UNIVERSIDADE DA BEIRA INTERIOR
Ciências da Saúde

Developing a NGS panel for diagnosis of Lyme disease and its co-infections

Eduardo Augusto Coelho

Dissertação para obtenção do Grau de Mestre em
Ciências Biomédicas
(2º ciclo de estudos)

Orientador: Prof.ª Doutora Isabel Theriaga Gonçalves
Co-orientador: Doutor Gonçalo Doria

Covilhã, outubro de 2018

Dedicatória

Este trabalho é dedicado a todos aqueles que, de alguma forma, fizeram parte do meu caminho e que me ajudaram a ser pessoa que sou hoje.

Agradecimentos

Em primeiro lugar, obviamente, gostaria de agradecer à minha família por me dar todas as ferramentas possíveis para atingir esta etapa. Para o meu Pai e para a minha Mãe, faltam-me as palavras para expressar o quão grato estou por todo o apoio e conforto que sempre me deram. Vocês são um verdadeiro exemplo para mim e acho que a melhor maneira de vos agradecer é provar, todos os dias, que o vosso esforço deu frutos. Aos meus irmãos, companheiros de brincadeira e de sofrimento, quero apenas relembrar que são indispensáveis na minha vida e que farei sempre o possível para vos ajudar, como sei que farão por mim. Aos meus avós, aos que me dão todo o apoio e aos que infelizmente não podem assistir a mais uma conquista na minha vida, agradeço por tudo o que fizeram por mim e por me fazerem sentir especial todos os dias em que estamos juntos. Gosto muito de todos vocês e obrigado por me acompanharem de perto o meu crescimento.

À minha namorada, Marta, quero agradecer pelo apoio incondicional que me dás sempre que preciso. Ajudaste-me muito a evoluir como pessoa e a melhorar a perceção do que me rodeia. Que venham muitos mais anos de felicidade, sempre de mãos dadas!

À minha orientadora, Professora Doutora Isabel Theriaga Gonçalves, gostaria de agradecer sinceramente por toda a disponibilidade e prontidão em ajudar-me sempre que eu necessitei.

Ao Professor Doutor Ilídio Correia, diretor do Mestrado em Ciências Biomédicas na UBI, gostaria de agradecer por me ter ajudado a abrir os horizontes, numa altura em que a motivação parecia escassa.

Queria agradecer a toda a equipa da STAB VIDA, principalmente ao CEO Orfeu Flores pela oportunidade que me deu em desenvolver este projeto em âmbito empresarial. Ao Gonçalo Doria, pilar no desenvolvimento deste projeto, o meu obrigado por todos os esclarecimentos e críticas construtivas que tanto me ajudaram a evoluir como investigador.

Por último, não por serem menos importantes, mas porque sei que leem esta página até ao fim, queria agradecer a todos os meus amigos que me têm acompanhado nos últimos anos. Ao Gustavo, fiel companheiro de casa, ao Rui e ao Miguel pela companhia e ajuda nesta reta final, ao Tiago por todos os conselhos, obrigado e um grande abraço. Aos meus irmãos da Desertuna, companheiros de trabalho e de diversão, muito obrigado por me darem tanta bagagem, tanto a nível pessoal como a nível profissional, para o meu futuro.

Muito obrigado a todos!

Resumo

A doença de Lyme, também conhecida como borreliose de Lyme, é a doença transmitida por carraças mais comum no hemisfério norte, com 300.000 novos casos estimados anualmente só nos Estados Unidos da América. Para além da transmissão de *Borrelia burgdorferi sensu lato* (*s.l.*), bactéria responsável pela doença de Lyme, as carraças do complexo *Ixodes ricinus* são vetores de outras infeções, sendo as mais comuns a babesiose e a anaplasmose granulocítica humana. A presença de coinfecções pode causar manifestações clínicas mais severas e o seu incorreto diagnóstico pode levar a um tratamento inapropriado. O único método aceite para o diagnóstico da borreliose de Lyme, atualmente, é um teste sorológico baseado numa abordagem de dois níveis no qual é realizado um imunoensaio enzimático complementado por um *western blot*. Este método indireto para a deteção de *Borrelia burgdorferi s.l.* carece de sensitividade na fase inicial da infeção, devido ao tempo necessário para os anticorpos serem produzidos.

Ao longo das últimas décadas, vários estudos usando métodos diretos, tais como cultura e *Polimerase Chain Reaction* (PCR), foram realizados com o objetivo de desenvolver um método de diagnóstico alternativo para esta doença infeciosa. Contudo, estes testes demonstraram uma taxa elevada de falsos negativos.

Neste estudo, foi criado um painel de *Next Generation Sequencing* (NGS), para a plataforma MiSeq da Illumina, que permite o diagnóstico simultâneo da doença de Lyme e das suas coinfecções mais frequentes. Este painel inclui sete pares de *primers* específicos para um fragmento de um gene de cada uma das espécies patogénicas incluídas, em regiões que permitem a distinção entre genoespécies. O painel foi testado na preparação das bibliotecas para sequenciação com amostras de sangue total e com amostras de ADN extraído do sangue total. A par do teste com o painel desenvolvido, com o intuito de avaliar a sensibilidade do mesmo, os dois tipos de amostras foram também testados com *primers* específicos para as regiões V3 e V4 do gene 16S do ARN ribossomal, amplamente usados na análise de microbiomas.

Devido à dificuldade em obter amostras de pacientes com doença de Lyme e com as outras infeções abrangidas pelo painel, neste trabalho, foram testadas cinco amostras de sangue de pacientes diagnosticados com babesiose, juntamente com os controlos positivo e negativo.

A condição que demonstrou melhores resultados foi aquela em que foi usado ADN extraído de sangue em combinação com o painel, com a qual foi possível identificar ADN de *Babesia microti* nas cinco amostras de pacientes infetados.

Apesar da necessidade de testar o método em amostras de pacientes com doença de Lyme e com as restantes infeções incluídas no painel desenvolvido, os resultados obtidos neste trabalho demonstram-se promissores para a futura utilização deste método como alternativa aos exames atuais, especialmente na fase inicial da infeção.

Palavras-chave

Doença de Lyme, diagnóstico, NGS, coinfecções.

Resumo Alargado

A doença de Lyme, também conhecida como borreliose de Lyme, é a doença transmitida por carraças mais comum no hemisfério norte, com 300.000 novos casos estimados anualmente só nos Estados Unidos da América. Para além da transmissão de *Borrelia burgdorferi sensu lato* (*s.l.*), bactéria responsável pela doença de Lyme, as carraças do complexo *Ixodes ricinus* são vetores de outras infeções, sendo as mais comuns a babesiose e a anaplasmose granulocítica humana. A presença de coinfecções pode causar manifestações clínicas mais severas e o seu incorreto diagnóstico pode levar a um tratamento inapropriado. O único método aceite pela *Food and Drug Administration* (FDA) para o diagnóstico da borreliose de Lyme, atualmente, é um teste sorológico baseado numa abordagem de dois níveis no qual é usado um imunoensaio enzimático complementado por um *Western blot*. Basicamente, neste método, se o imunoensaio enzimático der um resultado negativo, exclui-se a hipótese de doença. Caso o resultado seja positivo, a amostra de soro é submetida a um *Western blot* para deteção de anticorpos IgM ou IgG, consoante o tempo passado desde o início da infeção. Este método indireto para a deteção de *Borrelia burgdorferi s.l.* carece de sensitividade na fase inicial da infeção, devido ao tempo necessário para os anticorpos serem produzidos, uma vez que os anticorpos IgM e IgG podem demorar entre 2 a 4 e 4 a 6 semanas, respetivamente, a atingirem uma concentração mínima para ser detetada por este teste.

Ao longo das últimas décadas, vários estudos usando métodos diretos, tais como cultura e PCR, foram desenvolvidos com o objetivo de se alcançar um método de diagnóstico alternativo para esta doença infeciosa. Contudo, apesar dos vários tipos de amostras testadas, tais como, líquido cefalorraquidiano, líquido sinovial, sangue e urina, estes testes demonstraram uma taxa elevada de falsos negativos.

Neste estudo, foi criado um painel de *Next Generation Sequencing* (NGS), para a plataforma MiSeq da Illumina, que permite o diagnóstico simultâneo da doença de Lyme e das suas coinfecções mais frequentes. As espécies abrangidas por este painel são: *Borrelia burgdorferi s.l.*, responsável pela doença de Lyme, *Anaplasma phagocytophilum*, responsável por causar anaplasmose granulocítica humana, *Babesia microti*, responsável por causar babesiose, *Bartonella henselae*, responsável por causar bartonelose, *Coxiella burnetii*, causadora da febre Q, *Ehrlichia canis*, responsável por causar ehrlichiose e *Rickettsia rickettsii*, conhecida por causar a febre da carraça. Este painel inclui sete pares de *primers* específicos para um fragmento de um gene de cada uma das espécies patogénicas incluídas, em regiões que permitem a distinção entre genoespécies, informação que pode ser importante na compreensão das manifestações clínicas.

O painel foi testado na preparação das bibliotecas para sequenciação com amostras de sangue total e com amostras de ADN extraído do sangue total. A par do teste com o painel desenvolvido, com o intuito de avaliar a sensibilidade do mesmo, os dois tipos de amostras

foram também testados com *primers* específicos para as regiões V3 e V4 do gene 16S do ARN ribossomal, amplamente usados na análise de microbiomas.

Devido à dificuldade em obter amostras de pacientes com doença de Lyme e de pacientes com as outras infeções abrangidas pelo painel, neste trabalho, foram testadas cinco amostras de sangue de pacientes diagnosticados com babesiose,. Para além destas amostras, fornecidas pelo *Centers for Disease Control and Prevention* (CDC), dois controlos, um positivo e um negativo foram também testados. Como controlo negativo, foi usada uma amostra de sangue, escolhida de forma aleatória, de um grupo de indivíduos que não vivem em zonas endémicas para estas doenças e que não se recordam de terem sido mordidos por carraças. Para o controlo positivo, os fragmentos dos genes avaliados no painel foram amplificados individualmente a partir de ADN genómico de cada uma das espécies através de um PCR, usando só o respetivo par de *primers* específicos. Os fragmentos obtidos foram clonados e usados na transformação em células de *E.coli*. Posteriormente, estas células foram inseridas numa alíquota juntamente com sangue do controlo negativo, tentando mimetizar uma infeção.

No PCR da preparação das bibliotecas para a sequenciação, cada uma das amostras foi testada com quatro condições diferentes, relativamente ao tipo de amostra e aos *primers* usados. A condição que demonstrou melhores resultados foi aquela em que foi usado ADN extraído do sangue em combinação com o painel, na qual foi possível identificar ADN de *Babesia microti* nas cinco amostras de pacientes infetados. Na outra condição em que o painel foi usado, diretamente no sangue total, apenas foi possível detetar a presença do agente patogénico em três dos cinco pacientes e verificou-se a amplificação de produtos de PCR não específicos. Nas duas condições em que os *primers* específicos para o gene 16S do ARN ribossomal foram usados, só se observaram resultados positivos em duas das cinco amostras.

Apesar da necessidade de otimizar e testar o método em amostras de pacientes com doença de Lyme e com as restantes infeções incluídas no painel desenvolvido, os resultados obtidos neste trabalho demonstram-se promissores para a futura utilização deste método como alternativa aos exames atuais, especialmente na fase inicial da infeção.

Abstract

Lyme disease, also known as Lyme borreliosis, is the most common-tick borne disease in the northern hemisphere, with 300,000 cases estimated each year only in the United States. In addition to the transmission of *Borrelia burgdorferi sensu lato* (*s.l.*), the bacteria responsible for Lyme disease, ticks of the *Ixodes ricinus* complex are vectors for other infections, with the most common being babesiosis and human granulocytic anaplasmosis. The presence of co-infections may cause more severe clinical manifestations and their misdiagnosis may lead to an inappropriate treatment. The only accepted method to diagnose Lyme borreliosis, currently, is a serologic test based in a two-tier approach using an enzyme immunoassay (EIA) complemented with a Western immunoblot. This indirect method for *Borrelia burgdorferi s.l.* detection suffers from lack of sensitivity in the early stage of the disease, due to the time window needed for antibodies to be produced.

Over the past few decades, many studies have been carried using direct methods, such as culture and PCR, in order to develop an alternative diagnostic method for this infectious disease, however this tests also suffer from a high rate of false negatives.

In this study, a NGS panel was developed, for the Illumina's Miseq platform, which allows the simultaneous diagnose of Lyme disease and its most common co-infections. This panel includes seven specific primer pairs that target a gene fragment of each of the included pathogenic species, in regions that allow the genospecies distinction.

The panel was tested in the library preparation for sequencing using samples of whole blood and samples of extracted DNA from the whole blood. Along with the developed panel, in order to evaluate its sensibility, both types of samples were also tested with specific primers that target the V3 and V4 regions of the 16S ribosomal RNA gene, widely used in microbiome analysis.

Due to the difficulty to obtain samples from patients with Lyme disease and with the other infections covered by the panel, in this study, five samples of whole blood from patients diagnosed with babesiosis were tested, along with a positive and a negative controls.

The condition that has shown better results was the one using extracted DNA from the blood combined with the panel, which detected *Babesia microti* DNA in all five samples of the infected patients.

Despite the need to test this method in samples from patients with Lyme disease and the remaining infections included in the developed panel, the results obtained in this study are promising for the future use of this method as an alternative to the current tests, especially in the initial phase of infection.

Keywords

Lyme disease, diagnosis, NGS, co-infections.

Table of contents

List of Figures

List of Tables

List of Abbreviations

U.S.A.	United States of America
bp	base pairs
EM	Erythema migrans
s.l.	*Sensu lato*
CDC	Centers for Disease Control and Prevention
PCR	Polymerase chain reaction
ACA	Acrodermatitis chronica atrophicans
spp.	Species
s.s.	*Sensu stricto*
FDA	United States Food and Drug Administration
EIA	Enzyme immunoassay
IFA	Immunofluorescent assay
WB	Western blot
IgM	Immunoglobulin M
IgG	Immunoglobulin G
HGP	Human Genome Project
NGS	Next generation sequencing
AGE	Agarose gel eletrophoresis
TAE	Tris-Acetate-EDTA
V	Volt

I. INTRODUCTION

1. Lyme disease

Lyme disease is a zoonosis, which is a disease from animals that can be transmitted to humans. This disorder has become a major concern in the last four decades among the medical community in the United States of America(U.S.A.) and in Europe. Although Lyme Disease was initially considered an inflammatory joint disorder, soon became identified as a multisystemic disorder affecting as well the skin, nervous system and heart (1).

1.1. Identification of a new pathology

Lyme arthritis, the initial designation of the disorder that nowadays is preferentially referred to as Lyme disease or Lyme borreliosis, was first suggested as an unrecognized pathology in 1975. The first suspicions started when two mothers, from Old Lyme, Connecticut, within a month, informed the State Health Department and Yale Rheumatology Clinic about the strange prevalence of arthritis in a small community. The first mother only reported children cases, who have been diagnosed with juvenile rheumatoid arthritis, and both mothers highlighted the fact that most of people suffering from this symptoms lived in the same neighborhood or close together (2). By geographic clustering of reported cases and, in some patients, their association with the characteristic skin lesion, Steere *et al.* refer that Lyme arthritis has been affecting people in eastern Connecticut since 1972, with the majority of the new cases taking place in the summer and early fall (2-4).

1.2. Pathogen

Borrelia burgdorferi is a gram negative bacteria member of eubacterial phylum *Spirochaetes*. This phylum's name is due to the morphology of its organisms which show a spiral body. This irregularly coiled spirochetes range from 10 to 30 μm in length and from 0.18 to 0.25 μm in diameter (5).

This spirochete was first isolated in 1982 by Burgdorfer *et al.* by dissection of adult *Ixodes dammini*, now known as *Ixodes scapularis*, collected in Shelter Island, New York, a known endemic region of Lyme disease. More than a half of the ticks (61%) contained spirochetes, which were principally distributed in the midgut. No spirochetes were found in the salivary glands (5).

The theory suggested by Burgdorfer *et al.*, that this spirochete was the etiological agent of Lyme disease, was solidly supported by subsequent isolation of identical spirochetes from blood, skin, and cerebrospinal fluid of patients with signs and symptoms suggestive of Lyme disease (6, 7).

The complete genome of *Borrelia burgdorferi* (strain B31), was first sequenced by Fraser *et al.,* in 1997. It consists in a linear chromosome with 910,725 base pairs (bp) and 12 linear and 9 circular plasmids which summed contain 610,694 bp (8, 9). The genome encodes few

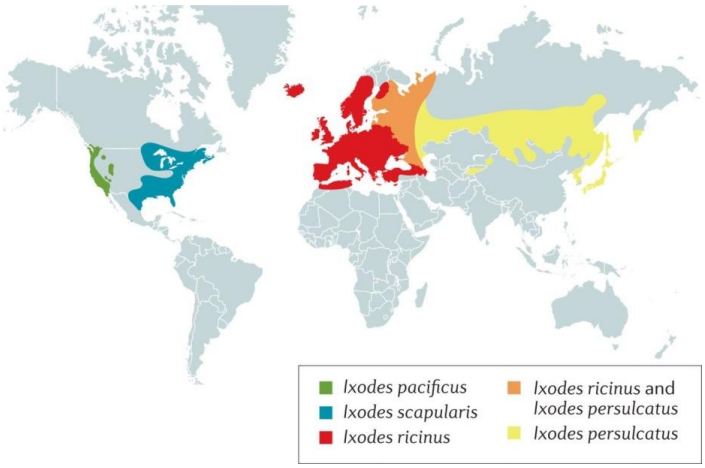
proteins with biosynthetic activity, which makes the bacteria dependent on the host to acquire its nutritional needs. Besides, no recognizable toxins are encoded by *Borrelia burgdorferi* genome, its pathogenic effect is caused by adhesion to host cells, migration through tissues and evasion of immune clearance (8-10).

1.3. Vector

The identification of erythema migrans (EM) as a symptom of this disorder, a skin lesion that has previously been associated with tick bites in Europe (11), though any arthritis episodes have been related with it, suggested that the vector responsible for the disease transmission was this arthropod. This theory was supported by the data obtained by geographical clustering of epidemiologic studies. The patients were from rural regions with heavily wooded areas, the peak onset of new cases being reported was between summer and early fall and the occurrence of the disorder onset in elements of the same family usually didn't occur in the same year. Besides, some patients remembered a tick bite in the region where the skin rash appeared and one of them even took the tick to identification (*Ixodes scapularis*) (12).

Nowadays, there are four different species of ticks that are known to be competent vectors for Lyme Disease, with all of them belonging to *Ixodes ricinus* complex. This complex is a paraphyletic group with a geographical distribution throughout almost the entire globe (13). This ticks' saliva has the ability to inhibit the alternative pathway of the complement of the host, which will lead to an absence of efficient rejection by the host, increasing the chance of a successful blood meal (14).

In the U.S.A., the blacklegged tick, *Ixodes scapularis*, is the responsible for the cases in eastern and upper midwestern regions while *Ixodes pacificus* is the vector in Pacific Coast. Across the Atlantic, *Ixodes ricinus* and *Ixodes persulcatus* are the principal vectors in Europe and Asia, respectively, but unlike U.S.A., there are regions where both are present (15-17) (Figure 1).



1.3.1. Figure 1 - Geographical distribution of the four ticks with greater importance in transmission of Lyme disease worldwide (Adapted from(17)).**The life cycle of *Ixodes ricinus* complex**

The life cycle of the four ticks early referred, has a duration of two to three years, and present different seasonality. From the eggs, ticks pass through three developmental stages: larvae, nymph and adult taking one blood meal at each one of them. Since there is no evidence for transovarial transmission of *B. burgdorferi sensu lato* (*s.l.*), ticks depend on infected hosts to be infected. After acquiring the pathogen, ticks are able to transmit it to the next host (16, 18, 19).

Nymphs, which are the principal responsible for transmitting the pathogen to humans, due to its small size that may go unnoticed, are active from early spring to mid-summer for *I. ricinus*, *I. persulcatus* and *I. pacificus* with the possibility of a second peak in the autumn for *I. ricinus*, while *I. scapularis* has its peak from early summer to early fall. The fact that the peaks of nymphal and larvae stages activity differ by about three months, leads to the wide transmission of *B. burgdorferi s.l.* in the hosts of this two stages which are mainly small mammals, such as mice and shrews, and birds. Adult ticks are known to feed in larger mammals, being the most commonly described the white-tailed deer which shows great importance in supporting tick populations (16, 18, 20).

1.4. Co-infections

Ixodes ricinus is the most widespread and abundant ixodid tick in western Europe and is frequently associated with bites in humans (21). This small hard tick is known to be a vector for a large variety of pathogenic species concerning both physicians and veterinaries. The most frequently reported co-infections of *Borrelia burgdorferi s.l.* in humans are caused by *Babesia microti*, known to cause babesiosis, and *Anaplasma phagocytophilum*, the responsible agent for human granulocytic anaplasmosis (22, 23). The presence of co-infections can cause greater disease severity and the misdiagnosis may lead to inappropriate treatment (23). There are other species that have already been reported as co-infections, although with less

frequency that the ones described before. These include *Rickettsia monacensis* and *Rickettsia helvetica*, that cause spotted fever ricketsiosis, *Bartonella henselae*, responsible for cat-scratch disease and *Coxiella burnetii*, the agent of Q fever (24-28).

1.5. Epidemiology

Lyme disease is the most common tick-borne disease in the northern hemisphere. This disorder can affect both genders and all ages, but in a surveillance report performed in the U.S.A., children and older adults are the most affected, with males showing a slightly higher incidence (15). This disorder is estimated to present 300,000 new cases every year just in the United States, but despite the high incidence rates, few cases of death have been reported(16).

In the United States of America Lyme Disease has become a notifiable disorder in 1991. Since then, surveillance reports for Lyme Disease have been performed by the Centers for Disease Control and Prevention (CDC). The latest report describes a total of 275,589 cases of Lyme between 2008 and 2015, with approximately 76% of the cases confirmed and the other 24% marked as probably. The highest peak of disease reported cases, was observed in the beginning of July in all the years covered by the study. Although states with high risk of infection tend to stabilize or even decrease the number of reported cases, neighbor states' events have been increasing(15).

Unlike the information in U.S.A., epidemiology of Lyme Disease is still very unclear due to not being included, until June of the present year, in the list of diseases with epidemiologic surveillance in Europe. However, through the analysis of published studies of Lyme borreliosis in different countries in Europe, Skyes and Makiello, calculated the weighted mean incidence rate obtaining the value of 22.05/100,000 person-years. The incidence rates for this disease show a wide variation not only between countries, but also in regions of the same country. The country with more cases reported is Sweden, with 464/100,000 person-years while in Italy only 0.001/100,000 person-years cases were described(29). That being said, approximately 91,000 new cases are estimated to occur in Western Europe every year.

Hereafter, incidence of Lyme borreliosis in European Union will be better understood, since there will be an standardization in surveillance which will open the way to a better understanding of data acquired from the different countries.

1.6. Clinical Manifestations

1.6.1. Erythema migrans

The most characteristic symptom of Lyme Disease is the formation of a skin rash, called erithrema migrans. This lesion was first described in Europe by Lipschütz (30), but it was Afzelius the first to suggest that this symptom was related to a tick bite(11).

This skin lesion occurs in about 75% of the patients diagnosed with Lyme borreliosis and appears from 3 to 20 days after the tick bite (12, 15). EM, most of the times begins in the tick bite site as a red papule or macule and expands forming a red ring, with partial central clearing, resulting in a bull's eye shape, its characteristic form. However, sometimes central clearing is not verified, resulting in a large red spot which makes it more difficult to distinguish between similar lesions (Figure 2). Although it occurs frequently in the thigh, groin, and axilla, EM has been observed at any region of the body, with a diameter of at least 5 cm, being the most common size observed of about 16 cm (12, 31, 32). Generally, only one lesion appears, however, multiple skin rashes occurring simultaneously have been reported in approximately 20% and 10% of patients from U.S.A. and Europe respectively (31). This symptom, even in untreated patients, tends to fade within 3 to 4 weeks, but it may reappear (32, 33).

Although initially this symptom was described as exclusive for Lyme Disease patients, cases of tick bite, outside the *Ixodes ricinus* complex, shown an identical skin lesion, and so, besides being a good indicator of Lyme borreliosis, no conclusive diagnosis must be taken by its identification (34).

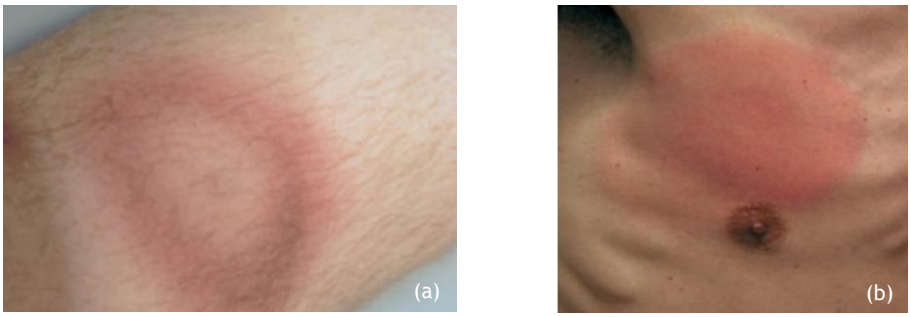


Figure 2- Examples of Erythema migrans. (a)- EM on the lower leg presenting central clearing. (b)- EM on the right breast without central clearing. (Adapted from (35)).

1.6.2. Stages of Lyme disease

In 1989, for clinical purposes, Steere described three stages to characterize the development of the illness, that are still used nowadays, being that two of them represent early disease and the other late disease (33).

Early Localized Infection:

In this stage, *B. burgdorferi* spreads locally in the skin resulting in EM for the majority of the patients. In Europe, EM has been described to expand slower than the cases reported in the U.S. and usually, no other symptoms are reported along with the skin lesion (36, 37). The two principal species causing Lyme disease in Europe, *B.afzelii* and *B.garinii*, present different evolution in the skin lesion provoked. *B.garinii* EM tends to spread faster than the one caused by *B.afzelii* and it's usually itchy, while *B.afzelii*, sometimes cause a rare skin manifestation,

known as borrelial lymphocytoma, located often in the ear lobe in children and on the nipple in adults (38).

The symptoms that usually occur at this stage are: fever, fatigue, malaise, headache, myalgias, arthralgias and regional lymphadenopathy (17, 33). According to Berger *et al.*, is at this stage, more than any other, that *B.burgdorferi* can be cultured from the skin lesions with higher success rate (39). In Early localized infection the mean response of peripheral blood mononuclear cells, to *B.burgdorferi* antigens, is low and specific antibodies to the spirochete are lacking (40, 41).

Early Disseminated Infection:

Days to weeks after its transmission to the patient, *B.burgdorferi* sensu stricto may spread through blood or lymph and affect multiple systems like musculoskeletal, neurologic, lymphatic and respiratory. It can also affect the eyes, heart, liver and kidneys (33). The appearance of multiple EM is a sign that the spirochete is disseminating(42). Despite of the variety of regions that can be affected by this spirochete, the most frequently described in this phase are skin, nervous and musculoskeletal systems (32).

At this stage, patients may start to have symptoms of acute Lyme neuroborreliosis, which include headaches and mild neck stiffness as manifestations of lymphocytic meningitis, radiculoneuritis and cranial neuropathy (43). This last one, can lead to Bell's palsy, a common manifestation described at this stage, affecting one or both sides of the face (42, 43). Several weeks after the infection onset, some untreated patients can develop cardiac abnormalities, with the most common being asymptomatic atrioventricular block from first to third degree, which are often reversible after antibiotic treatment (44, 45).

In Europe, *B.garinii* infection is more associated with dissemination to both central and peripheral nervous systems. One of its manifestations is Bannwarth syndrome, characterized by painful radiculoneuritis and lymphocytic pleocytosis in the cerebrospinal fluid, which is often followed with peripheral paresis (46, 47).

Neurological abnormalities can also occur in infections that have *B.afzelii* as its cause, but the clinical features are usually less specific and harder to diagnose (46). In the case *B.afzelii*, which rarely disseminates through other organs, the skin, not only on the region of the tick bite, is the most affected (17).

Late infection:

In this stage the spectrum of disease manifestations varies more than at any other phase when typical Europe and U.S.A. cases are compared (17, 42). In U.S.A., the typical manifestations are intermittent swelling of large joints accompanied by pain, with knees being the most affected, that can occur for several years. However, in some cases, patients present persistent synovitis for 4 to 5 years (17). The negative results of polymerase chain reaction (PCR) in this patients propose that there is no active infection, suggesting that in genetically susceptible patients, *B.burgdorferi* may induce an immune response with

autoreactive features that continues to occur, for months to years, after the bacteria has been killed (33, 48). This post-infectious persistence of the symptoms is described as antibiotic refractory Lyme arthritis (49).

Arthritis caused by Lyme borreliosis in late infection in Europe may also occur, but this manifestation tends to appear more frequently in the early stages of the disease (42). Acrodermatitis chronica atrophicans (ACA) is probably the most common manifestation of late disease and is almost exclusively caused by *B.afzelii* infection (42). It can occur several years after the tick bite, which most of the times leads the patient to not associate it with the event. This clinical manifestation, that occurs principally in elderly women, starts with a inflammatory phase, resulting in a characteristically bluish-red discoloration of the skin, often in one extremity of the body, that may continue through years to decades (50). Although the culture of *B.burgdorferi* spirochetes is rare in patients at late stage of the disease, a previous study reported the successful isolation of spirochetes from a patient with ACA for more than 10 years, suggesting that spirochetes may survive in human body for extensive periods of time (51).

Patients with late infection, with *B.garinii* as the infectious agent, tend to present severe chronic encephalomyelitis, resulting in cranial nerve paralysis, cognitive difficulties or paraparesis. In the U.S.A. chronic neurological disease has also been reported, though with less-severe abnormalities. In both cases, diagnosis is supported by the presence of intrathecal antibodies (52, 53).

1.7. Diagnosis

With the exception of EM which is diagnosed clinically, the other manifestations of Lyme borreliosis are normally diagnosed accordingly the characteristic clinical symptoms of the disease along with serological testing (54).

The methods that are currently used in laboratorial diagnosis of Lyme disease are divided in two different approaches: the direct methods, to detect *B. burgdorferi s.l.*, and the indirect ones that detect an immunological response against this pathogen (55).

1.7.1. Direct methods

The direct methods used in diagnosis include culture of *B. burgdorferi* and PCR. This approaches are still challenging in obtaining a correct diagnosis due to a low amount of this bacteria in most clinical samples, resulting in a low sensitivity. Although this tests can give important information of the infection, currently, none of them is used as a common practice for Lyme disease diagnosis (55).

Culture:

Culture of *Borrelia* species (spp.) has been essential for comprehension of Lyme borreliosis, and remains the gold standard for diagnosis confirmation. Despite its importance, culture is not a common practice to diagnose Lyme disease for several reasons. This include the long incubation time and low sensitivity of this method, due to the scarcity of bacterial burden in patients and difficult growth of the spirochete (55). Due to slowly replication of *Borrelia,* cultures are only considered negative after 8 to 12 weeks (56).

Positive cultures depends of the specimen, the stage of the disease and the *Borrelia* species involved in the infection. Culture of skin biopsies from EM presents a sensitivity of 40 to 60% (55). For infections caused by *B.burgdorferi sensu stricto* (*s.s.*), better results are obtained in skin biopsies from patients recently infected with small EM lesion, while in *B.afzelii* infection, successful culture occurs principally from skin biopsies from larger lesions, in patients infected up to 30 days (57, 58).

Polymerase chain reaction:

PCR has been a method with great importance for detection of microorganisms in various types of samples obtained from patients. PCR detection of *B.burgdorferi s.l.* has been of interest for Lyme disease study for almost three decades. This method presents high specificity in *Borrelia* spp. detection and likewise culture, presents the advantage of detecting infection sooner than serological tests (59). Despite being a good support for confirmation of diagnosis made from serological tests, this method is not a common practice in laboratorial diagnosis due to lack of sensitivity (18, 55). Thus, a negative PCR test does not exclude the possibility of having the disease and a positive result may not necessarily mean that the patient has an active infection (60).
The first PCR used in specific amplification of *Borrelia burgdorferi s.s.* from culture was first reported in 1989 (61), and since then, many studies have been carried to detect *Borrelia* species in patients' samples. This include skin biopsies, blood, cerebrospinal fluid, synovial fluid and urine. The sensitivity of PCR detection varies a lot accordingly to the specimens used and the time they were collected (55). Also different target genes have been described used in *Borrelia* spp. detection PCR, being the most frequently used the16S ribosomal RNA (rrs), the flagellin (flaB), recA and p66 genes encoded on the chromosome and the ospA gene encoded on a linear plasmid (62).
PCR testing in skin biopsy samples of patients presenting EM has good sensitivity results (≈69%, ranging from 36% to 88%), but in this cases, clinical diagnosis its usually enough to confirm the infection (18, 59, 62). However, this test may be a good option to obtain a clear diagnosis in patients with dubious shapes of EM and it can test if there is presence of co-infections in patients reporting unusual clinical manifestations (63).

Blood and cerebrospinal fluid PCR detection, described in MEDLINE-indexed studies from 1991 to 2003, have a low mean sensitivity (≈14%, ranging from 0% to 100% and 38% ranging from 12% to 100% respectively) (59). However, this lack of sensitivity may be associated with incorrect timing in sample collecting accordingly to the stage of the disease, since recent studies have shown much better results (63, 64).

PCR detection of *Borrelia* spp. in synovial fluid present a high sensitivity, (mean 78%, ranging from 42% to100%), and has been used to support diagnosis from serological testing in late stages of Lyme disease, when patients suffer from arthritis (18, 59). In case of urine samples, diagnosis by PCR showed poor results, thus, this specimen is not reliable for correct diagnosis (18, 55).

1.7.2. Indirect methods

The indirect methods are based on the detection of the host's immunological system response against to the microorganism causing the disease. Regarding Lyme disease, currently, the antibody-based assays are the only method approved by the United States Food and Drug Administration (FDA) for laboratorial diagnostic tests (55, 65). However, the practices used in different laboratories, and different interpretation of the test results may lead to low specificity of this method. In 1995, in order to improve specificity of serological testing in Lyme disease diagnosis, Centers for Disease Control and Prevention presented a two-tier approach consisting in an enzyme immunoassay (EIA) or, less frequently, an immunofluorescent assay (IFA), complemented with a Western immunoblot (WB). Basically, if a result is negative by a sensitive EIA or IFA no further test is needed, on the other hand, if this test is positive or equivocal, the sample must be submitted to a standardized Western immunoblot for detection of immunoglobulin M (IgM) or immunoglobulin G (IgG) antibodies to *B.burgdorferi* in serum. In the first 4 weeks of infections both IgM and IgG testing are recommended, however, after this period only IgG test should be performed. To be considered positive, the IgM WB must have at least 2 of the three signature bands, while in the IgG WB 5 of 10 signature bands are needed (66).

The main limitation of two-tier serological tests is that in early localized infection, many false negative results are obtained, clearly due to the time window of the specific antibodies to be produced. IgM antibodies can take 2-4 weeks to be produced in quantities that enable test detection, while IgG antibodies take 4-6 weeks. Consequently, in this stage of the disease (stage 1), serological testing presents a relatively low sensitivity in patients with early localized infection, approximately 46%, but in patients with stage 2 or stage 3 of the infection, early disseminated and late disease, the sensitivity of this method increases to approximately 90% and 99%, respectively (65). False positives results may be obtained in patients with disease that are known to produce antibodies that cross react in serological tests for *B.burgdorferI* (67).

1.8. Treatment

For treatment of patients with early localized (stage 1) or early disseminated disease (stage 2), presenting EM and associated symptoms, without specific neurologic symptoms or advance atrioventricular heart block, doxycycline, amoxicillin or cufuroxine axetil have shown remarkable effectiveness , thus, this antibiotics are the principal recommended. Doxycycline is usually the antimicrobial agent recommended in this stages of the disease, since it presents also the advantage of being effective for the treatment of some co-infections like Human granulocytic anaplasmosis. In the case of early Lyme disease presenting acute neurologic symptoms, intravenously ceftriaxone or cefotaxime, are often prescribed (45).

Macrolide antibiotics should not be used as first-line therapy since they present less effectiveness, however, for patients who are intolerant or should not take the antibiotics previously referred, macrolide antibiotics like azithromycin, clarithromycin and erythromycin can be used (45).

2. Next Generation Sequencing

2.1. Sequencing overview

In 1990, a very ambitious project, named The Human Genome Project (HGP), has been launched with the purpose of sequencing, with high accuracy, almost entirely the euchromatic part of the human genome. This project was carried out by the International Human Genome Sequencing Consortium, a collaboration between twenty centers distributed in six different countries. This project was able to assemble approximately 99% of the euchromatic sequence of the human genome, in a total of 2,85 billion nucleotides with an error associated of only 1 event per 100,000 bases (68). The HGP was carried out using Sanger sequencing and it took 13 years to its completion, with an estimated cost of 3 billion dollars (68, 69).

With successful results, this project was able to provide reference sequences, not only for the human genome but also for simpler organisms with smaller genomes (68). However, it consumed huge amount of time and resources. Therefore, the demand for faster, higher throughput, and cheaper technologies increased significantly among the scientific community (69, 70). The reference sequences obtained, led to the development of new approaches to re-sequencing in which smaller reads are mapped to the reference to indentify genetic variation (71). This new approaches for sequencing became known as Next Generation Sequencing (NGS).

Different NGS methods were thus developed after the completion of the HGP, providing many improves to Sanger sequencing, being the most important the ability of sequencing millions of DNA fragments simultaneously (massively parallel sequencing) with a high throughput (70). The differences of the launched NGS platforms are mainly based in the approach used in the

sequencing reaction. Second generation sequencing instruments can be categorized in 4 different types accordingly to the type of sequencing used. This include pyrosequencing, sequencing by synthesis, sequencing by ligation and ion semiconductor sequencing. The third generation of sequencing presents novel approaches with the ability to sequence at a single molecule level. These systems bring several advantages in some fields, highlighting the long read sequences obtained (up to hundred thousand bp), the portability and speed of this devices, and the possibility of collecting and analyzing sequencing data in real time. However, although this methods have been arousing curiosity among the scientific community, this systems are not still widely used due to some accuracy problems(70, 72).

The creation of NGS platforms has made sequencing accessible to more labs, rapidly increasing the amount of research regarding nucleic acids. These instruments have shown a lot of applications in fields such as genetic diseases research, personalized medicine and clinical diagnostics. Nowadays, due to being cost effective, the NGS instruments that are most commonly used in clinical diagnostics are Illumina's MiSeq and the Ion Personal Genome Machine (PGM), which use sequencing by synthesis and ion semiconductor sequencing methods, respectively(73).

2.2. Illumina MiSeq workflow

The Illumina sequencing workflow is similar in the various instruments, suitable for different applications, developed by the company. This workflow is divided in four steps: sample preparation, cluster generation, sequencing and data analysis (71).

2.2.1. Sample preparation

In the sample preparation step, adaptors are added to the extremities of the DNA fragments. This adaptors provide a complementary region to the insertion of the sequencing binding site, the indexes that will allow the correct read attribution to the respective sample and the region complementary to the fixed oligos present in the flow cell. This insertion is performed by an eight cycle PCR (71).

2.2.2. Cluster generation

The cluster generation step is where the DNA fragment is amplified. Illumina MiSeq uses a flow cell with one lane, which is coated with two different oligos. The adaptor region of one of the strands of the DNA fragment hybridizes with one of this oligos. Then, a DNA polymerase generates the complement of the hybridized strand creating a double stranded molecule. This molecule is after denatured, and only the complementary sequence stays fixed to the cell while the original template is washed. After this process the bridge amplification starts to occur. Basically, the strand that is fixed to the cell bends over and the adaptor region present in the opposite extremity of the fixed region hybridizes with the second type of fixed oligo.

The DNA polymerase will then generate a double stranded bridge that after being denatured results in two single stranded copies which at this point are both fixed to the flow cell (Figure 3). This process is repeated sequentially for millions of clusters simultaneously resulting in a massive amplification of the fragments. Finally, to begin the sequencing step, the reverse reads are washed (71).

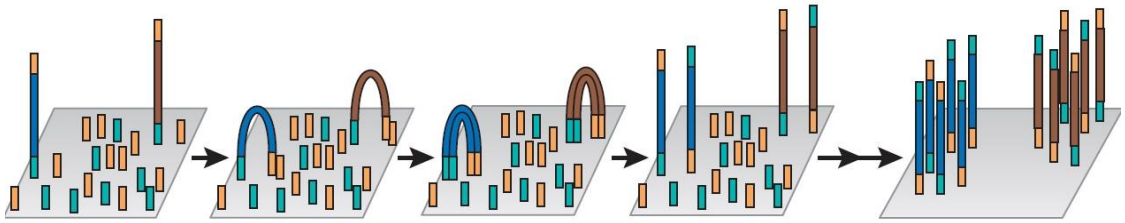


Figure 3- Illumina's cluster generation process: the complementary strand of the original template that hybridized with the first type of oligo is fixed to the flow cell. This strand bends over and hybridization of the second type of oligo with the complementary extremity of the strand occurs. The DNA polymerase generates a double stranded bridge which, after being denatured, results in two strands fixed to the cell. This process is repeated sequentially for millions of clusters simultaneously resulting in massive amplification of all the fragments (Adapted from (72)).

2.2.3. Sequencing

In the Illumina's MiSeq sequencing by synthesis approach all four nucleotides are added simultaneously. This nucleotides are reversibly fluorescent labeled and have the 3'-OH group chemically blocked, allowing only one base incorporation at a time. The forward read sequencing starts with the extension of the first sequencing primer. After nucleotide incorporation, the remaining nucleotides are washed away and the signal from laser-induced excitation of the fluorophores is read from each cluster, with an associated quality value for each base call. The fluorescent molecule and the terminator group are then cleaved and washed away and a new cycle commences. The length of the read is determined by the number of cycles. After finishing the forward read, the attached strand folds over, bridge amplification occurs and after being denatured, the forward read is washed. Reverse reads are then obtained through the same process as forward reads. In the end of the sequencing run, a base calling algorithm assigns the sequences and a quality value (phred score) to each read. The error rate in this system can increase as the reaction proceeds due to incomplete removal of the fluorescent molecule which will cause background noise in the acquired signal. Thus, it is important to bear in mind the chosen size of the fragment, in order to obtain an overlap of at least 50 bp between forward and reverse reads to overcome the lack of quality of the reads ends (71).

2.2.4. Data analysis

At the end of the sequence process, millions of reads were generated. MiSeq sequencer automatically attributes the reads to the correct sample in the library pool, based on the combination of indexes used. The first step in the analysis must be a quality control of the reads, in order to understand what parameters to use in filtering to obtain accurate results. After filtering, forward and reverse paired reads are merged to obtain a contiguous sequence. Further analysis should be performed conveniently for the intended application (71).

II. AIM

Several studies have been showing that the two-tier serological testing for Lyme disease, the only type of test approved for its diagnosis by FDA, show inaccuracy in detecting Lyme disease in the early localized stage of the infection. Since the antibodies against the pathogen can take weeks to be produced, efforts have been made, throughout the last decades, in order to present direct methods as an alternative to the diagnosis at this stage of the disease. This methods include culture and PCR. Despite different approaches that have been described, concerning the specimen and the methods used, the results obtained with this tests also suffer from a high rate of false negatives.

The present study intends to develop a method suitable for the diagnosis of Lyme borreliosis, specially in an early phase of the infection, using Illumina's MiSeq Next Generation Sequencing platform. The reason why this technology was chosen was based in its ability to sequence DNA present in low concentrations and due to being cost-effective making it applicable to diagnostic practices.

Since many reports have described cases of patients infected with more than one pathogen transmitted by *Ixodes ricinus* complex, this study aims to create a panel capable of detecting not only the presence of *Borrelia* spp. but also the presence of the most common co-infections. This panel will target fragments of genes that enable the determination of the genospecies responsible for the infection, which may give important information for the understanding of the clinical manifestations and to appropriate treatment.

III. MATERIAL AND METHODS

1.Biological material

1.1. Genomic DNA

Genomic DNA from *Borrelia* spp. and from the species responsible for the co-infections of Lyme disease were obtained by contacting other researchers/institutes. *Borrelia burgdorferi*, *Borrelia afzelii* and *Ehrlichia canis* genomic DNAs were provided by Instituto de Higiene e Medicina Tropical (IHMT-UNL). Genomic DNA of *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* HGE1 was provided by Prof. Dr. Ulrike Munderloh from Department of Entomology, University of Minnesota. Genomic DNA of *Bartonella henselae* Houston-1 and *Bartonella henselae* Marseille were provided by Prof. Dr. Volkhard Kempf, from Institute for Medical Microbiology and Infection Control, Frankfurt. Genomic DNA of *Coxiella burnetti* was provided by Prof. Dr. Federico Capuano from Department of Food Inspection from Istituto Zooprofilattico Sperimentale del Mezzogiorno. Genomic DNA of *Rickettsia rickettsii* was provided by Tina Clark, Microbiologist at Laboratory of Intracellular Parasites, NIAID, NIH. Genomic DNA of *Babesia microtii* Gray was purchased from the ATCC-LGC Standards Partnership (Spain).

1.2. Blood samples

Blood samples from six anonymized individuals were used in this work. Samples from five patients with Babesiosis were provided by CDC (Atlanta, U.S.A.). This patients' blood has been tested positive, in CDC, for the presence of *Babesia* species by qPCR. The sixth sample, used as the negative control, is the blood of an individual belonging to a group who never reported a tick bite and from a non-endemic area, which means it is most likely not to have Lyme disease nor co-infections. All blood samples were collected in EDTA vacutainer tubes.

1.2.1. Positive control

For the positive control used in this work, once the purpose was to evaluate the seven primer pairs from the panel, an aliquot of the negative control was spiked with approximately 100,000 *E.coli* cells from each of the ten different transformed cells containing the amplicon of the different genes selected from each species. The *E.coli* cells transformation procedure, better explained later in this chapter, was successful for all the fragments of the target genes, with exception of the fragment of the 18S gene from *Babesia microti*, since the results from Sanger sequencing after cell transformation showed the loss of the majority of the region where the primer should hybridize (data not shown). The estimated number of cells/mL was calculated using the OD value at 600nm of each sample, obtained with Nanodrop equipment, multiplied by 8×10^8, as suggested Agilent Genomics BioCalculator.

2. Cell preparation for positive control and PCR

2.1. Target Genes

Table 1- Identification of the species gathered for this study, the target gene for each species and the size of the amplicon obtained in PCR. (* Primer sequences are in the Table 2, Section 3.2.)

Genus	Species	Strain	Target Gene	Primers* (Size of the amplicon, bp)
Borrelia	*burgdorferi*	B31	Flagellin (fla)	BorF+BorR(429)
	afzelii	-		
Babesia	*microti*	Gray	18S rRNA	BabF+BabR(512)
Anaplasma	*phagocytophilum*	-	Molecular chaperone (groEL)	AnaF+AnaR(480)
Ehrlichia	*chaffeensis*	-		EhrF+EhrR(477)
	canis	-		
Bartonella	*henselae*	Houston-1	Riboflavin synthase (ribC)	BarF+BarR(437)
	henselae	Marseille		
Rickettsia	*rickettsii*	-	Cytrate synthase (gltA)	RickF+RickR(454)
Coxiella	*burnetii*	-	Isocitrate dehydrogenase (icd)	CoxF+CoxR(450)

The interest genes chosen to be sequenced from each of the ten species were amplified, from the respective genomic DNAs, through a standard PCR using the species specific designed primers, in the following conditions:

[]	**Reagents**	**Volume**
10 x	PCR Reaction Buffer	2.5 µL
25 mM	$MgCl_2$	1.5 µL
5 mM	dNTP	1µL
10 mM	Fwd Primer	1 µL
10 mM	Rev Primer	1 µL
10 U/µL	Surf Hot Taq Polymerase	0.2 µL
1-100 ng/ µL	Extracted DNA	2 µL
	Water (mQ)	15.8 µL
	Total	25 µL

PCR program

		Time	Temp.	
1.	Initial Denaturation	15 min	95ºC	
2.	Denaturation	30 sec	95ºC	35x
3.	Annealing	30 sec	55ºC	
4.	Elongation	1 min	72ºC	
5.	Final Elongation	5 min	72ºC	
6.	Hold	∞	4ºC	

To confirm if amplification occurred, the PCR products were visualized through an agarose gel electrophoresis (AGE) (conditions described in section 3).

All the samples containing a band, in the gel, for the expected size were purified using Magnetic Beads (MCLAB, San Francisco, U.S.A.) to remove, mostly, the primer-dimers formed in the reaction. The purification was performed accordingly to the manufacturer's protocol, using a 1:1.8 DNA/Magnetic Beads ratio. To ascertain if successful purification occurred, another AGE was performed.

The PCR products were then sequenced by Sanger method at STAB VIDA Lda., in order to verify the correct amplification of the desired sequences. All the sequences generated by Sanger sequencing throughout this work were assembled and analyzed with Sequencher 4.10.1 and FinchTV Version1.4.0 Software, respectively.

2.2. Primers

Species specific primers were designed and analyzed using Oligo Explorer 1.1.1 and Oligo Analyzer 1.0.2 software respectively, and finally a primer BLAST was performed in NCBI platform. The primers were acquired from STAB VIDA, Lda. (Portugal). Since the downstream application intended was NGS sequencing using Illumina MiSeq System, a common overhang adaptor was inserted in all the primers as described in 16S Metagenomic Sequencing Library Preparation (74). Forward adaptor sequence: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG and reverse adaptor sequence: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG. The MiSeq Reagent v3 was used with the 600 cycle kit to obtain read lengths of 2x300bp. Primers were designed to obtain amplicons with a size around 450 bp, with the purpose of having an overlap of at least 50 bp between forward and reverse reads.

Table 2- Designation, sequences and melting temperatures of the primers designed and 16S primers used in 16S metagenomic sequencing library preparation(74). All sequences include the common forward and reverse adaptors needed for hybridization to the flow cell in NGS. Tm refers exclusively to the melting temperature of the primer sequence designed to hybridize with the target, while Tm* refers to the melting temperature of the whole primer, include the adaptor sequence.

Designation	Sequence (5' to 3')	Tm (°C)	Tm* (°C)
BorF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCAGTTCAATCAGGTAACG	56.6	75.53
BorR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCACCAGARAAHAGATTTGC	57.8	73.2
BabF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAATGATGGKRAYCTAAACC	56.7	73.1
BabR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGTTAAATACGAATGCCC	56.6	73.3
AnaF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTCACTGTAGCGATTAGTAAGC	59.5	75.0
AnaR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATGTTCTTGTCTCCATTTGC	58.7	73.9
EhrF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGGACCTAAAGGACTTACTG	59.3	75.4
EhrR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCWGAAATAGTAGCAACTTGAGC	59.6	74.8
BarF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTTTGCTGTWGARGCRTG	59.4	76.0
BarR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCATYTCAAGYGTATGRCG	59.0	74.2
RickF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGATATAAGTAGGGTATCTGCGG	56.8	74.0
RickR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATTCTAATAGCGGTAAGTTCG	58.0	73.5
CoxF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGTTAGACCTGTATGTTTGCC	58.6	74.7
CoxR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATGTTACCTTTATGGACGAGC	59.9	74.5
16SF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG	63.0	78.4
16SR	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC	50.9	72.7

2.3. Cloning

The cloning process was performed using the CloneJET PCR Cloning Kit, purchased from Thermo Scientific (Waltham, MA, USA). Briefly, pJET1.2/blunt is a linearized cloning vector capable to accept inserts from 6 bp to 10 kb. The Sticky-End Cloning Protocol, which consists in a blunting reaction of the amplicons and a ligation reaction to the cloning vector, was followed as described by the manufacturer, with exception of the added volume of nuclease free water, due to lack of starting concentration of DNA.

2.4. Transformation

The transformation step was performed using *E.coli* NZY5α competent cells, purchased from NZYTech (Lisboa, Portugal). This cells show similar properties to DH5α, which are suitable for high efficiency transformation. The transformation protocol was performed accordingly to the manufacturer's. Briefly, 10μL of the ligation reaction were added to a tube containing 100μL of competent cells. After the incubation and heat-shock, SOC medium was added and 250μL of the transformed cells were inoculated on LB agar (Miller) plates. A competent cells control plasmid solution provided with the NZY5α was used as a positive control and milli-Q H2O was used as the negative control. The cell culture was incubated in CO_2 incubator overnight at 37ºC with 5% CO_2.

2.4.1. Confirmation of correct insertion

From each of the ten LB agar plates, two different colonies were picked and inoculated in 2 mL of LB Broth medium containing 100 μg/mL ampicillin and incubated at 37ºC o/n with agitation. Then, 2μL of each tube were used directly on a standard PCR using the pJET1.2F and pJET1.2R pair of primers provided in the CloneJET PCR Cloning Kit. After seeing the PCR products in the AGE, one of the two colonies of each species was selected to sequence by Sanger. All sequences were confirmed to be completely inserted in *E.coli* cells with exception of both *Babesia microti* colonies, which lacked the final region of the fragment, needed for primer to hybridize.
To verify the presence of the primers in the sequences, pDRAW32 1.0 Revision 1.1.133 ACACLONE Software was used.

3. Agarose Gel Electrophoresis

All the AGE performed throughout this work used 1,5% agarose gel prepared in Tris-Acetate-EDTA (TAE). The nucleic acids were stained with GelRed (1:50000) which is a stable fluorophore that intercalates the nucleic acids without impair their migration in the gel. The complex dsDNA-GelRed when excited with UV light emits fluorescence that is captured by the camera present in the trasilluminator equipment. The conditions of the AGE were 120 Volt (V) for 20 minutes in TAE 1x.

4. DNA Extraction

Extraction of genomic DNA from the blood samples was performed with the GE illustra blood genomicPrep Mini Spin Kit purchased from GE Healthcare (Little Chalfont, United Kingdom). The DNA extraction was performed as described in the manufacturer's protocol for samples between 50-300µl. Briefly, in this procedure, a chaotropic agent is used to extract de DNA from nucleated blood cells and promote the selective binding of DNA to the silica membrane column. The protease used in the kit is proteinase K, which is active even in the presence of detergents and chelating agents, like EDTA. The kit also provides a low ionic strength elution buffer which allows the DNA to be stored. From each sample of whole blood, 200µl were used in the DNA extraction procedure, since this is the indicated volume to obtain the optimal performance.

5. Library Preparation for NGS sequencing

The seven samples, five from Babesiosis patients and the positive and negative controls, were submitted to four different conditions: whole blood used in multiplex PCR with the primers designed (B+MX), extracted DNA also used in multiplex PCR (DNA+M) and extracted DNA with the multiplex PCR and the PCR using 16S primers.

5.1. Amplicon PCR

The four distinct PCR had different reagents and PCR programs. The reactions using whole blood were performed using the enzyme Hemo KlenTaq (New England BioLabs, Massachusetts, U.S.A.), while in the reactions using extracted DNA the enzymes Surf HotTaq DNA Polymerase (STAB VIDA, Portugal) and KAPA Hifi HotStart ReadyMix (KAPA Biosystems, were used for multiplex PCR and 16S PCR respectively.

5.1.1. Whole blood + Multiplex conditions (B+Mx)

[]	Reagents	Volume
5 x	Hemo KlenTaq Reaction Buffer	10 µL
10 mM	dNTP	1µL
10 µM	Fwd Multiplex Primers	1.5 µL
10 µM	Rev Multiplex Primers	1.5 µL
n.a.	Hemo KlenTaq	4 µL
	Whole Blood	10 µL
	Water (mQ)	22 µL
	Total	50 µL

PCR program

		Time	Temp.	
1.	Initial Denaturation	3 min	95ºC	
2.	Denaturation	20 sec	95ºC	35x
3.	Annealing	30 sec	45ºC	
4.	Elongation	1 min	68ºC	
5.	Final Elongation	10 min	68ºC	
6.	Hold	∞	4ºC	

5.1.2. Extracted DNA + Multiplex conditions (DNA+Mx)

[]	Reagents	Volume
10 x	PCR Reaction Buffer	2.5 µL
25 mM	$MgCl_2$	1.5 µL
5 mM	dNTP	1µL
10 µM	Fwd Multiplex Primers	1 µL
10 µM	Rev Multiplex Primers	1 µL
10U/µL	Surf Hot Taq Polymerase	0.2 µL
10-50 ng/µL	Extracted DNA	2 µL
	Water (mQ)	15.8 µL
	Total	25 µL

PCR program

		Time	Temp.	
1.	Initial Denaturation	15 min	95ºC	
2.	Denaturation	30 sec	95ºC	35x
3.	Annealing	30 sec	62ºC	
4.	Elongation	1 min	72ºC	
5.	Final Elongation	5 min	72ºC	
6.	Hold	∞	4ºC	

5.1.3. Whole blood + 16S primers conditions (B+16S)

[]	Reagents	Volume
5 x	Hemo KlenTaq Reaction Buffer	10 µL
10 mM	dNTP	1µL
1 µM	Fwd 16S Primer	1.5 µL
1 µM	Rev 16S Primer	1.5 µL
n.a.	Hemo KlenTaq	4 µL
	Whole Blood	10 µL
	Water (mQ)	22 µL
	Total	50 µL

PCR program

		Time	Temp.	
1.	Initial Denaturation	3 min	95ºC	
2.	Denaturation	20 sec	95ºC	35x
3.	Annealing	30 sec	45ºC	
4.	Elongation	1 min	68ºC	
5.	Final Elongation	10 min	68ºC	
6.	Hold	∞	4ºC	

5.1.4. Extracted DNA + 16S primers conditions (DNA+16S)

[]	Reagents	Volume
2 x	KAPA HiFi HotStart Ready Mix	12.5 µL
1 µM	Fwd 16S Primer	5 µL
1 µM	Rev 16S Primer	5 µL
10-50 ng/µL	Extracted DNA	2.5 µL
	Total	25 µL

PCR program

		Time	Temp.	
1.	Initial Denaturation	3 min	95ºC	
2.	Denaturation	30 sec	95ºC	35x
3.	Annealing	30 sec	55ºC	
4.	Elongation	30 sec	72ºC	
5.	Final Elongation	5 min	72ºC	
6.	Hold	∞	4ºC	

5.1.5. Amplification verification and purification

The resulting 28 PCR products were then submitted to an AGE to verify the amplification. Next, 10µL of each PCR product were purified using 18µL of AxyPrep Mag PCR Clean-up magnetic beads (Axygen, Corning, USA) following the manufacturers' protocol. After completing purification of the products, another AGE was performed to guarantee that the products hadn't been lost during the process.

5.2. Index PCR

A second PCR was performed to prepare samples to Illumina MiSeq sequencing, as described in Metagenomic Sequencing Library Preparation (74), to insert the indexes into the samples. When preparing this second PCR, a different Nextera XT Index 1 primers and Nextera XT Index

2 primers were used for each sample as it is mandatory for the correct attribution of the reads to each sample in the sequencing process.

[]	Reagents	Volume
2 x	Kapa HiFi HotStart ReadyMix	12.5 µL
	Nextera XT Index Primer 1	2.5 µL
	Nextera XT Index Primer 2	2.5 µL
	DNA	2.5 µL
	Water (mQ)	5 µL
	Total	25 µL

PCR Program

		Time	Temp.	
1.	Initial Denaturation	3 min	95ºC	
2.	Denaturation	30 sec	95ºC	8x
3.	Annealing	30 sec	55ºC	
4.	Elongation	30 sec	72ºC	
5.	Final Elongation	5 min	72ºC	
6.	Hold	∞	4ºC	

The AGE, purification and subsequent AGE of the samples are repeated, as described previously, after the second PCR. The next step of the library preparation was the quantification of the 28 samples and was performed with Qubit 2.0 Fluorometer using Qubit dsDNA BR Assay Kit (Invitrogen, California, EUA), accordingly the manufacturer's protocol. Briefly, Qubit assay is highly selective four double-stranded DNA and provides high accuracy determining sample concentration from 100 pg/µL to 1000 ng/µL. The working solution was prepared with a dilution of Qubit dsDNA BR Reagent 1:200 in Qubit dsDNA BR Buffer and the value of the sum between sample and working solution was always 200 µL. The calibration of the equipment was performed using 10 µL of each of the two standard DNAs provided in the kit, while 5 µL of each of the 28 samples were used for quantification. Finally, the DNA concentration was calculated in nM (equation 1), accordingly the size of the DNA amplicons and dilutions in 10 mM Tris pH 8.5 were made based on the coverage intended to each sample. For the 14 reactions using the panel of primers designed, dilutions were made to obtain 600,000 reads. For the seven samples from extracted DNA with 16S primers, dilutions were made to obtain 200,000 reads, while the samples of whole blood with 16S primers, were diluted to obtain 400,000 reads, simply due to the fact that they were sequenced in another sequencing run with less samples. Samples were then sequenced in Illumina MiSeq with the MiSeq Reagent kit v3 in the 600 cycles format.

$$\frac{DNA\ concentration\ in\ ng/\mu L}{660\ g/mol \times average\ library\ size} 10^6 = concentration\ in\ nM$$

Equation 1- Conversion of DNA concentration from ng/µL to nM, based on the size of DNA amplicons (74).

6.Data Analysis

6.1. Software

6.1.1. Qiime2 2018.6

Raw data originated in sequencing process was analyzed with the open source software Qiime2 2018.6. The first step in this analysis consisted in importing the data and simultaneously demultiplexing it. The following step was the filtration of the data. The forward and reverse reads were trimmed by 17 and 21 bases respectively, to reduce chimera occurrence. The reads were trunked in the number of the base where the average phred score was still above 20 and before the first accentuated decay peak.

The analysis was performed Qiime software was used through Linux command line the applying the following commands:

Importing Data:

```
> qiime tools import \
        --type 'SampleData[PairedEndSequencesWithQuality]' \
        --input-path Desktop/folder \
        --source-format CasavaOneEightSingleLanePerSampleDirFmt \
        --output-path folder1
```

Data filtering:

```
> qiime dada2 denoise-paired \
        --p-trim-left-f 17 \
        --p-trim-left-r 21 \
        --i-demultiplexed-seqs 16Sdemux.qza \
        --output-dir dada2 \
        --p-n-threads 4 \
        --p-trunc-len-f 292 \
        --p-trunc-len-r 227
```

Obtain the nucleotide sequence of the representative sequences:

```
> qiime feature-table tabulate-seqs \
        --i-data representative_sequences.qza \
        --o-visualization representative_sequences.qzv
```

After this steps, the fasta file with all the representative sequences and its correspondent nucleotide sequences was downloaded.

Obtain the frequency of each representative sequence:

```
> qiime feature-table summarize \
        --i-table dada2/table.qza \
        --o-visualization table.qzv
```

After this steps, the fasta file with all the representative sequences and its correspondent frequency was downloaded.

6.1.2. BLAST2GO

The BLAST of the representative sequences stemming from the 28 samples was performed separately with Blast2GO 5.2.0 Software. The BLAST program used was blastn (task megablast) with the following parameters:

-BLAST expectation value (E-Value): 1.0E-50
-Number of BLAST hits: 20
-Low complexity filter: on
-Word size: 28
-HSP cutoff: 33

The features that had no assigned result, were submitted to a second BLAST with the same parameters with exception of the E-value which was changed to 1.0E-25.

6.1.3. Python

Two tabular format files were created for each sample, one containing the frequency of each feature of the representative sequences while the other contained the feature of the representative sequence and its correspondent BLAST result. With Python's IDLE 3.6.5, a program was created to generate a file attributing the frequency of each feature to the respective BLAST result.

IV. RESULTS

1. Library preparation results

1.1. Amplicon PCR and purification

The amplification products of *Babesia* spp. and possible other microorganisms responsible for co-infections, covered by the designed primers panel, was analysed through visualization of the AGE of each condition for the five blood samples of patients with babesiosis provided by CDC and for negative and positive controls, blood of an individual living in a non-endemic area who never reported a tick bite and an aliquot of the same blood spiked with the transformed *E.coli* cells containing the fragment of the target genes, respectively. After the purification of the PCR products, another AGE was performed to ensure no errors occurred in this step. The results are shown in the figures below.

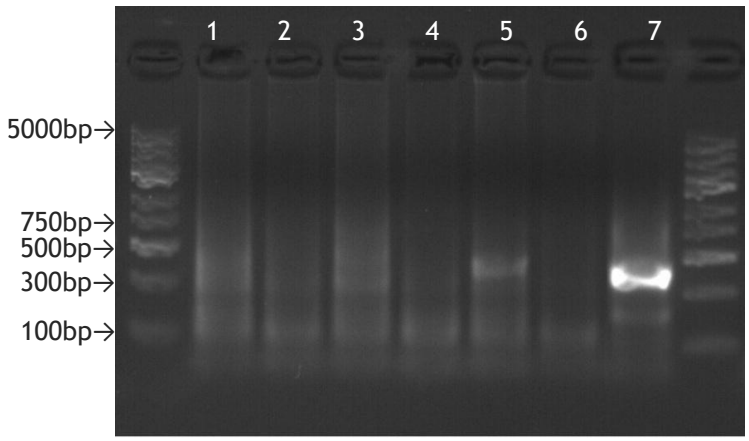


Figure 4- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with B+Mx condition: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

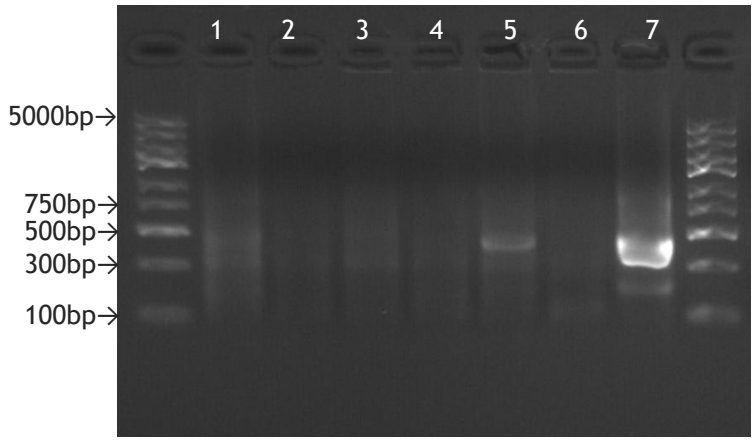


Figure 5- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with B+Mx condition, after purification: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

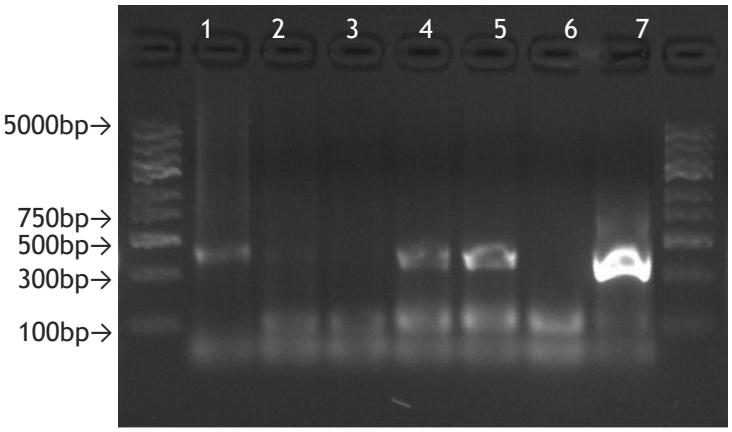


Figure 6- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with DNA+Mx condition: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

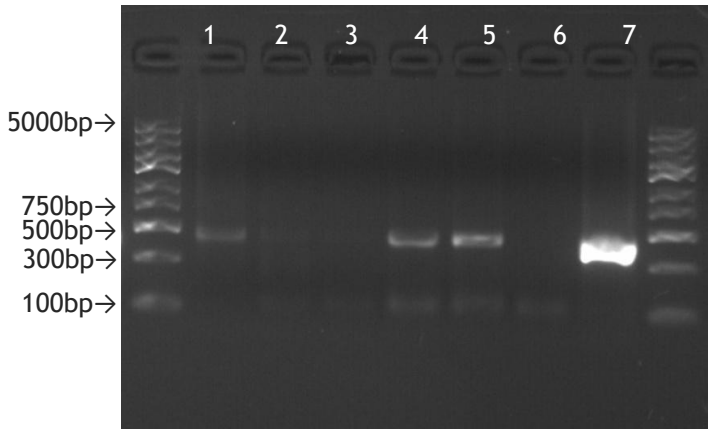


Figure 7- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with DNA+Mx condition, after purification: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

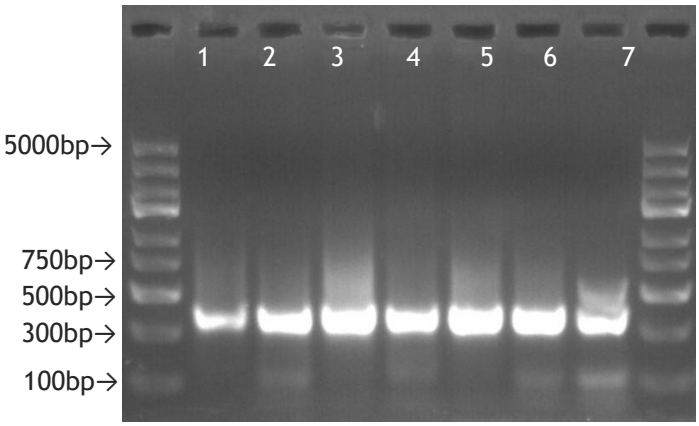


Figure 8- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with B+16S condition: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

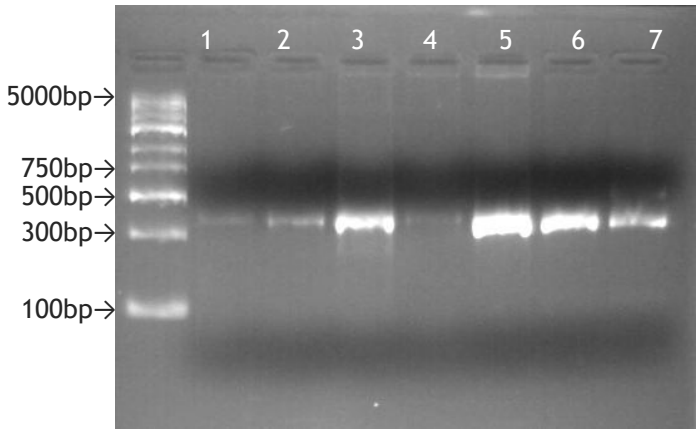


Figure 9- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with B+16S condition, after purification: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

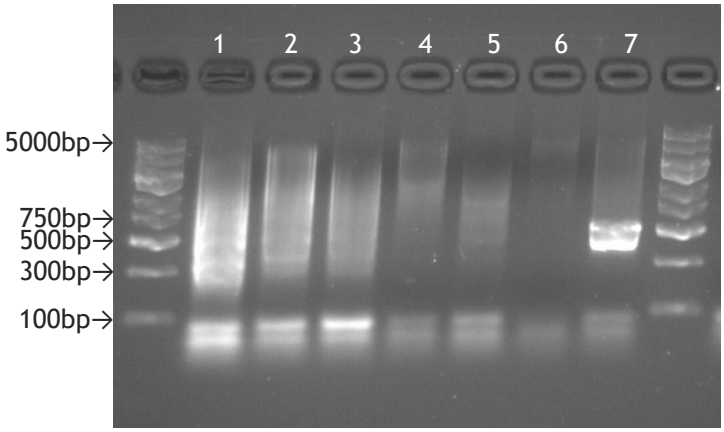


Figure 10- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with DNA+16S condition: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

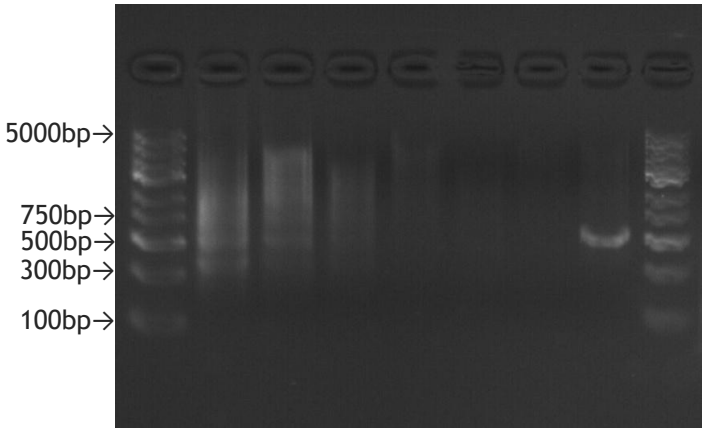


Figure 11- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained with DNA+16S condition, after purification: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

1.2. Index PCR and purification

To analize the second PCR results, used to insert the different indexes in each sample for the sequencing process, AGE were performed, like in the first PCR. This second PCR was successful, as well as the purification step that followed it, for all PCR products obtained from the amplicon PCR of the four conditions. The results are shown in the figures below (AGE of the B+16S condition was not included due to an error when saving).

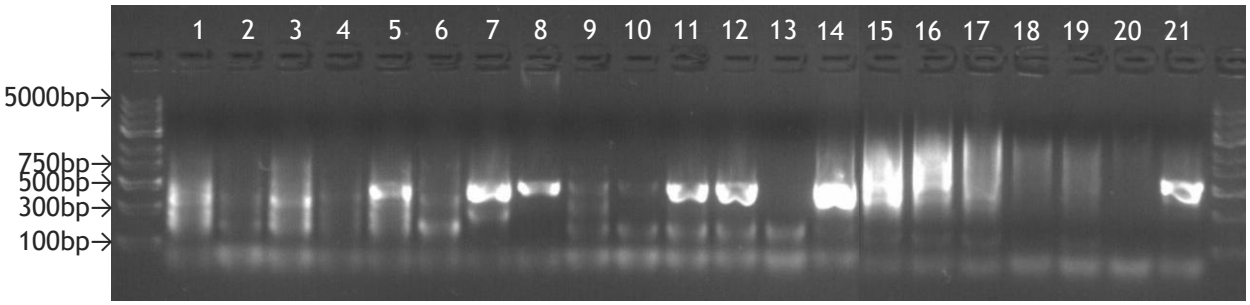


Figure 12- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained in the second PCR of library preparation. B+Mx condition: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control; DNA+Mx condition: 8-Sample 1, 9-Sample 2, 10-Sample 3, 11-Sample 4, 12-Sample 5, 13-Negative control,14-Positive control; DNA+16S condition: 15-Sample 1, 16-Sample 2, 17-Sample 3, 18-Sample 4, 19-Sample 5, 20-Negative control, 21-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

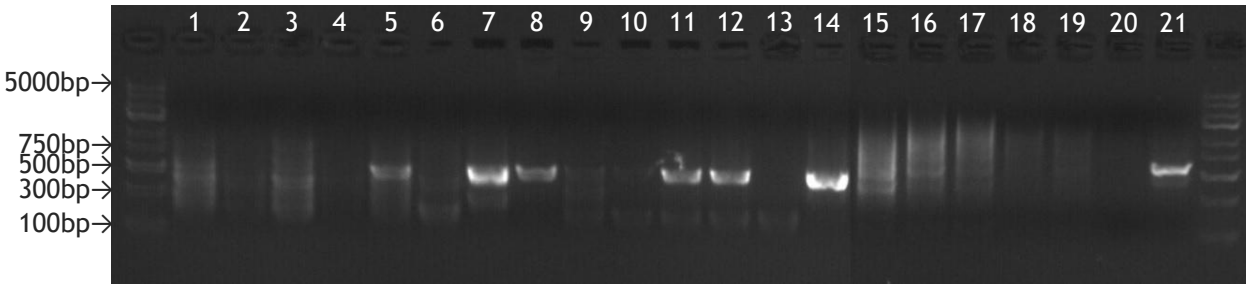


Figure 13- Agarose gel (1.5%) electrophoresis, stained with GelRed, of the PCR products obtained in the second PCR of library preparation, after purification: B+Mx condition: 1-Sample 1, 2-Sample 2, 3-Sample 3, 4-Sample 4, 5-Sample 5, 6-Negative control, 7-Positive control; DNA+Mx condition: 8-Sample 1, 9-Sample 2, 10-Sample 3, 11-Sample 4, 12-Sample 5, 13-Negative control,14-Positive control; DNA+16S condition: 15-Sample 1, 16-Sample 2, 17-Sample 3, 18-Sample 4, 19-Sample 5, 20-Negative control, 21-Positive control. Conditions: 120V for 20 minutes in TAE 1x.

2. NGS results

The libraries prepared were submitted to NGS sequencing, using Illumina Miseq with MiSeq reagent Kit V3 (600-cycles, 2x300bp format). The results, obtained by bioinformatics analyses, were used to create a bar graph for each of the seven samples analysed, including the four different conditions (Section 2.3.), in order to facilitate comparison between them.

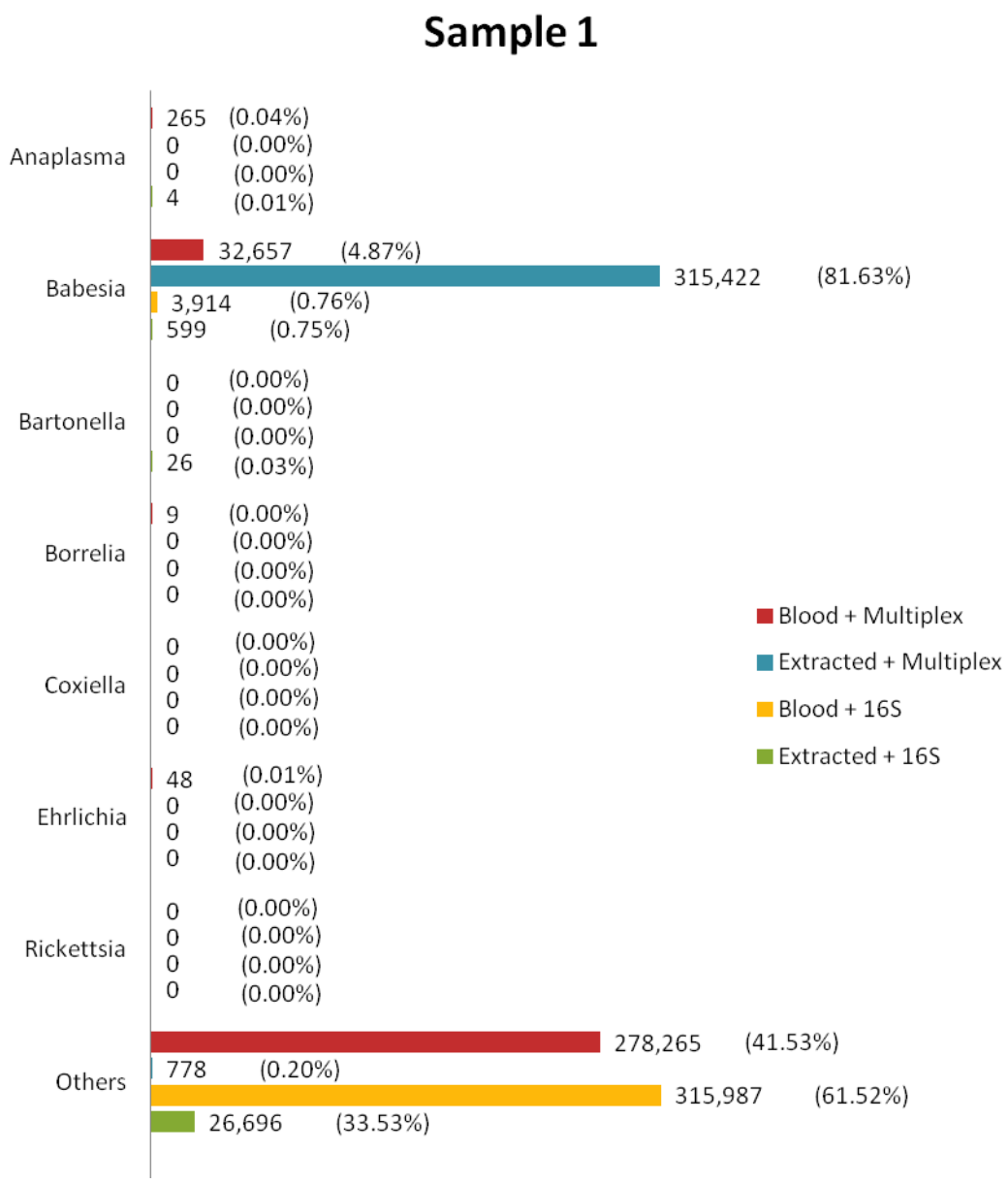
2.1. Unspecific amplification

Unlike the expected, for the samples that were submitted to the PCR with the multiplex primers designed, a high number of reads did not align to any of the seven species considered in this panel. This unspecific amplification occurred mainly in the tests performed directly in whole blood. To facilitate comprehension of the condition specificity, the number of reads attributed to other than the species covered in the experiment were also included in the bar graphs.

2.2. Cut-off definition

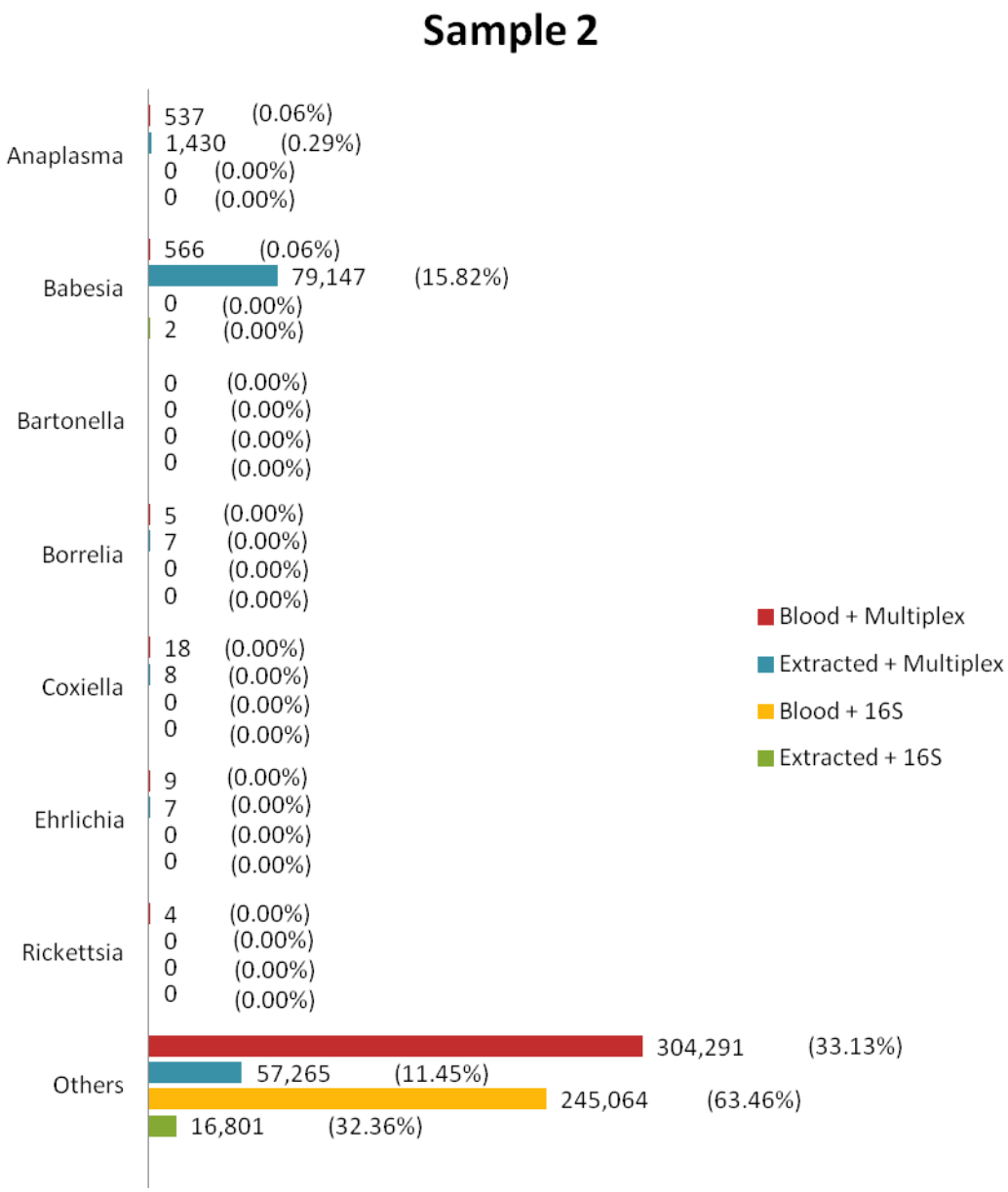
In the Next Generation Sequencing, where a large number of libraries can be pooled and sequenced simultaneously during a single sequencing run, index hopping can occur(75). Basically, index hopping is when reads originated from one sample are assigned to another sample when demultiplexing is done. That being said, not all reads shown in the bar charts can be assumed to be present in the sample. Based in comparison of the results from all samples a cut-off was defined. All targets that present a percentage lower than 0.35% in total reads, were not considered as being present in the sample but rather as errors in sequencing process.

2.3. Bar graphs



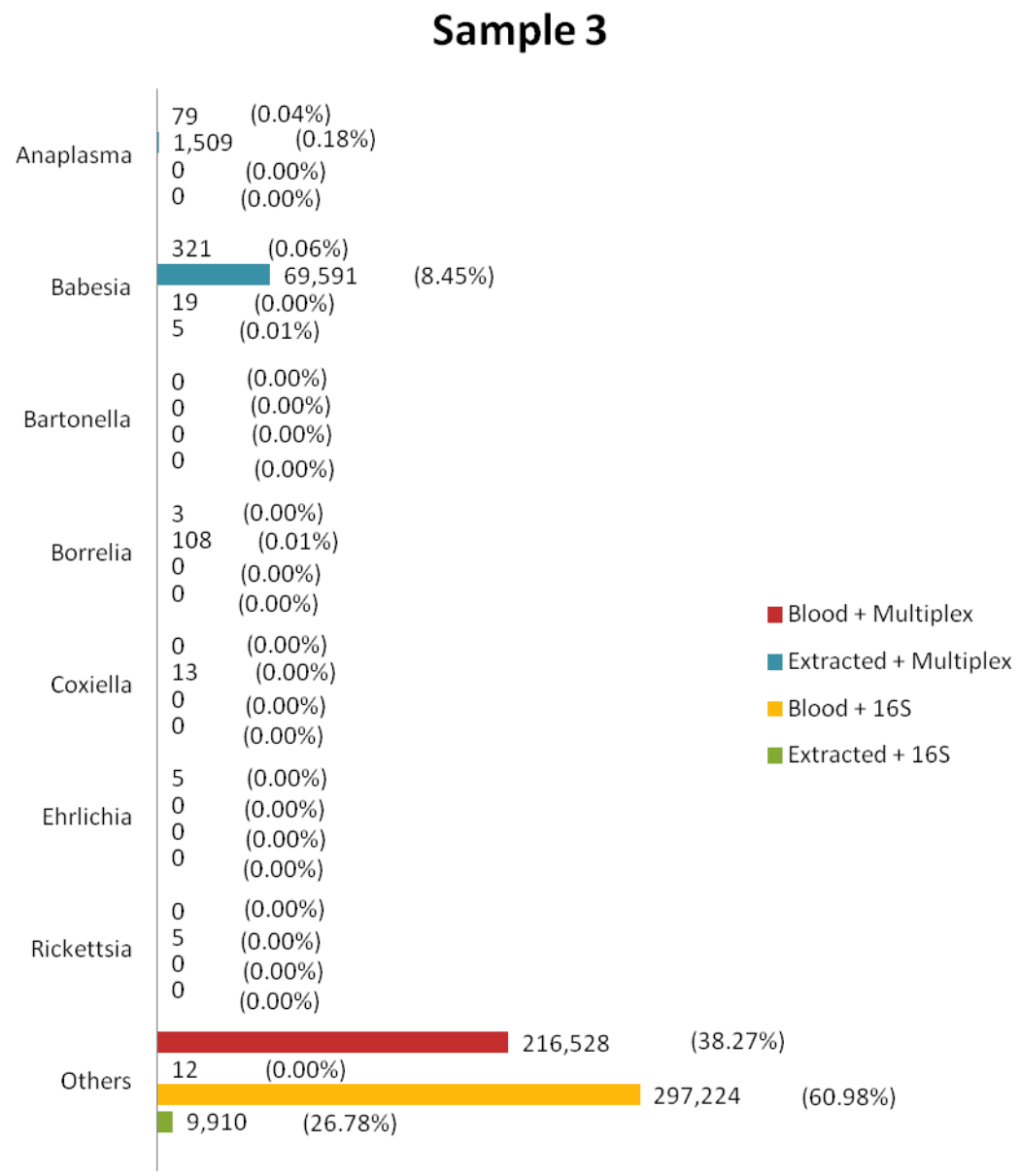
Graph 1- Number of reads for each species targeted and respective percentage in total number of raw reads from sample 1, obtained from the analysis of the four different conditions used.

In sample 1, *Babesia microti* DNA was detected with all four conditions.



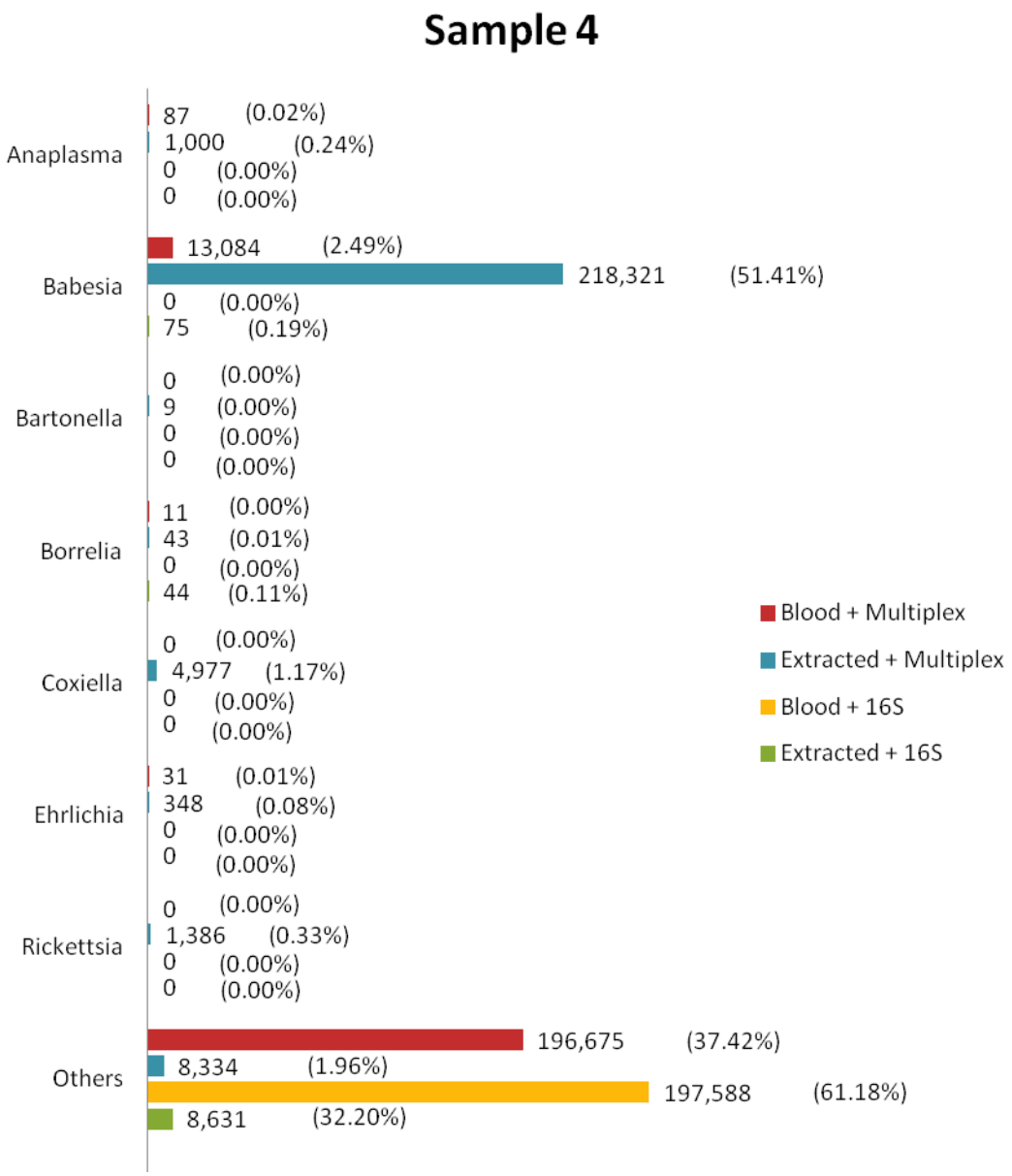
Graph 2- Number of reads for each species targeted and respective percentage in total number of raw reads from sample 2, obtained from the analysis of the four different conditions used.

In sample 2, *Babesia microti* DNA was only detected through the condition using extracted DNA and the primers panel.



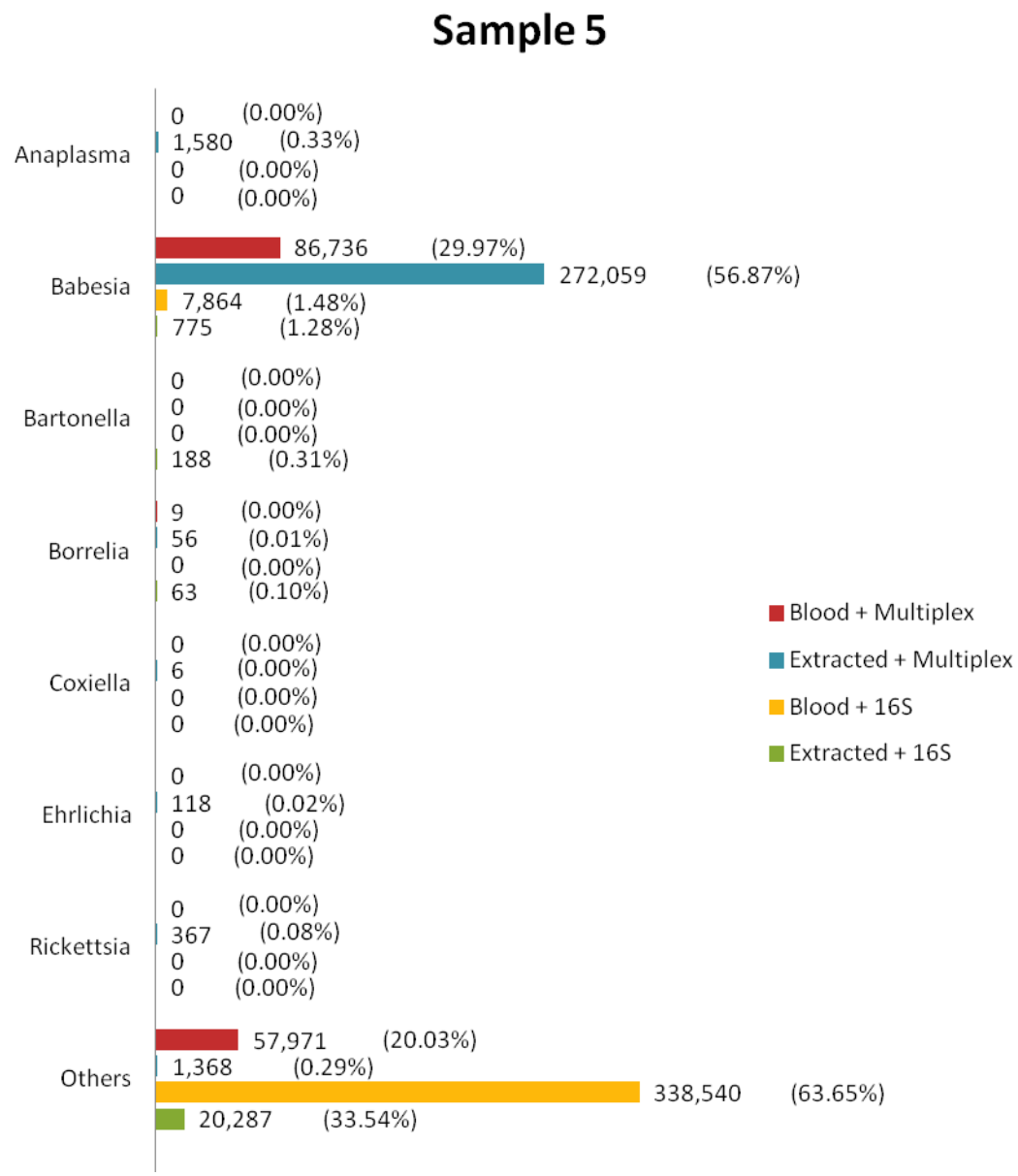
Graph 3- Number of reads for each species targeted and respective percentage in total number of raw reads from sample 3, obtained from the analysis of the four different conditions used.

In sample 3, like sample 2, *Babesia microti* DNA was only detected with the condition using extracted DNA and the multiplex PCR (DNA+Mx).



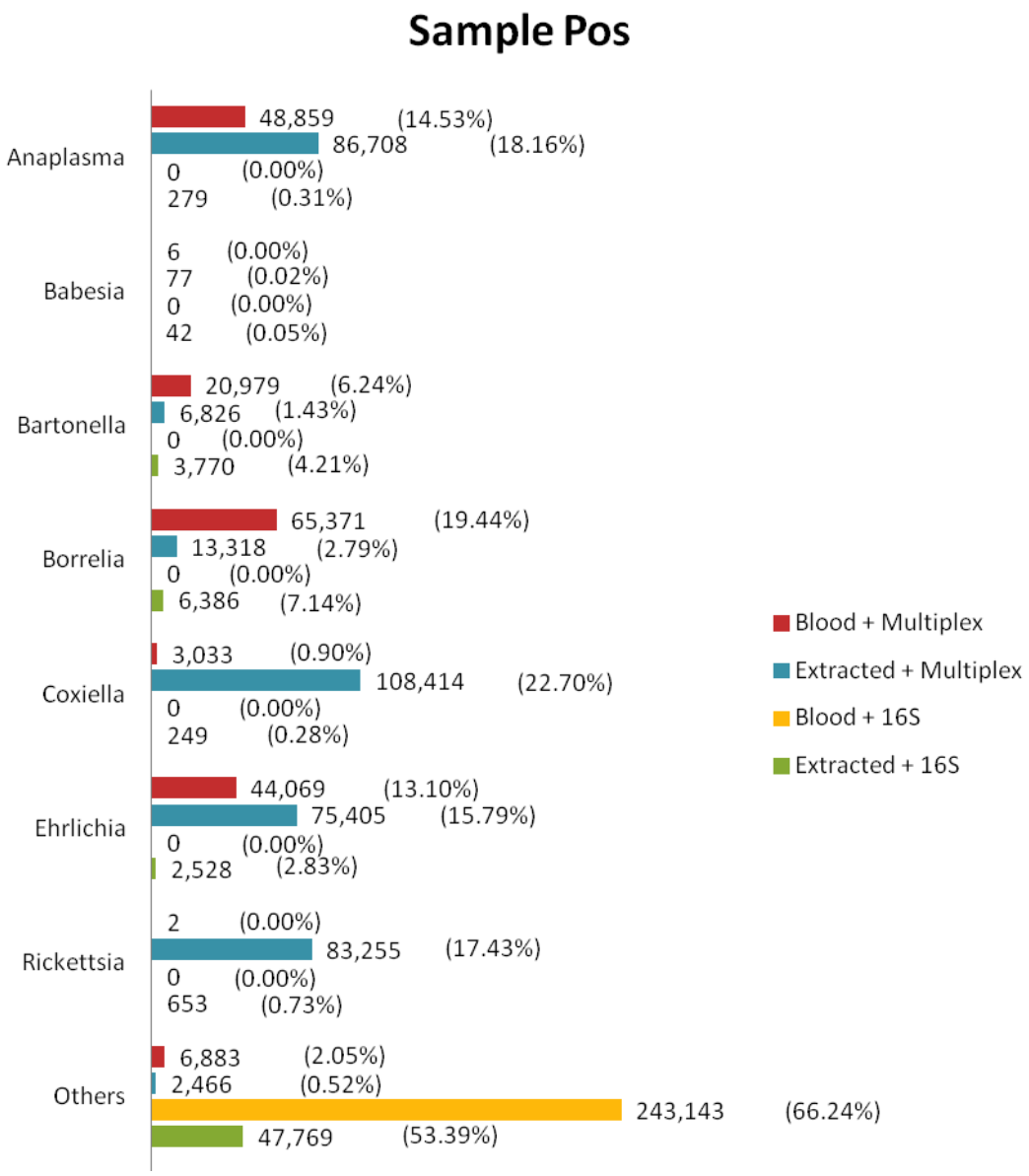
Graph 4- Number of reads for each species targeted and respective percentage in total number of raw reads from sample 4, obtained from the analysis of the four different conditions used.

For sample 4, both conditions using the multiplex PCR (B+Mx and DNA+Mx) were able to detect *Babesia microti* DNA. Furthermore, in the condition using extracted DNA (DNA+Mx), a co-infection with *Coxiella burnetii* was detected.



Graph 5- Number of reads for each species targeted and respective percentage in total number of raw reads from sample 5, obtained from the analysis of the four different conditions used.

In sample 5, *Babesia microtii* DNA was detected trough all four conditions.

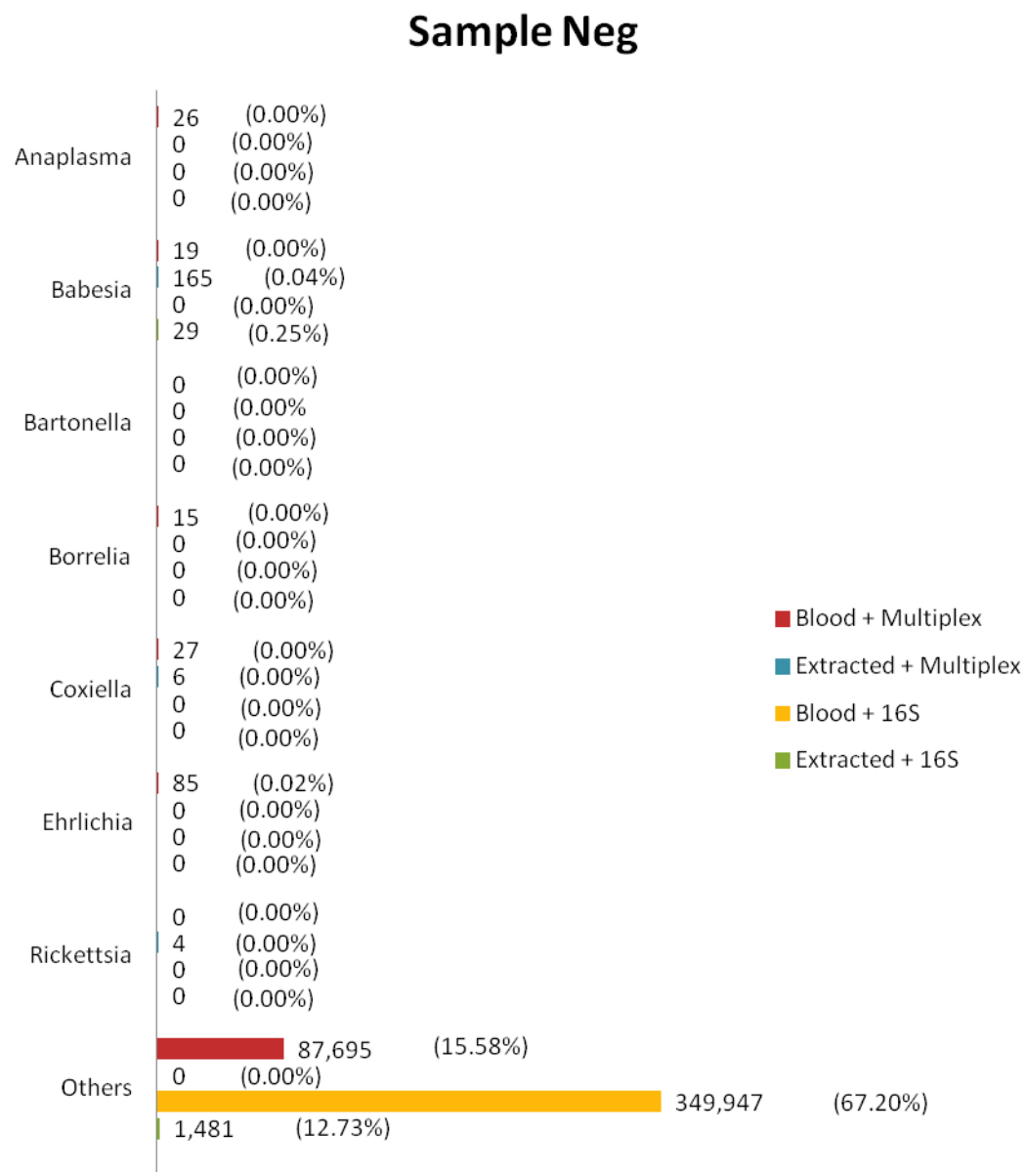


Graph 6- Number of reads for each species targeted and respective percentage in total number of raw reads from the positive sample, obtained from the analysis of the four different conditions used.

In the positive sample, as expected, *Babesia microtii* was never detected due to unsuccessful insertion of the 18S rRNA gene in *E.coli* plasmid, previously confirmed by Sanger sequencing. In the two conditions using the primers panel, all other target genes were detected, with exception of *Rickettsia rickettsii* gltA gene using whole blood (B+Mx). This result can not suggest inefficiency of *Rickettsia* primers in the panel developed since detection in the condition using extracted DNA was verified.

In the conditions using 16S primers, none of the species should be detected, since a fragment of each target genes was used, not genomic DNA, in sample spiking. However, 4 of the 7 species were detected in the DNA+16S condition. This result can only be explained by some

contamination during library preparation. The high percentage of other species, as it could be anticipated, was due to amplification of 16S rRNA gene from E.coli cells used in spike.



Graph 7- Number of reads for each species targeted and respective percentage in total number of raw reads from the negative sample, obtained from the analysis of the four different conditions used.

As expected in the negative sample, none of the species tested were detected. The few number of reads that were observed are due to index-hopping.

V. DISCUSSION AND CONCLUSION

1. Comparison of the four different conditions

1.1. General overview

With the results of PCR amplification observed in the electrophoresis gel, the pathogen burden in Samples 1 and 5 seems to be higher than the in other samples. For this two samples, all conditions used in this study were able to detect the pathogen's DNA. However, in samples showing less *Babesia* DNA, accordingly to the intensity of the bands verified in the electrophoresis gel, both conditions using whole blood (B+MX and B+16S) and the DNA+16S were inapt to correctly detect the presence of *Babesia microti* DNA. Therefore, from the conditions studied, PCR with DNA extracted from the blood and with the primers panel designed, was the condition that led to a lower limit of detection.

The two conditions using whole blood in the first PCR of library preparation were performed with the intention of cutting one step in the process, the DNA extraction, amplifying the pathogens' sequences directly from whole blood. With this method not only the process would be "hands-on" less time consuming but also the possibility of losing pathogen's DNA during the extraction process would be excluded. The results were the expected.

The use of 16S metagenomic sequencing library preparation was performed with the intention of providing the complete microbiome of the sample. This approach could be particularly useful for patients from endemic areas for Lyme disease who don't recall a tick bite and didn't develop any skin lesion, since it could provide information of the presence of others pathogenic species not included in the designed panel. However, in patients with diseases with a low burden of bacteria, such as Lyme borreliosis, this method is more prone to obtain false negative results than the one using the targeted primers.

1.2. Whole blood used in multiplex PCR (B+Mx)

Using this condition three of the five samples of babesiosis patients were positive (samples 1, 4 and 5). The results obtained with this condition, compared with the results obtained in the other condition using the primers panel, suggest the existence of components in the whole blood that seems to inhibit the amplification of the desired product in the PCR.

1.3. Extracted DNA used in multiplex PCR (DNA+Mx)

For the four conditions used in this study, the condition using extracted DNA and multiplex PCR with the target primers, was the one that showed better results. With this condition, *Babesia microtii* DNA was detected in all five samples from patients with human babesiosis. Although in positive control no *Babesia* DNA was detected, the observed results were due to the incorrect insertion of 18S rRNA gene fragment in *E.coli* cells used in spiking, as previously described.

1.4. Whole blood used in 16S PCR (B+16S)

With this condition, detection of DNA from *Babesia* was only obtained in samples 1 and 5. This test, along with the other condition using whole blood, seem to be ineffective for the diagnosis of Lyme disease.

1.5. Extracted DNA used in 16S PCR (DNA+16S)

In this condition, as well as in the B+16S, only samples 1 and 5 have given a positive result in *Babesia* DNA detection. However, the number of reads obtained in the sequencing process for this condition was lower than the expected and the negative results may be due to masking from species present in more number in the sample.

2. Unspecific amplification in multiplex PCR

In both conditions using multiplex PCR with the primers panel for amplification of *Babesia* spp. and possible co-infections, unspecific amplification has been found. The BLAST performed showed the presence of many reads attributed to artificial human sequences. This was not expected due to the specificity of the primers of the panel designed to target the selected regions of the previously referred genes. In order to understand the specificity of the primers, randomly selected reads correspondent to different regions of the human genome that were obtained in the sequencing results, were aligned to the 14 primers designed. No conclusive results can justify the amplification of these sequences. However, since the primers had an adaptor sequence needed for the ligation of the amplified sequences to the flow cell in NGS, a new alignment was performed using the complete sequence of the primers. With this alignment, a complementary region was found between the adaptor region of the primer and the beginning of the sequence, explaining our results.

The results shown that the unspecific amplification occurred mainly in the samples submitted to whole blood PCR. The great variation of unspecific amplification between this methods may be explained by the difference of the annealing temperature used in PCR programs. The PCR performed with whole blood samples had a much lower annealing temperature which may allow non-specific binding to occur.

To avoid this problem, an alternative of two separate PCRs may be used. The first PCR would only use the specific primers without the adaptor needed for NGS. As a result, with this first PCR only the fragments targeted should be amplified. Then, the PCR products would be submitted to a second PCR using primers with the adaptor making the amplicons suitable for further NGS process.

3.Conclusion

The intention of this study was to detect the presence of DNA from the pathogenic species covered by the panel, through Next Generation Sequencing, as a method to diagnose Lyme disease and the most common co-infections reported. Until the starting day of this project, that I know, no information regarding this diagnostic approach has been reported. However, in April of the present year, the U.S. National Institute of Health (NIH), posted a clinical trial with the title "Next Generation Sequencing detection of Lyme disease" that intends to be concluded in March of 2020. This highlights the relevance of this study.

Unfortunately, the only samples obtained, generously provided by CDC, were from babesiosis patients, which are known to have higher pathogen burden. Although NGS allows a low limit of detection, being Lyme borreliosis a disease known to have few spirochete burden, it is important to access the correct body site, specimen and stage of the disease when collecting the sample for testing. Despite not having samples from patients with Lyme borreliosis, the present study has shown great importance in the comprehension of this approach for diagnosis of the tick-borne diseases included. It is also important to bear in mind that a positive result using this method does not necessarily prove active infection once DNA of dead microorganisms may persist in patients' body for some time. Thus, this approach is not recommended soon after the antibiotic treatment to evaluate if it was successful.

The condition that used DNA extracted from the blood sample and the primers panel designed was the one that showed the best results, being able to detect *Babesia* DNA in all the samples from patients with babesiosis.

To conclude, this study has shown promising results in detection of DNA from species responsible for tick-borne diseases through Next Generation Sequencing, with the ambition of filling the gap of misdiagnosis in early Lyme disease.

VI. FUTURE PERSPECTIVES

From the four conditions used in this work, the ones that used direct whole blood in library preparation appear not to be suitable for detection of the pathogens covered by the panel. As for the condition using extracted DNA with the 16S metagenomic library preparation, a new test will be performed to evaluate if the lack of results is due to low coverage obtained in this sequencing run. The approach using extracted DNA and the panel of primers designed in the present dissertation was able to detect the presence of DNA of *Babesia* in the five samples from patients with babesiosis, which indicates that this may be the condition to carry forward.

Despite being an excellent kick-off to the comprehension of the method, our results do not prove yet that this approach will be suitable to detect *Borrelia* spp. in samples from patients with Lyme borreliosis. For further testing, and to validate the designed panel, samples from Lyme disease patients, especially from early infection stages, and from patients infected with the co-infections covered by the panel, will be tested.

VII. REFERENCES

1. Steere AC, Malawista SE, Hardin JA, Ruddy S, Askenase PW, Andiman WA. Erythema chronicum migrans and Lyme arthritis: the enlarging clinical spectrum. Annals of internal medicine. 1977;86(6):685-98.

2. Steere AC, Malawista SE, Snydman DR, Shope RE, Andiman WA, Ross MR, et al. An epidemic of oligoarticular arthritis in children and adults in three connecticut communities. Arthritis & Rheumatism. 1977;20(1):7-17.

3. Steere AC, Hardin JA, Malawista SE. Erythema chronicum migrans and Lyme arthritis: cryoimmunoglobulins and clinical activity of skin and joints. Science. 1977;196(4294):1121-2.

4. Steere AC, Malawista SE, Hardin JA, Ruddy S, Askenase W, Andiman WA. Erythema chronicum migrans and Lyme arthritis. The enlarging clinical spectrum. Ann Intern Med. 1977;86(6):685-98.

5. Burgdorfer W, Barbour A, Hayes S, Benach J, Grunwaldt E, Davis J. Lyme disease-a tick-borne spirochetosis? Science. 1982;216(4552):1317-9.

6. Steere AC, Grodzicki RL, Kornblatt AN, Craft JE, Barbour AG, Burgdorfer W, et al. The spirochetal etiology of Lyme disease. New England Journal of Medicine. 1983;308(13):733-40.

7. Benach JL, Bosler EM, Hanrahan JP, Coleman JL, Habicht GS, Bast TF, et al. Spirochetes isolated from the blood of two patients with Lyme disease. New England Journal of Medicine. 1983;308(13):740-2.

8. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, et al. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature. 1997;390:580.

9. Casjens S, Palmer N, Van Vugt R, Mun Huang W, Stevenson B, Rosa P, et al. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi. Molecular microbiology. 2000;35(3):490-516.

10. Steere AC, Coburn J, Glickstein L. The emergence of Lyme disease. The Journal of clinical investigation. 2004;113(8):1093-101.

11. Afzelius A. Erythema chronicum migrans. Acta Derm Venereol. 1921;2:120-5.

12. Steere AC, Broderick TF, Malawista SE. Erythema chronicum migrans and Lyme arthritis: epidemiologic evidence for a tick vector. American journal of epidemiology. 1978;108(4):312-21.

13. Xu G, Fang QQ, Keirans JE, Durden LA. Molecular phylogenetic analyses indicate that the Ixodes ricinus complex is a paraphyletic group. Journal of Parasitology. 2003;89(3):452-7.

14. Ribeiro JC. Ixodes dammini: salivary anti-complement activity. Experimental parasitology. 1987;64(3):347-53.

15. Schwartz AM, Hinckley AF, Mead PS, Hook SA, Kugeler KJ. Surveillance for Lyme Disease—United States, 2008–2015. MMWR Surveillance Summaries. 2017;66(22):1.

16. Mead PS. Epidemiology of Lyme disease. Infectious Disease Clinics. 2015;29(2):187-210.

17. Steere AC, Strle F, Wormser GP, Hu LT, Branda JA, Hovius JW, et al. Lyme borreliosis. Nature Reviews Disease Primers. 2016;2:16090.
18. Stanek G, Wormser GP, Gray J, Strle F. Lyme borreliosis. The Lancet. 2012;379(9814):461-73.
19. Rollend L, Fish D, Childs JE. Transovarial transmission of Borrelia spirochetes by Ixodes scapularis: A summary of the literature and recent observations. Ticks and Tick-borne Diseases. 2013;4(1):46-51.
20. Lane R, Piesman J, Burgdorfer W. Lyme borreliosis: relation of its causative agent to its vectors and hosts in North America and Europe. Annual review of entomology. 1991;36(1):587-609.
21. Cotté V, Bonnet S, Le Rhun D, Le Naour E, Chauvin A, Boulouis H-J, et al. Transmission of Bartonella henselae by Ixodes ricinus. Emerging infectious diseases. 2008;14(7):1074.
22. Horowitz HW, Aguero-Rosenfeld ME, Holmgren D, McKenna D, Schwartz I, Cox ME, et al. Lyme disease and human granulocytic anaplasmosis coinfection: impact of case definition on coinfection rates and illness severity. Clinical Infectious Diseases. 2012;56(1):93-9.
23. Diuk-Wasser MA, Vannier E, Krause PJ. Coinfection by Ixodes tick-borne pathogens: ecological, epidemiological, and clinical consequences. Trends in parasitology. 2016;32(1):30-42.
24. Medlock JM, Hansford KM, Bormane A, Derdakova M, Estrada-Peña A, George J-C, et al. Driving forces for changes in geographical distribution of Ixodes ricinus ticks in Europe. Parasites & vectors. 2013;6(1):1.
25. Wielinga PR, Gaasenbeek C, Fonville M, de Boer A, de Vries A, Dimmers W, et al. Longitudinal analysis of tick densities and Borrelia, Anaplasma, and Ehrlichia infections of Ixodes ricinus ticks in different habitat areas in The Netherlands. Applied and environmental microbiology. 2006;72(12):7594-601.
26. Sanogo YO, Zeaiter Z, Caruso G, Merola F, Shpynov S, Brouqui P, et al. Bartonella henselae in Ixodes ricinus ticks (Acari: Ixodida) removed from humans, Belluno province, Italy. Emerging infectious diseases. 2003;9(3):329.
27. Müller A, Reiter M, Schötta AM, Stockinger H, Stanek G. Detection of Bartonella spp. in Ixodes ricinus ticks and Bartonella seroprevalence in human populations. Ticks and Tick-borne Diseases. 2016;7(5):763-7.
28. Hildebrandt A, Straube E, Neubauer H, Schmoock G. Coxiella burnetii and Coinfections in Ixodes ricinus Ticks in Central Germany. Vector-Borne and Zoonotic Diseases. 2011;11(8):1205-7.
29. Sykes RA, Makiello P. An estimate of Lyme borreliosis incidence in Western Europe. Journal of Public Health. 2016;39(1):74-81.
30. Lipschütz B. Über eine seltene Erythemform (Erythema chronicum migrans). Archiv für Dermatologie und Syphilis. 1913;118(1):349-56.

31. Tibbles CD, Edlow JA. Does this patient have erythema migrans? Jama. 2007;297(23):2617-27.

32. Steere AC, Bartenhagen NH, Craft JE, Hutchinson GJ, Newman JH, Rahn DW, et al. The early clinical manifestations of Lyme disease. Annals of internal medicine. 1983;99(1):76-82.

33. Steere AC. Lyme Disease. New England Journal of Medicine. 1989;321(9):586-96.

34. Goddard J. Not All Erythema Migrans Lesions Are Lyme Disease. The American Journal of Medicine. 2017;130(2):231-3.

35. Stanek G, Fingerle V, Hunfeld KP, Jaulhac B, Kaiser R, Krause A, et al. Lyme borreliosis: Clinical case definitions for diagnosis and management in Europe. Clinical Microbiology and Infection. 2011;17(1):69-79.

36. Strle F, Nadelman RB, Cimperman J, Nowakowski J, Picken RN, Schwartz I, et al. Comparison of culture-confirmed erythema migrans caused by Borrelia burgdorferi sensu stricto in New York State and by Borrelia afzelii in Slovenia. Annals of internal medicine. 1999;130(1):32-6.

37. Strle F, Ružić-Sabljić E, Logar M, Maraspin V, Lotrič-Furlan S, Cimperman J, et al. Comparison of erythema migrans caused by Borrelia burgdorferi and Borrelia garinii. Vector-Borne and Zoonotic Diseases. 2011;11(9):1253-8.

38. MÜLLEGGER RR. Dermatological manifestations of Lyme borreliosis. European Journal of Dermatology. 2004;14(5):296-309.

39. Berger BW, Kaplan MH, Rothenberg IR, Barbour AG. Isolation and characterization of the Lyme disease spirochete from the skin of patients with erythema chronicum migrans. Journal of the American Academy of Dermatology. 1985;13(3):444-9.

40. Sigal LH, Steere AC, Freeman DH, Dwyer JM. Proliferative responses of mononuclear cells in Lyme disease. Reactivity to Borrelia burgdorferi antigens is greater in joint fluid than in blood. Arthritis & Rheumatism. 1986;29(6):761-9.

41. Shrestha M, Grodzicki RL, Steere AC. Diagnosing early lyme disease. The American Journal of Medicine. 1985;78(2):235-40.

42. Auwaerter PG, Aucott J, Dumler JS. Lyme borreliosis (Lyme disease): molecular and cellular pathobiology and prospects for prevention, diagnosis and treatment. Expert reviews in molecular medicine. 2004;6(2):1-22.

43. Pachner AR, Steere AC. The triad of neurologic manifestations of Lyme disease Meningitis, cranial neuritis, and radiculoneuritis. Neurology. 1985;35(1):47-.

44. McAlister HF, Klementowicz PT, Andrews C, Fisher JD, Feld M, Furman S. Lyme carditis: an important cause of reversible heart block. Annals of internal medicine. 1989;110(5):339-45.

45. Wormser GP, Dattwyler RJ, Shapiro ED, Halperin JJ, Steere AC, Klempner MS, et al. The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. Clinical Infectious Diseases. 2006;43(9):1089-134.

46. Strle F, Ružić-Sabljić E, Cimperman J, Lotrič-Furlan S, Maraspin V. Comparison of findings for patients with Borrelia garinii and Borrelia afzelii isolated from cerebrospinal fluid. Clinical infectious diseases. 2006;43(6):704-10.
47. Ogrinc K, Lusa L, Lotrič-Furlan S, Bogovič P, Stupica D, Cerar T, et al. Course and Outcome of Early European Lyme Neuroborreliosis (Bannwarth Syndrome): Clinical and Laboratory Findings. Clinical Infectious Diseases. 2016;63(3):346-53.
48. Carlson D, Hernandez J, Bloom BJ, Coburn J, Aversa JM, Steere AC. Lack of Borrelia burgdorferi DNA in synovial samples from patients with antibiotic treatment–resistant Lyme arthritis. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 1999;42(12):2705-9.
49. Steere AC, Angelis SM. Therapy for Lyme arthritis: strategies for the treatment of antibiotic-refractory arthritis. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 2006;54(10):3079-86.
50. Asbrink E. Cutaneous manifestations of Lyme borreliosis. Scand J Infect Dis Suppl. 1991;77:44-50.
51. Åsbrink E, Hovmark A. Successful cultivation of spirochetes from skin lesions of patients with erythema chronicum migrans Afzelius and acrodermatitis chronica atrophicans. Acta Pathologica Microbiologica Scandinavica Series B: Microbiology. 1985;93(1-6):161-3.
52. Oschmann P, Dorndorf W, Hornig C, Schäfer C, Wellensiek H, Pflughaupt K. Stages and syndromes of neuroborreliosis. Journal of neurology. 1998;245(5):262-72.
53. Logigian EL, Steere AC. Clinical and electrophysiologic findings in chronic neuropathy of Lyme disease. Neurology. 1992;42(2):303-.
54. Steere AC. Lyme Disease. New England Journal of Medicine. 2001;345(2):115-25.
55. Marques AR. Laboratory diagnosis of Lyme disease: advances and challenges. Infectious Disease Clinics. 2015;29(2):295-307.
56. Wormser GP, Nowakowski J, Nadelman RB, Bittker S, Cooper D, Pavia C. Improving the yield of blood cultures for patients with early Lyme disease. Journal of clinical microbiology. 1998;36(1):296-8.
57. Liveris D, Wang G, Girao G, Byrne DW, Nowakowski J, McKenna D, et al. Quantitative detection of Borrelia burgdorferi in 2-millimeter skin samples of erythema migrans lesions: correlation of results with clinical and laboratory findings. Journal of clinical microbiology. 2002;40(4):1249-53.
58. Strle F, Lusa L, Ružić-Sabljić E, Maraspin V, Furlan SL, Cimperman J, et al. Clinical characteristics associated with Borrelia burgdorferi sensu lato skin culture results in patients with erythema migrans. PloS one. 2013;8(12):e82132.
59. Aguero-Rosenfeld ME, Wang G, Schwartz I, Wormser GP. Diagnosis of Lyme borreliosis. Clinical microbiology reviews. 2005;18(3):484-509.
60. Li X, McHugh GA, Damle N, Sikand VK, Glickstein L, Steere AC. Burden and viability of Borrelia burgdorferi in skin and joints of patients with erythema migrans or Lyme arthritis. Arthritis & Rheumatism. 2011;63(8):2238-47.

61. Rosa PA, Schwan TG. A specific and sensitive assay for the Lyme disease spirochete Borrelia burgdorferi using the polymerase chain reaction. Journal of infectious diseases. 1989;160(6):1018-29.

62. Wang G. Direct detection methods for Lyme Borrelia, including the use of quantitative assays. Vector-Borne and Zoonotic Diseases. 2002;2(4):223-31.

63. Eshoo MW, Crowder CC, Rebman AW, Rounds MA, Matthews HE, Picuri JM, et al. Direct molecular detection and genotyping of Borrelia burgdorferi from whole blood of patients with early Lyme disease. PLoS One. 2012;7(5):e36825.

64. Liveris D, Schwartz I, McKenna D, Nowakowski J, Nadelman R, DeMarco J, et al. Quantitation of cell-associated borrelial DNA in the blood of Lyme disease patients with erythema migrans. European journal of clinical microbiology & infectious diseases. 2012;31(5):791-5.

65. Waddell LA, Greig J, Mascarenhas M, Harding S, Lindsay R, Ogden N. The accuracy of diagnostic tests for Lyme disease in humans, a systematic review and meta-analysis of North American research. PloS one. 2016;11(12):e0168613.

66. Control CfD, Prevention. Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. MMWR Morbidity and mortality weekly report. 1995;44(31):590.

67. Magnarelli LA, Anderson JF, Johnson RC. Cross-Reactivity in Serological Tests for Lyme Disease and Other Spirochetal Infections. The Journal of Infectious Diseases. 1987;156(1):183-8.

68. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931.

69. Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. The Journal of investigative dermatology. 2013;133(8):e11.

70. Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends in genetics. 2014;30(9):418-26.

71. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. nature. 2008;456(7218):53.

72. Shendure J, Ji H. Next-generation DNA sequencing. Nature biotechnology. 2008;26(10):1135.

73. Buermans HPJ, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2014;1842(10):1932-41.

74. Amplicon P, Clean-Up P, Index P. 16S Metagenomic Sequencing Library Preparation. 2013.

75. Inc I. Effects of index misassignment on multiplexing and downstream analysis. 2017.