5	70.5	PAK
Sindhi	CENTRAL_SOUTH_ASIA	24	CSA	5	25.5	69	PAK
Melanesian	OCEANIA	13	MEL	6	-6	155	IMEL
Papuan	OCEANIA	17	MEL	6	-4	143	PNG
BantuKenya	AFRICA	11	EAF	7	-3	37	EAFR
MbutiPygmy	AFRICA	13	EAF	7	1	29	EAFR
BantuSouthWest	AFRICA	3	SAF	9	-21	18.7	OTH
BantuSouthEast	AFRICA	5	SAF	9	-28.4	27.6	OTH
Mandenka	AFRICA	24	WAF	8	12	-12	WAFR
Yoruba	AFRICA	22	WAF	8	8	5	WAFR
BiakaPygmy	AFRICA	27	WAF	8	4	17	OTH
San	AFRICA	6	SAF	9	-21	20	OTH
Han	EAST_ASIA	34	CCH	10	32.5	114	CHN1
Han-NChina	EAST_ASIA	10	CCH	10	39	114	CHN1
Dai	EAST_ASIA	10	SCH	11	21	100	MMR
Lahu	EAST_ASIA	8	SCH	11	22	100	MMR
Daur	EAST_ASIA	10	NEA	12	48.5	124	NEA1
Hezhen	EAST_ASIA	9	NEA	12	47.5	133.5	NEA1
Oroqen	EAST_ASIA	9	NEA	12	50.5	126.5	NEA1
She	EAST_ASIA	10	CCH	10	27	119	CHN2
Miao	EAST_ASIA	10	CCH	10	28	109	CEA
Tujia	EAST_ASIA	10	CCH	10	29	109	CEA
Yi	EAST_ASIA	10	SCH	11	28	103	CHSTH2
Mongola	EAST_ASIA	10	NEA	12	48.5	119	NEA1
Naxi	EAST_ASIA	9	SCH	11	26	100	CHSTH1
Xibo	EAST_ASIA	9	WCH	13	43.5	81.5	CAS
Uygur	EAST_ASIA	10	WCH	13	44	81	CAS
Cambodian	EAST_ASIA	10	MSEA	14	12	105	CAB
Japanese	EAST_ASIA	29	EAS	15	38	138	NEA2
Tu	EAST_ASIA	10	CCH	10	36	101	OTH
Yakut	EAST_ASIA	25	NEA	12	63	129.5	OTH
TundraNentsi	EAST_ASIA	14	NRUS	2	67	65	OTH

APPENDIX S2.5

Table S2.5

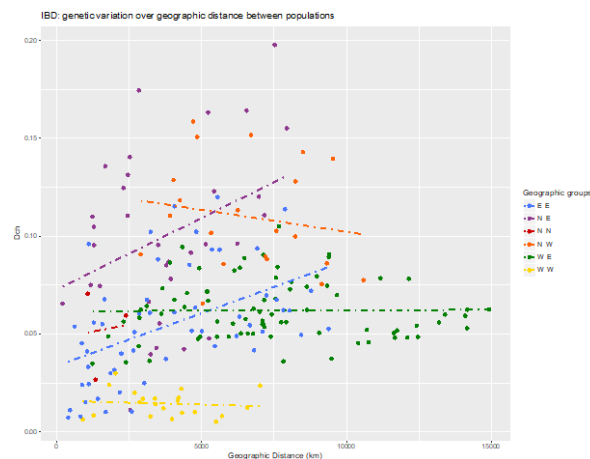
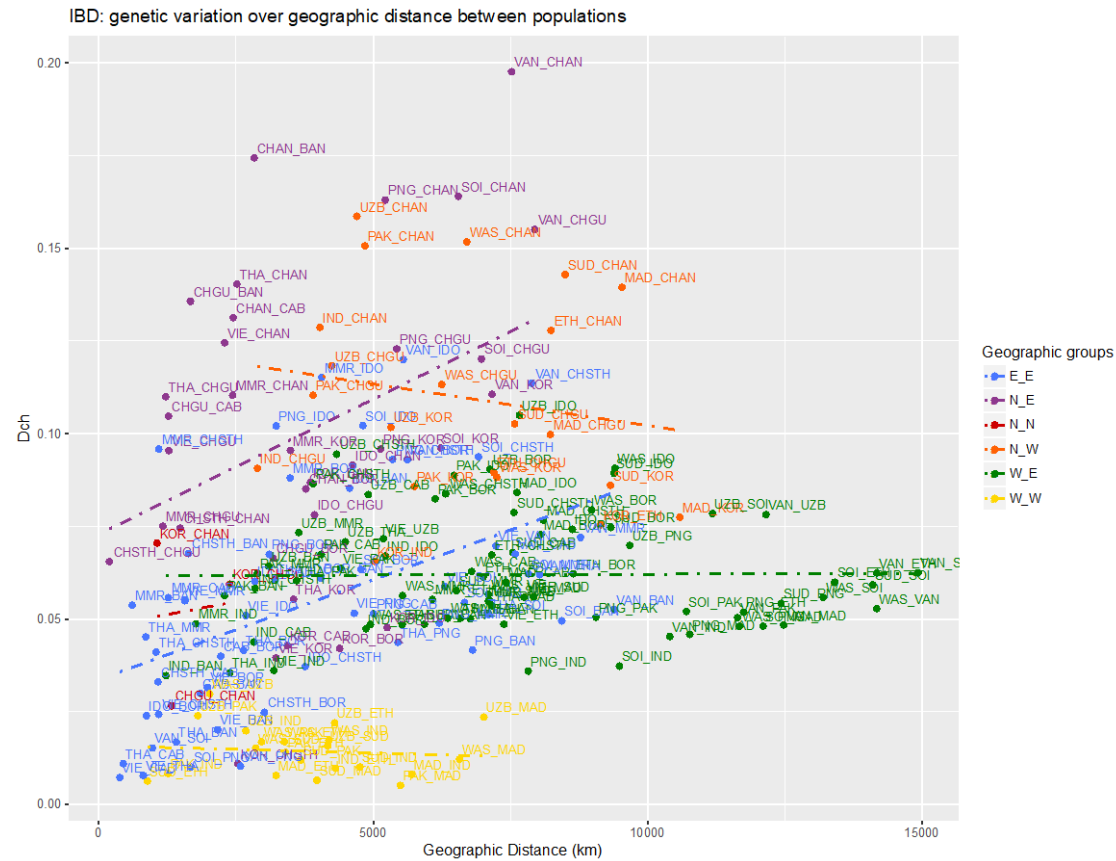
Table of *P. vivax* population categories, region of origin, number of samples, coordinates, and geographic grouping. There are three geographic divisions: West of Bangladesh (W), East of India (E), and North of Southern China (N). The matrix shows geographic distances between populations in kilometers (km) on the upper half, and population genetic differentiation (D_{CH}) on the bottom half.

Population	Region	n	Longitude	Latitude	GeoGroup	BAN	BOR	CAB	CHAN	CHGU	CHSTH	ETH	IDO	IND	KOR	MAD	MMR	PAK	PNG	SOI	SUD	THA	UZB	VAN	VIE	WAS
BAN	Bangladesh	8	90.36	23.68	E		4045	1845	2828	1638	1634	5528	4571	1222	3854	6771	607	2790	6808	8434	5931	1411	3094	9382	2165	4945
BOR	Borneo	4	117.87	-1.00	E	0.0394		2220	3771	3172	3014	8640	868	4866	4386	8053	3485	6128	3106	4774	9329	2637	7126	5617	1979	8976
CAB	Cambodia	62	104.75	14.16	E	0.0091	0.0219		2446	1269	1076	7001	2840	2825	3435	7415	1267	4050	5004	6659	7543	446	4907	7586	380	6790
CHAN	China Anhui	36	117.39	32.92	N	0.1297	0.0526	0.1025		1326	1482	8238	4623	4027	1059	9530	2436	4847	5218	6550	8495	2513	4700	7518	2291	6703
CHGU	China Guizhou	26	106.75	25.42	N	0.0981	0.0332	0.0775	0.0189		193	7196	3928	2889	2382	8224	1166	3901	5427	6970	7574	1219	4245	7940	1274	6245
CHSTH	China South	34	106.42	23.71	E	0.0457	0.0020	0.0269	0.0519	0.0456		7157	3757	2844	2531	8097	1090	3900	5351	6916	7562	1040	4330	7882	1091	6318
ETH	Ethiopia	9	40.49	9.15	W	0.0354	0.0388	0.0379	0.0990	0.0661	0.0448		8627	4313	9145	3279	6077	3400	11746	13405	885	6590	4294	14171	7381	3379
IDO	Indonesia	47	116.32	-8.65	E	0.0688	0.0060	0.0516	0.0680	0.0524	0.0214	0.0600		5220	5254	7617	4060	6473	3229	4810	9386	3207	7662	5546	2677	9404
IND	India	32	78.96	20.59	W	0.0277	0.0378	0.0330	0.1013	0.0691	0.0446	0.0004	0.0610		5029	5699	1766	1263	7821	9482	4757	2383	2678	10401	3190	4191
KOR	Korea	35	127.68	37.39	N	0.0502	0.0072	0.0305	0.0481	0.0409	0.0023	0.0446	0.0273	0.0446		10585	3491	5746	5135	6228	9319	3548	5320	7161	3226	7249
MAD	Madagascar	66	46.17	-19.35	W	0.0361	0.0466	0.0396	0.1058	0.0732	0.0515	0.0019	0.0668	0.0012	0.0494		7095	5495	10764	12100	3972	7160	7015	12476	7745	6572
MMR	Myanmar	32	95.96	21.91	E	0.0355	0.0638	0.0393	0.0941	0.0626	0.0745	0.0373	0.0944	0.0340	0.0704	0.0383		2892	6202	7828	6513	856	3641	8775	1567	5532
PAK	Pakistan	54	67.77	25.38	W	0.0382	0.0487	0.0413	0.1084	0.0760	0.0542	0.0044	0.0679	0.0077	0.0512	0.0013	0.0407		9054	10703	3678	3604	1803	11636	4404	2964
PNG	Papua New Guinea	190	145.45	-5.72	E	0.0314	0.0564	0.0390	0.1288	0.0983	0.0764	0.0368	0.0909	0.0316	0.0750	0.0346	0.0369	0.0353		1678	12426	5450	9672	2582	4654	11669
SOI	Solomon Islands	21	160.16	-9.65	E	0.0333	0.0512	0.0371	0.1239	0.0921	0.0731	0.0385	0.0879	0.0325	0.0697	0.0342	0.0400	0.0335	0.0024		14101	7103	11177	975	6300	13195
SUD	Sudan	34	34.43	14.45	W	0.0353	0.0463	0.0410	0.1061	0.0732	0.0511	0.0010	0.0680	0.0008	0.0509	0.0018	0.0379	0.0044	0.0375	0.0392		7114	4171	14924	7919	2857
THA	Thailand	44	100.99	15.87	E	0.0027	0.0287	0.0030	0.1146	0.0864	0.0349	0.0330	0.0552	0.0278	0.0398	0.0356	0.0345	0.0379	0.0352	0.0368	0.0349		4491	8032	808	6354
UZB	Uzbekistan	9	64.59	41.38	W	0.0473	0.0557	0.0515	0.1162	0.0832	0.0613	0.0141	0.0790	0.0108	0.0603	0.0151	0.0495	0.0157	0.0491	0.0506	0.0122	0.0460		12152	5183	2019
VAN	Vanuatu	11	166.96	-15.38	E	0.0435	0.0687	0.0465	0.1436	0.1119	0.0865	0.0484	0.0996	0.0413	0.0799	0.0404	0.0510	0.0383	0.0049	0.0020	0.0486	0.0463	0.0599		7234	14170
VIE	Vietnam	17	108.28	14.06	E	0.0033	0.0165	0.0014	0.1012	0.0748	0.0226	0.0301	0.0423	0.0260	0.0281	0.0343	0.0380	0.0367	0.0372	0.0382	0.0335	0.0003	0.0444	0.0485		7099
WAS	Western Asia	31	40.78	39.53	W	0.0305	0.0433	0.0355	0.1083	0.0759	0.0485	0.0047	0.0650	0.0037	0.0477	0.0040	0.0386	0.0058	0.0356	0.0361	0.0049	0.0304	0.0172	0.0423	0.0293	

APPENDIX S2.6

Figure S2.6a

Correlations between geographic distance and population differentiation, divided into different cross-geographic groupings depending on the geographic location of the sampled population. East (E) denotes anything to the East from and including Bangladesh, except for the three North (N) populations to the north of southern China. West (W) denotes anything to the west from India, inclusive. Each data point is coloured according to the location of the two populations being compared (e.g. an population in the East compared to one in the west is green, east compared with east is blue, etc.



GeoGroup	n	r	p
E_E	55	0.48	0.000
N_E	33	0.35	0.047
N_N	3	na	na
N_W	21	-0.18	0.447
W_E	77	0.01	0.932
W_W	21	-0.09	0.687
Overall	210	0.17	0.013

Figure 2.6b & Table 2.6 Geographic distance versus genetic differentiation (D_{ch})

Figure 2.6b: Correlations between geographic distance and population differentiation, divided into different cross-geographic groupings depending on the geographic location of the sampled population, as shown in Figure 7. 'East' denotes anything to the east of and including Bangladesh, except for the three 'North' populations to the north of southern China. 'West' denotes anything to the west from India, inclusive. Each data point is coloured according to the location of the two populations being compared (e.g. as given in legend). Correlation slopes (l_m) for each group are included.

Table 2.6: Group size (n), Pearson correlation coefficients (r) and p-value (p) comparing within and between each geographical group (GeoGroup) in Figure 9.

Supplementary Table S2.7. List of nonsynonymous SNPs identified in Pvs47 by Hupalo *et al.* 2016, Tachibana *et al.* 2015, and Woo *et al.* 2013. The reference allele frequency / alternative allele frequency is indicated (ref/alt), while "ref in all" indicates no observed SNPs at this location in this population.

Gene name	Mutation	Madagascar		India		India		Myanmar		Cambodia		Thailand		Thailand		Korea		Indonesia		Papua New Guinea		Vanuatu		(Country)
		Hupalo	Hupalo	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	Hupalo	Tachibana	
Pvs47		1																						
	Phe22Leu	0/1	0/2	0/2	0/3	0/8	0/3	0/1	0/21	0/1	0/21	1/9	0/40	0/3	0/15	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9
	Phe24Leu	0/1	0/2	0/2	0/3	0/8	0/3	0/1	0/21	0/1	0/21	0/10	0/40	0/3	0/15	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9
	Lys27Glu	0/1	0/2	0/2	0/3	0/7	0/3	0/1	0/21	0/1	0/21	0/10	0/40	0/3	0/15	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9
	Leu29Ile	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Asp31Asn	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	39/1	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Ser57Thr	0/1	1/1	0/3	0/3	ref in all	0/3	1/1	12/5	3/7	9/1	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Ser62Asn	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Leu82Val	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Asp156Gly	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Thr193Ile	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Val230Ile	ref in all	2/1	0/3	0/3	2/5	0/3	0/1	1/16	0/2	0/19	5/5	1/39	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
	Met233Ile	0/1	0/3	0/3	0/3	0/7	0/3	0/2	0/19	0/2	0/19	0/10	1/39	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
	Phe237Ile	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Glu240Asp	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Ile262Thr	ref in all	0/1	0/3	0/3	2/2	0/3	ref in all	4/1	ref in all	4/1	8/2	0/40	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
	Ile262Lys	ref in all	ref in all	ref in all	ref in all	5/2	0/3	0/1	10/4	0/1	10/4	5/5	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Ile273Val	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all	ref in all
	Ile273Met	ref in all	ref in all	ref in all	ref in all	6/1	0/3	0/1	9/8	0/1	9/8	6/4	8/32	2/1	1/4	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9
	Ala373Val	ref in all	3/1	ref in all	ref in all	6/1	0/3	0/1	9/8	0/1	9/8	6/4	8/32	2/1	1/4	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9

Figures S2.8a & S2.8b, Table S2.8

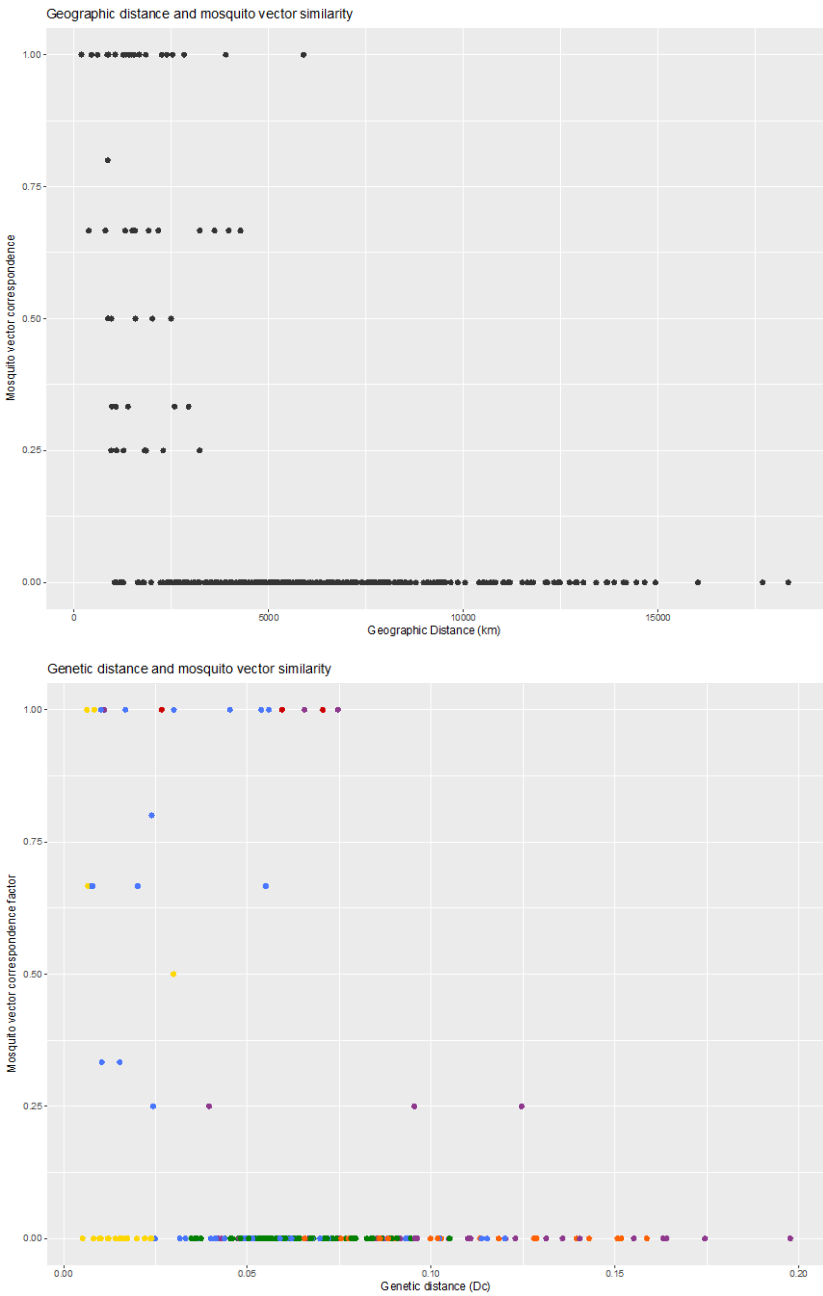


Figure S2.8.
a) Geographic distance and vector correspondence between sites.
Shows the correlation between vector species overlap and distance.

b) Genetic differentiation and vector correspondence between sites.
The figure shows the correlation between population differentiation (D_{CH}) and vector species overlap. Points are coloured according to cross-geographic categories, as shown in the Geography section (see Fig. 7).

Table S2.8: Correlation coefficients for Figs. 2.8
Sample size, Spearman signed rank correlation coefficients and p-values for the correlations between vector similarity with geographic distance (Vec_Geo) and with population genetic differentiation.

GeoGroup	n	rs	p
Vec_Geo	300	-0.54	0.000
Vec_Gen	210	-0.35	0.000

APPENDIX S2.9

Table S2.9

Table of linear model F-statistic values for Pv mitochondrial DNA population genetic variation (PvDch) compared with: geography (Geo) & vector (Vec) for all Pv populations, and human genetic variation (HsDch) geography (Geo) & vector (Vec) for Pv populations corresponding to Hs populations.

***, **, *, . denote levels of significance

ss<10 indicates where sample size is smaller than 10

Dch Pop Category	ImData	F-statistic	Independant Variables	DF-used	DF	p-value	
PvMt	ImPv4	12.750	Geo, Vec	2	207	0.000	***
	ImxPv4	8.927	Geo, Vec, GeoVec	3	206	0.000	***
PvMtSubsets	ImPv4WW	2.753	Geo, Vec	2	18	0.091	.
	ImPv4EE	12.250	Geo, Vec	2	52	0.000	***
	ImPv4EENN	11.490	Geo, Vec	2	88	0.000	***
	ImxPv4EE	8.143	Geo, Vec, GeoVec	3	51	0.000	***
	ImxPv4EENN	7.689	Geo, Vec, GeoVec	3	87	0.000	***
PvHs	ImPvHs	3.580	Hs, Geo, Vec	3	74	0.018	**
	ImxPvHs	2.000	Hs, Geo, Vec, HsGeo, HsVec, GeoVec, HsGeoVec	7	70	0.067	.
	ImxPvHs -a:b:c	2.240	Hs, Geo, Vec, HsGeo, HsVec, GeoVec	6	71	0.049	*
PvHsSubsets	ImPvHsWE	2.313	Hs, Geo (all Vec 0)	2	10	0.150	
	ImPvHsWW	1.382	Hs, Geo, Vec	3	6	0.336	ss<10
	ImPvHsEE	8.115	Hs, Geo, Vec	3	2	0.112	ss<10
	ImPvHsWENW	4.808	Hs, Geo (all Vec 0)	2	15	0.024	**
	ImPvHsEENENN	1.941	Hs, Geo, Vec	3	14	0.169	

APPENDIX S2.10

Table S2.10

Table of linear model analyses of variance (ANOVA) values for Pv mitochondrial DNA population genetic variation (PvDch) compared with: geography (Geo) & vector (Vec) for all Pv populations (PvAll), and human genetic variation (HsDch) geography (Geo) & vector (Vec) for Pv populations corresponding to Hs populations (PvHs). Analyses of geographic regional subsets are included (West and East of Bangladesh).

***, **, *, . denote levels of significance

PvPop

ImPv***

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Geo	1	0.008	0.008	6.820	0.010 **
Vec	1	0.023	0.023	18.684	0.000 ***
Residuals	207	0.253	0.001		

ImxPv***

Im (.x.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Geo	1	0.008	0.008	6.828	0.010 **
Vec	1	0.023	0.023	18.706	0.000 ***
Geo:Vec	1	0.002	0.002	1.246	0.266
Residuals	206	0.252	0.001		

ImPvWW

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Geo	1	0.000	0.000	0.205	0.656
Vec	1	0.000	0.000	5.301	0.033 *
Residuals	18	0.001	0.000		

ImPvEE***

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Geo	1	0.010	0.010	17.559	0.000 ***
Vec	1	0.004	0.004	6.942	0.011 *
Residuals	52	0.031	0.001		

ImPvEENN***

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Geo	1	0.017	0.017	12.419	0.001 ***
Vec	1	0.014	0.014	10.562	0.002 **
Residuals	88	0.119	0.001		

ImxPvEE***

Im (.x.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Geo	1	0.010	0.010	17.313	0.000 ***
Vec	1	0.004	0.004	6.845	0.012 *
Geo:Vec	1	0.000	0.000	0.270	0.605
Residuals	51	0.031	0.001		

ImxPvEENN***

Im (.x.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Geo	1	0.017	0.017	12.317	0.001 ***
Vec	1	0.014	0.014	10.475	0.002 **
Geo:Vec	1	0.000	0.000	0.276	0.601
Residuals	87	0.119	0.001		

PvHsPop

ImHsPv**

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.003	0.003	1.892	0.173
Geo	1	0.001	0.001	0.364	0.548
Vec	1	0.012	0.012	8.485	0.005 **
Residuals	74	0.103	0.001		

ImxHsPv

Im (.x.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.003	0.003	1.875	0.175
Geo	1	0.001	0.001	0.361	0.550
Vec	1	0.012	0.012	8.411	0.005 **
HsDch:Geo	1	0.000	0.000	0.022	0.883
HsDch:Vec	1	0.000	0.000	0.333	0.566
Geo:Vec	1	0.003	0.003	2.367	0.128
HsDch:Geo:Vec	1	0.001	0.001	0.628	0.431
Residuals	70	0.099	0.001		

ImxHsPv -a:b:c*

Im (.x. -a:b:c)	Df	Sum	Mean	F	Pr(>F)
HsDch	1	0.003	0.003	1.885	0.174
Geo	1	0.001	0.001	0.363	0.549
Vec	1	0.012	0.012	8.455	0.005 **
HsDch:Geo	1	0.000	0.000	0.022	0.883
HsDch:Vec	1	0.000	0.000	0.335	0.565
Geo:Vec	1	0.003	0.003	2.379	0.127
Residuals	71	0.100	0.001		

ImHsPvWE

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.000	0.000	3.453	0.093 .
Geo	1	0.000	0.000	1.172	0.304
Residuals	10	0.001	0.000		

ImHsPvWW

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.000	0.000	0.049	0.832
Geo	1	0.000	0.000	2.872	0.141
Vec	1	0.000	0.000	1.224	0.311
Residuals	6	0.000	0.000		

ImHsPvEE

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.000	0.000	0.001	0.983
Geo	1	0.001	0.001	14.124	0.064 .
Vec	1	0.001	0.001	10.221	0.085 .
Residuals	2	0.000	0.000		

ImHsPvNE

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.000	0.000	0.015	0.910
Geo	1	0.000	0.000	0.112	0.761
Residuals	3	0.007	0.002		

ImHsPvNW

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.001	0.001	0.464	0.566
Geo	1	0.001	0.001	0.718	0.486
Residuals	2	0.002	0.001		

ImHsPvWENW**

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.005	0.005	9.603	0.007 **
Geo	1	0.000	0.000	0.013	0.911
Residuals	15	0.009	0.001		

ImHsPvEENN

Im (.+.)	Df	SumSq	MeanSq	Fvalue	Pr(>F)
HsDch	1	0.000	0.000	0.334	0.573
Geo	1	0.001	0.001	0.725	0.409
Vec	1	0.004	0.004	4.765	0.047 *
Residuals	14	0.012	0.001		

APPENDIX S2.11

Table S2.11

Table of linear model analyses of variance coefficients for Pv mitochondrial DNA population genetic variation (PvDch) compared with: geography (Geo) & vector (Vec) for all Pv populations (PvAll), and human genetic variation (HsDch) geography (Geo) & vector (Vec) for Pv populations corresponding to Hs populations (PvHs). Analyses of geographic regional subsets are included (West and East of Bangladesh).

***, **, *, . denote levels of significance

PvAll

ImPv***

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.06874	0.006	12.138	<2.00E-16	***
Geo	2.921E-07	0.000	0.346	0.730	
Vec	-0.04156	0.010	-4.322	0.000	***

ImxPv***

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.06846	0.006	12.086	<2e-16	***
Geo	3.528E-07	0.000	0.417	0.677	
Vec	-0.02576	0.017	-1.506	0.134	.
Geo:Vec	-1.116E-05	0.000	-1.116	0.266	

ImPvWW

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.02305	0.005	4.902	0.000	***
Geo	-1.77E-06	0.000	-1.683	0.110	.
Vec	-0.01116	0.005	-2.302	0.033	*

ImPvEE***

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.04968	0.009	5.827	0.000	***
Geo	2.855E-06	0.000	1.782	0.081	.
Vec	-0.02964	0.011	-2.635	0.011	*

ImPvEENN***

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.07108	0.010	7.249	0.000	***
Geo	2.185E-06	0.000	1.119	0.266	
Vec	-0.04107	0.013	-3.250	0.002	**

ImxPvEE***

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.0493	0.009	5.720	0.000	***
Geo	2.946E-06	0.000	1.816	0.075	.
Vec	-0.02042	0.021	-0.971	0.336	
Geo:Vec	-7.625E-06	0.000	-0.520	0.605	

ImxPvEENN***

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.07048	0.010	7.110	0.000	***
Geo	2.332E-06	0.000	1.178	0.242	
Vec	-0.03116	0.023	-1.371	0.174	
Geo:Vec	-7.246E-06	0.000	-0.525	0.601	

PvHs

ImHsPv**

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.06923	0.015	4.484	0.000	***
HsDch	0.1493	0.115	1.298	0.198	
Geo	-2.386E-06	0.000	-1.461	0.148	
Vec	-0.04252	0.015	-2.913	0.005	**

ImxHsPv

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	5.44E-02	3.12E-02	1.741	0.086	
HsDch	2.27E-01	1.90E-01	1.19	0.238	
Geo	5.77E-07	5.92E-06	0.097	0.923	
Vec	-1.19E-01	2.25E-01	-0.531	0.597	
HsDch:Geo	-1.44E-05	2.86E-05	-0.502	0.617	
HsDch:Vec	1.08E+00	1.98E+00	0.544	0.588	
Geo:Vec	5.57E-05	9.63E-05	0.579	0.565	
HsDch:Geo:Vec	-7.18E-04	9.06E-04	-0.793	0.431	

ImxHsPv-abc*

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	5.33E-02	3.11E-02	1.714	0.0908	
HsDch	2.32E-01	1.90E-01	1.221	0.226	
Geo	7.74E-07	5.90E-06	0.131	0.896	
Vec	5.05E-02	6.80E-02	0.743	0.4599	
HsDch:Geo	-1.53E-05	2.85E-05	-0.537	0.5927	
HsDch:Vec	-4.49E-01	4.58E-01	-0.98	0.3305	
Geo:Vec	-1.99E-05	1.29E-05	-1.542	0.1274	

ImHsPvWE

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.07174	0.012	5.839	0.000	***
HsDch	-0.01266	0.100	-0.127	0.902	
Geo	-1.022E-06	0.000	-1.083	0.304	
Vec	NA	NA	NA	NA	(all Vec = 0)

ImHsPvWW

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.02388	0.008	2.888	0.028	*
HsDch	0.04791	0.061	0.780	0.465	
Geo	-2.772E-06	0.000	-1.803	0.121	
Vec	-0.008332	0.008	-1.106	0.311	

ImHsPvEE

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.2062	0.046	4.469	0.047	*
HsDch	-0.6616	0.171	-3.863	0.061	.
Geo	-2.867E-06	0.000	-0.808	0.504	
Vec	-0.07402	0.023	-3.197	0.086	.

ImHsPvNE

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.08959	0.086	1.036	0.376	
HsDch	-0.1288	0.655	-0.197	0.857	
Geo	0.0000359	0.000	0.334	0.760	
Vec	NA	NA	NA	NA	(all Vec = 0)

ImHsPvNW

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.1483	0.049	3.017	0.095	.
HsDch	0.1758	0.632	0.278	0.807	
Geo	-1.052E-05	0.000	-0.847	0.486	
Vec	NA	NA	NA	NA	(all Vec = 0)

ImHsPvWENN***

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.1305	0.021	6.178	0.000	***
HsDch	-0.3258	0.200	-1.627	0.125	
Geo	-2.664E-07	0.000	-0.114	0.911	
Vec	NA	NA	NA	NA	(all Vec = 0)

ImHsPvEENN

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
{Intercept}	0.1184	0.031	3.803	0.002	**
HsDch	-0.2496	0.214	-1.164	0.264	
Geo	-1.49E-06	0.000	-0.345	0.735	
Vec	-0.03927	0.018	-2.183	0.047	*

Figure S2.12



Figure S2.12 Genetic differentiation in human versus *P. vivax* populations

Genetic differentiation values (D_{ch}) between human populations compared to those between *P. vivax* populations. The abbreviation of the populations compared is given in each data point. Data points are coloured according to cross-regional comparison, with regions divided similar to those in Figures 7 & 9: east, west and north, coloured yellow, blue and red respectively. The dotted line gives the lm correlation between the two datasets.

Table S2.13

Overview of populations used for the *P. vivax* (Pv) and human (Hs) comparison, listing the names given to the populations when comparing both ("Combined Populations"), the names of the most similar *Pv* populations often used in this chapter ("Pv Pop Name"), the regions making up these populations ("Regions"), the number of samples ("Pv N") and all the same for the human populations.

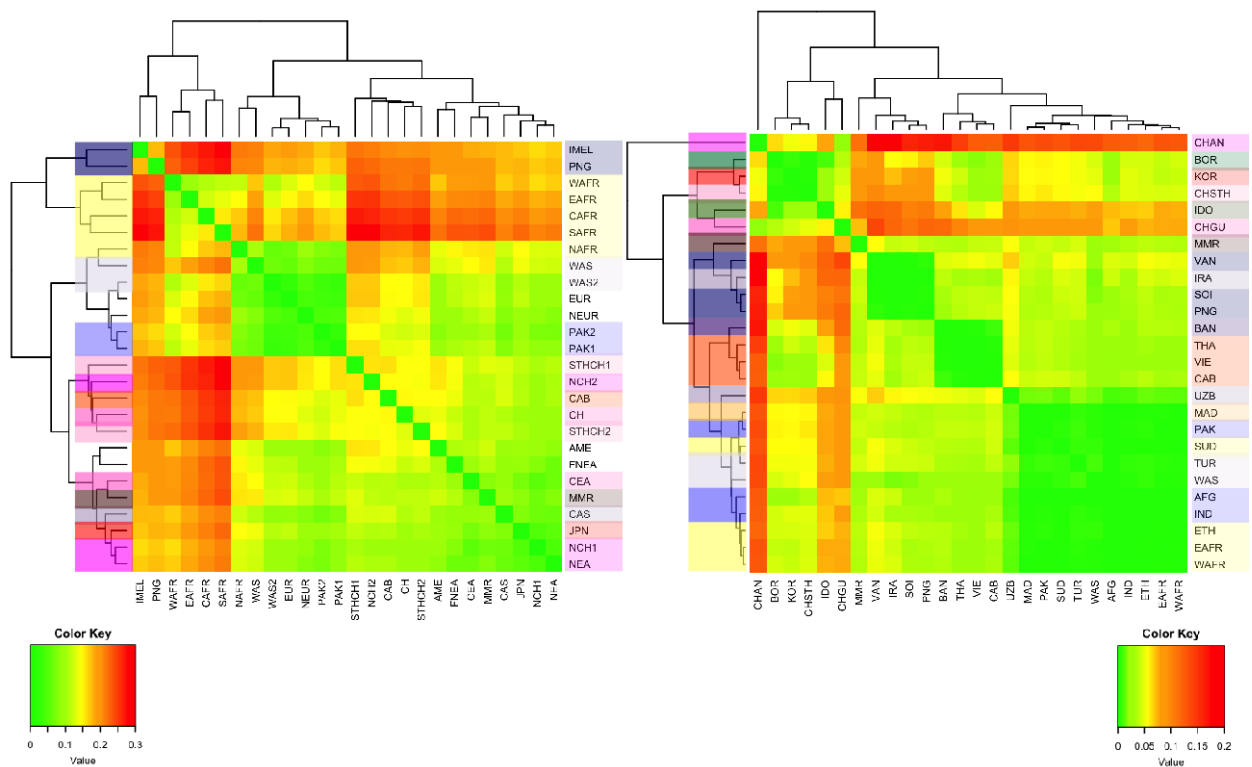
Combined Populations	Pv Pop Name	Regions	Pv N	Hs Pop Name	Ethnicity	Hs N
CAB	CAB	Cambodia	62	CAB	Cambodian	10
CAS	CAS	Uzbekistan	9	CAS	Yi, Uygur	19
CEA	CEA	China (Guizhou)	26	CEA	Miao, Tujia	20
CHN1		N/A		NCH1	Han, Han-NChina	44
CHN2	CHN	China (Anhui)	36	NCH2	She	10
CHSTH1	CHSTH	China (Guangxi, Hainan, Southern, Yunnan)	34	STHCH1	Naxi	9
CHSTH2		N/A		STHCH2	Yi	10
EAFR	EAFR	Ethiopia, Kenya, Rwanda, Tanzania	12	EAFR	BantuKenya, MbutiPygmy	24
IMEL	IMEL	Solomon Islands, Vanuatu	32	IMEL	Melanesian	13
MMR	MMR	Myanmar	32	MMR	Dai, Lahu	18
NEA1				NEA	Daur, Hezhen, Oroqen, Mongola	38
NEA2	NEA	All Korea	35	JPN	Japanese	29
PAK	PAK	All Pakistan	54	PAK	Balochi, Brahui, Burusho, Hazara, Kalash, Makrani, Pathan, Sindhi	194
PNG	PNG	All PNG	190	PNG	Papuan	17
WAFR	WAFR	Mauritania, Nigeria, Sao Tome	4	WAFR	Mandenka, Yoruba	46
WAS	WAS	Armenia, Azerbaijan, Turkey	31	WAS	Adygei	17

APPENDIX S2.14

Figure S2.14a

Heatmap of D_{ch} values of human (Hs) populations on the left and *P. vivax* (Pv) populations on the right.

The interpopulation diversity for the original human population data is given as a heatmap of the D_{ch} values between populations in Appendix S2.14, together with the *Pv* D_{ch} heatmap. There are some notable differences and similarities between the two. The first is that though for humans the Melanesian clade falls into a category on its own, splitting directly from African clades, for *Pv* the Melanesian clade branches off the main clade and is closely connected to Myanmar with some close relations to mainland Southeast Asia. Instead, in *Pv*, the more separated clades are those from Eastern Asia (China, Korea) and island Southeast Asia. Similarities between the two include that there are two to three distinct groupings in the eastern Asian populations, with the southern Chinese populations grouping with at least one North Asian population and the Central East Asian populations (CEA in humans and CHGU in *Pv*) are not grouped with southern China. There is also a strong grouping of West, Central and South Asian populations in both datasets, which includes all African populations in *Pv*, and contains one African population (NAFR) in humans. Island Southeast Asia cannot be compared as there was no data for ISEA human populations. Mainland Southeast Asia only has one population representing the human data (CAB), which falls within the East Asian clade that also contains southern China. For *Pv* mainland Southeast Asia is grouped only with Bangladesh, and forms a clade between the African/Asian branch and Melanesia. Taken overall, there is similarity between the general structure of both datasets.



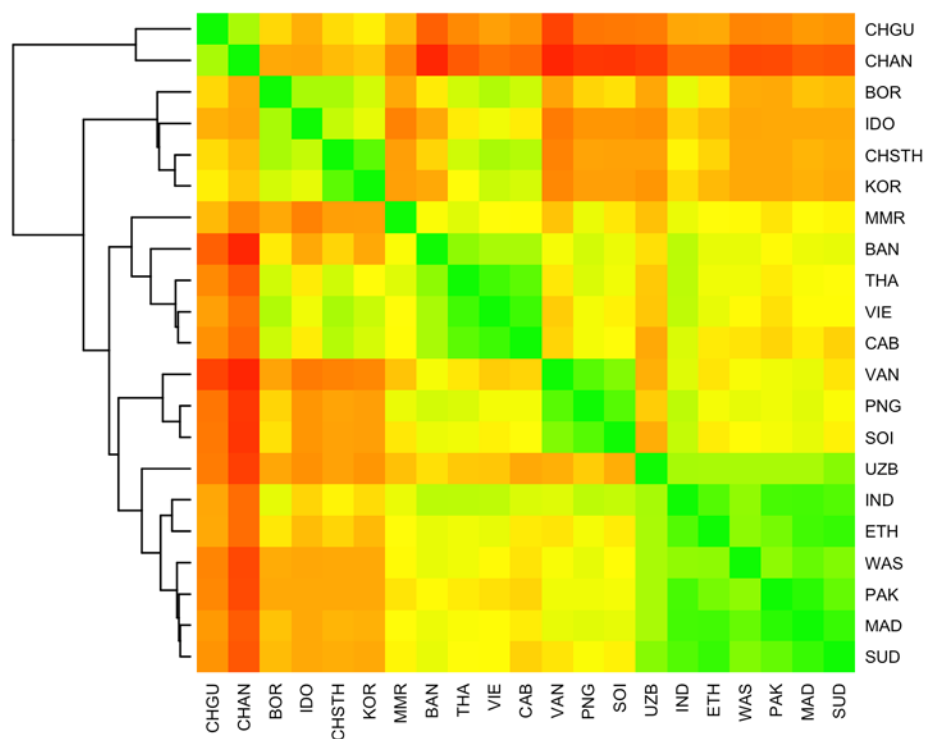


Figure S2.14b. *P. vivax* population differentiation

Heatmap and dendrogram of D_{CH} interpopulation differentiation values. Populations used as in Table 2, and Appendix S2.2 column 'Country 4 S#>4'.

Figure S2. 15

Overview of data and references used for date estimates in Figure 17.

Total	Lineage/Location	Species	Youngest Age (kyl)	Mean/Median	Oldest Age (kyl)	Reference
African-Non-African Split - Gen Human (Lin M)	African to Non-African (L3 - M)	Human	53000	56000	59000	Atkinson et al. 2009
	Hap M	Human	41300	46600	51900	Joann et al. 2012 (Soares et al. 2009)
	Hap M	Human	50000	79000	98700	Joann et al. 2012
	M (Eurasian)	Human	57700	64800	71900	Melamar et al. 2003
African-Non-African Split - Gen Human (Lin N)	African to Non-African (L3 - M)	Human	43300	61600	98700	Atkinson et al. 2009
	Hap N	Human	50000	57000	64000	Joann et al. 2012 (Soares et al. 2009)
	Hap N	Human	39800	42400	48000	Joann et al. 2012
	M (Eurasian)	Human	59600	78500	99200	Melamar et al. 2003
Africa - Gen HBV	Africa A1	HBV	38800	61200	99200	Paraskevisev et al. 2012
	Africa A2	HBV	4330	7900	12000	Paraskevisev et al. 2012
	Africa A3	HBV	3400	5900	8800	Paraskevisev et al. 2012
	Africa A4	HBV	2400	5100	7900	Paraskevisev et al. 2012
West Africa - Gen HBV	Africa A1	HBV	100	1300	3300	Paraskevisev et al. 2012
	Africa A2	HBV	100	5050	12000	Paraskevisev et al. 2012
	Africa A3	HBV	1800	4400	7300	Paraskevisev et al. 2012
	Africa A4	HBV	45000	45000	45000	Slafkin & Radino 2016 (Gallego Llorente et al. 2015)
Ethiopia - Arch Human	Ethiopia: Mota Cave, South East	Human	90000	105000	120000	Bae et al. 2017 (R. Grün & Stringer 1991, R. Grün et al. 2005)
	Israel: Skhul, Qafzeh	Human	75000	75000	75000	Bae et al. 2017 (Petraglia et al. 2011)
	Saudi Arabia: Jebel Qafzar	Human activity	80000	90000	100000	Bae et al. 2017 (Goren et al. 2015)
	Saudi Arabia: Mundadan	Human activity	105000	105000	105000	Bae et al. 2017 (Bose et al. 2011)
Arab Peninsula - Arch Human	Oman: Aghail Al Anwal	Human activity	175000	175000	175000	Bae et al. 2017 (Amirag et al. 2011)
	United Arab Emirates: Jebel Faya C	Human activity	75000	98750	125000	Bae et al. 2017 (Blaizot & Petraglia 2017)
	India: Koldi, 18R Dune	Human activity	80000	88000	96000	Bae et al. 2017 (M. Petraglia et al. 2007)
	India: Jawaharapuram	Human activity	75000	84000	96000	Joann et al. 2012 (Soares et al. 2009)
Thailand - Gen Human	Hap M21a (Orang Asli, Neolithic W Malaysia, Sabah, Thailand)	Human	27600	25500	29400	Joann et al. 2012
	Hap M21a (Orang Asli, Neolithic W Malaysia, Sabah, Thailand)	Human	16900	31100	46600	Joann et al. 2012
	Thailand: Mohn Kiew	Human	25600	25800	26000	Morley 2017 (Pookajorn 1996, Matsumura & Pookajorn 2005)
	Laos: Tam Pa Ling	Human	46000	48500	51000	Demeter et al. 2012
Vietnam - Gen Human	Hap M22 (Orang Asli, Proto Malays, Vietnamese, S Chinese)	Human	46000	51500	63000	Morley 2017 (Demeter et al. 2012a, b, 2015)
	Hap M22 (Orang Asli, Proto Malays, Vietnamese, S Chinese)	Human	72500	25500	29500	Joann et al. 2012 (Soares et al. 2009)
	SE Asia & E Asia B1/B2	HBV	4330	8000	12800	Paraskevisev et al. 2012
	SE Asia B3	HBV	1700	4100	6900	Paraskevisev et al. 2012
SE Asia - HBV Genetics	SE Asia B3-B9	HBV	4700	8500	13200	Paraskevisev et al. 2012
	SE Asia C5	HBV	3600	6600	10600	Paraskevisev et al. 2012
	Vietnam: Mai (P)	Human	1700	6800	13200	Morley 2017 (multiple ref.)
	Vietnam: Duan (P) (Man Dac)	Human	46000	54500	63000	Morley 2017 (Bacon et al. 2006, 2008)
Vietnam - Arch H. spp. unknown	Vietnam: Long Truong	Human	60000	70000	80000	Morley 2017 (Bacon et al. 1996, Bacon et al. 2008)
	Vietnam: Hang Nam	Human	80000	110000	140000	Morley 2017 (Bacon et al. 2006, Demell 2014)
			46000	75125	140000	

cont.

cont.

Total	Lineage/location	Species	Most Recent Age (kyr)	Mean/Median	Latest Age (kyr)	Reference
Malaysia - Gen Human	Hap B4b (Ternuan Malaysia, ancestral type South China)	Human	21700	24650	27200	Joann et al. 2012 [Soares et al. 2009]
	Hap B4b (Ternuan Malaysia, ancestral type South China)	Human	20600	24600	26500	Joann et al. 2012
	Hap B4b.1a2 (Ternuan Malaysia, ancestral type South China)	Human	8100	9150	10200	Joann et al. 2012 [Soares et al. 2009]
	Hap B4b.1a2 (Ternuan Malaysia, ancestral type South China)	Human	8000	13200	20000	Joann et al. 2012
	Hap B5 (Ternuan Malaysia)	Human	11700	13200	14700	Joann et al. 2012 [Soares et al. 2009]
	Hap B5 (Ternuan Malaysia)	Human	9800	20900	35500	Joann et al. 2012
	Hap M0b1 (Ternuan Malaysia/Philippines, Sumatra, S China)	Human	12900	14600	16200	Joann et al. 2012 [Soares et al. 2009]
	Hap M0b1 (Ternuan Malaysia/Philippines, Sumatra, S China)	Human	6800	18600	30600	Joann et al. 2012
	Hap M0c2 (Ternuan Malaysia, South China)	Human	17800	20100	22400	Joann et al. 2012 [Soares et al. 2009]
	Hap M0c2 (Ternuan Malaysia, South China)	Human	14500	27600	40500	Joann et al. 2012
	Hap M21 (Orang Asli, Ternuan (Vietnamese Cham))	Human	22900	25800	28700	Joann et al. 2012 [Soares et al. 2009]
	Hap M21 (Orang Asli, Ternuan (Vietnamese Cham))	Human	16800	21000	24100	Joann et al. 2012 [Soares et al. 2009]
	Hap M22 (Orang Asli, Ternuan, Philippines, Sumatra, Sumba)	Human	13400	27300	42200	Joann et al. 2012
	Hap M22 (Orang Asli, Ternuan, Philippines, Sumatra, Sumba)	Human	13700	25500	29400	Joann et al. 2012 [Soares et al. 2009]
	Hap M21a (Orang Asli, Neprito W Malaysia, Saba, Thailand)	Human	27600	31100	46600	Joann et al. 2012
	Hap M22 (Orang Asli, Proto Malays, Vietnamese, S Chinese)	Human	16900	25500	28500	Joann et al. 2012 [Soares et al. 2009]
Malaysia - Gen Human	Hap M22 (Orang Asli, Proto Malays, Vietnamese, S Chinese)	Human	22500	34400	51500	Joann et al. 2012
	Hap M22 (Orang Asli, Proto Malays, Vietnamese, S Chinese)	Human	17500	23506	25600	Joann et al. 2012
	Hap B5 (Southern China)	Human	4800			
	Hap B5 (Southern China)	Human	28600	32700	35800	Joann et al. 2012 [Soares et al. 2009]
Southern China - Gen Human	Hap B5 (Southern China)	Human	37100	54700	75600	Joann et al. 2012
	Hap M0/4 (Southern China)	Human	31300	36300	39300	Joann et al. 2012 [Soares et al. 2009]
	Hap M0/4 (Southern China)	Human	34100	52300	73500	Joann et al. 2012
	Hap M0/4 (Southern China)	Human	28600	43625	73500	Joann et al. 2012
Southern China - Arch <i>H. sapiens</i>	Southern China: Tianyuan Cave	Human	80000	100000	170000	Liu et al. 2015, Nature Letters
	Southern China: Lujang	Human	68000	68000	68000	Morley 2017 (Shen et al. 2009)
	Southern China: Lujang	Human	111000	125000	139000	Morley 2017 (Shen et al. 2009)
	Southern China: Zheending	Human	100000	100000	100000	Morley 2017 (Liu et al. 2010)
Southern China - Arch <i>H. sapiens</i>	Southern China: Maludang	Human	68000	98250	139000	
	Southern China: Longlin	Human	13000	15500	18000	Morley 2017 (Liu et al. 2012, 2015)
	Southern China: Longlin	Human	11000	11000	11000	Morley 2017 (Liu et al. 2012, 2015)
	Southern China: Longlin	Human	11000	13250	18000	Morley 2017 (Liu et al. 2012, 2015)
Southern China - Gen HBV	E. Asia B2	HBV	2500	5500	9000	Parakriset et al. 2012
	E. Asia C1	HBV	3600	6400	9600	Parakriset et al. 2012
	SE Asia & E Asia B1/B2	HBV	4300	8000	12800	Parakriset et al. 2012
	SE Asia & E Asia B1/B2	HBV	2500	6633	12800	Parakriset et al. 2012
Central China - Arch <i>H. sapiens</i>	China: Huanglong Cave, Hubei NE/Central China	Human	80000	90000	100000	Bae et al. 2017 (Liu et al. 2010)
	China: Tianyuan Cave, Beijing	Human	40000	40000	40000	Slackin & Racimo 2016 (Fu et al. 2013)
Malaysia (Seleter) - Gen Human	Hap B4c (Seleter Malaysia)	Human	22900	25740	28600	Joann et al. 2012 [Soares et al. 2009]
	Hap B4c (Seleter Malaysia)	Human	20400	34000	53000	Joann et al. 2012
	Hap B4c2 (Seleter Malaysia)	Human	8900	9460	10560	Joann et al. 2012 [Soares et al. 2009]
	Hap B4c2 (Seleter Malaysia)	Human	13300	13300	24100	Joann et al. 2012
Indonesia - Arch <i>H. floresiensis</i>	Hap G1x (Korcans, Han Chinese, Seleter Malaysia)	Human	10700	12100	13500	Joann et al. 2012 [Soares et al. 2009]
	Hap G1x (Korcans, Han Chinese, Seleter Malaysia)	Human	5400	15100	25500	Joann et al. 2012
	Hap G1x (Korcans, Han Chinese, Seleter Malaysia)	Human	3400	18617	53000	Joann et al. 2012
	Hap G1x (Korcans, Han Chinese, Seleter Malaysia)	Human	3400	18617	53000	Joann et al. 2012
Indonesia - Arch <i>H. floresiensis</i>	Indonesia: Liang Bua	<i>H. floresiensis</i>	50000	120000	190000	Morley 2017 (multiple ref.)
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	13800	15560	17500	Joann et al. 2012 [Soares et al. 2009]
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	11100	21100	31500	Joann et al. 2012
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	11500	12960	14900	Joann et al. 2012 [Soares et al. 2009]
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	8500	17200	26900	Joann et al. 2012
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	19300	21700	24100	Joann et al. 2012 [Soares et al. 2009]
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	13700	27300	42200	Joann et al. 2012
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	12900	14600	16200	Joann et al. 2012 [Soares et al. 2009]
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	6800	18600	30600	Joann et al. 2012
	Hap F1a1 (Blakpoh-Bonco, Sumatra)	Human	6800	18625	42200	Joann et al. 2012

cont.

cont.

Total	Lineage/Location	Species	Most Recent Age (kyr)	Mean/Median	Latest Age (kyr)	Reference
Indonesia - Arch <i>H. sapiens</i>	Indonesia: Lida Ajer Cave, Sumatra	Human	63000	68000	73000	Westaway et al. 2017
	Indonesia: Punung, Java	Human	81000	103500	126000	Morley 2017 (Storn et al. 2005, Westaway et al. 2007)
	Indonesia: Wajulik, Java	Human	29000	33000	37000	Morley 2017 (Dubois 1991, Storn et al. 2013)
Indonesia (Austronesian) - Gen Human	Hap E1b [Selekar, Bidayuh Borneo Malaysia]	Human	5100	5750	6400	Jinam et al. 2012 (Soares et al. 2009)
	Hap E1b [Selekar, Bidayuh Borneo Malaysia]	Human	1500	8200	15700	Jinam et al. 2012
	Hap M8a6a (Malaysian, incl Bidayuh & Selekar)	Human	4900	5500	6100	Jinam et al. 2012 (Soares et al. 2009)
	Hap M8a6a (Malaysian, incl Bidayuh & Selekar)	Human	2600	8200	14900	Jinam et al. 2012
	Hap M7/G3a (Bidayuh Malaysia)	Human	4900	5550	6200	Jinam et al. 2012 (Soares et al. 2009)
	Hap M7/G3a (Bidayuh Malaysia)	Human	900	9300	19200	Jinam et al. 2012
			900	7083	19200	
	Indonesia B7	HBV	2500	4800	7400	Parakeriset al. 2012
	Indonesia B8	HBV	400	3200	6000	Parakeriset al. 2012
	Indonesia B3, B5, B7/B9	HBV	3700	6500	10000	Parakeriset al. 2012
Indonesia - Gen HBV	Indonesia D5	HBV	1800	4200	6800	Parakeriset al. 2012
			400	4675	10000	
Philippines - Gen Human	Hap M7/B1 (Ternuan Malaysia/Philippines, Sumatra, S China)	Human	12900	14600	16300	Jinam et al. 2012 (Soares et al. 2009)
	Hap M7/B1 (Ternuan Malaysia/Philippines, Sumatra, S China)	Human	6800	18600	30600	Jinam et al. 2012
	Hap M22 (Orang Asli, Ternuan, Philippines, Sumatra, Sumba)	Human	19300	21700	24100	Jinam et al. 2012 (Soares et al. 2009)
	Hap M22 (Orang Asli, Ternuan, Philippines, Sumatra, Sumba)	Human	13700	27300	42200	Jinam et al. 2012
			6800	20550	42200	
Philippines - Arch <i>H. sapiens</i>	Philippines: Tabon Cave, Palawan	Human	16000	31500	47000	Morley 2017 (Fox 1970, D��troit et al. 2004, Lewis 2007)
Philippines - Arch <i>H. spp. unknown</i>	Philippines: Callao cave, Luzon	Human	66500	66700	66900	Morley 2017 (Mijares et al. 2010)
	Hap B1a (Bidayuh Malaysia)	Human	24100	27100	30100	Jinam et al. 2012 (Soares et al. 2009)
	Hap B1a (Bidayuh Malaysia)	Human	23700	37100	52500	Jinam et al. 2012
	Hap E [Selekar, Bidayuh Borneo Malaysia]	Human	23700	26100	29000	Jinam et al. 2012 (Soares et al. 2009)
	Hap E [Selekar, Bidayuh Borneo Malaysia]	Human	20500	36300	52100	Jinam et al. 2012
Borneo (postglacial) - Gen Human			20500	31650	52500	
	Hap B3b2 (Bidayuh Malaysia)	Human	8000	10000	11100	Jinam et al. 2012 (Soares et al. 2009)
	Hap B5b2 (Bidayuh Malaysia)	Human	5800	16300	27100	Jinam et al. 2012
	Hap F1a7c (Bidayuh Borneo, Sumatra)	Human	13800	15550	17300	Jinam et al. 2012 (Soares et al. 2009)
	Hap F1a7c (Bidayuh Borneo, Sumatra)	Human	11100	21100	31500	Jinam et al. 2012
	Hap M7/4b (Bidayuh Borneo, Hani S China)	Human	20400	29000	37600	Jinam et al. 2012 (Soares et al. 2009)
	Hap M7/4b (Bidayuh Borneo, Hani S China)	Human	20800	35400	49600	Jinam et al. 2012
	Hap M20 (Bidayuh Malaysia, S China)	Human	4400	4950	5500	Jinam et al. 2012 (Soares et al. 2009)
	Hap M20 (Bidayuh Malaysia, S China)	Human	1600	8500	16900	Jinam et al. 2012
	Hap B4a1a (Bidayuh Malaysia)	Human	8100	9650	10000	Jinam et al. 2012 (Soares et al. 2009)
	Hap B4a1a (Bidayuh Malaysia)	Human	3200	12300	22200	Jinam et al. 2012
			1600	15615	49600	
Borneo (Polynesian) - Gen Human	Niah Cave, Borneo	Human	50000	50000	50000	Soares et al. 2016, Bellwood & Glover 2006 (Hunt et al. 2007)
	Borneo: Niah Cave	Human	45100	45900	46700	Summichr��pes et al. 2010
	Borneo: Niah Cave	Human	46000	46000	46000	Morley 2017 (multiple ref.)
Borneo - Arch <i>H. sapiens</i>			45100	47300	50000	

cont.

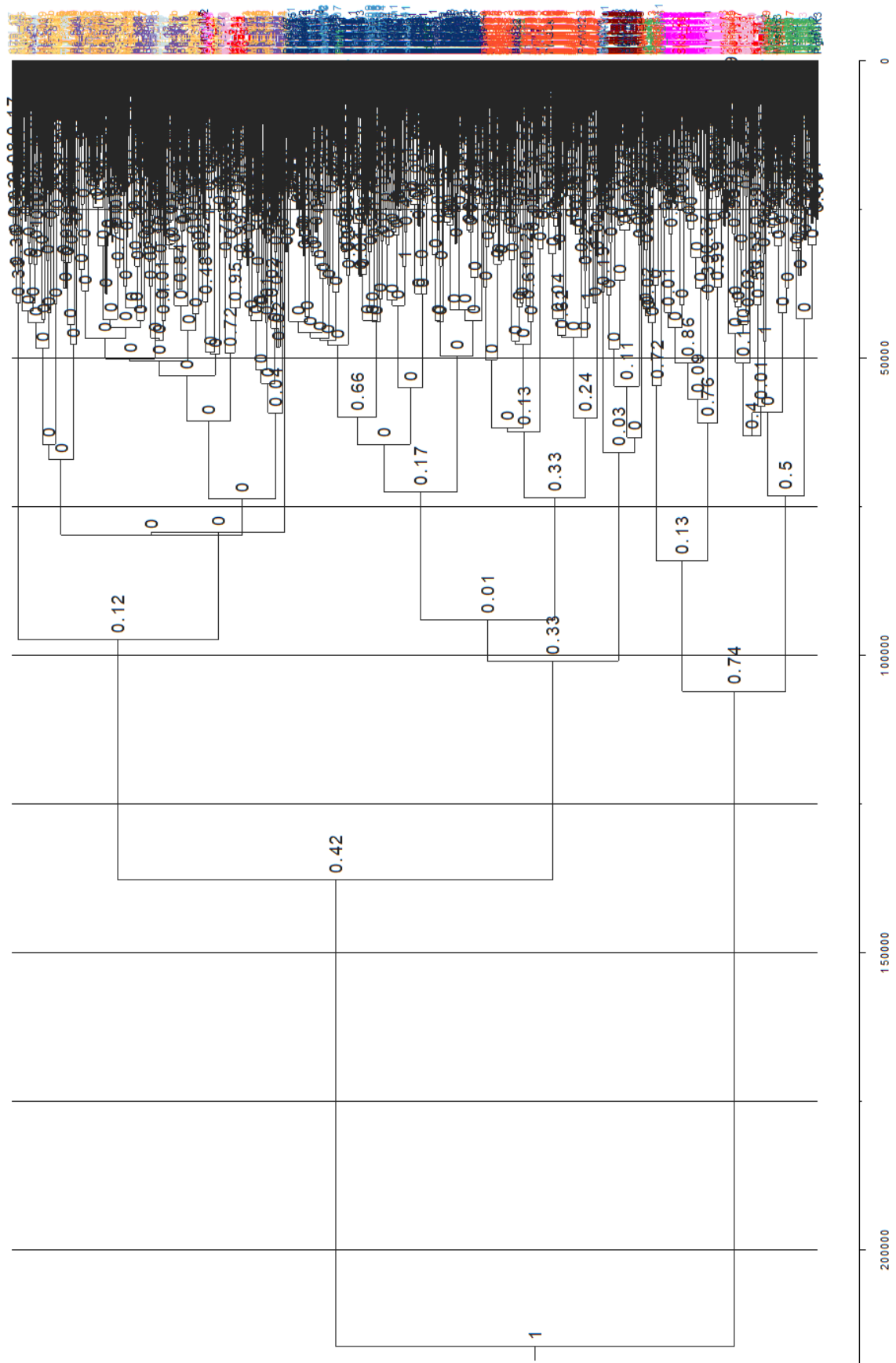
cont.

Total	Lineage/Location	Species	Most Recent Age (kyr)	Mean/Median	Latest Age (kyr)	Reference
Melanesia (Papuan) - Gen Human	Aus/Mel R	Human	46600	58000	66400	Friedlaender et al. 2007
	Aus/Mel IN	Human	45700	53000	60300	Friedlaender et al. 2007
	Aus/Mel M	Human	45500	53000	60500	Friedlaender et al. 2007
	M27 (QRnt577-nt16022)	Human	70100	84000	98700	Friedlaender et al. 2007
	M27 (Syn. Trans)	Human	39400	53100	66800	Friedlaender et al. 2007
	M28 (QRnt577-nt16022)	Human	44300	57800	71300	Friedlaender et al. 2007
	M28 (Syn. Trans)	Human	32300	47300	62300	Friedlaender et al. 2007
	M29 (QRnt577-nt16022)	Human	40700	56500	72300	Hudjashow et al. 2007
	M29 (Syn. Trans)	Human	26300	42300	58300	Friedlaender et al. 2007
	P1 (QRnt577-nt16022)	Human	43000	53000	63000	Friedlaender et al. 2007
	P1 (Syn. Trans)	Human	32400	44000	55600	Friedlaender et al. 2007
	P2 (QRnt577-nt16022)	Human	34300	47100	59900	Friedlaender et al. 2007
	P2 (Syn. Trans)	Human	24600	38300	52000	Friedlaender et al. 2007
	P3 (QRnt577-nt16022)	Human	62700	75000	87800	Friedlaender et al. 2007
	P3 (Syn. Trans)	Human	45200	59500	73800	Friedlaender et al. 2007
	PMG Coast, Highland & Bougainville (Branch 2)	Human	36000	45000	54000	Ingman & Gyllenstein 2003
	Australia & MG Highland (1c)	Human	37000	46000	55000	Ingman & Gyllenstein 2003
	Australia & MG Highland (1d)	Human	23000	31000	39000	Ingman & Gyllenstein 2003
	Australo-Papuan Divergence from others	Human	51000	58000	72000	Malaspinas et al. 2016
			23000	52753	98700	
AUS Amhem Land - Arch Human	AUS: Amhem Land Escarpment	Human	59300	65000	70700	Clarkson et al. 2017
	PMG Viliakae - Arch	Human	42970	45850	48630	
	PMG Airport Mound - Arch	Human	42520	43845	45170	
	PMG Kosape - Arch	Human	37970	39540	41110	
	PMG Huan Peninsula - Arch	Human	35000	37500	40000	
	PMG Nambé - Arch	Human	17500	23650	30000	
	Bismarck-Buang Merabak - Arch	Human	43100	43995	44890	
Melanesia (Papuan) - Arch Human	Bismarck-Male Rakum - Arch	Human	38950	40025	41100	
	New Britain, New Ireland - Arch	Human	38000	39000	40000	
			17300	39173	48690	
Oceania & Southeast Asia - Gen <i>M. tuberculosis</i>	Oceania & Southeast Asia Split	<i>M. tuberculosis</i>	48000	67000	88000	Comas et al. 2013
Sahul-Asia Split - Gen <i>H. pylori</i>	Sahul-Asia Lineage Split	<i>H. pylori</i>	31000	34000	37000	Moodley et al. 2012

cont.

cont.

Total	Lineage/Location	Species	Most Recent Age (kyr)	Mean/Median	Latest Age (kyr)	Reference
Melanesia (Austronesian) - Gen Human	Indonesia, B4a1a1a	Human	2000	4000	6000	Ingman & Gyllenstein 2003
	Bismarck B4a1a1a	Human	4500	6650	8850	
	PNG Coast & Polynesian Island (Branch 1a)	Human	7000	11000	15000	
			2000	7217	15000	
	Vanuatu	Human	3300	3300	3300	Bedford et al 2006 (Scares et al 2016) Riehl & Hunt 2008 Sheppard 2011 Sheppard 2011 Sheppard et al. 2016 Sheppard et al. 2016 Sheppard 2011 Sheppard 2011 Sheppard et al. 2009 Sheppard et al. 2009
	Western Polynesia (Tonga & Samoa)	Human	2900	2900	2900	
	SOI/Santa Cruz	Human	3141	3192	3243	
	PNG Missau - Arch	Human	3380	3470	3560	
	Tonga - Arch	Human	2340	2510	2680	
	VAN Teouma - Arch	Human	2740	2925	3110	
	VAN Teouma - Arch	Human	2925	2961	2997	
	VAN Mokué - Arch	Human	2932	2982	3032	
	PNG Manus - Arch	Human	3800	3875	4150	
	PNG Manus - Arch	Human	2050	2385	2720	
Melanesia (Lapita) - Arch Human			2050	3050	4150	
Melanesia (Pacific Islands) - Gen HBV	Polynesia C3	HBV	5900	10000	15100	Paraskevise et al. 2012
	Remote Oceania C3	HBV	4500	9500	16300	Paraskevise et al. 2012
	Near & Remote Oceania D4	HBV	2400	6200	11000	Paraskevise et al. 2012
	Remote Oceania C3	HBV	5400	9700	14400	Paraskevise et al. 2012
	Near & Remote Oceania D4	HBV	4800	6300	7700	Paraskevise et al. 2012
	Vanuatu C3	HBV	3300	6100	9500	Paraskevise et al. 2012
	Aborigines C4	HBV	700	4000	7000	Paraskevise et al. 2012
	Indonesia/Micronesia C6	HBV	5100	9200	14100	Paraskevise et al. 2012
	Pacific D4	HBV	4800	6300	7700	Paraskevise et al. 2012
			700	7478	16300	
Melanesia/Oceania - Arch Pigs	Flares - Arch	<i>Sus celebensis</i>	7000	7000	7000	Dohney et al. 2008
	New Guinea/Moluccas - Arch	<i>Sus spp.</i>	3500	3500	3500	Dohney et al. 2008
Australia/PNG - Arch & Gen Dogs	Australia	Dingo mtDNA	4600	5000	18300	Oskarsson et al. 2012
	Australia	Dingo arch	3500	3500	3500	Oskarsson et al. 2012



CHAPTER III

APPENDIX S3.1

Table S3.1

Mastermix and thermocycle are the same for the amplification of both halves of the mitochondrial genome.

S1.1. Mastermix reagents, concentrations and volumes for the PCR amplification of the *P. vivax* mitochondrial genome.

S1.2. Thermocycle protocol for the PCR amplification of the *P. vivax* mitochondrial genome.

S3.1.1

MASTERMIX		
	[final concentration]	x1 volume [ul]
ddH ₂ O	(total vol.-(MM+DNA))	20.85
Buffer (10x + MgCl)	1x	3
dNTPs	0.25mM	0.75
Taq/Enzyme (Roche FSHF)	2.0U	0.4
Primer Forward	0.5uM	1.5
Primer Reverse	0.5uM	1.5
SubTotal		28
DNA		2
TOTAL		30

S3.1.2

THERMOCYCLE			
	Temp.	Time	
10x	94 C	2 min.	
	94 C	30 sec.	
	61 C	1 min.	
	68 C	3 min. 30 sec.	
30x	94 C	30 sec.	
	61 C	1 min.	
	68 C	3. min 30 sec.	+ 20 sec/cycle
	68 C	7 min.	
	4 C	%	forever
PP1		PvMt_0323F	PvMt_3193R
PP2		PvMt_2946F	PvMt_0426R

Table S3.1.3

3.1.3. Primers used for amplification and sequencing of the *P. vivax* mitochondrial genome.

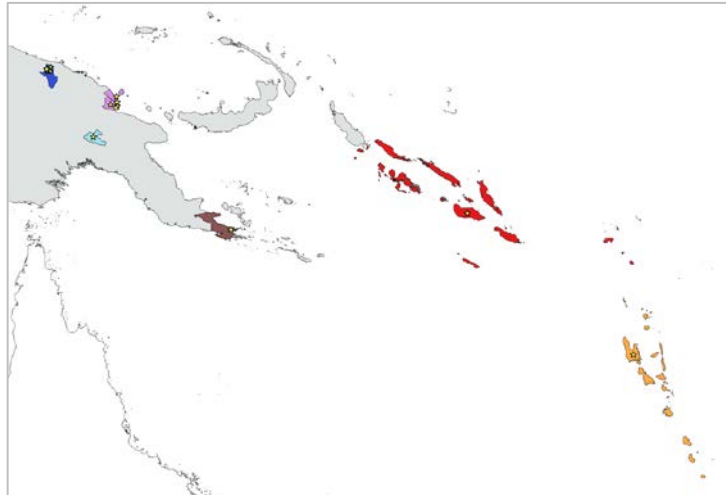
Position	Primer Name	Primer Sequence 5' - 3'
External	PvMtN_0323F	ACGCTGACTTCCTGGCTAAACTTC
External	PvMtN_0426R	GGAAAATTATAGACCGAACCTTGGAC
External	PvMtN_2946F	CCTTGTACACACCGCTCGTCAC
External	PvMtN_3193R	CCATGTCAGGCGTTAAAAGCGTTC
External Internals	PvMtN_0329F	CTTCCTGGCTAAACTCCCAATG
External Internals	PvMtN_0424R	ATAGACCGAACCTTGGACTC
External Internals	PvMtN_2947F	GTACACACCGCTCGTCACG
External Internals	PvMtN_3192R	TCAGGCGTTAAAAGCGTTCG
Internal	PvMt_1582F	GGGAACAAACTGCCTCAAGA
Internal	PvMt_1699R	GCAAACACTAGCGGTGGAAT
Internal	PvMt_3576F	ATGCCAGGATTATTGGAGGA
Internal	PvMt_5409R	GACTTAATAGATTGGATAGAAG
Extra Internal	PvMtE_1019F	GCAGGACGGAGATTACCC
Extra Internal	PvMtN_2111F	ATGACCATTCACTATTAGTATCA
Extra Internal	PvMtE_3570R	GAAGTAATTACGAATCCTCC
Extra Internal	PvMtE_3915F	ACATCIGTAATGCTATTATTAACA
Extra Internal	PvMtE_3892R	ATGTTAATAATAGCATTACAGATG
Extra Internal	PvMtE_3864F	ACTACTGTAATGCATCTAAGA
Extra Internal	PvMtE_4061R	AATATACACTCAGGATGTCCA
Extra Internal	PvMtN_4115F	CAGCATTGGTGTATTAGTCA
Extra Internal	PvMtE_4685F	GACGTATTCCTGATTATCCAGA
Extra Internal	PvMtE_4927R	ATAAGATATTCTGGAGTATAACG
Extra Internal	PvMtE_4952F	GTTATACCTCAGAAATATCTTATG
Extra Internal	PvMtE_5384F	GGTAGCACAAATCCTTTAGG
Extra Internal	PvMtE_5610F	ACAATACCTAGTAAACAGCAG
Extra Internal	PvMtN_5967R	TTAACATAATTATAACCTTACGGT

APPENDIX S3.2

Table S3.2: Geographical positions and ethnography for the known sample sites in Melanesia. Samples were taken from both coastal (mostly Austronesian) and inland (Papuan) populations, thus covering a representation of different human ethnicities. Only 7 samples from Vanuatu and 4 samples from the Solomon Islands were taken from previously sequenced data. Ethnographic information was taken from Lewis *et al.* 2015 and Merriweather *et al.* 1999, mtDNA and NRY data* from, Kayser *et al.* 2008 (Coastal, given here as Alexishafen; Highlands, given here as Sigimaru), Merriweather *et al.* 1999 (Madang); Vilar *et al.* 2008 (Wingei, close to the area given here as Maprik), van Oven *et al.* 2014 (Alotau), Delfin *et al.* 2012 (SOI**), Duggan *et al.* 2014 (SOI), Merriweather *et al.* 1999 (VAN**).

*NB 1: Only data of known genetic heritage is given. Missing percentages are unknown or undetermined data in the literature.

**NB 2: human reference samples across all populations, malaria samples only from Austronesian speaking populations).



Name	Province	Country	Longitude	Latitude	Ethnography	Language	Human mtDNA	Human NRY	Sample #
Sigimaru	Chimbu	Papua New Guinea	144.819	-6.494	Trans-NewGuinea	Papuan	92% Papuan	99% Papuan	11
Ilaita	East Sepik	Papua New Guinea	142.933	-3.683	Toricelli_Mufian	Papuan	100% Papuan; 0% Austro; 0% Other		7
Ingambils	East Sepik	Papua New Guinea	142.919	-3.677	Toricelli_Mufian	Papuan			1
Kamanakor	East Sepik	Papua New Guinea	142.933	-3.717	Toricelli_Mufian	Papuan			5
Sunuhu	East Sepik	Papua New Guinea	142.929	-3.732	Toricelli_Mufian	Papuan			5
W-Gwiningi	East Sepik	Papua New Guinea	143.047	-3.748	Sepik_Ambulas	Papuan			4
W-Kitikum	East Sepik	Papua New Guinea	142.986	-3.775	Toricelli_Mufian	Papuan			1
W-Kunjingini	East Sepik	Papua New Guinea	143.033	-3.783	Sepik_Ambulas	Papuan			2
W-Maprik	East Sepik	Papua New Guinea	143.055	-3.627	Sepik_Ambulas	Papuan			2
W-Nindigo	East Sepik	Papua New Guinea	142.994	-3.798	Toricelli_Mufian	Papuan			6
W-Patigo	East Sepik	Papua New Guinea	143.03	-3.79	Sepik_Ambulas	Papuan			5
W-Tatamba	East Sepik	Papua New Guinea	142.965	-3.722	Toricelli_Mufian	Papuan			2
W-Wisokum	East Sepik	Papua New Guinea	142.969	-3.757	Toricelli_Mufian	Papuan			4
Alexishafen	Madang	Papua New Guinea	145.817	-5.083	Austronesian & Trans-NewGuinea	Both	68% Austro	82% Papuan	25
Madang	Madang	Papua New Guinea	145.793	-5.233	Austronesian	Austronesian	24% Austro		3
Malala	Madang	Papua New Guinea	145.38	-4.47	Trans-NewGuinea	Papuan			10
Mugil Mission	Madang	Papua New Guinea	145.776	-4.835	Austronesian & Trans-NewGuinea	Both			15
Utu	Madang	Papua New Guinea	145.539	-5.161	Trans-NewGuinea	Papuan			11
Yagaum	Madang	Papua New Guinea	145.75	-5.3	Austronesian	Austronesian			2
Alotau	Milne Bay	Papua New Guinea	150.459	-10.316	Austronesian	Austronesian	55% Austro	79% Papuan	2
Solomon Islands	Total All	Solomon Islands	160.156	-9.646	Austronesian	Austronesian	60-80% Austro	73-78% Papuan	21
	Ngela	Solomon Islands					25% Papuan; 75% Austro; 0% Other		-3
	Guadalcanal	Solomon Islands					22% Papuan; 76% Austro; 2% Other		-12
	Malaika	Solomon Islands					33% Papuan; 67% Austro; 0% Other		-1
Vanuatu		Vanuatu	166.959	-15.377	Oceanic Austronesian	Austronesian	46% Austro		10

APPENDIX S3.3

Methods: Phylogenetic Analyses Set-Up

Table S3.3:

To estimate the nucleotide substitution model of best fit, *jModelTest 2.1.7 v20141120* (Guindon & Gascuel 2003, Posada 2008, Darriba *et al.* 2012) was employed. All results pointed towards models with unequal base frequencies. A significant proportion of invariant sites, no heterogeneity across sites (gamma) and unequal substitution rates were estimated to have optimal fit. Several different nucleotide substitution (GTR (Tavaré *et al.* 1986) and HKY) and evolutionary clock models (strict clock and relaxed clock log normal) were run in *BEASTv2.1.3* and compared in *Tracer v1.6* (Rambaut *et al.* 2014) through the Model Comparison AICM calculation (S.E. estimated from 1000 bootstrap replicates; Baele *et al.* 2012). An overview of the major runs and results can be seen in Appendix S3.3. Results for the strict clock were consistently stronger (higher AICM values). For each second model setting (round 2) two runs of 7.5×10^8 were computed and the results were combined in *LogCombiner v2.1.3* (Bouckaert *et al.* 2014). *TreeAnnotator* was used to combine generated trees with the Maximum Clade Credibility Tree setting, with Common Ancestor node heights, 75% posterior probability limit and a 20% burn-in. Using Bayesian Skyline models implemented in *Tracer v1.6* (Drummond *et al.* 2011, Ho & Shapiro 2011, Drummond & Bouckaert 2015) population size estimates over time were calculated.

Round	EvolutionaryStudyModel	SubstitutionModel	BranchRateModel	Gamma	ProportionInv(est)	GroupSize	AICM_Value
1a	CBS	GTR	SC	0	0.95	7	21841.804 (+/- 0.479)
1b	CBS	HKY	SC	0	0.95	7	21860.133 (+/- 0.934)
1c	CBS	GTR	RCLN	0	0.95	7	22162.754 (+/- 1.067)
1d	CBS	HKY	RCLN	0	0.95	7	22156.984 (+/- 1.038)
2a	CBS	GTR	SC	0	0.8	8	21889.079 (+/- 0.869)
2b	CBS	HKY	SC	0	0.8	8	21874.249 (+/- 0.78)
2c	CBS	GTR	SC	0	0.8	6	21881.611 (+/- 1.013)
2d	CBS	HKY	SC	0	0.8	6	21871.59 (+/- 0.75)

APPENDIX S3.4

BEAST settings for estimating Papuan vs Austronesian optimal MRCA introduction times and molecular clock rates of Melanesian *Plasmodium vivax*. Given in bold are the age range estimates for the Melanesian clade that the mutation speed is fitted to.

Panel	Variable	Subvariable	Papuan Settings	Austro Settings
AICM_Value	Likelihoods_1000btstrp		24560.826 (+/-0.438)	24556.052 (+/- 0.695)
ClockModel			SC	SC
	Clock.Rate		1.00E+00	1.00E+00
		Estimate	Yes	Yes
Initialization	bPopSizes	Upper	5.00E+08	5.00E+08
	bPop&bGroup	Dimensions	6	6
Priors	Tree.t	Type	CoalescentBayesianSkyline	CoalescentBayesianSkyline
		StartTree	RandomTree	RandomTree
	MarkovChainPopSizes.t		Jeffreys	Jeffreys
	ClockRate	Distribution	LogNormal	LogNormal
		M	4.40E-09	4.40E-09
		S	2	2
	GammaShape	Distribution	Exponential	Exponential
	Kappa.s	Distribution	LogNormal	LogNormal
MEL.prior	Distribution		Laplace Distribution	Laplace Distribution
	Monophyletic		Yes	Yes
	Mu		45000	6000
		Scale	5000	1000
		Offset	0	0
		5%Quantile	33500	3700
		Median	45000	6000
		95%Quantile	56500	8300
MCMC	ChainLength		750000000	750000000
	StoreEvery		100000	100000
	PreBurnin		100000	100000
#RunsCombined			3	3

Bayesian Skyline Plot of Estimated Population Sizes

Overview of population size estimates over time, for the Melanesian and East & Southeast Asian region. Bayesian Skyline Plots for median population size estimates over time (years ago) for the entire region. 95% HPD intervals are given in blue.

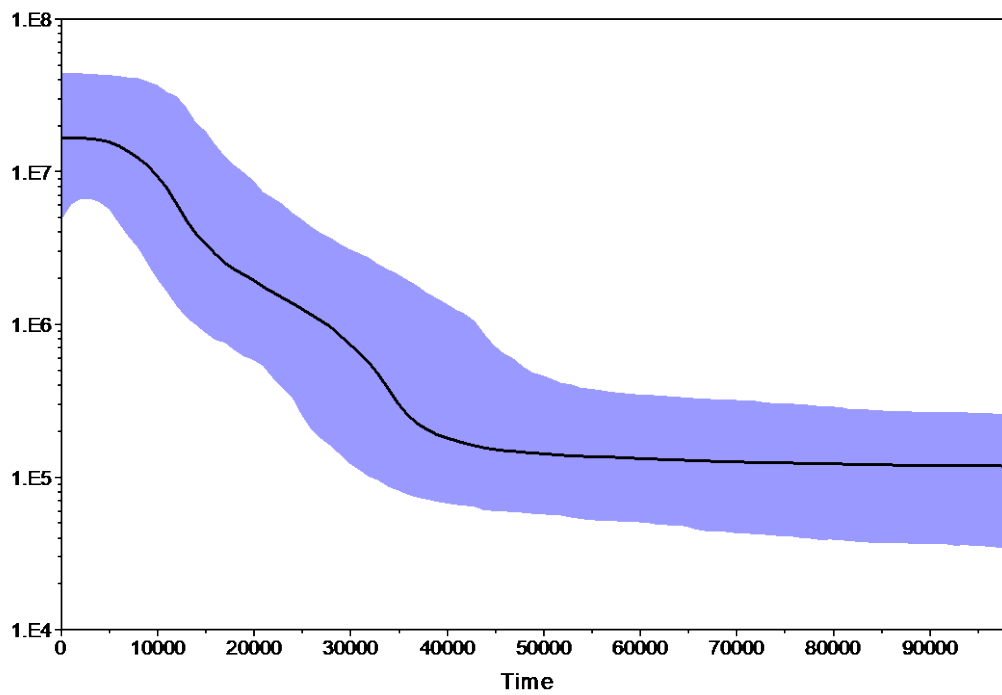


Table S3.6.1: Within population diversity and differentiation statistics per region, calculated in *dnasp5*. For three regions (Total, Southeast Asia and Melanesia) comparative data has been given taken from Taylor *et al.* 2013.

Geographic Region	Sample size (n)	Haplotypes (h)	Haplotype Diversity (Hd)	Neutrality Test (Tajima's D (P))
Total	506	202	0.969	-2.336 (<0.01)
<i>Global (Taylor et al. 2013)</i>	731	357	0.985	-2.576 (0.000)
Melanesia	221	97	0.865	-2.576 (<0.001)
<i>Melanesia (Taylor et al. 2013)</i>	72	37	0.929	-2.035 (0.002)
Island SE Asia	51	30	0.908	-2.381 (<0.01)
Mainland SE Asia	72	40	0.955	-2.200 (<0.01)
<i>Southeast Asia (Taylor et al. 2013)</i>	154	81	0.968	-2.109 (0.000)
South China	59	16	0.899	0.695 (>0.1)
NE Asia	70	25	0.865	-0.838 (>0.1)
Myanmar	32	17	0.885	-1.792 (0.1-0.05)

Table S3.6.2: Interregional population differentiation. Deletions in the sequence were considered as a fifth state. Fst values are given in the bottom left, Kst* values in the top right half of the table. Kst* values were all $p > 0.001$ for 1000 repetitions permutation test, calculated in *dnasp5*. The four lowest Kst* values have been underlined; the four highest values have been put in bold font, to highlight the geographical connections and boundaries respectively for *P. vivax* mitochondrial genetics.

Kst*	Melanesia	Island SEAsia	Mainland SEAsia	South China	NE Asia	Myanmar
Melanesia		0.191	<u>0.117</u>	0.169	0.215	<u>0.115</u>
Island SE Asia			0.18	<u>0.124</u>	0.178	0.275
Mainland SE Asia				0.136	0.19	0.128
South China					<u>0.036</u>	0.187
NE Asia						0.199
Myanmar						

Figure S3.7

MCC tree of round 1a GTR CBS SC BEAST2 Bayesian Phylogeny, with posterior values of the branch nodes. Posterior values show support for the placement of a branch across all trees generated and range between 0 (no consistent support) and 1 (found in all trees). The low values seen here are due to the limited - or absent - amount of variation between the sequences, which means branches can easily swap between trees. Usually support below 0.7 is seen as low, but here even support of 0.1 is strong. This also causes some of the branches inverse, as precise time estimation becomes harder.

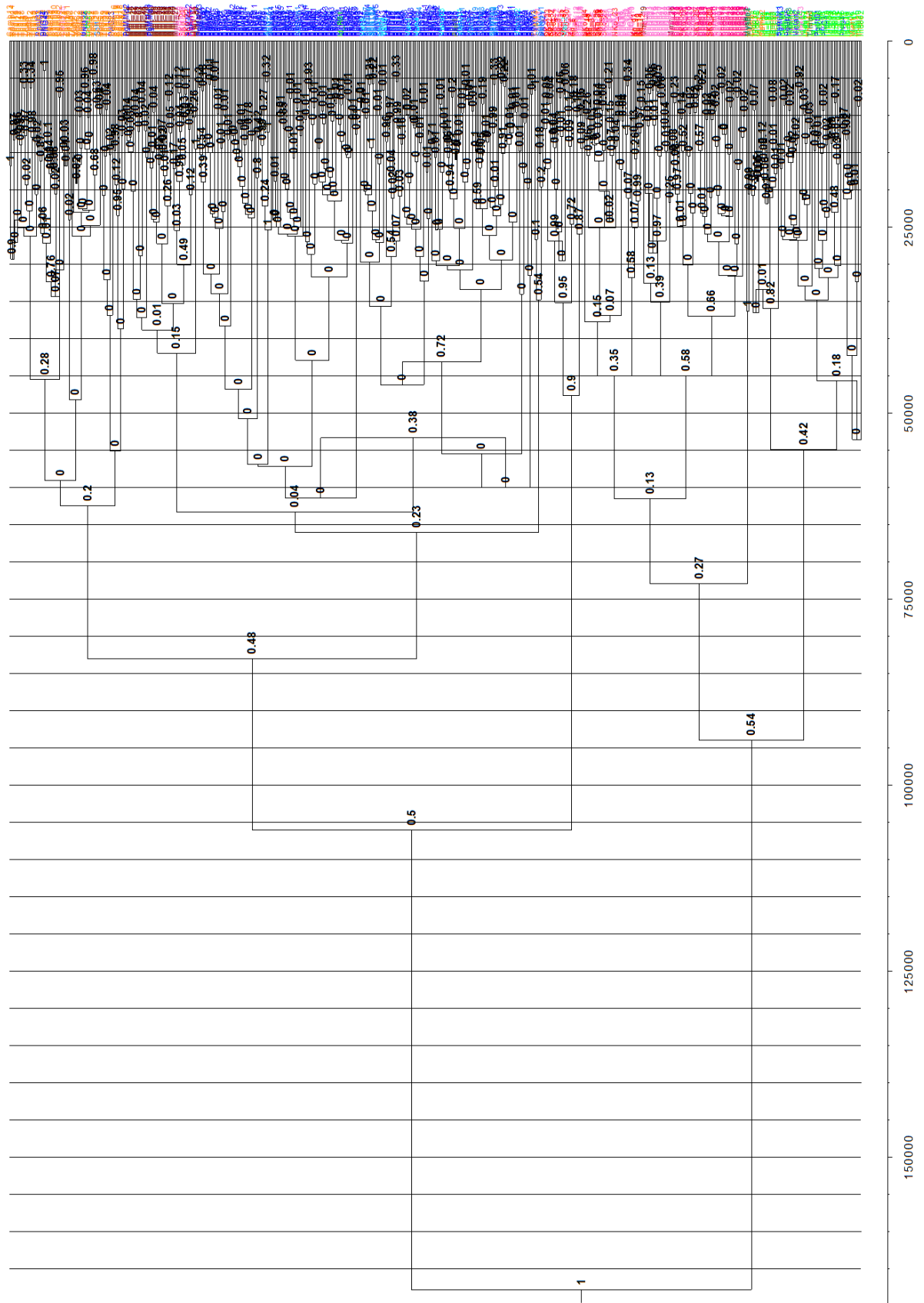


Table S3.8

Overview of calculated divergence time estimates compared to those across published data of several species and archaeological estimates. Dates from this paper are at the bottom.

Lineage/Location	Species	Data Type	Method	Reference	Latest Age (kya)	Median/Average Age (kya)	Earliest Age (kya)
Aus/Mel R	Human	mtDNA	Phyl tree rho statistic	Hudjashov <i>et al.</i> 2007	49500	59000	66500
Aus/Mel N	Human	mtDNA	Phyl tree rho statistic	Hudjashov <i>et al.</i> 2007	45700	53000	60300
Aus/Mel M	Human	mtDNA	Phyl tree rho statistic	Hudjashov <i>et al.</i> 2007	46500	53000	60500
M28 (CR-nt577-nt16022)	Human	mtDNA & gDNA	Phyl tree rho statistic	Friedlaender <i>et al.</i> 2007	44300	57800	71300
M29 (CR-nt577-nt16022)	Human	mtDNA & gDNA	Phyl tree rho statistic	Friedlaender <i>et al.</i> 2007	40700	56500	72300
P1 (CR-nt577-nt16022)	Human	mtDNA & gDNA	Phyl tree rho statistic	Friedlaender <i>et al.</i> 2007	43000	53000	63000
P2 (CR-nt577-nt16022)	Human	mtDNA & gDNA	Phyl tree rho statistic	Friedlaender <i>et al.</i> 2007	34300	47100	59500
P3 (CR-nt577-nt16022)	Human	mtDNA & gDNA	Phyl tree rho statistic	Friedlaender <i>et al.</i> 2007	62200	75000	87800
PNG Vahiauv	Human	human site	radiocarbon dating	Summthayes <i>et al.</i> 2010	42570		48090
PNG Airport Mound	Human	human site	radiocarbon dating	Summthayes <i>et al.</i> 2010	42520		45170
PNG Kosye	Human	human site	radiocarbon dating	Summthayes <i>et al.</i> 2010	37570		41110
PNG Huon Peninsula	Human	stone Tools	tephrochronology, re-dating, strat. data	Groube <i>et al.</i> 1986	35000		40000
PNG Morobe Rockshelter	Human	stone Tools	radiocarbon dating	Evans & Mountain in Pawley <i>et al.</i> 2005	17300		30000
Bismarcks Buang Merabak	Human	human site	radiocarbon dating	Summthayes <i>et al.</i> 2010	43100		44890
Bismarcks Matankupum	Human	human site	radiocarbon dating	Summthayes <i>et al.</i> 2010	38550		41100
New Britain, New Ireland	Human	shell tools, chert		Specht in Pawley <i>et al.</i> 2005	38000		40000
Indonesia, Batajajala	Human	mtDNA	ML & molecular clock	Saeres <i>et al.</i> 2011	2000	4000	6000
Bismarcks, Bialakala	Human	mtDNA	ML & molecular clock	Saeres <i>et al.</i> 2011	4500	6650	8850
PNG Mussau	Human	Lapita pottery	radiocarbon dating	Sheppard 2011	3380	3470	3560
Tonga	Human	human remains	radiocarbon dating	Sheppard <i>et al.</i> 2016	7340	7510	7680
Vanuatu Teouma	Human	human remains	radiocarbon dating	Skoglund <i>et al.</i> 2016	2740	2925	3110
Vanuatu Teouma	Human	Lapita pottery	radiocarbon dating	Sheppard 2011	2525	2561	2597
Vanuatu Makué	Human	Lapita pottery	radiocarbon dating	Sheppard 2011	2932	2982	3032
Sahul-Asia Lineage Split	<i>H. pylori</i>	DNA	Bayesian MCMC	Moodley <i>et al.</i> 2009	31000		37000
Melanesia Lineage Age (TMRCA)	<i>P. vivax</i>	mtDNA	Bayesian Multispecies Coalescent model	Taylor <i>et al.</i> 2013	27000	60000	117000
SE Asia Lineage Age (TMRCA)	<i>P. vivax</i>	mtDNA	Bayesian Multispecies Coalescent model	Taylor <i>et al.</i> 2013	34000	98000	195000
Mel-Asia Lineage Age (TMRCA)	<i>P. vivax</i>	mtDNA	Coalescent Bayesian Skyline model (S3.2d)	This paper	38002	61894	96067
Melanesia Lineage Age (TMRCA)	<i>P. vivax</i>	mtDNA	Coalescent Bayesian Skyline model (S3.2d)	This paper	39505.5249	54809.90336	78081.7627
					LowerHeight95%HPD	MedianHeight	UpperHeight95%HPD
					38002	61894	96067
					39505.5249	54809.90336	78081.7627
							78885
							69812

Consensus sequence generated from all *P. vivax* whole mitochondrial genome sequences available on Genbank (November 2014). The truncated sequence at the end is shown between parentheses.

TTCCACTACCAAAATATAATCTCCTGTTCTAATGTTCTAGGGTTTTTCGCGTTTTTTCAGGAGAAATCCGTATATCGATGTTTTTAAATCTACGCTATTGGATTCAACGTCCAGGACTTCTGACGTTTAATAACGATTCTACTTCCAGCAGCCATTTTGGTTTCAGCTACAAGTTCACTGTCAACTACCATGTTACGACTTCGCACCGACTGTTTTTCTTTACCTCAGAGTCGATCAGGAAGGTTTCATCCTTAATCTCGTAACCATGCCAACACATAAGAAGTTTTCAGGGAAGTTAAGGTGCTCAGGGTCTTACCGTCGGGCCGTATGATTCACATATTCATGGATAATCTATTTATTAGGAGTCTCACACTAGCGACAATGGGGAAGTCGTTACACCGTTTCATGCAGGACGGAGATTACCCGACAAGGAATTTTGTACCTTAGGACCGTTTACGATACAGCCGCCGTTTATCATTGATGCCGGGCAGATGTCAGTAACCTGAACTATTTCATCAGGAATTATCAGTGACTTGTGTTGAACCTTACAGACGCTTCCAGAATATAAATCTTATAAATGGTAGCGCCGGTTTCCGGGTATCCAATCCAGTGCTCCATTCAAGGCATAGAGACTCAGCCTGTGTTCAACTTTGTAGGATATATAATATAGCTTTTGGTATCTCGTAATGTAGAACATAATGGGTGACCGTCAAATCCTTTTCATTAAAGAGTGGACTAATGCCAGCCAACACCATCCAATTTGATTGGGAATAATCTGTGTTACAATACTTTTTGATCCCAGGCTGGTAAAAAATGTAACCTTTAGCCCAAAAGAATAGAAACAGATGCCAGGCCAAAAACCCAAAAATAGAGCTATGACGCTATCAATTTGACAAGGCAGATAAATCTTTTCATAGAACTTAACGTTTCATCCTCCATACATAAAACCGGTAGATAGGGAAACAACTGCCCTCAAGACGTTCTTAACCCAGCTCAGCGATCGCTTCTAACGGTGAACCTCTCATTCCAAATGGAACCTTGTCAAGTTCAAATAGATTGGTAAGGTATAGCGTTTACTATCGAATGAAACAATGTATCCACCGCTAGTGTTGCTTCTAACATTCCACTTGCTTATAACTGTATGACGTAACTCCAGGCAAGAAAAATGACCGTCAAAACGGAATCAGTTAACTATGGATAGCTGATACTAACAAATTTATCATTACTCAAGTCAGCATAGTCTATATGAAGGTTTCTATGGAAACACACTTCCCTTCTCGCCATTTGATAGCGGTTAACCTTCTTTTCTTACGTACTCTAGCTTTTAAACACAATATTATTGTCTATACTAGATACTATAGTTGAAACAGGACATATACATATTCATTATTCTGAATAGAAAAAGAACTCTATAATAACCATACAATTTCAACAAAAAGCAGTATAATATTGTAATTTGATCAGTATGAGGTAATACAATAAAGATATACCATATAATGAATATGACCATTCACTATTAGTATCATACATTTCAACTATTCTTATAAAATATATTAATAATAATAAACCACCTATAACATGTGAAAAATGTAACCTGTAAACACAATAAAATAATGTACCAGATATAGCATCATTTATACATATATCCTAAATGTAAGTATTCAGTTGTTGAAGTGATGCAAAACATTCTCCAAGTAAGTATATTATAAATAACAATACTAGAAATTTCTAACTCATTCTTTTCTATTAGAAATTGTAAACATGCTGTCATACATGATGCACTTGCTAAAAATAATGTAATAGTTAGAAATTAACATTCTGATGATGTTAAAAATAATTCCTTCATTATATAATGGATGGAGATAAACTAAAAATGTAATAACCCCAAAAAATATGTAATAAATAATATTGCTTCAGATATAATTATAGATAACATTCTGATACTAATGATGAATAATAGATAAAAACTTTCTCTAATAGAAATATACAATAATTATAAAAATAAGTATGAATGTAAATAAGATACCCACAGAAAAATTTTTAAAGATGTTCCATATAATGATGTTAATGAAGGATAAGATACTAAATGTGCTTAAATATTATTATAGTTACTAAATATAAAAAATATATTATAAGAACGGTGATTTTGTGTGCCGTTAACATATAACGGTAAGAAGGTTCCGCGGGGATAACAGGTTATAGTATATATAGAGCTCAAATCTTATATACTATTGGCACCTCCATGTCGTCTCATCGCAGCCTTGCAATAAAATTAATATTATTAGCGTGATTTGTCCTGTACACACCGCTCGTCACGCAAAATTTATATTACTAAGATAAAGAACTCCAGGCGTTAACTGTAGAGTTGAGATGGAACAGCCGGAAGGAAATATTACGTCCAAATGATAAGAATATATAGAAATATACTAGCATGGGACTAAAAAATGTTATGTTGTTGGTTTAAGCCCTTTTACCATACAGAGATCGCGTACTTTGGACTGAAAAAAGCTGTGAGGAAACTACATTAAAGGAACCTCGACTGGCCTACAATATAAGAACGAAACGCTTTTAAACGCTGACATGGATGGATAAATCTCGACTTTTCCAAAGTATAACCGCTGTCGCTGGGACTGTATGGATCGAATCTTACTTATTCATATCCAAGCTCCTATTGTTAATTATATATTATTTTTTTGTTTTCAATAGATATACACTTATTACAAATGCAATCATAAAACCTTTAGGTCTATACTATTATGGTTTTTCATTTTTATTGGTAGTTATGGTTTTTTTATTATCTGTTATTTTACGTACAGAAATATATTCTTCTTTTAAAGAAATAATTGCACAGAAAAATGTTAACTTATATAATATGATATTTTACATTACATGGAATTATTATGATATTCTTAAATAATAGCCAGGATTATTGGAGGATTCGGTAATTACTTCTACCAATTTTATGTGGTTCTCCAGAACTTGCCATATCCAAGAATTAATAGTATATCTTTATTTATACAACCAATAGCTTTTATATTAGTCATTTTATCTACAGCAGCAGAAATTTGGAGGAGTACTGGATGGACCTTATATCCACCAATTAAGTACATCACTTATGTCTTTATCTCTGTGCAAGTAGATGTTATCATTGTTGGTCTTTTAGTATCTGGTATTGCTAGTATTATGTCTTCTTAAATTTTATTACTACTGTAATGCATCTAAGATCTAAAGGTTTAACTTGGTATATTAAGTGTATCTACATGGTCATTAATAATTACATCTGTAATGCTATTATAACATTACCTGTTTTAACAGGTGGTGTGTTAATGTTATTATCAGATTACATTTTAAACATTATTTTTGATCCTACATTGCTGGAGATCCTATTTTATATCAACATCTATTTTGGTTTTTGGACATCCTGAAGTGTATATTTAATATTTACCAGCATTTGGTGTTATTAGTCATGTAATATCTACAAATTTATTGTAGAAGTTTATTGTGAATCAATCTATGATTTTAGCAATGAGTTGTATTGCTATATTAGGAAGTGTGTATGGGCTCATCATATGTATACTACAGGTTTAGAAGTAGATACAAGAGCAATTTTTTACATCTACAACATATTAATATCTATACCTACTGGAAACAAAAATATTAAATTGGATATGTACATATATGGGTAGTAATTTTGGTAACTCATAGTTCTCTTTTATTATCATTACTATTTATATGTACATTTACTTTTGGTGGTACTACAGGAGTAATATTAGGTAATGCAGCTATTGATATTGCATTACATGATACTTACTATGTAATCGCTCATTTCCATTTTGTATTATCTATAGGTGCAATTATTGGATTGTTTACATTAGTAAGTAGTTTCAAGAAAACTTTTTGGTAAACATTACGTGAAAAATCTATAAATAATATTATGGTCAATCTATTTTTTATTGGAGTTGATTAAACATTCTTACCTATGCATTTTCTTGGATTTAATGTAATGCCTAGACGTATTCCTGATTATCCAGACGCTTAAATGGATGGAATATGATTGTTCAATTGGATCAACAATGACTTTATTTGGTTTATTTTAAATAATATAAAATATTTTTGTTTATATGAATTATTATTCTATTAATTTAGCAAAAGCACATTTATTAATTAACCATGCCCATTAATTAATTTCTTATGGAATTATGGATTTCCTTTAGGTATTATTTCTTTATACAAATATTAACAGGTGTATTTTAGCAAGTCGTTATACTCCAGAAATATCTTATGCAATTATAGTATACAACATATTTTAAAGAGAATTATGGAGTGGATGGTGTTTAGATATATGCATGCTACAGGTGCATCTCTGTATTCTTATTAACCTATTATACACATTTTAAAGAGGATTAATTAATTCATATTTATATTTACCATTATCATGGATATCAGGATTAATCATATTTGCATTATTTATAGTTACAGCTTTTATAGGTTACGTTTACCTTGGGGACAAATGAGTTATTGGGGTCAACTGTTATAAATAATTTATTATCTCAATTCAGTATTAGTAATTTGGTTATGTGGAGGATATACTGTAAGTGATCCTACAATTAAGAGTTTTTGTATTACATTTTATACCTTCCCTTTGTAGCATTATGATTGATTTATACATATATTCTTTTACATTTACATGGTAGCACAATCCTTTAGGGTATGATACAGCTTTAAAAATACCTTCTATCCAAAATCTATTAAGTCTTGATGTAAAAAGGATTTAATAATATATTTATATTCTTATTACAAAGATTTTTTGGTATAATTCATATCTCATCCAGATAATGCTATATTAATTAACATATGTTACACCAATTCAAATTTGTTCTGAATGGTACTTTTTACCATTTTATGCAATGTTAAAAACAATACCTAGTAAAACAGCAGGTTTATTAATGTTTTAGCATCTTTACAATTATTATTTTATTAGCTGAACAAAGAAGTTTAACTACTATAATACAATTTAAATGACATTTGGTGCTAGAGAATATTCAGTACCAATGATATGGTTTATGTGTTTCATTCTATGCTTTATTATGGATTGGATGTCAGTTACCACAAGATTTTTTATTTTATATGGTCGTTTATTATAATATCATTTCTCTCTAGTGAGTATTTGCACTTGTTCATTATAAAAGAACACATTATGATTACAGCTCCCAAGCAAACTTTAAATTAACAAGGCTACGATAAGACGATATCTCTGAATATTGAGCAGAAACAATACAGACCGTAAGGTTATAATTATGTTAAAGTTAGGATAATGAAATTATAGTTACCATAGCTGATAGTGGATGCTTCGATATATTATATTTAAGTATTAATCGAGATTACATGCTCAGCCGCCAAAAATATAACGATATTATTACCGTACAAGCCGTTAGCAAGACATGATAGGGAGTTGACAAGTTAAAGAAGTTCTGGTTTATAATAGATACGTTTTTAAAGTTAGGATGTATGGGATAATTGTAGTACACCTTAATTGGTTTAACTTTATTTGGTTTTATAGATATTATGTTCCGTATTGCATGCCTGGTGTGTTTAAATTAAGACGCTGACTTCTGGCTAAACTTCCCAATGATATATCTTCAAAATAGATTTTCGCAGAAAAACCGCTATATTTTCATGTTTGATTGACCTTAAACCGCTAATTACGAATCTTCAAGAATATTTCAAGAGTCCAAGGTTCCGGTCTATAATTTTCTGTGTAAGAGGATTCTCTCCACACTTCAATTTCGTACT)

CHAPTER IV

APPENDIX S4.1

Table S4.1

Table giving an overview of data from 72 publications of *Plasmodium* spp. genetics, ranging from 1984 till the present (2018). The table lists the genome studied (Genome), the specific region/gene/genome studied (*Specific*), the species involved (*Species*), the size of the genetic sequence used (Size) in kilo base pairs (kbp), the type of research undertaken (*Function*) and the publication the information was taken from (*Reference*). When a study involved more than one species or specific gene these have been listed on separate rows.

The table is continued on the next page.

*MOI: Multiplicity of infection.

Genome	Specific	Species	Size (kbp)	Function	Reference
Nuclear	Actin	<i>Plasmodium</i> sp	1.3	Phylogenetics (interspecies)	McCutchen et al. 1984
Nuclear	Tubulin	<i>Plasmodium</i> sp	0.4	Phylogenetics (interspecies)	McCutchen et al. 1984
Nuclear	Dihydrofolate reductase (DHFR)	<i>Plasmodium</i> sp	0.7	Phylogenetics (interspecies)	McCutchen et al. 1984
Nuclear	Thymidylate synthetase (TS)	<i>Plasmodium</i> sp	1.1	Phylogenetics (interspecies)	McCutchen et al. 1984
N/A	Biological and morphological features	Sporozoa	N/A	Phylogeny (interspecies)	Barta 1989
Nuclear	small subunit rRNA (SSU rRNA)	Apicomplexans, Dinoflagellates, Ciliates	0.5	Phylogenetics (interspecies)	Johnson et al. 1987
Nuclear	small subunit rRNA (SSU rRNA)	Apicomplexans, Dinoflagellates, Ciliates	2	Phylogenetics (interspecies)	Gajadhar et al. 1991
Nuclear	18S small subunit rRNA (18S SSU rRNA)	<i>Plasmodium</i> sp	1.2	Phylogenetics (interspecies)	Waters et al. 1991
Nuclear	small subunit rRNA (SSU rRNA)	Apicomplexans, Dinoflagellates, Ciliates	1.2	Phylogenetics (interspecies)	Wolters 1991
Nuclear	circumsporozoite protein (CSP) gene	<i>Plasmodium falciparum</i>	0.18	Phylogenetics (interspecies)	Doalan et al. 1992
Nuclear	18S small subunit rRNA (18S SSU rRNA)	<i>Plasmodium</i> sp	1.2	Phylogenetics (interspecies)	Siddall & Barta 1992
Nuclear	18S small subunit rRNA (18S SSU rRNA)	<i>Plasmodium</i> sp	1.2	Molecular Diagnostics	Snounou et al. 1993
Nuclear	18S small subunit rRNA (18S SSU rRNA) Type A	<i>Plasmodium</i> sp	0.5	Phylogenetics (interspecies)	Escalante & Ayala 1994
Apicoplast	open reading frame 470 (ORF 470)	<i>Plasmodium falciparum</i>	1.4	Phylogenetics (interspecies)	Williamson et al. 1994
Nuclear	circumsporozoite protein (CSP) gene	<i>Plasmodium</i> sp	1.1	Phylogenetics (interspecies)	Escalante et al. 1995
Nuclear	circumsporozoite protein (CSP) gene	<i>Plasmodium</i> sp	1.1	Phylogenetics (interspecies)	McCutchen et al. 1996
Nuclear	small subunit rRNA (SSU rRNA)	<i>Plasmodium</i> sp	2.1	Phylogenetics (interspecies)	Qari et al. 1996
Apicoplast	Whole	<i>Plasmodium falciparum</i>	35	Gene function	Wilson et al. 1996
Nuclear	18S small subunit rRNA (18S SSU rRNA)	<i>Plasmodium</i> sp	0.5	Molecular Diagnostics	Kimura et al. 1997
Mitochondrial	cytochrome b (cytb)	<i>Plasmodium</i> sp	1.1	Phylogenetics (interspecies)	Escalante et al. 1998
Nuclear	Merozoite surface protein 1 (msp1)	<i>Plasmodium vivax</i>	5	Population structure & MOI* & Phylo (within species)	Lim et al. 2000
Nuclear	Microsatellites MS	<i>Plasmodium</i> sp	0.1	Phylogenetics	Barta 2001
Mitochondrial	cytochrome c oxidase subunit 1 (COXI)	<i>Plasmodium</i> sp	1.4	Phylogenetics (interspecies)	Barta 2001
Nuclear	beta-actin gene	<i>Plasmodium</i> sp	1.5	Phylogenetics (interspecies)	Barta 2001
Nuclear	ribosomal ITS regions	<i>Plasmodium</i> sp	1.1	Phylogenetics	Barta 2001
Nuclear	adenylosuccinate lyase (ASL) gene	<i>Plasmodium</i> sp	1.4	Phylogenetics (interspecies)	Kedzierski et al. 2001
Nuclear	18S small subunit rRNA (18S SSU rRNA)	<i>Plasmodium</i> sp	0.5	Phylogenetics (interspecies)	Rathore et al. 2001
Apicoplast	caseinolytic protease C (ClpC)	<i>Plasmodium</i> sp	0.6	Phylogenetics (interspecies)	Rathore et al. 2001
Mitochondrial	cytochrome b (cytb)	<i>Plasmodium</i> sp	1.1	Phylogenetics (interspecies)	Rathore et al. 2001
Mitochondrial	Whole mt	<i>Plasmodium y. yoelii</i>	6	Gene function	Carlton et al. 2002
Mitochondrial	Whole mt	<i>Plasmodium falciparum</i>	6	Gene function	Carlton et al. 2002
Whole	Whole	<i>Plasmodium falciparum</i>	22,900	Drug design target	Carlton et al. 2002
Whole	Whole	<i>Plasmodium y. yoelii</i>	23,100	Drug design target	Carlton et al. 2002
Apicoplast	Whole ap	<i>Plasmodium falciparum</i>	35	Gene function & Whole genome	Gardner et al. 2002
Nuclear	Whole	<i>Plasmodium falciparum</i>	22,800	Gene function & Whole genome	Gardner et al. 2002
Mitochondrial	cytochrome b (cytb)	<i>Plasmodium</i> sp	1.1	Phylogenetics (interspecies)	Perkins & Schall 2002
Nuclear	circumsporozoite protein (CSP) gene	<i>Plasmodium vivax</i>	1.1	Population structure & MOI*	Cui et al. 2003
Nuclear	Merozoite surface protein 3 alpha (msp3a)	<i>Plasmodium vivax</i>	1.5	Population structure & MOI*	Cui et al. 2003
Mitochondrial	Whole	<i>Plasmodium falciparum</i>	6	Phylogenetics (within species)	Joy et al. 2003
Nuclear	Microsatellites MS	<i>Plasmodium vivax</i>	0.1	Phylogenetics (within species)	Leclerc et al. 2004
Apicoplast	elongation factor Tu (EF-Tu)	<i>Plasmodium</i> sp	1.2	Phylogenetics (interspecies)	Escalante et al. 2005
Nuclear	beta-tubulin	<i>Plasmodium</i> sp	1.3	Phylogenetics (interspecies)	Escalante et al. 2005
Nuclear	cell division cycle 2 (CDC-2)	<i>Plasmodium</i> sp	1.5	Phylogenetics (interspecies)	Escalante et al. 2005
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Jongwutives et al. 2005
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Mu et al. 2005
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Mu et al. 2005
Nuclear	Microsatellites MS	<i>Plasmodium vivax</i>	0.1	Phylogenetics (within species)	Imwong et al. 2006
Nuclear	Merozoite surface protein 3 beta (msp3b)	<i>Plasmodium vivax</i>	4	Population structure & MOI*	Yang et al. 2006
Nuclear	Microsatellites MS	<i>Plasmodium vivax</i>	0.1	Population structure & MOI*	Ferreira et al. 2007
Whole	Whole	<i>Plasmodium falciparum</i>	4,250	Population structure & Drug resistance	Mu et al. 2007
Whole	Whole	<i>Plasmodium falciparum</i>	20,950	Population structure & Drug resistance	Volkman et al. 2007
Mitochondrial	partial cytochrome b (cytb)	<i>Plasmodium</i> sp	1	Phylogenetics (interspecies)	Martinsen et al. 2008
Mitochondrial	partial cytochrome c oxidase subunit I (coxI)	<i>Plasmodium</i> sp	1	Phylogenetics (interspecies)	Martinsen et al. 2008
Apicoplast	partial caseinolytic protease C (ClpC)	<i>Plasmodium</i> sp	1	Phylogenetics (interspecies)	Martinsen et al. 2008
Nuclear	partial asl gene	<i>Plasmodium</i> sp	0	Phylogenetics (interspecies)	Martinsen et al. 2008
Nuclear	circumsporozoite protein (CSP) gene	<i>Plasmodium</i> sp	1.1	Drug design target	Plasmeyer et al. 2009
Nuclear	Merozoite surface protein 3 alpha (msp3a)	<i>Plasmodium vivax</i>	1.5	Population structure & Patient relapse	Véron et al. 2009
Nuclear	Merozoite surface protein 1 b5 (msp1b5)	<i>Plasmodium vivax</i>	0.7	Population structure & Patient relapse	Véron et al. 2009
Nuclear	Merozoite surface protein 1 (msp1)	<i>Plasmodium</i> sp	5	Phylogenetics (interspecies)	Birkenmeier et al. 2010
Nuclear	dihydrofolate reductase (pvdhfr) gene	<i>Plasmodium vivax</i>	1.9	Drug resistance	Miao et al. 2010
Nuclear	dihydropteroate synthase (pvdhps) gene	<i>Plasmodium vivax</i>	3	Drug resistance	Miao et al. 2010
Apicoplast	rns gene, small subunit rRNA (SSUrRNA)	<i>Plasmodium</i> sp	1.4	Phylogenetics (interspecies)	Mitsui et al. 2010
Apicoplast	rml gene, large subunit rRNA (LSUrRNA)	<i>Plasmodium</i> sp	2.7	Phylogenetics (interspecies)	Mitsui et al. 2010
Apicoplast	clpc gene, caseinolytic protease C (ClpC)	<i>Plasmodium</i> sp	2	Phylogenetics (interspecies)	Mitsui et al. 2010
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Culleton et al. 2011
Nuclear	Merozoite surface protein 1 (msp1)	<i>Plasmodium falciparum</i>	5	Population structure & MOI*	Mwingira et al. 2011
Nuclear	Glutamate rich protein (glurp)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Mwingira et al. 2011
Nuclear	Merozoite surface protein 2 (msp2)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Mwingira et al. 2011

cont.

Genome	Specific	Species	Size (kbp)	Function	Reference
Apicoplast	Whole ap	<i>Plasmodium ovale</i>	35	Phylogenetics (interspecies)	Arisue et al. 2012
Whole	>70% of exonic positions, 86,158 SNPs	<i>Plasmodium falciparum</i>	15,500	Population structure & Phylo (within species)	Manske et al. 2012
Nuclear	Microsatellites MS	<i>Plasmodium vivax</i>	0.1	Phylogenetics (within species)	Miao et al. 2012
Nuclear	Single nucleotide polymorphisms	<i>Plasmodium vivax</i>	6	Phylogenetics	Miao et al. 2012
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Miao et al. 2012, BMC Evol B
Whole	Whole	<i>Plasmodium vivax</i>	26,900	Comparative genome analyses	Neafsey et al. 2012
Whole	Whole	<i>Plasmodium falciparum</i>	22,900	Comparative genome analyses	Neafsey et al. 2012
Whole	Whole	<i>Plasmodium cynomolgi</i>	26,200	Comparative genome analyses	Tachibana et al. 2012
Whole	Whole	<i>Plasmodium vivax</i>	26,900	Comparative genome analyses	Tachibana et al. 2012
Whole	Whole	<i>Plasmodium knowlesi</i>	23,700	Comparative genome analyses	Tachibana et al. 2012
Nuclear	NANP	<i>Plasmodium falciparum</i>	0.1	Population structure & MOI*	Bowman et al. 2013
Nuclear	Merozoite surface protein 1 (msp1)	<i>Plasmodium falciparum</i>	5	Population structure & MOI*	Bowman et al. 2013
Nuclear	Glutamate rich protein (glurp)	<i>Plasmodium falciparum</i>	6	Population structure & MOI* & Drug target	Bowman et al. 2013
Nuclear	Merozoite surface protein 2 (msp2)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Bowman et al. 2013
Nuclear	42 microsatellite markers	<i>Plasmodium vivax</i>	0.1	Population structure & MOI*	Sutton 2013
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Taylor et al. 2013
Apicoplast	apical membrane antigen 1 (AMA1)	<i>Plasmodium vivax</i>	1.3	Population structure & Drug target	Arnott et al. 2014
Apicoplast	apical membrane antigen 1 (AMA1)	<i>Plasmodium falciparum</i>	1.3	Population structure & Drug target	Arnott et al. 2014
Nuclear	Merozoite surface protein 1 (msp1)	<i>Plasmodium falciparum</i>	5	Population structure & MOI*	Congpuong et al. 2014
Nuclear	Glutamate rich protein (glurp)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Congpuong et al. 2014
Nuclear	Merozoite surface protein 2 (msp2)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Congpuong et al. 2014
Mitochondrial	cytochrome c oxidase subunit III (cox3)	<i>Plasmodium vivax</i>	0.8	Molecular diagnostics	Isozumi et al. 2014
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Iwagami et al. 2014
Apicoplast	caseinolytic protease C (clpC) gene	<i>Plasmodium vivax</i>	2	Phylogenetics (inter- & intraspecies)	Liu et al. 2014
Nuclear	lactate dehydrogenase (ldh) gene	<i>Plasmodium vivax</i>	0.7	Phylogenetics (inter- & intraspecies)	Liu et al. 2014
Nuclear	cell division cycle 2-related kinase (crk2) gene	<i>Plasmodium vivax</i>	1.5	Genetic diversity & Phylo (inter- & intraspecies)	Liu et al. 2014
Nuclear	b-tubulin (b-tub) gene	<i>Plasmodium vivax</i>	0.7	Genetic diversity & Phylo (inter- & intraspecies)	Liu et al. 2014
Nuclear	adenylosuccinate lyase (asl) gene	<i>Plasmodium vivax</i>	0.8	Genetic diversity & Phylo (inter- & intraspecies)	Liu et al. 2014
Mitochondrial	fragment D mtDNA	<i>Plasmodium vivax</i>	2.4	Phylogenetics (inter- & intraspecies)	Liu et al. 2014
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (inter- & intraspecies)	Liu et al. 2014
Mitochondrial	Whole mt	<i>Plasmodium vivax</i>	6	Phylogenetics (inter- & intraspecies)	Liu et al. 2014
Mitochondrial	Whole mt	<i>Plasmodium falciparum</i>	6	Genetic marker for epidemiology	Preston et al. 2014
Apicoplast	Whole ap	<i>Plasmodium falciparum</i>	35	Genetic marker for epidemiology	Preston et al. 2014
Nuclear	21 genes	<i>Haemosporida</i>		Phylogenetics (interspecies)	Borner et al. 2015
Nuclear	Microsatellites MS	<i>Plasmodium vivax</i>	0.1	Population structure & Phylo (within species)	Jennison et al. 2015
Nuclear	Microsatellites MS	<i>Plasmodium vivax</i>	0.1	Population structure & Phylo (within species)	Koepfli et al. 2015
Nuclear	5' 3' regions Csp	<i>Plasmodium sp</i>	0.2	Phylogenetics (interspecies)	Prugnolle et al. 2015
Nuclear	18S gene type A	<i>Plasmodium sp</i>	0.8	Phylogenetics (interspecies)	Prugnolle et al. 2015
Mitochondrial	Whole	<i>Plasmodium sp</i>	6	Phylogenetics (interspecies)	Prugnolle et al. 2015
Whole	Whole	<i>Plasmodium vivax</i>	26,900	Population structure & Phylo & Drug resistance	Hupalo et al. 2016
Whole	Whole	<i>Plasmodium vivax</i>	541	Phylogenetics & aDNA	Gelabert et al. 2016
Whole	Whole	<i>Plasmodium falciparum</i>	1,893	Phylogenetics & aDNA	Gelabert et al. 2016
Whole	Whole	<i>Plasmodium vivax</i>	26,900	Population structure & Phylo & Drug resistance	Pearson et al. 2016
Nuclear	Merozoite surface protein 1 (msp1)	<i>Plasmodium falciparum</i>	5	Population structure & MOI*	Yavo et al. 2016
Nuclear	Merozoite surface protein 2 (msp2)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Yavo et al. 2016
Whole	Whole	<i>Plasmodium vivax</i>	26,900	Population structure & Drug resistance	Benavente et al. 2017
Whole	Whole	<i>Plasmodium vivax</i>	26,100	Population structure & Phylo (within species)	Chen et al. 2017
Nuclear	Merozoite surface protein 1 (msp1)	<i>Plasmodium falciparum</i>	5	Population structure & MOI*	Soe et al. 2017
Nuclear	Glutamate rich protein (glurp)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Soe et al. 2017
Nuclear	Merozoite surface protein 2 (msp2)	<i>Plasmodium falciparum</i>	6	Population structure & MOI*	Soe et al. 2017
Nuclear	Microsatellites MS	<i>Plasmodium vivax</i>	0.1	Population structure & MOI*	Fala et al. 2018
Mitochondrial	Whole	<i>Plasmodium vivax</i>	6	Phylogenetics (within species)	Rodrigues et al. 2018

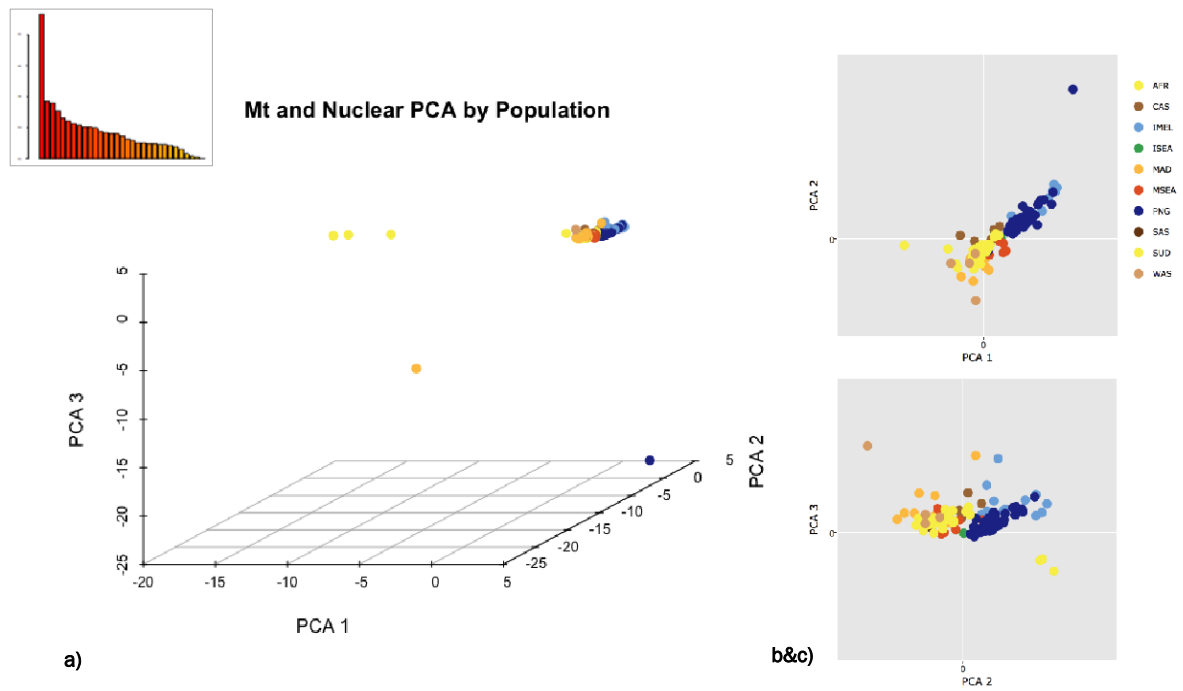


Figure S4.2

Principle Component Analyses of a) combined mitochondrial genome and nuclear microsatellite data, according to the first three dimensions. The inserted bar graph indicates the levels of resolution provided by each principle component, up to 30 dimensions. b & c) shows the PCA for the combined data according to two axes for each graph: axes 1 & 2 in b), axes 2 & 3 in c). Colours are according to geographic origin as seen in *Figure 2 of Chapter 4*.

ACKNOWLEDGEMENTS

*“Don’t look back
We have found our destiny
[...]
Forever, together”
– BTS, DNA*

I would like to start by thanking all the participants who gave their consent for sample collection, and in so doing who donated their trust, time and blood to research, often whilst in a vulnerable state. There is very little the medical world could do without these people, whose contributions are invaluable on many levels.

In addition, I would like to thank the researchers and research teams who went out of their way to collect the blood samples, on more than one occasion involving trips to dangerous and inaccessible places, and were willing to share these samples or their data with me. Such people include the following researchers and their teams: Ogobara Doumbo, Amadou Niangaly, Safiatou Niaré, Sarah Auburn *et al.*, Nhien Nguyen Thanh Thuy *et al.*, Anto Eijkman *et al.*, Carlo Severini *et al.*, Ingrid Felger, Alyssa Barry, Harin Karunaweera, Andrea Waltmann, Céline Barnadas and Cristian Koepfli. Also a special thank you to Wasan Forsyth who provided the best positive control sample I ever had. Every one of these people has been beyond helpful and supportive, and it has been a pleasure communicating and working with them.

Others whose contributions have been invaluable to this thesis and my time at Oxford have been the funding bodies who put faith and funding into my work. This includes the Oxford-Wellcome Trust ISSF (UK), School of Archaeology (Oxford, UK), Cultuurfonds (NL) including the Reiman de Bas Fonds (NL), Stichting de Fundatie van de Vrijvrouwe van Renswoude te 's-Gravenhage (NL), European Science Foundation (Europe), Merton College (UK), Meyerstein Grant (Oxford, UK), University Hardship Fund (Oxford, UK) and Hendrik Muller Fonds (NL). In addition the NHMRC (AUS) provided to Ivo Mueller helped fund part of this work.

Of course I would very much like to thank my examiners and supervisors, both official and unofficial. Examiners Profs. Deirdre Joy and Sunetra Gupta were sensible, sensitive and thorough, both during the viva voce and in their suggestions afterwards. Much thanks goes out to them for taking the time and effort to go through my thesis so thoroughly and make very helpful suggestions. Regarding supervisors, Amy Bogaard took me in when I had nowhere left to go, and has been an

unbelievable source of unending support and inspiration. She has guided me with much dependable patience, insight, understanding, questioning and robust editing skills, in a field where she was often in the dark herself. I have always felt comfortable walking into her office, and she has been one of those very special supervisors who would not only listen, but also ask, about anything that affected or could affect my studies. She also wrote endless recommendations to varying deadlines for funding applications, courses and conferences. Lastly, she was also great company at college dinners.

Ivo Mueller was another saviour in need. He adopted me into his lab in Australia when I was struggling to find support for my research in Oxford, and has been enthusiastic about the project from the very first e-mail I sent him. Our meeting can only be called serendipitous: I was looking for supervisors and a lab, and he was looking for a researcher, to do what amounted to almost exactly the same research topic. Generally his approach has been hands-off, but he was always there at times it mattered. His vast insight into many different topics has been an incredible resource, and where his expertise lacked he always found someone for me to approach.

This included my last official supervisor, Alyssa Barry. Her support has been mainly in the technicalities of population genetics, and in that she has been very helpful. She has responded quickly and with insight whenever asked for support and has been a great help in editing chapters. In addition her postdoc, now independent, Abby Harrison has been invaluable and took the form of my trainer and advisor in the lab and with much of the initial data processing and analyses. Though not on paper, Abby has been my main supervisor in several respects at the start of this thesis. She has also been a good friend, a crazy late-working desk neighbour who would supply necessities such as choosing and ordering primers, enzymes, wine, pizza and concert tickets. And a fellow rowing enthusiast, though from a rival college. She taught me not only how to work well in the lab, but also how to deal with difficult situations outside of the office. Without her, I would not have learned the tools with which to survive this thesis.

This brings me to my other lab-mates, both from the Mueller and the Barry side of the bench, in Australia. From the fantastic team of postdocs, Cristian Koepfli was another great source of learning and amusement. He taught me qPCR, supplied samples and data, kept the lab operating smoothly, threw Christmas parties for the international ‘orphans’ and dragged us to the Australian Open and Asian Football Cup. Céline Barnadas was the patient background support who could help with anything, Connie Li Wai Suen was the stats wiz and life of lunch times. Alicia Arnott was the flash of colour, productivity and positive energy any lab would be jealous to have, Zahra Razook would know a million different ways to troubleshoot any single problem in the lab, including knowing exactly how to cheer you up if you felt down or frustrated. Zahra’s seemingly endless patience with

everyone only matches my admiration for it. Though my work never overlapped with Rhea Longley's, we shared a mutual love for our weekly coffee and vegan muffin at the University of Melbourne, and I much enjoyed her company, including at outreach events.

Jess Brewster, Elisheba Malau and Rakshmei Keoh were the laboratory assistant support dream team, who kept the labs running smoothly and have each helped me out whenever I ran short on time or hands. Both Eli and my MSc student Tamarah Koleala were invaluable in processing blood samples. If not for them this project would have taken much longer to complete. And all of them were great company both in and outside of the lab.

Great company too were my fellow PhD/DPhil students and much appreciated friends, Yi Wan Quah, Camila França and Sofonias Tessema in Melbourne and Veronica Walker Vadillo, Phacha Phanomvan and Margaret Veall in Oxford. Vero, who has been with me on this journey from the start, made it all a lot of fun with not only chocolate and coffee, but intense discussions about academia and the world in general. Thank you all for sharing in the joys and pains, the chocolate and churros, the espresso martinis, the support time at Society Café, the time shared in the office and lab: I could not have done this without you!

This too goes for my friends from college. Melissa Jansen van Rensburg, Jody Reimer, Alice Baumgartner: thank you for being the best set of housemates I have had, and more. Here is hoping for some properly adventurous Holywell 4/5 reunions in the future that include more of us. To both my Anna's (Stepashova and Sanniti): words fall short to describe how much I have enjoyed and appreciated your friendship throughout. Whether for bubble tea, cocktails, coffee, shopping, walks or dinner, you were and are always the perfect companion(s). Outside of academia, thank you to José Reinders, who has shared smiles and tears, mostly from a distance. I know we both prefer having tea face-to-face, but making do with Skype and messaging has pulled us through somehow.

During the year and a half of waiting for the viva, surviving the viva, and working on my major corrections, I met more inspiring people who have supported me through and cheered me on. One of the first was Tianqi Leng: you have brought both fun and encouragement from the moment I met you. The time we had in Oxford was too short! Then Liyang Han: I don't even know how to say a proper 'thank you' to you – you are such a warm, kind, supportive friend, who was, and remains, an amazing partner in business. You bring different points of view, always important and well-considered – and the amount of care and respect I have for you is hard to put into words. My amazing cheer-leading squad – bringing sources of inspiration, new insights and unexpected wisdom, sometimes over breakfast, sometimes over dinner, sometimes over drinks, sometimes over late-night sushi: Jessamyn Chiu, Justin Ko, Linny Ng, Annie Tseng and Parth Ahya (you might not all

know each other, but you did all inspire me). Nayoon Heo and Jeonghyeok Lee for an incredible summer of recovery, joy, laughs and good food. The best hospitality from great friends. They were the first to introduce me to the rest of the Gurye family, which included the heart-warming families at Teaeut and Moooru. Last in this list from the past year is a big thanks to Trevor Lau, who has supported me with his care, generosity, trust, kindness, adventure, friendship and editing skills – against many odds. Your company and how you cheered me on, and up, were valuable beyond words.

And then of course my family, who have had to bear most of the burden passed on by a daughter and a sister going through a DPhil. My grandfather Bonpa, who was an inspiration and a support, both before and after his passing. My brother Tjeerd, and his family, who was a source of encouragement, understanding, good conversations and even helped finance some of my sanity. My sister Marjolein, who will never know how much her encouragement and admiration have helped me through. There is nothing quite like wanting to prove that you really are as cool as your younger sister thinks you are. And my parents, Pieter and Renée, who have stood by me in all ways imaginable, functioning not only as supportive and loving parents, but as soundboards, funders and proof-readers.

In the end my DPhil journey would have been nigh impossible without my partner in crime, Raoul Ribot, who has had to see me leave him for the other side of the world for long periods of time, and support me from a huge physical distance. Yet he managed to do so, not only with loving patience, but again as a proof-reader and occasional funder whilst pursuing his own tough career in academia.

Lastly, I would like to thank some people I have never met. The last year and a half of my DPhil was especially tough, but the encouraging words and examples of the ultra-hard-working musicians in the guise of G-dragon and the members of BTS have pulled me through in the end. Their music and lyrics have inspired me many times to give just a little bit more, work a little bit longer, run a little bit faster. And crawl if I could no longer walk. Together with my dance teacher Ellie Aldegheri and part of the team at 1Million Dance, they have seen me through in a better state both mentally and physically. The fact that BTS brought out an album titled ‘*Love Yourself*’ with the title track ‘*DNA*’ exactly during the uphill battle of the last six months felt like more than just coincidence. Maybe fate? Though I will be happy to change my alarm tune once handing this in. My mornings have had enough of combining every beat, whistle and letter from your *DNA* sequence with my genetic ones, in my dreams. I think I can *let go* now.

--

DNA - *By BTS*

Translation

I recognized you as soon as I saw you
As if we've called each other
The DNA in my blood is telling me
That it's you who I've been looking for

Our encounter is a mathematical formula
The law of religion, the ways of the universe
The evidence of fate given to me
You are the origin of my dreams
Take it take it
My hand held out to you is destiny

Don't worry love
Because all of this is not coincidence
We're completely different baby
Because we've found our destiny

From the day the universe was formed
Past infinite worlds
Probably in our past life and in our next lives
We'll be together forever, DNA

Because all of this is not coincidence
Because we've found our destiny
DNA

I want it this love I want it real love
I'm only focusing on you
It pulls me harder
From the start, my DNA wants you
This is fate, I love us
Only we are true lovers

Every time I see her, I am amazed
It's so strange,
my breath keeps stopping
Is this the love that I've only heard about?
Because from the start, my heart only beats towards you

Don't worry love
Because all of this is not coincidence
We're completely different baby
Because we've found our destiny

From the day the universe was formed
Past infinite worlds
Probably in our past life and in our next lives
We'll be together forever, DNA

Because all of this is not coincidence
Because we've found our destiny
DNA

Don't look back
Because we've found our destiny
Don't regret it baby
Forever
Forever
Forever
Because we're forever
together

Don't worry love
Because all of this is not coincidence
We're completely different baby
Because we've found our destiny, DNA

La la la la la
La la la la la
Because this isn't a coincidence, DNA

La la la la la
La la la la la
Because this isn't a coincidence
DNA

Korean

첫눈에 널 알아보게 됐어
서를 불러왔던 것처럼
내 혈관 속 DNA가 말해줘
내가 찾아 헤매던 너라는 걸

우리 만남은 수학의 공식
종교의 율법 우주의 섭리
내게 주어진 운명의 증거
너는 내 꿈의 출처
Take it take it
너에게 내민 내 손은 정해진 숙명

걱정하지 마 Love
이 모든 건 우연이 아니니까
우리 완전 달라 Baby
운명을 찾아낸 돌이니까

우주가 생긴 그 날부터 계속
무한의 세기를 넘어서 계속
우리 전생에도 아마 다음 생에도
영원히 함께니까 DNA

이 모든 건 우연이 아니니까
운명을 찾아낸 돌이니까
DNA

I want it this love I want it real love
난 너에게만 집중해
좀 더 세계 날 이끄네
태초의 DNA가 널 원하는데
이건 필연이야 I love us
우리만이 True lovers

그녀를 볼 때마다 소스라치게 놀라
신기하게 자꾸만 숨이 멎는 게
참 이상해 설마
이런 게 말로만 들던 사랑이란 감정일까
애초부터 내 심장은 널 향해 뛰니까

걱정하지 마 Love
이 모든 건 우연이 아니니까
우리 완전 달라 Baby
운명을 찾아낸 돌이니까

우주가 생긴 그 날부터 계속
무한의 세기를 넘어서 계속
우리 전생에도 아마 다음 생에도
영원히 함께니까 DNA

이 모든 건 우연이 아니니까
운명을 찾아낸 돌이니까
DNA

돌아보지 말아
운명을 찾아낸 우리니까
후회하지 말아 Baby
영원히
영원히
영원히
영원히
함께니까

걱정하지 마 Love
이 모든 건 우연이 아니니까
우리 완전 달라 Baby
운명을 찾아낸 돌이니까, DNA

La la la la la
La la la la la
우연이 아니니까, DNA

La la la la la
La la la la la
우연이 아니니까
DNA

Romanised Korean

cheonnune neol araboge dwaesseo
seorol bulleowatteon geotcheoreom
nae hyeolgwan sok DNAGA malhaejweo
naega chaja hemaedeon neoraneun geol

uri mannameun suhage gongshik
jonggyoye yulbeob ujuye seobri
naege jueojin unmyeonge jeunggeo
neoneun nae kkume chulcheo
Take it take it
neoege naemin nae soneun jeonghaejin sungmyeong

geokjeonghaji ma Love
i modeun geon uyeoni aninikka
urin wanjeon dalla Baby
unmyeongeul chajanaen durinikka

ujuga saenggin geu nalbuteo gyesok
muhane segireul neomeoseo gyesok
urin jeonsaengedo ama daeum saengedo
yeongweoni hamkkenikka DNA

i modeun geon uyeoni aninikka
unmyeongeul chajanaen durinikka
DNA

I want it this love I want it real love
nan neogeman jibjunghae
jom deo sege nal ikkeune
taechoye DNAGA neol weonaneunde
igeon piryeoniya I love us
urimani True lovers

geunyeoreul bol ttaemada souseurachige nolla
shingihage jakkuman sumi meonneun ge
cham isanghae seolma
ireon ge malloman deutteon sarangiran gamjeongikka
aechobuteo nae shimjangeun neol hyanghae ttwinikka

geokjeonghaji ma Love
i modeun geon uyeoni aninikka
urin wanjeon dalla Baby
unmyeongeul chajanaen durinikka

ujuga saenggin geu nalbuteo gyesok
muhane segireul neomeoseo gyesok
urin jeonsaengedo ama daeum saengedo
yeongweoni hamkkenikka DNA

i modeun geon uyeoni aninikka
unmyeongeul chajanaen durinikka
DNA

doraboji mara
unmyeongeul chajanaen urinikka
huhoehaji mara Baby
yeongweoni
yeongweoni
yeongweoni
yeongweoni
hamkkenikka

geokjeonghaji ma Love
i modeun geon uyeoni aninikka
urin wanjeon dalla Baby
unmyeongeul chajanaen durinikka, DNA

La la la la la
La la la la la
uyeoni aninikka, DNA

La la la la la
La la la la la
uyeoni aninikka
DNA

REFERENCES & BIBLIOGRAPHY

- Abdullah, N.R., Barber, B.E., William, T., Norahmad, N.A., Satsu, U.R., Muniandy, P.K., Ismail, Z., Grigg, M.J., Jelip, J., Piera, K., et al. (2013). *Plasmodium vivax* population structure and transmission dynamics in Sabah Malaysia. *PLoS ONE* 8, e82553.
- Achtman, M. (2012). Insights from genomic comparisons of genetically monomorphic bacterial pathogens. *Phil Trans R Soc B* 367, 860–867.
- Agostini, H.T., Yanagihara, R., Davis, V., Ryschkewitsch, C.F., and Stoner, G.L. (1997). Asian genotypes of JC virus in Native Americans and in a Pacific Island population: Markers of viral evolution and human migration. *PNAS* 94, 14542–14546.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans Automat Contr* 19, 716–723.
- Albsheer, M.M.A., Pestana, K., Ahmed, S., Elfaki, M., Gamil, E., Ahmed, S.M., Ibrahim, M.E., Musa, A.M., Lo, E., and Hamid, M.M.A. (2019). Distribution of Duffy phenotypes among *Plasmodium vivax* infections in Sudan. *Genes* 10, 437.
- Alekseyenko, A.V., Lee, C.J., and Suchard, M.A. (2008). Wagner and Dollo: A stochastic duet by composing two parsimonious solos. *Syst Biol* 57, 772–784.
- Alilio, M.S., Mills, A., and Breman, J.G. (2004). Conquering the intolerable burden of malaria: What's new, what's needed: A summary. *Am J Trop Med Hyg* 71, 1–15.
- Altmann, D.M., Balloux, F., and Boyton, R.J. (2012). Diverse approaches to analysing the history of human and pathogen evolution: how to tell the story of the past 70 000 years. *Phil Trans R Soc B* 367, 765–769.
- Amore, G., Bertolotti, L., Hamer, G.L., Kitron, U.D., Walker, E.D., Ruiz, M.O., Brawn, J.D., and Goldberg, T.L. (2010). Multi-year evolutionary dynamics of West Nile virus in suburban Chicago, USA, 2005–2007. *Phil Trans R Soc B* 365, 1871–1878.
- Andersen, K.G., Shylakhter, I., Tabrizi, S., Grossman, S.R., Happi, C.T., and Sabeti, P.C. (2012). Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. *Phil Trans R Soc B* 367, 868–877.
- Anderson, A., Clark, G.R., Leach, F., and O'Connor, S. (2008). Islands of inquiry: colonisation, seafaring and the archaeology of maritime landscapes (Canberra, A.C.T.: ANU E Press).
- Anderson, Atholl (2009). The rat and the octopus: Initial human colonization and the prehistoric introduction of domestic animals to Remote Oceania. *Biol Invasions* 11, 1503–1519.
- Andriopoulos, P., Economopoulou, A., Spanakos, G., and Assimakopoulos, G. (2013). A local outbreak of autochthonous *Plasmodium vivax* malaria in Laconia, Greece—a re-emerging infection in the southern borders of Europe? *Int J Infect Dis* 17, e125–e128.
- Aplin, K.P., Suzuki, H., Chinen, A.A., Chesser, R.T., ten Have, J., Donnellan, S.C., Austin, J., Frost, A., Gonzalez, J.P., Herbreteau, V., et al. (2011). Multiple geographic origins of commensalism and complex cispersal history of Black Rats. *PLoS ONE* 6, e26357.
- Arisue, N., and Hashimoto, T. (2015). Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitol Int* 64, 254–259.
- Arisue, N., Hashimoto, T., Mitsui, H., Palacpac, N.M.Q., Kaneko, A., Kawai, S., Hasegawa, M., Tanabe, K., and Horii, T. (2012). The *Plasmodium* apicoplast genome: Conserved structure and close relationship of *P. ovale* to rodent malaria parasites. *Mol Biol Evol* 29, 2095–2099.
- Armitage, S.J., Jasim, S.A., Marks, A.E., Parker, A.G., Usik, V.I., and Uerpmann, H.P. (2011). The southern route “Out of Africa”: Evidence for an early expansion of modern humans into Arabia. *Science* 331, 453–456.
- Arnott, A., Barnadas, C., Senn, N., Siba, P., Mueller, I., Reeder, J.C., and Barry, A.E. (2013). High genetic diversity of *Plasmodium vivax* on the north coast of Papua New Guinea. *Am J Trop Med Hyg* 89, 188–194.
- Arnott, A., Wapling, J., Mueller, I., Ramsland, P.A., Siba, P.M., Reeder, J.C., and Barry, A.E. (2014). Distinct patterns of diversity, population structure and evolution in the AMA1 genes of sympatric *Plasmodium falciparum* and *Plasmodium vivax* populations of Papua New Guinea from an area of similarly high transmission. *Malar J* 13, 233.
- Atkinson, Q.D., Gray, R.D., and Drummond, A.J. (2008). MtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. *Mol Biol Evol* 25, 468–474.
- Atkinson, Q.D., Gray, R.D., and Drummond, A.J. (2009). Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc Royal Soc B* 276, 367–373.
- Attenborough, R.D., and Alpers, M.P. (1992). *Human Biology in Papua New Guinea: The Small Cosmos*. (Oxford, UK: Clarendon Press).
- Attenborough, R.D., Burkot, T.R., and Gardner, D.S. (1997). Altitude and the risk of bites from mosquitoes infected with malaria and filariasis among the Mianmin people of Papua New Guinea. *Trans Roy Soc Trop Med Hyg* 91, 8–10.
- Auburn, S., and Barry, A.E. (2016). Dissecting malaria biology and epidemiology using population genetics and genomics. *Int J Parasitol*.
- Austin, C.C. (1999). Lizards took express train to Polynesia. *Nature* 397, 113–114.
- Ayala, F.J., and Coluzzi, M. (2005). Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *PNAS* 102 Suppl 1, 6535–6542.
- Bacon, A.-M., Demeter, F., Roussé, S., Long, V.T., Düringer, P., Antoine, P.-O., Thuy, N.K., Mai, B.T., Huong, N.T.M., Dodo, Y., et al. (2006). New palaeontological assemblage, sedimentological and chronological data from the Pleistocene Ma U’Oi cave (northern Vietnam). *Palaeogeogr Palaeoclimatol Palaeoecol* 230, 280–298.
- Bacon, A.-M., Demeter, F., Düringer, P., Helm, C., Bano, M., Vu The Long, Kim Thuy, N.T., Antoine, P.-O., Thi Mai, B., Huong, N.T.M., et al. (2008). The Late Pleistocene Duoi U’Oi cave in northern Vietnam: Palaeontology, sedimentology, taphonomy and palaeoenvironments. *Quat Sci Rev* 27, 1627–1654.
- Bae, C.J., Wang, W., Zhao, J., Huang, S., Tian, F., and Shen, G. (2014). Modern human teeth from Late Pleistocene Luna Cave (Guangxi, China). *Quatern Int* 354, 169–183.

- Bae, C.J., Douka, K., and Petraglia, M.D. (2017). On the origin of modern humans: Asian perspectives. *Science* 358, eaai9067.
- Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., and Alekseyenko, A.V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 29, 2157–2167.
- Baird, J.K., Basri, H., Weina, P., Maguire, J.D., Barcus, M.J., Picarema, H., Elyazar, I.R.F., Ayomi, E., and Sekartuti (2003). Adult Javanese migrants to Indonesian Papua at high risk of severe disease caused by malaria. *Epidemiol Infect* 131, 791–797.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376.
- Balkenhol, N., and Waits, L.P. (2009). Molecular road ecology: exploring the potential of genetics for investigating transportation impacts on wildlife. *Mol Ecol* 18, 4151–4164.
- Ballard, C. (2003). Writing (pre)history: narrative and archaeological explanation in the New Guinea Highlands. *Arch Oceania* 38, 135–148.
- Ballard, J.W., and Dean, M.D. (2001). The mitochondrial genome: mutation, selection and recombination. *Curr Opin Gen Dev* 11, 667–672.
- Balloux, F., Handley, L.-J.L., Jombart, T., Liu, H., and Manica, A. (2009). Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proc Royal Soc B* 276, 3447–3455.
- Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16, 37–48.
- Baranova, A., Sergiev, V., Morozova, L., Turbabina, N., and Morozov, E. (2019). Imported *Plasmodium vivax* malaria in the Russian Federation from Western Sub-Saharan Africa. *J Trop Med* 2019, 1–5.
- Barber, P.H., Palumbi, S.R., Erdmann, M.V., and Moosa, M.K. (2000). A marine Wallace's line?: Biogeography. *Nature* 406, 692–693.
- Barnadas, C., Timinao, L., Javati, S., Iga, J., Malau, E., Koepfli, C., Robinson, L.J., Senn, N., Kiniboro, B., Rare, L., et al. (2015). Significant geographical differences in prevalence of mutations associated with *Plasmodium falciparum* and *Plasmodium vivax* drug resistance in two regions from Papua New Guinea. *Malar J* 14, 399.
- Baron, J.M., Higgins, J.M., and Dzik, W.H. (2011). A revised timeline for the origin of *Plasmodium falciparum* as a human pathogen. *J Mol Evol* 73, 297–304.
- Barracough, T.G., and Nee, S. (2001). Phylogenetics and speciation. *Trends Ecol Evol* 16, 391–399.
- Barry, A.E., Leliwa, A., Choi, M., Nielsen, K.M., Hartl, D.L., and Day, K.P. (2003). DNA sequence artifacts and the estimation of time to the most recent common ancestor (TMRCA) of *Plasmodium falciparum*. *Mol Biochem Parasitol* 130, 143–147.
- Barta, J.R. (1989). Phylogenetic analysis of the Class Sporozoa (Phylum Apicomplexa Levine, 1970): Evidence for the independent evolution of heteroxenous life cycles. *J Parasitol* 75, 195.
- Barta, J.R. (2001). Molecular approaches for inferring evolutionary relationships among protistan parasites. *Vet Parasitol* 101, 175–186.
- Battle, K.E., Karhunen, M.S., Bhatt, S., Gething, P.W., Howes, R.E., Golding, N., Van Boeckel, T.P., Messina, J.P., Shanks, G., Smith, D.L., et al. (2014). Geographical variation in *Plasmodium vivax* relapse. *Malar J* 13, 144.
- Bayley, R., Cameron, R., and Lucas, C. (2013). *The Oxford Handbook of Sociolinguistics* (Oxford, UK: Oxford University Press).
- Becker, N., Petric, D., Zgomba, M., Boase, C., Dahl, C., Lane, J., and Kaiser, A. (2003). *Mosquitoes and their control* (New York, Boston, Dordrecht, London, Moscow: Kluwer Academic/Plenum Publishers).
- Bedford, T., and Hartl, D.L. (2009). Optimization of gene expression by natural selection. *PNAS* 106, 1133–1138.
- Beebe, N.W., and Cooper, R.D. (2002). Distribution and evolution of the *Anopheles punctulatus* group (Diptera: Culicidae) in Australia and Papua New Guinea. *Int J Parasitol* 32, 563–574.
- Beebe, N.W., Cooper, R.D., Morrison, D.A., and Ellis, J.T. (2000a). A phylogenetic study of the *Anopheles punctulatus* group of malaria vectors comparing rDNA sequence alignments derived from the mitochondrial and nuclear small ribosomal subunits. *Mol Phylogenetics Evol* 17, 430–436.
- Beebe, N.W., Cooper, R., Foley, D.H., and Ellis, J.T. (2000b). Populations of the south-west Pacific malaria vector *Anopheles farauti* s.s. revealed by ribosomal DNA transcribed spacer polymorphisms. *Heredity* 84, 244–253.
- Beeson, J.G., and Duffy, P.E. (2005). The immunology and pathogenesis of malaria during pregnancy. *Curr Top Microbiol Immunol* 297, 187–227.
- Beeson, J.G., Osier, F.H.A., and Engwerda, C.R. (2008). Recent insights into humoral and cellular immune responses against malaria. *Trends Parasitol* 24, 578–584.
- Behar, D.M., Vilems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., et al. (2008). The dawn of human matrilineal diversity. *Am J Hum Genet* 82, 1130–1140.
- Behbahani, A., Dutton, T.J., Raju, A.K., Townson, H., and Sinkins, S.P. (2004). Polymorphic microsatellite loci in the mosquito *Aedes polynesiensis*. *Mol Ecol Notes* 4, 59–61.
- Belaich, A., and Belaich, J.P. (1976). Microcalorimetric study of the anaerobic growth of *Escherichia coli*: growth thermograms in a synthetic medium. *J Bacteriol* 125, 14–18.
- Belinky, F., Cohen, O., and Huchon, D. (2010). Large-scale parsimony analysis of Metazoan indels in protein-coding genes. *Mol Biol Evol* 27, 441–451.
- Belkin, John N. (1962a). *The mosquitoes of the South Pacific (Diptera, Culicidae)* (Berkeley: University of California Press).
- Belkin, John N. (1962b). *The mosquitoes of the South Pacific (Diptera, Culicidae)* (Berkeley: University of California Press).
- Belle, E.M.S., Piganeau, G., Gardner, M., and Eyre-Walker, A. (2005). An investigation of the variation in the transition bias among various animal mitochondrial DNA. *Gene* 355, 58–66.

- Bellwood, P. (1994). An archaeologist's view of language macrofamily relationships. *Ocean Linguist* 33, 391–406.
- Bellwood, P. (2001). Early agriculturalist population diasporas? Farming, languages, and genes. *Annu Rev Anthropol* 30, 181–207.
- Bellwood, P. (2005). Domesticated pigs in Eastern Indonesia. *Science* 309, 381a–381a.
- Bellwood, P. (2009). The dispersals of established food-producing populations. *Curr Anthropol* 50, 621–626.
- Bellwood, P. (2011). Holocene population history in the Pacific region as a model for worldwide food producer dispersals. *Curr Anthropol* 52, S363–S378.
- Bellwood, P., and Diamond, J. (2005). On explicit “replacement” models in Island Southeast Asia: a reply to Stephen Oppenheimer. *World Archaeology* 37, 503–506.
- Bellwood, P., and Glover, I. (2006). Southeast Asia: foundations for an archaeological history. In *Southeast Asia: From Prehistory to History*, I. Glover, and P. Bellwood, eds. (New York, USA: RoutledgeCurzon), p.
- Bellwood, P., and Sanchez-Mazas, A. (2005). Human migrations in continental East Asia and Taiwan: Genetic, linguistic, and archaeological evidence. *Curr Anthropol* 46, 480–484.
- Bellwood, P., Chambers, G., Ross, M., and Hung, H. (2011). Are ‘cultures’ inherited? Multidisciplinary perspectives on the origins and migrations of Austronesian-speaking peoples prior to 1000 BC. In *Investigating Archaeological Cultures*, B.W. Roberts, and M.V. Linden, eds. (Springer New York), pp. 321–354.
- Bellwood, P.S., Fox, J.J., and Tryon, D.T. (1995). *The Austronesians: Historical and Comparative Perspectives* (Canberra, Australia: Dept. of Anthropology, Research School of Pacific and Asian Studies, ANU).
- Benavente, E.D., Ward, Z., Chan, W., Mohareb, F.R., Sutherland, C.J., Roper, C., Campino, S., and Clark, T.G. (2017). Genomic variation in *Plasmodium vivax* malaria reveals regions under selective pressure. *PLoS ONE* 12, e0177134.
- Benedict, M.Q., Levine, R.S., Hawley, W.A., and Lounibos, L.P. (2007). Spread of the tiger: Global risk of invasion by the mosquito *Aedes albopictus*. *Vector Borne Zoonotic Dis* 7, 76–85.
- Bensch, S., Pérez-Tris, J., Waldenström, J., and Hellgren, O. (2004). Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? *Evolution* 58, 1617–1621.
- Bensch, S., Hellgren, O., Križanauskienė, A., Palinauskas, V., Valkiūnas, G., Outlaw, D., and Ricklefs, R.E. (2013). How can we determine the molecular clock of malaria parasites? *Trends Parasitol* 29, 363–369.
- Bethke, L.L., Zilversmit, M., Nielsen, K., Daily, J., Volkman, S.K., Ndiaye, D., Lozovsky, E.R., Hartl, D.L., and Wirth, D.F. (2006). Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of *Plasmodium falciparum*. *Mol Biochem Parasitol* 150, 10–24.
- Beugin, M.-P., Gayet, T., Pontier, D., Devillard, S., and Jombart, T. (2018). A fast likelihood solution to the genetic clustering problem. *Meth Ecol Evol*.
- Bhatia, G., Patterson, N., Sankararaman, S., and Price, A.L. (2013). Estimating and interpreting FST: The impact of rare variants. *Genome Res* 23, 1514–1521.
- Bhatt, S., Katzourakis, A., and Pybus, O.G. (2010). Detecting natural selection in RNA virus populations using sequence summary statistics. *Infect Genet Evol* 10, 421–430.
- Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K.E., Moyes, C.L., Henry, A., Eckhoff, P.A., et al. (2015). The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* 526, 207–211.
- Blackburn, T.M. (2005). Response to comment on “Avian extinction and mammalian introductions on oceanic islands.” *Science* 307, 1412b–1412b.
- Blackburn, T.M., and Duncan, R.P. (2001). Determinants of establishment success in introduced birds. *Nature* 414, 195–197.
- Blackburn, T.M., Cassey, P., Duncan, R.P., Evans, K.L., and Gaston, K.J. (2004). Avian extinction and mammalian introductions on oceanic islands. *Science* 305, 1955–1958.
- Blackburn, T.M., Cassey, P., and Lockwood, Julie.L. (2008). The island biogeography of exotic bird species. *Glob Ecol Biogeogr* 17, 246–251.
- Blasco, B., Leroy, D., and Fidock, D.A. (2017). Antimalarial drug resistance: linking *Plasmodium falciparum* parasite biology to the clinic. *Nat Med* 23, 917–928.
- Blench, R. (2011). Was there an Austroasiatic presence in Island Southeast Asia prior to the Austronesian expansion? *BIPPA* 30, 133–144.
- Blench, R. (2014). Origins of ethnolinguistic identity in Southeast Asia. In *Handbook of East and Southeast Asian Archaeology*, J. Habu, P. Lape, J. Olsen, and J. Zhichun, eds. (USA: Springer), p.
- Blench, Roger (2004). Ancient connections between insular S.E. Asia and West Africa in the light of ethnobotanical and other evidence.
- Blinkhorn, J., and Petraglia, M.D. (2017). Environments and Cultural Change in the Indian Subcontinent: Implications for the Dispersal of *Homo sapiens* in the Late Pleistocene. *Curr Anthropol* 58, S463–S479.
- Bonhomme, F., Orth, A., Cucchi, T., Rajabi-Maham, H., Catalan, J., Boursot, P., Auffray, J.-C., and Britton-Davidian, J. (2011). Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proc Royal Soc B* 278, 1034–1043.
- Bopp, S.E.R., Manary, M.J., Bright, A.T., Johnston, G.L., Dharia, N.V., Luna, F.L., McCormack, S., Plouffe, D., McNamara, C.W., Walker, J.R., et al. (2013). Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. *PLoS Genet* 9, e1003293.
- Borner, J., Pick, C., Thiede, J., Kolawole, O.M., Kingsley, M.T., Schulze, J., Cottontail, V.M., Wellinghausen, N., Schmidt-Chanasit, J., Bruchhaus, I., et al. (2016). Phylogeny of haemosporidian blood parasites revealed by a multi-gene approach. *Mol Phylogenetics Evol* 94, 221–231.
- Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., et al. (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478, 506–510.

- Bosssdorf, O., Auge, H., Lafuma, L., Rogers, W.E., Siemann, E., and Prati, D. (2005). Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* 144, 1–11.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10, e1003537.
- Bousema, J.T., Roeffen, W., van der Kolk, M., de Vlas, S.J., van de Vegte-Bolmer, M., Bangs, M.J., Teelen, K., Kurniawan, L., Maguire, J.D., Baird, J.K., et al. (2006). Rapid onset of transmission-reducing antibodies in Javanese migrants exposed to malaria in Papua, Indonesia. *Am J Trop Med Hyg* 74, 425–431.
- Bouwman, A., and Rühli, F. (2016). Archaeogenetics in evolutionary medicine. *J Mol Med*.
- Bowman, N.M., Congdon, S., Mvalo, T., Patel, J.C., Escamilla, V., Emch, M., Martinson, F., Hoffman, I., Meshnick, S.R., and Juliano, J.J. (2013). Comparative population structure of *Plasmodium falciparum* circumsporozoite protein NANP repeat lengths in Lilongwe, Malawi. *Sci Rep* 3.
- Boyd, Mark F (1949). *Malaria: a comprehensive survey of all aspects of this group of diseases from a global standpoint* (Philadelphia: Saunders).
- Brabin, L., Brabin, B.J., Dimitrakakis, M., and Gust, I. (1989). Factors affecting the prevalence of infection with hepatitis B virus among non-pregnant women in the Alexishafen area of Papua New Guinea. *Ann Trop Med Parasitol* 83, 365–374.
- Brace, C.L. (2000). *Evolution in an anthropological view: collected essays* (Rowman & Littlefield).
- Bright, A., Tewhey, R., Abeles, S., Chuquiyauri, R., Llanos-Cuentas, A., Ferreira, M.U., Schork, N.J., Vinetz, J.M., and Winzeler, E.A. (2012). Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics* 13, 262.
- Britton-Davidian, J., Catalan, J., Lopez, J., Ganem, G., Nunes, A.C., Ramalhinho, M.G., Auffray, J.C., Searle, J.B., and Mathias, M.L. (2007). Patterns of genic diversity and structure in a species undergoing rapid chromosomal radiation: an allozyme analysis of house mice from the Madeira archipelago. *Heredity* 99, 432–442.
- Brockmann, D. (2010). Human Mobility and Spatial Disease Dynamics. In *Reviews of Nonlinear Dynamics and Complexity*, (Wiley-VCH Verlag GmbH & Co. KGaA), pp. 1–24.
- Brockmann, D., and Theis, F. (2008). Money circulation, trackable items, and the emergence of universal human mobility patterns. *IEEE Pervasive Comput* 7, 28–35.
- Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature* 439, 462–465.
- Bromham, L., and Woolfit, M. (2004). Explosive radiations and the reliability of molecular clocks: island endemic radiations as a test case. *Syst Biol* 53, 758–766.
- Brooks, D.R., and McLennan, D.A. (1992). The evolutionary origin of *Plasmodium falciparum*. *J Parasitol* 78, 564.
- Brown, J.H., and Sax, D.F. (2004). An essay on some topics concerning invasive species. *Austral Ecol* 29, 530–536.
- Brown, J.H., and Sax, D.F. (2005). Biological invasions and scientific objectivity: reply to Cassey et al. (2005). *Austral Ecol* 30, 481–483.
- Brown, J.E., McBride, C.S., Johnson, P., Ritchie, S., Paupy, C., Bossin, H., Lutomia, J., Fernandez-Salas, I., Ponlawat, A., Cornel, A.J., et al. (2011). Worldwide patterns of genetic differentiation imply multiple “domestications” of *Aedes aegypti*, a major vector of human diseases. *Proc Royal Soc B* 278, 2446–2454.
- Brown, K.M., DePristo, M.A., Weinreich, D.M., and Hartl, D.L. (2009). Temporal constraints on the incorporation of regulatory mutants in evolutionary pathways. *Mol Biol Evol* 26, 2455–2462.
- Bruce, M.C., Macheso, A., Kelly-Hope, L.A., Nkhoma, S., McConnachie, A., and Molyneux, M.E. (2008). Effect of transmission setting and mixed species infections on clinical measures of malaria in Malawi. *PLoS ONE* 3, e2775.
- Bruce-Chwatt, L.J., and de Zulueta, J. (1980). *The Rise and Fall of malaria in Europe. A Historico-Epidemiological Study* (Oxford: Oxford University Press).
- Bryant, J.E., Holmes, E.C., and Barrett, A.D.T. (2007). Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog* 3, e75.
- Bustamante, C.D., Nielsen, R., and Hartl, D.L. (2003). Maximum likelihood and Bayesian methods for estimating the distribution of selective effects among classes of mutations using DNA polymorphism data. *Theor Popul Biol* 63, 91–103.
- Byrne, J.P. (2008). *Encyclopedia of Pestilence, Pandemics, and Plagues (ABC-CLIO)*.
- Byrne, M., Steane, D.A., Joseph, L., Yeates, D.K., Jordan, G.J., Crayn, D., Aplin, K., Cantrill, D.J., Cook, L.G., Crisp, M.D., et al. (2011). Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *J Biogeogr* 38, 1635–1656.
- Camerini, J.R. (1993). Evolution, biogeography, and maps: An early history of Wallace’s Line. *Isis* 84, 700–727.
- Campagna, L., Benites, P., Loughheed, S.C., Lijtmaer, D.A., Di Giacomo, A.S., Eaton, M.D., and Tubaro, P.L. (2011). Rapid phenotypic evolution during incipient speciation in a continental avian radiation. *Proc Royal Soc B*.
- Campo, B., Vandal, O., Wesche, D.L., and Burrows, J.N. (2015). Killing the hypnozoite – drug discovery approaches to prevent relapse in *Plasmodium vivax*. *Pathog Glob Health* 109, 107–122.
- Camuffo, D., and Jones, P.D. (2002). *Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources* (Kluwer Academic Publisher: Dordrecht/Boston/London).
- Cann, H.M. (2002). A human genome diversity cell line panel. *Science* 296, 261b–2262.
- Carlson, J.S., Martínez-Gómez, J.E., Cornel, A., Loiseau, C., and Sehgal, R.N.M. (2011). Implications of plasmodium parasite infected mosquitoes on an insular avifauna: The case of Socorro Island, México. *J Vect Ecol* 36, 213–220.
- Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Perte, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., Selengut, J.D., Koo, H.L., et al. (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419, 512–519.
- Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S.V., Merino, E.F., Amedeo, P., et al. (2008). Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455, 757–763.
- Carlton, J.M., Sina, B.J., and Adams, J.H. (2011). Why is *Plasmodium vivax* a neglected tropical disease? *PLoS Negl Trop Dis* 5, e1160.

- Carlton, J.M., Perkins, S.L., and Deitsch, K.W. (2013). *Malaria Parasites: Comparative Genomics, Evolution and Molecular Biology* (Norfolk, UK: Caister Academic Press).
- Carpenter, M.L., Buenrostro, J.D., Valdiosera, C., Schroeder, H., Allentoft, M.E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., et al. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *The American Journal of Human Genetics* 93, 852–864.
- Carranza, S., and Arnold, E.N. (2006). Systematics, biogeography, and evolution of Hemidactylus geckos (Reptilia: Gekkonidae) elucidated using mitochondrial DNA sequences. *Mol Phylogenetics Evol* 38, 531–545.
- Carter, R. (2003). Speculations on the origins of *Plasmodium vivax* malaria. *Trends Parasitol* 19, 214–219.
- Carter, R., and Mendis, K.N. (2002). Evolutionary and historical aspects of the burden of malaria. *Clin Microbiol Rev* 15, 564–594.
- Cassey, P., Blackburn, T.M., Sol, D., Duncan, R.P., and Lockwood, J.L. (2004). Global patterns of introduction effort and establishment success in birds. *Proc Royal Soc B* 271, S405–S408.
- Cassey, P., Blackburn, T.M., Duncan, R.P., and Gaston, K.J. (2005a). Causes of exotic bird establishment across oceanic islands. *Proc Royal Soc B* 272, 2059–2063.
- Cassey, P., Blackburn, T.M., Duncan, R.P., and Chown, S.L. (2005b). Concerning invasive species: Reply to Brown and Sax. *Austral Ecol* 30, 475–480.
- Cavalli-Sforza, L.L., and Edwards, A.W.F. (1967). Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 19, 233–257.
- Cayres-Vallinoto, I.M.V., Vallinoto, A.C.R., Azevedo, V.N., Machado, L.F.A., Ishak, M. de O.G., and Ishak, R. (2012). Human JCV infections as a bio-anthropological marker of the formation of Brazilian Amazonian populations. *PLoS ONE* 7, e46523.
- Chang, H.-H., Park, D.J., Galinsky, K.J., Schaffner, S.F., Ndiaye, D., Ndir, O., Mboup, S., Wiegand, R.C., Volkman, S.K., Sabeti, P.C., et al. (2012). Genomic sequencing of *Plasmodium falciparum* malaria parasites from Senegal reveals the demographic history of the population. *Mol Biol Evol* 29, 3427–3439.
- Chappell, J. (1993). Late Pleistocene coasts and human migrations in the Austral region. In *A Community of Culture: The People and Prehistory of the Pacific*, (Canberra, Australia: Dept. of Prehistory, Research School of Pacific Studies, ANU), pp. 43–48.
- Charlwood, J.D., Birley, M.H., Dagoro, H., Paru, R., and Holmes, P.R. (1985). Assessing survival rates of *Anopheles farauti* (Diptera: Culicidae) from Papua New Guinea. *J Anim Ecol* 54, 1003–1016.
- Chaubey, G., Metspalu, M., Choi, Y., Mägi, R., Romero, I.G., Soares, P., Oven, M. van, Behar, D.M., Rootsi, S., Hudjashov, G., et al. (2010). Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. *Mol Biol Evol* 28, 1013–1024.
- Chen, R., and Vasilakis, N. (2011). Dengue — Quo tu et quo vadis? *Viruses* 3, 1562–1608.
- Chen, B., Harbach, R.E., and Butlin, R.K. (2004). Genetic variation and population structure of the mosquito *Anopheles jeyporiensis* in southern China. *Mol Ecol* 13, 3051–3056.
- Chen, B., Pedro, P.M., Harbach, R.E., Somboon, P., Walton, C., and Butlin, R.K. (2011). Mitochondrial DNA variation in the malaria vector *Anopheles minimus* across China, Thailand and Vietnam: evolutionary hypothesis, population structure and population history. *Heredity* 106, 241–252.
- Chen, S.-B., Wang, Y., Kassegne, K., Xu, B., Shen, H.-M., and Chen, J.-H. (2017). Whole-genome sequencing of a *Plasmodium vivax* clinical isolate exhibits geographical characteristics and high genetic variation in China-Myanmar border area. *BMC Genomics* 18.
- Chenet, S.M., Tapia, L.L., Escalante, A.A., Durand, S., Lucas, C., and Bacon, D.J. (2012). Genetic diversity and population structure of genes encoding vaccine candidate antigens of *Plasmodium vivax*. *Malar J* 11, 68.
- Chookajorn, T., Costanzo, M.S., Hartl, D.L., and Deitsch, K.W. (2007a). Malaria: a peek at the var variorum. *Trends Parasitol* 23, 563–565.
- Chookajorn, T., Dzikiowski, R., Frank, M., Li, F., Jiwani, A.Z., Hartl, D.L., and Deitsch, K.W. (2007b). Epigenetic memory at malaria virulence genes. *PNAS* 104, 899–902.
- Chu, H., Li, C., Guo, X., Zhang, H., Luo, P., Wu, Z., Wang, G., and Zhao, T. (2018). The phylogenetic relationships of known mosquito (Diptera: Culicidae) mitogenomes. *Mitochondrial DNA* 29, 31–35.
- Clark, J.T., and Kelly, K.M. (1993). Human genetics, paleoenvironments, and malaria: Relationships and implications for the settlement of Oceania. *Am Anthropol* 95, 612–630.
- Coimbra, C.E.A. (1988). Human settlements, demographic pattern, and epidemiology in lowland Amazonia: The case of Chagas's disease. *Am Anthropol* 90, 82–97.
- Colizza, V., and Vespignani, A. (2010). The flu fighters. *Phys World* 23, 26–30.
- Coluzzi, M., Sabatini, A., Torre, A. della, Deco, M.A.D., and Petrarca, V. (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* 298, 1415–1418.
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet advance online publication*.
- Congpuong, K., Sukaram, R., Prompan, Y., and Dornae, A. (2014). Genetic diversity of the msp-1, msp-2, and glurp genes of *Plasmodium falciparum* isolates along the Thai-Myanmar borders. *Asian Pac J Trop Biomed* 4, 598–602.
- Consortium, T.H.P.-A.S. (2009). Mapping human genetic diversity in Asia. *Science* 326, 1541–1545.
- Conway, D.J. (2003). Tracing the dawn of *Plasmodium falciparum* with mitochondrial genome sequences. *Trends Genet* 19, 671–674.
- Conway, D.J. (2007). Molecular epidemiology of malaria. *Clin Microbiol Rev* 20, 188–204.
- Cooper, R.D., Waterson, D.G.E., Frances, S.P., Beebe, N.W., and Sweeney, A.W. (2002). Speciation and distribution of the members of the *Anopheles punctulatus* (Diptera: Culicidae) Group in Papua New Guinea. *J Med Entomol* 39, 16–27.
- Cordaux, R., and Stoneking, M. (2003). South Asia, the Andamanese, and the Genetic Evidence for an “Early” Human Dispersal out of Africa. *Am J Hum Genet* 72, 1586–1590.

- Cormier, L.A. (2011). *The Ten-Thousand Year Fever: Rethinking Human and Wild Primate Malaras* (Walnut Creek, USA: Left Coast Press).
- Cornejo, O.E., and Escalante, A.A. (2006). The origin and age of *Plasmodium vivax*. *Trends Parasitol* 22, 558–563.
- Cowman, A.F., Healer, J., Marapana, D., and Marsh, K. (2016). Malaria: Biology and disease. *Cell* 167, 610–624.
- Cox, M.P., Karafet, T.M., Lansing, J.S., Sudoyo, H., and Hammer, M.F. (2010). Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian–Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. *Proc Royal Soc B* 277, 1589–1596.
- Cox-Singh, J., Lau, T.Y., Singh, B., and Lee, K.S. (2006). *Plasmodium vivax*: diversity on the mitochondrial genome in Malaysian Borneo.
- Crispo, E., Moore, J.-S., Lee-Yaw, J.A., Gray, S.M., and Haller, B.C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals. *Bioessays* 33, 508–518.
- Crotti, D. (2005). A history of malaria from and for a popular point of view. *Infez Med* 13, 265–270.
- Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., and Scozzari, R. (2011). A revised root for the human Y chromosomal phylogenetic tree: The origin of patrilineal diversity in Africa. *Am J Hum Genet* 88, 814–818.
- Cuevas, J.M., Geller, R., Garijo, R., López-Aldegue, J., and Sanjuán, R. (2015). Extremely high mutation rate of HIV-1 in vivo. *PLoS Bio* 13, e1002251.
- Cui, L., Escalante, A.A., Imwong, M., and Snounou, G. (2003a). The genetic diversity of *Plasmodium vivax* populations. *Trends Parasitol* 19, 220–226.
- Cui, L., Mascorro, C.N., Fan, Q., Rzomp, K.A., Khuntirat, B., Zhou, G., Chen, H., Yan, G., and Sattabongkot, J. (2003b). Genetic diversity and multiple infections of *Plasmodium vivax* malaria in Western Thailand. *Am J Trop Med Hyg* 68, 613–619.
- Culleton, R.L., and Ferreira, P.E. (2012). Duffy Phenotype and *Plasmodium vivax* infections in humans and apes, Africa. *Emerg Infect Dis* 18, 1704–1705.
- Culleton, R., Coban, C., Zeyrek, F.Y., Cravo, P., Kaneko, A., Randrianarivelojosia, M., Andrianaranjaka, V., Kano, S., Farnert, A., Arez, A.P., et al. (2011). The origins of African *Plasmodium vivax*: Insights from mitochondrial genome sequencing. *PLoS ONE* 6, e29137.
- Cunliffe, B.W., Gosden, C., and Joyce, R.A. (2009). *The Oxford Handbook of Archaeology* (Oxford, UK: Oxford University Press).
- Curtin, P.D. (1968). Epidemiology and the slave trade. *Polit Sci Q* 83, 190–216.
- Curtin, P.D. (1990). The end of the “White man’s grave”? Nineteenth-century mortality in West Africa. *J Interdiscip Hist* 21, 63.
- Curtin, P.D. (1994). Malarial immunities in nineteenth-century west Africa and the Caribbean. *Parassitologia* 36, 69–82.
- Czarnecki, J., Friedlaender, J.S., and Stoner, G. (2007). Viral Phylogeny and Human Migration in the Southwest Pacific. In *Population Genetics, Linguistics, and Culture History in the Southwest Pacific*, J.S. Friedlaender, ed. (Oxford University Press), pp. 171–180.
- Dahl, C. (1977). Taxonomy and geographic distribution of Swedish Culicidae (Diptera, Nematocera). *Entomol Scand* 8.
- Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth* 9, 772–772.
- Dash, A.P., Valecha, N., Anvikar, A.R., and Kumar, A. (2008). Malaria in India: Challenges and opportunities. *J Biosc* 33, 583–592.
- Dávalos, L.M., and Perkins, S.L. (2008). Saturation and base composition bias explain phylogenomic conflict in *Plasmodium*. *Genomics* 91, 433–442.
- Davidson, I. (2013). Peopling the last new worlds: The first colonisation of Sahul and the Americas. *Quatern Int* 285, 1–29.
- Davies, D.M., Jones, P., and Mantle, D. (1976). The kinetics of formation of horseradish peroxidase compound I by reaction with peroxobenzoic acids. pH and peroxo acid substituent effects. *Biochem J* 157, 247–253.
- Dawson, G.J., Desai, S.M., Birkenmeyer, L., and Muerhoff, A.S. (2010). Isolation and characterization of the MSP1 genes from *Plasmodium malariae* and *Plasmodium ovale*. *Am J Trop Med Hyg* 82, 996–1003.
- Deeks, W.E. (1925). *Malaria: Its Cause, Prevention and Cure* (Boston, Mass.: United Fruit Company).
- Delatte, H., Bagny, L., Brengue, C., Bouetard, A., Paupy, C., and Fontenille, D. (2011). The invaders: Phylogeography of dengue and chikungunya viruses *Aedes* vectors, on the South West islands of the Indian Ocean. *Infect Genet Evol* 11, 1769–1781.
- Delfin, F., Myles, S., Choi, Y., Hughes, D., Illek, R., van Oven, M., Pakendorf, B., Kayser, M., and Stoneking, M. (2012). Bridging Near and Remote Oceania: mtDNA and NRY variation in the Solomon Islands. *Mol Biol Evol* 29, 545–564.
- Delgado-Ratto, C., Gamboa, D., Soto-Calle, V.E., Van den Eede, P., Torres, E., Sánchez-Martínez, L., Contreras-Mancilla, J., Rosanas-Urgell, A., Rodríguez Ferrucci, H., Llanos-Cuentas, A., et al. (2016). Population genetics of *Plasmodium vivax* in the Peruvian Amazon. *PLoS Negl Trop Dis* 10, e0004376.
- Demeter, F., Bacon, A., Thuy, N.K., Long, V.T., Matsumura, H., Nga, H.H., Schuster, M., Huong, N.M., and Coppens, Y. (2004). An Archaic *Homo Molar* from Northern Vietnam. *Curr Anthropol* 45, 535–541.
- Demeter, F., Bacon, A.-M., Nguyen, K.T., Long, V.T., Düringer, P., Roussé, S., Coppens, Y., Matsumura, H., Dodo, Y., Nguyen, M.H., et al. (2005). Discovery of a second human molar and cranium fragment in the late Middle to Late Pleistocene cave of Ma U’Oi (Northern Vietnam). *J Hum Evol* 48, 393–402.
- Dempwolff Bericht über eine Malaria-Expedition nach Deutsch-Neu-Guinea. *Z Hyg Infektionskr.*
- Denham, T. (2003). Archaeological evidence for mid-Holocene agriculture in the interior of Papua New Guinea: a critical review. *Arch Oceania* 38, 159–176.
- Denham, T. (2005). Agricultural origins and the emergence of rectilinear ditch networks in the Highlands of New Guinea. In *Papuan Past: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, A. Pawley, R. Attenborough, J. Golson, and R. Hide, eds. (Canberra, ACT, Australia: Pacific Linguistics), p.
- Denham, T., Fullagar, R., and Head, L. (2009). Plant exploitation on Sahul: From colonisation to the emergence of regional specialisation during the Holocene. *Quatern Int* 202, 29–40.

- Denham, T.P., Haberle, S.G., Lentfer, C., Fullagar, R., Field, J., Therin, M., Porch, N., and Winsborough, B. (2003). Origins of agriculture at Kuk Swamp in the highlands of New Guinea. *Science* 301, 189–193.
- Desai, M., ter Kuile, F.O., Nosten, F., McGready, R., Asamo, K., Brabin, B., and Newman, R.D. (2007). Epidemiology and burden of malaria in pregnancy. *The Lancet Infectious Diseases* 7, 93–104.
- Dev, V., and Sharma, V.P. (2013). The Dominant Mosquito Vectors of Human Malaria in India. In *Anopheles Mosquitoes - New Insights into Malaria Vectors*, S. Manguin, ed. (InTech), p.
- Diamond, J., and Bellwood, P. (2003). Farmers and their languages: The first expansions. *Science* 300, 597–603.
- Didham, R.K. (2005). Comment on “Avian extinction and mammalian introductions on oceanic islands.” *Science* 307, 1412a–1412a.
- Dimopoulou, D., Piperaki, E.-T., Zerva, L., Tsagris, V., Asprogeraka, S., Kalantzis, D., and Papaevangelou, V. (2017). Concurrent autochthonous malaria caused by *Plasmodium vivax* in father and son in Greece. *J Trop Pediatr* 63, 74–77.
- Dixo, M., Metzger, J.P., Morgante, J.S., and Zamudio, K.R. (2009). Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. *Biol Conserv* 142, 1560–1569.
- Dobney, K., Cucchi, T., and Larson, G. (2008). The pigs of Island Southeast Asia and the Pacific: New evidence for taxonomic status and human-mediated dispersal. *AP* 47, 59–74.
- Donoghue, H.D. (2009). Human tuberculosis – an ancient disease, as elucidated by ancient microbial biomolecules. *Microb Infect* 11, 1156–1162.
- Donohue, M., and Denham, T. (2010). Farming and language in Island Southeast Asia: reframing Austronesian history. *Curr Anthropol* 51, 223–256.
- Doolan, D.L., Saul, A.J., and Good, M.F. (1992). Geographically restricted heterogeneity of the *Plasmodium falciparum* circumsporozoite protein: relevance for vaccine development. *Infect Immun* 60, 675–682.
- Doolan, D.L., Dobaño, C., and Baird, J.K. (2009). Acquired immunity to malaria. *Clin Microbiol Rev* 22, 13–36.
- Drake, J.M., and Lodge, D.M. (2004). Global hot spots of biological invasions: evaluating options for ballast-water management. *Proc Royal Soc B* 271, 575–580.
- Dronamraju, K.R. (2004). *Infectious Disease and Host-Pathogen Evolution* (Cambridge University Press).
- Drummond, A.J., and Bouckaert, R.R. (2015). *Bayesian Evolutionary Analysis with BEAST* (Cambridge University Press).
- Drummond, A.J., Rambaut, A., and Xie, W. (2011). Bayesian Skyline Plot.
- Duggan, A.T., Evans, B., Friedlaender, F.R., Friedlaender, J.S., Koki, G., Merriwether, D.A., Kayser, M., and Stoneking, M. (2014). Maternal history of Oceania from complete mtDNA genomes: Contrasting ancient diversity with recent homogenization due to the Austronesian expansion. *Am J Hum Genet* 94, 721–733.
- Duncan, R.P., and Forsyth, D.M. (2006). Modelling population persistence on islands: mammal introductions in the New Zealand archipelago. *Proc Royal Soc B* 273, 2969–2975.
- Duncan, R.P., Cassey, P., and Blackburn, T.M. (2009). Do climate envelope models transfer? A manipulative test using dung beetle introductions. *Proc Royal Soc B* 276, 1449–1457.
- Dunn, A.M. (2009). Chapter 7 Parasites and Biological Invasions. In *Advances in Parasitology*, (Elsevier), pp. 161–184.
- Dupinay, T., Restorp, K., Leutscher, P., Rousset, D., Chemin, I., Migliani, R., Magnius, L., and Norder, H. (2010). High prevalence of hepatitis B virus genotype E in Northern Madagascar indicates a West-African lineage. *J Med Virol* 82, 1515–1526.
- Duplantier, J.-M., Orth, A., Catalan, J., and Bonhomme, F. (2002). Evidence for a mitochondrial lineage originating from the Arabian peninsula in the Madagascar house mouse (*Mus musculus*). *Heredity* 89, 154–158.
- Durham, W.H. (1991). *Coevolution: Genes, Culture, and Human Diversity* (Stanford University Press).
- Duval, L., and Arie, F. (2012). Ape *Plasmodium* parasites as a source of human outbreaks. *Clin Microbiol Infect* 18, 528–532.
- Eden, M.J. (1993). Swidden cultivation in forest and savanna in lowland Southwest Papua New Guinea. *Hum Ecol* 21, 145–166.
- Edinur, H.A., Dunn, P.P.J., Hammond, L., Selwyn, C., Velickovic, Z.M., Lea, R.A., and Chambers, G.K. (2012). Using HLA loci to inform ancestry and health in Polynesian and Maori populations. *Tissue Antigens* 80, 509–522.
- Eklom, T. (1938). Les races suédoises de l'*Anopheles maculipennis* et leur rôle épidémiologique. *Bull Soc Path Exot* 31.
- Eklom, T., and Ströman, R. (1932). Geographical and biological studies of *Anopheles maculipennis* in Sweden from an epidemiological point of view. *K Sven Vetensk Akad Handl* 11.
- Endicott, P., Gilbert, M.T.P., Stringer, C., Lalueza-Fox, C., Willerslev, E., Hansen, A.J., and Cooper, A. (2003). The genetic origins of the Andaman Islanders. *Am J Hum Genet* 72, 178–184.
- Enserink, M. (2008). A mosquito goes global. *Science* 320, 864–866.
- Escalante, A.A., and Ayala, F.J. (1994). Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *PNAS* 91, 11373–11377.
- Escalante, A.A., Barrio, E., and Ayala, F.J. (1995). Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. *Mol Biol Evol*.
- Escalante, A.A., Freeland, D.E., Collins, W.E., and Lal, A.A. (1998a). The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. *PNAS* 95, 8124–8129.
- Escalante, A.A., Lal, A.A., and Ayala, F.J. (1998b). Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* 149, 189–202.
- Escalante, A.A., Cornejo, O.E., Freeland, D.E., Poe, A.C., Durrego, E., Collins, W.E., and Lal, A.A. (2005). A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *PNAS* 102, 1980–1985.
- Estoup, A., and Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what? *Mol Ecol* 19, 4113–4130.
- Evans, A.G., and Wellems, T.E. (2002). Coevolutionary genetics of *Plasmodium* malaria parasites and their human hosts. *Integr Comp Biol* 42, 401–407.

- Evans, B., and Mountain, M.-J. (2005). Pasin bilong tumbuna: Archaeological evidence for early human activity in the highlands of New Guinea. In *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, A. Pawley, R. Attenborough, J. Golson, and R. Hide, eds. (Canberra, ACT, Australia: Pacific Linguistics), p.
- Excoffier, L., and Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* 7, 745–758.
- Excoffier, L., Smouse, P.E., and Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- Facon, B., Pointier, J.-P., Jarne, P., Sarda, V., and David, P. (2008). High genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Curr Biol* 18, 363–367.
- Failloux, A., Raymond, M., Ung, A., Chevillon, C., and Pasteur, N. (1997). Genetic differentiation associated with commercial traffic in the Polynesian mosquito, *Aedes polynesiensis* Marks 1951. *Biol J Linnean Soc* 60, 107–118.
- Failloux, A.-B., Raymond, M., Ung, A., Glaziou, P., Martin, P.M.V., and Pasteur, N. (2009). Variation in the vector competence of *Aedes polynesiensis* for *Wuchereria bancrofti*. *Parasitology* 111, 19.
- Farrell, A., Coleman, B.I., Benenati, B., Brown, K.M., Blader, I.J., Marth, G.T., and Gubbels, M.-J. (2014). Whole genome profiling of spontaneous and chemically induced mutations in *Toxoplasma gondii*. *BMC Genomics* 15, 354.
- Feng, X., Carlton, J.M., Joy, D.A., Mu, J., Furuya, T., Suh, B.B., Wang, Y., Barnwell, J.W., and Su, X.-Z. (2003). Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *PNAS* 100, 8502–8507.
- Ferreira, M.U., and Hartl, D.L. (2007). *Plasmodium falciparum*: Worldwide sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-2 (MSP-2). *Exp Parasitol* 115, 32–40.
- Ferreira, M.U., and Rodrigues, P.T. (2014). Tracking malaria parasites in the eradication era. *Trends Parasitol* 30, 465–466.
- Ferreira, M., Carlton, J.M., Ribolla, P.E.M., Escalante, A.A., Volkman, S.K., Uplekar, S., Roos, D.S., Hupalo, D.N., Harb, O.S., Pereira Alves, J.M., et al. (2015). Population genetics, evolutionary genomics, and genome-wide studies of malaria: A view across the international centers of excellence for malaria research. *Am J Trop Med Hyg* 93, 87–98.
- Ferreira, M.U., Karunaweera, N.D., da Silva-Nunes, M., da Silva, N.S., Wirth, D.F., and Hartl, D.L. (2007a). Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. *J Infect Dis* 195, 1218–1226.
- Ferreira, M.U., Karunaweera, N.D., Silva-Nunes, M. da, Silva, N.S. da, Wirth, D.F., and Hartl, D.L. (2007b). Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. *J Infect Dis* 195, 1218–1226.
- Ficetola, G.F., Mazel, F., and Thuiller, W. (2017). Global determinants of zoogeographical boundaries. *Nat Ecol Evol* 1, 0089.
- Fincher, C.L., and Thornhill, R. (2008). A parasite-driven wedge: Infectious diseases may explain language and other biodiversity. *Oikos* 117, 1289–1297.
- Fisher, N., Abd Majid, R., Antoine, T., Al-Helal, M., Warman, A.J., Johnson, D.J., Lawrenson, A.S., Ranson, H., O'Neill, P.M., Ward, S.A., et al. (2012). Cytochrome b mutation Y268S conferring atovaquone resistance phenotype in malaria parasite results in reduced parasite bc1 catalytic turnover and protein expression. *J Biol Chem* 287, 9731–9741.
- Fitzpatrick, S.M., and Callaghan, R. (2008). Seafaring simulations and the origin of prehistoric settlers in Madagascar. In *Islands of Inquiry Colonisation, Seafaring and the Archaeological of Maritime Landscapes*, (Canberra, ACT, Australia: ANU E Press), p.
- Fletcher, H.A., Donoghue, H.D., Holton, J., Pap, I., and Spigelman, M. (2003). Widespread occurrence of *Mycobacterium tuberculosis* DNA from 18th–19th century Hungarians. *Am J Phys Anthropol* 120, 144–152.
- Flint, J., Hill, A.V.S., Bowden, D.K., Oppenheimer, S.J., Sill, P.R., Serjeantson, S.W., Bana-Koiri, J., Bhatia, K., Alpers, M.P., Boyce, A.J., et al. (1986). High frequencies of α -thalassaemia are the result of natural selection by malaria. *Nature* 321, 744–750.
- Flint, J., Harding, R.M., Clegg, J.B., and Boyce, A.J. (1993). Why are some genetic diseases common? *Hum Genet* 91, 91–117.
- Flint, J., Harding, R.M., Boyce, A.J., and Clegg, J.B. (1998). The population genetics of the haemoglobinopathies. *Bail Cl Haem* 11, 1–51.
- Fluxus-Engineering Network 4.6.
- Fola, A.A., Abby Harrison, G.L., Hazairin, M.H., Barnadas, C., Hetzel, M.W., Iga, J., Siba, P.M., Mueller, I., and Barry, A.E. (2017). Higher complexity of infection and genetic diversity of *Plasmodium vivax* than *Plasmodium falciparum* across all malaria transmission zones of Papua New Guinea. *Am J Trop Med Hyg* 16–0716.
- Fola, A.A., Nate, E., Abby Harrison, G.L., Barnadas, C., Hetzel, M.W., Iga, J., Siba, P., Mueller, I., and Barry, A.E. (2018). Nationwide genetic surveillance of *Plasmodium vivax* in Papua New Guinea reveals heterogeneous transmission dynamics and routes of migration amongst subdivided populations. *Infect Genet Evol* 58, 83–95.
- Foley, D.H., Meek, S.R., and Bryan, J.H. (1994). The *Anopheles punctulatus* group of mosquitoes in the Solomon Islands and Vanuatu surveyed by allozyme electrophoresis. *Med Vet Entomol* 8, 340–350.
- Foley, D.H., Rueda, L.M., and Wilkerson, R.C. (2007). Insight into global mosquito biogeography from country species records. *J Med Entomol* 44, 554–567.
- Fontanillas, P., Landry, C.R., Wittkopp, P.J., Russ, C., Gruber, J.D., Nusbaum, C., and Hartl, D.L. (2010). Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. *Mol Ecol* 19, 212–227.
- FORD, D.H., KANTOUNIS, S., and LAWRENCE, R. (1959). The localization of I 131-labeled triiodothyronine in the pituitary and brain of normal and thyroidectomized male rats. *Endocrinology* 64, 977–991.
- Forster, P. (2004). Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Phil Trans R Soc B* 359, 255–264.
- Förster, D.W., Gündüz, I., Nunes, A.C., Gabriel, S., Ramalhinho, M.G., Mathias, M.L., Britton-Davidian, J., and Searle, J.B. (2009). Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Mol Ecol* 18, 4477–4494.
- Fox, J.J. (2006). *Origins, Ancestry and Alliance: Explorations in Austronesian Ethnography* (ANU E Press).

- Fried, M., Muga, R.O., Misore, A.O., and Duffy, P.E. (1998). Malaria elicits type 1 cytokines in the human placenta: IFN- γ and TNF- α associated with pregnancy outcomes. *J Immunol* 160, 2523–2530.
- Friedlaender, J.S. (2007a). *Genes, Language, & Culture History in the Southwest Pacific* (Oxford, UK: Oxford University Press).
- Friedlaender, J.S. (2007b). *Population Genetics, Linguistics, and Culture History in the Southwest Pacific* (Oxford University Press).
- Friedlaender, J.S., Friedlaender, F.R., Hodgson, J.A., Stoltz, M., Koki, G., Horvat, G., Zhadanov, S., Schurr, T.G., and Merriwether, D.A. (2007). Melanesian mtDNA complexity. *PLoS ONE* 2, e248.
- Friedlaender, J.S., Friedlaender, F.R., Reed, F.A., Kidd, K.K., Kidd, J.R., Chambers, G.K., Lea, R.A., Loo, J.-H., Koki, G., Hodgson, J.A., et al. (2008). The genetic structure of Pacific Islanders. *PLoS Genet* 4, 0173–0190.
- Fu, Q., Rudan, P., Pääbo, S., and Krause, J. (2012). Complete mitochondrial genomes reveal Neolithic expansion into Europe. *PLoS ONE* 7, e32473.
- Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. *Nature* 534, 200–205.
- Fullagar, R., Field, J., Denham, T., and Lentfer, C. (2006). Early and mid Holocene tool-use and processing of taro (*Colocasia esculenta*), yam (*Dioscorea* sp.) and other plants at Kuk Swamp in the highlands of Papua New Guinea. *J Archaeol Sci* 33, 595–614.
- Gabriel, S.I., Stevens, M.I., Mathias, M. da L., and Searle, J.B. (2011). Of Mice and ‘Convicts’: Origin of the Australian House Mouse, *Mus musculus*. *PLoS ONE* 6, e28622.
- Gagneux, S. (2012). Host–pathogen coevolution in human tuberculosis. *Phil Trans R Soc B* 367, 850–859.
- Gajadhar, A.A., Marquardt, W.C., Hall, R., Gunderson, J., Ariztia-Carmona, E.V., and Sogin, M.L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Cryptosporidium parvum* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol Biochem Parasitol* 45, 147–154.
- Gallup, J.I., and Sachs, J.D. (2001). The economic burden of malaria. *The American Journal of Tropical Medicine and Hygiene* 64, 85–96.
- Gamain, B., Smith, J.D., Viebig, N.K., Gysin, J., and Scherf, A. (2007). Pregnancy-associated malaria: Parasite binding, natural immunity and vaccine development. *International Journal for Parasitology* 37, 273–283.
- García-Ramos, G., and Rodríguez, D. (2002). Evolutionary speed of species invasions. *Evolution* 56, 661.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511.
- Garnham, P.C.C. (1966). *Malaria Parasites and Other Haemosporidia* (Oxford, UK: Blackwell Scientific Publications).
- Garnham, P.C.C. (1980). Malaria in its various vertebrate hosts. In *Malaria*, J.P. Kreier, ed. (New York: Academic Press), p.
- Garnham, P.C.C. (1988). Malaria parasites of man: life-cycles and morphology (excluding ultrastructure). In *Malaria Principles and Practice of Malariology*, W.H. Wernsdorfer, and I. McGregor, eds. (Edinburgh: Churchill Livingstone), p.
- Gelabert, P., Sandoval-Velasco, M., Olalde, I., Fregel, R., Rieux, A., Escosa, R., Aranda, C., Paaijmans, K., Mueller, I., Gilbert, M.T.P., et al. (2016). Mitochondrial DNA from the eradicated European *Plasmodium vivax* and *P. falciparum* from 70-year-old slides from the Ebro Delta in Spain. *PNAS* 113, 11495–11500.
- Genton, B., Al-Yaman, F., Mgone, C.S., Alexander, N., Panu, M.M., Alpers, M.P., and Mokela, D. (1995). Ovalocytosis and cerebral malaria. *Nature* 378, 564–565.
- Gething, P.W., Van Boeckel, T.P., Smith, D.L., Guerra, C.A., Patil, A.P., Snow, R.W., and Hay, S.I. (2011a). Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasit Vectors* 4, 92.
- Gething, P.W., Patil, A.P., Smith, D.L., Guerra, C.A., Elyazar, I.R., Johnston, G.L., Tatem, A.J., and Hay, S.I. (2011b). A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar J* 10, 378.
- 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., et al. (2012). A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl Trop Dis* 6, e1814.
- Gibbons, A. (2001). The Peopling of the Pacific. *Science* 291, 1735–1737.
- Gignoux, C.R., Henn, B.M., and Mountain, J.L. (2011). Rapid, global demographic expansions after the origins of agriculture. *PNAS* 108, 6044–6049.
- Gjesing, B. (1975). The electrolyte pattern of the gastric secretion from the vervet monkey. *West Afr J Pharmacol Drug Res* 2, 10–16.
- Global Partnership to Roll Back Malaria, World Health Organization, and UNICEF (2009). *World Malaria Report 2009* (Geneva: World Health Organization).
- Gomes, S.M., Bodner, M., Souto, L., Zimmermann, B., Huber, G., Strobl, C., Röck, A.W., Achilli, A., Olivieri, A., Torroni, A., et al. (2015). Human settlement history between Sunda and Sahul: a focus on East Timor (Timor-Leste) and the Pleistocene mtDNA diversity. *BMC Genomics* 16, 70.
- Gomes, S.M., van Oven, M., Souto, L., Morreira, H., Brauer, S., Bodner, M., Zimmermann, B., Huber, G., Strobl, C., Röck, A.W., et al. (2017). Lack of gene–language correlation due to reciprocal female but directional male admixture in Austronesians and non-Austronesians of East Timor. *Eur J Hum Genet* 25, 246–252.
- Gonçalves, L.A., Cravo, P., and Ferreira, M.U. (2014). Emerging *Plasmodium vivax* resistance to chloroquine in South America: an overview. *Memórias Do Instituto Oswaldo Cruz* 109, 534–539.
- Gonder, M.K., Mortensen, H.M., Reed, F.A., Sousa, A. de, and Tishkoff, S.A. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 24, 757–768.
- Gonzalez, M.C., Hidalgo, C.A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature* 453, 779–782.
- Gosden, C. (2009). The Pacific Islands. In *The Oxford Handbook of Archaeology*, B.W. Cunliffe, C. Gosden, and R.A. Joyce, eds. (Oxford, UK: Oxford University Press), pp. 898–925.
- Gosden, C. (2010). When Humans Arrived in the New Guinea Highlands. *Science* 330, 41–42.

- Gourjon, G., Boëtsch, G., and Degioanni, A. (2011). Gender and population history: sex bias revealed by studying genetic admixture of Ngazidja population (Comoro Archipelago). *Am J Phys Anthropol* 144, 653–660.
- Govella, N.J., and Ferguson, H. (2012). Why use of interventions targeting outdoor biting mosquitoes will be necessary to achieve malaria elimination. *Front Physiol* 3.
- Gray, R.D., Drummond, A.J., and Greenhill, S.J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323, 479–483.
- Green, R. (2000). A range of disciplines support a dual origin for the bottle gourd in the Pacific. *J Polynesian Soc* 109, 191–197.
- Green, R.C. (1991). The Lapita Cultural Complex: Current evidence and proposed models. *BIPPA* 11, 295–305.
- Groube, L.M. (1993). Contradictions and malaria in Melanesian and Australian prehistory. In *A Community of Culture: The People and Prehistory of the Pacific*, (Canberra, Australia: Dept. of Prehistory, Research School of Pacific Studies, ANU), pp. 164–186.
- Groube, L., Chappell, J., Muke, J., and Price, D. (1986). A 40,000 year-old human occupation site at Huon Peninsula, Papua New Guinea. *Nature* 324, 453–455.
- Groucutt, H.S., and Petraglia, M.D. (2012). The prehistory of the Arabian peninsula: Deserts, dispersals, and demography. *Evol Anthr* 21, 113–125.
- Groucutt, H.S., Petraglia, M.D., Bailey, G., Scerri, E.M.L., Parton, A., Clark-Balzan, L., Jennings, R.P., Lewis, L., Blinkhorn, J., Drake, N.A., et al. (2015). Rethinking the dispersal of *Homo sapiens* out of Africa. *Evol Anthr* 24, 149–164.
- Gruenberg, M., Moniz, C.A., Hofmann, N.E., Wampfler, R., Koepfli, C., Mueller, I., Monteiro, W.M., Lacerda, M., de Melo, G.C., Kuehn, A., et al. (2018). *Plasmodium vivax* molecular diagnostics in community surveys: pitfalls and solutions. *Malar J* 17.
- Grün, R., and Stringer, C.B. (1991). Electron spin resonance dating and the evolution of modern humans. *Archaeometry* 33, 153–199.
- Grün, R., Stringer, C., McDermott, F., Nathan, R., Porat, N., Robertson, S., Taylor, L., Mortimer, G., Eggins, S., and McCulloch, M. (2005). U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *J Hum Evol* 49, 316–334.
- Guerra, C.A., Snow, R.W., and Hay, S.I. (2006a). A global assessment of closed forests, deforestation and malaria risk. *Ann Trop Med Parasitol* 100, 189–204.
- Guerra, C.A., Snow, R.W., and Hay, S.I. (2006b). Mapping the global extent of malaria in 2005. *Trends in Parasitology* 22, 353–358.
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52, 696–704.
- Gunasekera, A.M., Wickramarachchi, T., Neafsey, D.E., Ganguli, I., Perera, L., Premaratne, P.H., Hartl, D., Handunnetti, S.M., Udagama-Randeniya, P.V., and Wirth, D.F. (2007). Genetic diversity and selection at the *Plasmodium vivax* apical membrane antigen-1 (*PvAMA-1*) locus in a Sri Lankan population. *Mol Biol Evol* 24, 939–947.
- Gunawardena, S., Karunaweera, N.D., Ferreira, M.U., Phone-Kyaw, M., Pollack, R.J., Alifrangis, M., Rajakaruna, R.S., Konradsen, F., Amerasinghe, P.H., Schousboe, M.L., et al. (2010). Geographic structure of *Plasmodium vivax*: Microsatellite analysis of parasite populations from Sri Lanka, Myanmar, and Ethiopia. *Am J Trop Med Hyg* 82, 235–242.
- Gunawardena, S., Ferreira, M.U., Kapilanda, G.M.G., Wirth, D.F., and Karunaweera, N.D. (2014). The Sri Lankan paradox: High genetic diversity in *Plasmodium vivax* populations despite decreasing levels of malaria transmission. *Parasitology* 141, 880–890.
- Gunnarsdóttir, E.D., Li, M., Bauchet, M., Finstermeier, K., and Stoneking, M. (2011). High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res* 21, 1–11.
- Gunz, P., Bookstein, F.L., Mitteroecker, P., Stadlmayr, A., Seidler, H., and Weber, G.W. (2009). Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *PNAS* 106, 6094–6098.
- Gupta, S., Snow, R.W., Donnelly, C.A., Marsh, K., and Newbold, C. (1999). Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nat Med* 5, 340–343.
- Gutsevich, A.V., Monchadskii, A.S., and Shtakelberg, A.A. (1974). Mosquitoes. Family Culicidae. Fauna of the USSR. *Diptera* 3.
- Guyatt, H.L., and Snow, R.W. (2001). Malaria in pregnancy as an indirect cause of infant mortality in sub-Saharan Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 95, 569–576.
- Hagelberg, E., Goldman, N., Lió, P., Whelan, S., Schiefenhövel, W., Clegg, J.B., and Bowden, D.K. (1999). Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc Royal Soc B* 266, 485–492.
- Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T., and Fox, J.A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecol Lett* 8, 1114–1127.
- Halligan, J.J., Waters, M.R., Perrotti, A., Owens, I.J., Feinberg, J.M., Bourne, M.D., Fenerty, B., Winsborough, B., Carlson, D., Fisher, D.C., et al. (2016). Pre-Clovis occupation 14,550 years ago at the Page-Ladson site, Florida, and the peopling of the Americas. *Sci Adv* 2, e1600375–e1600375.
- Hamad, A.A., El Hassan, I.M., El Khalifa, A.A., Ahmed, G.I., Abdelrahim, S.A., Theander, T.G., and Arnot, D.E. (2000). Chronic *Plasmodium falciparum* infections in an area of low intensity malaria transmission in the Sudan. *Parasitology* 120, 447–456.
- Hamilton, W.D. (1982). Pathogens as Causes of Genetic Diversity in their Host Populations. In *Population Biology of Infectious Diseases*, R.M. Anderson, and R.M. May, eds. (Berlin Heidelberg: Springer), pp. 269–296.
- Hamilton, A.M., Zug, G.R., and Austin, C.C. (2010). Biogeographic anomaly or human introduction: a cryptogenic population of tree skink (Reptilia: Squamata) from the Cook Islands, Oceania. *Biol J Linnean Soc* 100, 318–328.
- Hänfling, B., and Kollmann, J. (2002). An evolutionary perspective of biological invasions. *Trends Ecol Evol* 17, 545–546.

- Hardouin, E.A., Chapuis, J.-L., Stevens, M.I., van Vuuren, J., Quillfeldt, P., Scavetta, R.J., Teschke, M., and Tautz, D. (2010). House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evol Biol* 10, 325.
- Harlan, J.R. (1976). *Origins of African Plant Domestication* (Walter de Gruyter).
- Harrell, F.E. (2018). Hmisc: Harrell Miscellaneous.
- Harrus, S., and Baneth, G. (2005). Drivers for the emergence and re-emergence of vector-borne protozoal and bacterial diseases. *Int J Parasitol* 35, 1309–1318.
- Hartl, D.L. (2004). The origin of malaria: Mixed messages from genetic diversity. *Nat Rev Micro* 2, 15–22.
- Hartl, D.L., Volkman, S.K., Nielsen, K.M., Barry, A.E., Day, K.P., Wirth, D.F., and Winzeler, E.A. (2002). The paradoxical population genetics of *Plasmodium falciparum*. *Trends Parasitol* 18, 266–272.
- Hasan, A.U., Suguri, S., Fujimoto, C., Itaki, R.L., Harada, M., Kawabata, M., Bugoro, H., and Albino, B. (2008). Genetic diversity in two sibling species of the *Anopheles punctulatus* group of mosquitoes on Guadalcanal in the Solomon Islands. *BMC Evol Biol* 8, 318.
- Hay, S.I., Sinka, M.E., Okara, R.M., Kabaria, C.W., Mbithi, P.M., Tago, C.C., Benz, D., Gething, P.W., Howes, R.E., Patil, A.P., et al. (2010). Developing global maps of the dominant *Anopheles* vectors of human malaria. *PLoS Med* 7, e1000209.
- Haynes, J.D., Diggs, C.L., Hines, F.A., and Desjardins, R.E. (1976). Culture of human malaria parasites *Plasmodium falciparum*. *Nature* 263, 767–769.
- Hedrick, P.W. (2011). Population genetics of malaria resistance in humans. *Heredity* 107, 283–304.
- Heller, M., Edelstein, P., and Mayer, M. (1975). Membrane-bound enzymes. III. Protease activity in leucocytes in relation to erythrocyte membranes. *Biochim. Biophys. Acta* 413, 472–482.
- Henn, B.M., Cavalli-Sforza, L.L., and Feldman, M.W. (2012). The great human expansion. *PNAS* 109, 17758–17764.
- Henry-Hallidin, C.N., Nadesakumaran, K., Keven, J.B., Zimmerman, A.M., Siba, P., Mueller, I., Hetzel, M.W., Kazura, J.W., Thomsen, E., Reimer, L.J., et al. (2012). Multiplex assay for species identification and monitoring of insecticide resistance in *Anopheles punctulatus* group populations of Papua New Guinea. *Am J Trop Med Hyg* 86, 140–151.
- Hernberg, C.A. (1947a). Clinical observations on malaria tertiana in Finland, and on difference between autumn and spring malaria. *Ada Med Scand* 78.
- Hernberg, C.A. (1947b). The epidemiology in malaria tertiana in Finland during the years 1941–1945. *Acta Med Scand* 77.
- Herrera, S., Vinetz, J.M., Moss, W.J., Mueller, I., Nankabirwa, J., Cui, L., Mohapatra, P.K., Arevalo-Herrera, M., Guzman, M., Kazura, J.W., et al. (2015). Malaria epidemiology and control within the International Centers of Excellence for Malaria Research. *Am J Trop Med Hyg* 93, 5–15.
- Hershkovitz, I., Weber, G.W., Quam, R., Duval, M., Grün, R., Kinsley, L., Ayalon, A., Bar-Matthews, M., Valladas, H., Mercier, N., et al. (2018). The earliest modern humans outside Africa. *Science* 359, 456–459.
- Hertzberg, M., Mickleson, K.N., Serjeantson, S.W., Prior, J.F., and Trent, R.J. (1989). An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *Am J Hum Genet* 44, 504–510.
- Hii, J.L.K., Smith, T., Mai, A., Ibam, E., and Alpers, M.P. (2000). Comparison between anopheline mosquitoes (Diptera: Culicidae) caught using different methods in a malaria endemic area of Papua New Guinea. *Bull Entomol Res* 90.
- Hijikata, M., Keicho, N., Duc, L.V., Maeda, S., Hang, N.T.L., Matsushita, I., and Kato, S. (2017). Spoligotyping and whole-genome sequencing analysis of lineage 1 strains of *Mycobacterium tuberculosis* in Da Nang, Vietnam. *PLoS ONE* 12, e0186800.
- Hikosaka, K., Watanabe, Y., Kobayashi, F., Waki, S., Kita, K., and Tanabe, K. (2011). Highly conserved gene arrangement of the mitochondrial genomes of 23 Plasmodium species. *Parasitol Int* 60, 175–180.
- Hill, A.V.S. (2012). Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Phil Trans R Soc B* 367, 840–849.
- Hill, E. (2010). Sunda to Sahul: Trans-Wallacean distribution of recent salticid genera (Araneae: Salticidae). *Peckhamia* 80, 1–60.
- Ho, S.Y.W., and Shapiro, B. (2011). Skyline-plot methods for estimating demographic history from nucleotide sequences: INVITED TECHNICAL REVIEW. *Mol Ecol Resour* 11, 423–434.
- Hofreiter, M., Paijmans, J.L.A., Goodchild, H., Speller, C.F., Barlow, A., Fortes, G.G., Thomas, J.A., Ludwig, A., and Collins, M.J. (2015). The future of ancient DNA: Technical advances and conceptual shifts: Prospects & Overviews. *BioEssays* 37, 284–293.
- Holderegger, R., and Di Giulio, M. (2010). The genetic effects of roads: A review of empirical evidence. *Basic and Applied Ecology* 11, 522–531.
- Holmes, E.C. (2009). *The evolution and emergence of RNA viruses* (Oxford: Oxford University Press).
- Holmes, E.C., Nee, S., Rambaut, A., Garnett, G.P., and Harvey, P.H. (1995). Revealing the history of infectious disease epidemics through phylogenetic trees. *Phil Trans R Soc B* 349, 33–40.
- Holsinger, K.E., and Mason-Gamer, R.J. (1996). Hierarchical analysis of nucleotide diversity in geographically structured populations. *Genetics* 142, 629–639.
- van Holst Pellekaan, S. (2013). Genetic evidence for the colonization of Australia. *Quatern Int* 285, 44–56.
- Hope, G. (1983). The vegetational changes of the last 20,000 years at Telefomin, Papua New Guinea. *Singapore Journal of Tropical Geography* 4, 25–33.
- Hope, G.S. (1996). Quaternary change and the historical biogeography of Pacific Islands. In *The Origin and Evolution of Pacific Island Biotas, New Guinea to Eastern Polynesia: Patterns and Processes*, A. Keast, and S.E. Miller, eds. (Amsterdam: SPB Publishing), pp. 165–190.
- Hope, G., and Golson, J. (1995). Late Quaternary change in the mountains of New Guinea. *Antiquity* 69, 818–830.
- Hostetler, J.B., Lo, E., Kanjee, U., Amaratunga, C., Suon, S., Sreng, S., Mao, S., Yewhalaw, D., Mascarenhas, A., Kwiatkowski, D.P., et al. (2016). Independent origin and global distribution of distinct *Plasmodium vivax* Duffy Binding Protein gene duplications. *PLoS Negl Trop Dis* 10, e0005091.

- Howes, R.E., Battle, K.E., Satyagraha, A.W., Baird, J.K., and Hay, S.I. (2013). G6PD deficiency: Global distribution, genetic variants and primaquine therapy. *Adv Parasitol* 81, 133–201.
- Howes, R.E., Reiner Jr., R.C., Battle, K.E., Longbottom, J., Mappin, B., Ordanovich, D., Tatem, A.J., Drakeley, C., Gething, P.W., Zimmerman, P.A., et al. (2015). *Plasmodium vivax* transmission in Africa. *PLoS Negl Trop Dis* 9, e0004222.
- Howes, R.E., Battle, K.E., Mendis, K.N., Smith, D.L., Cibulskis, R.E., Baird, J.K., and Hay, S.I. (2016). Global epidemiology of *Plasmodium vivax*. *Am J Trop Med Hyg* 95, 15–34.
- Hsieh, J., Hsing, Y.C., Hsu, T., Li, P.J., Li, K., and Tsang, C. (2011). Studies on ancient rice—where botanists, agronomists, archeologists, linguists, and ethnologists meet. *Rice* 4, 178–183.
- Hubbe, M., Neves, W.A., and Harvati, K. (2010). Testing Evolutionary and Dispersion Scenarios for the Settlement of the New World. *PLoS ONE* 5, e11105.
- Hudjashov, G., Kivisild, T., Underhill, P.A., Endicott, P., Sanchez, J.J., Lin, A.A., Shen, P., Oefner, P., Renfrew, C., Villems, R., et al. (2007). Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *PNAS* 104, 8726–8730.
- Hudson, R.R. (2000). A new statistic for detecting genetic differentiation. *Genetics* 155, 2011–2014.
- Hudson, R.R., Boos, D.D., and Kaplan, N.L. (1992a). A statistical test for detecting geographic subdivision. *Mol Biol Evol* 9, 138–151.
- Hudson, R.R., Slatkin, M., and Maddison, W.P. (1992b). Estimation of levels of gene flow from DNA sequence data. *Genetics* 132, 583–589.
- Hughes, A.L., and Verra, F. (2001). Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc Royal Soc B* 268, 1855–1860.
- Huldén, L., Huldén, L., and Heliövaara, K. (2005). Endemic malaria: an “indoor” disease in northern Europe. Historical data analysed. *Malar J* 4, 19.
- Hupalo, D.N., Luo, Z., Melnikov, A., Sutton, P.L., Rogov, P., Escalante, A., Vallejo, A.F., Herrera, S., Arévalo-Herrera, M., Fan, Q., et al. (2016). Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat Genet* 48, 953–958.
- Imwong, M., Sudimack, D., Pukrittayakamee, S., Osorio, L., Carlton, J.M., Day, N.P.J., White, N.J., and Anderson, T.J.C. (2006). Microsatellite variation, repeat array length, and population history of *Plasmodium vivax*. *Mol Biol Evol* 23, 1016–1018.
- Imwong, M., Nair, S., Pukrittayakamee, S., Sudimack, D., Williams, J.T., Mayxay, M., Newton, P.N., Kim, J.R., Nandy, A., Osorio, L., et al. (2007). Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America. *Int J Parasitol* 37, 1013–1022.
- Imwong, M., Hien, T.T., Thuy-Nhien, N.T., Dondorp, A.M., and White, N.J. (2017a). Spread of a single multidrug resistant malaria parasite lineage (PfPailin) to Vietnam. *Lancet Infect Dis* 17, 1022–1023.
- Imwong, M., Suwannasin, K., Kunasol, C., Sutawong, K., Mayxay, M., Rekol, H., Smithuis, F.M., Hlaing, T.M., Tun, K.M., van der Pluijm, R.W., et al. (2017b). The spread of artemisinin-resistant *Plasmodium falciparum* in the Greater Mekong subregion: a molecular epidemiology observational study. *Lancet Infect Dis* 17, 491–497.
- Ingman, M., and Gyllenstein, U. (2003). Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* 13, 1600–1606.
- Ingman, M., Kaessmann, H., Pääbo, S., and Gyllenstein, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708–713.
- Irwin, G. (1994). *The Prehistoric Exploration and Colonisation of the Pacific* (Cambridge University Press).
- Iwagami, M., Hwang, S.-Y., Fukumoto, M., Hayakawa, T., Tanabe, K., Kim, S.-H., Kho, W.-G., and Kano, S. (2010). Geographical origin of *Plasmodium vivax* in the Republic of Korea: haplotype network analysis based on the parasite’s mitochondrial genome. *Malar J* 9, 184.
- Jablonski, N.G., and Chaplin, G. (2012). Human skin pigmentation, migration and disease susceptibility. *Phil Trans R Soc B* 367, 785–792.
- Jaenson, T.G., and Ameneshewa, B. (1991). Prehibernation diet and reproductive condition of female *Anopheles messeae* in Sweden. *Med Vet Entomol* 5.
- Jakobsson, M., Edge, M.D., and Rosenberg, N.A. (2013). The relationship between FST and the frequency of the most frequent allele. *Genetics* 193, 515–528.
- Jansen, C.C., and Beebe, N.W. (2010). The dengue vector *Aedes aegypti*: what comes next. *Microb Infect* 12, 272–279.
- Jennison, C., Arnott, A., Tessier, N., Tavul, L., Koepfli, C., Felger, I., Siba, P.M., Reeder, J.C., Bahlo, M., Mueller, I., et al. (2015). *Plasmodium vivax* populations are more genetically diverse and less structured than sympatric *Plasmodium falciparum* populations. *PLoS Negl Trop Dis* 9, e0003634.
- Jin, L., and Su, B. (2000). Natives or immigrants: modern human origin in east Asia. *Nat Rev Genet* 1, 126–133.
- Jinam, T.A., Hong, L.-C., Phipps, M.E., Stoneking, M., Ameen, M., Edo, J., and Saitou, N. (2012). Evolutionary history of continental Southeast Asians: “Early Train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA Data. *Mol Biol Evol* 29, 3513–3527.
- Jobling, M.A. (2012). The impact of recent events on human genetic diversity. *Phil Trans R Soc B* 367, 793–799.
- Johnson, A.M., Murray, P.J., Illana, S., and Baverstock, P.J. (1987). Rapid nucleotide sequence analysis of the small subunit ribosomal RNA of *Toxoplasma gondii*: evolutionary implications for the Apicomplexa. *Mol Biochem Parasitol* 25, 239–246.
- Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11, 94.
- Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Ferreira, M.U., Kanbara, H., and Hughes, A.L. (2005). Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. *Mol Biol Evol* 22, 1733–1739.
- Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Kretzli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., et al. (2003). Early origin and recent expansion of *Plasmodium falciparum*. *Science* 300, 318–321.

- Joy, D.A., Mu, J., Jiang, H., and Su, X. (2006). Genetic diversity and population history of *Plasmodium falciparum* and *Plasmodium vivax*. *Parassitologia* 48, 561–566.
- Joy, D.A., Gonzalez-Ceron, L., Carlton, J.M., Gueye, A., Fay, M., McCutchan, T.F., and Su, X. -z. (2008). Local adaptation and vector-mediated population structure in *Plasmodium vivax* malaria. *Mol Biol Evol* 25, 1245–1252.
- Jung, S., and Marron, J.S. (2009). PCA consistency in high dimension, low sample size context. *The Annals of Statistics* 37, 4104–4130.
- Kaessmann, H., and Paabo, S. (2002). The genetical history of humans and the great apes. *Journal of Internal Medicine* 251, 1–18.
- Kamgang, B., Brengues, C., Fontenille, D., Njiokou, F., Simard, F., and Paupy, C. (2011). Genetic structure of the tiger mosquito, *Aedes albopictus*, in Cameroon (Central Africa). *PLoS ONE* 6, e20257.
- Kaneko, A. (2004). Hypothesis: Malaria biodiversity and control on Island Melanesia. *Int Congress Series* 1267, 88–97.
- Kaneko, A., Taleo, G., Kalkoa, M., Yaviong, J., Reeve, P.A., Ganczakowski, M., Shirakawa, C., Palmer, K., Kobayakawa, T., and Björkman, A. (1998). Malaria epidemiology, glucose 6-phosphate dehydrogenase deficiency and human settlement in the Vanuatu Archipelago. *Acta Trop* 70, 285–302.
- Kano, F.S., de Souza, A.M., de Menezes Torres, L., Costa, M.A., Souza-Silva, F.A., Sanchez, B.A.M., Fontes, C.J.F., Soares, I.S., de Brito, C.F.A., Carvalho, L.H., et al. (2018). Susceptibility to *Plasmodium vivax* malaria associated with DARC (Duffy antigen) polymorphisms is influenced by the time of exposure to malaria. *Sci Rep* 8.
- Karafet, T.M., Hallmark, B., Cox, M.P., Sudoyo, H., Downey, S., Lansing, J.S., and Hammer, M.F. (2010). Major East-West division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol* 27, 1833–1844.
- Karafet, T.M., Mendez, F.L., Sudoyo, H., Lansing, J.S., and Hammer, M.F. (2015). Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. *Eur J Hum Genet* 23, 369–373.
- Karunaweera, N.D., Ferreira, M.U., Munasinghe, A., Barnwell, J.W., Collins, W.E., King, C.L., Kawamoto, F., Hartl, D.L., and Wirth, D.F. (2008). Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. *Gene* 410, 105–112.
- Katzourakis, A., and Gifford, R.J. (2010). Endogenous Viral Elements in Animal Genomes. *PLoS Genet* 6, e1001191.
- Kayser, M. (2010). The human genetic history of Oceania: Near and Remote views of dispersal. *Curr Biol* 20, R194–R201.
- Kayser, M., Brauer, S., Weiss, G., Underhill, Peter A., Roewer, L., Schiefenhövel, W., and Stoneking, M. (2000). Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 10, 1237–1246.
- Kayser, M., Brauer, S., Weiss, G., Underhill, P.A., Roewer, L., Schiefenhövel, W., and Stoneking, M. (2001a). Correction: Melanesian origin of Polynesian Y-chromosomes. *Curr Biol* 11, I–II.
- Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P.A., and Stoneking, M. (2001b). Independent histories of human Y chromosomes from Melanesia and Australia. *Am J Hum Genet* 68, 173–190.
- Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M., and Stoneking, M. (2003). Reduced Y-Chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72, 281–302.
- Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L.A., Moyse-Faurie, C., Rutledge, R.B., Schiefenhövel, W., Gil, D., et al. (2006). Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol* 23, 2234–2244.
- Kayser, M., Choi, Y., Oven, M. van, Mona, S., Brauer, S., Trent, R.J., Suarkia, D., Schiefenhövel, W., and Stoneking, M. (2008a). The impact of the Austronesian expansion: Evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol Biol Evol* 25, 1362–1374.
- Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X., Nürnberg, P., Trent, R.J., and Stoneking, M. (2008b). Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am J Hum Genet* 82, 194–198.
- Kedzierski, L., Escalante, A.A., Isea, R., Black, C.G., Barnwell, J.W., and Coppel, R.L. (2002). Phylogenetic analysis of the genus *Plasmodium* based on the gene encoding adenylosuccinate lyase. *Infect Genet Evol* 1, 297–301.
- Keenihan, S.H., Gramzinski, R., Ratiwayanto, S., Hadiputranto, H., Riberu, W., Soebianto, S., Rusjdy, F., Syafruddin, D., Kartikasari, A., Djojsubroto, M., et al. (2003). *Plasmodium falciparum*. Mechanisms of innate and acquired protection against *Plasmodium falciparum* in Javanese transmigrant adults and children newly resident in malaria-endemic Northwest Papua. *Adv Exp Med Biol* 531, 83–102.
- Keller, I., Bensasson, D., and Nichols, R.A. (2007). Transition-Transversion Bias Is Not Universal: A Counter Example from Grasshopper Pseudogenes. *PLoS Genet* 3, e22.
- Keller, R.P., Drake, J.M., Drew, M.B., and Lodge, D.M. (2011). Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network. *Diversity and Distributions* 17, 93–102.
- Kellermann, V., van Heerwaarden, B., Sgro, C.M., and Hoffmann, A.A. (2009). Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science* 325, 1244–1246.
- Kelly, K.M. (1990). Gm polymorphisms, linguistic affinities, and natural selection in Melanesia. *Curr Anthropol* 31, 201–219.
- Kelly, K.M. (1999). Malaria and immunoglobulins in Pacific prehistory. *Am Anthropol* 101, 806–809.
- Kelly, D.W., Paterson, R.A., Townsend, C.R., Poulin, R., and Tompkins, D.M. (2009). Parasite spillback: A neglected concept in invasion ecology? *Ecology* 90, 2047–2056.
- Kermack, W.O., and McKendrick, A.G. (1933). Contributions to the mathematical theory of epidemics. III. Further studies of the problem of endemicity. *Proc Royal Soc A* 141, 94–122.
- Kettle, D.S. (1995). *Medical and veterinary entomology* (Cambridge: CAB International).
- Kidgell, C., Volkman, S.K., Daily, J., Borevitz, J.O., Plouffe, D., Zhou, Y., Johnson, J.R., Le Roch, K.G., Sarr, O., Ndir, O., et al. (2006). A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog* 2, e57.
- Kilian, B., Özkan, H., Walther, A., Kohl, J., Dagan, T., Salamini, F., and Martin, W. (2007). Molecular Diversity at 18 Loci in 321 Wild and 92 Domesticated Lines Reveal No Reduction of Nucleotide Diversity during *Triticum monococcum* (Einkorn) Domestication: Implications for the Origin of Agriculture. *Mol Biol Evol* 24, 2657–2668.

- Killeen, G.F., Marshall, J.M., Kiware, S.S., South, A.B., Tusting, L.S., Chaki, P.P., and Govella, N.J. (2017). Measuring, manipulating and exploiting behaviours of adult mosquitoes to optimise malaria vector control impact. *BMJ* 2, e000212.
- Kimura, M., Shimizu, Y., Settheetham-Ishida, W., Soemantri, A., Tiwawech, D., Romphruk, A., Duangchan, P., and Ishida, T. (1998). Twenty-seven base pair deletion in erythrocyte band 3 protein gene responsible for Southeast Asian ovalocytosis is not common among Southeast Asians. *Hum Biol* 70, 993–1000.
- Kimura, M., Tamam, M., Soemantri, A., Nakazawa, M., Ataka, Y., Ohtsuka, R., and Ishida, T. (2003). Distribution of a 27-bp deletion in the band 3 gene in South Pacific islanders. *J Hum Genet* 48, 642–645.
- Kirch, P.V. (2000). On the road of the winds: an archaeological history of the Pacific Islands before European contact (Berkeley: University of California Press).
- Kirch, P.V. (2010). Peopling of the Pacific: A holistic anthropological perspective. *Ann Rev Anthropol* 39, 131–148.
- Kirch, P.V., and Hunt, T.L. (1997). Historical Ecology in the Pacific Islands: Prehistoric Environmental and Landscape Change (New Haven, USA: Yale University Press).
- Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S.E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *Am J Trop Med Hyg* 70, 486–498.
- Kitchen, A., Miyamoto, M.M., and Mulligan, C.J. (2008). Utility of DNA viruses for studying human host history: Case study of JC virus. *Mol Phylogenetics Evol* 46, 673–682.
- Kittichai, V., Koepfli, C., Nguitragool, W., Sattabongkot, J., and Cui, L. (2017). Substantial population structure of *Plasmodium vivax* in Thailand facilitates identification of the sources of residual transmission. *PLoS Negl Trop Dis* 11, e0005930.
- Klingbjaer, P., and Moberg, A. (2003). A composite monthly temperature record from Tornedalen in northern Sweden, 1802–2002. *Int J Clim* 23.
- Klinkenberg, E., Takken, W., Huibers, F., and Touré, Y.T. (2003). The phenology of malaria mosquitoes in irrigated rice fields in Mali. *Acta Trop* 85, 71–82.
- Koepfli, C., and Mueller, I. (2017). Malaria epidemiology at the clone level. *Trends Parasitol* 33, 974–985.
- Koepfli, C., Mueller, I., Siba, P., and Felger, I. (2012). *Plasmodium vivax* in Papua New Guinea: High diversity and gene flow among endemic populations signal roadblocks for elimination. *Malar J* 11, O24.
- Koepfli, C., Timinao, L., Antao, T., Barry, A.E., Siba, P., Mueller, I., and Felger, I. (2013). A large *Plasmodium vivax* reservoir and little population structure in the South Pacific. *PLoS ONE* 8, e66041.
- Koepfli, C., Rodrigues, P.T., Antao, T., Orjuela-Sánchez, P., Van den Eede, P., Gamboa, D., van Hong, N., Bendezu, J., Erhart, A., Barnadas, C., et al. (2015). *Plasmodium vivax* diversity and population structure across four continents. *PLoS Negl Trop Dis* 9, e0003872.
- Kolb, A.W., Ané, C., and Brandt, C.R. (2013). Using HSV-1 Genome Phylogenetics to Track Past Human Migrations. *PLoS ONE* 8, e76267.
- Kölzsch, A., and Blasius, B. (2011). Indications of marine bioinvasion from network theory. *The European Physical Journal B* 84, 601–612.
- Korsinczyk, M., Chen, N., Kotecka, B., Saul, A., Rieckmann, K., and Cheng, Q. (2000). Mutations in *Plasmodium falciparum* cytochrome that are associated with atovaquone resistance are located at a putative drug-binding site. *Antimicrob Agents Chemother* 44, 2100–2108.
- Kozuch, O., and Mayer, V. (1975a). Pig kidney epithelial (PS) cells: a perfect tool for the study of flaviviruses and some other arboviruses. *Acta Virol.* 19, 498.
- Kozuch, O., and Mayer, V. (1975b). Pig kidney epithelial (PS) cells: a perfect tool for the study of flaviviruses and some other arboviruses. *Acta Virol.* 19, 498.
- Krotoski, W.A. (1985). Discovery of the hypnozoite and a new theory of malarial relapse. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 79, 1–11.
- Krotoski, W.A. (1989). The hypnozoite and malarial relapse. *Prog Clin Parasitol* 1, 1–19.
- Kupferschmidt, K. (2016). After 40 years, the most important weapon against mosquitoes may be failing. *Science*.
- Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., et al. (2001). Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *The Lancet* 357, 1225–1240.
- Kwiatkowski, D.P. (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet* 77, 171–192.
- Lalremruata, A., Ball, M., Bianucci, R., Welte, B., Nerlich, A.G., Kun, J.F.J., and Pusch, C.M. (2013). Molecular identification of *Falciparum* malaria and human Tuberculosis co-infections in mummies from the Fayum Depression (Lower Egypt). *PLoS ONE* 8, e60307.
- Lambrinos, J.G. (2004). How interactions between ecology and evolution influence contemporary invasion dynamics. *Ecology* 85, 2061–2070.
- Larruga, J.M., Marrero, P., Abu-Amro, K.K., Golubenkov, M.V., and Cabrera, V.M. (2017). Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area. *BMC Evol Biol* 17.
- Larson, G., Cucchi, T., Fujita, M., Matisoo-Smith, E., Robins, J., Anderson, A., Rolett, B., Spriggs, M., Dolman, G., Kim, T.-H., et al. (2007). Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in island southeast Asia and Oceania. *PNAS* 104, 4834–4839.
- Lavergne, S., and Molofsky, J. (2007). Increased genetic variation and evolutionary potential drive the success of an invasive grass. *PNAS* 104, 3883–3888.
- Lawniczak, M.K.N., Emrich, S.J., Holloway, A.K., Regier, A.P., Olson, M., White, B., Redmond, S., Fulton, L., Appelbaum, E., Godfrey, J., et al. (2010). Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330, 512–514.
- Leclerc, M.C., Durand, P., Gauthier, C., Patot, S., Billotte, N., Menegon, M., Severini, C., Ayala, F.J., and Renaud, F. (2004). Meager genetic variability of the human malaria agent *Plasmodium vivax*. *PNAS* 101, 14455–14460.
- Lee, C.E. (2002). Evolutionary genetics of invasive species. *Trends Ecol Evol* 17, 386–391.

- Lee, R.S., and Behr, M.A. (2016). The implications of whole-genome sequencing in the control of tuberculosis. *Ther Adv Infect Dis* 3, 47–62.
- Levander, K.M. (1902). Mittheilungen über *Anopheles claviger* Fabr. in Finland. *Acta Soc Fauna Flora Fenn* 21.
- Levine, J.M., and D'Antonio, C.M. (2003). Forecasting Biological Invasions with Increasing International Trade. *Conserv Biol* 17, 322–326.
- Lewis, M.P., Simons, G.F., and Fennig, C.D. (2015). *Ethnologue: languages of the world* (Dallas, Tex: SIL International).
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Linder, J. (1717). *Tänkar om Frossan och Kin-Kina barken* (Joh. L. S. Horn: Stockholm).
- Linderholm, A. (2016). Ancient DNA: the next generation - chapter and verse. *Biol J Linnean Soc* 117, 150–160.
- Lindroth, S. (1978). *Svensk lärdoms historia. Frihetstiden* (Norstedts Förlag AB: Stockholm).
- Liò, P., and Goldman, N. (1998). Models of molecular evolution and phylogeny. *Genome Res* 8, 1233–1244.
- Lipson, M., Loh, P.-R., Patterson, N., Moorjani, P., Ko, Y.-C., Stoneking, M., Berger, B., and Reich, D. (2014). Reconstructing Austronesian population history in Island Southeast Asia. *Nat Commun* 5, 4689.
- Liu, W., Jin, C.-Z., Zhang, Y.-Q., Cai, Y.-J., Xing, S., Wu, X.-J., Cheng, H., Edwards, R.L., Pan, W.-S., Qin, D.-G., et al. (2010a). Human remains from Zhirendong, South China, and modern human emergence in East Asia. *PNAS* 107, 19201–19206.
- Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson, J.D., Keele, B.F., Ndjongo, J.-B.N., Sanz, C.M., Morgan, D.B., Locatelli, S., et al. (2010b). Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467, 420–425.
- Liu, W., Wu, X., Pei, S., Wu, X., and Norton, C.J. (2010c). Huanglong Cave: A Late Pleistocene human fossil site in Hubei Province, China. *Quatern Int* 211, 29–41.
- Liu, W., Li, Y., Shaw, K.S., Learn, G.H., Plenderleith, L.J., Malenke, J.A., Sundaraman, S.A., Ramirez, M.A., Crystal, P.A., Smith, A.G., et al. (2014). African origin of the malaria parasite *Plasmodium vivax*. *Nat Commun* 5, 3346.
- Livingstone, F.B. (1958). Anthropological implications of sickle cell gene distribution in West Africa. *Am Anthropol* 60, 533–562.
- Livingstone, F.B. (1964). Aspects of the population dynamics of the abnormal hemoglobin and glucose-6-phosphate dehydrogenase deficiency genes. *Am J Hum Genet* 16, 435–450.
- Livingstone, F.B. (1971). Malaria and human polymorphisms. *Annu Rev Genet* 5, 33–64.
- Livingstone, F.B. (2009). *Abnormal Hemoglobins in Human Populations* (New Jersey, USA: Transaction Publishers).
- Loaiza, J.R., Scott, M.E., Bermingham, E., Sanjur, O.I., Wilkerson, R., Rovira, J., Gutiérrez, L.A., Correa, M.M., Grijalva, M.J., Birnberg, L., et al. (2010). Late Pleistocene environmental changes lead to unstable demography and population divergence of *Anopheles albimanus* in the northern Neotropics. *Mol Phylogenetics Evol* 57, 1341–1346.
- Loaiza, J.R., Bermingham, E., Sanjur, O.I., Scott, M.E., Bickersmith, S.A., and Conn, J.E. (2012). Review of genetic diversity in malaria vectors (Culicidae: Anophelinae). *Infect Genet Evol* 12, 1–12.
- Lockwood, J. (2007). *Invasion ecology* (Malden MA: Blackwell Pub.).
- Lockwood, J.L., Cassey, P., and Blackburn, T. (2005). The role of propagule pressure in explaining species invasions. *Trends Ecol Evol* 20, 223–228.
- Logue, K., Chan, E.R., Phipps, T., Small, S.T., Reimer, L., Henry-Halldin, C., Sattabongkot, J., Siba, P.M., Zimmerman, P.A., and Serre, D. (2013). Mitochondrial genome sequences reveal deep divergences among *Anopheles punctulatus* sibling species in Papua New Guinea. *Malar J* 12, 64.
- Logue, K., Small, S.T., Chan, E.R., Reimer, L., Siba, P.M., Zimmerman, P.A., and Serre, D. (2015). Whole-genome sequencing reveals absence of recent gene flow and separate demographic histories for *Anopheles punctulatus* mosquitoes in Papua New Guinea. *Mol Ecol* 24, 1263–1274.
- Lokki, J., Saura, A., Korvenkontio, P., and Ulmanen, I. (1979). Diagnosing adult *Anopheles* mosquitoes. *Aquilo Ser Zool* 20.
- Loo, J.-H., Trejaut, J.A., Yen, J.-C., Chen, Z.-S., Lee, C.-L., and Lin, M. (2011). Genetic affinities between the Yami tribe people of Orchid Island and the Philippine Islanders of the Batanes archipelago. *BMC Genet* 12, 21.
- Loy, D.E., Liu, W., Li, Y., Learn, G.H., Plenderleith, L.J., Sundaraman, S.A., Sharp, P.M., and Hahn, B.H. (2017). Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. *Int J Parasitol* 47, 87–97.
- Lum, J.K. (2007). Contributions of population origins and gene flow to diversity of neutral and malaria selected autosomal genetic loci of Pacific Island populations. In *Genes, Language and Culture History in the Southwest Pacific*, J.S. Friedlaender, ed. (Oxford, UK: Oxford University Press), p.
- Lynch, M., and Crease, T.J. (1990). The analysis of population survey data on DNA sequence variation. *Mol Biol Evol* 7, 377–394.
- Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005a). Tracing modern human origins - Response. *Science* 309, 1995–1996.
- Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005b). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308, 1034–1036.
- MacDonals, G. (1957). *The Epidemiology and Control of Malaria* (London: Oxford University Press).
- Malaspinas, A.-S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., Crawford, J.E., et al. (2016). A genomic history of Aboriginal Australia. *Nature* 538, 207–214.
- Markov, P.V., van de Laar, T.J., Thomas, X.V., Aronson, S.J., Weegink, C.J., van den Berk, G.E., Prins, M., Pybus, O.G., and Schinkel, J. (2012). Colonial history and contemporary transmission shape the genetic diversity of hepatitis C virus genotype 2 in Amsterdam. *J. Virol.* 86, 7677–7687.
- Markus, M.B. (2012). Dormancy in mammalian malaria. *Trends Parasitol* 28, 39–45.

- Marnocha, E., Pollinger, J., and Smith, T.B. (2011). Human-induced morphological shifts in an island lizard. *Evol Applic* 4, 388–396.
- Martens, P., Kovats, R., Nijhof, S., Devries, P., Livermore, M., Bradley, D., Cox, J., and McMichael, A. (1999). Climate change and future populations at risk of malaria. *Global Environmental Change* 9, S89–S107.
- Martin, E., Moutailler, S., Madec, Y., and Failloux, A.-B. (2010). Differential responses of the mosquito *Aedes albopictus* from the Indian Ocean region to two chikungunya isolates. *BMC Ecol* 10, 8.
- Martini, I.P. (2010). *Landscapes and Societies* (Springer).
- Martinsen, E.S., Perkins, S.L., and Schall, J.J. (2008). A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): Evolution of life-history traits and host switches. *Mol Phylogenetics Evol* 47, 261–273.
- Maruvka, Y.E., Shnerb, N.M., Solomon, S., Yaari, G., and Kessler, D.A. (2011). Slicing and dicing the genome: A statistical physics approach to population genetics. *J Stat Phys* 142, 1302–1316.
- Matisoo-Smith, E. (2007). Animal translocations, genetic variation, and the human settlement of the Pacific. In *Genes, Language and Culture History in the Southwest Pacific*, J.S. Friedlaender, ed. (Oxford, UK: Oxford University Press), p.
- Matisoo-Smith, E. (2009). The commensal model for human settlement of the Pacific 10 years on—what can we say and where to now? *J. Isl. Coast Arch.* 4, 151–163.
- Matisoo-Smith, E., and Robins, J.H. (2004). Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *PNAS* 101, 9167–9172.
- Mayr, E. (1944). Wallace's Line in the light of recent zoogeographic studies. *The Quarterly Review of Biology* 19, 1–14.
- McCorriston, J. (2006). Farmers Speak Farnese (Bellwood's First Farmers). *Curr Anthropol* 47, 1047–1048.
- McCutchan, T., Dame, J., Miller, L., and Barnwell, J. (1984). Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. *Science* 225, 808–811.
- McCutchan, T.F., Kissinger, J.C., Touray, M.G., Rogers, M.J., Li, J., Sullivan, M., Braga, E.M., Krettli, A.U., and Miller, L.H. Comparison of circumsporozoite proteins from avian and mammalian malaras: biological and phylogenetic implications. *PNAS* 11889.
- McDevitt, A.D., Rambau, R.V., O'Brien, J., McDevitt, C.D., Hayden, T.J., and Searle, J.B. (2009). Genetic variation in Irish pygmy shrews *Sorex minutus* (Soricomorpha: Soricidae): implications for colonization history. *Biol J Linnean Soc* 97, 918–927.
- McGready, R., Lee, S., Wiladphaingern, J., Ashley, E., Rijken, M., Boel, M., Simpson, J., Paw, M., Pimanpanarak, M., Mu, O., et al. (2012). Adverse effects of falciparum and vivax malaria and the safety of antimalarial treatment in early pregnancy: A population-based study. *The Lancet Infectious Diseases* 12, 388–396.
- McNeill, W.H. (1998). *Plagues and Peoples* (New York, USA: Anchor Press : Doubleday).
- Mehlotra, R.K., Lorry, K., Kastens, W., Miller, S.M., Alpers, M.P., Bockarie, M., Kazura, J.W., and Zimmerman, P.A. (2000). Random distribution of mixed species malaria infections in Papua New Guinea. *Am J Trop Med Hyg* 62, 225–231.
- Mehta, A., Mason, P.J., and Vulliamy, T.J. (2000). Glucose-6-phosphate dehydrogenase deficiency. *Bail Cl Haem* 13, 21–38.
- Menard, D., Barnadas, C., Bouchier, C., Henry-Halldin, C., Gray, L.R., Ratsimbao, A., Thonier, V., Carod, J.-F., Domarle, O., Colin, Y., et al. (2010). *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *PNAS* 107, 5967–5971.
- Mendes, C., Dias, F., Figueiredo, J., Mora, V.G., Cano, J., de Sousa, B., do Rosário, V.E., Benito, A., Berzosa, P., and Arez, A.P. (2011). Duffy negative antigen is no longer a barrier to *Plasmodium vivax* – Molecular evidences from the African west coast (Angola and Equatorial Guinea). *PLoS Negl Trop Dis* 5, e1192.
- Menendez, C. (1995). Malaria during pregnancy: A priority area of malaria research and control. *Parasitology Today* 11, 178–183.
- Menéndez, C., D'Alessandro, U., and ter Kuile, F.O. (2007). Reducing the burden of malaria in pregnancy by preventive strategies. *The Lancet Infectious Diseases* 7, 126–135.
- Merriwether, D.A., Friedlaender, J.S., Mediavilla, J., Mgone, C., Gentz, F., and Ferrell, R.E. (1999). Mitochondrial DNA variation is an indicator of Austronesian influence in Island Melanesia. *Am J Phys Anthropol* 110, 243–270.
- Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T.P., et al. (2004). Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5, 26.
- Mgone, C.S., Koki, G., Paniu, M.M., Kono, J., Bhatia, K.K., Genton, B., Alexander, N.D.E., and Alpers, M.P. (1996). Occurrence of the erythrocyte band 3 (AE1) gene deletion in relation to malaria endemicity in Papua New Guinea. *Trans. Roy. Soc. Trop. Med. Hyg.* 90, 228–231.
- Miao, M., Yang, Z., Cui, L., Ahlum, J., Huang, Y., and Cui, L. (2010). Different allele prevalence in the dihydrofolate reductase and dihydropteroate synthase genes in *Plasmodium vivax* populations from China. *Am J Trop Med Hyg* 83, 1206–1211.
- Miao, M., Yang, Z., Patch, H., Huang, Y., Escalante, A.A., and Cui, L. (2012). *Plasmodium vivax* populations revisited: Mitochondrial genomes of temperate strains in Asia suggest ancient population expansion. *BMC Evol Biol* 12, 22.
- Miles, A., Harding, N.J., Bottà, G., Clarkson, C.S., Antão, T., Kozak, K., Schrider, D.R., Kern, A.D., Redmond, S., Sharakhov, I., et al. (2017). Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*.
- Miller, L.H., Mason, S.J., Clyde, D.F., and McGinniss, M.H. (1976). The Resistance Factor to *Plasmodium vivax* in Blacks: The Duffy-Blood-Group Genotype, *FyFy*. *N Engl J Med* 295, 302–304.
- Miller, L.H., Baruch, D.I., Marsh, K., and Doumbo, O.K. (2002). The pathogenic basis of malaria. *Nature* 415, 673–679.
- Mirabal, S., Herrera, K.J., Gayden, T., Regueiro, M., Underhill, P.A., Garcia-Bertrand, R.L., and Herrera, R.J. (2012). Increased Y-chromosome resolution of haplogroup O suggests genetic ties between the Ami aborigines of Taiwan and the Polynesian Islands of Samoa and Tonga. *Gene* 492, 339–348.

- Mitsui, H., Arisue, N., Sakihama, N., Inagaki, Y., Horii, T., Hasegawa, M., Tanabe, K., and Hashimoto, T. (2010). Phylogeny of Asian primate malaria parasites inferred from apicoplast genome-encoded genes with special emphasis on the positions of *Plasmodium vivax* and *P. fragile*. *Gene* 450, 32–38.
- Molina-Cruz, A., and Barillas-Mury, C. (2014). The remarkable journey of adaptation of the *Plasmodium falciparum* malaria parasite to New World anopheline mosquitoes. *Mem Inst Oswaldo Cruz* 109, 662–667.
- Molina-Cruz, A., Canepa, G.E., Kamath, N., Pavlovic, N.V., Mu, J., Ramphul, U.N., Ramirez, J.L., and Barillas-Mury, C. (2015). *Plasmodium* evasion of mosquito immunity and global malaria transmission: The lock-and-key theory. *PNAS* 112, 15178–15183.
- Mona, S., Tommaseo-Ponzetta, M., Brauer, S., Sudoyo, H., Marzuki, S., and Kayser, M. (2007). Patterns of Y-chromosome diversity intersect with the Trans-New Guinea Hypothesis. *Mol Biol Evol* 24, 2546–2555.
- Mona, S., Grunz, K.E., Brauer, S., Pakendorf, B., Castri, L., Sudoyo, H., Marzuki, S., Barnes, R.H., Schmidtke, J., Stoneking, M., et al. (2009). Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol Biol Evol* 26, 1865–1877.
- Montenegro, Á., Callaghan, R.T., and Fitzpatrick, S.M. (2016). Using seafaring simulations and shortest-hop trajectories to model the prehistoric colonization of Remote Oceania. *PNAS* 113, 12685–12690.
- Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H.M., Breurec, S., Wu, J.-Y., Maady, A., Bernhöft, S., Thiberge, J.-M., Phuanukoonnon, S., et al. (2009). The peopling of the Pacific from a bacterial perspective. *Science* 323, 527–530.
- Moodley, Y., Linz, B., Bond, R.P., Nieuwoudt, M., Soodyall, H., Schlebusch, C.M., Bernhöft, S., Hale, J., Suerbaum, S., Mugisha, L., et al. (2012). Age of the association between *Helicobacter pylori* and man. *PLoS Pathog* 8, e1002693.
- Moritz, C., Case, T., Bolger, D., and Donnellan, S. (1993). Genetic diversity and the history of pacific island house geckos (*Hemidactylus* and *Lepidodactylus*). *Biol J Linnean Soc* 48, 113–133.
- Morley, M.W. (2017). The geoarchaeology of hominin dispersals to and from tropical Southeast Asia: A review and prognosis. *Journal of Archaeological Science* 77, 78–93.
- Morley, M.W., Goldberg, P., Sutikna, T., Tocheri, M.W., Prinsloo, L.C., Jatmiko, Saptomo, E.W., Wasisto, S., and Roberts, R.G. (2017). Initial micromorphological results from Liang Bua, Flores (Indonesia): Site formation processes and hominin activities at the type locality of *Homo floresiensis*. *J Archaeol Sci* 77, 125–142.
- Morwood, M.J., Soejono, R.P., Roberts, R.G., Sutikna, T., Turney, C.S.M., Westaway, K.E., Rink, W.J., Zhao, J.-x., van den Bergh, G.D., Due, R.A., et al. (2004). Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature* 431, 1087–1091.
- Morwood, M.J., Brown, P., Jatmiko, Sutikna, T., Wahyu Saptomo, E., Westaway, K.E., Awe Due, R., Roberts, R.G., Maeda, T., Wasisto, S., et al. (2005). Further evidence for small-bodied hominins from the Late Pleistocene of Flores, Indonesia. *Nature* 437, 1012–1017.
- Msaïdie, S., Ducourneau, A., Boetsch, G., Longepied, G., Papa, K., Allibert, C., Yahaya, A.A., Chiaroni, J., and Mitchell, M.J. (2011). Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *Eur J Hum Genet* 19, 89–94.
- Mu, J., Awadalla, P., Duan, J., McGee, K.M., Joy, D.A., McVean, G.A.T., and Su, X. (2005a). Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biology* 3, e335.
- Mu, J.B., Joy, D.A., Duan, J.H., Huang, Y.M., Carlton, J., Walker, J., Barnwell, J., Beerli, P., Charleston, M.A., Pybus, O.G., et al. (2005b). Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol Biol Evol* 22, 1686–1693.
- Mueller, I., Galinski, M.R., Baird, J.K., Carlton, J.M., Kochar, D.K., Alonso, P.L., and del Portillo, H.A. (2009). Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infect Dis* 9, 555–566.
- Müller, I., Bockarie, M., Alpers, M., and Smith, T. (2003). The epidemiology of malaria in Papua New Guinea. *Trends Parasitol* 19, 253–259.
- Müller, I., Genton, B., Rare, L., Kiniboro, B., Kastens, W., Zimmerman, P., Kazura, J., Alpers, M., and Smith, T.A. (2009a). Three different *Plasmodium* species show similar patterns of clinical tolerance of malaria infection. *Malar J* 8, 158.
- Müller, I., Widmer, S., Michel, D., Maraga, S., McNamara, D.T., Kiniboro, B., Sie, A., Smith, T.A., and Zimmerman, P.A. (2009b). 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.
- Mundfrom, D.J., Shaw, D.G., and Ke, T.L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing* 5, 159–168.
- Mwingira, F., Nkwengulila, G., Schoepflin, S., Sumari, D., Beck, H.-P., Snounou, G., Felger, I., Oliaro, P., and Mugittu, K. (2011). *Plasmodium falciparum* msp1, msp2 and glurp allele frequency and diversity in sub-Saharan Africa. *Malar J* 10, 79.
- Nagy, L.G., Kocsubé, S., Csanádi, Z., Kovács, G.M., Petkovits, T., Vágvölgyi, C., and Papp, T. (2012). Re-mind the gap! Insertion – deletion data reveal neglected phylogenetic potential of the nuclear ribosomal internal transcribed spacer (ITS) of Fungi. *PLoS ONE* 7, e49794.
- Neafsey, D.E., Hartl, D.L., and Berriman, M. (2005). Evolution of noncoding and silent coding sites in the *Plasmodium falciparum* and *Plasmodium reichenowi* genomes. *Mol Biol Evol* 22, 1621–1626.
- Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner, D.A., Lukens, A., Rosen, D., Daniels, R., Houde, N., et al. (2008). Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol* 9, R171.
- Neafsey, D.E., Galinsky, K., Jiang, R.H.Y., Young, L., Sykes, S.M., Saif, S., Gujja, S., Goldberg, J.M., Young, S., Zeng, Q., et al. (2012). The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat Genet* 44, 1046–1050.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *PNAS* 70, 3321–3323.
- Nei, M., and Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS* 76, 5269–5273.

- Nielsen, K.M., Kasper, J., Choi, M., Bedford, T., Kristiansen, K., Wirth, D.F., Volkman, S.K., Lozovsky, E.R., and Hartl, D.L. (2003). Gene conversion as a source of nucleotide diversity in *Plasmodium falciparum*. *Mol Biol Evol* 20, 726–734.
- Norberg, P., Kasubi, M.J., Haarr, L., Bergström, T., and Liljeqvist, J.-Å. (2007). Divergence and recombination of clinical Herpes Simplex Virus Type 2 isolates. *J Virol* 81, 13158–13167.
- Nosten, F., McGready, R., Simpson, J., Thwai, K., Balkan, S., Cho, T., Hkirijsaroen, L., Looareesuwan, S., and White, N. (1999). Effects of *Plasmodium vivax* malaria in pregnancy. *The Lancet* 354, 546–549.
- Novak, S.J. (2007). The role of evolution in the invasion process. *PNAS* 104, 3671–3672.
- Novembre, J., and Han, E. (2012). Human population structure and the adaptive response to pathogen-induced selection pressures. *Phil Trans R Soc B* 367, 878–886.
- Noviyanti, R., Coutrier, F., Utami, R.A.S., Trimarsanto, H., Tirta, Y.K., Trianty, L., Kusuma, A., Sutanto, I., Kosasih, A., Kusriastuti, R., et al. (2015). Contrasting transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum*: Implications for malaria control and elimination. *PLoS Negl Trop Dis* 9, e0003739.
- Nozais, J.P. (1988). Malaria in the Mediterranean world. The historical and present-day distribution. *Bull Soc Pathol Exot Filiales* 81, 854–860.
- O’Brien, M.J., and Laland, K.N. (2012). Genes, culture, and agriculture. An example of human niche construction. *Curr Anthropol* 53, 434–470.
- Obsomer, V., Defourny, P., and Coosemans, M. (2007). The *Anopheles dirus* complex: Spatial distribution and environmental drivers. *Malar J* 6, 26.
- Obsomer, V., Defourny, P., and Coosemans, M. (2012). Predicted distribution of major malaria vectors belonging to the *Anopheles dirus* Complex in Asia: Ecological niche and environmental influences. *PLoS ONE* 7, e50475.
- Ochomo, E., Chahilu, M., Cook, J., Kinyari, T., Bayoh, N.M., West, P., Kamau, L., Osangale, A., Ombok, M., Njagi, K., et al. (2017). Insecticide-treated nets and protection against insecticide-resistant malaria vectors in Western Kenya. *Emerg Infect Dis* 23, 758–764.
- O’Connor, S., Clark, G.R., Leach, F., Anderson, A., and ANU E Press. (2008). Islands of inquiry colonisation, seafaring and the archaeological of maritime landscapes (Canberra, ACT, Australia: ANU E Press).
- Ogden, T.H., and Rosenberg, M.S. (2007). How should gaps be treated in parsimony? A comparison of approaches using simulation. *Mol Phylogenetics Evol* 42, 817–826.
- Ohba, S.Y., Matsuo, T., and Takagi, M. (2012). Mosquitoes and other aquatic insects in fallow field biotopes and rice paddy fields. *Med Vet Entomol* 96–103.
- Oppenheimer, S. (2006). Response to Peter Bellwood and Jared Diamond (2005): “On explicit ‘replacement’ models in Island Southeast Asia: A reply to Stephen Oppenheimer.” *World Archaeology* 38, 715–717.
- Oppenheimer, S. (2012). Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Phil Trans R Soc B* 367, 770–784.
- O’Rourke, D.H. (2009). Human migrations: The two roads taken. *Curr Biol* 19, R203–R205.
- Oskarsson, M.C.R., Klutsch, C.F.C., Boonyaparakob, U., Wilton, A., Tanabe, Y., and Savolainen, P. (2011). Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proc Royal Soc B* 279, 967–974.
- van Oven, M., Brauer, S., Choi, Y., Ensing, J., Schiefenhövel, W., Stoneking, M., and Kayser, M. (2014). Human genetics of the Kula Ring: Y-chromosome and mitochondrial DNA variation in the Massim of Papua New Guinea. *Eur J Hum Genet* 22, 1393–1403.
- Pääbo, S., Gifford, J.A., and Wilson, A.C. (1988). Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Research* 16, 9775–9787.
- Paquette, A.M., Harahap, A., Laosombat, V., Patnode, J.M., Satyagraha, A., Sudoyo, H., Thompson, M.K., Yusoff, N.M., and Wilder, J.A. (2015). The evolutionary origins of Southeast Asian Ovalocytosis. *Infect Genet Evol* 34, 153–159.
- Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S.Y.W., Belshaw, R., Allain, J.-P., and Hatzakis, A. (2013). Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* 57, 908–916.
- Parham, P., Norman, P.J., Abi-Rached, L., and Guethlein, L.A. (2012). Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Phil Trans R Soc B* 367, 800–811.
- Parise, M., Nahlen, B., Menendez, C., and Steketee, R. (2001). The burden of malaria in pregnancy in malaria-endemic areas. *The American Journal of Tropical Medicine and Hygiene* 64, 28–35.
- Patel, S.S., King, C.L., Mgone, C.S., Kazura, J.W., and Zimmerman, P.A. (2004). Glycophorin C (Gerbig antigen blood group) and band 3 polymorphisms in two malaria holoendemic regions of Papua New Guinea. *Am J Hematol* 75, 1–5.
- Paul, R.E.L., Diallo, M., and Brey, P.T. (2004). Mosquitoes and transmission of malaria parasites – not just vectors. *Malar J* 3.
- Pawley, A., Attenborough, R., Golson, J., and Hide, R. (2005). *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples* (Canberra, Australia: Pacific Linguistics).
- Pearson, R.D., Amato, R., Auburn, S., Miotto, O., Almagro-Garcia, J., Amaratunga, C., Suon, S., Mao, S., Noviyanti, R., Trimarsanto, H., et al. (2016). Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat Genet* 48, 959–964.
- Peel, M.C., Finlayson, B.L., and McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644.
- Peng, M.-S., Quang, H.H., Dang, K.P., Trieu, A.V., Wang, H.-W., Yao, Y.-G., Kong, Q.-P., and Zhang, Y.-P. (2010). Tracing the Austronesian footprint in mainland Southeast Asia: A perspective from mitochondrial DNA. *Mol Biol Evol* 27, 2417–2430.
- Pennisi, E. (2001). Malaria’s beginnings: On the heels of hoes? *Science* 293, 416–417.
- Pepperell, C.S., Casto, A.M., Kitchen, A., Granka, J.M., Cornejo, O.E., Holmes, E.C., Birren, B., Galagan, J., and Feldman, M.W. (2013). The role of selection in shaping diversity of natural *M. tuberculosis* populations. *PLoS Pathog* 9, e1003543.

- Perkins, S.L. (2008). Molecular systematics of the three mitochondrial protein-coding genes of malaria parasites: corroborative and new evidence for the origins of human malaria. *Mitochondrial DNA* 19, 471–478.
- Perkins, S.L., and Schall, J. (2002). A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J Parasitol* 88, 972–978.
- Petruglia, M., Korisettar, R., Boivin, N., Clarkson, C., Ditchfield, P., Jones, S., Koshy, J., Lahr, M.M., Oppenheimer, C., Pyle, D., et al. (2007). Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. *Science* 317, 114–116.
- Petruglia, M.D., Haslam, M., Fuller, D.Q., Boivin, N., and Clarkson, C. (2010). Out of Africa: New hypotheses and evidence for the dispersal of *Homo sapiens* along the Indian Ocean rim. *Ann Hum Biol* 37, 288–311.
- Petruglia, M.D., Alsharekh, A.M., Crassard, R., Drake, N.A., Groucutt, H., Parker, A.G., and Roberts, R.G. (2011). Middle Paleolithic occupation on a Marine Isotope Stage 5 lakeshore in the Nefud Desert, Saudi Arabia. *Quat Sci Rev* 30, 1555–1559.
- Petren, K., Grant, B.R., and Grant, P.R. (1999). A phylogeny of Darwin's finches based on microsatellite DNA length variation. *Proc Royal Soc B* 266, 321–329.
- Pickford, M.D., and Turner, J.H. (2012). Standard Operating Procedure for In-house Preparation of 131I-rituximab for Radioimmunotherapy of Non-Hodgkin's Lymphoma. *World J Nucl Med* 11, 105–109.
- Piper, P.J., Hung, H., Campos, F.Z., Bellwood, P., and Santiago, R. (2009). A 4000 year-old introduction of domestic pigs into the Philippine Archipelago: implications for understanding routes of human migration through Island Southeast Asia and Wallacea. *Antiquity* 83, 687–695.
- Pitblado, B.L. (2011). A tale of two migrations: Reconciling recent biological and archaeological evidence for the Pleistocene peopling of the Americas. *J Arch Res* 19, 327–375.
- Pocock, M.J.O., Hauffe, H.C., and Searle, J.B. (2005). Dispersal in house mice. *Biol J Linnean Soc* 84, 565–583.
- Poirier, P., Doderer-Lang, C., Atchade, P.S., Lemoine, J.-P., de l'Isle, M.-L.C., Abou-bacar, A., Pfaff, A.W., Brunet, J., Arnoux, L., Haar, E., et al. (2016). The hide and seek of *Plasmodium vivax* in West Africa: report from a large-scale study in Beninese asymptomatic subjects. *Malar J* 15.
- Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25, 1253–1256.
- Preston, M.D., Campino, S., Assefa, S.A., Echeverry, D.F., Ocholla, H., Amambua-Ngwa, A., Stewart, L.B., Conway, D.J., Borrmann, S., Michon, P., et al. (2014). A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. *Nat Commun* 5.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26, 1641–1650.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, e9490.
- Pritchard, J.K., and Cox, N.J. (2002). The allelic architecture of human disease genes: common disease–common variant... or not? *Hum Mol Genet* 11, 2417–2423.
- Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Pritchard, J.K., Pickrell, J.K., and Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Curr Biol* 20, R208–R215.
- Prugnolle, F., Manica, A., and Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Curr Biol* 15, R159–R160.
- Prugnolle, F., Rougeron, V., Becquart, P., Berry, A., Makanga, B., Rahola, N., Arnathau, C., Ngoubangoye, B., Menard, S., Willaume, E., et al. (2013). Diversity, host switching and evolution of *Plasmodium vivax* infecting African great apes. *PNAS* 110, 8123–8128.
- Putaporntip, C., Jongwutiwes, S., Sakihama, N., Ferreira, M.U., Kho, W.-G., Kaneko, A., Kanbara, H., Hattori, T., and Tanabe, K. (2002). Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. *PNAS* 99, 16348–16353.
- Pybus, O.G., Markov, P.V., Wu, A., and Tatem, A.J. (2007). Investigating the endemic transmission of the hepatitis C virus. *Int J Parasitol* 37, 839–849.
- Qari, S.H., Shi, Y.P., Pieniazek, N.J., Collins, W.E., and Lal, A.A. (1996). Phylogenetic relationship among the malaria parasites based on small subunit rRNA gene sequences: monophyletic nature of the human malaria parasite, *Plasmodium falciparum*. *Mol Phylogenetics Evol* 6, 157–165.
- Rabe, T., Jambou, R., Rabarijaona, L., Raharimalala, L., Rason, M.A., Arie, F., and Dhermy, D. (2002). South-East Asian ovalocytosis among the population of the Highlands of Madagascar: a vestige of the island's settlement. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 96, 143–144.
- Raghavan, M., Steinrucken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Avila-Arcos, M.C., Malaspina, A.-S., et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349, aab3884–aab3884.
- Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *PNAS* 102, 15942–15947.
- Rambaut, A., Suchard, M.A., Xie, D., and Drummond, A.J. (2014). Tracer v1.6, Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334, 94–98.
- Rathore, D., Wahl, A.M., Sullivan, M., and McCutchan, T.F. (2001). A phylogenetic comparison of gene trees constructed from plastid, mitochondrial and genomic DNA of *Plasmodium* species. *Mol Biochem Parasitol* 114, 89–94.

- Rayner, J.C., Liu, W., Peeters, M., Sharp, P.M., and Hahn, B.H. (2011). A plethora of *Plasmodium* species in wild apes: a source of human infection? *Trends Parasitol* 27, 222–229.
- Redd, A.J., and Stoneking, M. (1999). Peopling of Sahul: mtDNA Variation in Aboriginal Australian and Papua New Guinean Populations. *Am J Hum Genet* 65, 808–828.
- Redd, A.J., Takezaki, N., Sherry, S.T., McGarvey, S.T., Sofro, A.S., and Stoneking, M. (1995). Evolutionary history of the COII/tRNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 12, 604–615.
- Reiff, D.M., Kaneko, A., Taleo, G., Amos, M., and Lum, J.K. (2007). Population structure and gene flow of *Anopheles farauti* s.s. (Diptera: Culicidae) among ten sites on five islands of Vanuatu: Implications for malaria control. *J Med Entomol* 44, 601–607.
- Reiter, P. (2001). Climate change and mosquito-borne diseases. *Environ Health Perspect* 109.
- Remm, H.J. (1957). Materialy po faune i ekologii komarov (Diptera, Culicidae) Estonskoi SSR. *Entomol Obozr* 36.
- Renkonen, K.O. (1944). Über das Vorkommen von Malaria in Finnland. *Ada Med Scand* 69.
- Reuter, J.A., Spacek, D.V., Pai, R.K., and Snyder, M.P. (2016). Simul-seq: combined DNA and RNA sequencing for whole-genome and transcriptome profiling. *Nat Meth* 13, 953–958.
- Reyes-Centeno, H. (2016). Out of Africa and into Asia: Fossil and genetic evidence on modern human origins and dispersals. *Quatern Int* 416, 249–262.
- Reyes-Centeno, H., Hubbe, M., Hanihara, T., Stringer, C., and Harvati, K. (2015). Testing modern human out-of-Africa dispersal models and implications for modern human origins. *J Hum Evol* 87, 95–106.
- Reynolds, C.H. (1975). The NAD-linked alpha-glycerophosphate dehydrogenase of trypanosomes. *Biochem Biophys Res Commun* 67, 538–543.
- Rheindt, F.E., and Edwards, S.V. (2011). Genetic introgression: An integral but neglected component of speciation in birds. *The Auk* 128, 620–632.
- Rich, S.M., Licht, M.C., Hudson, R.R., and Ayala, F.J. (1998). Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *PNAS* 95, 4425–4430.
- Rich, S.M., Leendertz, F.H., Xu, G., LeBreton, M., Djoko, C.F., Aminake, M.N., Takang, E.E., Diffo, J.L.D., Pike, B.L., Rosenthal, B.M., et al. (2009). The origin of malignant malaria. *PNAS* 106, 14902–14907.
- Ricklefs, R.E., and Outlaw, D.C. (2010). A molecular clock for malaria parasites. *Science* 329, 226–229.
- Rieth, T.M., Hunt, T.L., Lipo, C., and Wilmshurst, J.M. (2011). The 13th century polynesian colonization of Hawai'i Island. *J Archaeol Sci* 38, 2740–2749.
- Roberts, R.G., Westaway, K.E., Zhao, J.-x., Turney, C.S.M., Bird, M.I., Rink, W.J., and Fifield, L.K. (2009). Geochronology of cave deposits at Liang Bua and of adjacent river terraces in the Wae Racang valley, western Flores, Indonesia: a synthesis of age estimates for the type locality of *Homo floresiensis*. *J Hum Evol* 57, 484–502.
- Rocha, S., Posada, D., Carretero, M.A., and Harris, D.J. (2007). Phylogenetic affinities of Comoroan and East African day geckos (genus *Phelsuma*): Multiple natural colonisations, introductions and island radiations. *Mol Phylogenetics Evol* 43, 685–692.
- Rocha, S., Carretero, M.A., and Harris, D.J. (2010). On the diversity, colonization patterns and status of *Hemidactylus* spp. (Reptilia: Gekkonidae) from the Western Indian Ocean islands. *The Herpetological Journal* 20, 83–89.
- Rodrigues, P.T., Alves, J.M.P., Santamaria, A.M., Calzada, J.E., Xayavong, M., Parise, M., Silva, A.J. da, and Ferreira, M.U. (2014). Using mitochondrial genome sequences to track the origin of imported *Plasmodium vivax* infections diagnosed in the United States. *Am J Trop Med Hyg* 90, 1102–1108.
- Rodrigues, P.T., Valdivia, H.O., Oliveira, T.C., Alves, J.M.P., Duarte, A.M.R.C., Cerutti-Junior, C., Buery, J.C., Brito, C.F.A., Souza, J.C., Hirano, Z.M.B., et al. (2017). Human migration and the spread of malaria parasites to the New World.
- Roetzer, A., Diel, R., Kohl, T.A., Rückert, C., Nübel, U., Blom, J., Wirth, T., Jaenicke, S., Schuback, S., Rüscher-Gerdes, S., et al. (2013). Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 10, e1001387.
- Rogers, D.J., and Randolph, S.E. (2003). Studying the global distribution of infectious diseases using GIS and RS. *Nat Rev Micro* 1, 231–237.
- Rogerson, S.J., Hviid, L., Duffy, P.E., Leke, R.F., and Taylor, D.W. (2007). Malaria in pregnancy: Pathogenesis and immunity. *The Lancet Infectious Diseases* 7, 105–117.
- Romaine, S. (2013). Linguistics and ecological diversity. In *The Oxford Handbook of Sociolinguistics*, R. Bayley, R. Cameron, and C. Lucas, eds. (Oxford, UK: Oxford University Press), p.
- Rosanas-Urgell, A., Lin, E., Manning, L., Rarau, P., Laman, M., Senn, N., Grimberg, B.T., Tavul, L., Stanicic, D.I., Robinson, L.J., et al. (2012). Reduced risk of *Plasmodium vivax* malaria in Papua New Guinean children with Southeast Asian Ovalocytosis in two cohorts and a case-control study. *PLoS Med* 9, e1001305.
- Rose, J.I., Usik, V.I., Marks, A.E., Hilbert, Y.H., Galletti, C.S., Parton, A., Geiling, J.M., Černý, V., Morley, M.W., and Roberts, R.G. (2011). The Nubian Complex of Dhofar, Oman: an African middle stone age industry in Southern Arabia. *PLoS ONE* 6, e28239.
- Rosenberg, N.A. (2002). Genetic structure of human populations. *Science* 298, 2381–2385.
- Roskoski, R., Lim, C.T., and Roskoski, L.M. (1975). Human brain and placental choline acetyltransferase: purification and properties. *Biochemistry* 14, 5105–5110.
- Roy, S.W., and Irimia, M. (2008). Origins of human malaria: rare genomic changes and full mitochondrial genomes confirm the relationship of *Plasmodium falciparum* to other mammalian parasites but complicate the origins of *Plasmodium vivax*. *Mol Biol Evol* 25, 1192–1198.
- Rozas, J., and Rozas, R. (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput Appl Biosci* 11, 621–625.
- Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497.

- Rutledge, G.G., Böhme, U., Sanders, M., Reid, A.J., Cotton, J.A., Maiga-Ascofare, O., Djimdé, A.A., Apinjoh, T.O., Amenga-Etego, L., Manske, M., et al. (2017). *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature* 542, 101–104.
- Ryan, J.R., Stoute, J.A., Amon, J., Dunton, R.F., Mtalib, R., Koros, J., Owour, B., Luckhart, S., Wirtz, R.A., Barnwell, J.W., 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.
- Ryschkewitsch, C.F., Friedlaender, J.S., Mgone, C.S., Jobes, D.V., Agostini, H.T., Chima, S.C., Alpers, M.P., Koki, G., Yanagihara, R., and Stoner, G.L. (2000). Human polyomavirus JC variants in Papua New Guinea and Guam reflect ancient population settlement and viral evolution. *Microb Infect* 2, 987–996.
- Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R., and Martin, W. (2002). Genetics and geography of wild cereal domestication in the Near East. *Nat Rev Genet* 3, 429–441.
- Samarawickrema, W.A., Sone, F., and Cummings, R.F. (2009). Seasonal abundance, diel biting activity and parity of *Aedes polynesiensis* marks and *A. samoanus* (Grünberg) (Diptera: Culicidae) in Samoa. *Bull Entomol Res* 77, 191.
- Sangula, A.K., Belsham, G.J., Muwanika, V.B., Heller, R., Balinda, S.N., Masembe, C., and Siegismund, H.R. (2010). Evolutionary analysis of foot-and-mouth disease virus serotype SAT 1 isolates from east africa suggests two independent introductions from southern africa. *BMC Evol Biol* 10, 371.
- Sastu, U.R., Abdullah, N.R., Norahmad, N.A., Saat, M.N.F., Muniandy, P.K., Jelip, J., Tikuson, M., Yusof, N., and Sidek, H.M. (2016). Mutations of *pvdhfr* and *pvdhps* genes in *vivax* endemic-malaria areas in Kota Marudu and Kalabakan, Sabah. *Malar J* 15.
- Sattabongkot, J., Tsuboi, T., Zollner, G.E., Sirichaisinthop, J., and Cui, L. (2004). *Plasmodium vivax* transmission: chances for control? *Trends Parasitol* 20, 192–198.
- Schmedes, S.E., Patel, D., Kelley, J., Udhayakumar, V., and Talundzic, E. (2019). Using the *Plasmodium* mitochondrial genome for classifying mixed-species infections and inferring the geographical origin of *P. falciparum* parasites imported to the U.S. *PLoS ONE* 14, e0215754.
- Schmidt, E.R., Foley, D.H., Bugoro, H., and Bryan, J.H. (2007). A morphological study of the *Anopheles punctulatus* group (Diptera: Culicidae) in the Solomon Islands, with a description of *Anopheles (Cellia) irenicus* Schmidt, sp.n. *Bull Entomol Res* 93.
- van Schouwenburg, P.A., Davenport, E.E., Kienzler, A.-K., Marwah, I., Wright, B., Lucas, M., Malinauskas, T., Martin, H.C., Lockstone, H.E., Cazier, J.-B., et al. (2015). Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders. *Clin Immunol* 160, 301–314.
- Searle, J.B., Kotlik, P., Rambau, R.V., Markova, S., Herman, J.S., and McDevitt, A.D. (2009a). The Celtic fringe of Britain: insights from small mammal phylogeography. *Proc Royal Soc B* 276, 4287–4294.
- Searle, J.B., Jones, C.S., Gündüz, İ., Scascitelli, M., Jones, E.P., Herman, J.S., Rambau, R.V., Noble, L.R., Berry, R.J., Giménez, M.D., et al. (2009b). Of mice and (Viking?) men: Phylogeography of British and Irish house mice. *Proc Royal Soc B* 276, 201–207.
- Seehausen, O., Takimoto, G., Roy, D., and Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol Ecol* 17, 30–44.
- Senn, N., Rarau, P., Stanisic, D.I., Robinson, L., Barnadas, C., Manong, D., Salib, M., Iga, J., Tarongka, N., Ley, S., et al. (2012). Intermittent preventive treatment for malaria in Papua New Guinean infants exposed to *Plasmodium falciparum* and *P. vivax*: A randomized controlled trial. *PLoS Med* 9, e1001195.
- Serjeantson, S.W. (1984). Migration and admixture in the Pacific: Insights provided by human leukocyte antigens. *J Pac Hist* 19, 160–171.
- Serjeantson, S.W., Ryan, D.P., and Thompson, A.R. (1982). The colonization of the Pacific: The story according to human leukocyte antigens. *Am J Hum Genet* 34, 904–918.
- Serjeantson, S.W., Kirk, R.L., and Booth, P.B. (1983). Linguistic and genetic differentiation in New Guinea. *J Hum Evol* 12, 77–92.
- Serjeantson, S.W., Board, P.G., and Bhatia, K.K. (1992). Population genetics in Papua New Guinea: A perspective on human evolution. In *Human Biology in Papua New Guinea: The Small Cosmos*, R.D. Attenborough, and M.P. Alpers, eds. (Oxford, UK: Clarendon Press), pp. 198–233.
- Service, M.W. (1973). The biology of *Anopheles claviger* Meigen (Diptera: Culicidae) in southern England. *Bull Entomol Res* 63.
- Shackelton, L.A., Rambaut, A., Pybus, O.G., and Holmes, E.C. (2006). JC virus evolution and its association with human populations. *J Virol* 80, 9928–9933.
- Sharp, P.M., and Simmonds, P. (2011). Evaluating the evidence for virus/host co-evolution. *Curr Opin Virol* 1, 436–441.
- Sharp, P.M., Liu, W., Learn, G.H., Rayner, J.C., Peeters, M., and Hahn, B.H. (2011). Source of the human malaria parasite *Plasmodium falciparum*. *PNAS* 108, E744–E745.
- Shen, H.-M., Chen, S.-B., Wang, Y., Xu, B., Abe, E.M., and Chen, J.-H. (2017). Genome-wide scans for the identification of *Plasmodium vivax* genes under positive selection. *Malar J* 16.
- Sheppard, P.J. (2011). Lapita colonization across the Near/Remote Oceania boundary. *Curr Anthropol* 52, 799–840.
- Sheppard, P.J., Thomas, T., Summerhayes, G., New Zealand Archaeological Association, and Lapita Conference 2007 : Honiara, S.I. (2009). *Lapita: Ancestors and Descendants* (Auckland: New Zealand Archaeological Association).
- Shiraishi, Y., Fujimoto, A., Furuta, M., Tanaka, H., Chiba, K., Boroevich, K.A., Abe, T., Kawakami, Y., Ueno, M., Gotoh, K., et al. (2014). Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. *PLoS ONE* 9, e114263.
- Short, K.H., and Petren, K. (2011). Multimodal dispersal during the range expansion of the tropical house gecko *Hemidactylus mabouia*. *Ecol Evol* 1, 181–190.
- Shortt, H.E. (1948). The life cycle of *Plasmodium cynomolgi* in its insect and mammalian hosts. *Trans Roy Soc Trop Med Hyg* 42, 227–230.

- Shortt, H.E., and Garnham, P.C.C. (1948). The pre-erythrocytic development of *Plasmodium cynomolgi* and *Plasmodium vivax*. *Trans Roy Soc Trop Med Hyg* 41, 785–795.
- Shortt, H.E., Bray, R.S., and Cooper, W. (1954). Further notes on the tissue stages of *Plasmodium cynomolgi*. *Trans Roy Soc Trop Med Hyg* 48, 122–131.
- Siddall, M.E., and Barta, J.R. (1992). Phylogeny of *Plasmodium* species: Estimation and inference. *J Parasitol* 78, 567.
- Sidhu, A.B.S., Verdier-Pinard, D., and Fidock, D.A. (2002). Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by *pfprt* mutations. *Science* 298, 210–213.
- Sievers, R. (1891). Om frossan i Finland. *Finska Läkaresällskapets Handlingar* 33.
- Silva, J.C., Egan, A., Friedman, R., Munro, J.B., Carlton, J.M., and Hughes, A.L. (2011). Genome sequences reveal divergence times of malaria parasite lineages. *Parasitology* 138, 1737–1749.
- Simberloff, D. (2009). The role of propagule pressure in biological invasions. *Ann Rev Ecol Evol Syst* 40, 81–102.
- Simmons, M.P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol* 49, 369–381.
- Simonson, T.S., Xing, J., Barrett, R., Jerah, E., Loa, P., Zhang, Y., Watkins, W.S., Witherspoon, D.J., Huff, C.D., Woodward, S., et al. (2011). Ancestry of the Iban is predominantly Southeast Asian: Genetic evidence from autosomal, mitochondrial, and Y chromosomes. *PLoS ONE* 6, e16338.
- Simpson, G.G. (1977). Too many lines; The limits of the Oriental and Australian zoogeographic regions. *Proc Am Philos Soc* 121, 107–120.
- Sinka, M.E., Bangs, M.J., Manguin, S., Chareonviriyaphap, T., Patil, A.P., Temperley, W.H., Gething, P.W., Elyazar, I.R.F., Kabaria, C.W., Harbach, R.E., et al. (2011). The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasit Vectors* 4, 89.
- Sinka, M.E., Bangs, M.J., Manguin, S., Rubio-Palis, Y., Chareonviriyaphap, T., Coetzee, M., Mbogo, C.M., Hemingway, J., Patil, A.P., Temperley, W.H., et al. (2012). A global map of dominant malaria vectors. *Parasites & Vectors* 5, 69.
- Sirelius, U.T. (1921). Suomen kansanomaista kulttuuria (Kustannusosakeyhtiö Otava: Helsinki).
- Siska, V., Jones, E.R., Jeon, S., Bhak, Y., Kim, H.-M., Cho, Y.S., Kim, H., Lee, K., Veselovskaya, E., Balueva, T., et al. (2017). Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Science Advances* 3, e1601877.
- Sivén, V.O. (1904). Om frossan i Helsingfors i belysning af den moderna myggteorin. *Fin Lakaresallsk Handl.* 47.
- Skoglund, P., and Mathieson, I. Ancient genomics: A new view into human prehistory and evolution.
- Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*.
- Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., et al. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature* 538, 510–513.
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science* 236, 787–792.
- Slatkin, M., and Racimo, F. (2016). Ancient DNA and human history. *PNAS* 113, 6380–6387.
- Smith, T., Bhatia, K., Prasad, M., Koki, G., and Alpers, M. (1994). Altitude, language, and class I HLA allele frequencies in Papua New Guinea. *Am J Phys Anthropol* 95, 155–168.
- Snounou, G., and Pérignon, J.-L. (2013). Malariotherapy – Insanity at the Service of Malariology. In *Advances in Parasitology*, (Elsevier), pp. 223–255.
- Snow, D.R., and Lanphear, K.M. (1988). European contact and Indian depopulation in the Northeast: The timing of the first epidemics. *Ethnohistory* 35, 15–33.
- Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., and Richards, M.B. (2009). Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am J Hum Genet* 84, 740–759.
- Soares, P., Rito, T., Trejaut, J., Mormina, M., Hill, C., Tinkler-Hundal, E., Braid, M., Clarke, D.J., Loo, J.-H., Thomson, N., et al. (2011). Ancient voyaging and Polynesian origins. *Am J Hum Genet* 88, 239–247.
- Soares, P., Alshamali, F., Pereira, J.B., Fernandes, V., Silva, N.M., Afonso, C., Costa, M.D., Musilová, E., Macaulay, V., Richards, M.B., et al. (2012). The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol Biol Evol* 29, 915–927.
- Soares, P.A., Trejaut, J.A., Rito, T., Cavadas, B., Hill, C., Eng, K.K., Mormina, M., Brandão, A., Fraser, R.M., Wang, T.-Y., et al. (2016). Resolving the ancestry of Austronesian-speaking populations. *Hum Genet* 135, 309–326.
- Soe, T.N., Wu, Y., Tun, M.W., Xu, X., Hu, Y., Ruan, Y., Win, A.Y.N., Nyunt, M.H., Mon, N.C.N., Han, K.T., et al. (2017). Genetic diversity of *Plasmodium falciparum* populations in southeast and western Myanmar. *Parasit Vectors* 10.
- Sokhna, C., Ndiath, M.O., and Rogier, C. (2013). The changes in mosquito vector behaviour and the emerging resistance to insecticides will challenge the decline of malaria. *Clin Microbiol Infect* 19, 902–907.
- Sokolova, M.I., and Snow, K.R. (2002). Malaria vectors in European Russia. *Eur Mosq Bull* 12.
- Song, K.J., Lee, K.N., Lim, C.S., Kim, S.H., Kwon, S.I., and Song, J.W. (2000). Analysis of *Plasmodium vivax* merozoite surface protein-1 gene sequences from resurgent Korean isolates. *Am J Trop Med Hyg* 62, 261–265.
- Specht, J. (2005). Revisiting the Bismarcks: Some alternative views. In *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, A. Pawley, R. Attenborough, J. Golson, and R. Hide, eds. (Canberra, ACT, Australia: Pacific Linguistics), p.
- Spriggs, M. (1993). Island Melanesia: The last 10,000 years. In *A Community of Culture: The People and Prehistory of the Pacific*, M. Spriggs, D. Yen E., W. Ambrose, R. Jones, A. Thorpe, and A. Andrews, eds. (Canberra, Australia: Dept. of Prehistory, Research School of Pacific Studies, ANU), pp. 187–205.
- Spriggs, M. (1997). *The Island Melanesians* (Oxford, UK: Blackwell).
- Spriggs, M. (2011). Archaeology and the Austronesian expansion: where are we now? *Antiquity* 85, 510–528.
- Spriggs, M., Yen, D. E., Ambrose, W., Jones, R., Thorpe, A., and Andrews, A. (1993). *A Community of Culture: The People and Prehistory of the Pacific* (Canberra, Australia: Dept. of Prehistory, Research School of Pacific Studies, ANU).

- Stachowicz, J.J., Fried, H., Osman, R.W., and Whitlatch, R.B. (2002). Biodiversity, invasion resistance, and marine ecosystem function: Reconciling pattern and process. *Ecology* 83, 2575–2590.
- Stanisic, D.I., Mueller, I., Betuela, I., Siba, P., and Schofield, L. (2010). Robert Koch redux: Malaria immunology in Papua New Guinea. *Parasite Immunol.* 32, 623–632.
- Steel, M., Huson, D., and Lockhart, P.J. (2000). Invariable sites models and their use in phylogeny reconstruction. *Syst Biol* 49, 225–232.
- Steiper, M.E., Walsh, F., and Zichello, J.M. (2012). The SLC4A1 gene is under differential selective pressure in primates infected by *Plasmodium falciparum* and related parasites. *Infect Genet Evol* 12, 1037–1045.
- Stemshorn, K.C., Reed, F.A., Nolte, A.W., and Tautz, D. (2011). Rapid formation of distinct hybrid lineages after secondary contact of two fish species (*Cottus* sp.). *Mol Ecol* 20, 1475–1491.
- Stewart, J.R., and Stringer, C.B. (2012). Human evolution Out of Africa: The role of refugia and climate change. *Science* 335, 1317–1321.
- Stewart, M., Louys, J., Price, G.J., Drake, N.A., Groucutt, H.S., and Petraglia, M.D. (2017). Middle and Late Pleistocene mammal fossils of Arabia and surrounding regions: Implications for biogeography and hominin dispersals. *Quatern Int.*
- Stigzelius, L.H. (1864). Anteckningar om Malaria epidemierna på Åland under sednaste århundrade (Faculty of Medicine: PhD thesis. University of Helsinki).
- Stoneking, M., and Delfin, F. (2010). The human genetic history of East Asia: Weaving a complex tapestry. *Curr Biol* 20, R188–R193.
- Stoneking, M., and Krause, J. (2011). Learning about human population history from ancient and modern genomes. *Nat Rev Genet* 12, 603–614.
- Stoneking, M., Jorde, L.B., Bhatia, K., and Wilson, A.C. (1989). Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124, 717–733.
- Stoneking, M., Sherry, S.T., Redd, A.J., and Vigilant, L. (1992). New approaches to dating suggest a recent age for the human mtDNA ancestor. *Phil Trans R Soc B* 337, 167–175.
- Stoner, G.L., Jobes, D.V., Fernandez Cobo, M., Agostini, H.T., Chima, S.C., and Ryschkewitsch, C.F. (2000). JC virus as a marker of human migration to the Americas. *Microb Infect* 2, 1905–1911.
- Storey, A.A., Ramirez, J.M., Quiroz, D., Burley, D.V., Addison, D.J., Walter, R., Anderson, A.J., Hunt, T.L., Athens, J.S., Huynen, L., et al. (2007). Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *PNAS* 104, 10335–10339.
- Storey, A.A., Athens, J.S., Bryant, D., Carson, M., Emery, K., deFrance, S., Higham, C., Huynen, L., Intoh, M., Jones, S., et al. (2012). Investigating the global dispersal of chickens in prehistory using ancient mitochondrial DNA Signatures. *PLoS ONE* 7, e39171.
- Storey, A.A., Clarke, A.C., Ladefoged, T., Robins, J., and Matisoo-Smith, E. (2013). DNA and Pacific commensal models: Applications, construction, limitations, and future prospects. *The Journal of Island and Coastal Archaeology* 8, 37–65.
- Strode, C., Donegan, S., Garner, P., Enayati, A.A., and Hemingway, J. (2014). The impact of pyrethroid resistance on the efficacy of insecticide-treated bed nets against African Anopheline mosquitoes: Systematic review and meta-analysis. *PLoS Med* 11, e1001619.
- Stucki, D., Malla, B., Hostettler, S., Huna, T., Feldmann, J., Yeboah-Manu, D., Borrell, S., Fenner, L., Comas, I., Coscollà, M., et al. (2012). Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. *PLoS ONE* 7, e41253.
- Su, X., Ferdig, M.T., Huang, Y., Huynh, C.Q., Liu, A., You, J., Wootton, J.C., and Wellems, T.E. (1999). A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* 286, 1351–1353.
- Su, X.-Z., Mu, J., and Joy, D.A. (2003). The “Malaria’s Eve” hypothesis and the debate concerning the origin of the human malaria parasite *Plasmodium falciparum*. *Microb Infect* 5, 891–896.
- Summerhayes, G.R., Leavesley, M., Fairbairn, A., Mandui, H., Field, J., Ford, A., and Fullagar, R. (2010). Human adaptation and plant use in highland New Guinea 49,000 to 44,000 years ago. *Science* 330, 78–81.
- Sundell, N.M., and Durrett, R.T. (2001). Exponential distance statistics to detect the effects of population subdivision. *Theor Popul Biol* 60, 107–116.
- Sutikna, T., Tocheri, M.W., Morwood, M.J., Saptomo, E.W., Jatmiko, Awe, R.D., Wasisto, S., Westaway, K.E., Aubert, M., Li, B., et al. (2016). Revised stratigraphy and chronology for Homo floresiensis at Liang Bua in Indonesia. *Nature* 532, 366–369.
- Sutton, P.L. (2013). A call to arms: On refining *Plasmodium vivax* microsatellite marker panels for comparing global diversity. *Malar J* 12, 447.
- Swellengrebel, N.H., and de Buck, A. (1938). Malaria in the Netherlands (Holkema Ltd: Amsterdam).
- Tachibana, S.-I., Sullivan, S.A., Kawai, S., Nakamura, S., Kim, H.R., Goto, N., Arisue, N., Palacpac, N.M.Q., Honma, H., Yagi, M., et al. (2012). *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat Genet* 44, 1051–1055.
- Takken, W., Geene, R., Adam, W., Jetten, T.H., and van der Velden, J.A. (2001). Distribution and dynamics of larval populations of *Anopheles messeae* and *A. atroparvus* in the delta of the rivers Rhine and Meuse, the Netherlands. *Ambio* 31.
- Talbi, C., Lemey, P., Suchard, M.A., Abdelatif, E., Elharrak, M., Jalal, N., Faouzi, A., Echevarria, J.E., Vazquez Moron, S., Rambaut, A., et al. (2010). Phylogenetics and Human-Mediated Dispersal of a Zoonotic Virus RID B-2481-2009. *PLoS Pathog* 6.
- Tanabe, K., Escalante, A., Sakihama, N., Honda, M., Arisue, N., Horii, T., Culleton, R., Hayakawa, T., Hashimoto, T., Longacre, S., et al. (2007). Recent independent evolution of msp1 polymorphism in *Plasmodium vivax* and related simian malaria parasites. *Mol Biochem Parasitol* 156, 74–79.
- Tanabe, K., Mita, T., Jombart, T., Eriksson, A., Horibe, S., Palacpac, N., Ranford-Cartwright, L., Sawai, H., Sakihama, N., Ohmae, H., et al. (2010). *Plasmodium falciparum* accompanied the human expansion out of Africa. *Curr Biol* 20, 1283–1289.

- Tatem, A.J. (2006). Global traffic and disease vector dispersal. *PNAS* 103, 6242–6247.
- Tatem, A.J., and Hay, S.I. (2007). Climatic similarity and biological exchange in the worldwide airline transportation network. *Proc Royal Soc B* 274, 1489–1496.
- Tatem, A.J., Rogers, D.J., and Hay, S.I. (2006). Global transport networks and infectious disease spread. *Adv Parasitol* 62, 293–343.
- Tavaré, S. (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect. Math Life Sci.* 17, 57–86.
- Taylor, J.E., Pacheco, M.A., Bacon, D.J., Beg, M.A., Machado, R.L.D., Fairhurst, R.M., Herrera, S., Kim, J.-Y., Menard, D., Póvoa, M.M., et al. (2013). The evolutionary history of *Plasmodium vivax* as inferred from mitochondrial genomes: Parasite genetic diversity in the Americas. *Mol Biol Evol* 30, 2050–2064.
- Taylor, W.H., Taylor, M.L., Balch, W.E., and Gilchrist, P.S. (1976). Purification of properties of dihydroorotase, a zinc-containing metalloenzyme in *Clostridium oroticum*. *J Bacteriol* 127, 863–873.
- Templeton, A. (2002). Out of Africa again and again. *Nature* 416, 45–51.
- Terrell, J.E., Kelly, K.M., and Rainbird, P. (2001). Foregone conclusions? In search of “Papuan” and “Austronesians.” *Curr Anthropol* 42, 97–124.
- The Genographic Consortium, Nagle, N., Ballantyne, K.N., van Oven, M., Tyler-Smith, C., Xue, Y., Wilcox, S., Wilcox, L., Turkalov, R., van Oorschot, R.A.H., et al. (2017). Mitochondrial DNA diversity of present-day Aboriginal Australians and implications for human evolution in Oceania. *J Hum Genet* 62, 343–353.
- Theджа, M.D., Muljono, D.H., Nurainy, N., Sukowati, C.H.C., Verhoef, J., and Marzuki, S. (2011). Ethnogeographical structure of hepatitis B virus genotype distribution in Indonesia and discovery of a new subgenotype, B9. *Arch Virol* 156, 855–868.
- Thomsen, E.K., Koimbu, G., Pulford, J., Jamea-Maiasa, S., Ura, Y., Keven, J.B., Siba, P.M., Mueller, I., Hetzel, M.W., and Reimer, L.J. (2016). Mosquito behaviour change after distribution of bednets results in decreased protection against malaria exposure. *J Infect Dis* jiw615.
- Thorsby, E. (2012). The Polynesian gene pool: an early contribution by Amerindians to Easter Island. *Phil Trans R Soc B* 367, 812–819.
- Tibayrenc, M. (2011). *Genetics and Evolution of Infectious Diseases* (Elsevier).
- Timmermann, A., and Friedrich, T. (2016). Late Pleistocene climate drivers of early human migration. *Nature* 538, 92–95.
- Tirakarn, S., Riangrunroj, P., Kongsaree, P., Imwong, M., Yuthavong, Y., and Leartsakulpanich, U. (2012). Cloning and heterologous expression of *Plasmodium ovale* dihydrofolate reductase-thymidylate synthase gene. *Parasitol Int* 61, 324–332.
- Tishkoff, S.A., Dietzsch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K., Bonn -Tamir, B., Santachiara-Benerecetti, A.S., Moral, P., and Krings, M. (1996). Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271, 1380–1387.
- Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drouiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., et al. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. *Science* 293, 455–462.
- Tomokazu Takasaka, T.K. (2006). Phylogenetic analysis of major African genotype (Af2) of JC Virus: implications for origin and dispersals of modern Africans. *Am J Phys Anthropol* 129, 465–472.
- Tonione, M.A., Reeder, N., and Moritz, C.C. (2011). High genetic diversity despite the potential for stepping-stone colonizations in an invasive species of gecko on Moorea, French Polynesia. *PLoS One* 6, e26874.
- Tryon, D. (1995). Proto-Austronesian and the major Austronesian subgroups. In *The Austronesians: Historical and Comparative Perspectives*, P. Bellwood, J.J. Fox, and D. Tryon, eds. (Canberra, Australia: ANU E Press), p. 19.
- Tryon, D.T. (2011). *Comparative Austronesian Dictionary, An Introduction to Austronesian Studies* (Berlin, Boston: De Gruyter Mouton).
- Tsukahara, T., Hombhanje, F.W., Lum, J.K., Hwaihwanje, I., Masta, A., Kaneko, A., and Kobayakawa, T. (2006). Austronesian origin of the 27-bp deletion of the erythrocyte band 3 gene in East Sepik, Papua New Guinea inferred from mtDNA analysis. *J Hum Genet* 51, 244–248.
- Ubalee, R., Kim, H.-C., Schuster, A.L., McCardle, P.W., Phasomkusolsil, S., Takhampunya, R., Davidson, S.A., Lee, W.-J., and Klein, T.A. (2016). Vector competence of *Anopheles klein* and *Anopheles sinensis* (Diptera: Culicidae) from the Republic of Korea to vivax malaria-infected blood from patients from Thailand. *J Med Entomol* 53, 1425–1432.
- Ulbricht, W., and Wagner, H.H. (1975). The influence of pH on equilibrium effects of tetrodotoxin on myelinated nerve fibres of *Rana esculenta*. *J. Physiol. (Lond.)* 252, 159–184.
- Urdaneta-Marquez, L., and Failloux, A.-B. (2011). Population genetic structure of *Aedes aegypti*, the principal vector of dengue viruses. *Infect Genet Evol* 11, 253–261.
- Utrio, P. (1976). Identification key to Finnish mosquito larvae (Diptera, Culicidae). *Ann Agric Fenn* 15.
- Utrio, P. (1978). Distribution of mosquitoes (Diptera, Culicidae) attracted by man in Finland in early July. *Notulae Entomol* 58.
- Utrio, P. (1979). Geographic distribution of mosquitoes (Diptera, Culicidae) in eastern Fennoscandia. *Notulae Entomol* 59.
- Vattula, K. (1983). *The Economic History of Finland 3. Historic statistics* (Kustannusosakeyhti  Tammi: Helsinki).
- Vazeille-Falcoz, M., Mousson, L., Rodhain, F., Chungue, E., and Failloux, A.B. (1999). Variation in oral susceptibility to dengue type 2 virus of populations of *Aedes aegypti* from the islands of Tahiti and Moorea, French Polynesia. *Am J Trop Med Hyg* 60, 292–299.
- Vellend, M., Harmon, L.J., Lockwood, J.L., Mayfield, M.M., Hughes, A.R., Wares, J.P., and Sax, D.F. (2007). Effects of exotic species on evolutionary diversification. *Trends Ecol Evol* 22, 481–488.
- Vences, M., Wanke, S., Vieites, D.R., Branch, W.R., Glaw, F., and Meyer, A. (2004). Natural colonization or introduction? Phylogeographical relationships and morphological differentiation of house geckos (*Hemidactylus*) from Madagascar. *Biol J Linnean Soc* 83, 115–130.

- Verhave, J.P. (2013). Experimental, therapeutic and natural transmission of *Plasmodium vivax* tertian malaria: scientific and anecdotal data on the history of Dutch malaria studies. *Parasit Vectors* 6, 19.
- Vernot, B., Tucci, S., Kelso, J., Schraiber, J.G., Wolf, A.B., Gittelman, R.M., Dannemann, M., Grote, S., McCoy, R.C., Norton, H., et al. (2016). Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* 352, 235–239.
- Véron, V., Legrand, E., Yrinesi, J., Volney, B., Simon, S., and Carme, B. (2009). Genetic diversity of *msp3a* and *msp1_b5* markers of *Plasmodium vivax* in French Guiana. *Malar J* 8, 40.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A. (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253, 1503–1507.
- Vilar, M.G., Kaneko, A., Hombhanje, F.W., Tsukahara, T., Hwaihwanje, I., and Lum, J.K. (2008). Reconstructing the origin of the Lapita Cultural Complex: mtDNA analyses of East Sepik Province, PNG. *J Hum Genet* 53, 698–708.
- Villabona-Arenas, C.J., and Zanolto, P.M. de A. (2011). Evolutionary history of Dengue virus type 4: Insights into genotype phylodynamics. *Infect Genet Evol* 11, 878–885.
- Vina, M.A.F., Hollenbach, J.A., Lyke, K.E., Sztein, M.B., Maier, M., Klitz, W., Cano, P., Mack, S., Single, R., Brautbar, C., et al. (2012). Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. *Phil Trans R Soc B* 367, 820–829.
- Vinkenoog, R., Veldhuisen, B., Sperança, M.A., del Portillo, H.A., Janse, C., and Waters, A.P. (1995). Comparison of introns in a *cdc2*-homologous gene within a number of *Plasmodium* species. *Mol Biochem Parasitol* 71, 233–241.
- Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F., and Hartl, D.L. (2001). Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* 293, 482–484.
- Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., et al. (2006). A genome-wide map of diversity in *Plasmodium falciparum*. *Nat Genet* 39, 113–119.
- Volkman, S.K., Lozovsky, E., Barry, A.E., Bedford, T., Bethke, L., Myrick, A., Day, K.P., Hartl, D.L., Wirth, D.F., and Sawyer, S.A. (2007). Genomic heterogeneity in the density of noncoding single-nucleotide and microsatellite polymorphisms in *Plasmodium falciparum*. *Gene* 387, 1–6.
- Volkman, S.K., Neafsey, D.E., Schaffner, S.F., Park, D.J., and Wirth, D.F. (2012). Harnessing genomics and genome biology to understand malaria biology. *Nat Rev Genet* 13, 315–328.
- Vonlanthen, P., Bittner, D., Hudson, A.G., Young, K.A., Müller, R., Lundsgaard-Hansen, B., Roy, D., Di Piazza, S., Largiadere, C.R., and Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* 482, 357–362.
- Wall, J.D., Lohmueller, K.E., and Plagnol, V. (2009). Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol Biol Evol* 26, 1823–1827.
- Walsh, J.C., Venter, O., Watson, J.E.M., Fuller, R.A., Blackburn, T.M., and Possingham, H.P. (2011). Exotic species richness and native species endemism increase the impact of exotic species on islands. *Global Ecology and Biogeography* no.
- Waltmann, A., Darcy, A.W., Harris, I., Koepfli, C., Lodo, J., Vahi, V., Piziki, D., Shanks, G.D., Barry, A.E., Whittaker, M., et al. (2015). High rates of asymptomatic, sub-microscopic *Plasmodium vivax* infection and disappearing *Plasmodium falciparum* malaria in an area of low transmission in Solomon Islands. *PLoS Negl Trop Dis* 9, e0003758.
- Waltmann, A., Koepfli, C., Tessier, N., Karl, S., Fola, A., Darcy, A.W., Wini, L., Harrison, G.L.A., Barnadas, C., Jennison, C., et al. (2018). Increasingly inbred and fragmented populations of *Plasmodium vivax* associated with the eastward decline in malaria transmission across the Southwest Pacific. *PLoS Negl Trop Dis* 12, e0006146.
- Wang, G.-D., Zhai, W., Yang, H.-C., Wang, L., Zhong, L., Liu, Y.-H., Fan, R.-X., Yin, T.-T., Zhu, C.-L., Poyarkov, A.D., et al. (2015). Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res*.
- Wang, S., Lewis, C.M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M.V., Molina, J.A., Gallo, C., et al. (2007). Genetic variation and population structure in Native Americans. *PLoS Genet* 3, e185.
- Waters, A.P., Higgins, D.G., and McCutchan, T.F. (1991). *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *PNAS* 88, 3140–3144.
- Waters, A.P., Higgins, D.G., and McCutchan, T.F. (1993a). Evolutionary relatedness of some primate models of *Plasmodium*. *Mol Biol Evol* 10, 914–923.
- Waters, A.P., Higgins, D.G., and McCutchan, T.F. (1993b). The phylogeny of malaria: A useful study. *Parasitol Today* 9, 246–250.
- Watkins, W.S., Ricker, C.E., Bamshad, M.J., Carroll, M.L., Nguyen, S.V., Batzer, M.A., Harpending, H.C., Rogers, A.R., and Jorde, L.B. (2001). Patterns of Ancestral Human Diversity: An Analysis of Alu-Insertion and Restriction-Site Polymorphisms. *Am J Hum Genet* 68, 738–752.
- Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.
- Wellems, T.E., and Plowe, C.V. (2001). Chloroquine-resistant malaria. *J Infect Dis* 184, 770–776.
- Wernsdorfer, W.H. (1980). The importance of malaria in the world. In *Malaria*, J.P. Kreier, ed. (New York: Academic Press), p.
- Wesseling, J.G., Snijders, P.J.F., van Someren, P., Jansen, J., Smits, M.A., and Schoenmakers, J.G.G. (1989). Stage-specific expression and genomic organization of the actin genes of the malaria parasite *Plasmodium falciparum*. *Mol Biochem Parasitol* 35, 167–176.
- Westaway, K.E., Roberts, R.G., Sutikna, T., Morwood, M.J., Drysdale, R., Zhao, J.-x., and Chivas, A.R. (2009). The evolving landscape and climate of western Flores: an environmental context for the archaeological site of Liang Bua. *J Hum Evol* 57, 450–464.
- White, G.B. (1978). Systematic reappraisal of the *Anopheles maculipennis* complex. *Mosq Syst* 7.
- White, N.J. (2004). Antimalarial drug resistance. *J Clin Invest* 113, 1084–1092.
- White, N.J. (2011). Determinants of relapse periodicity in *Plasmodium vivax* malaria. *Malar J* 10, 297.

- White, M.T., Shirreff, G., Karl, S., Ghani, A.C., and Mueller, I. (2016). Variation in relapse frequency and the transmission potential of *Plasmodium vivax* malaria. *Proc Royal Soc B* 283, 20160048.
- WHO (2012). Malaria Factsheet.
- Wiesmann, U.N., DiDonato, S., and Herschkowitz, N.N. (1975). Effect of chloroquine on cultured fibroblasts: release of lysosomal hydrolases and inhibition of their uptake. *Biochem. Biophys. Res. Commun.* 66, 1338–1343.
- Wilder, J.A., Stone, J.A., Preston, E.G., Finn, L.E., Ratcliffe, H.L., and Sudoyo, H. (2009). Molecular population genetics of SLC4A1 and Southeast Asian Ovalocytosis. *J Hum Genet* 54, 182–187.
- Wilder, J.A., Cox, M.P., Paquette, A.M., Alford, R., Satyagraha, A.W., Harahap, A., and Sudoyo, H. (2011). Genetic continuity across a deeply divergent linguistic contact zone in North Maluku, Indonesia. *BMC Genet* 12, 100.
- Wilkinson, R.L. (1994). Yellow fever: Ecology, epidemiology, and role in the collapse of the Classic lowland Maya civilization. *Med Anthropol* 16, 269–294.
- William, T., Rahman, H.A., Jelip, J., Ibrahim, M.Y., Menon, J., Grigg, M.J., Yeo, T.W., Anstey, N.M., and Barber, B.E. (2013). Increasing Incidence of *Plasmodium knowlesi* Malaria following Control of *P. falciparum* and *P. vivax* Malaria in Sabah, Malaysia. *PLoS Negl Trop Dis* 7, e2026.
- Williams, N. (2009). Malaria's primate links. *Curr Biol* 19, R668–R669.
- Williams, T.N. (2006). Human red blood cell polymorphisms and malaria. *Curr Opin Microbiol* 9, 388–394.
- Williams-Blangero, S., Criscione, C.D., VandeBerg, J.L., Correa-Oliveira, R., Williams, K.D., Subedi, J., Kent, J.W., Williams, J., Kumar, S., and Blangero, J. (2012). Host genetics and population structure effects on parasitic disease. *Phil Trans R Soc B* 367, 887–894.
- Williamson, D.H., Gardner, M.J., Preiser, P., Moore, D.J., Rangachari, K., and Wilson, R.J.M. (1994). The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. *Mol Gen Genet* 243, 249–252.
- Willis, K.J., Gillson, L., and Brncic, T.M. (2004). How “Virgin” Is Virgin Rainforest? *Science* 304, 402–403.
- Wilmshurst, J.M., Anderson, A.J., Higham, T.F.G., and Worthy, T.H. (2008). Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. *PNAS* 105, 7676–7680.
- Wilmshurst, J.M., Hunt, T.L., Lipo, C.P., and Anderson, A.J. (2011). High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *PNAS* 108, 1815–1820.
- Wilson, (Iain) R.J.M., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W., et al. (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol* 261, 155–172.
- Wilson, J.R.U., Dormontt, E.E., Prentis, P.J., Lowe, A.J., and Richardson, D.M. (2009). Something in the way you move: dispersal pathways affect invasion success. *Trends Ecol Evol* 24, 136–144.
- de Winter, J.C.F., Dodou, D., and Wieringa, P.A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behav Res* 44, 147–181.
- Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human infectious diseases. *Nature* 447, 279–283.
- Wolters, J. (1991). The troublesome parasites — molecular and morphological evidence that Apicomplexa belong to the dinoflagellate-ciliate clade. *Biosystems* 25, 75–83.
- Worathumrong, N., and Grimes, A.J. (1975). The effect of o-salicylate upon pentose phosphate pathway activity in normal and G6PD-deficient red cells. *Br. J. Haematol.* 30, 225–231.
- World Health Organisation (2008). World Malaria Report 2008 (Luxembourg: WHO).
- World Health Organisation (2018). World Malaria Report 2018 (Luxembourg: WHO).
- Wright, S. (1965). The interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution* 19, 395–420.
- Wurtz, N., Mint Lekweiry, K., Bogreau, H., Pradines, B., Rogier, C., Ould Mohamed Salem Boukhary, A., Hafid, J.E., Ould Ahmedou Salem, M.S., Trape, J.-F., Basco, L.K., et al. (2011). Vivax malaria in Mauritania includes infection of a Duffy-negative individual. *Malar J* 10, 336.
- Xu, S., and Stoneking, M. (2012). Reply to Denham and Donohue: Asian-Papuan genetic admixture is in excellent agreement with Austronesian dispersal in eastern Indonesia. *PNAS* 109, E2578–E2578.
- Xu, S., Pugach, I., Stoneking, M., Kayser, M., and Jin, L. (2012). Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *PNAS* 109, 4574–4579.
- Yada, A., Carnevale, P., Barro, D., Roisin, A., Cot, M., Abel, L., and Feingold, J. (1993). Risk factors of malaria infection during pregnancy in Burkina Faso: Suggestion of a genetic influence. *The American Journal of Tropical Medicine and Hygiene* 48, 358–364.
- Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T.J., Aubouy, A., Balloux, F., Besnard, P., Bogreau, H., et al. (2012). Multiple independent introductions of *Plasmodium falciparum* in South America. *PNAS* 109, 511–516.
- Yang, M.A., and Fu, Q. (2018). Insights into modern human prehistory using ancient genomes. *Trends Genet* 34, 184–196.
- Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 17, 32–43.
- Yang, Z., Miao, J., Huang, Y., Li, X., Putapornpit, C., Jongwutiwes, S., Gao, Q., Udomsangpetch, R., Sattabongkot, J., and Cui, L. (2006). Genetic structures of geographically distinct *Plasmodium vivax* populations assessed by PCR/RFLP analysis of the merozoite surface protein 3 β gene. *Acta Trop* 100, 205–212.
- Yavo, W., Konaté, A., Mawili-Mboumba, D.P., Kassi, F.K., Tshibola Mbuyi, M.L., Angora, E.K., Menan, E.I.H., and Bouyou-Akotet, M.K. (2016). Genetic polymorphism of msp1 and msp2 in *Plasmodium falciparum* isolates from Côte d'Ivoire versus Gabon. *J Parasitol Res* 2016, 1–7.
- Yen, D.E. (1963). The New Zealand kumara or sweet potato. *Econ. Bot.* 17, 31–45.
- Zanini, F., Puller, V., Brodin, J., Albert, J., and Neher, R.A. (2017). In vivo mutation rates and the landscape of fitness costs of HIV-1. *Virus Evol* 3.

- Zarowiecki, M., Loaiza, J.R., and Conn, J.E. (2011a). Towards a new role for vector systematics in parasite control. *Parasitology* 138, 1723–1729.
- Zarowiecki, M., Walton, C., Torres, E., McAlister, E., Htun, P.T., Sumrandee, C., Sochanta, T., Dinh, T.H., Ng, L.C., and Linton, Y.-M. (2011b). Pleistocene genetic connectivity in a widespread, open-habitat-adapted mosquito in the Indo-Oriental region. *J Biogeogr* 38, 1422–1432.
- Zhang, Y., Chen, F., Fonseca, N.A., He, Y., Fujita, M., Nakagawa, H., Zhang, Z., Brazma, A., and Creighton, C. (2017). Whole genome and RNA sequencing of 1,220 cancers reveals hundreds of genes deregulated by rearrangement of cis-regulatory elements. *Biorxiv*.
- Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J Comp Biol* 7, 203–214.
- Zhong, D., Bonizzoni, M., Zhou, G., Wang, G., Chen, B., Vardo-Zalik, A., Cui, L., Yan, G., and Zheng, B. (2011). Genetic diversity of *Plasmodium vivax* malaria in China and Myanmar. *Infect Genet Evol* 11, 1419–1425.
- Zilversmit, M.M., Volkman, S.K., DePristo, M.A., Wirth, D.F., Awadalla, P., and Hartl, D.L. (2010). Low-complexity regions in *Plasmodium falciparum*: Missing links in the evolution of an extreme genome. *Mol Biol Evol* 27, 2198–2209.
- Zimmerman, P.A., Woolley, I., Masinde, G.L., Miller, S.M., McNamara, D.T., Hazlett, F., Mgone, C.S., Alpers, M.P., Genton, B., Boatin, B.A., et al. (1999). Emergence of FY*A null in a *Plasmodium vivax*-endemic region of Papua New Guinea. *PNAS* 96, 13973–13977.