Vrije Universiteit Brussel
Katholieke Universiteit Leuven
Universiteit Antwerpen

Interuniversity Program Molecular Biology (IPMB)

CONTRIBUTIONS TO UNRAVELLING THE NATURE OF THE FALSE POSITIVITY IN CIRCUMSPOROZOITE *Plasmodium falciparum* ELISA

Thesis submitted in partial fulfillment of the requirements for the Degree of
Master of Science in Molecular Biology

Francisco Javier Morales Yánez

Promoter: Prof. Marc Coosemans **Supervisor: Dr. Lies Durnez**

Unit of Medical Entomology
Institute of Tropical Medicine Antwerp
Academic year 2012-2013

ACKNOWLEDGMENTS

I would like to thank first to the Interuniversity Program Molecular Biology (IPMB) for giving me the chance to boost my career and for the amazing theoretic and practical training provided during the last two years. Professor Edilbert VanDriessche and Rudi Willemns have been doing a wonderful job coordinating the program. For them, my sincere acknowledgments. Special thanks for the Belgian Technical Cooperation (BTC) for funding my scholarship in Belgium and for its outstanding collaboration with the South.

Thanks to Professor Marc Coosemans, head of the Unit of Medical Entomology for accepting me as a master student and giving me he opportunity to be part of the wonderful group of researcher in his unit. I would like to express my most sincere acknowledgments to Dr. Lies Durnez for the excellent guidance and training through the development and writing of this thesis. It was an honor to work with such an exceptional and high-level researcher. Special thanks also to Leen Denis for the excellent technical assistance in the laboratory.

My sincere acknowledgements to the personnel of the Center of Medical Genetics from the Antwerp University Hospital for their excellent technical support in Next Generation Sequencing. Many thanks also to Geert Vandeweyer and Victor Cavalcanti for their assistance in bioinformatics analysis.

This thesis is dedicated to my family and friends, who have supported me during the hard-working days of laboratory and writing work.

Francisco

Table of Contents

LIST OF ABBREVIATIONS

BB	Blocking Buffer
BLAST	Basic Local Alignment Search Tool
CSP	Circumsporozoite Protein
CSP ELISA	Circumsporozoite Enzyme Linked Immune Assay
EIR	Entomological Inoculation Rate
FPABD	False Positive Abdomen
FPHT	False Positive Head/Thorax
IRS	Indoor Residual Spraying
LLIN	Long Lasting Impregnated Bednets
MALDI TOF/TOF	Matrix-assisted desorption/ionization time of flight/time of flight
NGS	Next Generation Sequencing
OD	Optical Density
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
Pf CSP	Circumsporozoite Protein of *Plasmodium falciparum*
Pf CSP ELISA	*Plasmodium falciparum* Circumsporozoite Enzyme linked Immune Assay
PSIRED	Prediction Secondary Structure
RBC	Red Blood Cells
rpm	Revolutions per minute
RT	Room Temperature
STABD	Anopheles stephensi Abdomen
STHT	Anopheles stephensi Head/Thorax
TN	True Negative
TTGE	Temporal Temperature gradient electrophoresis
WGS	Whole Genome Sequencing
WHO	World Health Organization

ABSTRACT

INTRODUCTION: *Plasmodium falciparum* malaria, a vector-borne disease caused by the bite of *Anopheles* mosquitoes is a public health problem worldwide. Vector control remains as an efficient method to block malaria transmission. Vector incrimination, being a crucial factor for efficient vector control, nowadays mainly relies on detection of the parasite's circumsporozoite protein in ELISA. However, false positivity in *P. falciparum* Circumsporozoite ELISA has been reported in Africa and Southeast Asia. The principal objective of this research is to determine the nature of the agent that provokes false positivity in Pf CSP ELISA through *in silico* analysis of Pf CSP, Next Generation Sequencing of bacterial 16s rDNA, detection of Microsporidia parasites in mosquitoes and isolation of the protein responsible for false positivity. **MATERIALS AND METHODS:** *In silico* analysis of Pf CSP was performed through BLAST search. Primers Bakt_341F/Bakt_805R attached to Illumina® adaptors were used to amplify bacterial 16s rDNA in 10 false positive samples from Vietnam and Cambodia, 1 laboratory reared *An. stephensi* and 2 true negative mosquitoes. Samples were identified into 16s rDNA database from GenBank with the Galaxy workflow modified for metagenomic studies. Cross-reaction of Microsporidia in Pf CSP ELISA was evaluated with *Aedes aegypti* infected with *Vavraia culicis* and *Edhazardia aedis.* Microsporidia were detected by specific amplification of 18s rDNA. Protein capture from 6 false positive samples was carried out using Dynabeads® M270 Epoxy antibody coupling kit and capture monoclonal antibodies 2A10. Isolated protein was separated through SDS-PAGE and identified by MALDI TOF/TOF analysis. **RESULTS:** *In silico* analysis of Pf CSP demonstrates some degree of identity with the domain Bacteria and a disordered NANPL region present in the Microsporidia *E. aedis*. Amplification and sequencing of 16s rDNA of false positive samples did not indicate a single species of bacteria present in the majority of false positive samples analyzed and absent in negative controls. Microsporidia did not show cross reactivity in Pf CSP ELISA and only one sample was found positive in head/thorax and abdomen for detection of Microsporidia by PCR. MALDI TOF/TOF analysis showed that actin cross-reacted with capture monoclonal antibodies in Pf CSP ELISA. **DISCUSSION**: *In silico* and laboratory analysis can not confirm Microsporidia, nor one single bacterial

species as the cross-reactive protein. The remarkable diversity of species supports the hypothesis that more than one bacterial species could cross react with monoclonal antibodies in Pf CSP ELISA. The finding that the actin protein can be recognized by anti-Pf CSP monoclonal antibodies supports the animal origin theory of the causative agent of false positivity. Identification of this agent remains unclear, nevertheless important insights are provided towards a feasible explanation of false positivity in Pf CSP ELISA.

1. INTRODUCTION

1.1. General overview of malaria

Malaria is a parasitic disease caused by hemoprotozoan *Plasmodium* species that is transmitted through the bite of female *Anopheles* mosquitoes. The genus *Plasmodium* comprises four species of medical importance: *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*. *P. falciparum* is responsible for the most severe clinical presentation of the disease, cerebral malaria, while *P. vivax* causes a more benign disease, rarely fatal, but highly spread worldwide, except in sub-Saharan Africa, where the absence of Duffy antigen in red blood cells (RBC) does not allow a high level of transmission in *P. vivax* [1][2] [3](Figure 1).

The complex life cycle of the parasite demonstrates the high level of adaptation of the parasite to its mammalian and arthropod hosts. The infective state of the parasite is the sporozoite, which is injected by the *Anopheles* mosquito into the skin of the vertebrate host during a blood meal. Once in the blood, the parasite infects the hepatocyte and performs several rounds of replication into these cells resulting in the schizonts forms of the parasite. The hepatic cell is disrupted and thousands of merozoites are released to the blood and then locate into the red blood cell where several rounds of replications take place that ultimately lead to the destruction of the erythrocytes releasing new merozoites. The majority of them infect new red blood cells achieving the erythrocytic cycle of the parasite [4]. A few parasites develop into sexual stages (gametocytes), which are ingested during a blood meal by a female *Anopheles* mosquito. In the midgut of the mosquito, microgametocyte and macrogametocyte fuse to form the zygote, which then develops into an ookinete, that permeates the epithelial wall of midgut cells. The parasite encysts immediately under the basal lamina of the epithelial cells giving place to the oocyst. These events are called the “early sporogonic cycle” and usually lasts at least two days while the mosquito is digesting the blood meal [5]. The oocysts remains at this stage for a period called the “mid sporogonic cycle” followed by successive rounds of replications

that leads to the release and invasion of sporozoites to the haemocoel and finally to the salivary glands ("late sporogonic cycle")[5] [6] [7]. (Figure 2).

1.2. Malaria control

Although important advances in controlling the disease have been achieved in the last years, malaria is still a challenging public health issue in many countries (Figure 1) [3]. According to the World Health Organization (WHO) 3,3 billion of people live at high risk of acquiring the disease. The number of cases has decreased globally in the last years and some countries report a downward trend up to 50%. Nevertheless 216 millions of cases were detected in 2010; 81% of them in Africa. Malaria claimed 655.000 lives in 2010, 91% children under five years old in Africa[3].

Malaria control is based on four aspects: i) Accurate and early diagnosis; ii) Rapid and effective treatment iii) Intermittent preventive treatment and iv) Vector control. The first three strategies are aimed to stop the cycle of the parasite in the human host, a key factor in the life cycle because unlike other vector-borne diseases there is no non-human reservoir for *P. falciparum* or *P. vivax*. Another key aspect of malaria control is blocking the host-vector contact to prevent humans from becoming infected. The contact of the vector with the human host depends on several factors, some of them related to the degree of anthropophagy and the biting activity of the mosquito, which are highly variable according to species. *Anopheles gambiae sensu lato*, the most important vector in Africa has an innate anthropophilic and endophilic behavior, explaining its high capacity to perpetuate the cycle of *P. falciparum* [8]. In South-East Asia, *An. dirus* A, one of the most important vectors of *P. falciparum* is highly anthropophilic, exophilic and with a preference to bite outdoors, while *An. minimus* A presents a more zoophilic behavior [9]. These striking differences are important parameters to implement malaria control projects based on Long Lasting Impregnated Bednets (LLIN) and indoor residual insecticide spray (IRS) because these strategies are effective only when mosquitoes bite or rest indoors during times that people are sleeping under bednets.

1.3. Vector Incrimination and false positivity in Circumsporozoite ELISA

Vector incrimination, meaning pin-pointing anophelines that can transmit the malaria parasite successfully, is a very important indicator in vector controls programs as different anopheline species can exhibit different behavior, and thus should be targeted differently. The gold standard test for incrimination of anopheline mosquito species as vectors of malaria relies on the detection of sporozoites, the infective stage of the parasite, in salivary glands through microscopic observation of the parasites in freshly dissected salivary glands [10] [11], nevertheless the procedure is time consuming and needs to be performed in the field with a microscope by trained personnel [12]. For these reasons the Circumsporozoite (CSP) Enzyme Linked Immune Assay (CSP-ELISA) to detect parasites in field mosquito samples according to the technique described by Wirtz *et al.* [13] has been developed as a highly sensitive test to determine the Entomological Inoculation Rate (EIR), a measure of the intensity of the malaria transmission which is calculated by multiplying the mosquito biting rate and the sporozoite rate in each species of mosquitoes. [10]. Other methods of detection include the Polymerase Chain Reaction (PCR) [14] [15], a technique that has been postulated as highly sensitive and able to avoid drawbacks of CSP-ELISA such as overestimation of EIR due to dissemination of sporozoites through wings and legs of the mosquitoes [11]. The latter problem can be easily tackled with an accurate separation and analysis of the different parts of the mosquito; nevertheless detection of DNA through PCR is not specific for sporozoite stage and remains as a confirmatory test. High levels of overestimation in EIR for *P. falciparum* has been reported recently in *An. karwari, An. barbirostris, An. kochi, An. jeyporiensis, and An. peditaeniatusi* in Bangladesh [16] and in *An. barbirostris, An. maculatus and An. minimus s.l.* in Cambodia and Vietnam [17]. False positives have been also described in *An. gambiae* and *An. moucheti* in Gabon and Tchad [18]. A common factor in these studies is that false positivity is correlated with zoophilic behavior of the vector and only in the thoracic portion of the mosquito. Heating at 100^{o}C for 10 minutes efficiently eliminates false positives, as has been described by [17] and [19]. A possible cross-reaction with other parasites related with the genus *Plasmodium* such us Haemosporidia, Trypanosomatidae, Piroplasmida, and Haemogregarina has already

ruled out [17]. For the reasons indicated above, currently CSP-ELISA needs to be confirmed through PCR or heating of the samples before ELISA analysis, nevertheless the real nature of the factor associated with false positivity remains unknown.

The primary structure of *P. falciparum* CSP (Pf CSP) has been previously described [20]. It has 442 amino acids and it is the most abundant surface protein of *P. falciparum* during its stage as sporozoites [21]. The protein has an N-terminus that binds to sulfate proteoglycans, a region composed by a four amino acid repeat region (NANP/NVDP) and the C-terminus which corresponds to a thrombospondin-like type I repeat (TSR) domain [22]. The central region is formed by 43 repeats of the NANP sequence and 4 NVDP of the sequence NVDP, nevertheless these number can vary according to the strain of the parasite [19]. The NANP repeated region is the main antigenic determinant in the protein and consequently, the responsible of the recognition of monoclonal antibodies in the CSP-ELISA. The recombinant *P. falciparum* protein R32tet$_{32}$ CS protein used in the CSP-ELISA as positive control is composed by 30 Asn-Ala-Asn-Pro and two Asn-Val-Asp-Pro tetrapeptide repeats [23].

1.4. Bacterial proteins as possible cross reacting agents in CSP-ELISA

In the context of the false positivity detected in Pf CSP ELISA, a cross-reaction of the NANP region of Pf CSP with a bacterial protein present in the head/thorax portion of the mosquito has not been documented so far; nevertheless, preliminary experiments carried out to isolate and identify a cross-reacting protein present in head/thorax lysates of false positive specimens indicate that bacterial agents could be implicated in the false positives recently identified in Southeast Asia (Unpublished data).

The approach to identify complex bacterial communities in *Anopheles* mosquitoes has dramatically changed in the last years upon the availability of more accurate and sophisticated technologies in metagenomics. Before the arose of sequencing technologies, culture-dependent methods have demonstrated that species composition differs according to *Anopheles* species, geographical localization and the kind of habitat where mosquitoes are captured [24]. Unfortunately, these methods are not able to capture the real diversity of bacteria because roughly 50% of the total

taxa could be not detected under culture-based methods in laboratory reared *An. stephensi* mosquitoes [24]. Identification of bacteria through analysis of 16s rDNA is nowadays the most accurate way to study bacterial communities regardless the origin and complexity of the sample [25]. Two different approaches have been developed in metagenomic studies based in 16s rDNA. The first one is the PCR amplification of 16s rDNA with primers targeting hypervariable regions V1-V9 of the gene. In this case, careful selection of the primers to be used as well as special care to avoid overestimation of taxa must be considered due to artifacts in PCR reaction that could lead to chimeric sequences [26]. Secondly, assembly of whole genome sequences (WGS) from bacterial communities has been also used to identify bacteria in different ecological niches, nevertheless this technique exhibits poor recovery of microorganism that are present in small quantities within a bacterial community [27]. Next Generation Sequencing (NGS) has brought metagenomic studies to an unprecedented level, allowing rapid, accurate and detailed description of bacterial communities in a wide variety of environments including the *Anopheles* midgut [28] [29] [30] [31]. Metagenomic studies in *Anopheles* mosquitoes have been developed mainly to identify a target to block transmission through paratransgenesis, a process aimed to impede the parasite development through symbiotic bacteria of the midgut [32]. Due to restitution of gut epithelium during transition from aquatic to terrestrial environment, there is a limited diversity of microbiota in adult mosquitoes compared with mammalian hosts [30]. Vertical transmission of bacteria in the midgut has been also described in the symbiont *Thorsellia anopheles* [33] and *Asaia sp*. [34]. Analysis of 16S rDNA hypervariable regions V3, V4 and V6 through 454 pyrosequencing showed that *Proteobacteria* is by far the most abundant Phyllum in field mosquitoes, while *Flavobacterium* and *Elizabethkinggia spp.* are predominant in laboratory mosquitoes [30]. These studies have been carried out mainly to characterize the microbiota of the midgut, while the composition of bacteria present in the salivary glands remains somehow unexplored. One study performed in *An. stephensi* demonstrates that 25% of the 16S rDNA copies in salivary glands corresponded to the acetic bacteria *Asaia sp*., which indicates this is the dominant bacterium in the salivary gland of this species in this study [34].

1.5. Microorganisms other than bacteria can successfully colonize Anopheles mosquitoes

Although bacteria is probably the most abundant domain of microorganisms that are able to colonize mosquitoes as symbiont or pathogens; Microsporidia, a phylum of fungi also inhabits the fat body of the invertebrate host and can indeed affect the survival of adult and young stages [35]. The role of Microsporidia in the biology of insects was first described in populations of mosquitoes infected with the *Amblyospora*, which is able to parasite around eight genera of mosquitoes [36][37]. Typically, the life cycle of these parasites includes at least three morphologically different spore stages in the insect host and in the copepod intermediate host through horizontal transmission and alternatively by transovarial infection (horizontal transmission)[38]. The infection takes place mainly in the fat organ of the mosquito nevertheless it is spread through the muscles and periovarian zone [39]. Although *Amblyospora* represents the largest group of Microsporidia which is able to infect mosquitoes, the genera *Brachiola* (*Nosema), Vavraia* and *Edhazardia* have been largely studied mainly due to their potential role controlling mosquitoes populations [40]. Nosematosis, the infection caused by *Brachiola algerae* has been described in *An. stephensi* because it leads to a devastating decrease of survival of adult stages, and also due to the deleterious effects in the development of *Plasmodium* through the sporogonic cycle [41]. It has only one type of dyplokariotic spore, which is able to colonize and infect mosquito cell lines and even mammalian cells [42]. *B. algerae* has been also found in naturally infected populations of *An. gambiae* and *An. albimanus* [43] [44], which suggests a potential application in vector control, nevertheless the infection has been reported mainly as a problem in laboratory mosquitoes colonies. Other species that can be affected include *Aedes aegypti* and *Culex pipiens* [38]. Similar to *B. algerae*, *Vavraia culicis* has wide host-range specificity, being able to successfully colonize *Culex spp*, *Anopheles spp*. and in a limited manner *Aedes spp.* [45]. In contrast to *B. algerae*, *V. culicis* is a multisporous uninucleate parasite which can block the development of *P. berghei* in *An. gambiae* [39], decrease notably the survival rate of young mosquitoes [46] and even prevent the development of resistance to insecticides in a stress-induced manner [47].

The natural host of *Edhazardia aedis* is the yellow-fever mosquito *Ae. aegypti*, which is the only arthropod able to develop transovarial transmission decreasing its reproductive capacity, nevertheless *An. quadrimaculatus*, *Orthopodomyia signifera* and *Toxorhynchites rutilus rutilus* can be infected without affecting the larval progeny [45][48]. The life cycle of *E. aedis* lacks an intermediary copepod host and the transmission is done through spores liberated to the aquatic environment from transovarially infected dead larvae, which are subsequently ingested by non infected larvae, perpetuating in that way the cycle of the fungi [45].

2. AIM OF THE RESEARCH AND SPECIFIC OBJECTIVES

Given the complexity of the microbiota in *Anopheles* mosquitoes, the possibility to identify a non-insect protein responsible for the false positivity in the CSP-ELISA seems to be very unlikely, especially because it is not possible to restrict the analysis to only one phylum, genus or species of microorganisms and also because the biology of the mosquito gut and salivary glands inhabitants is not fully understood.

The main objective of this research is to identify the agent responsible for false positivity in Pf CSP ELISA through four different but related approaches. Therefore, the specific objectives of this thesis are:

- First, to make an *in silico* analysis of Pf CSP through BLAST.
- Second, to describe the bacterial microbiota of *Anopheles* 16s rDNA through NGS.
- Third, to investigate the presence of Microsporidia in false positive mosquitoes through ELISA test and PCR detection.
- Finally, to isolate and identify the protein responsible for false positivity through a protein capture approach.

3. MATERIALS AND METHODS:

3.1. *In silico* analysis of Pf CSP

3.1.1. BLAST analysis of Pf CSP

In order to identify protein sequences with similarity to Pf CSP, BLAST analysis [49] was performed with the complete 442 aminoacids sequence of Pf CSP (Accession Number: AAA29554.1) in non-redundant database excluding the taxa *Plasmodium* available in GenBank through the PubMed website http://www.ncbi.nlm.nih.gov/genbank/ [50]. The scores of each retrieved sequence were calculated with linear gap penalty and an expected threshold of 10. Bacterial and fungal sequences were recorded and manually analyzed in order to identify candidates with similar structure to Pf CSP in the NANP region.

3.1.2. Secondary structure and disorder prediction of NANP repetitions

Secondary structure prediction of protein EJW02257.1 from *E. aedis* was obtained through PSIPRED v3.3 [51] [52]. Since the PSIPRED platform does not allow predictions larger than 1500 aminoacids residues, the sequence was split in two parts, one from residue 1 to residue 1400 and the other from 1401 to 2704. To predict disordered regions, IUPred was used under default parameters [53] [54]. Alternatively, in order to identify favorable interactions with globular proteins ANCHOR method was used [53] [55].

3.2. Mosquito samples preparation

3.2.1. False positive mosquitoes used for molecular analysis

The samples of false positive mosquitoes used for molecular analysis were collected through human landing collection in different villages in Cambodia and Vietnam [17]. We randomly selected 10 out of 142 samples from mosquitoes that were false-positive in CSP-ELISA for *P. falciparum* described by [13]. Specific characteristics of the samples are summarized in Table 1. In addition, and in order to compare microbiota diversity, 2 true negative mosquitoes (negative in CSP-ELISA and PCR) were analyzed together with 1 laboratory-reared negative *An. stephensi* mosquito.

These mosquitoes were processed exactly in the same way as the false positive samples. Six false positive mosquitoes collected in a different study were used in a pool for the protein capture protocol. Characteristics of these samples are shown in Table 2.

All mosquitoes were prepared in the same way. The head/thorax portion, abdomen and legs of each mosquito were separated and put in different tubes. The legs were not used in this study and were stored in silica gel. Head/Thorax and abdomen parts were crashed and lysated with 50 µl of Blocking Buffer-Igepal CA 630 (0.5% Casein technical, from bovine milk Sigma-Aldrich, 0.05% Igepal CA-630and 0.1N Na(OH) in PBS, pH 7.4) using sterile pestles. The pestles were carefully rinsed with 150 µl of Blocking Buffer (0.5% Casein technical, from bovine milk Sigma-Aldrich, 0.05%, and 0.1N Na(OH) in PBS, pH 7.4) to reach a final volume of 200 µl per sample.

3.2.2. CSP-ELISA for P. falciparum

The protocol of Wirtz et al. [13] to detect *Plasmodium* parasites through sandwich ELISA was applied to test the samples. Shortly, 96 wells-ELISA Polyvinyl Chloride plates (Corning® Sigma-Aldrich) were coated with 50 µl of 4 µg/µl capture monoclonal antibodies Pf-2A10-28 (KPL, Gaithersburg, USA) and incubated for 1 hour. After the incubation the liquid in the wells was discarded without washing. Then, 200 µl of Blocking Buffer (BB) (0.5% Casein technical, from bovine milk Sigma-Aldrich, 0.05%, 0.1N Na(OH) in PBS, pH 7.25) was added to the wells and incubated for 1 hour. Then, plates were emptied without washing. In the next step 50 µl of mosquito lysate were added to the plate. Recombinant CSP protein of *P. falciparum* (provided by CDC, Atlanta, USA) was used as positive control in two wells per plate. Negative controls were *An. stephensi* mosquitoes reared in the laboratory in at least five wells. Samples and controls were incubated for two hours and then washed 2 times with PBS-Tween20 0.05% and thawed. After that, 50 µl of 1 µg/µl monoclonal peroxidase labeled antibody (CDC, USA) was added to each well, incubated for 1 hour and washed with PBS-Tween20 2 times. To develop the plates 100 µl of peroxidase substrate (KPL, Gaithersburg, USA) were added and incubated for 30 minutes in dark room.

Color reaction was measured visually from 0 to 3. 0 means no color and 3 strong color reaction (as strong as the positive control).

3.3 Amplification and sequencing of 16s rDNA

3.3.1. In silico Analysis of 16s rDNA primers

To identify bacteria in head thorax of mosquitoes we analyzed five different sets of primers to amplify 16s rDNA of the domain Bacteria (Table 3) [56] [57] [58] [59] [60]. An independent search for highly similar sequences was performed with each pair of primers through nucleotide Megablast [49] into the 16s rDNA sequences for Bacteria and Archaea database through the PubMed website http://www.ncbi.nlm.nih.gov/genbank/[50]. The scores of each retrieved sequence were calculated with linear gap penalty and an expected threshold of 10. The parameters taken into account in this analysis are listed in Table 4 [49]. Considering that the total score represents the sum of scores of the alignment in both primers and the maximum score is a measure of the best alignment of one of the two primers, the difference between the total score and the maximum score was calculated for each sequence. If the value was equal to zero, that means that only one primer was aligned in the target sequence. In that case, the sequence was eliminated, performing in that way an initial cleaning of non-specific sequences. Then, repeated and unmatched sequences were eliminated in order to obtain an accurate database to compare five sets of primers. Three best sets of primers in terms of number of sequences were then analyzed with ANOVA one-way test for differences in alignment parameters, previous elimination of outliers. A general overview of the methodology *in silico* to analyze the five sets of primers is presented in Figure 3.

3.3.2. DNA extraction

To extract DNA from mosquito samples, QIAamp® DNA Micro Kit (Qiagen) for small DNA volumes was used on 20 µl of the lysate of each part of the mosquitoes according to the instructions of the manufacturer. Briefly, 80 µl of ATL buffer was added to each sample to reach a final volume of 100 µl. Then, each tube was digested with 10 µl of Proteinase K and added 100 µl of buffer AL. After short vortex the samples were incubated at 56°C for 10 minutes shaking. After that, the tubes were incubated with

ethanol 100% for 3 minutes at Room Temperature (RT) and briefly centrifuged. The lysates were then transferred to a QIAamp® Minielute column and centrifuged for 1 minute at 800 revolutions per minute (rpm). Washing steps were done with 500 µl of AW1 and AW2 Buffers for 1 minute at 8000 rpm each one. Finally, the membrane was dried for 3 minutes at 14000 rpm and the DNA eluted in 50 µl of Buffer AE in a sterile tube.

3.3.3. Amplification of 16s rDNA from Bacteria

Two out of five sets of primers analyzed *in silico* were tested to amplify 16srDNA of bacteria [56][60][57][58][59]. An adapter sequence was attached at the 5'end of the forward and reverse primers (5' F-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' and 5'R- 'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3') according to the protocol available from Illumina® [61]. Due to best performance *in silico* as well as successful amplification with the adapter sequences, only Herlemann's primers were used for sequencing purposes. All protocols were tested with DNA from *Mycobacterium tuberculosis*, *An. stephensi* lysates and in false positivity samples. Methodology followed for standardization of protocols is depicted in Figure 4 in a TProfesional Thermocycler (Biometra). PCR products were run for 40 minutes in 2% Agarose gels and stained with Ethidium Bromide solution during 20 minutes in order to visualize amplicon in a UV transiluminator InGenius LHR gel analysis and documentation system (Syngene). Photos were stored and analyzed with GenSnap (Syngene).

3.3.4. Identification of bacterial communities through Next Generation sequencing

Amplicons of 24 PCR reactions were sequenced in the MiSeq® System following the manufacturers instructions, using MiSeq® reagent kit v2 for 500 cycles [62] . For sample preparation, Nextera® XT Sample Preparation kit for 24 samples with indexes i5 and i7 was used according the instructions of the manufacturer available in [63] (Annex 1). Due to different species of bacteria were amplified in the same sample, the tagmentation step in the protocol was eliminated. Instead, 5 µl of miliQ water replacing ATM buffer (containing tagmentation enzyme) were added to each sample. A general overview of the sample preparation is depicted in Figure 5 [61]. 30 µl of

denatured and diluted PhiX v3 adapted-ligated library was added to the pool of samples as positive control to reach a total volume of 600 µl in the final library, according the recommendations of the manufacturer [64].

3.3.5. Cleaning of data and workflow

For quality-filtering of sequences retrieved from sequencing, reverse and forward primers were trimmed, and low quality readings were eliminated in such a way that only readings with a Q score higher than 30 were analyzed (99,99% of base call accuracy) [65]. For data analysis, an adapted workflow for end-pair metagenomic analysis from Galaxy metagenomic pipeline was used [66]. Briefly, FASTQ files were transformed to the FASTA format and each sequence was analyzed through BLAST with the Megablast algorithm under default parameters into 16s rDNA database. The best 30 hits in both readings were compared and paired with a default mismatch between species of 1%. The final output is the species identification together with the number of matching reads. In case of one reading (forward or reverse) had the maximum rate of mismatches allowed by the workflow and the other one overcame the threshold of mismatching (1%), only the first one was taken for species identification.

3.4. Detection of Microsporidia

3.4.1. CSP-ELISA to detect cross-reaction with Microsporidia

Cross-reaction of Microsporidia parasites was investigated through CSP-ELISA for *P. falciparum* as described in [13]. For this analysis, *Aedes aegypti* larvae infected with *E. aedis*, mosquitoes and larvae infected with *V. culicis* and pure spores of *V. culicis* (kindly provided by Prof. Yannis Michalakis, Montpellier Research Center on Infection, MIVEGEC) were tested through ELISA CSP for *P. falciparum*. This procedure emulated exactly the methodology to detect *P. falciparum* parasites in field mosquitoes through CSP-ELISA described above.

3.4.2. Detection and identification of Microsporidia *through PCR analysis*

To detect Microsporidia parasites in false positive samples, primers 18F/1492R were used to amplify 18s rDNA according to the protocol described by [38]. Briefly, reaction mixtures were prepared in a total volume per reaction of 50 µl, using 0.25

units of Taq DNA Polymerase (Qiagen), 2.5 µM of each dNTP (Qiagen), 25 pmol of each primer, 0.5 µl of Bovine Serum Albumin (BSA), 5 µl of 10X Buffer Qiagen and 5 µl of the DNA extraction. PCR conditions were as follows: 94ºC (3 minutes) and 35 cycles of 94ºC (45 seconds), 45ºC (30 seconds) and 72ºC (90 seconds). PCR products were run 30 minutes in 1% Agarose gels and stained with Ethidium Bromide solution during 20 minutes in order to visualize amplicons. *An. stephensi* infected with *Nosema sp*. (from a disseminated infection of a mosquitoes colony identified though microscopy) as well as *V. culicis* spores were used as positive controls. MiliQ water was used as negative control. Species identification was carried out by sequencing the PCR products of positive samples (VIB, Antwerp, Belgium).

3.4.3. Sequence analysis of Microsporidia*:*

Sequences of positive samples and positive control were aligned through CodonCode Aligner. In short, forward sequence was aligned with reverse-complement of the reverse reading, low quality-readings were automatically eliminated and a consensus sequence was obtained. These sequences were blasted in GenBank database with linear gap cost and an expected threshold of 10.

3.5. Identification of protein responsible for the false positivity

3.5.1. Coupling reaction with P. falciparum A210 Monoclonal antibodies.

In order to isolate a protein that cross-reacts with monoclonal antibodies against *P. falciparum,* we used Dynabeads® M-270 Epoxy Antibody Coupling Kit (Life Technologies-Invitrogen) according to the instructions of the manufacturer [67]. Briefly, capture monoclonal antibodies 2A10 (CDC, USA) were dissolved in 1ml of distillated autoclaved UV-irradiated miliQ water. The solution was centrifuged and the supernatant was collected. 12.5 mg of Dynabeads® M-270 Epoxy were weighed in separate sterile Eppendorf tubes and 1ml of C1 buffer was added to each tube, vortexed for a few seconds before elimination of the supernatant. The amounts of monoclonal antibodies, C1 and C2 buffers for the next coupling steps were calculated according the manufacturer recommendations [67]. The tubes were then placed horizontally in a shaking incubator for 16 hours at 37ºC prior to elimination of the supernatant. Washing steps were performed with 0.8 µl of buffers HB, LB (once each

one) and SB (twice) in this order. Finally the pellet was resuspended in SB buffer adjusting a concentration of 30 mg of beads/ml, which represents one third of the total capacity of coupling of the beads.

3.5.2. Immunoprecipitation

Coupled beads were tested for their capacity to bind Pf CSP prior to sample testing. Firs, two sterile Eppendorf tubes were blocked with 900 µl of BB for 60 minutes. Then, 20 µl of coupled Dynabeads® were added to each tube and washed 2 times in BB. 200 µl of BB were used as negative control while a mixture of 150 µl of BB and 50 µl of recombinant antigens provided by CDC Atlanta, USA (2 pg/µl) was used as positive control. The tubes were placed in a horizontal incubator for 90 minutes at RT. Supernatant was collected and the beads were washed three times with 900 µl of BB. Three extra washing steps with PBS-Tween20 0.5% were performed, the first one with an incubation of 5 minutes and the following 3 without incubation. In the next step, 200 µl of conjugated antibody (1:1000) were added to each tube and incubated in a rotator for 60 minutes. Supernatant was discarded and the tubes were washed 3 times in 900 µl of PBS-Tween20, 100 µl of substrate solution was added followed by incubation at dark room of 30 minutes. The liquid was changed to an ELISA plate and read with a filter of 405nm. For the analysis of the samples, the same methodology described above was used with the difference that the procedure was performed in six tubes using 250 µl of coupled Dynabeads®. One tube was resuspended with 400 µl of lysate of false positive samples and two tubes were used as positive control (350 µl of BB + 50 µl of thorax/head portion lysate of one *An. gambiae* mosquito, kindly provided by Wagenigen University) and negative control (400 µl of BB). 70 µl of each tube were separated to perform ELISA as indicated above in the remaining three blocked tubes. The residual 830 µl of supernatant was discarded and the beads were resuspended in 20 µl of PBS and 20 µl of loading dye (0,125M Tris HCl pH 6.8, 4% Sodium Duodecyl Sulphate, 20% Sucrose and 0,04% Bromophenol Blue). This mixture was boiled and shaken at 95^{o}C for 15 minutes. The tubes were placed in a magnet and the supernatant was stored in new sterile tubes for further analysis with Western Blot and Coomassie Blue staining.

3.5.3. SDS-Polyacrylamide gel and Western Blot

20 µl of the protein isolated through immunoprecipitation were loaded in two different wells in a precast gel for Polyacrylamide electrophoresis (Ready Gel, Bio-Rad). The gel was placed in an electrophoresis chamber (Mini-PROTEAN Tube Cell #165-2961, Bio-Rad) filled with 800 ml of Buffer Tris-Glycine-SDS (Bio-Rad GmbH, Heidemannstraβe, Munchen, Germany) and run for 50 minutes. The gel was cut in two parts, one for Western Blot and the other one for Coomassie Blue staining. For Western Blot, the proteins were transferred to a Protran™ nitrocellulose membrane (Whatman) in a Mini Trans-Blot® Cell and Criterion™ Blotter (Bio-Rad) filled with transfer buffer (25 mM Tris, 192 mM glycine, pH 8.3, 20% methanol) with a steering magnet and an block of cold ice for 1 hour. Once the protein was transferred to the membrane, it was blocked with BB for 1 hour and washed three times in PBS for 5 minutes in an agitator. The membrane was then submerged in 50 ml a solution containing 100 µl of Pf2A10 monoclonal antibodies (2pg/µl) in 50 ml of PBS and incubated for 1 hour in an agitator. Next, the membrane was washed 3 times for 5 minutes in PBS and then the secondary antibody, rabbit anti-mouse IgG1-Horse Radish Peroxidase (Abd Serotec, Bio-Rad) 0,1% was added and incubated for 1 hour in agitation. After three washing steps with PBS, developer solution (18mg HCN, 6ml methanol, 30 ml PBS and 18 µl of H_2O_2) was added and incubated for 15 minutes prior to reading the pattern of bands.

3.5.4. Coomassie Blue staining of the protein

For Coomassie Blue staining the gel was fixated overnight in a 200 ml of a solution containing 50% Methanol (MERCK, ACS grade), 100mM ammonium acetate and 10% acetic acid. Then, the gel was stained with 0.025% of Coomassie dye in 10% acetic acid for 1 hour. After that, 5 distaining steps were done with 50 ml of 10% acetic acid. The gel was put between two transparent plastic sheets and scanned in a photocopier machine. Those bands that were present in the sample lane and not in the control lanes were cut and stored in separated sterile Eppendorf tubes at -20^{o}C.

3.5.5. Matrix-assisted laser desorption/ionization Time of Flight MALDI TOF/TOF analysis

MALDI TOF/TOF was performed on the cut bands in a 4800 MALDI TOF/TOF™ Analyzer (Applied Biosystems) according to the manufacturer indications [68]. Briefly, 2 bands cut from Coomassie Blue staining were washed in 100 µl of Pestanal Grade water in a 96-well plate and shaken for 5 minutes. The water was removed and the plugs were shrink twice in 100 µl of 95% acetonitrile (Pestanal grade) to eliminate the water. The plugs were then embedded in dithiotretiol (6.66mM DTT, 0.05M NH_4HCO_3) and incubated at 56°C for 45 minutes. The liquid was discarded and the plugs washed in 100 µl of 95% acetonitrile (Pestanal grade) twice for 5 minutes. Next, the plugs were embedded in iodoacetamid (55mM IAM, 0.05M NH_4HCO_3), incubated at RT in the dark and washed in 100 µl of 95% acetonitrile twice for 5 minutes. Protein was digested with 20 µl of digest buffer composed by 10 µg of Trypsin (Promega) in 50mMCH_3COOH mixed with 790 µl 0.05 NH_4HCO_3, 10% CH_3N on an ice block. The plate was placed at 4°C for 30 minutes to allow the gel plugs to swell properly. Next, 10 µl of replacement buffer (1ml 0.05M NH_4HCO3, 10% CH_3CN) was added to the plugs, incubated 3.5 hours at 37°C and stored at -20°C.

Matrix of α-cyano-4-hydroxycinnamic acid in a large spatula tip was washed twice in 1ml of water and briefly vortexed and centrifuged. The matrix was then mixed with 600 µl of acetonitrile and vortexed. 63 pmole of Glu-1-Fibrinopeptide B Peptide Mass Standard (PROTEA BIOSCIENCES) was used as calibrator. Crystallization of the matrix was checked observing crystallization of 1200nl of the matrix after drying at 42°C in a maldi plate (Applied Biosciences). To spot the matrix and the sample ZipTip® Pipette Tips (Millipore) were loaded with 20 µl of acetonitrile 95% to activate the coating and then eluted. The activated tip was then loaded with the sample pipetting 4-5 times to allow binding of peptides. The tip was finally desalted with 10 µl of Water LC-MS CHROMASOLV® (Sigma-Aldrich), charged with 1.2 µl of matrix and spotted in the maldi plate. Alignment of sequences retrieved by MALDI TOF/TOF analysis was done through CLUSTALW2 [69] [70].

3.6. Data analysis

Sequences retrieved from BLAST for *in silico* analysis of primers were analyzed with Microsoft® Excel and Microsoft ®Access in order to eliminate repeated and unmatched sequences. Matching and comparison of sequences were done through Accession numbers of each sequence. For statistical analysis of the best three sets of primers selected based on number of sequences, one-way ANOVA test was used in each of the five parameters listed in Table 4 through StatPlus™. For metagenomic data analysis, FASTA and FASTQ files were converted to Microsoft® Excel worksheets and comparisons between them were performed in Microsoft® Access.

4. RESULTS

4.1. In silico analysis of Pf CSP

BLAST analysis of Pf CSP resulted in 81 sequences with significant alignments. 77% of the sequences belonged to the Kingdom Bacteria; 21% to Eukarya and 2% to Fungi (Table 5). The detailed list of sequences is presented in Annex 2. Although the majority of sequences belong to Prokarya, one sequence of a hypothetical protein without known function from *Edhazardia aedis* (Accession Number: EJW02257.1) contains a NANP repetition between the residues 1605-1962 containing 23 residues NANPL overlapped with 24 residues NANP. In contrast to Pf CSP the NANP repetition is not continuous and it spans in a segment of around 357 residues.

4.2. Secondary structure and disorder prediction of EJW02257.1

Secondary structure prediction through PSIPRED of the two parts of the above-mentioned protein demonstrated only few segments with secondary structure. The NANPL repeats did not show secondary structure (Figure 6). The repetitive region demonstrated also a high degree of disorder as predicted by IUPred (Figure 7). Anchor method to identify segments to fold upon binding to globular protein indicates that the NANPL region has an interaction prone behavior that allows this segment to bind with other proteins (Figure 8).

4.3. *In silico* analysis of primers for Bacteria

The number of sequences that matched with each pair of primers before and after preliminary cleaning is listed in Table 6. Only 1346 sequences were common to 5 sets of primers analyzed, which clearly indicate differences between them. After initial cleaning, repeated sequences were identified and eliminated in order to obtain a real number of sequences specifically matched in each set of primers (Table 7).

Based on these data, it was evident that the maximum number of sequences aligned with the primers corresponded to the pair 356F'/1064R', 968F'/1381R' and

Bakt_341F/Bakt_805R'. The amount of sequences recognized by the three sets of primers was significantly higher than Lane and Dojka primers.

Sequences were then analyzed through one way-ANOVA test in order to describe what set of primers showed better properties in terms of alignment scores. Cross match analysis between primers from Winsley, Roudiere and Herlemann showed that they are able to recognize 6952 different sequences, although with differences between the parameters analyzed. Best scores were found in Herlemann's primers (Table 8).
A detailed analysis of the sequences retrieved with Roudiere's primers showed 112 sequences representing outliers that severely skewed the distribution of the sequences. These 112 outlier-sequences had e-values equal or above 1. To be able to compare more homogenous sequence data sets these sequences were eliminated from the analysis together with 586 high e-scores in Winsley's primers. This restriction reduced the data set to 6297 sequences. This step was not done in Herlemann's primers due to all the sequences showed e-scores less than 1(Table 9).
In addition, in order to calculate an expected size of bands in each sequence, an independent BLAST analysis of each primer was carried out adjusting the search for short length sequences. The average expected size of amplicons in Herlemann's primers was 437.85 bp (SD: 92.62), while in Winsley's primers the size of the amplicon was 668.550bp (SD: 149.464).

4.4. PCR amplification and sequencing of V3-V4 regions of Bacteria

PCR of 16s rDNA in samples with Winsley's protocol showed an amplified segment of 700 bp (regions V6-V9) [57] and for Herlemann's protocol the size of the amplicon was 500 bp [59] (regions V3-V4) . Amplification with Winsley's primers attached to adaptors was not successful; consequently only regions V3-V4 amplified with Herlemann's primers were sequenced.

Sequencing of samples through Illumina MiSeq ® instrumentation showed a total yield of the sequencing reaction of 7.8 Gb with 79.2% of readings with a Q score higher than 30. Cluster density was 1,160 +/- 48 and 84.79% of indexes were successfully identified in all the samples (Minimum percentage: 1.32 and Maximal

percentage: 8) (Table 10). PhiX v3 adapted-ligated library showed an alignment of 5.42%.

After quality filtering, a total of 5,516,982 readings were successfully identified in 16s rDNA database of GenBank. The number of readings in each sample as well as in controls is depicted in Figure 9.

4.5. Microbial diversity

In order to analyze bacterial microbiota, samples were divided as follows: False positives Abdomen (FPABD); False positives head/thorax (FPHT); True negatives (TN), *An. stephensi* head/thorax (STHT) and *An. stephensi* abdomen (STABD). Number of readings in each group of samples was: FPADP = 2,283,753; FPHT = 2,383,875; TN = 582,154; STABD = 156,645 and STHT 110,555. To summarize the data in this specific analysis, readings with a percentage less than 1 with respect the total amount of readings in the subset of samples was classified as Others. A detailed list of the species found in each group is presented in Annex 3. It is important to note that one single reading can pair with more than one sequence in 16s rDNA database, consequently several possibilities of species can be possible for one single reading.

Bacterial microbiota in FPABD was composed predominantly by *Pseudomonas libanensis* (31.21%)and *Stenotrophomonas rhizophila* (17.27%) (Figure 10). There was a clear predominance of *Pseudomonas sp.* in this group of samples. This taxa represents more than 50% of the bacterial composition in the abdominal portion of FPABD. 4,389 different species were identified in this group.

In FPHT a more diverse species composition was found, nevertheless it was also evident that the genera *Pseudomonas* and *Stenotrophomonas* are the most abundant bacterial genera in head/thorax. Contrary to FPABD the group of microorganisms assigned as Others in FPHT, represents 31% of total diversity compared to only 18,74% in FPABD (Figure 11). Diversity in terms of species occurrence was higher than FPADB (4,799 species identified).

The most abundant species in TN group was the Gram-Negative bacteria *Herbaspirillum rhizosphaerae* strain UMS-37 (17,38%). *P. veronii* occupied the second place in abundance of readings (16,86%) followed by *Undibacterium pigrum*. (Figure 12) 2,281 species were identified within this group.

STABD showed a more homogenous bacterial composition with *S. rhizophila* as the most abundant species (54.4%), followed by *Acinetobacter schindleri* (5.52%) and *Propionibacterium acnes* (4.17%) (Figure 13). 1,751 species were identified in this group.

Finally, bacterial composition on STHT was composed mainly by *S. rhizophila* (37.41%); *P. acnes* (11.69%) and *Staphylococcus capitis subsp. Urealyticus* (7.82%) (Figure 14). In this group 1,251 species were identified.

The common core in all groups was obtained to have a broad overview of bacterial composition in all five groups. This core was composed by 511 species belonging to 73 different genera (Table 11). A detailed list of species of this core is presented in Annex 4.

4.5.1. Identification of possible candidates responsible for false positivity

A paired analysis of samples was carried out in order to compare microbiota in different segments of each mosquito. First, a list of species was obtained from head/thorax and abdomen of each mosquito. Second, both lists were cross-matched in order to identify species that occurred only in head/thorax portions. Then, each list was again compared with a species pool found in STHT, STABD and TN (Annex 5). This analysis allowed identifying species that occurred only in head/thorax portions and not in any of the controls analyzed in each mosquito (Annex 6). These lists were then compared in order to obtain bacterial species that were present from 1 up to 10 samples analyzed. Under these parameters the maximum number of hits was 7 out of 10 samples for *Atopococcus tabaci* strain :CCUG 48253 and *Sphingopyxis witflariensis* strain W-50. There were not bacterial species present only in head/thorax portion in all 10 samples analyzed and not in controls. The complete list of species found only in

head/thorax portions in False Positive samples obtained through paired analysis is presented in Annex 7.

4.6. Cross-reaction of Pf CSP with Microsporidia

Ae. aegypti infected with *E. aedis*, larvae and mosquitoes infected with *V. culicis* as well as pure spores of *V. culicis* were negative in Pf CSP ELISA.

2 out of 10 false positive samples were positive for PCR analysis to Microsporidia. One was positive in the head/thorax portion as well as in the abdomen. The other one was positive only in the head/thorax portion. From two controls tested, only the mosquito infected with *Nosema sp*. was positive.

From 4 samples with amplification of 18s rDNA in the Microsporidia PCR assay, 3 amplicons were successfully sequenced. One of the alignments corresponded to the mite *Limnesia sp*. The rest of sequences belonged to the Fungi Kingdom. Best alignment scores for each sample are presented in Table 12.

4.7. Isolation of cross-reactive protein through Dynabeads.

The coupling test of Dynabeads® with monoclonal antibodies against Pf CSP resulted in a weak color reaction according to OD reading. Positive control had a reading of 0,257. Negative control had a reading of 0,077. OD readings were as follows: Positive control= 2,571; negative control =0,076, and pool of false positive samples tested = 0,071. The very low OD in the reading indicates the extremely small amount (if any) of cross-reactive protein present in the samples.

4.9. Western Blot and SDS analysis.

Western Blot analysis did not show any band in the samples. Only a 50KDa band was observed in the three lanes, which corresponds to the heavy chain of immunoglobulin G from the monoclonal antibodies (Figure 15).

Coomassie Blue staining on the other hand showed two bands that were not present in the positive and negative controls (Figure 16), which were submitted to MALDI TOF/TOF analysis.

4.10. MALDI TOF/TOF Analysis

Two bands isolated from the Coomassie blue staining were successfully identified through MALDI TOF/TOF analysis. The list of peptides with significant protein score is presented in Table 13. The heaviest band with an approximate mass of 250KDa showed high similarity with the myosin protein of several organisms. The lightest band with an approximate mass of 45KDa was related with actin protein. Interestingly, sequences retrieved for the lightest band showed correspondence with BLAST analysis of Pf CSP (accession numbers: gi|187282496, gi|260785919 and gi|156388857). Alignment through CLUSTALW2 was done in both sets of proteins together with the complete sequence of CSP *P. falciparum* protein in order to detect conserved residues or regions of high similarity to CSP of *P. falciparum*. For the lightest protein, CLUSTALW2 showed conserved aminoacids in positions 289 (P), 297 (N), 313 (P), 329 (N), 367 (K) and 372 (L). The NANP repetition, which spans from residues 134 to 318 did not show a clear pattern of similarity with the proteins identified through MALDI TOF/TOF in its whole extension, nevertheless residues located in positions 282 to 381 demonstrated a cluster of similarity with CSP *P. falciparum* protein (Figure 17). Furthermore, a restricted alignment of gi|187282496, gi|260785919, gi|156388857 and Pf CSP showed several conserved residues dispersed in the whole protein, although there was no conclusive evidence of a region with a high degree of similarity to the NANP repetition (Figure 18).

Alignment through CLUSTALW2 of the heavy protein with the Pf CSP did not show conserved residues.

5. DISCUSSION

5.1. Pf CSP structure and sequence analysis

Although the structure of Pf CSP has been previously defined [22], similarity of a protein that could cross-react with monoclonal antibodies in CSP-ELISA has not been described so far. BLAST analysis of Pf CSP demonstrated that the NANP/NVDP repetition in Pf CSP is a unique and distinctive feature of the parasite. The absence of similarity with other protein in any microorganism, other than *Plasmodium sp*. together with the high immune response elicited by the NANP repeat is a key component of vaccine development against *P. falciparum* and has been broadly reviewed by [71] and [72] . Based on substitution analysis of peptides, it is well known that binding of monoclonal antibodies to Pf CSP is determined by recognition of three kinds of peptides: NANP, NVDP and NDPD [73]. These repeats are located in different parts of the Pf CSP according to the Table 14 [21]. Among these short sequences the most similar one which was found in BLAST analysis in the present study was the repetition NANPL, from the Microsporidia *E. aedis*, which spans in a discontinuous manner through a protein of unknown function (Accession number EJW02257.1). To our knowledge this is the first time that a similar sequence to the NANP repetition is identified in a parasite of *Anopheles* mosquitoes. *E. aedis* has as its primary host the Dengue fever vector *Ae. aegypti* [74], nevertheless the colonization of *An. quadrimaculatus* is possible through horizontal infection of larva, as demonstrated in [75]. Considering that false positivity has been described only in the head/thorax portion of Anopheles mosquitoes [17], our findings based on PCR detection of Microsporidia together with negative results in Pf CSP-ELISA carried out in mosquitoes and larva infected with *E. aedis* and *V. culicis*, can not confirm Microsporidia as the responsible of false positivity. Moreover, one of the samples sequenced was identified as *Limnesia sp*., a mite that is able to colonize *An. implexus* as described in [76]. This finding can be explained by the report of cross-amplification of arthropod DNA [77].

Prediction of secondary structure of the hypothetical protein EJW02257.1 indicates that the NANPL region does not have a defined secondary structure to be compared to NANP repetition of Pf CSP which exhibits a stem-like structure composed by β turns [22] [78]. In addition to these results, disorder prediction of EJW02257.1 demonstrated that the NANPL region is the most disordered part of the protein. In this sense it is important to note that although NANP region of *P. falciparum* has 43 repetitions, they are located somewhat dispersed between residues 134-317 and interrupted by NDPN and NVDP sequences. On the contrary, EJW02257.1 exhibits an almost perfect repetition of NANPL residues in positions 1605-1962. This finding corroborates previous research indicating that perfect repetitions are usually related with high degree or disorder and cytoplasmic localization [79]. In agreement with the high level of disorder, ANCHOR prediction indicates that NANPL region has a high tendency to bind globular proteins, although the real function of the protein remains unknown.

On the other hand, Pf CSP BLAST analysis demonstrates some degree of similarity with bacterial proteins particularly from *Streptococcus agalactiae*, an encapsulated facultative anaerobe Gram Positive bacteria responsible of infections in pregnant women [80]. In *Anopheles* mosquitoes, *S. agalactiae* has been previously described in the gut of *An. funestus*, *An. gambiae* and *An. stephensi* [81], nevertheless its role in the gut homeostasis or possible effects in the transmission of malaria have not been investigated. From the rest of bacteria species identified in BLAST analysis of Pf CSP only the genus *Enterococcus* has been identified in *Anopheles* [81].

Eukaryal sequences were also found in the BLAST search and they correspond to a wide range of organisms that ranges from simple invertebrates such as the lancelet *Nematostella vectensis* to complex vertebrates like the yak *Bos grunienns mutus*. These somehow odd findings arise the question if a cross-reaction protein from animals which *Anopheles* mosquitoes use as blood source can indeed result in false positivity in the CSP-ELISA . This possibility has been addressed by [82] and [19], nevertheless not all mosquitoes fed with bovine or swine blood result positive in CSP ELISA. Due to the reasons briefly explained above, interpretation of the results of BLAST search of

Pf CSP alone are inconclusive, remarking the necessity of an experimental approach that could give some insights to the *in silico* results.

5.2. 16s rDNA as a reliable tool to analyze complex bacterial communities

The usefulness of 16s rDNA as a marker of diversity of bacteria relies on the high degree of conservation among phyla and the possibility to easily amplify fragments of the gene through PCR [83]. The first approach developed to analyze bacterial sequences of 16s rDNA was the Denaturing Gradient Gel Electrophoresis (DGGE), which is able to separate sequences with a resolution of one single nucleotide [58]. The fast and gripping development of new sequencing technologies such as pyrosequencing and next generation sequencing (NGS) has left apart this technique, although it has remained as a genotyping tool particularly in food science [84] [85]. One of the most difficult aspects of 16s rDNA analysis is the selection of the region to be amplified. Since 16s rDNA is composed by several regions, namely V1 to V9 with an approximate size of 1.5Kb [86], the advantages of amplifying different segments of the gene have been extensively studied in different bacterial habitats such as soil, human gut and sea environments.

The number of sequences retrieved by each set of primers explored in this thesis showed significant differences among them, supporting previous studies indicating that the selection of oligonucleotides in the study of complex microbial communities is critical to obtain an accurate representation of genus and species [87][88]. A considerable amount of research has been done in the last years in order to find the most appropriate set of primers in metagenomic studies, nevertheless there is no consensus in one single pair of primers that can be used for different bacterial communities [89]. The best approach must be a careful selection based in the phyla expected in each environment that could avoid the possible bias that some regions could exhibit in a given bacterial community [90]. Our research indicates that Herlemann's primers are superior in terms of number of sequences and genus identified *in silico*. The region amplified by this set of primers corresponds to V3-V4, which has being postulated together with V6 as the most accurate regions in metagenomic studies [91]. In agreement with other studies the classical Lane's

primers used to amplify V3-V6 16s rDNA demonstrates low capacity to recover a maximum amount of sequences and genera, probably because the recent update in the ribosomal sequences available in GenBank [57] [90].

5.3. Diversity of species

Although the sequencing reaction was overclustered (optimal density of cluster is 800 K/mm^2), the bacterial composition in the five groups demonstrated clear differences between them, which are explained by the dissimilar settings they come from. In FPABD the most abundant genus identified was *Pseudomonas*, a Gram-Negative, Oxidase positive non *Enterobacteria* that has been previously described as a common microorganism in the midgut of *An. stephensi* and *An. maculipennis* from the field [32]. The other important species identified in this study correspond to *Stenotrophomonas rhizophila* and *Pantoea dispersa*, both previously described in *An. funestus* and *An. gambiae sensu lato* [92] [93]. *Propionibacterium acnes* is an anaerobic Gram-Positive bacteria that has previously detected in *An. gambiae* playing a roll in the oxidative stress induced by blood meal feed [31]. In comparison to FPHT the relative abundance of each species was slightly different although *Pseudomonas* and *Stenotrophomonas* dominate the bacterial community in this part of the mosquito. Interestingly, the amount of bacteria grouped as Others in FPHT was higher than in FPABD, showing a higher species diversity in this location. This group includes mainly Gram-Positive such as the spoiled-food *Brochothrix thermosphacta* [94] as well as *Staphylococcus, Streptococcus* and *Lactobacillus*. The role of this group of microorganisms in the physiology of the *Anopheles* mosquito has not been investigated.

Interestingly the main species present in TN was the Nitrogen-fixing bacteria *H. rhizosphaerae*, a common inhabitant of ground water [95]. To our knowledge it is the first time that this bacteria is described as an important component of the microbiota in *Anopheles*. The reduced diversity of TN could be explained by the fact that only two mosquitoes were analyzed in this group.

In the laboratory reared *An. stephensi*, the composition of bacteria was notoriously different than in the other groups (Figure 14). *Pseudomonas* is not the predominant genus in this group, which could be explained by the fact that this group came from a laboratory environment. In this group, *Stenotrophomonas* was by far more abundant than the rest of bacteria. In contrast to the report of [30], *Elizabethkingia* was not found in the samples analyzed in the present study.

The recent description of microbiota in anopheline mosquitoes in Thailand and Vietnam through TTGE by [96] provides an excellent reference to compare microbial diversity in our study. From 31 genera of bacteria described by [96] only *Diplorickettsia* and *Chromobacterium* were not found in the present study (Table 15). *Elizabethkingia* was not found in TN, although it was present in FPABD and FPHT. In our study a total of 1301 genera were identified in all five groups. This data demonstrated the outstanding powerful of NGS compared with classic genotyping tools such as TTGE.

An important difference found in our study is the fact that *Asaia*, a genus widely described as symbionts, and postulated as a good candidate for paratransgenesis in the control of malaria [34] [96] was not found as a dominant species in any of the groups analyzed. Nevertheless, it was present in FPABD, FPHT and TN.

Although the scope of this study is not to give a detailed analysis of every genus or species found, it is important to mention the incidental finding of *Rickettsia felis* in a sample from Pang Rolim (Cambodia) (sample 6HZG0033T *An. minimus*). *R. felis* has been detected through qPCR in *Ae. albopictus* from Gabon [97] suggesting a possible role in the transmission and epidemiology of rickettsiosis in Africa. More recently, infected *An. gambiae* from Côte d'Ivoire, Gabon, and Senegal indicate that *Rickettsia* is indeed an emerging pathogen that could be transmitted through mosquitoes [98]. To our knowledge, this is the first report of Rickettsia in *Anopheles* mosquitoes from Asia.

5.4. Possible microorganisms identified through Next generation sequencing responsible of false positivity

Paired analysis of samples did not show genera or species present only in head/thorax in all the samples analyzed, nevertheless 1775 species belonging to 595 genera of bacteria were identified as unique in FPHT (Annex 7). The most frequent bacteria found in all the samples were *Atopococcus tabaci* strain :CCUG 48253 and *Sphingopyxis witflariensis* strain W-50. Both were found in 7 out of 10 paired samples. The first one is a Gram–positive contaminant of tobacco closely related to the *Clostridium* subphyla [99]. The second one is a Gram-Negative bacteria isolated from wastewater from Germany and phylogeneticaly related to the genus *Sphingomonas* [100]. None of them have been previously described as pathogens or symbionts of *Anopheles* mosquitoes. Other microorganisms frequently identified in the samples include the genera *Lactobacillus, Sphingomonas* and *Streptococcus* (6 out of 10). The genus *Lactobacillus* has been previously described in *Drosophila melanogaster* mainly due to the rol of *L. plantarum* in the systemic growth of the host [101]. Its role in the physiology of *Anopheles* mosquitoes has not been investigated. *Sphingomonas* has been detected in *An. maculatus* and *An. dirus* from Vietnam [96]. Due to the taxonomic similarity with *Sphingopyxis*, this group of microorganisms represents a promising candidate of false positivity in this study. On the other hand, the finding of *Streptococcus dysgalactiae* in 6/10 samples together with the BLAST search of Pf CSP that indicate some degree of similarity with a protein of *S. agalactiae* also provides some degree of evidence that this genus could be implicated in false positivity. These findings indicate that a most specific approach to detect a limited group of bacteria based in the preliminary list provided in this thesis is needed to confirm these results. The data in this thesis also imply that one single species of bacteria can not be implicated as responsible of false positivity. However, it is important to note that number of readings in each sample was not homogenous and consequently underestimation of species in some samples could be possible (Figure 9). Even in this scenario the question that arises is if false positivity could be the result of not only one species of bacteria but maybe different genera or species that are present scattered in a population of mosquitoes. If that is the case and with such a high number of species

that can be incriminated as responsible of false positivity, new approaches to investigate the presence of the selected genus/species in a more detailed and specific manner are needed to elucidate the most likely species that can be responsible of false positivity.

5.5. Protein analysis and microorganism identified as possible responsible of false positivity.

MALDI TOF/TOF analysis of protein extracted from false positive samples showed two proteins with cross reactivity with monoclonal antibodies against Pf CSP. The heaviest protein corresponded to myosin protein; nevertheless no conserved residues compared to Pf CSP were detected, indicating that further experiments evaluating cross-reactivity with this protein need to be done to explain this finding. The light protein peptides identified showed correspondence with three sequences from *S. purpuratus*, *B. floridae* and *N. vectensis*. The first organism is a sea urchin and the peptide sequence is related with muscle actin [102]. *B. floridae* is a lancelet that has been studied mainly for evidence of phylogenetic relationship between tunicates and vertebrates [103]. Lancelets are also a source of food in some parts of Asia where they are harvested for human and animal consumption [104], which could somehow explains the finding of this protein in *Anopheles* mosquitoes captured in the field. The sequence of *B. floridae* codifies to actin protein. Finally, the sequence of *N. vectensis*, a starlet sea anemone, represents also the protein actin. A closer analysis of these proteins aligned with Pf CSP, showed several conserved residues that could indicate some degree of similarity and somehow explain recognition of monoclonal antibodies. Nevertheless, none of the organisms is related with the biology of *Anopheles* mosquitoes, although they share a key characteristic, which is the identification of actin as the common protein. Actin is a ubiquitous protein of eukaryotic cells and although it possess a high degree of conservation among phyla, it is codified by a supergene that contains at least six isoforms of the protein [105]. It is hard to explain the real meaning of these results, but what seems clear is that monoclonal antibodies recognize a protein from the family of actin, which is present in eukaryotic cells. This statement would support the hypothesis that false positivity is related with animal blood feed [106] [82]. Although actin is a very conserved

protein among Eukarya, slight differences in its structure can be crucial in the recognition of monoclonal antibodies in Pf CSP ELISA, consequently this hypothesis cannot be discarded as an explanation of false positivity. Actin is also a very dynamic protein, which interacts with several components of the eukaryotic cell to maintain its shape and regulate cellular movement [107], that means that the protein can interact with other elements of the cell particularly divalent cations (Mg^{2+} and Ca^{2+}) that could ultimately change the antigenic structure of the protein [108].

6. CONCLUSIONS

The determination and study of false positivity in Pf CSP ELISA is challenging and although this research provides important insights regarding the etiological agent responsible of false positivity, its nature remains unclear. Upon the result of our investigation there are two key aspects to be considered: the first one is the evaluation of NGS as a powerful tool to investigate microbiota in *Anopheles* mosquitoes. The results of this investigation provide a wide but more straightforward scenario to further investigate the hypothesis that a bacterium (or a group of them) is responsible of false positivity. A detailed and dedicated analysis of bacteria found only in head/thorax portions of the mosquito *in silico* and *in vivo* could give more clues to confirm that the causal agent of false positivity is from bacterial origin. On the other hand, a fine-tune standardization of the methodology to sequencing through Illumina® platform is needed. This is necessary in order to eliminate possible drawbacks (such as overclustering) that could limit the diversity obtained through NGS. NGS offers as well the possibility to replicate the approach used in this thesis with other microorganism like fungi, virus or ectoparasites that could be present only in head/thorax portion of *Anopheles* mosquitoes.

The second aspect to be addressed is the cross-reactivity of actin with monoclonal antibodies against Pf CSP, a finding that could be further investigated testing monoclonal antibodies against Pf CSP for cross-reaction with purified actin protein. This experiment can elucidate in a better way the findings *in silico* and from cross-reacting protein analysis performed in this thesis.

7. INDEX OF TABLES

8. INDEX OF FIGURES

9. REFERENCES

1. Miller LH, Mason SJ, Clyde DF, McGinniss MH: **The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy.** *The New England journal of medicine* 1976, **295**:302–4.

2. **Duffy Phenotype and Plasmodium vivax infections in Humans and Apes, Africa - Vol. 18 No. 10 - October 2012 - Emerging Infectious Disease journal - CDC**. .

3. **WHO | World Malaria Report 2012**. .

4. Crutcher JM HSL, Baron S editor: **Malaria**. In *Medical Microbiology 4th edition* . Galveston (TX): University of Texas Medical Branch at Galveston; 1996.

5. Zollner GE, Ponsa N, Garman GW, Poudel S, Bell JA, Sattabongkot J, Coleman RE, Vaughan JA: **Population dynamics of sporogony for Plasmodium vivax parasites from western Thailand developing within three species of colonized Anopheles mosquitoes.** *Malaria journal* 2006, **5**:68.

6. Hillyer JF, Barreau C, Vernick KD: **Efficiency of salivary gland invasion by malaria sporozoites is controlled by rapid sporozoite destruction in the mosquito haemocoel**. *Int J Parasitol* 2007, **37**:673–681.

7. Phillips RS: **Current status of malaria and potential for control.** *Clinical microbiology reviews* 2001, **14**:208–26.

8. Dekker T, Takken W, Braks MA: **Innate preference for host-odor blends modulates degree of anthropophagy of Anopheles gambiae sensu lato (Diptera: Culicidae)**. *J Med Entomol* 2001, **38**:868–871.

9. Trung HD, Bortel W V, Sochantha T, Keokenchanh K, Briet OJ, Coosemans M: **Behavioural heterogeneity of Anopheles species in ecologically different localities in Southeast Asia: a challenge for vector control**. *Trop Med Int Health* 2005, **10**:251–262.

10. Doolan D, Beier J: **Vector Incrimination and Entomological Inoculation Rates**. In *Malaria Methods and Protocols*. *Volume 72*. Humana Press; 2002:3–11.

11. Beier JC, Koros JK: **Anatomical dissemination of circumsporozoite protein in wild Afrotropical Anopheles affects malaria sporozoite rate determination by ELISA**. *Med Vet Entomol* 1991, **5**:81–85.

12. WHO: **Part II. Methods and Techiniques.** . In *Manual of Practical Entolomogy on Malaria*. Edited by diseases WHO division of M and other parasitic. Geneva: WHO; 1975.

13. Wirtz RA, Burkot TR, Graves PM, Andre RG: **Field evaluation of enzyme-linked immunosorbent assays for Plasmodium falciparum and Plasmodium vivax sporozoites in mosquitoes (Diptera: Culicidae) from Papua New Guinea**. *J Med Entomol* 1987, **24**:433–437.

14. Wilson MD, Ofosu-Okyere A, Okoli AU, McCall PJ, Snounou G: **Direct comparison of microscopy and polymerase chain reaction for the detection of Plasmodium sporozoites in salivary glands of mosquitoes**. *Trans R Soc Trop Med Hyg* 1998, **92**:482–483.

15. Bass C, Nikou D, Blagborough AM, Vontas J, Sinden RE, Williamson MS, Field LM: **PCR-based detection of Plasmodium in Anopheles mosquitoes: a comparison of a new high-throughput assay with existing methods**. *Malar J* 2008, **7**:177.

16. Bashar K, Tuno N, Ahmed TU, Howlader AJ: **False positivity of circumsporozoite protein (CSP)-ELISA in zoophilic anophelines in Bangladesh.** *Acta tropica* 2013, **125**:220–5.

17. Durnez L, Van Bortel W, Denis L, Roelants P, Veracx A, Trung HD, Sochantha T, Coosemans M: **False positive circumsporozoite protein ELISA: a challenge for the estimation of the entomological inoculation rate of malaria and for vector incrimination**. *Malar J* , **10**:195.

18. Sylla EH, Kun JF, Kremsner PG: **Mosquito distribution and entomological inoculation rates in three malaria-endemic areas in Gabon**. *Trans R Soc Trop Med Hyg* 2000, **94**:652–656.

19. Somboon P, Morakote N, Koottathep S, Trisanarom U: **Detection of sporozoites of Plasmodium vivax and Plasmodium falciparum in mosquitoes by ELISA: false positivity associated with bovine and swine blood**. *Trans R Soc Trop Med Hyg* 1993, **87**:322–324.

20. Dame JB, Williams JL, McCutchan TF, Weber JL, Wirtz RA, Hockmeyer WT, Maloy WL, Haynes JD, Schneider I, Roberts D, et al.: **Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite Plasmodium falciparum**. *Science* 1984, **225**:593–599.

21. Lockyer MJ, Schwarz RT: **Strain variation in the circumsporozoite protein gene of Plasmodium falciparum**. *Molecular and Biochemical Parasitology* 1987, **22**:101–108.

22. Plassmeyer ML, Reiter K, Shimp RL, Kotova S, Smith PD, Hurt DE, House B, Zou X, Zhang Y, Hickman M, Uchime O, Herrera R, Nguyen V, Glen J, Lebowitz J, Jin AJ, Miller LH, MacDonald NJ, Wu Y, Narum DL: **Structure of the Plasmodium falciparum Circumsporozoite Protein, a Leading Malaria Vaccine Candidate**. *The Journal of Biological Chemistry* 2009, **284**:26951–26963.

23. Young JF, Hockmeyer WT, Gross M, Ballou WR, Wirtz RA, Trosper JH, Beaudoin RL, Hollingdale MR, Miller LH, Diggs CL, et al.: **Expression of Plasmodium falciparum circumsporozoite proteins in Escherichia coli for potential use in a human malaria vaccine**. *Science* 1985, **228**:958–962.

24. Rani A, Sharma A, Rajagopal R, Adak T, Bhatnagar RK: **Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected Anopheles stephensi-an Asian malarial vector.** *BMC microbiology* 2009, **9**:96.

25. Petti CA, Polage CR, Schreckenberger P: **The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods.** *Journal of clinical microbiology* 2005, **43**:6123–5.

26. Shah N, Tang H, Doak TG, Ye Y: **Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics.** *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing* 2011:165–76.

27. Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E, Salamov A, Grigoriev I V, Suciu D, Levine SR, Markowitz VM, Rigoutsos I, Tringe SG, Bruce DC, Richardson PM, Lidstrom ME, Chistoserdova L: **High-resolution metagenomics targets specific functional types in complex microbial communities.** *Nature biotechnology* 2008, **26**:1029–34.

28. Shokralla S, Spall JL, Gibson JF, Hajibabaei M: **Next-generation sequencing technologies for environmental DNA research.** *Molecular ecology* 2012, **21**:1794–805.

29. Maccaferri S, Biagi E, Brigidi P: **Metagenomics: key to human gut microbiota.** *Digestive diseases (Basel, Switzerland)* 2011, **29**:525–30.

30. Boissiere A, Tchioffo MT, Bachar D, Abate L, Marie A, Nsango SE, Shahbazkia HR, Awono-Ambene PH, Levashina EA, Christen R, Morlais I: **Midgut microbiota of the malaria mosquito vector Anopheles gambiae and interactions with Plasmodium falciparum infection**. *PLoS Pathog* , **8**:e1002742.

31. Wang Y, Gilbreath TM, Kukutla P, Yan G, Xu J: **Dynamic gut microbiome across life history of the malaria mosquito Anopheles gambiae in Kenya.** *PloS one* 2011, **6**:e24767.

32. Dinparast Djadid N, Jazayeri H, Raz A, Favia G, Ricci I, Zakeri S: **Identification of the midgut microbiota of An. stephensi and An. maculipennis for their application as a paratransgenic tool against malaria**. *PLoS One* , **6**:e28484.

33. Briones AM, Shililu J, Githure J, Novak R, Raskin L: **Thorsellia anophelis is the dominant bacterium in a Kenyan population of adult Anopheles gambiae mosquitoes**. *ISME J* 2008, **2**:74–82.

34. Favia G, Ricci I, Damiani C, Raddadi N, Crotti E, Marzorati M, Rizzi A, Urso R, Brusetti L, Borin S, Mora D, Scuppa P, Pasqualini L, Clementi E, Genchi M, Corona S, Negri I, Grandi G, Alma A, Kramer L, Esposito F, Bandi C, Sacchi L, Daffonchio D: **Bacteria of the genus Asaia stably associate with Anopheles stephensi, an Asian malarial mosquito vector**. *Proc Natl Acad Sci U S A* 2007, **104**:9047–9051.

35. Koella JC, Lorenz L, Bargielowski I, Joanne PW: **Chapter 12 Microsporidians as Evolutionary Proof Agents of Malaria Control?** In *Advances in Parasitology*. *Volume Volume 68*. Academic Press; 2009:315–327.

36. Micieli M V, Marti GA, GarcÃa JJ, Tranchida MC, Becnel JJ: **Epizootiological studies of Amblyospora camposi (Microsporidia: Amblyosporidae) in Culex renatoi (Diptera: Culicidae) and Paracyclops fimbriatus fimbriatus (Copepoda: Cyclopidae) in a bromeliad habitat**. *Journal of Invertebrate Pathology* 2007, **94**:31–37.

37. Vossbrinck CR, Andreadis TG, Debrunner-Vossbrinck BA: **Verification of intermediate hosts in the life cycles of microsporidia by small subunit rDNA sequencing**. *J Eukaryot Microbiol* 1998, **45**:290–292.

38. Vossbrinck CR, Debrunner-Vossbrinck BA: **Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations.** *Folia parasitologica* 2005, **52**:131–142; discussion 130.

39. Bargielowski I, Koella JC: **A Possible Mechanism for the Suppression of Plasmodium berghei Development in the Mosquito Anopheles gambiae by the Microsporidian Vavraia culicis**. *PLoS ONE* 2009, **4**:5.

40. Sweeney AW, Becnel JJ: **Potential of microsporidia for the biological control of mosquitoes.** *Parasitology today Personal ed* 1991, **7**:217–220.

41. Schenker W, Maier WA, Seitz HM: **The effects of Nosema algerae on the development of Plasmodium yoelii nigeriensis in Anopheles stephensi.** *Parasitology Research* 1992, **78**:56–59.

42. Coyle CM, Weiss LM, Rhodes LV 3rd, Cali A, Takvorian PM, Brown DF, Visvesvara GS, Xiao L, Naktin J, Young E, Gareca M, Colasante G WM: **Fatal myositis due to the**

microsporidian Brachiola algerae, a mosquito pathogen. *New England Journal of Medicine* 2004, **351**:42–47.

43. V. Robert TT& PC: **Successful eradication of a microsporidian, Nosema sp., in a mosquito colony**. *Ann Soc belge Med trop* 1993, **73**:71–72.

44. Avery SW, Undeen AH: **The isolation of microsporidia and other pathogens from concentrated ditch water.** *Journal of the American Mosquito Control Association* 1987, **3**:54–8.

45. Becnel JJ, White SE, Shapiro AM: **Review of microsporidia-mosquito relationships: from the simple to the complex.** *Folia parasitologica* 2005, **52**:41–50.

46. Lorenz LM, Koella JC: **The microsporidian parasite Vavraia culicis as a potential late life-acting control agent of malaria**. *Evolutionary Applications* 2011, **4**:783–790.

47. Koella JC, Saddler A, Karacs TPS: **Blocking the evolution of insecticide-resistant malaria vectors with a microsporidian**. *Evolutionary Applications* 2012, **5**:283–292.

48. Becnel JJ, Garcia JJ, Johnson MA: **Edhazardia aedis (Microspora: Culicosporidae) effects on the reproductive capacity of Aedes aegypti (Diptera: Culicidae).** *Journal of medical entomology* 1995, **32**:549–53.

49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**:403–10.

50. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic acids research* 2005, **33**(Database issue):D34–8.

51. Jones DT: **Protein secondary structure prediction based on position-specific scoring matrices.** *Journal of molecular biology* 1999, **292**:195–202.

52. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics (Oxford, England)* 2000, **16**:404–5.

53. Dosztányi Z, Mészáros B, Simon I: **ANCHOR: web server for predicting protein binding regions in disordered proteins.** *Bioinformatics (Oxford, England)* 2009, **25**:2745–6.

54. Dosztányi Z, Csizmók V, Tompa P, Simon I: **The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins.** *Journal of molecular biology* 2005, **347**:827–39.

55. Mészáros B, Simon I, Dosztányi Z: **Prediction of protein binding regions in disordered proteins.** *PLoS computational biology* 2009, **5**:e1000376.

56. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR: **Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses**. *Proc Natl Acad Sci U S A* 1985, **82**:6955–6959.

57. Winsley T, van Dorst JM, Brown M V, Ferrari BC: **Capturing greater 16S rRNA gene sequence diversity within the domain Bacteria**. *Appl Environ Microbiol* , **78**:5938–5941.

58. Roudiere L, Lorto S, Tallagrand E, Marchandin H, Jeannot JL, Jumas-Bilak E: **[Molecular fingerprint of bacterial communities and 16S rDNA intra-species heterogeneity: a pitfall that should be considered]**. *Pathol Biol (Paris)* 2007, **55**:434–440.

59. Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF: **Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea.** *The ISME journal* 2011, **5**:1571–1579.

60. Dojka MA, Hugenholtz P, Haack SK, Pace NR: **Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation**. *Appl Environ Microbiol* 1998, **64**:3869–3877.

61. **High-speed, multiplexed 16s Amplicon sequencing on the MiSeq® system** [http://res.illumina.com/documents/products/appnotes/appnote_miseq_16s.pdf]

62. **MiSeq® Reagent Kit v2** [http://supportres.illumina.com/documents/myillumina/cebf8b82-b1d9-4384-a64b-002db4193cbe/miseqreagentkit_v2_reagentprepguide_15034097_b.pdf]

63. **Nextera® XT** [http://supportres.illumina.com/documents/myillumina/900851dc-01cf-4b70-9e95-d590531c5bd4/nextera_xt_sample_preparation_guide_15031942_c.pdf]

64. **Using a PhiX Control for HiSeq® Sequencing Runs** [http://res.illumina.com/documents/products/technotes/technote_phixcontrolv3.pdf]

65. **Quality Scores for Next Generation Sequencing** [http://res.illumina.com/documents/products/technotes/technote_q-scores.pdf]

66. Kosakovsky Pond S, Wadhawan S, Chiaromonte F, Ananda G, Chung W-Y, Taylor J, Nekrutenko A: **Windshield splatter analysis with the Galaxy metagenomic pipeline.** *Genome research* 2009, **19**:2144–53.

67. **Dynabeads® Antibody Coupling Kit**
[http://tools.invitrogen.com/content/sfs/manuals/dynabeads_antibody_couplingkit_man.pdf]

68. Biosystems A: **Applied Biosystems/MDS SCIEX 4800 MALDI TOF/TOFTM Analyzer Hardware Guide**. 2005.

69. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics (Oxford, England)* 2007, **23**:2947–8.

70. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R: **A new bioinformatics analysis tools framework at EMBL-EBI.** *Nucleic acids research* 2010, **38**(Web Server issue):W695–9.

71. Herrington DA, Clyde DF, Davis JR, Baqar S, Murphy JR, Cortese JF, Bank RS, Nardin E, DiJohn D, Nussenzweig RS: **Human studies with synthetic peptide sporozoite vaccine (NANP)3-TT and immunization with irradiated sporozoites.** *Bulletin of the World Health Organization* 1990, **68 Suppl**:33–7.

72. PM G, H G: **Vaccines for preventing malaria in the pre-erythrocytic phase**. 2009.

73. Burkot TR, DA ZW, Geysen HM, Wirtz RA, Saul A: **Fine specificities of monoclonal antibodies against the Plasmodium falciparum circumsporozoite protein: recognition of both repetitive and non-repetitive regions**. *Parasite Immunology* 1991, **13**:161–170.

74. Becnel JJ, Sprague V, Fukuda T, Hazard EI: **Development of Edhazardia aedis (Kudo, 1930) n. g., n. comb. (Microsporida: Amblyosporidae) in the mosquito Aedes aegypti (L.) (Diptera: Culicidae).** *The Journal of protozoology* , **36**:119–30.

75. Becnel JJ, Johnson MA: **Mosquito host range and specificity of Edhazardia aedis (Microspora: Culicosporidae).** *Journal of the American Mosquito Control Association* 1993, **9**:269–74.

76. Silver JB: *Mosquito Ecology: Field Sampling Methods (Google eBook)*. Springer; 2007:1499.

77. Tokarev Y, Sitnikova N, Pistone D: **Microsporidia PCR detection artifacts due to non-specific binding of the universal microsporidia primers to the rDNA of arthropod hosts**. *Buletinul Academiei de Stiinte a Moldovei Stiintele vietii* 2010, **no.1 p. 72**.

78. Matsushima N, Yoshida H, Kumaki Y, Kamiya M, Tanaka T, Izumi Y, Kretsinger RH: **Flexible structures and ligand interactions of tandem repeats consisting of**

proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins. *Current protein & peptide science* 2008, **9**:591–610.

79. Jorda J, Xue B, Uversky VN, Kajava A V: **Protein tandem repeats - the more perfect, the less structured.** *The FEBS journal* 2010, **277**:2673–82.

80. Patterson. MJ: **Streptococcus**. In *Medical Microbiology*. 4th Editio. Edited by Baron S editor. Galveston (TX): University of Texas Medical Branch at Galveston; 1996.

81. Christophides MG and GK: *Anopheles Mosquitoes - New Insights into Malaria Vectors*. InTech; 2013.

82. Lochouarn L, Fontenille D: **ELISA detection of malaria sporozoites: false-positive results in Anopheles gambiae s.l. associated with bovine bloodmeals.** *Transactions of the Royal Society of Tropical Medicine and Hygiene* , **93**:101–2.

83. Medlin L, Elwood HJ, Stickel S, Sogin ML: **The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions**. *Gene* 1988, **71**:491–499.

84. Dolci P, Zenato S, Pramotton R, Barmaz A, Alessandria V, Rantsiou K, Cocolin L: **Cheese surface microbiota complexity: RT-PCR-DGGE, a tool for a detailed picture?** *International journal of food microbiology* 2013, **162**:8–12.

85. Feligini M, Panelli S, Buffoni JN, Bonacina C, Andrighetto C, Lombardi A: **Identification of microbiota present on the surface of Taleggio cheese using PCR-DGGE and RAPD-PCR.** *Journal of food science* 2012, **77**:M609–15.

86. **16sr RNA Gene** [http://greengenes.lbl.gov/cgi-bin/JD_Tutorial/nph-16S.cgi]

87. Armougom F, Raoult D: **Exploring Microbial Diversity Using 16S rRNA High-Throughput Methods**. *Journal of Computer Science & Systems Biology* 2009, **02**:74–92.

88. Wang Y, Qian P-Y: **Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies.** *PloS one* 2009, **4**:e7401.

89. Soergel DAW, Dey N, Knight R, Brenner SE: **Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences.** *The ISME journal* 2012, **6**:1440–4.

90. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO: **Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.** *Nucleic acids research* 2013, **41**:e1.

91. Chakravorty S, Helb D, Burday M, Connell N, Alland D: **A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria.** *Journal of microbiological methods* 2007, **69**:330–9.

92. Lindh JM, Terenius O, Faye I: **16S rRNA gene-based identification of midgut bacteria from field-caught Anopheles gambiae sensu lato and A. funestus mosquitoes reveals new species related to known insect symbionts.** *Applied and environmental microbiology* 2005, **71**:7217–23.

93. Straif SC, Mbogo CN, Toure AM, Walker ED, Kaufman M, Toure YT, Beier JC: **Midgut bacteria in Anopheles gambiae and An. funestus (Diptera: Culicidae) from Kenya and Mali.** *Journal of medical entomology* 1998, **35**:222–6.

94. Kilcher S, Loessner MJ, Klumpp J: **Brochothrix thermosphacta bacteriophages feature heterogeneous and highly mosaic genomes and utilize unique prophage insertion sites.** *Journal of bacteriology* 2010, **192**:5441–53.

95. Connon SA, Tovanabootr A, Dolan M, Vergin K, Giovannoni SJ, Semprini L: **Bacterial community composition determined by culture-independent and -dependent methods during propane-stimulated bioremediation in trichloroethene-contaminated groundwater.** *Environmental microbiology* 2005, **7**:165–78.

96. Sylvie Manguin, Chung Thuy Ngo, Krajana Tainchum W, Juntarajumnong, Theeraphap Chareonviriyaphap A-L, Jumas-Bilak M and E: **Anopheles mosquitoes - New insights into malaria vectors**. Edited by Manguin S. InTech; 2013.

97. **Rickettsia felis in Aedes albopictus Mosquitoes, Libreville, Gabon - Vol. 18 No. 10 - October 2012 - Emerging Infectious Disease journal - CDC**. .

98. Socolovschi C, Pages F, Ndiath MO, Ratmanov P, Raoult D: **Rickettsia species in African Anopheles mosquitoes.** *PloS one* 2012, **7**:e48254.

99. Collins MD, Wiernik A, Falsen E, Lawson PA: **Atopococcus tabaci gen. nov., sp. nov., a novel Gram-positive, catalase-negative, coccus-shaped bacterium isolated from tobacco.** *International journal of systematic and evolutionary microbiology* 2005, **55**(Pt 4):1693–6.

100. Kämpfer P, Witzenberger R, Denner EBM, Busse H-J, Neef A: **Sphingopyxis witflariensis sp. nov., isolated from activated sludge.** *International journal of systematic and evolutionary microbiology* 2002, **52**(Pt 6):2029–34.

101. Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F: **Lactobacillus plantarum promotes Drosophila systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing.** *Cell metabolism* 2011, **14**:403–14.

102. Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH: **Gene structure in the sea urchin Strongylocentrotus purpuratus based on transcriptome analysis.** *Genome research* 2012, **22**:2079–87.

103. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, Benito-Gutiérrez EL, Dubchak I, Garcia-Fernàndez J, Gibson-Brown JJ, Grigoriev I V, Horton AC, de Jong PJ, Jurka J, Kapitonov V V, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, et al.: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453**:1064–71.

104. **Amphioxus (Lancelet)**

105. Herman IM: **Actin isoforms.** *Current opinion in cell biology* 1993, **5**:48–55.

106. Somboon P, Morakote N, Koottathep S, Trisanarom U: **Detection of sporozoites of Plasmodium vivax and Plasmodium falciparum in mosquitoes by ELISA: false positivity associated with bovine and swine blood**. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1993, **87**:322–324.

107. Ballestrem C, Wehrle-Haller B, Imhof BA: **Actin dynamics in living mammalian cells.** *Journal of cell science* 1998, **111 (Pt 1**:1649–58.

108. Méjean C, Hué HK, Pons F, Roustan C, Benyamin Y: **Cation binding sites on actin: a structural relationship between antigenic epitopes and cation exchange.** *Biochemical and biophysical research communications* 1988, **152**:368–75.

10. INDEX OF ANNEXES