Genetic characteristics of *Plasmodium vivax* from Northern Mali

By

Moussa DJIMDE
DJMMOU001

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree

Master of Science in Medicine: Bioinformatics

Submitted on June 30, 2018
Supervisor: Prof Nicola MULDER
Co-Supervisors: Prof Abdoulaye DJIMDE
Dr Antoine DARA

Integrative Biomedical Sciences
Computational Biology (CBIO) Division
****
Faculty of Health Sciences
University of Cape Town

Acknowledgements

I wish to thank all patients who participated in the study from which data allowed us to make this analysis. I would like to extend great thanks to my supervisor Professor Nicola Mulder and my co-supervisors Prof Abdoulaye Djimdé and Dr Antoine Dara for the constant support and encouragement. Thanks are also extended to all the Computational Biology (CBIO) Group for accepting me in the team, for the support and for providing the academic environment that was required to complete this study.

This work was supported through the DELTAS Africa Initiative [grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107740/Z/15/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.
I am indebted to all the Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) team for the constant assistance during my training.

I can never thank enough my Mentor Professor Ogobara K. Doumbo for his constant support, encouragements and teachings. Professor, you left us prematurely petrifying us all. With our heart shattered, we implore our creator to welcome you in Paradise. Prof Doumbo was not only a Scientist, he was also a great humanist. May God restore your efforts. Through you, my deep thank go to the whole team of the **M**alaria **R**esearch and **T**raining **C**enter which is now becoming MRTC-Ogobara Doumbo.
I sincerely thank my colleague Aboubacrin Haidara for participating in data collection in the Kidal area during wartime.

I greatly thank my closes friends and colleagues for their encouragement and priceless supports during this training.

I sincerely thank Thomas Otto from Wellcome Trust Sanger Institute for his availability whenever I needed his advice or support.

To my family, I thank everyone tremendously for giving me the opportunity, constant support and love throughout my education. It’s invaluable.

Declaration

I, Moussa Djimdé, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: June 29, 2018

Contents

List of tables and figures

1- List of tables

2- List of figures

Abbreviation

ACT	Artemisinin – based combination therapy
AIDS	Acquired immunodeficiency syndrome
AL	Artemether – Lumefantrine
AS + AQ	Artesunate – Amodiaquine
BAM file	compressed binary version of a SAM file
bp	Base pair
BWA	Burrows-Wheeler Aligner
CBIO	Computational Biology Division
CDC	Centers for Disease Control and Prevention
cDNA	complementary deoxyribonucleic acid
CQ	Chloroquine
DEL	Deletions
DHA	Dihydroartemisinin
DNA	deoxyribonucleic acid
ENA	European Nucleotide Archive
F_{ST}	Wright's Fixation Index
IGV	Integrative Genome Viewer
INS	Insertions
IPTp	Intermittent preventive treatment in pregnancy
IRS	Indoor Residual Spraying
IST	Intermittent screening and treatment
km	Kilometre
MD	Doctor of Medicine
MRTC	Malaria Research and Training Center
MVIP	Malaria Vaccine Implementation Programme
NGS	Next Generation Sequencing
ng/µl	Nano gram per microliter
nM	Nanomolar
P. falciparum	*Plasmodium falciparum*
P. malariae	*Plasmodium malariae*
P. ovale	*Plasmodium ovale*
P. vivax	*Plasmodium vivax*

PCA	Principal Component Analysis
PCR	polymerase chain reaction
PharmD	Doctor of Pharmacy
PhD	Doctor of Philosophy
PlasmoDB	biological database for the genus Plasmodium
PQ	Primaquine
QT-NASBA	quantitative nucleic acid sequence based amplification
RDT	Rapid Diagnostic Test
RNA	Ribonucleic acid
SAM file	Sequence Alignment/Map file
SBS	Sequencing by synthesis
SMC	Seasonal Malaria Chemoprevention
SP	Sulfadoxine-Pyrimethamine
SNP	Single Nucleotide Polymorphism
sWGA	Selective Whole-Genome Amplification
UCT	University of Cape Town
USA	United State of America
VCF	Variant Call Format
VEP	Variant Effect Predictor
VSAs	Variant Surface Antigens
WBC	White Blood Cell
WHO	World Health Organization

People involved

Full name	Title/Position	Role	Address
Dr Moussa Djimde	MD, MSc Student	Candidate	University of Cape Town Faculty of Health Sciences Integrative Biomedical Sciences Computational Biology Division Mobile: +27 (0) 780466218 Emails: djmmou001@myuct.ac.za mdjimde@icermali.org
Prof Nicola Mulder	PhD, Head of CBIO, UCT	Supervisor	University of Cape Town Faculty of Health Sciences Integrative Biomedical Sciences Computational Biology Division Email: nicola.mulder@uct.ac.za
Prof Abdoulaye Djimde	PharmD, PhD	Co-Supervisor	Malaria Research & Training Center, University of Sciences, Techniques & Technology of Bamako PoBOX: 1805, Bamako, Mali Phone: (+223) 20228109 Email: adjimde@icermali.org
Dr Antoine Dara	PharmD, PhD	Co-Supervisor	University of Maryland School of Medicine, Baltimore Malaria Research & Training Center, University of Sciences, Techniques & Technology of Bamako PoBOX: 1805, Bamako, Mali Phone: (+223) 20228109 Email: adara@som.umaryland.edu tonydara@icermali.org
Prof Ogobara K. Doumbo	MD, PhD Director of MRTC	Mentor	Malaria Research & Training Center, University of Sciences, Techniques & Technology of Bamako PoBOX: 1805, Bamako, Mali Phone: (+223) 20228109 Email: okd@icermali.org

Abstract

Introduction:

The surprising presence of *P. vivax* in West Africa and their ability to infect a Duffy negative population is one more threat to public health. In order to contribute to malaria elimination efforts, there is a need to investigate the origin and characteristics of *P. vivax* population isolates in Northern Mali. Next Generation Sequence Analysis (NGSA) can help us understand parasite genetic characteristics although low parasite density is a challenge for whole genome sequencing (WGS). In the present work, we investigated if selective whole genome amplification (sWGA) can enrich *P. vivax* DNA extracted from Rapid Diagnostic Tests (RDTs) for Whole Genome Sequencing. We also investigated the origin and the susceptibility to antimalarial drugs of the strains isolated in Northern Mali.

Methods:

Parasite DNA was extracted from 267 RDTs using the QIAamp DNA mini kit, then nested PCR and 7 samples were positive for *P. vivax*. After sWGA, the whole genomes were sequenced using the Illumina platform. Next Generation Sequences Analysis was done followed by population differentiation analyses. Twenty-two additional *P. vivax* whole genomes from other parts of the World were downloaded from the European Nucleotide Archive for further Neighbour Joining analysis.

Results:

The sequences extracted from RDTs showed high contamination with human DNA (80%). From the parasite DNA, in total 69529 SNPs were found in the seven *P. vivax* strains of Northern Mali. The most significant p-values per SNP were carried by the chromosomes 2, 3, 4, 5, 12, 13 and 14. With regard to variant effects, the Transition/Transversion ratio was 1.1. The density of variants with a high effect was 1.62%. There was no mutation associated with antimalarial drugs resistance on *pvcrt-o* or *pvmdr-1* genes. Pairwise differentiation suggests a high degree of relatedness between *P. vivax* strains isolated in Northern Mali. The Neighbours-Joining analysis shows clearly that strains from Mali cluster together and are genetically distinct from those from Mauritania, which shares a border with Mali. The strains isolated in Northern Mali are genetically closer to those from Madagascar, India and Latina America.

Conclusion:

We did not identify mutations associated to the resistance to antimalarial drugs in *pvcrt-o* and *pvmdr-1* genes. This study confirms that *P. vivax* strains genetically distinct from those of

Mauritania are circulating in Mali. Finally, we conclude that SWGA is a feasible approach for *P. vivax* DNA enrichment for WGS despite the high proportion of human contamination.

Keywords: *Plasmodium vivax*, Northern Mali, Next Generation Sequencing Analysis.

CHAPTER 1

1. Introduction: General informations on malaria and Next Generation Sequencing (NGS)

1.1. General introduction

The malaria parasite was identified for the first time in 1880 by Charles Louis Alphonse Laveran. Dr Laveran was a French army surgeon stationed in Constantine, Algeria. For his discovery, Alphonse received the Nobel Prize in 1907 (Center for Disease Control and Prevention (CDC) 2017).

Malaria is a major public health problem. An infectious bite of a female of *Anopheles* mosquito is responsible for transmitting the disease, though the transmission can rarely also occur through transfusion of infected blood (Merriam-Webster 2015). Though also rare, it is possible that a woman can infect her foetus during pregnancy or delivery (Malhotra et al. 2006).

Among anopheles species, *Anopheles gambiae*, *Anopheles arabiensis* and *Anopheles funestus* are the primary vectors transmitting malaria in sub-Saharan Africa (Mzilahowa et al. 2012; Sinka et al. 2012).

Six species of *Plasmodium* are known to be responsible for malaria: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale (P.o. curtisi* and *P.o. wallikeri), Plasmodium malariae and Plasmodium knowlesi* (Medicines for Malaria Venture (MMV) 2016). Among these six species *P. falciparum* is responsible for the majority of malaria deaths (Medicines for Malaria Venture (MMV) 2016). *P. vivax* is the most widely distributed, causing approximately half of all clinical cases of malaria outside of Africa (World Health Organisation(WHO) 2016c). *P. vivax* is the predominant malaria parasite american continent (64%) following by the Eastern Mediterranean regions (40%) and South-East Asia (30%). More than the half of *P. vivax* malaria cases occur in the WHO South-East Asia Region(World Health Organisation(WHO) 2017d). In 1948, the presence of *P. vivax* was reported in Mauritania (Mint lekweiry et al. 2012), a country sharing a border with Mali. Recently, publications showed the high prevalence of *P. vivax* malaria in Nouakchott, Mauritania (Ba et al. 2016; M. S. Ould Ahmedou Salem et al. 2015; Ouldabdallahi Moukah et al. 2016). *P. vivax* malaria has also been reported in East African countries like Ethiopia, Eritrea and Sudan (Talha et al. 2015).

Malaria is a poverty linked disease, malaria-related mortality rates are highest in low income countries where the highest burden is in rural areas with financial constraints (World Health Organisation(WHO) 2013). In endemic areas, malaria is responsible for about a 1.3% reduction in economic growth (Sachs and Malaney 2002).

Despite the fact that malaria is an evitable and curable disease, in 2016 there were approximately 40.6 million malaria cases and about 18,700 malaria deaths, most of them occurring in Sub-Saharan Africa (World Health Organisation(WHO) 2017e). Some population groups like children under 5 years of age, pregnant women, non-immune migrants, travellers and patients with HIV/AIDS are more vulnerable to malaria, and developing severe disease, than others (World Health Organisation(WHO) 2016a).

1.2. Malaria parasites life cycle

The malaria parasite life cycle involves two types of hosts (Figure 1): humans and female Anopheles mosquitoes. It can proliferate in different cell types while circumvening its clearance by the host through immune system (Greenwood et al. 2008).

During a blood meal, the female Anopheles mosquito inoculates sporozoites into the human blood stream. As shown in Figure 1, after inoculation in human, sporozoites are rapidly taken up by the liver cells. They undergo several stages:

- Liver stage: The duration of the hepatic stage differs according to the parasite species. It is shorter for *P. falciparum*, ~6 days. With *P. vivax* the duration of the hepatic stage is ~8 days with the possibility of slow hepatic schizogonic multiplication (hypnozoites). This stage is ~9 days with *P. ovale,* with the possibility of hypnozoites, while for *P. malaria*, the hepatic stage duration is ~15 days (Yaro Jean-Baptiste Bibié; oral communication). Only *P. vivax* and *P. ovale* can remain dormant in the hepatocytes for months or several years (Medecine for Malaria Venture 2017). The merozoites are released into the bloodstream after the rupture of liver cells, and then they rapidly invade the red blood cells.
- Blood stage: At this stage trophozoites mature into schizonts. The rupture of schizonts releases merozoites. Some parasites differentiate into gametocytes (sexual erythrocytic stages). It is at the blood stage that the clinical manifestations of the disease appear.

During the blood meal, an Anopheles mosquito ingests the gametocytes (male and female). When gametocytes fertilize and multiply in the mosquito it is known as the sporogonic cycle. The zygotes then become ookinetes, the ookinetes and the zygotes are the only diploid form of the parasite. Ookinetes invade the midgut wall of the mosquito where they develop into oocysts. Mature sporozoites exit the oocysts and make their way to the mosquito's salivary glands, where the future inoculation to humans perpetuates the malaria life cycle. *P. malariae* is known to have the possibility to persist asymptomatically in the blood for decades (Collins and Jeffery 2007).

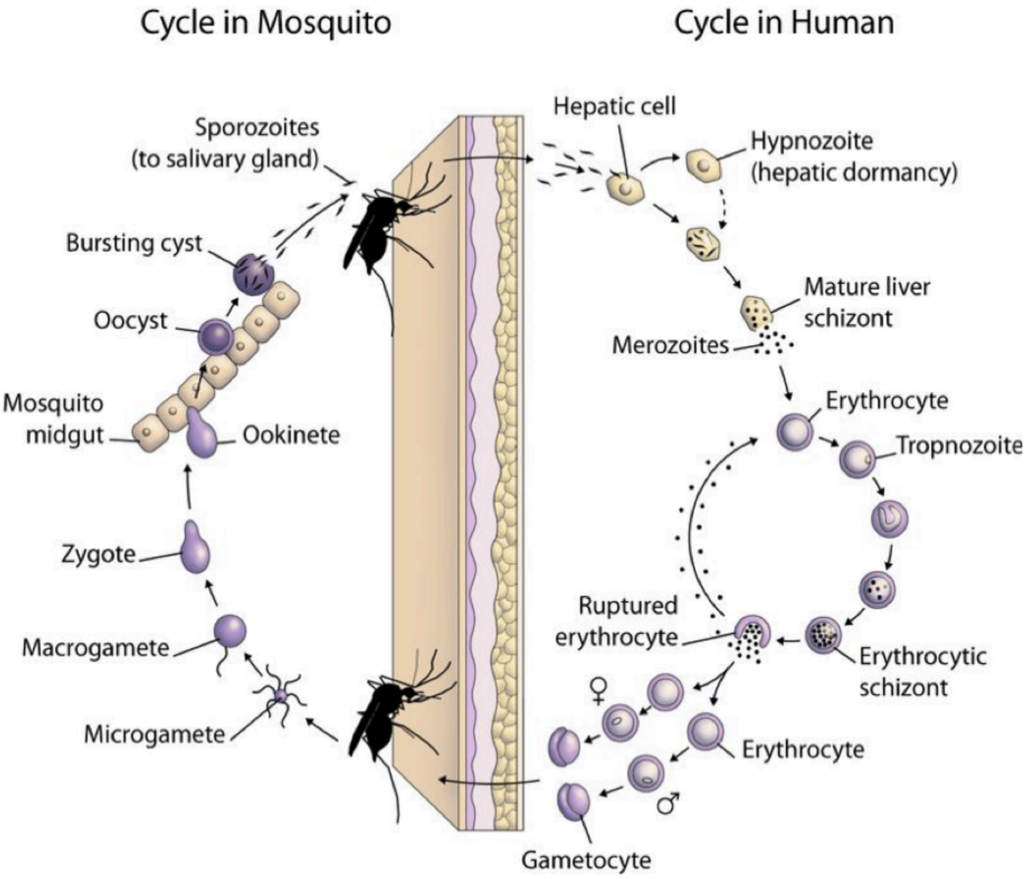


Figure 1: Life cycle of malaria parasites (Source: https://www.malariasite.com/life-cycle/; accessed 4th December 2017)

1.3. Particularity of *Plasmodium vivax* life cycle

As with the other *Plasmodium* species, after inoculation into the human body by female anopheles mosquitoes, sporozoites reach the hepatocytes and start the exoerythrocytic stage. Once in hepatocytes, *P. vivax* sporozoites can either enter in schizogonia which release merozoites into the bloodstream after thousands of mitotic replications in individual hepatocytes, or instead, sporozoites can differentiate into a dormant stage called a hypnozoite, with can be activated after several months or years (Krotoski 1985).

During the erythrocytic stages merozoites predominantly invade immature red blood cells, reticulocytes (Sturm 2006). Then, reticulocytes become enlarged and more deformed (Suwanarusk et al. 2004). In addition to producing specific proteins known as Schüfner's dots (Barnwell et al. 1990), some *P. vivax* parasites can differentiate into mature gametocytes. The Schüfner's dots appear as profuse speckling in Giemsa-stained blood smears.

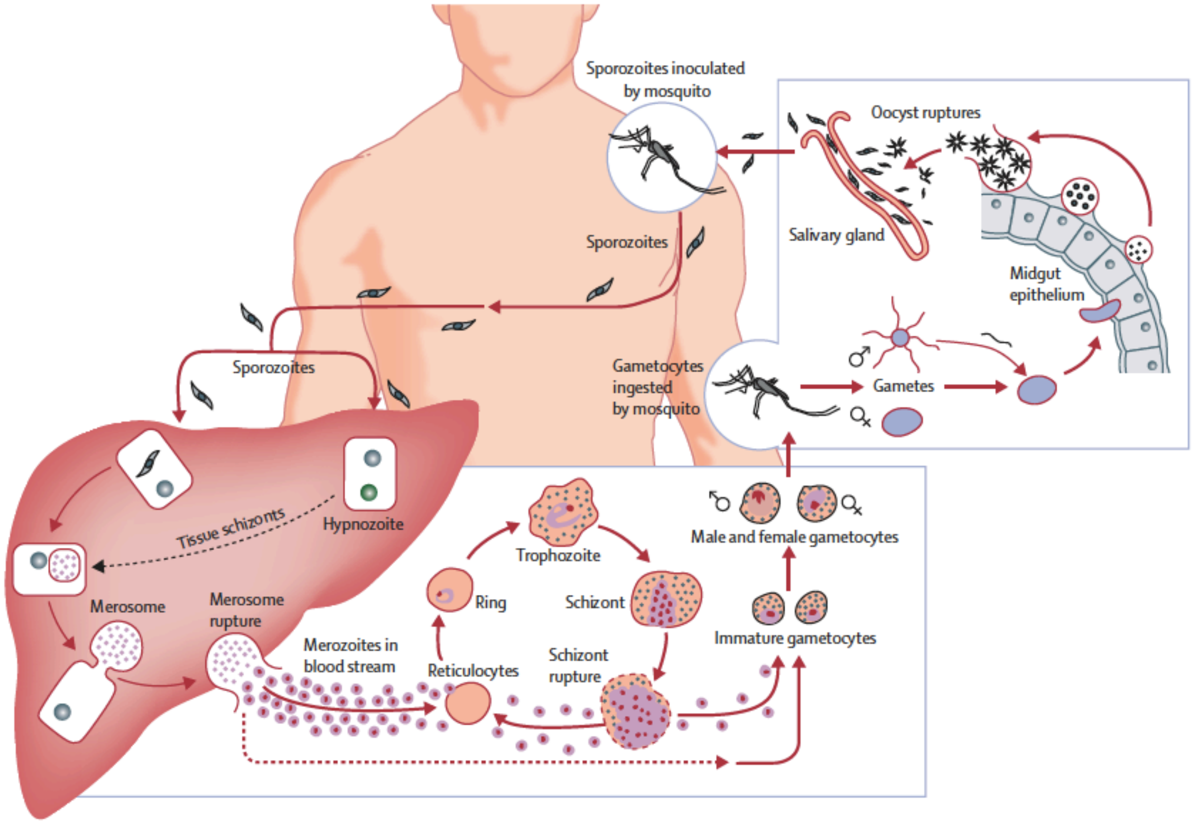


Figure 2: *Plasmodium vivax* life cycle (source: Mueller et al., 2009)

1.4. Clinical manifestations of malaria

At the blood stage, parasites cause massive bursting of erythrocytes causing the accumulation of toxic wastes that result in malaria clinical manifestations. Toxic wastes occuring from erythrocyte destruction trigger the production of cytokines and other soluble factors by macrophages and other cells lead to fevers and others clinical presentations. Many factors such as patient under five year old, Long term travelling from endemic area leading to the lost of naturally acquired immunity, the disease transmission intensity, host and parasite genetics play a role in the clinical presentation (Doolan, Dobaño, and Baird 2009; Kwiatkowski 2005; Reyburn et al. 2005). Malaria manifests in different forms:

- **Asymptomatic Malaria**: In high transmission areas, continuous exposure to malaria parasites can result in partial immunity that leads to asymptomatic carriers that can act as potential reservoirs of parasites, or gametocyte carriages resulting in persistent transmission (Bousema et al. 2004). In the Sikasso Region, Southern Mali, Save the Children reported 80% of asymptomatic malaria amongst school age children (Save the Children 2016).
- **Uncomplicated Malaria**: Destruction of erythrocytes leads to fever and flu-like symptoms, such as: chills, headache, muscle aches, tiredness, loss of appetite, abdominal pain, nausea, vomiting, diarrhoea, etc. These initial symptoms are non-specific to malaria. Malaria is considered uncomplicated when symptoms are present but there are no clinical or laboratory signs to indicate severity or vital organ dysfunction (Medicines for Malaria Venture 2017).
- **Severe Malaria**: Almost all forms of severe malaria are caused by *P. falciparum* (Snow et al. 2005; Svenson et al. 1995) and *P. vivax*(Genton et al. 2008; Tjitra et al. 2008). Severe malaria can be manifested by cerebral malaria associated with abnormal behavior, impaired consciousness, seizures, coma, or other neurological abnormalities; severe anemia due to massive destruction of erythrocytes; low blood pressure due to cardiovascular collapse; hemoglobinuria due to hemolysis; acute respiratory distress syndrome (because of deep breathing resulting from metabolic acidosis); acute renal failure, etc.

1.5. *Plasmodium vivax* malaria

Even though it is less virulent than *P. falciparum*, *P. vivax* malaria can lead to severe disease and death (Anstey et al. 2012; Baird 2007) even if the mechanisms of severe disease are not clearly understood. The pyrogenic threshold in *P. vivax* malaria is greater than that for *P. falciparum* (Price et al. 2007). In vivax malaria, generally, parasite density is low and the disease severity is not characterized by high parasite density (Anstey et al. 2009).

Unlike *P. falciparum*, this particular species can populate human blood with gametocytes in the absence of symptoms, which means that treatment of symptomatic cases does not necessarily help stop an outbreak (Vogel 2013). Causing no symptoms the parasite can stay dormant in the liver during several days or years (hypnozoites) allowing the parasite to survive in more temperate zones, where mosquitoes do not bite all year round (Gething et al. 2012). It has been proposed that a single infection can trigger six or more relapses a year. The relapses can be triggered by other infectious diseases including, *P. falciparum* malaria (White 2011). Robinson LJ et al. showed that four of every five *P. vivax* infections are due to reactivation of hypnozoites during a study caried out in Papua New Guinea (Robinson et al. 2015).

The major manifestations that can be caused by severe *P. vivax* are severe anaemia (Anstey et al. 2009, 2012; Kochar et al. 2010), acute renal injury (V B Kute et al. 2012; Vivek B. Kute et al. 2012), acute lung injury as well as respiratory distress (Anstey et al. 2007; Lança et al. 2012) and even cerebral malaria (Beg et al. 2002). Previous study conducted in the Brazilian Amazon reported two deaths among 24 *P. vivax* severe cases (Lança et al. 2012). From 17 severe *P. vivax* malaria cases, Lacerda *et al* reported 13 deaths (Lacerda et al. 2012). In 2015, a study revealed that the overall fatality rate due to vivax was 20-fold higher in India compared to Brazil (Siqueira et al. 2015). A greater risk to severe malaria has been found in inpatients for *P. vivax* compared to *P. falciparum* during a prospective study in Papua, Indonesia (Tjitra et al. 2008), where an other study showed a similar case-fataly rate for inpatients with *P. falciparum* and *P. vivax* malaria (Poespoprodjo et al. 2009).

1.6. Immunity against malaria

Immunity against malaria can be naturally acquired after multiple exposures and increases with age. This acquisition is a process that is put in place gradually. In areas where malaria is endemic, individuals are continuously exposed to infections; clinical immunity to malaria is usually acquired in adults (Hviid 2005). The immune response of the host plays an important role in the immunity conferred. It has been shown that malaria immunity is age and transmission intensity dependent (Langhorne et al. 2008; Marsh and Kinyanjui 2006). The mechanisms by which immunity against malaria is acquired are not well understood. However, there is ample evidence to suggest that antibodies to the variant surface antigens (VSAs) of the malaria parasite are responsible for the progressive development of immunity (Bousema et al. 2004; Day and Marsh 1991; Marsh and Howard 1986).

1.7. Prevention

1.7.1. Vaccine against malaria

A large number of malaria vaccine candidates have been developed (World Health Organisation (WHO) 2017c). Among them the most advanced is RTS,S/AS01, a recombinant protein-based malaria vaccine, that has been approved for use by European regulators in July 2015 (Wikipedia 2017). On the 24th April 2017, the World Health Organization Regional Office for Africa (WHO/AFRO) announced that three African countries (Ghana, Kenya, and Malawi) will partner with the WHO in the Malaria Vaccine Implementation Programme (MVIP) that will make the RTS,S vaccine available in selected areas in 2018 ((MVI) 2017). However, the RTS, S vaccine efficacy is limited (Malaria Vaccine Initiative 2016).

1.7.2. Intermittent preventive treatment in pregnancy (IPTp)

If malaria infection occurs during pregnancy, it can be a substantial risk for the mother, the foetus and the neonate (World Health Organisation (WHO) 2017). Malaria during pregnancy can lead to abortion, stillbirth, and low birth weight (McGready et al. 2012). Compared to non-pregnant women, pregnant women are 3 times more likely to suffer from severe malaria, and have a high mortality rate from severe disease (Monif GRG 2004). To reduce maternal malaria episodes, anaemia, and poor delivery outcome, the WHO recommends IPTp with sulfadoxine-pyrimethamine (World Health Organisation(WHO) 2017b). Recently a meta-analysis demonstrated that IPTp with 3 or more doses of sulfadoxine-pyrimethamine protects better against poor delivery outcomes than the standard 2-dose regimens (Kayentao et al. 2013). Recommendation are to observe a one month interval between doses of IPTp-SP, however, sulfadoxine-pyrimethamine is not indicated during the first trimester of pregnancy (World Health Organisation(WHO) 2017b).

1.7.3. Intermittent screening and treatment (IST)

A potential alternative to IPTp-SP, the Intermitent Screening and Treatment (IST) has been shown to be efficacious. In The Gambia, Mali, Burkina Faso and Ghana (countries with a low prevalence of resistance to sulfadoxine-pyrimethamine) IST with Artemether-Lumefantrine (AL) in pregnant women was found to be efficacious as well as IPTp-SP (Tagbor et al. 2015). The study showed that the mean haemoglobin level at the last antennal consultation was 10.97g/dL and 10.94g/dL in women who received the IPTp-SP and ISTp-AL respectively. As for placental infection, it was present in 24.5% of women in the IPTp-SP group and in 24.2%

of women in the ISTp-AL group (OR = 0.95 [95% CI 0.81, 1.12]) (Tagbor et al. 2015). IST is a good alternative, particularly, for pregnant HIV positive women receiving cotrimoxazole prophylaxis in whom SP is contraindicated.

1.7.4. Seasonal malaria chemoprevention (SMC)

Children under 5 years of age are very vulnerable to malaria, accounting for 70% of all malaria related deaths (World Health Organisation(WHO) 2016c). The intermittent administration of full treatment courses of antimalarial medicine to children, called SMC, is recommended in areas of highly seasonal malaria transmission during the transmission season. In order to prevent malarial episodes by maintaining therapeutic antimalarial drug concentrations in the blood, the WHO recommends SMC with sulfadoxine-pyrimethamine + amodiaquine in areas where *P. falciparum* is sensitive to both antimalarial drugs (World Health Organisation(WHO) 2017c). The efficacy of SMC was demonstrated in Mali by substantially reduced prevalence and incidence of malaria, as well as prevalence of anaemia (Diawara et al. 2017).

1.7.5. Long-lasting insecticide-treated nets (LLINs) and indoor residual spray (IRS)

LLINs and IRS are the two tools contributing to current malaria reduction (Bhatt et al. 2015). Both tools are effective against the biting of malaria vectors (World Health Organisation(WHO) 2016c). However, the effectiveness of IRS depends on malaria vectors resistance status and the product residual time on the wall surface (World Health Organisation(WHO) 2015b).

1.8. Malaria diagnosis

For effective management of malaria cases, early accurate and reliable diagnosis is essential. Despite the fact that early the diagnosis is the less expensive way to facilitate treatment, the clinical diagnosis used alone can lead to errors due to the non-specificity of malaria signs (Leslie et al. 2012). WHO guidelines on malaria treatment places particular emphasis on the significance of the confirmation of any clinical suspicion cases (World Health Organisation(WHO) 2010).

For malaria diagnosis, microscopy is the gold standard (Moody, Chiodini, and Work-2000), however, the identification of species other than falciparum, in particular vivax, remains a major challenge. Thick and thin smears are used respectively to detect infections and for species identification. The microscopist experience and the quality of the blood smears are very important for the accuracy of microscopy diagnostic of malaria (Wongsrichanalai et al. 2007). Many malaria endemic regions don't have enough qualified staff and good materials to facilitate reliable and accurate malaria diagnosis (Ishengoma et al. 2009; Mens et al. 2007; Snounou G et al. 1993), leading to presumptive treatments. The Rapid Diagnostic Test (RDT) seems to be a good alternative, however, the sensitivity and specificity of RDT are better with *P. falciparum* infection than that with *P. vivax* (World Health Organisation(WHO) 2012a).

Polymerase chain reaction (PCR) as well as serology are also known to be malaria diagnostic tools with good sensitivity and specificity (Mens et al. 2006, 2007; De Oliveira et al. 2009).

1.9. Treatment of malaria

Quinine has been the drug of last resort for the treatment of malaria, especially for the severe disease. Chloroquine was first synthesized in 1934 and has been the most widely used antimalarial drug. It has been the drug of choice for the treatment uncomplicated malaria and for chemoprophylaxis. Closely related to chloroquine, Amodiaquine is a relatively widely available compound. Some antibiotics such as Tetracycline and derivatives, doxycycline are very potent antimalarials and are used for both treatment and prophylaxis in areas where response to quinine has deteriorated.

1.9.1. The strategies currently recommended for malaria cases management:

The current tools recommended by the World Health Organization for the management of malaria are, on one hand, the prevention by the use of long-lasting insecticidal mosquito nets, indoor residual spraying for vector control, seasonal malaria chemoprevention, and intermittent preventive treatment for malaria in infants, and on the other hand, prompt access to diagnostic testing of suspected cases and the treatment of *P. falciparum* confirmed cases with effective artemisinin-based combination (ACT) therapies (World Health Organisation(WHO) 2016b).

In 2006 Mali changed its policy of antimalarial drug of first choice because of the emergence of malaria parasite resistance to chloroquine. Chloroquine (CQ) was effectively replaced by artemisinin-based combination therapy: artemether + lumefantrine (AL) and artesunate + amodiaquine (AS + AQ) in 2007. The two ACTs recommended by the Malian National Malaria Control Programme, Artemether-Lumefantrine (AL) and Artesunate-

Amodiaquine (AS + AQ), are known to be safe and effective(Djimdé et al. 2008; Sagara et al. 2006).

1.9.2. Treatment of *P. vivax* malaria

Efficient antimalarial treatment in *P. vivax* allows the host to clear peripheral asexual parasites, prevent recurrent infection, and also interrupt the cycle of transmission (Baird, Maguire, and Price 2012). For the treatment of uncomplicated *P. vivax* malaria, the WHO recommends CQ in most endemic countries or an artemisinin combination treatment plus Primaquine (PQ) in G6PD positive individuals (World Health Organisation(WHO) 2014). Almost 30 years after chloroquine-resistant *P. falciparum*, chloroquine-resistant *P. vivax* was first reported in 1989 (Rieckmann, Davis, and Hutton 1989). *P. vivax* resistance to chloroquine has not yet been reported in Sudan and Mauritania (M. Ould Ahmedou Salem et al. 2015; Schousboe et al. 2015), however *Pvmdr1* mutations are emerging. The *Pvmdr1* mutated codon 976F was shown to be associated with treatment failure in Southeast Asia and Ethiopia (Brega et al. 2005; Golassa et al. 2015). In some countries like Brazil, the current treatment for *P. vivax* malaria is based on Chloroquine and Primaquine (Naing et al. 2010). Prevalence of chloroquine resistance in the treatment of P. vivax malaria is becoming more and more higher in Asia (Figure 3).

Evidence on the role of pvcrt-o on the clinical response of CQ treatment has been described in a patient with severe malaria who had an increase in pvcrt-o gene expression 20-fold compared to 3 patients with *P. vivax* uncomplicated malaria (Fernández-Becerra et al. 2009). Also, a study conducted in Brazil showed that pvcrt-o expression was elevated in patients with poor response to CQ (Melo et al. 2014).

Category 1: > 10% recurrences by day 28

Category 2 : at least 5% recurrences by day 28, with the presence of whole-blood chloroquine concentrations greater than 100 nm

Category 3 : at least 5% recurrences by day 28, irrespective of confirmation of adequate blood chloroquine concentration

Figure 3: Prevalence of chloroquine resistance in the treatment of *P. vivax* malaria (Source: adapted from http://www.wwarn.org/vivax/surveyor/#0)

1.10. Management of severe malaria

The management of severe malaria includes emergency measures in addition to antimalarial drug administration (World Health Organisation(WHO) 2012b).

The treatment recommended by WHO is artesunate at 2.4mg/kg of body weight. It should be given intravenously or intramuscularly immediately, 12 and 24 hours later and then once a day. Alternative treatment: Artemether or quinine if artesunate is not available. Artemether should be administrated at 3.2 mg/kg per day.

Quinine is given at 20 mg/kg through intravenous infusion, then 10 mg/kg every 8 hours.

1.11. Next Generation Sequencing (NGS)

The availability of genomic sequences for pathogens can allow the extraction of strong epidemiologycal indicators; in addition, genetic information can also be a key to understanding the causes of a disease and the reactivity of pathogens to drugs. Therefore, the study of the malaria parasite genome is important for disease control and elimination. *Plasmodium* spp. genetic diversity revealed within genome sequencing projects contributed to enlightened the parasite biology and host-parasite interaction (Hemingway et al. 2016). Research on *P. vivax* malaria has been largely neglected leading to a general lack of knowledge about the biology of the plasmodial species. Currently *P. vivax* malaria is known to have considerable morbidity and mortality (Alexandre et al. 2010). Genome, transcriptome and proteome sequencing projects accomplished great achievements on *P. vivax* biology (Acharya et al. 2009, 2011; Alexandre et al. 2010).

One of the challenges in sequencing *P. vivax* DNA extracted from RDTs is low parasite density. To correct this challenge, selective whole genome amplification should be an approaches becoming routine for parasite surveillance once the economic costs outweigh the current cost benefits of targeted approaches (Auburn and Barry 2017).

The term next-generation sequencing (NGS) (Figure 4), describes a number of different modern sequencing technologies including: Illumina (Solexa), Roche 454, Ion torrent: Proton / PGM and SOLiD sequencing. These new technologies allow us to sequence DNA and RNA much more quickly and cheaply and as such have revolutionised the study of genomics and molecular biology (YourGenome 2017).

DNA sequencing using NGS include three general steps (Atdbio 2005):

- Library preparation: during this first step, libraries are created using random fragmentation of DNA either enzymatically or by sonication. DNA fragmentation is followed by ligation with custom linkers;
- Amplification: during the second step, the library is amplified using clonal amplification methods and PCR;
- Sequencing: finally, DNA is sequenced using one of several different approaches.

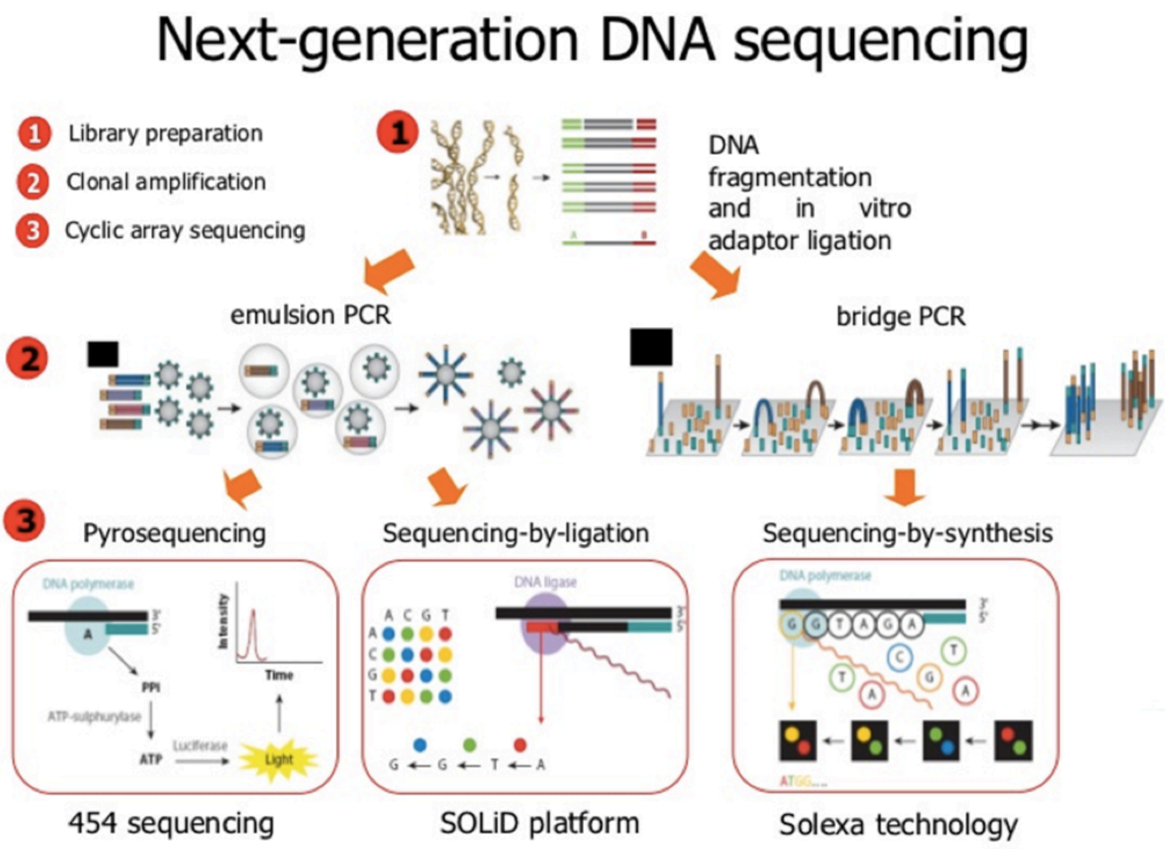


Figure 4: DNA sequencing steps (Source: https://www.slideshare.net/ueb52/introduction-to-next-generation-sequencing-v2)

1.12. Study justification

One of the characteristics that distinguish *Plasmodium vivax* from *Plasmodium falciparum* is the development of dormant hypnozoite forms in the liver that cause subsequent infections in the blood called relapses (Krotoski WA 1985). The World Health Organization (WHO) reported that more than 30% of the world's population is at risk of *P. vivax* infection. This parasite is mostly present in Asia and Latin America. In 2013, more than one million *P. vivax* malaria cases were recorded in four countries (Ethiopia, India, Indonesia and Pakistan) (Mizutani et al. 2016; World Health Organisation(WHO) 2015a). In Saharan Africa, Berber people, who are mostly Duffy-positive have been infected by *P. vivax* (Howes et al. 2015; Picot 2006). The circulation of *P. vivax* in Central Africa (Cameroon), as well as its ability to infect Duffy-negative individuals was described previously (Mbenda, Gouado, and Das 2016; Russo et al. 2017). This particular parasite has been thought to be absent from West Africa because of the Duffy blood group, a blood receptor considered indispensable for the invasiveness of this species into red blood cells. More than 90% of the population in this area are believed to be Duffy negative (Miller et al. 1976). However, in 2012 a report showed that 25 (28.4%) out of

88 samples analysed from Northern Mali contained mature stages of *P. vivax*, (M Bernabeu et al. 2012). Recently, *P. vivax* was found in four black, African women living near Bamako in Mali (Williams et al. 2016). It is not yet known where these cases originated. Hence the importance of studying these parasites in the global context of the malaria elimination agenda. The surprising appearance of *P. vivax* in west Africa and its ability to infect Duffy negative populations is one more threat to population health. Research on *P. vivax* malaria was neglected in West Africa because of the belief that it was relatively benign, it was believed that this species is not able to infect Duffy-negative individuals. In order to contribute to malaria control and elimination efforts, there is a need to investigate the genetic characteristics of these *P. vivax* isolates and the divergence in orthologues of genes conferring antimalarial drug resistance in *P. falciparum.* This is critical in understanding the potential resistance pattern of the parasite to these drugs in Northern Mali. Bioinformatics has the potential to help us understand parasite genetic characteristics.

1.13. Hypothesis and Objectives

1.13.1. Hypothesis

We hypothesize that using selective Whole Genome Amplification, DNA extracted from Rapid Diagnostic tests will allow us to determine the origin of *P. vivax* isolated from Northern Mali.

1.13.2. Objectives

1.13.2.1. General objective

To assess the genetic characteristics and origin of *P. vivax* isolates from Northern Mali.

1.13.2.2. Specific objectives:

- To determine variation levels and their effects in *pvcrt-o* and *pvmdr-1* genes in Northern Mali strains.
- To determine the genetic relationships between *P. vivax* field isolates collected from Northern Mali.
- To determine the evolutionary relationship of the *P. vivax* parasites from Mali with those from other parts of the World.

CHAPTER 2

2. Methods

2.1. Study sites

The analysis used data collected from a previous cross-sectional epidemiological study conducted in October 2015 in the region of Kidal in Northern Mali. As indicated in the Figure 5, Kidal is located at ~1,564 km in North-East of Bamako.

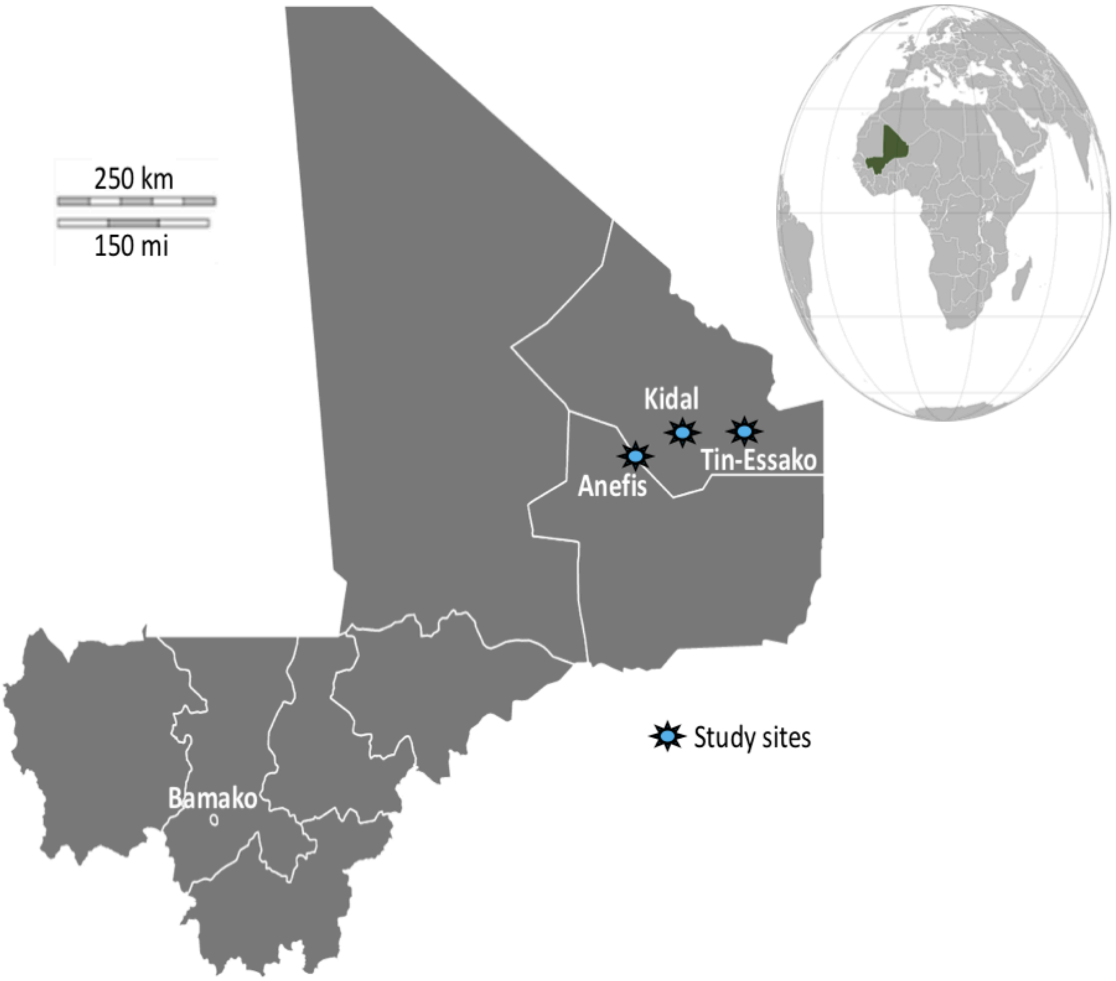


Figure 5: Map of Mali showing study sites

(Source: adapted from https://simplemaps.com/resources/svg-ml)

In Mali, malaria is the primary cause of morbidity and mortality, in particular amongst the vulnerable youth (Ministère de la Santé du Mali 2007; PMI 2016). Malaria is endemic in the central and southern regions of the country and epidemic in the northern regions (Mouhamadou GUEYE, Bourama FANE 2015). In 2012, the prevalence of malaria reported by the Demographic and Health Survey was 52% amongst Malian children under five years of age (PMI 2016). In Mali one in twenty children die before their fifth birthday because of malaria (IRIN 2007). In 2015, 5272 cases of malaria were reported for the region of Kidal in Mali compared to 2267 cases in 2014, an increase of 3005 cases (Ministère de la Santé et de l'Hygiène Publique 2015).

2.2. Study description

Following the emergence of malaria cases in the region of Kidal, the Ministry of Health and Public Hygiene and WHO organized an emergency meeting in Bamako. During this meeting, the MRTC proposed, in addition to an epidemiological cross-sectional study, a response to this emergency with the seasonal malaria chemoprevention (SMC) strategy, indoor residual spraying (IRS) of insecticide, and management of confirmed malaria cases. Malaria rapid diagnosis tests were performed on the patients presenting fever or with history of fever using the SD BIOLINE Malaria Ag *P. f* test. Parasite DNA was extracted from rapid diagnosis tests using the QIAamp DNA mini kit (Qiagen, Valencia, CA) (Annex 19). After nested PCR (Snounou G et al. 1993) (Annex 19), out of 267 samples, 7 were found to be positive for *P. vivax*. DNA from these parasites were subjected to selective Whole-Genome Amplification (sWGA) (Cowell et al. 2017) (Annex 20) followed by NGS (Behjati and Tarpey 2013) using Illumina Platform at MACROGNE in Corea.

2.3. Ethical considerations

The samples were collected anonymously during the management of a malaria epidemic in northern Mali, at the request of the Ministry of Health and Public Hygiene. The analysis of the samples has been approved both by the ethics committee of the Faculty of Medicine, Pharmacy and Dentistry at the University of Sciences, Techniques and Technologies of Bamako in Mali and the ethics committee in the University of Cape Town (UCT), South Africa.

2.4. Next Generation Sequencing

2.4.1. Library QC

Using a DNA 1000 chip, the template size distribution was checked by running on an Agilent Technologies 2100 Bioanalyzer in the order to verify the size of PCR enriched fragments. Using qPCR according to the Illumina qPCR Quantification Protocol Guide, prepared libraries were quantified. Roche's Rapid library standard Quantification solution allowed to generate a standard curve of fluorescence readings. Then, the library sample concentration was calculated.

As shown in table 1 all seven samples passed the Illumina library quality control.
After performing quality control (QC) the library construction was performed.

Table 1: Illumina library quality control

Library name	Library type	Conc. (ng/µl)	Conc. (nM)	Size (bp)	Result
1	TruSeq DNA PCR Free (350)	6.74	22.06	470	Pass
2	TruSeq DNA PCR Free (350)	6.42	21.01	470	Pass
3	TruSeq DNA PCR Free (350)	9.29	30.41	470	Pass
4	TruSeq DNA PCR Free (350)	4.71	15.41	470	Pass
5	TruSeq DNA PCR Free (350)	8.17	26.75	470	Pass
6	TruSeq DNA PCR Free (350)	5.43	17.77	470	Pass
7	TruSeq DNA PCR Free (350)	8.86	29.00	470	Pass

2.4.2. Whole genome sequencing

The Illumina Hiseq platform was used for the whole genome sequencing using the procedure shown in the Figure 6.

Figure 6: Sequencing Experiment Overview (Source: http://dna.macrogen.com)

2.4.2.1. Sample preparation

DNA was extracted from the samples as described above. After performing quality control (QC), the library construction proceeded.

2.4.2.2. Library construction

The sequencing library was prepared by random fragmentation of the DNA or cDNA sample followed by 5' and 3' adapter ligation, then, adapter-ligated fragments were PCR amplified and gel purified.

2.4.2.3. Sequencing

Illumina sequencing by synthesis (SBS) technology was used for the sequencing. The result was highly accurate base-by-base sequencing with few errors, even within repetitive sequence regions and homopolymers.

2.4.2.4. Raw data

Sequenced data were converted into raw data using fastq.gz file format for the analysis.

2.5. Data quality assessment

2.5.1. Quality assessment

The quality scores demonstrate quality and thus the usability of the data. We used fastQC for sequence quality assessment and considered a Phred quality score ≥ 20 as good qualities. More than 90% of the reads had a quality score > 20.

2.5.2. Verification of contamination and filtering

Using bowtie2-2.2.6 (Langmead and Salzberg 2012), we first mapped the *P. vivax* sequences to the human reference genome (Annex 1) in order to check for contamination with host DNA. Samtools (Li et al. 2009) (Version: 1.5) was used to compute the mapping rate (Annex 2). Thereafter, we extracted the reads that did not align with the human reference genome using Bam2fastq version1.1.0 (HudsonAlpha 2010)(Annex 3).

2.6. Reference genomes

The *P. vivax* reference genome (PlasmoDB-35_PvivaxP01_Genome.fasta) was downloaded from PlasmoDB (http://plasmodb.org/common/downloads/Current_Release/PvivaxP01/fasta/data/).

The human reference genome was downloaded from the European Nucleotide Archive database (ENA at https://www.ebi.ac.uk/ena/data/view/GCA_000001405.26) in fasta format. *P. vivax* whole genome sequences from Mauritania, Madagascar, Brazil, Peru and India were also downloaded from the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/data/search?query=).

2.7. Sample size

For this analysis, we used 7 *P. vivax* whole genome sequences from the region of Kidal in Mali, 8 whole genome sequences from Mauritania, 3 from Madagascar, 3 from India, 3 from Brazil and 5 from Peru (see Figure 7). Thus, a total of 29 *P. vivax* whole genome sequences, all sequenced on Illumina HiSeq 2000, were used.

Figure 7: World map showing the origins of the parasite sequence data (Source: adapted from https://commons.wikimedia.org/wiki/File:World_map_between_2003_and_2005.png)

2.8. Data Analysis procedure

2.8.1. Next Generation Sequencing Analysis

We ran the following pipelines:

- Quality Control
- Read Alignment

- Variant Identification
- Annotation

2.8.1.1. Read Alignment

The reads that did not align with the human reference genome were mapped onto the *P. vivax* reference genome using SMALT (http://www.sanger.ac.uk/resources/software/smalt) with default parameters to first extract the *P. vivax* reads. We evaluated the sequence conformity to the reference genome with Samtools (Annex 4). Next, the reads aligned with the *P. vivax* reference genome were extracted using Bam2fastq version1.1.0 (HudsonAlpha 2010)(Annex 5) and re-aligned with the *P. vivax* reference genome using SMALT (Annex 6).

2.8.1.2. Variant calling

Variant calling was done using Samtools-0.1.19 following 3 steps. We first indexed the reference genome; next, we used samtools (Li et al. 2009) to generate a BCF file; and finally, using bcftools-0.1.19, the variants were filtered (Annex 7).

Like the samples isolated from Northern Mali, all the data downloaded from the ENA were subjected to contamination filtering, alignment to the *P. vivax* reference genome and variant calling following the same pipeline (Annex 11, 12, 13, 14 and 15).

2.8.1.3. Annotation

An open source tool, snpEff 4.3t (build 2017-11-24 10:18) (Li et al. 2017) was used to annotate variants and predict the coding effects. The *P. vivax* genes of interest (chloroquine resistance transporter: *pvcrt-o* and multidrug resistance protein: *pvmdr-1*, were searched in the NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene/).

SnpSift version 4.3t (build 2017-11-24 10:18) was used to extract the variation positions within antimalarial drugs resistance genes descriptions in *EnsemblProtists (https://protists.ensembl.org/biomart/martview/d4bb190f94368e623c0f9ea7894abfeb)*.

2.8.2. Population differentiation

PLINK-version 1.90b5.1(Alexander and Lange 2011) was used to generate a pca file from the Mali *P. vivax* vcf file to measure the Northern Mali *P. vivax* population differentiation (Annex 8). The eigenvec file was used in R (version 3.4.2) to perform a principal component analysis (Annex 9) to determine the genetic distance among parasites collected in Northern Mali.

2.8.3. Neighbour Joining analysis

Similar to the analysis performed on Malian samples, we performed an NGS analysis on strains from other countries. The BAM files were used to generate common variant calling files. Then, positions overlapping with the strains of Northern Mali with the highest number of reads were used for Neighbour Joining analysis (Annex 15). Using PLINK 1.9, ped files were generated from variant calling files. The ped files were converted into fasta files (Annex 16). Using the MAFFT/7.305_intel16.0.1 program, we aligned the sequences (Annex 17) to determine the evolutionary relationships.

A Neighbour-Joining method (Saitou and Nei 1987) was used to study the evolutionary history of the taxa. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura, Nei, and Kumar 2004). This analysis involved 28 nucleotide sequences from Northern Mali, Mauritania, Madagascar, Brazil, Peru and India. All positions containing gaps and missing data were eliminated. The evolutionary analysis was done using MEGA7 (Kumar, Stecher, and Tamura 2016).

CHAPTER 3

3. Results

In order to determine the origin of the 7 *P. vivax* sequences isolated in Northern Mali, we first assessed the quality of the sequence data, processed it and did variant calling. Then we did pairwaise differentiation and evolutionary analysis using 22 additional existing *P. vivax* sequences from other regions of the World.

3.1. Determination of the data quality

3.1.1. NanoDrop reports

The selective Whole Genome Amplification allowed DNA enrichment for WGS. Before amplification, the DNA quantities were almost undetectable. For the second sample, from just traces, the quantity increased to exceed 15 ng/µl.

3.1.2. Sequence quality

The average GC content was about 41. More than 91% of the reads had a quality control (QC) score above 20. The mapping of our sequences with the human reference genome showed a high proportion of contamination with host DNA (Table 2). On average, 80% of the reads were human DNA.

Table 2: Sample quality summary statistics

	Total reads	GC content	Q20 (%)	Mapping rate on human reference genome (%)
Sample 1	64,319,870	41.07	91.78	70.15
Sample 2	66,094,594	43.05	91.68	57.42
Sample 3	57,033,898	39.19	92.53	97.34
Sample 4	65,600,582	42.06	91.62	55.63
Sample 5	68,643,540	39.78	92.8	95.07
Sample 6	73,421,330	40.49	92.24	87.53
Sample 7	65,239,736	40.24	92.32	94.21

After filtering the reads that did not align with the human reference genome, for the alignment with the *Plasmodium vivax* reference genome (Table 3), the sample from TinEssako (region of Kidal) had the most reads mapping to *P. vivax P01*.

The result illustrates that in addition to human DNA, the samples were highly contaminated with other organisms.

Table 3: Mapping rate with *P. vivax* reference genome after filtering out human DNA

	Origin	Total reads unaligned with human reference genome	Reads mapped to *P. vivax* reference genome	Mapping rate (%)
Sample 1	Kidal	17810776	755252	4.24
Sample 2	Kidal	26444138	756434	2.86
Sample 3	Kidal	159580	43340	27.16
Sample 4	Kidal	27747574	1117428	4.03
Sample 5	Kidal	144920	10053	6.94
Sample 6	Anefis	7174992	333354	4.65
Sample 7	Tin-Essako	1855108	1285572	69.30

Despite the high contamination with human and other DNA, the remaining sequences had a number of reads covered by the *P. vivax* reference genome. When all contamination was removed, the re-alignment with the *P. vivax* reference genome showed that the sample collected in TinEssako had the most reads aligning with the *P. vivax* reference genome. The deep coverage was above five for Kidal2 and Tin-Essako. Sample 5 which has lower x-coverage (Figure 8), and fewer than 4,000 reads, was not used for further analysis.

Table 4: *P. vivax* DNA extracted mapping rate with *P. vivax* reference genome

	Origin	Total reads extracted as *P. vivax* DNA	Reads mapped with *P. vivax* reference genome	Mapping rate (%)
Sample 1	Kidal	63654	60932	95.72
Sample 2	Kidal	66868	64729	96.80
Sample 3	Kidal	21300	21068	98.91
Sample 4	Kidal	80820	77164	95.48
Sample 5	Kidal	2886	**2837**	98.30
Sample 6	Anefis	49334	47938	97.17
Sample 7	Tin-Essako	1230732	1229987	99.94

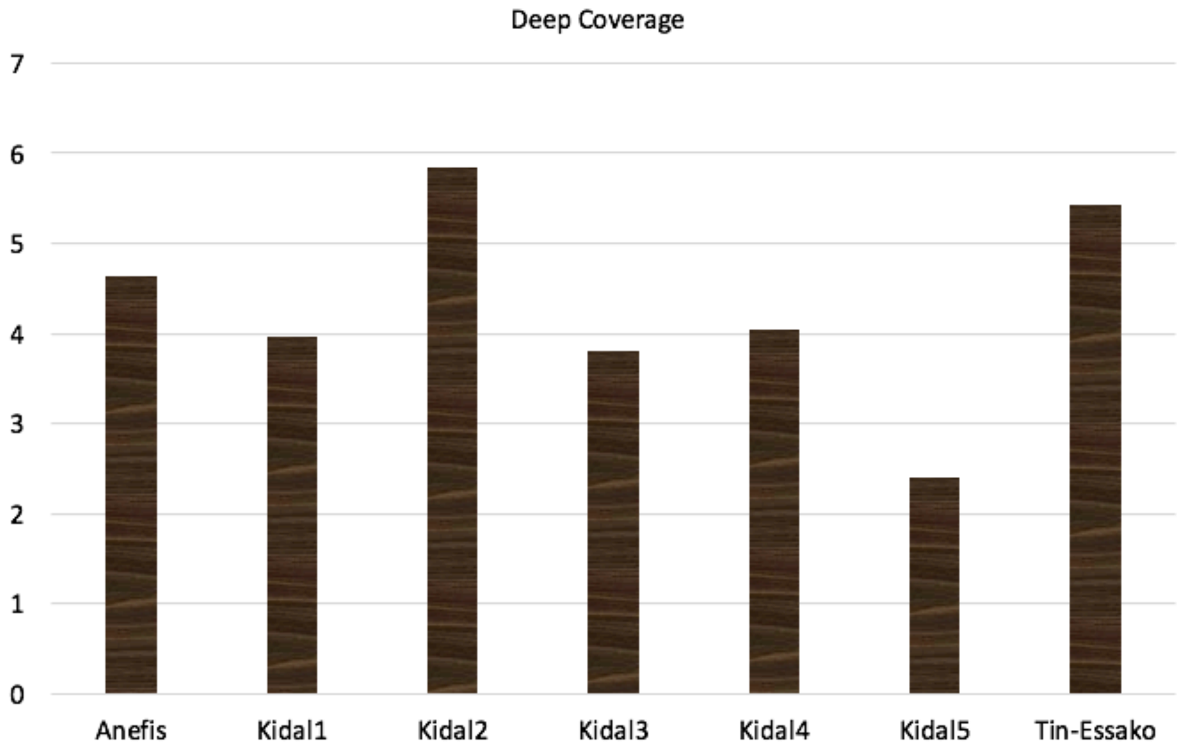


Figure 8 : Depth of coverage of the *P. vivax* sequences collected in Northern Mali

3.2. Variant annotation

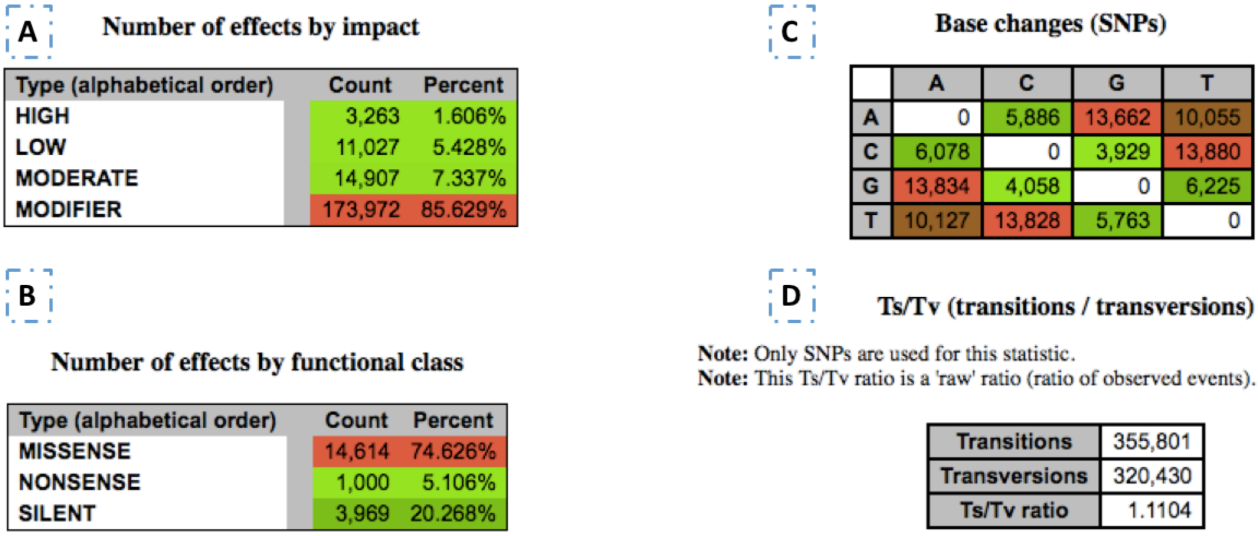
We then did variant annotation using snpEff for the 6 samples from the Northern Mali that had enough reads. Table 5 shows the chromosome length, the numbers of variants as well as the average number of bases that separate the variants. The variants are closer to each other in chromosome 4 where there is a variant on average each 196 bases. The density is highest for chromosome 4, followed by chromosome 10 with a variant density of 206. The greatest distance between the variants was observed in chromosome 14. The lowest number of variants was observed on chromosome 1, and the highest number is on chromosome 9.

Table 5: Variants number and density by chromosome.

Chromosome	Length	Variants	Variant density
Chr1	830,022	2,629	315
Chr2	755,035	3,527	214
Chr3	1,011,127	3,083	327
Chr4	876,622	4,459	196
Chr5	1,370,936	3,898	351
Chr6	1,033,388	2,715	380
Chr7	1,497,819	4,045	370
Chr8	1,678,596	3,886	431
Chr9	1,923,364	7,611	252
Chr10	1,419,739	6,868	206
Chr11	2,067,354	4,651	444
Chr12	3,004,884	6,445	466
Chr13	2,031,768	4,194	484
Chr14	3,120,417	5,968	522

Genetic variation provides a way for populations to evolve in response to environmental changes. In our analysis, unsurprisingly, the variants with high effects have the lowest proportion (1.6%) (table 6-A). However, the results show a high level of missense mutation (almost 75%) (table 6-B). The ratio of transitions/transversions was about 1.11 (table 6-D).

Table 6: Results of variant effect predictions



A

Number of effects by impact

Type (alphabetical order)	Count	Percent
HIGH	3,263	1.606%
LOW	11,027	5.428%
MODERATE	14,907	7.337%
MODIFIER	173,972	85.629%

B

Number of effects by functional class

Type (alphabetical order)	Count	Percent
MISSENSE	14,614	74.626%
NONSENSE	1,000	5.106%
SILENT	3,969	20.268%

C

Base changes (SNPs)

	A	C	G	T
A	0	5,886	13,662	10,055
C	6,078	0	3,929	13,880
G	13,834	4,058	0	6,225
T	10,127	13,828	5,763	0

D

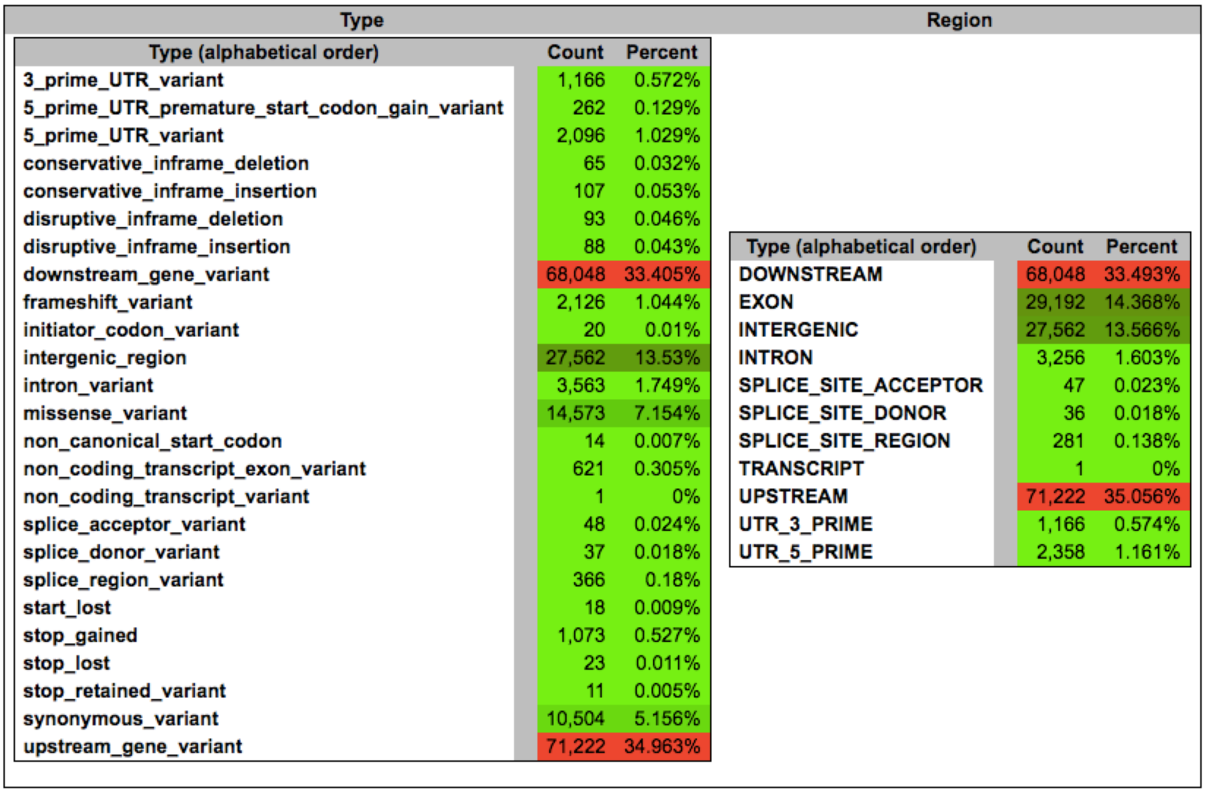
Ts/Tv (transitions / transversions)

Note: Only SNPs are used for this statistic.
Note: This Ts/Tv ratio is a 'raw' ratio (ratio of observed events).

Transitions	355,801
Transversions	320,430
Ts/Tv ratio	1.1104

With regard to the frequency of effects by type and localization (table 7), downstream (33.4%) and upstream (34.8%) regions of genes are carrying more than 67% of the variants. Variants were much more often observed on exons (14.4%) compared to introns (1.6%). Non-coding transcript variants were the least frequent (0.001%).

Table 7: Number of variant effects by type and region



Type

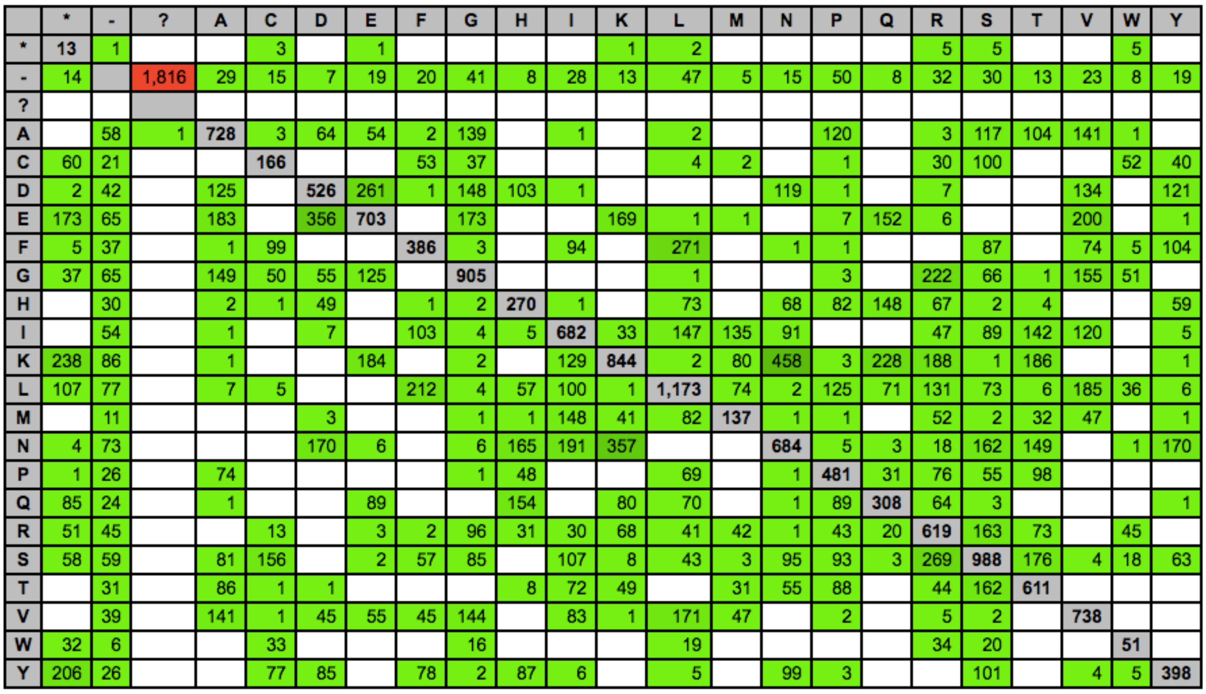
Type (alphabetical order)	Count	Percent
3_prime_UTR_variant	1,166	0.572%
5_prime_UTR_premature_start_codon_gain_variant	262	0.129%
5_prime_UTR_variant	2,096	1.029%
conservative_inframe_deletion	65	0.032%
conservative_inframe_insertion	107	0.053%
disruptive_inframe_deletion	93	0.046%
disruptive_inframe_insertion	88	0.043%
downstream_gene_variant	68,048	33.405%
frameshift_variant	2,126	1.044%
initiator_codon_variant	20	0.01%
intergenic_region	27,562	13.53%
intron_variant	3,563	1.749%
missense_variant	14,573	7.154%
non_canonical_start_codon	14	0.007%
non_coding_transcript_exon_variant	621	0.305%
non_coding_transcript_variant	1	0%
splice_acceptor_variant	48	0.024%
splice_donor_variant	37	0.018%
splice_region_variant	366	0.18%
start_lost	18	0.009%
stop_gained	1,073	0.527%
stop_lost	23	0.011%
stop_retained_variant	11	0.005%
synonymous_variant	10,504	5.156%
upstream_gene_variant	71,222	34.963%

Region

Type (alphabetical order)	Count	Percent
DOWNSTREAM	68,048	33.493%
EXON	29,192	14.368%
INTERGENIC	27,562	13.566%
INTRON	3,256	1.603%
SPLICE_SITE_ACCEPTOR	47	0.023%
SPLICE_SITE_DONOR	36	0.018%
SPLICE_SITE_REGION	281	0.138%
TRANSCRIPT	1	0%
UPSTREAM	71,222	35.056%
UTR_3_PRIME	1,166	0.574%
UTR_5_PRIME	2,358	1.161%

Table 8 summarizes the amino acid changes where the rows are reference amino acids. Phenylalanine (F) is replaced 263 times by Leucine (L), both are essential amino acids. Histidine (H) is replaced 143 times by Glutamine (Q), a non-essential amino acid. Isoleucine is highly replaced by Leucine (both are essentials). Leucine is more often replaced by Phenylalanine (F) and Methionine (M) by Isoleucine. The essential amino acid Threonine (T) is replaced 152 times by the non-essential Serine (S). Diagonals indicated using grey background color are synonymous mutations.

Table 8: Amino acid changes



	*	-	?	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
*	13	1			3		1					1	2					5	5			5	
-	14		1,816	29	15	7	19	20	41	8	28	13	47	5	15	50	8	32	30	13	23	8	19
?																							
A		58	1	728	3	64	54	2	139		1		2			120		3	117	104	141	1	
C	60	21			166			53	37				4	2		1		30	100			52	40
D	2	42		125		526	261	1	148	103	1				119	1		7			134		121
E	173	65		183		356	703		173			169	1	1		7	152	6			200		1
F	5	37		1	99			386	3		94		271		1	1			87		74	5	104
G	37	65		149	50	55	125		905				1			3		222	66	1	155	51	
H		30		2	1	49		1	2	270	1		73		68	82	148	67	2	4			59
I		54		1		7		103	4	5	682	33	147	135	91			47	89	142	120		5
K	238	86		1			184		2		129	844	2	80	458	3	228	188	1	186			1
L	107	77		7	5			212	4	57	100	1	1,173	74	2	125	71	131	73	6	185	36	6
M		11				3			1	1	148	41	82	137	1	1		52	2	32	47		1
N	4	73				170	6		6	165	191	357			684	5	3	18	162	149		1	170
P	1	26		74					1	48			69		1	481	31	76	55	98			
Q	85	24		1			89			154		80	70		1	89	308	64	3				1
R	51	45			13		3	2	96	31	30	68	41	42	1	43	20	619	163	73		45	
S	58	59		81	156		2	57	85		107	8	43	3	95	93	3	269	988	176	4	18	63
T		31		86	1	1				8	72	49		31	55	88		44	162	611			
V		39		141	1	45	55	45	144		83	1	171	47		2		5	2		738		
W	32	6			33				16				19					34	20			51	
Y	206	26			77	85		78	2	87	6		5		99	3			101		4	5	398

3.3. Genes related to resistance against antimalarial drugs in Northern Mali strains

The chloroquine resistance transporter (*pvcrt-o*) is found on Chromosome 1, and the multidrug resistance protein (*pvmdr-1*), on Chromosome 10. In total, 58 variants were found in the two genes from Northern Mali strains using SNPeff. In *pvcrt-o* we found one variant leading to a high effect on Chloroquine resistance associated protein (table 9). Two variations in the *pvmdr-1* gene whose one at the position 357357 have large effects on an uncharacterized protein (PVX_080090) (table 10).

From the *EnsemblProtists* variants database, we looked for *pvcrt-o* and *pvmdr-1*gene orthologues with the same variants in *Plasmodium falciparum* 3D7. From *EnsemblProtists* the variants we observed were: rs45338627, rs45319121, rs45319122, rs45365696, rs45319123, rs45365697 rs45319125, rs45365699, rs45338628 and rs45319126, all were carried by Chromosome 10. However, none of them expressed a functional or phenotypic description.

None of the variants observed in the gene *pvcrt-o* nor 65.5% (19 out of 29) of the variants observed on the *pvmdr-1* gene have been described yet in the *EnsemblProtists* variants database and therefore appear to be novel.

Table 9: Chloroquine resistance transporter (*pvcrt-o)* variants and their effects

Chromosome	Position	REF	ALT	Type of variant	Effect
chr1	326582	T	C	Synonymous	Low
chr1	326764	G	C	Missense	Moderate
chr1	327016	T	C	Upstream gene	Modifier
chr1	329317	G	A	Upstream gene	Modifier
chr1	330262	G	A	5′ UTR	Modifier
chr1	330541	TAT	T	5’UTR	Modifier
chr1	330818	G	GG	5’UTR	Modifier
chr1	330857	A	C	5’UTR	Modifier
chr1	330860	T	C	5’UTR	Modifier
chr1	330862	G	C	5’UTR	Modifier
chr1	330884	T	C	5’UTR	Modifier
chr1	331359	G	A	Synonymous	Low
chr1	331421	G	C	Missense	Moderate
chr1	332758	C	T	Upstream gene	Modifier
chr1	333045	T	C	Upstream gene	Modifier
chr1	333863	C	A	Upstream gene	Modifier
chr1	333864	A	C	Upstream gene	Modifier
chr1	333868	C	A	Upstream gene	Modifier
chr1	335228	T	TT	Upstream gene	Modifier
chr1	336610	A	C	Upstream gene	Modifier
chr1	336773	T	TT	Upstream gene	Modifier
chr1	336962	G	A	Upstream gene	Modifier
chr1	337343	C	T,A	Initiator codon	Low
chr1	338023	C	T	Synonymous	Low
chr1	338133	T	C	Missense	Moderate
chr1	338140	T	C	Synonymous	Low
chr1	338622	A	C	Missense	Moderate
chr1	338847	A	G	Missense	Moderate
chr1	339245	AA	A	Frameshift	High

Table 10: Multidrug resistance protein (*pvmdr-1)* variants and their effects

Chromosome	Position	REF	ALT	Type of variant	Effect
chr10	356721	C	T	Missense	Moderate
chr10	357357	T	TTT	Frameshift	High
chr10	357878	T	G	Upstream gene	Modifier
chr10	357885	G	C	Upstream gene	Modifier
chr10	357887	A	G	Upstream gene	Modifier
chr10	357890	A	T	Upstream gene	Modifier
chr10	357891	G	T	Upstream gene	Modifier
chr10	357894	A	G	Upstream gene	Modifier
chr10	357898	C	G	Upstream gene	Modifier
chr10	357900	G	A	Upstream gene	Modifier
chr10	357903	T	A	Upstream gene	Modifier
chr10	357988	T	G	Upstream gene	Modifier
chr10	358161	T	A	Upstream gene	Modifier
chr10	358819	A	C	5'UTR	Modifier
chr10	359147	A	AA	Frameshift	High
chr10	359988	C	T	Missense	Moderate
chr10	360536	T	G	Missense	Moderate
chr10	360602	A	G	Missense	Moderate
chr10	361889	A	G	Missense	Moderate
chr10	363144	C	T	Synonymous	Low
chr10	363676	G	A	Missense	Moderate
chr10	363970	G	T	Missense	Moderate
chr10	365622	T	C	Synonymous	Low
chr10	366628	C	A	Upstream gene	Modifier
chr10	366718	A	T	Upstream gene	Modifier
chr10	367286	G	A	Upstream gene	Modifier
chr10	369415	G	GGG	Upstream gene	Modifier
chr10	370450	T	G	Missense	Moderate
chr10	370990	C	T	Synonymous	Low

ALT= alternate allele, REF= reference allele, UTR= Untranslated region

3.4. Determination of the genetic relationships between *P. vivax* field isolates collected in Northern Mali

Using PLINK we investigated the genetic distance between *P. vivax* strains collected in the Kidal region. Table 11 shows the standard deviation of each component used for the prinincipal component analysis. The standard deviations of the components were ~0.45 for each component, and the proportions of variance were the same at 20%.

Table 11: The importance of components

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	0.4472	0.4472	0.4472	0.4472	0.4472
Proportion of Variance	0.2000	0.2000	0.2000	0.2000	0.2000
Cumulative Proportion	0.2000	0.4000	0.6000	0.8000	1.0000

*Comp= component

The principal component analysis (PCA) plot (Figure 9) shows that strains isolated from rural areas (Anefis and Tin-Essako) are genetically close to each other, and the strains collected in the Kidal district are clustering together. Only, the third strain collected in the Kidal district is genetically distant from the others.

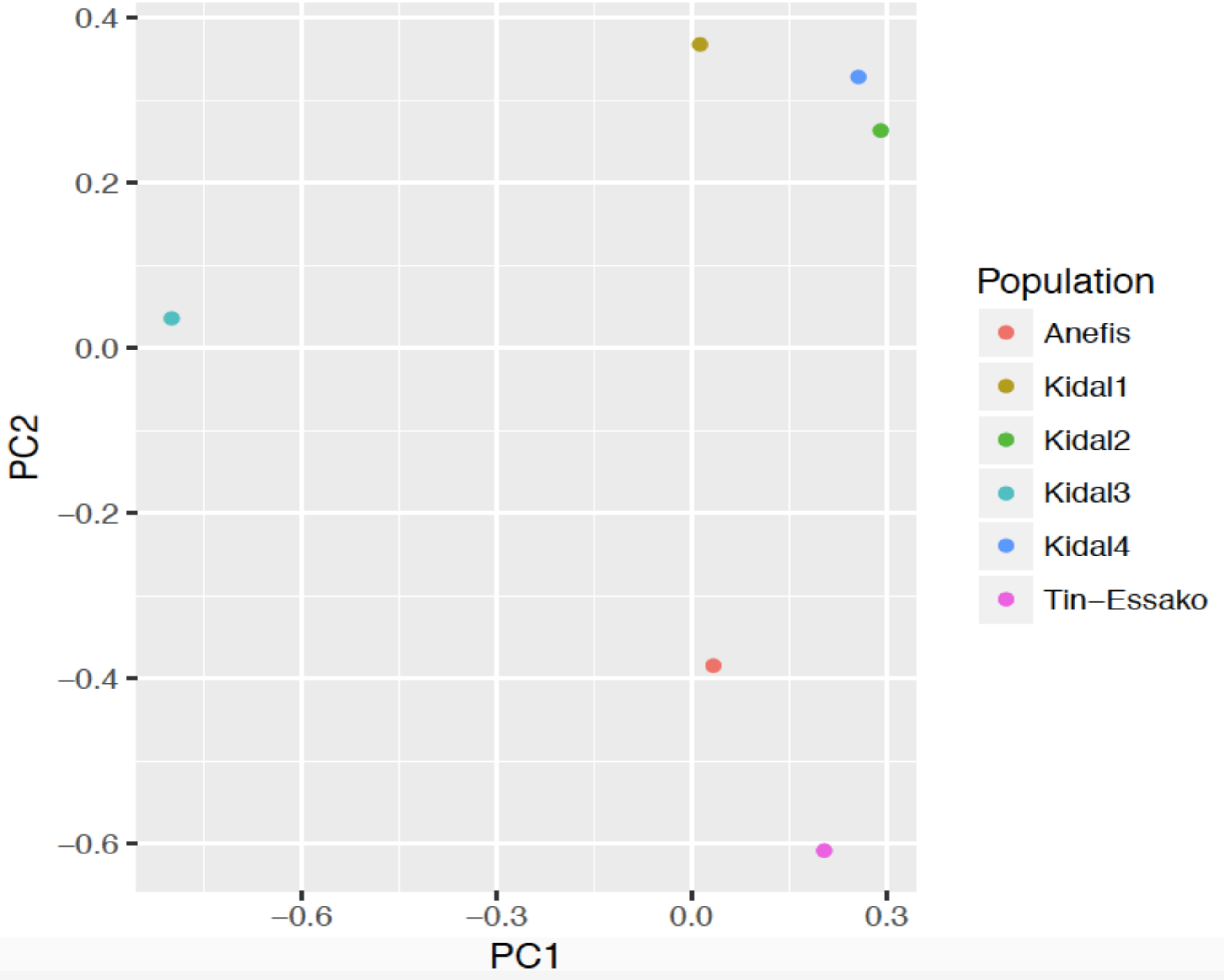


Figure 9: Genetic distance of *P. vivax* field isolates collected in Northern Mali

3.5. Determination of the genetic distance between *P. vivax* parasites from Northern Mali and those from other parts of the world

The Neighbour-Joining analysis was done using *P. vivax* genome isolates from Northern Mali and other countries in Africa, Asia and Latina America. The phylogenetic tree (Figure 10) shows clearly that almost all strains of Northern Mali are clustering together and are genetically distinct from those from Mauritania. Only the strain collected at Tin-Essako is moving away from the Mali isolates and is not close to any other strain used in this analysis. The phylogenetic tree demonstrates that *P. vivax* strains from Mauritania are clustering alone and are genetically distinct from those of other parts of the world. The 3 strains from Madagascar used in this analysis are genetically closely related to the strains from India. The Neighbour-Joining analysis shows that isolates from Latina America are clustering together and genetically close the strains isolated from India.

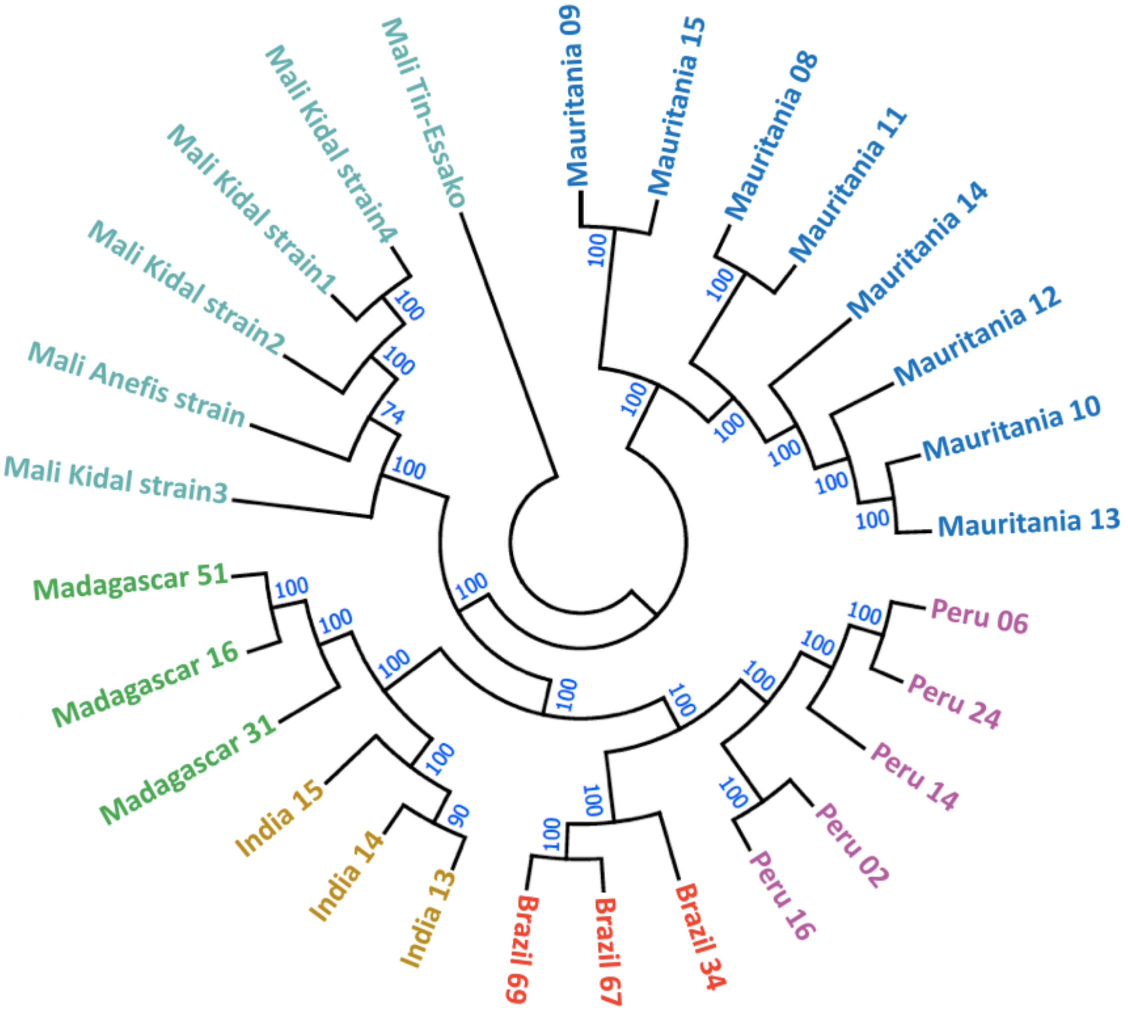


Figure 10: Evolutionary relationship of *P. vivax* taxa

CHAPTER 4

4. Discussion

4.1. Study context and data quality

This study was conducted in a context of emergency management why rapid diagnostic tests was used for malaria diagnostic. For long time, *P. vivax* has been thought absent from West-Africa, due to the absence of the Duffy antigen (Maria Bernabeu et al. 2012). However, scientific evidence shows that human *P. vivax* is recognized to be from Africa (Liu et al. 2014). Nowadays, only one route experimentally confirmed, using the Duffy antigen, is known to be used by *P. vivax* for invading reticulocytes (Horuk et al. 1993; Miller et al. 1976). Regardless of its pathway for entrance into reticulocytes and geographical origin, the presence of *P. vivax* in Africa is not sufficiently explored.

We have shown for the first to our knowledge that *P. vivax* whole genome can be recovered from a RDT, which is used for diagnostics. The result shows that selective whole-genome amplification (SWGA) is a good approach for *P. vivax* DNA enrichment for Whole Genome Sequencing. Cowell *et al*. demonstrated that sWGA is a robust tool for the enrichment of *P. vivax* DNA extracted from human blood samples and dried blood spots for the purpose of high-quality WGS (Cowell et al. 2017).

Our samples were, however significantly contaminated with the host's DNA and other organisms (Tables 2 and 3). This probably occured because the samples are from rapid diagnostic tests, and the depletion of white blood cells (WBC) was not possible before parasite DNA extraction. A previous study showed that, with over 80% human host DNA contamination, enzymatic treatment enriches *P. falciparum* DNA for more efficient WGS (Oyola et al. 2013). For our study, due to time constraints and availability of samples, we had to proceed with the data we got from our samples.

4.2. Single Nucleotide Polymorphisms and Indels

The coverage of Northern Mali sequences was not very deep, just over the 5x (Figure 8). Although the hurdle of SNP calling using low-coverage sequencing data has been recognized (Yu and Sun 2013), we still managed to get SNP's called from these sequences.

P. vivax isolates from northen Mali displayed extensive genetic variation. Mutation provides a way for a parasite population to evolve in response to environmental changes (Furusawa 2012). Northern Mali *P. vivax* strains have few variants with high effects but a high rate of missense mutation (Table 6). It is known that missense mutations can lead to non-functional proteins (Minde et al. 2011), and such mutations can lead to diseases in humans (Boillée, Vande Velde, and Cleveland 2006) or drug resistance in parasites.

4.3. Antimalarial drug resistance genes

Chloroquine resistance in *P. falciparum* was found to be strongly associated with the *Pfcrt* gene K76T mutation (Fidock et al. 2000). However, it has been described in previous publications that the *P. vivax* ortholog of *Pfcrt*, *Pvcg10* was not involved in chloroquine resistance in *P. vivax* (Nomura et al. 2001).

We did not identify variants in the *pvcrt-o* or *pvmdr-1* orthologs in *P. falciparum* that could lead to antimalarial resistance. This suggests that Chloroquine and Primaquine recommended in the treatment of *P. vivax* malaria (ALVING et al. 1953; López-Antuñano 1999) could remain effective against Northern Mali strains. The resistance of *P. vivax* to Chloroquine has not yet been documented in Mauritania or Sudan (M. Ould Ahmedou Salem et al. 2015; Schousboe et al. 2015). However, *P. vivax* resistance to Chloroquine has reached an alarming level in Indonesia, East Timor and Papua New Guinea (Baird 2009). Convincing evidence of *P vivax* resistance to chloroquine has been reported from Thailand (Phyo et al. 2011), Cambodia (Leang et al. 2013), Ethiopia (Yohannes et al. 2011) and South America (De Santana Filho et al. 2007). In somes countries like Indonesia, Malaysia and Cambodia ACT is advised for *P. vivax* malaria treatment (Grigg et al. 2016; Novotny et al. 2016; World Health Organisation(WHO) 2017a).

4.4. Genetic relationships between *P. vivax* field isolates collected in Northern Mali

Principal component analysis is a powerful method for discriminating populations that are genetically very close (Arora et al. 2011). In the PCA, we used 6 clinical samples among which, 4 were from the Kidal district, 1 from Anefis and 1 from TinEssako, from the region of Kidal. Anefis and TinEssako are based respectively 95 km southwest and 115 km east of the Kidal district. The plot in Figure 19 shows that the strains are clustering based on the sampling area. Only one strain from the Kidal district seemed genetically distinct from the others and could have been imported from another locality by a patient travelling from another village to the district. In a previous study, Hupalo *et al.* demonstrated that *P. vivax* genomic sequences cluster mostly according to their geographic origins (Hupalo et al. 2016). High genome-wide diversity has been found in relatively close geographic areas in *P. vivax* populations in the Americas (de Oliveira et al. 2017). It has been described in several publications that in Africa, local *P. vivax* populations are nearly as polymorphic as their *P. falciparum* counterparts (Chang et al. 2012; Mobegi et al. 2014; Prabhat Jha et al 2016).

The phylogenetic tree (Figure 10) shows that *P. vivax* strains from Northern Mali, which are genetically close to each other, seem not to have originated from patients from the neighbouring country, Mauritania, while the strains from Latina America are clustering together. The result confirms that *P. vivax* parasite populations are circulating in west Africa. *P. vivax* infection was described in pregnant women on the periphery of Bamako in 2016 (Williams et al. 2016) and recently in Duffy-negative children in Bandiagara (Niangaly et al. 2017), in the centre of Mali. The authors suggest that it is possible that *P. vivax* could have been introduced from the Sahara desert in the north (M Bernabeu et al. 2012; Koita et al. 2012). Evidence of *P. vivax* infection in Duffy negative populations was previously described in Western Kenya (Ryan et al. 2006).

Surprisingly, one strain isolated from Northern Mali, Tin-Essako, seemed to be genetically distinct from the other strains used in this analysis and could have a different ancestry. The strains from Madagascar are clustering with those of India and are genetically closer to the strains from India, Brazil and Peru.

Scientific evidence shows that the origin of human *P. vivax* is in Africa because it arose from within a *Plasmodium* species that infects chimpanzees and gorillas (Liu et al. 2014).

Though previously several authors assumed that *P. vivax* originated from Asia (Carlton, Das, and Escalante 2013; Escalante et al. 2005; Mu et al. 2005).

Regarding the spread of the *P. vivax* species, we can propose the two hypothesis. First, the African slave trade to the Americas and Europe in the 16th century could have favoured the migration of *Plasmodium vivax* to North America, India and Asia (Figure 11). secondly, merchants carrying goods by boat could also have spread the *P. vivax* species (Figure 12).

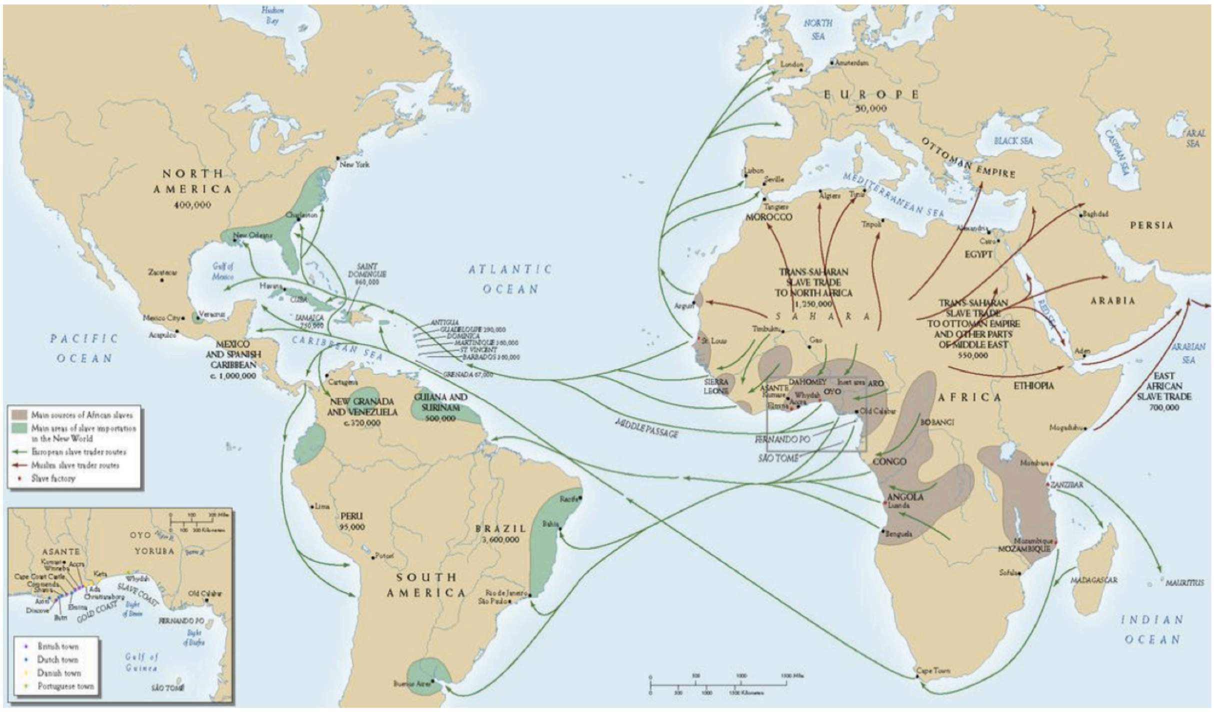


Figure 11: African slave trade possibly leading to the spread of *P. vivax* (Source: http://www.wwnorton.com/worlds/ch4/maps.htm)

Figure 12: World map showing a plausible path of migration of *P. vivax* from India to Madagascar and Mauritania (Source: adapted from https://commons.wikimedia.org/wiki/File:World_map_between_2003_and_2005.png)

CHAPTER 5

5. Conclusions and Limitations

5.1. Conclusions

This study aimed to investigate the usefullness of sWGA to enrich *P. vivax* DNA extract from RDTs, the origin of the strains isolated in Northern Mali and their susceptibility to antimalarial drugs.

The results obtained allow us to draw the following conclusions:

- For low parasite density, even when the DNA is extracted from a RDT, this study consolidates that selective Whole Genome Amplification is a useful alternative for *P. vivax* DNA enrichment for Whole Genome Sequencing.
- This study confirms that *Plasmodium vivax* strains circulating in Northern Mali are genetically distinct from those of Mauritania.
- We did not identify mutations associated to the resistance to antimalarial drugs in pvcrt-o and pvmdr-1 genes.

5.2. Limitations

We note some important limitations to this analysis. Genomic DNA was extracted from RDTs, which in addition to not allowing the depletion of white blood cells might have suffered from the known low parasite density inherent to *P. vivax*. Also, the sample size was smaller than that needed for some kinds of analyses, such as genome-wide association studies. A large sample size is needed to further investigate the susceptibility to antimalarials of strains of P. vivax isolated in northern Mali.

References

Malaria Vaccine Initiative (MVI), 2017. “RTS,S.” *http://www.malariavaccine.org/malaria-and-vaccines/first-generation-vaccine/rtss; Accessed 28 May 2017.*

Acharya, Pragyan et al. 2009. “A Glimpse into the Clinical Proteome of Human Malaria Parasites Plasmodium Falciparum and Plasmodium Vivax.” *Proteomics - Clinical Applications*.

Acharya, Pragyan et al. 2011. “Clinical Proteomics of the Neglected Human Malarial Parasite Plasmodium Vivax.” *PLoS ONE*.

Alexander, David H. and Kenneth Lange. 2011. “Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation.” *BMC Bioinformatics* 12.

Alexandre, Márcia A. et al. 2010. “Severe Plasmodium Vivax Malaria, Brazilian Amazon.” *Emerging Infectious Diseases*.

ALVING, A. S. et al. 1953. “Korean Vivax Malaria. II. Curative Treatment with Pamaquine and Primaquine.” *The American journal of tropical medicine and hygiene* 2(6):970–76.

Anstey, Nicholas M. et al. 2007. “Lung Injury in Vivax Malaria: Pathophysiological Evidence for Pulmonary Vascular Sequestration and Posttreatment Alveolar-Capillary Inflammation.” *The Journal of Infectious Diseases* 195(4):589–96. Retrieved (https://academic.oup.com/jid/article-lookup/doi/10.1086/510756).

Anstey, Nicholas M., Nicholas M. Douglas, Jeanne R. Poespoprodjo, and Ric N. Price. 2012. “Plasmodium Vivax: Clinical Spectrum, Risk Factors and Pathogenesis.” *Advances in Parasitology* 80:151–201.

Anstey, Nicholas M., Bruce Russell, Tsin W. Yeo, and Ric N. Price. 2009. “The Pathophysiology of Vivax Malaria.” *Trends in Parasitology* 25(5):220–27.

Arora, R., S. Bhatia, B. P. Mishra, and B. K. Joshi. 2011. “Population Structure in Indian Sheep Ascertained Using Microsatellite Information.” *Animal Genetics* 42(3):242–50.

Atdbio. 2005. “Next Generation Sequencing.” *http://www.atdbio.com/content/58/Next-generation-sequencing; Accessed 20 May 2017.*

Auburn, Sarah and Alyssa E. Barry. 2017. “Dissecting Malaria Biology and Epidemiology Using Population Genetics and Genomics.” *International Journal for Parasitology*.

Ba, Hampâté et al. 2016. “Widespread Distribution of Plasmodium Vivax Malaria in Mauritania on the Interface of the Maghreb and West Africa.” *Malaria Journal* 15(1):80. Retrieved (http://www.malariajournal.com/content/15/1/80).

Baird, J. Kevin. 2007. “Neglect of Plasmodium Vivax Malaria.” *Trends in Parasitology* 23(11):533–39.

Baird, J. Kevin. 2009. “Resistance to Therapies for Infection by Plasmodium Vivax.” *Clinical Microbiology Reviews* 22(3):508–34.

Baird, Kevin J., Jason D. Maguire, and Ric N. Price. 2012. “Diagnosis and Treatment of Plasmodium Vivax Malaria.” *Advances in Parasitology* 80:203–70.

Barnwell, John W., Paul Ingravallo, Mary R. Galinski, Yoshitsugo Matsumoto, and Masamichi Aikawa. 1990. “Plasmodium Vivax: Malarial Proteins Associated with the Membrane-Bound Caveola-Vesicle Complexes and Cytoplasmic Cleft Structures of Infected Erythrocytes.” *Experimental Parasitology* 70(1):85–99.

Beg, Masim A. et al. 2002. “Cerebral Involvement in Benign Tertian Malaria.” *American Journal of Tropical Medicine and Hygiene* 67(3):230–32.

Behjati, Sam and Patrick S. Tarpey. 2013. “What Is next Generation Sequencing?” *Archives of disease in childhood. Education and practice edition* 98(6):236–38. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/23986538%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3841808).

Bernabeu, M. et al. 2012. “Plasmodium Vivax Malaria in Mali: A Study from Three Different Regions.” *Malar J* 11:405.

Bernabeu, Maria et al. 2012. “Plasmodium Vivax Malaria in Mali: A Study from Three Different Regions.” *Malaria Journal* 11.

Bhatia, Gaurav, Nick Patterson, Sriram Sankararaman, and Alkes L. Price. 2013. “Estimating and Interpreting FST: The Impact of Rare Variants.” *Genome Research* 23(9):1514–21.

Bhatt, S. et al. 2015. “The Effect of Malaria Control on Plasmodium Falciparum in Africa between 2000 and 2015.” *Nature*.

Boillée, Séverine, Christine Vande Velde, and Don W. W. Cleveland. 2006. “ALS: A Disease of Motor Neurons and Their Nonneuronal Neighbors.” *Neuron* 52(1):39–59.

Bousema, J. T. et al. 2004. “Plasmodium Falciparum Gametocyte Carriage in Asymptomatic Children in Western Kenya.” *Malar J* 3:18.

Brega, S. et al. 2005. “Identification of the Plasmodium Vivax Mdr-like Gene (Pvmdr1) and Analysis of Single-Nucleotide Polymorphisms among Isolates from Different Areas of Endemicity.” *J Infect Dis* 191(2):272–77. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/15609238%5Cnhttp://jid.oxfordjournals.org/content/191/2/272.full.pdf).

C. de Oliveira, Thais et al. 2017. “Genome-Wide Diversity and Differentiation in New World Populations of the Human Malaria Parasite Plasmodium Vivax.” *PLoS Neglected Tropical Diseases* 11(7).

Carlton, Jane M., Aparup Das, and Ananias A. Escalante. 2013. “Genomics, Population Genetics and Evolutionary History of Plasmodium Vivax.” *Advances in Parasitology* 81:204–22.

Center for Disease Control and Prevention (CDC). 2017. “The History of Malaria, an Ancient Disease.” *https://www.cdc.gov/malaria/about/history/; Accessed 14 May 2017*.

Chang, Hsiao-Han et al. 2012. “Genomic Sequencing of Plasmodium Falciparum Malaria Parasites from Senegal Reveals the Demographic History of the Population.” *Molecular Biology and Evolution* 29(11):3427–39. Retrieved (https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss161).

Collins, William E. and Geoffrey M. Jeffery. 2007. “Plasmodium Malariae: Parasite and Disease.” *Clinical Microbiology Reviews* 20(4):579–92.

Cowell, Annie N. et al. 2017. “Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium Vivax from Unprocessed Clinical Samples.” *mBio* 8(1).

Day, K. P. and K. Marsh. 1991. “Naturally Acquired Immunity to Plasmodium Falciparum.” *Immunol Today* 12(3):A68-71. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/2069680).

Diawara, Fatou et al. 2017. “Measuring the Impact of Seasonal Malaria Chemoprevention as Part of Routine Malaria Control in Kita, Mali.” *Malaria Journal* 16(1):325. Retrieved (http://malariajournal.biomedcentral.com/articles/10.1186/s12936-017-1974-x).

Djimdé, Abdoulaye A. et al. 2008. “Efficacy, Safety, and Selection of Molecular Markers of Drug Resistance by Two ACTs in Mali.” *American Journal of Tropical Medicine and Hygiene* 78(3):455–61.

Doolan, Denise L., Carlota Dobaño, and J. Kevin Baird. 2009. “Acquired Immunity to Malaria.” *Clinical Microbiology Reviews*.

Escalante, A. A. et al. 2005. “A Monkey’s Tale: The Origin of Plasmodium Vivax as a Human Malaria Parasite.” *Proceedings of the National Academy of Sciences* 102(6):1980–85. Retrieved (http://www.pnas.org/cgi/doi/10.1073/pnas.0409652102).

Fernández-Becerra, Carmen et al. 2009. “Increased Expression Levels of the Pvcrt-o and Pvmdr1 Genes in a Patient with Severe Plasmodium Vivax Malaria.” *Malaria Journal*.

Fidock, David A. et al. 2000. “Mutations in the P. Falciparum Digestive Vacuole Transmembrane Protein PfCRT and Evidence for Their Role in Chloroquine Resistance.” *Molecular Cell* 6(4):861–71.

Furusawa, Mitsuru. 2012. “Implications of Fidelity Difference between the Leading and the Lagging Strand of DNA for the Acceleration of Evolution.” *Frontiers in Oncology* 2. Retrieved (http://journal.frontiersin.org/article/10.3389/fonc.2012.00144/abstract).

Genton, Blaise et al. 2008. “Plasmodium Vivax and Mixed Infections Are Associated with Severe Malaria in Children: A Prospective Cohort Study from Papua New Guinea.” *PLoS Medicine*.

Gething, Peter W. et al. 2012. “A Long Neglected World Malaria Map: Plasmodium Vivax Endemicity in 2010.” *PLoS Neglected Tropical Diseases* 6(9).

Golassa, Lemu, Berhanu Erko, Frederick N. Baliraine, Abraham Aseffa, and Göte Swedberg. 2015. “Polymorphisms in Chloroquine Resistance-Associated Genes in Plasmodium Vivax in Ethiopia.” *Malaria Journal* 14(1):164. Retrieved (http://www.malariajournal.com/content/14/1/164).

Greenwood, Brian M. et al. 2008. “Malaria: Progress, Perils, and Prospects for Eradication.” *Journal of Clinical Investigation* 118(4):1266–76.

Grigg, Matthew J. et al. 2016. “Efficacy of Artesunate-Mefloquine for Chloroquine-Resistant Plasmodium Vivax Malaria in Malaysia: An Open-Label, Randomized, Controlled Trial.” *Clinical Infectious Diseases*.

Hemingway, Janet et al. 2016. “Tools and Strategies for Malaria Control and Elimination: What Do We Need to Achieve a Grand Convergence in Malaria?” *PLoS Biology*.

Horuk, R. et al. 1993. “A Receptor for the Malarial Parasite Plasmodium Vivax: The Erythrocyte Chemokine Receptor.” *Science*.

Howes, Rosalind E. et al. 2015. “Plasmodium Vivax Transmission in Africa.” *PLoS Neglected Tropical Diseases* 9(11).

HudsonAlpha, Genomic Servives Lab at. 2010. “Bam2fastq.” *https://gsl.hudsonalpha.org/information/software/bam2fastq; Accessed 12 February 2018*.

Hupalo, Daniel N. et al. 2016. “Population Genomics Studies Identify Signatures of Global Dispersal and Drug Resistance in Plasmodium Vivax.” *Nature Genetics* 48(8):953–58.

Hviid, Lars. 2005. “Naturally Acquired Immunity to Plasmodium Falciparum Malaria in Africa.” *Acta Tropica* 95(3):270–75.

IRIN. 2007. “Beating Malaria Achievable This Year Gov’t Says.” *http://www.irinnews.org/report/70594/mali-beating-malaria-achievable-year-govt-says*.

Ishengoma, D. R. S. et al. 2009. “Health Laboratories in the Tanga Region of Tanzania: The Quality of Diagnostic Services for Malaria and Other Communicable Diseases.” *Annals of tropical medicine and parasitology* 103(5):441–53. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/19583914).

Kayentao, Kassoum et al. 2013. “Intermittent Preventive Therapy for Malaria during Pregnancy Using 2 vs 3 or More Doses of Sulfadoxine-Pyrimethamine and Risk of Low Birth Weight in Africa: Systematic Review and Meta-Analysis.” *JAMA : the journal of the American Medical Association* 309:594–604. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/23403684).

Kochar, Dhanpat Kumar et al. 2010. “Clinical Features of Children Hospitalized with Malaria--a Study from Bikaner, Northwest India.” *The American journal of tropical medicine and hygiene* 83(5):981–89. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2963956&tool=pmcentrez&rendertype=abstract).

Koita, Ousmane A. et al. 2012. “Effect of Seasonality and Ecological Factors on the

Prevalence of the Four Malaria Parasite Species in Northern Mali.” *Journal of Tropical Medicine*.

Krotoski WA. 1985. “Discovery of the Hypnozoite and a New Theory of Malarial Relapse.” *Trans R Soc Trop Med Hyg* 79: 1–11.

Krotoski, Wojciech 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.

Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. 2016. “MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.” *Molecular biology and evolution* 33(7):1870–74.

Kute, V. B. et al. 2012. “Outcome and Prognostic Factors of Malaria-Associated Acute Kidney Injury Requiring Hemodialysis: A Single Center Experience.” *Indian journal of nephrology* 22(1):33–38. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/22279340%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3263060).

Kute, Vivek B. et al. 2012. “Plasmodium Vivax Malaria-Associated Acute Kidney Injury, India, 2010-2011.” *Emerging Infectious Diseases* 18(5):842–45.

Kwiatkowski, D. P. P. 2005. “How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria.” *Am J Hum Genet* 77(2):171–90. Retrieved (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16001361).

Lacerda, Marcus V. G. et al. 2012. “Postmortem Characterization of Patients with Clinical Diagnosis of Plasmodium Vivax Malaria: To What Extent Does This Parasite Kill?” *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 55(8):e67-74. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/22772803).

Lança, Ellen Fátima Caetano et al. 2012. “Risk Factors and Characterization of Plasmodium Vivax-Associated Admissions to Pediatric Intensive Care Units in the Brazilian Amazon.” *PLoS ONE* 7(4).

Langhorne, Jean, Francis M. M. Ndungu, Anne-Marit Sponaas, and Kevin Marsh. 2008. “Immunity to Malaria: More Questions than Answers.” *Nature immunology* 9(7):725–32. Retrieved (http://www.nature.com/ni/journal/v9/n7/full/ni.f.205.html).

Langmead, Ben and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9(4):357–59.

Leang, Rithea et al. 2013. “Efficacy of Dihydroartemisinin-Piperaquine for Treatment of Uncomplicated Plasmodium Falciparum and Plasmodium Vivax in Cambodia, 2008 to 2010.” *Antimicrobial Agents and Chemotherapy*.

Leslie, Toby et al. 2012. “Overdiagnosis and Mistreatment of Malaria among Febrile Patients at Primary Healthcare Level in Afghanistan: Observational Study.” *BMJ (Clinical research ed.)* 345(jul24_2):e4389. Retrieved (http://www.bmj.com/content/345/bmj.e4389.full).

Li, Heng et al. 2009. “The Sequence Alignment / Map Format and SAMtools.” *Bioinformatics* 25(16):2078–79.

Li, Miaoxin et al. 2017. “Robust and Rapid Algorithms Facilitate Large-Scale Whole Genome Sequencing Downstream Analysis in an Integrative Framework.” *Nucleic Acids Research* 45(9).

Liu, Weimin et al. 2014. “African Origin of the Malaria Parasite Plasmodium Vivax.” *Nature Communications* 5.

López-Antuñano, Francisco Javier. 1999. “Is Primaquine Useful and Safe as True Exo-Erythrocytic Merontocidal, Hypnozoitocidal and Gametocidal Antimalarial Drug?” *Salud Publica de Mexico* 41(5):410–19.

Malaria Vaccine Initiative. 2016. “Fact Sheet: RTS,S Malaria Vaccine Candidate (Mosquirix^{TM}).” *http://www.malariavaccine.org/files/content/page/files/RTSS%20vaccine%20candidate%20Factsheet_FINAL.pdf.*

Malhotra, Indu et al. 2006. “Umbilical Cord-Blood Infections with Plasmodium Falciparum Malaria Are Acquired Antenatally in Kenya.” *The Journal of infectious diseases* 194(2):176–83. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/16779723).

Marsh, K. and R. J. Howard. 1986. “Antigens Induced on Erythrocytes by P. Falciparum: Expression of Diverse and Conserved Determinants.” *American Journal of Psychiatry* 231(10):150–53.

Marsh, K. and S. Kinyanjui. 2006. “Immune Effector Mechanisms in Malaria.” *Parasite Immunology* 28(1–2):51–60.

Mbenda, Huguette Gaelle Ngassa, Inocent Gouado, and Aparup Das. 2016. “An Additional Observation of Plasmodium Vivax Malaria Infection in Duffy-Negative Individuals from Cameroon.” *Journal of Infection in Developing Countries* 10(6):682–86.

McGready, R. 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(5):388–96.

Medecine for Malaria Venture. 2017. “Parasite Lifecycle.” *https://www.mmv.org/malaria-medicines/parasite-lifecycle; Accessed 4th December 2017.*

Medicines for Malaria Venture. 2017. “Definitions and Symptoms.” *https://www.mmv.org/malaria-medicines/definitions-and-symptoms; Accessed 15 May 2017.*

Medicines for Malaria Venture (MMV). 2016. “Five Eprecies.” *https://www.mmv.org/malaria-medicines/five-species. Accessed 5 April 2017.*

Melo, Gisely C. et al. 2014. “Expression Levels of Pvcrt-o and Pvmdr-1 Are Associated with Chloroquine Resistance and Severe Plasmodium Vivax Malaria in Patients of the Brazilian Amazon.” *PLoS ONE*.

Mens, P. et al. 2007. “Is Molecular Biology the Best Alternative for Diagnosis of Malaria to Microscopy? A Comparison between Microscopy, Antigen Detection and Molecular Tests in Rural Kenya and Urban Tanzania.” *Tropical Medicine and International Health* 12(2):238–44.

Mens, Petra F., Gerard J. Schoone, Piet A. Kager, and Henk D. F. H. Schallig. 2006. “Detection and Identification of Human Plasmodium Species with Real-Time Quantitative Nucleic Acid Sequence-Based Amplification.” *Malaria journal* 5:80. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/17018138%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1592503).

Merriam-Webster. 2015. “Malaria.” *https://www.merriam-webster.com/dictionary/malaria. Accessed 5 April 2015.*

Miller, L. H., S. J. Mason, D. F. Clyde, and M. H. McGinniss. 1976. “The Resistance Factor to Plasmodium Vivax in Blacks. The Duffy-Blood-Group Genotype, FyFy.” *The New England journal of medicine* 295(6):302–4. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/778616).

Minde, David P., Zeinab Anvarian, Stefan G. D. Rüdiger, and Madelon M. Maurice. 2011. “Messing up Disorder: How Do Missense Mutations in the Tumor Suppressor Protein APC Lead to Cancer?” *Molecular Cancer* 10.

Ministère de la Santé du Mali. 2007. *Plan d’accélération Des Activités de Lutte Contre Le Paludisme, Ministère de La Santé, République Du Mali, Avril 2007.*

Ministère de la Santé et de l’Hygiène Publique. 2015. “COMMUNIQUÉ DU

GOUVERNEMENT SUR LA SITUATION D'URGENCE DU PALUDISME AU MALI." *http://www.sante.gov.ml/index.php/actualites/communiques/item/2543-communique-du-gouvernement-sur-la-situation-d-urgence-du-paludisme-au-mali; Accessed 16 June 2017.*

Mint lekweiry, Khadijetou et al. 2012. "Molecular Surveillance of Drug-Resistant Plasmodium Vivax Using Pvdhfr, Pvdhps and Pvmdr1 Markers in Nouakchott, Mauritania." *Journal of Antimicrobial Chemotherapy* 67(2):367–74.

Mizutani, Masanori et al. 2016. "Development of a Plasmodium Berghei Transgenic Parasite Expressing the Full-Length Plasmodium Vivax Circumsporozoite VK247 Protein for Testing Vaccine Efficacy in a Murine Model." *Malaria Journal* 15(1):251. Retrieved (http://malariajournal.biomedcentral.com/articles/10.1186/s12936-016-1297-3).

Mobegi, Victor A. et al. 2014. "Genome-Wide Analysis of Selection on the Malaria Parasite Plasmodium Falciparum in West African Populations of Differing Infection Endemicity." *Molecular Biology and Evolution* 31(6):1490–99.

Monif GRG, Baker DA. 2004. *Infectious Disease in Obstetrics and Gynecology. 6th Ed. New York: Parthenon; 2004. Pp. 280–286.*

Moody, A. H., P. L. Chiodini, and General Haematology Work-. 2000. "Methods for the Detection of Blood Parasites." *Ckin Lab Haem* 22:189–202.

Mouhamadou GUEYE, Bourama FANE, Sekou FOFANA. 2015. *EVALUATION RAPIDE DE LA POSSESSION ET DE L'UTILISATION DE LA MOUSTIQUAIRE DE LONGUE DUREE APRES LA CAMPAGNE DE DISTRIBUTION DANS LES REGIONS DE SIKASSO, DE SEGOU ET DE MOPTI.*

Mu, Jianbing et al. 2005. "Host Switch Leads to Emergence of Plasmodium Vivax Malaria in Humans." *Molecular Biology and Evolution* 22(8):1686–93.

Mzilahowa, Themba, Ian M. Hastings, Malcolm E. Molyneux, and Philip J. McCall. 2012. "Entomological Indices of Malaria Transmission in Chikhwawa District, Southern Malawi." *Malaria journal* 11(1):380. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3536595&tool=pmcentrez&rendertype=abstract).

Naing, Cho, Kyan Aung, Daw-Khin Win, and Mak Joon Wah. 2010. "Efficacy and Safety of Chloroquine for Treatment in Patients with Uncomplicated Plasmodium Vivax Infections in Endemic Countries." *Transactions of the Royal Society of Tropical Medicine and Hygiene*.

Niangaly, Amadou et al. 2017. "Plasmodium Vivax Infections over 3 Years in Duffy Blood Group Negative Malians in Bandiagara, Mali." *American Journal of Tropical Medicine and Hygiene* 97(3):744–52.

Nomura, T. et al. 2001. "Evidence for Different Mechanisms of Chloroquine Resistance in 2 Plasmodium Species That Cause Human Malaria." *The Journal of infectious diseases* 183(11):1653–61.

Novotny, Joe, Amandeep Singh, Lek Dysoley, Siv Sovannaroth, and Huy Rekol. 2016. "Evidence of Successful Malaria Case Management Policy Implementation in Cambodia: Results from National ACTwatch Outlet Surveys." *Malaria Journal*.

De Oliveira, Alexandre Macedo et al. 2009. "Performance of Malaria Rapid Diagnostic Tests as Part of Routine Malaria Case Management in Kenya." *American Journal of Tropical Medicine and Hygiene* 80(3):470–74.

Ould Ahmedou Salem, Mohamed et al. 2015. "Efficacy of Chloroquine for the Treatment of Plasmodium Vivax in the Saharan Zone in Mauritania." *Malaria Journal* 14(1):39. Retrieved (http://www.malariajournal.com/content/14/1/39).

Ould Ahmedou Salem, Mohamed Salem et al. 2015. "Increasing Prevalence of Plasmodium Vivax among Febrile Patients in Nouakchott, Mauritania." *American Journal of Tropical*

Medicine and Hygiene 92(3):537–40.

Ouldabdallahi Moukah, Mohamed et al. 2016. "Malaria in Three Epidemiological Strata in Mauritania." *Malaria Journal* 15(1):204. Retrieved (http://malariajournal.biomedcentral.com/articles/10.1186/s12936-016-1244-3).

Oyola, Samuel O. et al. 2013. "Efficient Depletion of Host DNA Contamination in Malaria Clinical Sequencing." *Journal of Clinical Microbiology* 51(3):745–51.

Phyo, Aung Pyae et al. 2011. "Dihydroartemisinin-Piperaquine versus Chloroquine in the Treatment of Plasmodium Vivax Malaria in Thailand: A Randomized Controlled Trial." *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*.

Picot, S. 2006. "[Is Plasmodium Vivax Still a Paradigm for Uncomplicated Malaria?]." *Médecine et maladies infectieuses* 36(8):406–13. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/16842954).

PMI. 2016. "President's Malaria Initiative, Fighting Malaria and Saving Lives." *https://www.pmi.gov/docs/default-source/default-document-library/country-profiles/mali_profile.pdf?sfvrsn=18; Available on July 07, 2016.*

Poespoprodjo, Jeanne R. et al. 2009. "Vivax Malaria: A Major Cause of Morbidity in Early Infancy." *Clinical Infectious Diseases*.

Prabhat Jha et al. 2016. "Genomic Epidemiology of Artemisinin Resistant Malaria." *eLife* 5.

Price, Ric N. et al. 2007. "Vivax Malaria: Neglected and Not Benign." *American Journal of Tropical Medicine and Hygiene* 77(SUPPL. 6):79–87.

Reyburn, Hugh et al. 2005. "Association of Transmission Intensity and Age with Clinical Manifestations and Case Fatality of Severe Plasmodium Falciparum Malaria." *JAMA : the journal of the American Medical Association* 293(12):1461–70.

Rieckmann, K. H., D. R. Davis, and D. C. Hutton. 1989. "PLASMODIUM VIVAX RESISTANCE TO CHLOROQUINE?" *The Lancet*.

Robinson, Leanne J. et al. 2015. "Strategies for Understanding and Reducing the Plasmodium Vivax and Plasmodium Ovale Hypnozoite Reservoir in Papua New Guinean Children: A Randomised Placebo-Controlled Trial and Mathematical Model." *PLoS Medicine*.

Russo, Gianluca et al. 2017. "Molecular Evidence of Plasmodium Vivax Infection in Duffy Negative Symptomatic Individuals from Dschang, West Cameroon." *Malaria Journal* 16(1):74. Retrieved (http://malariajournal.biomedcentral.com/articles/10.1186/s12936-017-1722-2).

Ryan, Jeffrey R. et al. 2006. "Evidence for Transmission of Plasmodium Vivax among a Duffy Antigen Negative Population in Western Kenya." *American Journal of Tropical Medicine and Hygiene* 75(4):575–81.

Sachs, Jeffrey and Pia Malaney. 2002. "The Economic and Social Burden of Malaria." *Nature* 415(6872):680–85.

Sagara, Issaka et al. 2006. "A Randomized Trial of Artesunate-Sulfamethoxypyrazine-Pyrimethamine versus Artemether-Lumefantrine for the Treatment of Uncomplicated Plasmodium Falciparum Malaria in Mali." *American Journal of Tropical Medicine and Hygiene* 75(4):630–36.

Saitou, N. and M. Nei. 1987. "The Neighbour-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Mol Biol Evo* 4(4):406–25.

De Santana Filho, Franklin Simoes et al. 2007. "Chloroquine-Resistant Plasmodium Vivax, Brazilian Amazon [8]." *Emerging Infectious Diseases*.

Save the Children. 2016. "Malaria Control in Schools in Mali: Results from a Cluster Randomized Control Trial in Sikasso Region, Mali." *http://www.savethechildren.org/atf/cf/%7B9def2ebe-10ae-432c-9bd0-df91d2eba74a%7D/MALARIA%20CONTROL%20IN%20SCHOOLS%20IN%20MALI%*

20EN%20LOW%20FINAL.PDF; Accessed 14 May 2017.

Schousboe, Mette L. et al. 2015. "Multiple Origins of Mutations in the Mdr1 Gene—A Putative Marker of Chloroquine Resistance in P. Vivax." *PLoS Neglected Tropical Diseases* 9(11).

Sinka, Marianne E. et al. 2012. "A Global Map of Dominant Malaria Vectors." *Parasites & vectors* 5(January 2008):69. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3349467&tool=pmcentrez&rendertype=abstract).

Siqueira, André M. et al. 2015. "Characterization of Plasmodium Vivax-Associated Admissions to Reference Hospitals in Brazil and India." *BMC medicine* 13:57.

Snounou G et al. 1993. "High Sensitivity of Detection of Human Malaria Parasites by the Use of Nested Polymerase Chain Reaction." *Mol Biochem Parasitol* 61(2):315–20. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/8479452).

Snow, Robert W., Carlos A. Guerra, Abdisalan M. Noor, Hla Y. Myint, and Simon I. Hay. 2005. "The Global Distribution of Clinical Episodes of Plasmodium Falciparum Malaria." *Nature*.

Sturm, A. 2006. "Manipulation of Host Hepatocytes by the Malaria Parasite for Delivery into Liver Sinusoids." *Science* 313(5791):1287–90. Retrieved (http://www.sciencemag.org/cgi/doi/10.1126/science.1129720).

Suwanarusk, Rossarin et al. 2004. "The Deformability of Red Blood Cells Parasitized by Plasmodium Falciparum and P. Vivax." *The Journal of infectious diseases* 189(2):190–94.

Svenson, J. E., J. D. MacLean, T. W. Gyorkos, and J. Keystone. 1995. "Imported Malaria. Clinical Presentation and Examination of Symptomatic Travelers." *Arch Intern Med* 155(8):861–68.

Tagbor, Harry et al. 2015. "A Non-Inferiority, Individually Randomized Trial of Intermittent Screening and Treatment versus Intermittent Preventive Treatment in the Control of Malaria in Pregnancy." *PLoS ONE* 10(8).

Talha, Albadawi Abdelbagi et al. 2015. "Molecular Genetic Analysis of Plasmodium Vivax Isolates from Eastern and Central Sudan Using Pvcsp and Pvmsp-3?? Genes as Molecular Markers." *Infection, Genetics and Evolution* 32:12–22.

Tamura, K., M. Nei, and S. Kumar. 2004. "Prospects for Inferring Very Large Phylogenies by Using the Neighbor-Joining Method." *Proceedings of the National Academy of Sciences* 101(30):11030–35. Retrieved (http://www.pnas.org/cgi/doi/10.1073/pnas.0404206101).

Tjitra, Emiliana et al. 2008. "Multidrug-Resistant Plasmodium Vivax Associated with Severe and Fatal Malaria: A Prospective Study in Papua, Indonesia." *PLoS Medicine*.

Vogel, G. 2013. "The Forgotten Malaria." *Science* 342(6159):684–87. Retrieved (http://www.sciencemag.org/cgi/doi/10.1126/science.342.6159.684).

White, Nicholas J. 2011. "Determinants of Relapse Periodicity in Plasmodium Vivax Malaria." *Malaria Journal*.

Wikipedia. 2017. "RTS,S." *https://en.wikipedia.org/wiki/RTS,S; Accessed 28 May 2017.*

Williams, John et al. 2016. "Non-Falciparum Malaria Infections in Pregnant Women in West Africa." *Malaria Journal* 15(1):53. Retrieved (http://www.malariajournal.com/content/15/1/53).

Wongsrichanalai, Chansuda, Mazie J. Barcus, Sinuon Muth, Awalludin Sutamihardja, and Walther H. Wernsdorfer. 2007. "A Review of Malaria Diagnostic Tools: Microscopy and Rapid Diagnostic Test (RDT)." *American Journal of Tropical Medicine and Hygiene* 77(SUPPL. 6):119–27.

World Health Organisation(WHO). 2010. "Guidelines for the Treatment of Malaria - 2nd Edition." *http://apps.who.int/medicinedocs/en/d/Js19105en/; Accessed 15 May 2017.*

World Health Organisation(WHO). 2012a. *Malaria Rapid Diagnostic Test Performance: Summary of Results of WHO Product Testing of Malaria RDTs, Round 1–4 (2008–2012) Geneva, Switzerland: WHO.*

World Health Organisation(WHO). 2012b. "MANAGEMENT OF SEVERE MALARIA; A PRACTICAL HANDBOOK Third Edition." *http://apps.who.int/iris/bitstream/10665/79317/1/9789241548526_eng.pdf; Accessed 20 November 2017.*

World Health Organisation(WHO). 2013. "World Malaria Report 2012." *http://www.who.int/malaria/publications/world_malaria_report_2012/en/; consulted on 17 March 2017.*

World Health Organisation(WHO). 2014. "Guidelines for the Treatment of Malaria. 2nd Edition. Geneva, Switzerland: WHO; 2010." *http://apps.who.int/medicinedocs/en/d/Js19105en/; Accessed 15 May 2017; Accessed 2nd November 2017.*

World Health Organisation(WHO). 2015a. "Confronting Plasmodium Vivax Malaria." *http://www.searo.who.int/entity/malaria/confronting-plasmodium.pdf?ua=1; Accessed 10 December 2017.*

World Health Organisation(WHO). 2015b. "Indoor Residual Spraying: An Operational Manual for IRS for Malaria Transmission, Control and Elimination. Second Edition." *http://www.who.int/malaria/publications/atoz/9789241508940/en/; Accessed 24 August 2017.*

World Health Organisation(WHO). 2016a. "Malaria, Fact Sheet, Updated April 2016." *http://www.who.int/mediacentre/factsheets/fs094/en/; Available on November 15, 2016.*

World Health Organisation(WHO). 2016b. "Malaria: Global Fund Proposal Development." *http://www.who.int/malaria/publications/atoz/malaria_gf_proposal_dev_who_policy_brief_201106.pdf; available on 01 July 2016.*

World Health Organisation(WHO). 2016c. "World Malaria Report 2015." *http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158_eng.pdf; available on December 17, 2016.*

World Health Organisation(WHO). 2017a. "Indonesia." *http://www.who.int/malaria/publications/country-profiles/profile_idn_en.pdf; Accessed 31 August 2018.*

World Health Organisation(WHO). 2017b. "Intermittent Preventive Treatment in Pregnancy (IPTp)." *http://www.who.int/malaria/areas/preventive_therapies/pregnancy/en/; Accessed 7 December 2017.*

World Health Organisation(WHO). 2017c. "Seasonal Malaria Chemoprevention (SMC)." *http://www.who.int/malaria/areas/preventive_therapies/children/en/; Accessed 8 December 2017.*

World Health Organisation(WHO). 2017d. "Tables of Malaria Vaccine Projects Globally." *http://www.who.int/immunization/research/development/Rainbow_tables/en/; Accessed 28 May 2017.*

World Health Organisation(WHO). 2017e. "World Malaria Report 2017." *http://www.who.int/malaria/publications/world-malaria-report-2017/en/; Accessed on 29 January 2018.*

Yohannes, Ambachew M., Awash Teklehaimanot, Yngve Bergqvist, and Pascal Ringwald. 2011. "Confirmed Vivax Resistance to Chloroquine and Effectiveness of Artemether-Lumefantrine for the Treatment of Vivax Malaria in Ethiopia." *American Journal of Tropical Medicine and Hygiene.*

YourGenome. 2017. "Why Was There a Race to Sequence the Human Genome?" *http://www.yourgenome.org/stories/why-was-there-a-race-to-sequence-the-human-*

genome; Accessed 16 May 2017.

Yu, Xiaoqing and Shuying Sun. 2013. “Comparing a Few SNP Calling Algorithms Using Low-Coverage Sequencing Data.” *BMC Bioinformatics* 14(1).

Annex

Annex 1: Script used to align the sequences with human reference genome

```
#PBS -N Aligment_vivax_hum
#PBS -q UCTlong
#PBS -l nodes=1:ppn=32:series600
DIR=/researchdata/fhgfs/djmmou001/moussa_data

# NB, PLEASE READ THIS!
# There is a 2:1 correspondence between RAM and cores on the 600 series.
# You need to know how much RAM your job will consume before submitting it.
# Please set the ppn value above to be 1/2 the GB of RAM required.  For
# example a job needing 10GB of RAM should have ppn=5

# Please leave the hostname command here for troubleshooting purposes hostname

# Your science stuff goes here:
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/human_reference.fasta $DIR/human-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/human-ref -1 $DIR/1_1.fastq -2 $DIR/1_2.fastq -S $DIR/kidal1.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/human-ref -1 $DIR/2_1.fastq.gz -2 $DIR/2_2.fastq.gz -S $DIR/kidal2.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/human-ref -1 $DIR/3_1.fastq.gz -2 $DIR/3_2.fastq.gz -S $DIR/kidal3.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/human-ref -1 $DIR/4_1.fastq.gz -2 $DIR/4_2.fastq.gz -S $DIR/kidal4.sam
```

Annex 2: Script used to compute the sequences mapping rate with human reference genome

```
#PBS -N mapping_rate_hum
#PBS -q UCTlong
#PBS -l nodes=1:ppn=32:series600
DIR=/researchdata/fhgfs/djmmou001/moussa_data/bamfiles

# NB, PLEASE READ THIS!
# There is a 2:1 correspondence between RAM and cores on the 600 series.
# You need to know how much RAM your job will consume before submitting it.
# Please set the ppn value above to be 1/2 the GB of RAM required.  For
# example a job needing 10GB of RAM should have ppn=5

# Please leave the hostname command here for troubleshooting purposes hostname

# Your science stuff goes here:
/opt/exp_soft/samtoools-1.3.1/samtools view -bT PvivaxP01_Genome.fasta kidal1.sam > kidal1.bam
/opt/exp_soft/samtoools-1.3.1/samtools sort kidal1.bam > stdkidal1.bam
/opt/exp_soft/samtoools-1.3.1/samtools index stdkidal1.bam
/opt/exp_soft/samtoools-1.3.1/samtools flagstat stdkidal1.bam
/opt/exp_soft/samtoools-1.3.1/samtools view -bT PvivaxP01_Genome.fasta kidal2.sam > kidal2.bam
/opt/exp_soft/samtoools-1.3.1/samtools sort kidal2.bam > stdkidal2.bam
/opt/exp_soft/samtoools-1.3.1/samtools index stdkidal2.bam
/opt/exp_soft/samtoools-1.3.1/samtools flagstat stdkidal2.bam
/opt/exp_soft/samtoools-1.3.1/samtools view -bT PvivaxP01_Genome.fasta kidal3.sam > kidal3.bam
/opt/exp_soft/samtoools-1.3.1/samtools sort kidal3.bam > stdkidal3.bam
/opt/exp_soft/samtoools-1.3.1/samtools index stdkidal3.bam
/opt/exp_soft/samtoools-1.3.1/samtools flagstat stdkidal3.bam
/opt/exp_soft/samtoools-1.3.1/samtools view -bT PvivaxP01_Genome.fasta kidal4.sam > kidal4.bam
/opt/exp_soft/samtoools-1.3.1/samtools sort kidal4.bam > stdkidal4.bam
```

```
/opt/exp_soft/samtoools-1.3.1/samtools index stdkidal4.bam
/opt/exp_soft/samtoools-1.3.1/samtools flagstat stdkidal4.bam
/opt/exp_soft/samtoools-1.3.1/samtools view -bT PvivaxP01_Genome.fasta kidal5.sam > kidal5.bam
/opt/exp_soft/samtoools-1.3.1/samtools sort kidal5.bam > stdkidal5.bam
/opt/exp_soft/samtoools-1.3.1/samtools index stdkidal5.bam
/opt/exp_soft/samtoools-1.3.1/samtools flagstat stdkidal5.bam
/opt/exp_soft/samtoools-1.3.1/samtools view -bT PvivaxP01_Genome.fasta Anefis.sam > Anefis.bam
/opt/exp_soft/samtoools-1.3.1/samtools sort Anefis.bam > stdAnefis.bam
/opt/exp_soft/samtoools-1.3.1/samtools index stdAnefis.bam
/opt/exp_soft/samtoools-1.3.1/samtools flagstat stdAnefis.bam
/opt/exp_soft/samtoools-1.3.1/samtools view -bT PvivaxP01_Genome.fasta TinEssako.sam > TinEssako.bam
/opt/exp_soft/samtoools-1.3.1/samtools sort TinEssako.bam > TinEssako.bam
/opt/exp_soft/samtoools-1.3.1/samtools index stdTinEssako.bam
/opt/exp_soft/samtoools-1.3.1/samtools flagstat stdTinEssako.bam
```

Annex 3: Script used to extract unmapped reads to human genome

```
#!/bin/sh

# Extraction of unmapped reads from bam files

bam2fastq -o Anefis_filtered_R#.fastq --no-aligned Anefis.bam

bam2fastq -o Kidal1_filtered_R#.fastq --no-aligned kidal1.bam

bam2fastq -o Kidal2_filtered_R#.fastq --no-aligned kidal2.bam

bam2fastq -o Kidal3_filtered_R#.fastq --no-aligned kidal3.bam

bam2fastq -o Kidal4_filtered_R#.fastq --no-aligned kidal4.bam

bam2fastq -o Kidal5_filtered_R#.fastq --no-aligned kidal5.bam

bam2fastq -o TinEssako_filtered_R#.fastq --no-aligned TinEssako.bam
```

Annex 4: Script used to align the sequences with *P. vivax* reference genome

```
#!/bin/sh

# Alignment of sequences with P. vivax reference genome using SMALT

smalt index -k 14 -s 8 vivax_ref PlasmoDB-35_PvivaxP01_Genome.fasta

smalt map -o Anefis_map.sam vivax_ref Anefis_filtered_R_1.fastq Anefis_filtered_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Anefis_map.sam >
Anefis_map.bam
samtools sort Anefis_map.bam > Anefis_mapStd.bam
samtools index Anefis_mapStd.bam
samtools flagstat Anefis_mapStd.bam

smalt map -o Kidal1_map.sam vivax_ref Kidal1_filtered_R_1.fastq
Kidal1_filtered_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal1_map.sam >
Kidal1_map.bam
samtools sort Kidal1_map.bam > Kidal1_mapStd.bam
samtools index Kidal1_mapStd.bam
samtools flagstat Kidal1_mapStd.bam

smalt map -o Kidal2_map.sam vivax_ref Kidal2_filtered_R_1.fastq
Kidal2_filtered_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal2_map.sam >
Kidal2_map.bam
samtools sort Kidal2_map.bam > Kidal2_mapStd.bam
samtools index Kidal2_mapStd.bam
samtools flagstat Kidal2_mapStd.bam

smalt map -o Kidal3_map.sam vivax_ref Kidal3_filtered_R_1.fastq
Kidal3_filtered_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal3_map.sam >
```

```
Kidal3_map.bam
samtools sort Kidal3_map.bam > Kidal3_mapStd.bam
samtools index Kidal3_mapStd.bam
samtools flagstat Kidal3_mapStd.bam

smalt map -o Kidal4_map.sam vivax_ref Kidal4_filtered_R_1.fastq
Kidal4_filtered_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal4_map.sam >
Kidal4_map.bam
samtools sort Kidal4_map.bam > Kidal4_mapStd.bam
samtools index Kidal4_mapStd.bam
samtools flagstat Kidal4_mapStd.bam

smalt map -o Kidal5_map.sam vivax_ref Kidal5_filtered_R_1.fastq
Kidal5_filtered_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal5_map.sam >
Kidal5_map.bam
samtools sort Kidal5_map.bam > Kidal5_mapStd.bam
samtools index Kidal5_mapStd.bam
samtools flagstat Kidal5_mapStd.bam

smalt map -o TinEssako_map.sam vivax_ref TinEssako_filtered_R_1.fastq
TinEssako_filtered_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta TinEssako_map.sam >
TinEssako_map.bam
samtools sort TinEssako_map.bam > TinEssako_mapStd.bam
samtools index TinEssako_mapStd.bam
samtools flagstat TinEssako_mapStd.bam
```

Annex 5: Extraction of reads mapping with *P. vivax* reference genome

```
#!/bin/sh

# Extraction of mapped reads from bam file

# And then, extract the reads aligned with P. vivax

bam2fastq -o Anefis_map-vivax_R#.fastq --no-unaligned Anefis_map.bam

bam2fastq -o Kidal1_map-vivax_R#.fastq --no-unaligned Kidal1_map.bam

bam2fastq -o Kidal2_map-vivax_R#.fastq --no-unaligned Kidal2_map.bam

bam2fastq -o Kidal3_map-vivax_R#.fastq --no-unaligned Kidal3_map.bam

bam2fastq -o Kidal4_map-vivax_R#.fastq --no-unaligned Kidal4_map.bam

bam2fastq -o Kidal5_map-vivax_R#.fastq --no-unaligned Kidal5_map.bam

bam2fastq -o TinEssako_map-vivax_R#.fastq --no-unaligned TinEssako_map.bam
```

Annex 6: Script used to re-align the reads extracted with *P. vivax* reference genome

```
#!/bin/sh

# Alignment of sequences with P. vivax reference genome using SMALT

smalt map -o Anefis_smalt.sam smalt_vivax Anefis_map-vivax_R_1.fastq Anefis_map-vivax_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Anefis_smalt.sam > Anefis_smalt.bam
samtools sort Anefis_smalt.bam > Anefis_Stdsmalt.bam
samtools index Anefis_Stdsmalt.bam
samtools flagstat Anefis_Stdsmalt.bam

smalt map -o Kidal1_smalt.sam smalt_vivax Kidal1_map-vivax_R_1.fastq Kidal1_map-vivax_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal1_smalt.sam > Kidal1_smalt.bam
samtools sort Kidal1_smalt.bam > Kidal1_Stdsmalt.bam
samtools index Kidal1_Stdsmalt.bam
samtools flagstat Kidal1_Stdsmalt.bam

smalt map -o Kidal2_smalt.sam smalt_vivax Kidal2_map-vivax_R_1.fastq Kidal2_map-vivax_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal2_smalt.sam > Kidal2_smalt.bam
samtools sort Kidal2_smalt.bam > Kidal2_Stdsmalt.bam
samtools index Kidal2_Stdsmalt.bam
samtools flagstat Kidal2_Stdsmalt.bam

smalt map -o Kidal3_smalt.sam smalt_vivax Kidal3_map-vivax_R_1.fastq Kidal3_map-vivax_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal3_smalt.sam > Kidal3_smalt.bam
```

```
samtools sort Kidal3_smalt.bam > Kidal3_Stdsmalt.bam
samtools index Kidal3_Stdsmalt.bam
samtools flagstat Kidal3_Stdsmalt.bam

smalt map -o Kidal4_smalt.sam smalt_vivax Kidal4_map-vivax_R_1.fastq Kidal4_map-vivax_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal4_smalt.sam > Kidal4_smalt.bam
samtools sort Kidal4_smalt.bam > Kidal4_Stdsmalt.bam
samtools index Kidal4_Stdsmalt.bam
samtools flagstat Kidal4_Stdsmalt.bam

smalt map -o Kidal5_smalt.sam smalt_vivax Kidal5_map-vivax_R_1.fastq Kidal5_map-vivax_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta Kidal5_smalt.sam > Kidal5_smalt.bam
samtools sort Kidal5_smalt.bam > Kidal5_Stdsmalt.bam
samtools index Kidal5_Stdsmalt.bam
samtools flagstat Kidal5_Stdsmalt.bam

smalt map -o TinEssako_smalt.sam smalt_vivax TinEssako_map-vivax_R_1.fastq TinEssako_map-vivax_R_2.fastq
samtools view -bT PlasmoDB-35_PvivaxP01_Genome.fasta TinEssako_smalt.sam > TinEssako_smalt.bam
samtools sort TinEssako_smalt.bam > TinEssako_Stdsmalt.bam
samtools index TinEssako_Stdsmalt.bam
samtools flagstat TinEssako_Stdsmalt.bam
```

Annex 7: Script used for variant calling

```
#!/bin/sh

# Variant calling

samtools index PlasmoDB-35_PvivaxP01_Genome.fasta
samtools index Anefis_Stdsmalt.bam
samtools index Kidal1_Stdsmalt.bam
samtools index Kidal2_Stdsmalt.bam
samtools index Kidal3_Stdsmalt.bam
samtools index Kidal4_Stdsmalt.bam
samtools index TinEssako_Stdsmalt.bam

samtools mpileup -u -f PlasmoDB-35_PvivaxP01_Genome.fasta Anefis_Stdsmalt.bam
Kidal1_Stdsmalt.bam Kidal2_Stdsmalt.bam Kidal3_Stdsmalt.bam Kidal4_Stdsmalt.bam
TinEssako_Stdsmalt.bam > mali_vivax.bcf
bcftools view -v -c -g mali_vivax.bcf > mali_vivax.vcf
```

Annex 8: Use of plink for association study and principal component analysis

```
#!/bin/sh

# Creation of ped bed fam files using plink
/Users/mdjimde/Documents/plink_mac/plink --vcf mali_vivax.vcf --snps-only --biallelic-only strict --allow-extra-chr --keep-allele-order --make-bed --out vivax2

# Association study
/Users/mdjimde/Documents/plink_mac/plink --bfile Mali --assoc --allow-extra-chr --allow-no-sex --out Mali_assoc

# Generating pca file
/Users/mdjimde/Documents/plink_mac/plink --bfile Mali --pca --allow-extra-chr --out Mali_pca
```

Annex 9: Script used to plot PCA in R

```
#!/usr/bin/Rscript

pdf("PCAplot.pdf",width=5, height=4.5)

# PCA 2 Dimensions

# Work directory setting
setwd("~/Desktop/Folders/MSc_UCT/P_vivax/Seq_Data/Mali_vivax/")

# We can visualise files present in the work directory
dir()

# Creating the pdf file for result storing
pdf("PCAplot_vivax.pdf", width=5, height=3)

# PCA computing
pca_model <- read.table("snp_mali_update.pca.evec")

names(pca_model) <- c("SampleID", "PC1", "PC2", "PC3","PC4",
                 "PC5", "vivax_Strain")

df <- pca_model[, 2:6]

library(ggfortify)

autoplot(prcomp(df))

autoplot(prcomp(df), data = pca_model, colour = 'vivax_Strain')
```

Annex 10: Script used to filter contamination from vivax sequences from Mauritania, align with vivax reference genome and variant calling

```
#PBS -N Aligment_2_Variants
#PBS -q UCTlong
#PBS -l nodes=1:ppn=32:series600
DIR=/researchdata/fhgfs/djmmou001/Vivax_data

# NB, PLEASE READ THIS!
# There is a 2:1 correspondence between RAM and cores on the 600 series.
# You need to know how much RAM your job will consume before submitting it.
# Please set the ppn value above to be 1/2 the GB of RAM required.  For
# example a job needing 10GB of RAM should have ppn=5

# Please leave the hostname command here for troubleshooting purposes hostname

# Your science stuff goes here:
# Your science stuff goes here:
# First step: alignment with human reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/entry.fasta $DIR/Human-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332408_1.fastq.gz
-2 $DIR/SRR332408_2.fastq.gz -S $DIR/Maurit-H$
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332409_1.fastq.gz
-2 $DIR/SRR332409_2.fastq.gz -S $DIR/Maurit-H$
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332410_1.fastq.gz
-2 $DIR/SRR332410_2.fastq.gz -S $DIR/Maurit-H$
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332411_1.fastq.gz
-2 $DIR/SRR332411_2.fastq.gz -S $DIR/Maurit-H$
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332412_1.fastq.gz
-2 $DIR/SRR332412_2.fastq.gz -S $DIR/Maurit-H$
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332413_1.fastq.gz
-2 $DIR/SRR332413_2.fastq.gz -S $DIR/Maurit-H$
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332414_1.fastq.gz
-2 $DIR/SRR332414_2.fastq.gz -S $DIR/Maurit-H$
```

```
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332415_1.fastq.gz
-2 $DIR/SRR332415_2.fastq.gz -S $DIR/Maurit-H$

# Second step: generate bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332408.sam -o $DIR/Maurit-Hum_SRR332408.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332409.sam -o $DIR/Maurit-Hum_SRR332409.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332410.sam -o $DIR/Maurit-Hum_SRR332410.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332411.sam -o $DIR/Maurit-Hum_SRR332411.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332412.sam -o $DIR/Maurit-Hum_SRR332412.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332413.sam -o $DIR/Maurit-Hum_SRR332413.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332414.sam -o $DIR/Maurit-Hum_SRR332414.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Maurit-
Hum_SRR332415.sam -o $DIR/Maurit-Hum_SRR332415.bam

# Third step: filtering human DNA
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered08_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332408.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered09_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332409.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered10_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332410.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered11_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332411.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered12_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332412.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered13_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332413.bam
```

```
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered14_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332414.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt_filtered15_R#.fastq --no-aligned
$DIR/Maurit-Hum_SRR332415.bam

# Fourth step: align with vivax reference genome the read unaligned with human genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered08_R_1.fastq -2 $DIR/Maurt_filtered08_R_2.fastq -S
$DIR/Maurt_filtered08.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered09_R_1.fastq -2 $DIR/Maurt_filtered09_R_2.fastq -S
$DIR/Maurt_filtered09.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered10_R_1.fastq -2 $DIR/Maurt_filtered10_R_2.fastq -S
$DIR/Maurt_filtered10.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered11_R_1.fastq -2 $DIR/Maurt_filtered11_R_2.fastq -S
$DIR/Maurt_filtered11.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered12_R_1.fastq -2 $DIR/Maurt_filtered12_R_2.fastq -S
$DIR/Maurt_filtered12.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered13_R_1.fastq -2 $DIR/Maurt_filtered13_R_2.fastq -S
$DIR/Maurt_filtered13.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered14_R_1.fastq -2 $DIR/Maurt_filtered14_R_2.fastq -S
$DIR/Maurt_filtered14.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Maurt_filtered15_R_1.fastq -2 $DIR/Maurt_filtered15_R_2.fastq -S
$DIR/Maurt_filtered15.sam

# Fifth step: generate filtered bam files
```

```
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered08.sam -o $DIR/Maurt_filtered08.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered09.sam -o $DIR/Maurt_filtered09.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered10.sam -o $DIR/Maurt_filtered10.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered11.sam -o $DIR/Maurt_filtered11.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered12.sam -o $DIR/Maurt_filtered12.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered13.sam -o $DIR/Maurt_filtered13.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered14.sam -o $DIR/Maurt_filtered14.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Maurt_filtered15.sam -o $DIR/Maurt_filtered15.bam

# Sixth step: extract reads mapping with vivax reference genome
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt-vivax08_R#.fastq --no-unaligned $DIR/Maurt_filtered08.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt-vivax09_R#.fastq --no-unaligned $DIR/Maurt_filtered09.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt-vivax10_R#.fastq --no-unaligned $DIR/Maurt_filtered10.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt-vivax11_R#.fastq --no-unaligned $DIR/Maurt_filtered11.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt-vivax12_R#.fastq --no-unaligned $DIR/Maurt_filtered12.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt-vivax13_R#.fastq --no-unaligned $DIR/Maurt_filtered13.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Maurt-vivax14_R#.fastq --no-unaligned $DIR/Maurt_filtered14.bam

# Seventh step: align again with vivax reference genome
```

```
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax08_R_1.fastq -2 $DIR/Maurt-vivax08_R_2.fastq -S $DIR/Maurit08.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax09_R_1.fastq -2 $DIR/Maurt-vivax09_R_2.fastq -S $DIR/Maurit09.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax10_R_1.fastq -2 $DIR/Maurt-vivax10_R_2.fastq -S $DIR/Maurit10.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax11_R_1.fastq -2 $DIR/Maurt-vivax11_R_2.fastq -S $DIR/Maurit11.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax12_R_1.fastq -2 $DIR/Maurt-vivax12_R_2.fastq -S $DIR/Maurit12.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax13_R_1.fastq -2 $DIR/Maurt-vivax13_R_2.fastq -S $DIR/Maurit13.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax14_R_1.fastq -2 $DIR/Maurt-vivax14_R_2.fastq -S $DIR/Maurit14.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Maurt-
vivax15_R_1.fastq -2 $DIR/Maurt-vivax15_R_2.fastq -S $DIR/Maurit15.sam

# Eighth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit08.sam -o $DIR/Maurit08.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit09.sam -o $DIR/Maurit09.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit10.sam -o $DIR/Maurit10.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit11.sam -o $DIR/Maurit11.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit12.sam -o $DIR/Maurit12.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit13.sam -o $DIR/Maurit13.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit14.sam -o $DIR/Maurit14.bam
```

```
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Maurit15.sam -o $DIR/Maurit15.bam

# Ninth step: Sort the bam files
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Maurit08.bam $DIR/Maurit08_Std
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Maurit09.bam $DIR/Maurit09_Std
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Maurit10.bam $DIR/Maurit10_Std
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Maurit11.bam $DIR/Maurit11_Std
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Maurit12.bam $DIR/Maurit12_Std
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Maurit13.bam $DIR/Maurit13_Std
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Maurit14.bam $DIR/Maurit14_Std
```

Annex 11: Script used to filter contamination from Madagascar sequences, align with vivax reference genome and variant calling

```
#PBS -N Aligment_2_Variants
#PBS -q UCTlong
#PBS -l nodes=1:ppn=32:series600
DIR=/researchdata/fhgfs/djmmou001/Vivax_data

# NB, PLEASE READ THIS!
# There is a 2:1 correspondence between RAM and cores on the 600 series.
# You need to know how much RAM your job will consume before submitting it.
# Please set the ppn value above to be 1/2 the GB of RAM required.  For
# example a job needing 10GB of RAM should have ppn=5

# Please leave the hostname command here for troubleshooting purposes hostname

# Your science stuff goes here:
# Your science stuff goes here:
# First step: alignment with human reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/entry.fasta $DIR/Human-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR570031_1.fastq.gz
-2 $DIR/SRR570031_2.fastq.gz -S $DIR/Madag-Hum_SRR570031.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR572651_1.fastq.gz
-2 $DIR/SRR572651_2.fastq.gz -S $DIR/Madag-Hum_SRR572651.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR828416_1.fastq.gz
-2 $DIR/SRR828416_2.fastq.gz -S $DIR/Madag-Hum_SRR828416.sam

# Second step: generate bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Madag-
Hum_SRR570031.sam -o $DIR/Madag-Hum_SRR570031.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Madag-
Hum_SRR572651.sam -o $DIR/Madag-Hum_SRR572651.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Madag-
Hum_SRR828416.sam -o $DIR/Madag-Hum_SRR828416.bam
```

```
# Third step: filtering human DNA
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Madag_filtered31_R#.fastq --no-aligned
$DIR/Madag-Hum_SRR570031.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Madag_filtered51_R#.fastq --no-aligned
$DIR/Madag-Hum_SRR572651.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Madag_filtered16_R#.fastq --no-aligned
$DIR/Madag-Hum_SRR828416.bam

# Fourth step: align with vivax reference genome the read unaligned with human genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
$DIR/Madag_filtered16_R_1.fastq -2 $DIR/Madag_filtered16_R_2.fastq -S
$DIR/Madag16_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Madag_filtered31_R_1.fastq -2 $DIR/Madag_filtered31_R_2.fastq -S
$DIR/Madag31_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Madag_filtered51_R_1.fastq -2 $DIR/Madag_filtered51_R_2.fastq -S
$DIR/Madag51_filt.sam

# Fifth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Madag16_filt.sam -o $DIR/Madag16_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Madag31_filt.sam -o $DIR/Madag31_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Madag51_filt.sam -o $DIR/Madag51_filt.bam

# Sixth step: extract reads mapping with vivax reference genome
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Madag16-vivax_R#.fastq --no-unaligned
$DIR/Madag16_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Madag31-vivax_R#.fastq --no-unaligned
$DIR/Madag31_filt.bam
```

```
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Madag51-vivax_R#.fastq --no-unaligned $DIR/Madag51_filt.bam

# Seventh step: align again with vivax reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Madag16-vivax_R_1.fastq -2 $DIR/Madag16-vivax_R_2.fastq -S $DIR/Madagascar16.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Madag31-vivax_R_1.fastq -2 $DIR/Madag31-vivax_R_2.fastq -S $DIR/Madagascar31.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Madag51-vivax_R_1.fastq -2 $DIR/Madag51-vivax_R_2.fastq -S $DIR/Madagascar51.sam

# Eighth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Madagascar16.sam -o $DIR/Madag16.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Madagascar31.sam -o $DIR/Madag31.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Madagascar51.sam -o $DIR/Madag51.bam

# Ninth step: Sort the bam files
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Madag16.bam $DIR/Madagascar16_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Madag31.bam $DIR/Madagascar31_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Madag51.bam $DIR/Madagascar51_sorted
```

Annex 12: Script used to filter contamination from Peru sequences, align with vivax reference genome and variant calling

```
#PBS -N Aligment_2_Variants
#PBS -q UCTlong
#PBS -l nodes=1:ppn=32:series600
DIR=/researchdata/fhgfs/djmmou001/Vivax_data

# NB, PLEASE READ THIS!
# There is a 2:1 correspondence between RAM and cores on the 600 series.
# You need to know how much RAM your job will consume before submitting it.
# Please set the ppn value above to be 1/2 the GB of RAM required.  For
# example a job needing 10GB of RAM should have ppn=5

# Please leave the hostname command here for troubleshooting purposes hostname

# Your science stuff goes here:
# Your science stuff goes here:
# First step: alignment with human reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/entry.fasta $DIR/Human-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1
$DIR/SRR1568202_1.fastq.gz -2 $DIR/SRR1568202_2.fastq.gz -S $DIR/Peru-
Hum_SRR1568202.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1
$DIR/SRR1562606_1.fastq.gz -2 $DIR/SRR1562606_2.fastq.gz -S $DIR/Peru-
Hum_SRR1562606.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1
$DIR/SRR1562614_1.fastq.gz -2 $DIR/SRR1562614_2.fastq.gz -S $DIR/Peru-
Hum_SRR1562614.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1
$DIR/SRR1568216_1.fastq.gz -2 $DIR/SRR1568216_2.fastq.gz -S $DIR/Peru-
Hum_SRR1568216.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1
$DIR/SRR1562624_1.fastq.gz -2 $DIR/SRR1562624_2.fastq.gz -S $DIR/Peru-
```

```
Hum_SRR1562624.sam

# Second step: generate bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Peru-
Hum_SRR1568202.sam -o $DIR/Peru-Hum_SRR1568202.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Peru-
Hum_SRR1562606.sam -o $DIR/Peru-Hum_SRR1562606.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Peru-
Hum_SRR1562614.sam -o $DIR/Peru-Hum_SRR1562614.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Peru-
Hum_SRR1568216.sam -o $DIR/Peru-Hum_SRR1568216.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/Peru-
Hum_SRR1562624.sam -o $DIR/Peru-Hum_SRR1562624.bam

# Third step: filtering human DNA
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru02_filtered_R#.fastq --no-aligned
$DIR/Peru-Hum_SRR1568202.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru06_filtered_R#.fastq --no-aligned
$DIR/Peru-Hum_SRR1562606.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru14_filtered_R#.fastq --no-aligned
$DIR/Peru-Hum_SRR1562614.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru16_filtered_R#.fastq --no-aligned
$DIR/Peru-Hum_SRR1568216.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru24_filtered_R#.fastq --no-aligned
$DIR/Peru-Hum_SRR1562624.bam

# Fourth step: align with vivax reference genome the read unaligned with human genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Peru02_filtered_R_1.fastq -2 $DIR/Peru02_filtered_R_2.fastq -S $DIR/Peru02_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Peru06_filtered_R_1.fastq -2 $DIR/Peru06_filtered_R_2.fastq -S $DIR/Peru06_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
```

```
$DIR/Peru14_filtered_R_1.fastq -2 $DIR/Peru14_filtered_R_2.fastq -S $DIR/Peru14_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Peru16_filtered_R_1.fastq -2 $DIR/Peru16_filtered_R_2.fastq -S $DIR/Peru16_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Peru24_filtered_R_1.fastq -2 $DIR/Peru24_filtered_R_2.fastq -S $DIR/Peru24_filt.sam

# Fifth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru02_filt.sam -o $DIR/Peru02_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru06_filt.sam -o $DIR/Peru06_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru14_filt.sam -o $DIR/Peru14_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru16_filt.sam -o $DIR/Peru16_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru24_filt.sam -o $DIR/Peru24_filt.bam

# Sixth step: extract reads mapping with vivax reference genome
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru02-vivax_R#.fastq --no-unaligned
$DIR/Peru02_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru06-vivax_R#.fastq --no-unaligned
$DIR/Peru06_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru14-vivax_R#.fastq --no-unaligned
$DIR/Peru14_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru16-vivax_R#.fastq --no-unaligned
$DIR/Peru16_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Peru24-vivax_R#.fastq --no-unaligned
$DIR/Peru24_filt.bam

# Seventh step: align again with vivax reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Peru02-vivax_R_1.fastq
```

```
-2 $DIR/Peru02-vivax_R_2.fastq -S $DIR/Peru02.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Peru06-vivax_R_1.fastq
-2 $DIR/Peru06-vivax_R_2.fastq -S $DIR/Peru06.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Peru14-vivax_R_1.fastq
-2 $DIR/Peru14-vivax_R_2.fastq -S $DIR/Peru14.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Peru16-vivax_R_1.fastq
-2 $DIR/Peru16-vivax_R_2.fastq -S $DIR/Peru16.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Peru24-vivax_R_1.fastq
-2 $DIR/Peru24-vivax_R_2.fastq -S $DIR/Peru24.sam

# Eighth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru02.sam -o $DIR/Peru02.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru06.sam -o $DIR/Peru06.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru14.sam -o $DIR/Peru14.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru16.sam -o $DIR/Peru16.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Peru24.sam -o $DIR/Peru24.bam

# Ninth step: Sort the bam files
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Peru02.bam $DIR/Peru02_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Peru06.bam $DIR/Peru06_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Peru14.bam $DIR/Peru14_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Peru16.bam $DIR/Peru16_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Peru24.bam $DIR/Peru24_sorted
```

Annex 13: Script used to filter contamination from India sequences, align with vivax reference genome and variant calling

```
#PBS -N Aligment_2_Variants
#PBS -q UCTlong
#PBS -l nodes=1:ppn=32:series600
DIR=/researchdata/fhgfs/djmmou001/Vivax_data

# NB, PLEASE READ THIS!
# There is a 2:1 correspondence between RAM and cores on the 600 series.
# You need to know how much RAM your job will consume before submitting it.
# Please set the ppn value above to be 1/2 the GB of RAM required.  For
# example a job needing 10GB of RAM should have ppn=5

# Please leave the hostname command here for troubleshooting purposes hostname

# Your science stuff goes here:
# Your science stuff goes here:
# First step: alignment with human reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/entry.fasta $DIR/Human-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332913_1.fastq.gz -2 $DIR/SRR332913_2.fastq.gz -S $DIR/India-Hum_SRR332913.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332914_1.fastq.gz -2 $DIR/SRR332914_2.fastq.gz -S $DIR/India-Hum_SRR332914.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332915_1.fastq.gz -2 $DIR/SRR332915_2.fastq.gz -S $DIR/India-Hum_SRR332915.sam

# Second step: generate bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/India-Hum_SRR332913.sam -o $DIR/India-Hum_SRR332913.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/India-Hum_SRR332914.sam -o $DIR/India-Hum_SRR332914.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/India-Hum_SRR332915.sam -o $DIR/India-Hum_SRR332915.bam
```

```
# Third step: filtering human DNA
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/India13_filtered_R#.fastq --no-aligned
$DIR/India-Hum_SRR332913.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/India14_filtered_R#.fastq --no-aligned
$DIR/India-Hum_SRR332914.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/India15_filtered_R#.fastq --no-aligned
$DIR/India-Hum_SRR332915.bam

# Fourth step: align with vivax reference genome the read unaligned with human genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/India13_filtered_R_1.fastq -2 $DIR/India13_filtered_R_2.fastq -S
$DIR/India13_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/India14_filtered_R_1.fastq -2 $DIR/India14_filtered_R_2.fastq -S
$DIR/India14_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/India15_filtered_R_1.fastq -2 $DIR/India15_filtered_R_2.fastq -S
$DIR/India15_filt.sam

# Fifth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/India13_filt.sam -o $DIR/India13_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/India14_filt.sam -o $DIR/India14_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/India15_filt.sam -o $DIR/India15_filt.bam

# Sixth step: extract reads mapping with vivax reference genome
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/India13-vivax_R#.fastq --no-unaligned
$DIR/India13_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/India14-vivax_R#.fastq --no-unaligned
```

```
$DIR/India14_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/India15-vivax_R#.fastq --no-unaligned
$DIR/India15_filt.bam

# Seventh step: align again with vivax reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/India13-vivax_R_1.fastq
-2 $DIR/India13-vivax_R_2.fastq -S $DIR/India13.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/India14-vivax_R_1.fastq
-2 $DIR/India14-vivax_R_2.fastq -S $DIR/India14.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/India15-vivax_R_1.fastq
-2 $DIR/India15-vivax_R_2.fastq -S $DIR/India15.sam

# Eighth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/India13.sam -o $DIR/India13.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/India14.sam -o $DIR/India14.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/India15.sam -o $DIR/India15.bam

# Ninth step: Sort the bam files
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/India13.bam $DIR/India13_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/India14.bam $DIR/India14_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/India15.bam $DIR/India15_sorted
```

Annex 14: Script used to filter contamination from Brazil sequences, align with vivax reference genome and variant calling

```
#PBS -N Aligment_2_Variants
#PBS -q UCTlong
#PBS -l nodes=1:ppn=32:series600
DIR=/researchdata/fhgfs/djmmou001/Vivax_data

# NB, PLEASE READ THIS!
# There is a 2:1 correspondence between RAM and cores on the 600 series.
# You need to know how much RAM your job will consume before submitting it.
# Please set the ppn value above to be 1/2 the GB of RAM required.  For
# example a job needing 10GB of RAM should have ppn=5

# Please leave the hostname command here for troubleshooting purposes hostname

# Your science stuff goes here:
# First step: alignment with human reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/entry.fasta $DIR/Human-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332567_1.fastq.gz
-2 $DIR/SRR332567_2.fastq.gz -S $DIR/Brazil-Hum_SRR332567.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR332569_1.fastq.gz
-2 $DIR/SRR332569_2.fastq.gz -S $DIR/ Brazil-Hum_SRR332569.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/Human-ref -1 $DIR/SRR340134_1.fastq.gz
-2 $DIR/ SRR340134_2.fastq.gz -S $DIR/ Brazil-Hum_ SRR340134.sam

# Second step: generate bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/ Brazil-
Hum_SRR332567.sam -o $DIR/ Brazil-Hum_SRR332567.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/ Brazil-
Hum_SRR332569.sam -o $DIR/ Brazil-Hum_SRR332569.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/entry.fasta $DIR/ Brazil-
Hum_SRR340134.sam -o $DIR/ Brazil-Hum_ SRR340134.bam
```

```
# Third step: filtering human DNA
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Brazil67_filtered_R#.fastq --no-aligned
$DIR/Brazil-Hum_SRR332567.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Brazil69_filtered_R#.fastq --no-aligned
$DIR/Brazil-Hum_SRR332569.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Brazil34_filtered_R#.fastq --no-aligned
$DIR/Brazil-Hum_SRR340134.bam

# Fourth step: align with vivax reference genome the read unaligned with human genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta
$DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Brazil67_filtered_R_1.fastq -2 $DIR/Brazil67_filtered_R_2.fastq -S
$DIR/Brazil67_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Brazil69_filtered_R_1.fastq -2 $DIR/Brazil69_filtered_R_2.fastq -S
$DIR/Brazil69_filt.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1
$DIR/Brazil34_filtered_R_1.fastq -2 $DIR/Brazil34_filtered_R_2.fastq -S
$DIR/Brazil34_filt.sam

# Fifth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Brazil67_filt.sam -o $DIR/Brazil67_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Brazil69_filt.sam -o $DIR/Brazil69_filt.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-
35_PvivaxP01_Genome.fasta $DIR/Brazil34_filt.sam -o $DIR/Brazil34_filt.bam

# Sixth step: extract reads mapping with vivax reference genome
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Brazil67-vivax_R#.fastq --no-unaligned
$DIR/Brazil67_filt.bam
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Brazil69-vivax_R#.fastq --no-unaligned
$DIR/Brazil69_filt.bam
```

```
/opt/exp_soft/bam2fastq/bam2fastq -o $DIR/Brazil34-vivax_R#.fastq --no-unaligned $DIR/Brazil34_filt.bam

# Seventh step: align again with vivax reference genome
/opt/exp_soft/bowtie2-2.2.6/bowtie2-build $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/vivax-ref
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Brazil67-vivax_R_1.fastq -2 $DIR/Brazil67-vivax_R_2.fastq -S $DIR/Brazil67.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Brazil69-vivax_R_1.fastq -2 $DIR/Brazil69-vivax_R_2.fastq -S $DIR/Brazil69.sam
/opt/exp_soft/bowtie2-2.2.6/bowtie2 -p 4 -x $DIR/vivax-ref -1 $DIR/Brazil34-vivax_R_1.fastq -2 $DIR/Brazil34-vivax_R_2.fastq -S $DIR/Brazil34.sam

# Eighth step: generate filtered bam files
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Brazil67.sam -o $DIR/Brazil67.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Brazil69.sam -o $DIR/Brazil69.bam
/opt/exp_soft/samtools-0.1.19/samtools view -bT $DIR/PlasmoDB-35_PvivaxP01_Genome.fasta $DIR/Brazil34.sam -o $DIR/Brazil34.bam

# Ninth step: Sort the bam files
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Brazil67.bam $DIR/Brazil67_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Brazil69.bam $DIR/Brazil69_sorted
/opt/exp_soft/samtools-0.1.19/samtools sort $DIR/Brazil34.bam $DIR/Brazil34_sorted
```

Annex 15: Script used to generate common vcf file and extract the overlapping position and generate the ped file.

```
#!/bin/bash

# Generate the common vcf file
# Index the bam files
samtools index Anefis.bam
samtools index Brazil34.bam
samtools index Brazil67.bam
samtools index Brazil69.bam
samtools index India13.bam
samtools index India14.bam
samtools index India15.bam
samtools index Kidal1.bam
samtools index Kidal2.bam
samtools index Kidal3.bam
samtools index Kidal4.bam
samtools index Madagascar16.bam
samtools index Madagascar31.bam
samtools index Madagascar51.bam
samtools index Maurit08.bam
samtools index Maurit09.bam
samtools index Maurit10.bam
samtools index Maurit11.bam
samtools index Maurit12.bam
samtools index Maurit13.bam
samtools index Maurit14.bam
samtools index Maurit15.bam
samtools index Peru02.bam
samtools index Peru06.bam
samtools index Peru14.bam
samtools index Peru16.bam
```

```
samtools index Peru24.bam
samtools index TinEssako.bam

# Generate vcf file
samtools mpileup -u -f PlasmoDB-35_PvivaxP01_Genome.fasta Anefis.bam Brazil34.bam
Brazil67.bam Brazil69.bam India13.bam India14.bam India15.bam Kidal1.bam Kidal2.bam
Kidal3.bam Kidal4.bam Madagascar16.bam Madagascar31.bam Madagascar51.bam
Maurit08.bam Maurit09.bam Maurit10.bam Maurit11.bam Maurit12.bam Maurit13.bam
Maurit14.bam Maurit15.bam Peru02.bam Peru06.bam Peru14.bam Peru16.bam Peru24.bam
TinEssako.bam > All_vivax.bcf

bcftools view -v -c -g All_vivax.bcf > All_vivax.vcf

# Extract the overlapping position
bcftools query -f '%CHROM\t%POS\t%POS\n' TinEssako.vcf > TinEssako_positions.csv
bcftools view --targets-file TinEssako_positions.csv All_vivax.vcf > All_vivax_Overlap.vcf

# Generate the ped file
/home/moussa/Downloads/plink --vcf All_vivax_Overlap.vcf --snps-only --biallelic-only
strict --allow-extra-chr --keep-allele-order --recode --alleleACGT --make-bed --out All_vivax
```

Annex 16: Script used to convert ped to fasta file

```
# PED to FASTA converter, ped2fasta.pl
# Gets first 6 columns of each line as header line and the rest as
# the sequence replacing 0s with Ns and organizes it into a FASTA file

# Note 0s are for missing nucleotides defined by default in PLINK

# How to run: perl ped2fasta.pl "C:\path\to\PED_File"

use strict;
my $file = $ARGV[0]; # Path to PED file
my @columns;
my $sequence;
open(my $inf, "<", $file.".ped") or die $!;
open(my $ouf, ">", $file.".fa") or die $!;
while (my $row = <$inf>) {
	# Each row belongs to one individual
	chomp $row;
	@columns = split(" ", $row); # Splits columns of each line into an array
	print $ouf ">", join(" ", @columns[0..5]), "\n"; # Joins and prints first 6 columns of each line
	$sequence = join("", @columns[6..$#columns]); # Joins all nucleotides
	$sequence =~ s/0/N/g; # Replaces 0s with Ns
	print $ouf $sequence, "\n"; # Prints the sequence
}
close $inf;
close $ouf;
```

Annex 17: Script used to align sequences for evolutionary relationship analysis purpose

```
#!/bin/bash
#PBS -l select=2:ncpus=24
#PBS -l walltime=120:00:00
#PBS -P CBBI0818
#PBS -q normal
#PBS -o /mnt/lustre/users/mdjimde/moussa_data/stdout.txt
#PBS -e /mnt/lustre/users/mdjimde/moussa_data/stderr.txt
#PBS -N Align_Taxa
#PBS -M mdjimde@icermali.org
#PBS -m be

module add chpc/BIOMODULES
module add mafft/7.305_intel16.0.1

mafft /mnt/lustre/users/mdjimde/Phylogeny/ All_vivax.fa >
/mnt/lustre/users/mdjimde/Phylogeny/ All_vivax_aligned.fasta
```

Annex 18: Standard Operating Procedure (SOP) used for DNA extraction from RDTs

a- Materials and Equipment

- QiaAmp DNA Blood Midi Kit (100) including:
  - Qiagen Protease
  - Buffer AL (lysis buffer)
  - Buffer ATL (tissue lysis buffer)
  - QiaAmp Midi columns with collection tubes
  - Buffer AW1 (wash buffer 1)
  - Buffer AW2 (wash buffer 2)
  - Buffer AE
  - Collection tubes for eluted DNA
- Centrifuge with swinging bucket rotor and 15mL tube holders
- 70C water bath with floatation device
- Ethanol (96-100%)
- Proteinase K
- 2mL centrifuge tubes and tube racks
- 1.5mL microcentrifuge tubes and racks
- 5mL serological pipets and pipet pump
- 1000µL filter tips and P-1000 pipettor
- Pipet sheaths
- Pliers to break the RDTs
- Scissors to cut the strip
- Distilled water

b- Procedure

- The RDTs have been broken to extract the strip on which the drop of blood has migrated.
- The dried blood spot on the the strip were placed into a 1.5 ml microcentrifuge tube.
- Reconstitute Qiagen Protease (if not already done).
- Tissue lysis buffer (180 µl of Buffer ATL) was added in the tube. We then left the tube in incubation overnight at room temperature (15–25°C).
- The next day, the tube where briefly centrifuged after 10 minutes' incubation at 85°C to remove drop from lib.

- Proteinase K (20 µl) was added. The mix was vortexed and incubated at 56°C during 1 hour following by brief centrifuge.
- Then, lysis buffer (200 µl Buffer AL) was added and thoroughly vortex for mix and incubated during 10 minutes at 70°C and briefly centrifuged.
- 200 µl of 96-100% ethanol was added and mixed thoroughly using vortex following by a brief centrifuge.
- After applied from the microcentrifuge tubes to the QIAamp Mini spin column, the mixtures were centrifuged at 8000 rpm for 1 minute.
- After centrifugation, QIAamp Mini spin column was placed in 2 ml collection tubes. The filtrate was discarded.
- 500 µl of Buffer AW1 was added carefully (without wetting the rim) to the QIAamp Mini spin column and centrifuged at 8000 rpm for 1 minute.
- QIAamp Mini spin column was placed in 2 ml collection tubes and as previous step the filtrate was discarded.
- 500 µl of Buffer AW2 was added carefully (without wetting the rim) to the QIAamp Mini spin column and centrifuged at full speed 14,000 rpm for 3 minutes.
- QIAamp Mini spin column was placed in new 2 ml collection tubes and centrifuged at full speed for 1 minute.
- The QIAamp Mini spin column was finally placed in new 1.5 ml microcentrifuge tubes.
- The tubes containing the filtrate was discarded.
- The tubes were carefully opened and 100 µl distilled water was added.
- The tube was incubated at room temperature (15–25°C) for a minute before centrifugation. The speed used was 8000 rpm for 1 minute.

c- Limitations of the Procedure

The QiaAmp DNA Blood Midi Kit is more expensive than the Nucleon DNA Extraction Kit; however, it has the advantage of not requiring the use of organic solvents (e.g. chloroform).

Annex 19: SOP used for Nested Polymerase Chain Reaction (PCR)

a- Materials and Equipment

- Thermocycler
- Centrifuge
- PCR Buffer (+ MgCl2)
- 10xdNTPs
- Taq polymerase
- Water for PCR
- Bovine Serum Albumin (BSA)
- NEB Digestion Buffer
- ApoI Digestion Enzyme
- Sterile Gloves
- Tubes (1.5ml, 0.5ml, 0.2ml) and racks
- DNA
- pipettes and tips (1000µl, 200µl, 20µl and 10µl)
- Markers
- Water bath or incubator
- Primers for species diagnostic:

Plasmodium species	Primers
Plasmodium sp.	rPLU5: CCT GTT GTT GCC TTA AAC TTC rPLU6: TTA AAA TTG TTG CAG TTA AAA G
P. falciparum	rFAL1: TTA AAC TGG TTT GGG AAA ACC AAA TAT ATT rFAL2: ACA CAA TGA ACT CAA TCA TGA CTA CCC GTC
P. malariae	rMAL1: ATA ACA TAG TTG TAC GTT AAG AAT AAC CGC rMAL2: AAA ATT CCC ATG CAT AAA AAA TTA TAC AAA
P. vivax	rVIV1: CGC TTC TAG CTT AAT CCA CAT AAC TGA TAC rVIV2: ACT TCC AAG CCG AAG CAA AGA AAG TCC TTA
P. ovale curtisi	rOVA1: ATC TCT TTT GCT ATT TTT TAG TAT TGG AGA rOVA2: ATC TAA GAA TTT CAC CTC TGA CAT CTG

b- Procedure

- Preparation of the reaction mixture

Reagents	Initial concentration	Final concentration	Volume (µL)
Buffer/$MgCl_2$	5X/17.5 nM	1X/3.5 nM	5
dNTPs	10X/2nM	1X/200µM	2.5
Primers	100X/10µM	1X/100nM	0.25
Taq Polymerase	5U/µL	0.025U	0.125
PCR water	-	-	16.125
Total			24

- First amplification (PCR1) programme:
  - 95°C for 5 min;
  - 94°C for 30sec,
  - 58°C for 30sec,
  - and 72°C for 30sec;
  - 44 cycles go to 2;
  - 72°C for 5 min
- Second amplification (PCR2) programme:
  - 95°C for 5 min;
  - 94°C for 30sec,
  - 58°C for 30sec,
  - and 72°C for 30sec;
  - 19 cycles go to 2,
  - 72°C for 5 min

Annex 20: SOP used for Selective Whole Genome Amplification (sWGA)

a- Materials and Equipment

- Thermocycler
- Centrifuge
- Taq polymerase
- Water for PCR
- Sterile Gloves
- Tubes (1.5ml, 0.5ml, 0.2ml) and racks
- DNA extracted from RDT samples using the QIAamp DNA mini kit (Qiagen, Valencia, CA)
- Phi29 Buffer
- Phi29 DNA polymerase
- dNTPs
- BSA
- Phi29 polymerase
- primers
    - 5'-AACGAAGC*G*A-3'
    - 5'-ACGAAGCG*A*A-3'
    - 5'-ACGACGA*A*G-3'
    - 5'-ACGCGCA*A*C-3'
    - 5'-CAACGCG*G*T-3'
    - 5'-GACGAAA*C*G-3'
    - 5'-GCGAAAAA*G*G-3'
    - 5'-GCGAAGC*G*A-3'
    - 5'-GCGGAAC*G*A-3'
    - 5'-GCGTCGA*A*G-3'
    - 5'GGTTAGCG*G*C-3'and
    - 5'-AACGAAT*C*G-3'

b- Procedure

- Ramp down 35°C, the temperature was increased by 1°C each 10 minutes
- 16 hours at 30°C
- 10 minutes at 65°C
- Hold at 4°C

c- DNA quantification

NanoDrop 2000 UV-vis spectrophotometer (made in U.S.A.) was used to quantify DNA before and after sWGA.