Ana Cristina Rosado Pelerito Araújo Gonçalves

Licenciada em Engª Biotecnológica

Mestre em Microbiologia Clínica

Brucellosis: an emerging disease in Portugal

Dissertação para obtenção do Grau de Doutor em Biologia

Orientadora: Doutora Maria Sofia Cobra Lince Núncio Soares, Investigadora Auxiliar, Instituto Nacional Saúde Doutor Ricardo Jorge

Co-orientadora: Doutora Suzana Salcedo
Laboratory of Molecular Microbiology and Structural Biochemistry, University of Lyon

Co-orientadora: Doutora Isabel Maria Godinho de Sá Nogueira, Professora Associada com Agregação, Faculdade de Ciências e Tecnologia, Universidade nova de Lisboa

Júri:
Presidente: Paulo Manuel Assis Loureiro Limão-Vieira
Arguentes: Doutor Fernando Sivano Boinas
Doutora Isabel Maria dos Santos Leitão Couto
Vogais: Doutora Maria Manuela Spratley de Lemos Carolino
Doutora Maria Sofia Cobra Lince Núncio Soares

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março 2019

Ana Cristina Rosado Pelerito Araújo Gonçalves

Licenciada em Engª Biotecnológica

Mestre em Microbiologia Clínica

Brucelose: Uma infeção emergente em Portugal

Dissertação para obtenção do Grau de Doutor em Biologia

Para ti minha mãe,
a estrelinha mais brilhante...

Agradecimentos

Concluir um doutoramento implica uma jornada longa e difícil, tornando-se impossível levar a bom porto sem o apoio de uma rede de pessoas a quem quero deixar aqui os meus sinceros agradecimentos:

À minha orientadora Doutora Sofia Núncio, Responsável da Unidade de Resposta a Emergências e Biopreparação do Departamento de Doenças Infecciosas do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA), pelos ensinamentos científicos que me transmitiu ao longo destes anos, pelo apoio que sempre me deu, pela disponibilidade, pela ajuda, e amizade.

À minha co-orientadora Doutora Suzana Salcedo, que apesar de longe esteve sempre disponível para me responder na hora a qualquer email e pronta para me dar a conhecer um mundo, para mim, novo na ciência.

A todos os restantes membros da Comissão de Acompanhamento de Tese (CAT), por todas as sugestões concedidas. Em especial, à minha co-orientadora professora Ilda Sanches (que me acompanhou no inicio deste doutoramento), à professora Isabel Sá Nogueira, que passou a ser a minha co-orientadora, pela sua disponibilidade em responder a todas as minhas dúvidas. Agradeço ainda ao Doutor Paulo Nogueira.

Ao João Paulo, pelo conhecimento cientifico que me transmitiu ao longo destes anos, pela ajuda e ideias que me foi dando enquanto o trabalho ia avançando, pela sua disponibilidade, paciência e pelo carinho demonstrado.

A todos os colegas que colaboraram nos trabalhos que constam nesta tese, com os quais tive muito gosto em trabalhar e em partilhar a co-autoria dos mesmos.

A todos os outros colegas do INSA, que de alguma maneira nem que seja com um sorriso me apoiaram.

À Maria João Simões, Rita Matos e Mónica Oleastro pela amizade demonstrada.

Às minhas colegas e amigas Rita Cordeiro e Isabel Carvalho, pelo apoio, paciência e amizade.

À minha família, nomeadamente à minha irmãe, cunhado, pai e boadrasta, pelo apoio sempre prestado, pela força para continuar, por serem o meu Porto de Abrigo.

Aos meus amigos, obrigada por estarem sempre aqui.

E por último, mas com a maior importância, aos meus queridos filhos Lourenço e Constança, obrigada pela vossa compreensão quando a paciência me faltava, principalmente nestes últimos meses. Obrigada meus Amores.

Resumo

A brucelose humana é uma das zoonoses com maior incidência a nível mundial. Em Portugal, a brucelose é uma doença de notificação obrigatória, cuja casuística a coloca entre as três zoonoses com maior incidência. Apesar da sua importância, estudos da prevalência de brucelose e a realidade eco-epidemiológica desta doença em Portugal são escassos.

Com o presente estudo pretende-se fazer uma avaliação da situação epidemiológica da brucelose humana em Portugal e identificar quais as espécies associadas a casos humanos. Pretende-se também investigar qual a origem da infeção no Homem, utilizando estudos de tipagem molecular e novas metodologias de sequenciação de nova geração. Neste trabalho pretende-se ainda realizar estudos para avaliar o polimorfismo genético de vários fatores de virulência em estirpes de *Brucella* spp.

Os nossos resultados comprovam que a única espécie de *Brucella* associada a casos clínicos em Portugal é *Brucella melitensis* e que existe uma forte ligação epidemiológica entre vários casos estudados, cujo a análise originou pequenos clusters, podendo mesmo corresponder a pequenos surtos. Verificou-se também uma forte proximidade filogenética entre as estirpes isoladas na zona mediterrânica (Espanha, Grécia e Itália), devido à proximidade geográfica, cultural e alimentar. Neste trabalho, e utilizando uma abordagem inovadora, recorrendo às novas metodologias de sequenciação de nova geração, conseguimos validar a utilização MLVA – 16, o *gold standard* para a tipagem de *Brucella* spp., utilizando a extração dos genótipos *in silico*.

Globalmente, os resultados apresentados nesta tese contribuem para a melhor compreensão da situação da brucelose em Portugal. Os resultados poderão contribuir para a implementação de uma nova ferramenta laboratorial que permita melhorar a vigilância epidemiológica de brucelose, disponibilizando informação mais precisa e rápida aos decisores com responsabilidades na área da implementação de medidas de prevenção e controlo desta doença no nosso País, tanto na vertente de saúde humana como veterinária, em sintonia com a abordagem *One Health*.

Palavras chave: *Brucella* spp., *B. melitensis*, brucelose, sequenciação nova geração, tipagem molecular, epidemiologia, filogenia e virulência.

Abstract

Human brucellosis is one of the most common zoonosis worldwide. In Portugal, brucellosis is a notifiable disease in humans and the casuistic puts it among the three zoonosis with the highest incidence. Despite this, studies on prevalence of brucellosis in Portugal are scarce.

The present study intends to evaluate the epidemiological situation of human brucellosis in Portugal and to identify the species associated with human cases. It also intend to investigate the origin of infection in humans using molecular typing studies and whole genome sequencing approaches. In this work, we propose study the genetic polymorphism of several virulence factors in *Brucella* spp.

Our results showed that *Brucella melitensis* is the main species associated to human brucellosis and that there is a strong epidemiological link between many cases studies, whose originated small clusters and may even correspond to small outbreaks. In this study, it was possible to verify a strong phylogenetic proximity between isolated strains in the Mediterranean area (Spain, Greece and Italy) probably due to geographical, cultural and type of food proximity. In this work, using an advanced approach, new generation sequencing methodologies, we were able to validate the use of MLVA - 16, the gold standard for typing *Brucella* spp., using *in silico* extraction.

Globally, the findings presented in this PhD thesis contribute for better understanding of the brucellosis situation in Portugal. The results may contribute to the implementation of a new laboratory tool to improve the epidemiological surveillance of brucellosis. Furthermore, is providing more accurate and quick information to the decision makers with responsibilities in the area of the implementation of measures of prevention and control of this disease in our Country, both in human and veterinary health, in line with the One Health approach.

Key-words: *Brucella* spp., *B. melitensis*, brucellosis, next generation sequencing, molecular typing, epidemiology, phylogeny and virulence.

Table of Contents

Figure index

Table index

List of Abbreviations

ALM/ASM	Allele larger/Smaller than length mode
AMOS	Abortus, Melitensis, Ovis, Suis
BCV	*Brucella* – containing vacuole
Brucellacapt	Immunocapture-agglutination test
CDC	Centers for Disease Control and Prevention
cgMLST	core genome Multilocus sequence typing
DNA	Desoxyrribonucleic acid
DGS	General Directorate of Health
ELISA	Enzyme immunoassays
IFI	Indirect immunofluorescence
LPS	Lipopolysaccharide
MIC	Minimal inhibitory concentration
MLST	Multilocus sequence typing
MLVA	Multi-locus VNTR Analysis
NIH	Reference Laboratory at the Portuguese National Institute of Health
NIPH/NIPHEM	non-informative paralogous hit
NLRs	NOD–like receptors
OMPs	Outer membrane proteins
PAMPs	Pathogen–associated molecular patterns
PCR	Polymerase chain reaction
PL	Phospholipids
QS	Quorum sensing
RNA	Ribonucleic acid
SAT	Serum agglutination test
SNP	Single nucleotide polymorphims
TLRs	Toll-like receptors
VNTR	Variable Number Tandem Repeats
WHO	World health Organization
WGS	Whole Genome Sequencing

Notes of the author: thesis organization, format and outline

This thesis is organized into six chapters (listed below) that are presented as individual chapters (I to VI). In Chapter I, a general introduction to the subject of the thesis is presented, including the history and status of knowledge and the objectives of the work developed in this thesis. Chapters II to V describe research studies that were performed to achieve the delineated goals. The last chapter, summarizes all major contributions from this PhD project and suggest directions for future research. One of the chapters have already been published (the other two have been submitted for publication at the time this thesis was completed) in peer reviewed international journals, being presented in this thesis essentially as a reproduction of the content that was published.

The chapters were organized according to a rational taking into account the objectives delineated for this PhD work, and in agreement with the association between the scientific subjects addressed in each one, as the results obtained during one study influenced the progress of others. Considering the different article types and layouts adopted by the journals in which the manuscripts were published or submitted, the chapters (II to IV) were formatted in a single style, with all references listed together in the "References" section. Finally, annexes relative to each chapter were also compiled in a last section of this PhD dissertation in a section denominated "Supplemental material". To facilitate the access, a link to an online document is provided for some extensive supplementary data. Numbering of figures and tables is presented according to the number of the chapter. The specific contents of each chapter are enclosed in this PhD thesis, as follows:

Chapter I. Consist of a general introduction that intends to supply the reader with the state of the art in the subjects addressed in this doctoral dissertation around the brucellosis infection. On behalf of this, it is firstly given a global overview of the major aspects of the *Brucella* spp. taxonomy, biology, molecular epidemiology, pathogenesis and clinical presentation, followed by insights into the genetic and phylogenetic analysis. It ends with the description of the main objectives of this Ph.D. project, and includes the specific research questions that drove the investigations carried out on behalf of each chapter.

Chapter II. Consists of the following published scientific manuscript: Pelerito A, Cordeiro R, Matos R, Santos MA, Soeiro S, Santos J, Manita C, Rio C, Santo M, Paixão E, Nunes A, Núncio S. “Human brucellosis in Portugal-Retrospective analysis of suspected clinical cases of infection from 2009 to 2016”. PLoS One. 2017 Jul 10;12(7):e0179667. doi: 10.1371/journal.pone.0179667

Chapter III. Consists of the following scientific manuscript submitted for publication: Ana Pelerito, Alexandra Nunes, Sofia Núncio, João Paulo Gomes. “Genome-scale genetic relatedness among *Brucella melitensis* strains causing human infections in Portugal”.

Chapter IV. Consists of the following scientific manuscript submitted for publication: Ana Pelerito, Alexandra Nunes, Joana Isidro, Catarina Silva, Ferreira AC, Valdezate S, Sofia Núncio, Enrico Georgi, João Paulo Gomes. " Evaluation of an *in silico* approach for Multiple Locus Variable Number Tandem Repeat Analysis for genetic characterization of *Brucella* spp. "

Chapter V. Consists on an ongoing study entitled: "Genetic diversity of Type IV *Brucella* spp. effectors among *B. melitensis* strains circulating in Portugal"

Chapter VI. This chapter includes a global and conclusive overview of the subjects addressed throughout the chapters, highlighting the main results and conclusions achieved in this Ph.D. dissertation. New research questions raised on the course of this work that can be addressed in the future development of these investigations are also presented.

CHAPTER I

GENERAL INTRODUCTION

1.0 Introduction

1.1 The Genus *Brucella*

In 1886, in Malta, David Bruce (1855-1931), a British army surgeon, isolated a cocco-bacillus that he named "Micrococcus melitensis" from the spleen of a man who had died of "Malta Fever" (Moreno, 2014; Mantur *et al*., 2007; Rust *et al*., 2006). This disease was endemic in this country, but could be erroneously mistaken as another disease, especially malaria. Between 1901 and 1906, it affected 652 civilian and 605 military, with a death rate of 10.4% and 2.3%, respectively (De Ley *et al*., 1987). The human disease was associated with people that either consumed goat's milk or had close contact with goats, and soon the microorganism was isolated from these animals (Tonna and Tonna, 2005). In 1897 a similar microbe was isolated from the udder of cows, and in 1914 from swine (Murray *et al*., 2005; Carvalho *et al*., 1995). In about 1920 the genus was renamed *Brucella*, enrolling the species *B. melitensis*, *B. abortus*, and *B. suis* and marking the beginning of the history of brucellosis, one of the most extended bacterial zoonosis at a global level and a complex infection of animals and humans with worldwide impact (Wiat, 2013; Araf, 2010).

Brucella spp. are facultative intracellular, Gram-negative, non-motile, partially acid-fast coccobacilli that lack capsules, flagellae, endospores or native plasmids. The bacterium has 0.5-0.7 μm of diameter, 0.6 – 1.5 μm of length and is oxidase, catalase and urease positive (Young, 1995).

Based on 16S rRNA gene sequences, *Brucella* spp. are categorised as α–2 Proteobacteria and have close phylogenetic relationships with *Agrobacterium*, *Rickettsia*, *Rhizobium* and *Rhodobacter* (Bohlin *et al*., 2010; Whatmore *et al*., 2009; Scholz *et al*., 2008).

The genus *Brucella*, belongs to the family *Brucellaceacae* within the order *Rhizobiales* of the class Alphaproteobacteria, which is one of the largest and most diverse groups within the phylum Proteobacteria (Scholz *et al*., 2008). *Brucella* is currently classified according to differences in pathogenicity and host preference in 12 species (Scholz *et al*., 2008; Scholz *et al*., 2010; Marzetti, *et al*., 2013).

The most important *Brucella* species are *B. abortus* and *B. melitensis* (Table 1.1), which have, as preferential hosts, cattle and small ruminants (sheep and goats). *B. suis* was isolated for the first time from an aborted pigs fetus in Europe in 1909, and later on in the United States. For many years, it was believed that the agent was a highly pathogenic variant of *B. abortus* but in 1929, *B. suis* was finally considered a separate species. In 1956, Buddle and Boyce (Buddle, 1956) discovered *B. ovis*, the cause of epididymitis in rams. In 1957, Stoenner and Lackman isolated *B. neotomae* from desert wood rat in Utah in USA. In 1968, Carmicheal and Bruner (Lucero *et al*., 2005) discovered *B. canis* as the cause of an epidemic of abortions in beagles. However, human infections due to *B. canis* have also been reported (Marzetti *et al*., 2013). Since 2007, more species were included in the genus: *B. ceti* and *B. pinnipedialis*, which were isolated from marine mammals (Foster, *et al*., 2007), *B. microti* from voles (Scholz, *et al*.,

2008), *B. inopinata* from an inflamed breast implant of a 71 year – old patient in USA, and more recently, *B. papionis*, isolated from baboons and *B. vulpi*, isolated from foxes. The natural reservoir of *B. inopinata* still remains unclear (Scholz, *et al*., 2010; Scholz, *et al*., 2016; Eisenberg, *et al*., 2017). Other "atypical" *Brucella* strains have been isolated from diverse animal sources such as wild rodents, frogs and fish, and will likely be proposed as new species in a near future (Eisenberg, *et al*., 2017; Tiller, *et al*., 2010; Eisenberg, *et al*., 2012). *B. abortus*, *B. melitensis* and *B. suis* are recognised as the most economically significant pathogens of the group.

Table 1.1 – Species of *Brucella*, preferential hosts and pathogenicity for humans. [Adapted from Al Dahouk, 2013]

***Brucella* species**	**Natural host**	**Pathogenicity for humans**
B. melitensis	Sheep, goat and camels	High
B. abortus	Cattle	Moderate
B. suis	Swine	High
B. neotomae	Rodents	No
B. ovis	Ram	No
B. canis	Dog	Moderate
B. ceti	Cetaceans	Unknown
B. pinnipedialis	Pinnipeds	Unknown
B. microti	Soil, vole, fox	Unknown
B. inopinata	Unknown	High
B. papionis sp. *nov.*	Baboons	Unknown
B. vulpis sp. *nov.*	Foxes	Unknown

1.2 Pathogenicity and immune response

Brucella **s**pp. have the ability to avoid the killing mechanism and proliferate within the macrophages. These bacteria quickly translocate across the mucosal epithelium layer *in vivo* and are endocytosed by mucosal macrophages and dendritic cells. To be a successful pathogen *Brucella* **s**pp. requires four major steps during the infectious process, namely, adherence, invasion, establishment, and dissemination within the host. In the macrophages, *Brucella* **s**pp. cells survive and multiply, inhibiting phagosome–lysososme fusion and the accumulated bacteria are disseminated to other host cells (McDermott *et al*., 2013; Figueiredo *et al*., 2015)

During infection, the host has evolved mechanisms to recognize the presence of bacteria through an innate immune surveillance system, which is able to distinguish conserved "pathogen–associated molecular patterns" (PAMPs) through pathogen recognition receptors. These host receptors can be found in cell membranes (toll-like receptors, TLRs) or in the cytosol (NOD–like receptors, NLRs) and have the ability to detect bacterial products, such as the lipopolysaccharide (LPS), lipoteichoic acids, lipoproteins and flagelin, leading to the activation of the initial proinflammatory response. As a chronic

pathogen, *Brucella* has developed passive and active mechanisms to evade detection by both TLRs and NLRs in order to persist and cause long-lasting infection (Smith, 2018).

The mechanism involved in *Brucella spp.* entry into host cells still remain to be characterized (Gorvel, 2014), but its ability to successfully survive and replicate within different hosts cells explains their pathogenicity. In fact, extensive replication of *Brucella* spp. in placental treophoblasts is associated with abortion in their preferential hosts, and persistence in macrophages leads to chronic infections that are a hallmark of brucellosis in both animals and humans (Kim, 2015; Gorvel, 2014; Grilló *et al*., 2012; Roop *et al*., 2009).

In vitro studies were used as models to understand adhesion, internalization, intracellular trafficking, survival and replication of *Brucella* in susceptible hosts. After attachment to the surface of mucosal epithelial cells, *Brucella* induces a zipper–like mechanism for internalization. Binding promotes activation of small GTPases that trigger a signaling cascade that reorganizes the actin cytosqueleton to induce a host cell membrane rearrangement along the surface of the pathogen, which enhances invasion. Entry occurs within a few minutes after interaction, which requires full activation of a mitogen–activated protein kinase-signaling pathway (Rosseti, 2012). *Brucella* survive and replicate inside nonprofessional phagocytic cells up to 72 hours *in vitro* and move across the epithelium *in vivo* by subverting the mucosal epithelial barrier function to facilitate *Brucella* transepithelial migration (Rosseti *et al*., 2012). Simultaneously, this interaction initiates a minimal innate immune response with weak proinflamatory activity (Rosseti *et al*., 2012; Barquero-Calvo *et al*., 2007). Inside mononuclear phagocytic cells, *Brucella* reside in a special vacuole (*Brucella* – containing vacuole, BCV), modify intracellular trafficking, and transform the vacuole into a replicative compartment (Ruiz – Ranwez *et al*., 2015; Gomez *et al*., 2013; Lamontagne *et al*., 2010).

In the course of infection, invading *Brucella* spp. surviving the adaptation period gradually recover the expression of key genes involved in metabolic processes, and initiate replication concurrently with the resumption of expression of multiple virulence genes. Among the early transcription changes that contribute to adaptation, this bacterium has several clever strategies to establish and maintain a chronic infection, including inhibition of apoptosis of infected mononuclear cells and preventing maturation of dendritic cells (Wattam *et al*., 2014).

The *Brucella* outer membrane constitutes an important barrier for survival in hostile environments and is an accessible target for the interaction with the host and defense mechanisms of the immune system **(**Vizcaíno and Cloeckaert, 2012). Structurally, it is an asymmetrical lipid bilayer composed of LPS and other haptenic polysaccharides, such as hapten native (NH), proteins and phospholipids (PL), with the LPS molecules and PL located in the outer and inner leaflet, respectively. Like other Gram-negative bacteria, *Brucella* have LPS as a major component of their outer membrane and an important virulence factor (Smith, 2018). The *Brucella spp.* LPS possess unusual immunological properties such as low toxicity, high resistance to macrophage degradation and protection against immune responses, being a major virulence factor in *Brucella* (Lapaque *et al*, 2005; Supriya *et al*., 2019).

Since LPS is the most relevant antigen during infection and vaccination, LPS and LPS–related molecules are extensively used in immunological studies and in the diagnosis of brucellosis. Among Gram-negative bacteria, the genus *Brucella* is the unique in which some species express the smooth (S) – type LPS (*B. abortus*, *B. melitensis*, *B. suis*, *B. microti*, *B. neotome*, *B. ceti*, *B. pinnipedialis*, *B. inopinata* and *B. papionis*) and others have naturally rough (R) – type LPS (*B. canis* and *B. ovis*). The S-LPS and R-LPS differ mostly in the most external LPS moiety (the O-polysaccharide), which is not synthetized in rough *Brucella* species (Moreno and Myrion, 2006; Lapaque *et al*.2005). Colony morphology is termed "smooth" or "rough" dependending on the LPS structure (Moreno and Myrion, 2006).

Numerous outer and inner membranes, cytoplasmic and periplasmic protein antigens have also been characterized. Some are recognized by the immune system during infection and are potentially useful in diagnostic tests. Omp 25 is an outer membrane structural protein that is highly conserved in all *Brucella* species. It is associated with lipopolysaccharide components and is involved in protection against challenge with *Brucella* via both antibody and cell mediated responses. (Lapaque *et al*., 2005).

In recent years, various virulence factors besides LPS have been identified as essential for infection, including the β-cyclic glucan, the BvR/BvS, two component systems (TCS), some OMPs and the VirB type IV secretion system (T4SS) (Sellem *et al*., 2008, Martim-Martim *et al*., 2011). Quorum sensing (QS) is also known to be involved in the regulation of *Brucella spp.* virulence determinants mostly linked to the cell surface (T4SS, flagellum, Omps and exopolysaccharide) contributing to the adaptation of the metabolic network during the nutrient shift faced by *Brucella spp.* along its intracellular trafficking (Gorvel, 2014; Weeks *et al*., 2010; Rambow-Larsen *et al*., 2009; Letesson *et al*., 2002).

1.3 Epidemiology

Brucella spp. can be traced back in 2.8 million years by presumptive evidence of pathologic changes in a late Pliocene homini skeleton (D'Anastasio, R *et al*., 2009). Additionally, molecular tests demonstrated the presence of *B. melitensis* DNA in a 700 – year- old skeleton from medieval Italy (Kay *et al*., 2014).

Brucellosis is the most common bacterial zoonosis worldwide in terrestrial and marine environments. In addition to the hundreds of thousands of new infections reported annually, the disease is characterized by progression, in a significant percentage of patients, to residual pathology and chronicity. Annually, more than 500.000 new human cases de brucellosis are reported worldwide, but ranks as one of the seven most neglected diseases, according to the World Health Organization (Hull and Schumaker, 2018). During the last two decades, the epidemiology of brucellosis has changed significantly, with the emergence of new global outbreaks in association with major political/historical events, successful control of the disease in many parts of the Mediterranean, and the referral of epidemiological data from countries where brucellosis was endemic but in an unknown proportion (Pappas *et al*., 2006; Pappas *et al*., 2010). The majority of brucellosis cases are registered in

Mediterranean countries, South and Central America, Africa, Asia, Indian subcontinent, Eastern Europe and the Middle East (Pappas *et al*., 2006). Successful implementation of an animal vaccination program, in addition to testing and slaughter of animals which are suspected of (or test positive for) the disease has conferred "officially bovine brucellosis free" and/or "officially ovine and caprine (*Brucella melitensis*) free" status on a number of countries (Shevtsova *et al*., 2016). Control in humans is critical for prevention of spread to humans indeed, the majority of human cases in brucellosis – free regions are due to import by people who have travelled to endemic regions, as they may have had contact with wild animals or with imported products.

The geographical distribution of brucellosis is constantly changing, with the emergence and re-emergence of new outbreaks around the world. Reflecting the social, cultural, and economic policies that describe a changing global society, this pathology has been replicating this dynamics, making its control and eradication a constant challenge.

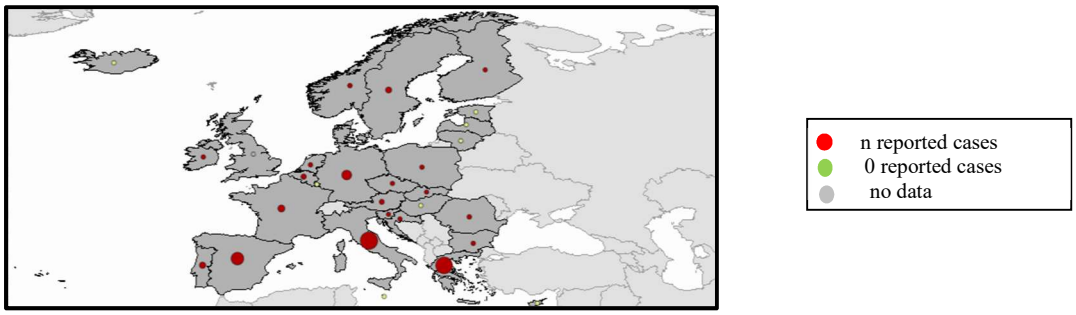


Figure 1.1 – Reported cases by European Centre for Disease Prevention and Control (ECDC - 2018)

The distribution of the *Brucella* species. *B. abortus*, *B. melitensis* and biovars 1-3 of *B. suis* have been virtually eliminated from livestock in many developed countries. However, some of these organisms are common in parts of Middle East, Asia and Latin America (Godfroid *et al*., 2010**)**. Feral pigs or wild boar continue to maintain *B. suis* biovars 1, 2 or 3 in many areas where *B. suis* is virtually absent from commercial swine, and a few foci of wildlife reservoirs for *B. abortus* or *B. melitensis* have been identified in limited areas. The distribution of some organisms including *B. microti*, *B. neotomae*, *B. vulpis*, *B. papions* and *B. inopinata*, is still poorly understood (Jaý *et al*., 2018).

1.4 Transmission, clinical presentation and treatment

Brucellosis is rarely transmitted from person to person, and transmission reports involve bone marrow transplantation, blood transfusion or sexual intercourse. Common sources of human infection include contact with animal abortion products, ingestion of unpasteurized dairy products from cows, small ruminants or camels, ingestion of undercooked meat, bone marrow or other uncooked meat products. Besides being an infectious disease, human brucellosis is also considered an occupational disease because it usually occurs during occupational activities that expose workers to *Brucella*

(Weinstein and Singh, 2009). Populations such as abattoir workers, veterinarians, lab technicians, hunters, farmers, and livestock producers are involved in such activities (Weiming *et al*., 2018). Occupational brucellosis is predominantly reported in the animal husbandry, agriculture, meat processing and vaccine production industries. In these occupational fields, exposure to *Brucella* can be prevented by maintaining good hygiene and using protective equipment (Weiming *et al*., 2018).

Four *Brucella* species, namely *B. abortus*, *B. melitensis*, *B. suis* and *B. canis* are classified as category B bioterrorism agents according to the Centers for Disease Control and Prevention (CDC), as they are moderately easy to disseminate, and they result in moderate morbidity and low mortality rates (Centers for Disease C. prevention, 2000). Extensive studies have been done in the past exploring the potential of *Brucella* spp. as biological weapon agents (Franz *et al*., 1997; Doganay and Doganay, 2013). In fact, there are several biological and pathogenic properties of *Brucella* spp. that make them useful agents of biological warfare. The major characteristic is that they are highly infectious via the aerosol route, with an infectious dose estimated at approximately 10-100 organisms (Mense *et al*., 2004). The mechanisms of transmission, through aerosols or food chains, makes them easily transmissible to both humans and animals. It has been estimated that the release of 50 kg of *B. suis* from a plane along a 2-10 km line upwind of a population center of 500,000 people would result in 500 deaths and 125,000 people being incapacitated (Doganay and Doganay, 2013).

Brucella spp. can entry into the human host by several ways: via inhalation, ingestion, contact with mucosa or puncture wounds such as needle sticks (Franco *et al*., 2007). This is followed by an incubation of 10-21 days (but can take as long as 12 months), a brief bacteraemia, and localization to the mononuclear phagocyte system (Brucellosis, 2009). The disease has several clinical presentations, depending on the species, the mode of transmission and the host immune response (Baldi and Giambartolomei, 2013). Asymptomatic infections are common. In symptomatic cases, the disease is extremely variable and the clinical symptoms may appear insidiously or abruptly. Typically, brucellosis begins as an acute febrile illness with nonspecific flu-like signs such as fever, headache, malaise, back pain, myalgia and generalized aches. Drenching sweats can occur, particularly at night (Pappas *et al*., 2005). Splenomegaly, hepatomegaly, coughing and pleuritic chest pain are sometimes seen. Gastrointestinal signs including anorexia, nausea, vomiting, diarrhea and constipation occur frequently in adults but less often in children. Acute and chronic brucellosis can lead to complications in multiple organ systems. The most common complications are arthritis, spondylithis, epididymoorchitis and chronic fatigue. Neurological signs occur in up to 5% of cases (Dean *et al*., 2012**)**. They may include personality changes, meningitides, encephalitis and peripheral neuropathy. Endocarditis is one of the most serious complication, and is often the cause of death in fatal cases.

Depending on the timing of treatment and severity of illness, recovery may take a few weeks to several months. Some forms of localized disease, such as endocarditis, may require surgery. Death from brucellosis is rare, occurring in only about 2% of all cases (Franco *et al*., 2007). The choice of a regimen and duration of antimicrobial therapy should be based on the site of the infection and the underlying

conditions of the patient (Solera, 2010). Treatment of brucellosis should involve antibiotics that can penetrate macrophages and can act in acidic intracellular environment, as *Brucella* is an intracellular pathogen. Also treatment with duration of less than 4 weeks have a high risk for relapse (Pappas *et al*., 2005). So, combination of two antibiotics for longer duration of therapy for at least six weeks is recommended. According to WHO guidelines, the recommended combination of treatment for human brucellosis is doxycycline along with either rifampicin or streptomycin, a recommendation that has been in place for more than a decade (Pappas *et al*., 2005).

Isolation of resistant *Brucella* strains highlights new difficulties for managing antibiotic therapy. It may actually mean we could have to give up the use of easily administered low cost oral antibiotics, which can be prescribed to children and pregnant women. Both dual therapy (doxycycline and rifampicin) and triple therapy (doxycycline, rifampicin, and gentamicin) have been already administered (Skalsky *et al*., 2018**)**, and the rationale behind the use of the combination of three antibiotics is the possibility of reducing the risk of brucellosis recurrence (Yousefi-Nooraie *et al*., 2012**)**. Time to diagnosis is crucial to choose the best therapy regimen and to avoid complications that lengthen the hospitalization, increasing the expenses for the national healthcare system. It is believed that, in a future scenario, where *Brucella* becomes more frequently resistant to antibiotics, the management of an already subtle disease will become more difficult, where setting up therapeutic regimens with multiple antibiotics could eventually reduce the risk of disease recurrence (Skalsky *et al*., 2018).

1.5 Diagnostics and typing of brucellosis

The laboratory investigation for brucellosis is usually done by isolation of the bacteria or detection of anti-*Brucella* antibodies. Although isolation of the bacteria is the "gold standard", it requires long incubation periods and is seldom successful (Jama'ayah *et al*., 2011; Rich *et al*., 2000; Robichand *et al*., 2004). Therefore, although culture has a definitive diagnostic value, serologic tests have a major role in brucellosis diagnosis.

Several serological tests have been used for the diagnosis of human brucellosis (Avijgan M, *et al*, 2019). The serum agglutination test (SAT) for brucellosis, developed by Wright *et al* in 1897 (Wright *et al*., 1997), is still the reference method to which other tests are compared. Other methods that have been developed since then include the Rose Bengal test, the complement fixation test, the indirect Coombs test, enzyme immunoassays (ELISA) (Gad El-Rab *et al*., 1998), and an immunocapture-agglutination test (Brucellacapt) (Orduña *et al*., 2000; Rubio *et al*., 2001). Usually, the sensitivity of the serological tests range from 65 to 95%, but the specificity in areas where brucellosis is endemic are low because a large part of the population has contact with infected animals or products of animal origin and develop and maintain antibodies against *Brucella* even without the occurrence of active infection. Moreover, most serological tests can produce cross-reactions with other bacteria and also exhibit important limitations with samples taken in the early phases of the disease, from persons exposed

professionally, from patients with a recent history of brucellosis, and from patients who relapse (Queipo – Ortuno, *et al*, 1997).

The routine identification and differentiation of *Brucella* species is typically based on phenotypic traits. However, it is a lengthy process that requires experience technicians, and is associated with a high risk of laboratory–acquired infection (Hinic *et al*., 2008). In order to overcome these difficulties, and in spite of the high degrees of genetic similarity among different species, several conventional and real time PCR assays have been established. The conventional polymerase chain reaction (PCR) and multiplex PCR typing are capable of identifying *Brucella* up to species level. On this regard, a multiplex PCR assay, AMOS PCR (AMOS is acronym from "abortus, melitensis, ovis, suis"), is able to identify *B. abortus*, *B. melitensis*, *B. ovis*, and *B. suis,* including the discrimination at biovar level, using a combination of different PCR primers. This PCR assay had successfully identified *B. abortus* biovars 5, 6, and 9 and some field strains of biovar 3 *B. abortus* (Hinic *et al*., 2008). However, this method needs at least 5–6 h to be completed and required post-amplification handling of PCR products, which may yield contaminations. For the rapid, sensitive and accurate detection of *Brucella* spp. the multiple insertion element IS711, which is stable in both number and position in the *Brucella* chromossomes, has been choosen as a target (Hinic *et al*., 2008). Also, a non-multiplex PCR for species differentiation has been developed, which is based on unique genetic loci of *B. melitensis*, *B. abortus*, *B. suis*, *B. ovis*, *B. canis* and *B. neotonae*. Nevertheless, the simple identification of genus and, in some cases, species by PCR assays, is adequate for purposes of diagnosis of human/animal disease or identification of food contamination but not for the tracing of outbreaks or bioterrorism attacks. To achieve these goals, whole-genome sequencing-based approaches are required, due to their higher discriminatory power (Santis, *et al*).

Identification and typing of *B. melitensis* are still traditionally performed with the use of biotyping techniques. This methodology, however, suffers from inconsistencies and requires handling of the live bacteria. For this reasons, PCR-based typing is now commonly used as an alternative to the culture dependent typing methods. The results of the classical biotyping schemes categorize *B. melitensis* into three biovars that are of limited epidemiological value, as they do not provide sufficient resolution between the isolates. In fact, *B. melitensis* is considered a monomorphic pathogen, which renders its differentiation at the strain level very difficult. To fulfill the objective of sub-species discrimination, Variable Number Tandem Repeats (VNTR) have been investigated in Multi-locus VNTR Analysis (MLVA) by various scientific groups since 2003 (Le Fleche *et al*., 2006; Mambres, *et al*., 2017; Sun *et al*., 2017). This *Brucella* typing scheme, using VNTRs based on 16 *loci* ("MLVA-16"), has been proven to have the ability to differentiate *Brucella* species, biovar and even the isolates. More importantly, there is an online database of MLVA-16 profiles available to all researchers allowing the comparison of *Brucella* strains at the worldwide scale (Le fleche *et al*., 2006; Sun *et al*., 2017; Mambres *et al*., 2017). Although MLVA has become a major molecular typing method to characterize several

pathogenic bacterial species, this methodology is laborious, time consuming and frequently the amplification of all *loci* cannot be achieved.

Overall, efficient and reliable surveillance programs are essential for detection and control of outbreaks and largely depend on collection and access to epidemiological data. Currently, epidemiological investigations rely on the availability of standardized and effective molecular typing methods and analysis tools that allow the public health laboratories to identify and trace an outbreak back to its source.

1.6 Comparative Genomics and phylogenetic analysis

Brucella species are characterized by extremely high levels of nucleotide similarity despite the notorious differences in host tropisms, microbial and disease phenotypes and pathogenicity. For many years molecular studies and the development of molecular typing tools were hampered by this lack of diversity. However, gradual progress was made in identifying useful markers and tools and this progress has been greatly accelerated in recent years by the availability of genome sequence (Parkhill and Wren, 2011; Bohlin *et al*., 2010).

Distinguishing individual bacterial lineages within a species, initially by phenotypic and subsequently by genotypic typing techniques, has been the cornerstone of infectious disease epidemiology, allowing the identification and tracking of the organisms responsible for infection and disease [Parkhill and Wren, 2011]. During the past decade, the understanding of evolution at the genomic level has been shaken to its core by many reports showing that genomes from closely related species can vary in terms of gene content [Williams *et al*., 2011].

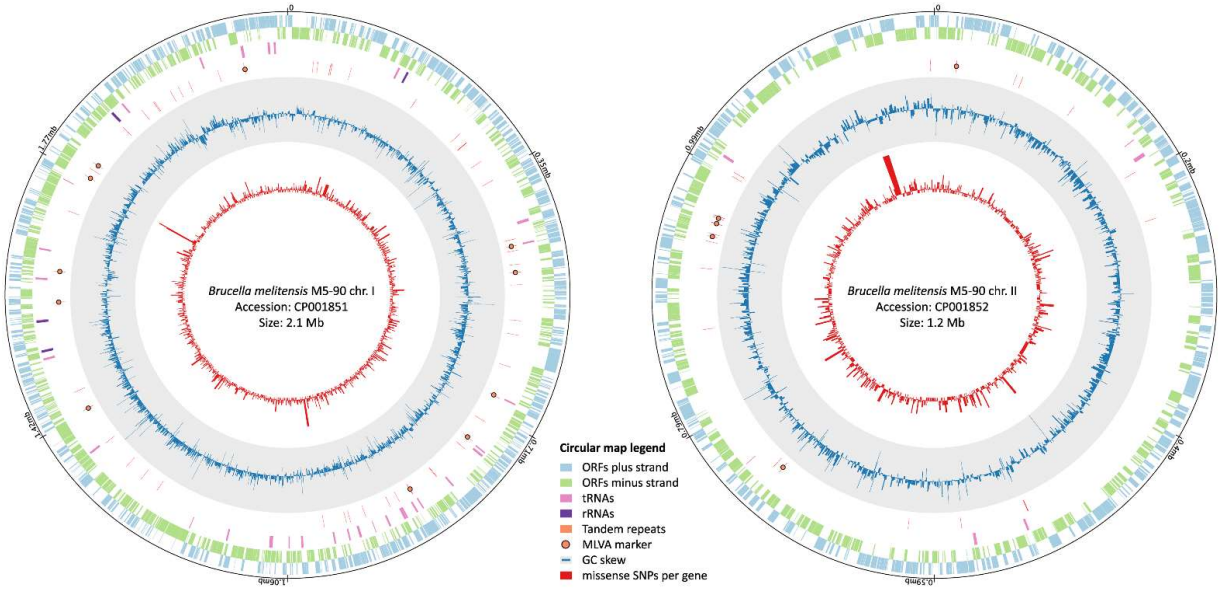


Figure 1. 2 – Circular view of the two chromosomes of Brucella melitensis displaying some of the major genetic features loci. (Georgi E *et al*., 2017)

In 2002, the first *Brucella* genomes became available with the publication of the *B. melitensis* 16M (DelVecchio *et al*., 2002) genome, followed a few months later by *B. suis* 1330 (Paulsen *et al*., 2002) and *B. abortus* biovar 1 strain 9-941 (Halling *et al*., 2005). The *Brucella* genome is composed of two circular chromosomes of approximately 2.1 and 1.2 Mb (Figure 1.2). Both chromosomes share similar GC content, a similar proportion of coding regions and equivalent housekeeping gene distribution (Sankarasubramanian, *et al*., 2017). The existence of numerous transposons, insertion elements and phage remnants suggest a vigorous contribution of these mobile genetic elements to evolution. Despite an evolutionary divergence and/or host specific adaptation, orthologous characteristics relevant to virulence do not appear to have undergone substantial change within genus. Although there are several examples of phage – mediated and other insertion/deletion events that may account for differences in virulence and host specificity, their contribution through small sequence changes (single nucleotide polymorphisms) in orthologous functions remains the primary potential source of distinction (Ficht, 2010).

To date, genome sequences from more than 770 different *Brucella* strains, representing all species, have been published either as complete genomes or as draft assemblies. The overriding conclusion derived from comparative genomics studies was that the genomes are all highly conserved, but with a variety of indels and recombination events. The major difference in chromosome I was previously identified in *B. suis* biovar 2 strain Thomsen (ATCC23445), where a 210kb segment of chromosome I has been translocated to chromosome II (Wattam *et al*., 2009). In contrast, chromosome II was found to be somewhat more variable as more internal rearrangements, including the 700 kb inversion in *B. abortus* genomes, have been described.

The comparison of the genomes of *B. melitensis* 16M, *B. suis* 1330 and *B. abortus* biovar 1 strain 9-941 confirmed just how closely related the three species are. Their genomes revealed extensive gene similarity and syntonic with the majority of genes (>90%) sharing 98-100% identity (Verger *et al*., 1985). More variable genes (<95% identity) were confined to genes encoding hypothetical genes and probable surface exposed proteins such as outer membrane proteins, membrane transporters, putative invasion and ShdA-like adhesins (Paulsen *et al*., 2002). Considering these early findings much research over subsequent years focused on the identification of molecular markers and suitable experimental approaches to discriminate between members of the genus. While in many cases simple identification as a member of the genus is adequate for practical purposes (e.g. diagnosis of human disease or identification of food contamination), in other cases, identification to species or subspecies level is needed. For example, most government sponsored eradication programs include regulations stipulating a species-specific response and sub-typing is essential to facilitate any degree of epidemiological trace back. As such, the simple identification of genus and, in some cases, species by PCR assays is adequate for purposes of diagnosis of human/animal disease or identification of food contamination but not for the tracing of outbreaks or bioterrorism attacks (Santis *et al*., 2011; Whatmore *et al*., 2007). It is

unequivocal that the advent of next generation sequencing has been a technological revolution that fulfils these requirements, as full bacterial genome sequencing has become highly accessible (Foster *et al*., 2009; Liu *et al*., 2012; Wattam *et al*., 2009). Unlike the taxonomically informative or canonical single nucleotide polymorphims (SNP) – based approaches, whole genome sequencing serves as a robust an unbiased method to resolve intraspecies relationships for closely related species such as *Brucella* spp. In this regard, Tan *et al* (Tan *et al*., 2015) draw a map of the global genetic diversity of *B. melitensis* strains isolated in different continents. In their study, *B. melitensis* isolates are represented by five genotypes: Mediterrean strains are identified as genotype I, Asian strains are classified as genotype II, genotype III is represented by strains of African descendent. The genotypes IV and V are assigned respectively to the European and American lines. (Tan *et al*., 2015).

In recent years, the typing methods have shifted towards genome-based approaches that finally allowed an accurate differentiation between *Brucella* isolates and establishment of a common consensus for the subtyping schemes of this pathogen. Rather than constructing phylogenetic inferences from a small portion of the genome (like MLST, MLVA) entire genomes can now be compared. Recent implementation of whole-genome single nucleotide polymorphism (SNP)-based typing has led to substantial improvements of both molecular subtyping and phylogenetic analyses in microbiology. The development of core- and whole-genome MLST schemes has been focused on a restrict number of bacterial pathogens, including *Brucella* spp. but their application may be tricky (Tan *et al*., 2015 Janowicz *et al*., 2018; Sankarasubramanian *et al*., 2018). Also, whole genome comparisons and phylogenetic analysis of *Brucella* have only been done on a limited scale. With the technological advances and decreased cost of Whole Genome Sequencing, new methods of pathogen typing, including gene-by-gene comparison using core genome multilocus sequence typing (cgMLST), as well as single-nucleotide polymorphism (SNP) calling based on a reference sequence analysis, are considered to be a suitable and more informative replacement of the gold standard typing schemes (Sankarasubramanian *et al*., 2019). Recently, two cgMLST schemas were developed for *Brucella*: one genus-specific (Sankarasubramanian *et al*., 2019) and the other exclusively for *B. melitensis* (Janowicz *et al*., 2018; Sankarasubramanian *et al*, 2019). However, despite the former is incorporated on a freely web Brucella platform (http:// www.dbtbrucellosis.in/brucellabase.html), it is based on a very small panel of 164 loci. On the other hand, the latter contains a wider core gene set of 2704 targets but is available on a pay-per-use platform.

Overall, genomic studies have already contributed substantially to our understanding of the biology of *Brucella* and have facilitated the development of new tools to identify and characterize members of the group. As we move into an era of availability of multiple genomes within individual species, progress should accelerate in a number of areas where genomic analysis offer huge promise. This provides the direct understanding of relationships between isolates and a framework for the generation of hypotheses for further biological investigation. Exploitation of existing and forthcoming

genome sequences will potentiate our understanding across multiple areas of *Brucella* biology, such as virulence and pathogenic processes associated with *Brucella* that remain relatively poorly understood.

1.7 Aim of the thesis

According to the described background, the major aims of this thesis can be pointed out as follows:

- contribute to a more accurate evaluation of the epidemiological situation of human brucellosis in Portugal;

- identify potential outbreaks and transmission links the brucellosis infection through the implementation of a whole genome sequencing (WGS) approach to give a step forward in the *Brucella* spp. surveillance in Portugal;

- evaluate the agreement between experimental (*i.e.*, wet lab based) and the developed *in silico* determination of MLVA for strains comprising several *Brucella* species in order to check the validity of such technological transition underlying the genetic characterization of *Brucella*;

- evaluate the genetic polymorphism of several virulence factors of *B. melitensis* strains.

CHAPTER II

HUMAN BRUCELLOSIS IN PORTUGAL – RETROSPECTIVE ANALYSIS OF SUSPECTED CLINICAL CASES OF INFECTION FROM 2009 TO 2016

The results presented in this chapter were previously published

Pelerito A, Cordeiro R, Matos R, Santos MA, Soeiro S, Santos J, Manita C, Rio C, Santo M, Paixão E, Nunes A, Núncio S. Human brucellosis in Portugal-Retrospective analysis of suspected clinical cases of infection from 2009 to 2016. PLoS One. 2017 Jul 10;12(7):e0179667. doi: 10.1371/journal.pone.0179667

Personal contribution

AP contributed to the design of the study, performed most of the experimental work, interpreted data and wrote the manuscript.

2 Human Brucellosis in Portugal – Retrospective analysis of suspected clinical cases of infection from 2009 to 2016

2.1 Abstract

Brucellosis is a zoonosis that is emerging in some regions of the world. Although brucellosis is a disease of obligatory declaration and is not eradicated in Portugal, no prevalence data is available in this country. In this study, we retrospectively analyzed the data available at the Reference Laboratory at the Portuguese National Institute of Health during the past 7 years (2009-2016) in order to get insight into the epidemiological scenario of brucellosis in Portugal. A total of 2313 biological samples from patients with clinical suspicion of brucellosis were subjected to immunological techniques for laboratory diagnosis. From 2010 to 2015, a subset of 259 samples was subjected to molecular methods. According to the available data, 167 out of 2313 (7.2%) samples had positive serology for *Brucella* spp. and 43 out of 259 samples (16.6%) were positive for *B. melitensis* by real time PCR, being classified as biovar 1 and 3. This study draws attention to the importance of integrating clinical and laboratory data of human cases in order to increase the efficacy of the response measures in case of outbreaks.

2.2 Introduction

Brucellosis is a worldwide zoonosis caused by the intracellular facultative bacteria of the *Brucella* genus (Young, 2005; Seleem *et al*., 2010). The later currently encloses 12 species, five of which (*B. abortus*, *B. suis*, *B. melitensis*, *B. ovis* and rarely *B. canis*) are the ones more commonly associated with human disease (Whatmore *et al*., 2014; Scholz *et al*., 2016). *B. melitensis* is the most virulent and has the largest public health impact in the EU due to its predominance in small ruminant populations (Young, 2005). Human brucellosis, also known as Malta fever, Undulating, Mediterranean, Gibraltar or Bang Disease, affects the well-being of people, not only as a disease in man and animals, but due to its economic impact, since it implies heavy losses in livestock farms. It also influences people´s life quality, especially those who live in rural areas, where contact with animals and the consumption of food and milk from homebred animal origin is more frequent and less controlled (Whatmore *et al*., 2009). The two most common ways of human infection are through the contact with infected animals or the ingestion of unpasteurized dairy products. Risk groups for this disease include individuals that work with unvaccinated infected animals, farmers, slaughterhouse workers and veterinarians. They get infected through direct contact or inhalation of aerosols produced by the infected animal tissue. This situation is frequently found in areas where brucellosis is endemic in ovine and bovine cattle, and it is usually associated with infection by *B. melitensis* (Moreno, 2014). Human brucellosis is a systemic disease that may affect any organ or system, in subacute, acute or chronic form. The disease has several clinical presentations, depending on the species, the mode of transmission and

also the host immune response (Baldi *et al*., 2013). The incubation period is difficult to determine in humans, ranging from one week to more than two months (usually 2-4 weeks) (referênciahttp://www.who.int/zoonoses/diseases/Brucellosissurveillance.pdf). Fever, night sweats, severe headache and body aches and other non-specific symptoms may occur. Acute and chronic brucellosis can lead to complications in multiple organ systems. The musculoskeletal system, central nervous system, respiratory tract, the liver, heart, gastrointestinal and genitourinary tracts can all be affected. Untreated brucellosis has a fatality rate of 5% (Franco *et al*., 2007).

The inclusion of *Brucella* spp. in the list of agents with the potential to be used as a biological weapon increased the concern of the authorities responsible for human and animal health (CDC, 1999; Jacobs *et al*., 2004) and made reference laboratories ensure constant improvement and update their laboratory methods for diagnosis and early detection of the pathogen in both environmental, food and biological samples (Araj, 2010; Pappas *et al*., 2006; Tzaneva *et al*., 2007). On this regard, it is also important to have the complete information regarding phenotype and genotype of the strains that are most prevalent in each geographic region.

The laboratory diagnosis is based on the use of direct methods, such as the isolation of the causative agent for culture analysis and detection of nucleic acids by molecular methods, as well as indirect methods such as the detection of specific antibodies. However, the immunological diagnosis of human brucellosis does not differentiate the species of the genus *Brucella* spp. (Araj, 2010). Recently, several molecular methods were developed, including real time PCR, which reveals great potential for direct and rapid identification of species of the genus *Brucella* spp. (Gopaul *et al*., 2008).

In Portugal, brucellosis is a notifiable disease, and one of the three most frequent zoonosis. Human cases are reported in all regions of continental Portugal, as shown in the 2011-2014 report of the General Directorate of Health (DGS) [Doenças de Declaração Obrigatória 2009-2012). Nevertheless, there is no published study with data on the prevalence and incidence of human brucellosis in Portugal, so the real prevalence of brucellosis in Portugal is unknown. Moreover, for the vast majority of the reported cases it has not been possible to identify which *Brucella* species caused the infection.

This is not done and this lack of information may have serious impact in the identification of the sources of infection, impairs the identification of the most important reservoir hosts and also the implementation of timely and adequate measures that could promote the prevention and/or mitigation of the impact of this infection in the population.

The aim of this study was to contribute to a more accurate evaluation of the epidemiological situation of human brucellosis in Portugal, through the analysis of data available at the Department of Infectious Diseases at the Portuguese National Institute of Health (NIH), gathered between 2009 and 2016.

2.3 Methods

Between 2009 and 2016, 2571 samples from patients with clinical suspicion of brucellosis were received at the Reference Laboratory at the Portuguese National Institute of Health (NIH) for diagnostic purposes. Samples were analyzed by immunological techniques, except for a subset of 259 samples (collected between 2010 and 2016) that were instead analyzed by a combination of molecular methods in agreement with the clinicians' request. In the present study, we conducted a retrospective analysis of all data collected on those samples. No informed consent was obtained from each participant as, besides the information regarding gender and age, no further information was available to the laboratory and no tests besides the ones requested by the clinicians were performed. This procedure is in agreement with the Portuguese law No. 12/2005 of 26 January). The ethical commission of National Institute of Health also approved this study and the anonymity of the patients was maintained.

The immunological diagnosis of brucellosis infection was made using serological methods for antibodies' detection based on agglutination techniques (Rose Bengal (Vircell, Granada, Spain), Wright, 2-mercaptoethanol (Fortness, Diagnostic, UK), Coombs test, indirect immunofluorescence (IFI) and immunoenzymatic assays (Brucella Elisa Igm/IgG Testkit, Virotech, Russelsheim, Germany)). All samples of sera and/or cerebrospinal fluid (CSF) were analyzed at least by two of the mentioned above immunological techniques. According to the Reference Laboratory at the Portuguese NIH guidelines, we considered a positive serological result when we observed a simultaneously positive result for one agglutination technique and one IFI or ELISA. No bacteriological cultures or PCR techniques were attempted in the serologically positive cases.

The molecular methods of brucellosis infection were performed in a tandem fashion. First, an "in house" real time PCR using hydrolysis probes was used to detect and identify the species of *Brucella* genus from blood samples, CSF, biopsies and strains isolated from blood cultures. For the rapid, sensitive and accurate detection of *Brucella* spp., the multiple IS711insertion elements were chosen as they are conserved in both number and position in the *Brucella* chromosomes (Hiníc *et al*., 2008). For species differentiation, primers and Taqman probes were designed within the following ORFs: BMEII0466 gene for *B. melitensis*, BruAb2_0168 gene for *B. abortus* (Hinic *et al*., 2008).

Finally to distinguish *Brucella* biovars, a molecular characterization of the *rpoB* gene was also performed. In contrast to the *16S* rRNA locus, which lacks sufficient sequence variability for differentiation of *Brucella* spp, the *rpoB* gene shows sufficient polymorphism to differentiate all *Brucella* species and their biovars; the exceptions are *B. abortus* biovars 1 and 4 and *B. abortus* biovars 5, 6 and 9, which show the same *rpoB* sequence (Marianelli *et al*., 2006).

Brucella strains were subjected to whole-genome sequencing on a MiSeq Illumina platform (Illumina) for other purposes than the ones of the present study, but allowing us to perform the *in silico* extraction of the *rpoB* . All 4134 bp *rpo*B gene sequences were retrieved from each draft genome and were compared with that of the published *B. melitensis* 16M genome (Marianelli *et al*., 2006). *B.*

melitensis strains are classified in three *rpo* types (biovar 1, biovar 2 and biovar 3) according to the presence or absence of mutations in *rpo* gene targeting the specific codon residues 629, 985, 1249 and 1309. Basically, a strain was classified as phenotypically belong to biovar 1, if *rpoB* is 100% identical to that of the *B. melitensis* 16M genome. The presence of nucleotide substitutions GCG to GTG at codon 629, GCC to GTC at codon position 985 and CTG to CTA at codon position 1309 underlies the classification as biovar 2. The existence of the nucleotide substitution ATG to ATA at codon position 1249 leads to the classification as biovar 3 (Marianelli *et al*., 2006).

Statistical analysis was performed using descriptive analysis and associations were tested using Chi-squared test. It was considered a 5% significance level to reject the null hypothesis of the tests. Statistical analyses were computed using software R version 3.3.2.

As this study constitutes a retrospective analysis, some of its methodological limitations regard to the lack of information that would allow a more complete analysis of risk factors (such as occupation and residence area of the patients). Also, as the analyzed samples had been sent to the lab with the clinician request for a specific diagnostic method, we respected such request hampering the use of a single method for all samples.

2.4 Results

Between January 2009 and December 2016, 2313 biological samples from patients with clinical suspicion of brucellosis were analyzed by immunological techniques and 7.2% (167/2313) had positive serology for *Brucella* spp. The distribution of infection rate by year ranged from 5% (2012) to 10.7% (2009) for the 2313 samples analyzed by immunological methods, (Fig. 2.1 and S2.1 Table).

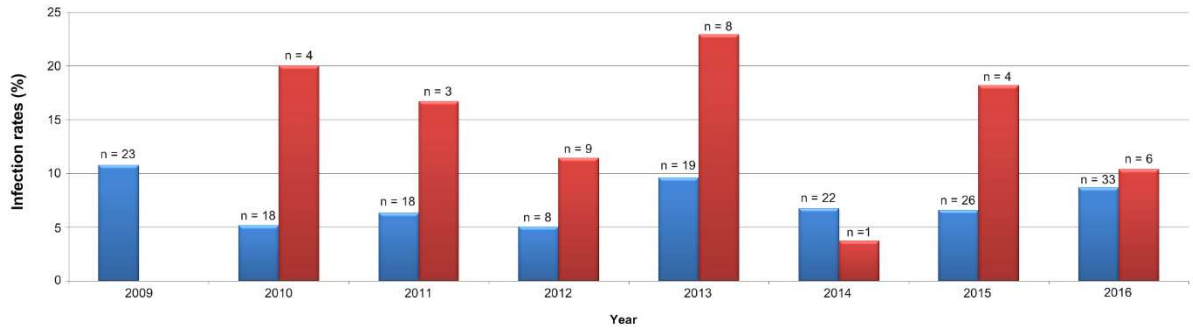


Figure. 2.1 Brucellosis infection rate between 2009 and 2016. Distribution of Brucellosis cases identified in the Portuguese National Institute of Health in the period between 2009 and 2016, by immunological techniques (blue bars) and molecular biology (red bars). The "n" above each bar corresponds to the number of positive samples. Infection rate per year was defined as the number of positive cases / total number of patients.

Of the 167 patients that yielded positive serology for *Brucella* spp, 61.7% (103/167) were male and 38.3% (64/167) were female (p=0.014) (data not shown). The age was known for 98.8% (165/167) of the cases, of which half (57.7%, 95/165) are between 26-65 years. Although the distribution by age

groups showed an irregular pattern, we found that 5.4% of the positive cases belong to children <5 years (Fig. 2.2 and S2.1 Table).

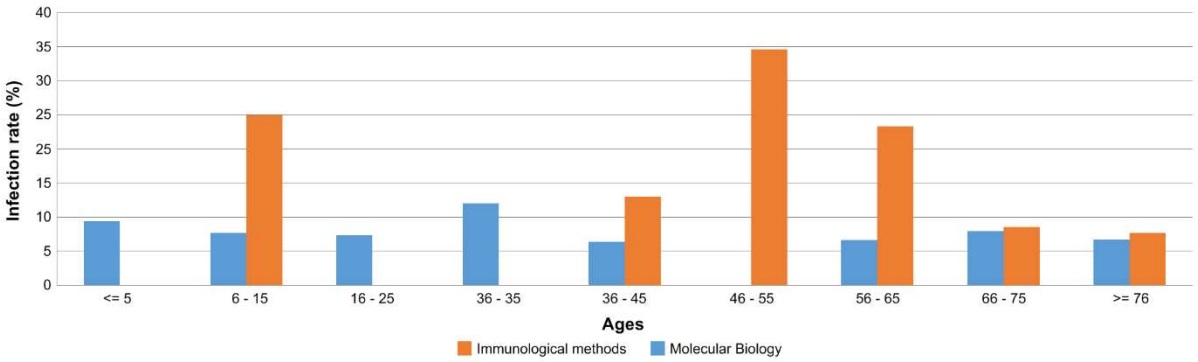


Figure.2.2 - Brucellosis infection rate by age groups. Distribution of Brucellosis infection rate by age groups performed by immunological techniques (blue bars) and molecular biology (red bars).

Between the years 2010 and 2016, 259 samples were tested by real time PCR methods, and 16.6% (43/259) were positive for *Brucella* spp., being *Brucella melitensis* the only species identified in the analyzed cases (Fig.2.1). The distribution of infection rate by year ranged from 3.7% (2014) to 22.9% (2013).

Concerning this subset of samples, a higher prevalence of positive samples for *Brucella* spp. was observed in males (p=0.007), similar to the scenario observed for the immunological methods. Regarding age distributions, in average, the age of the infected patients was 48.5 years (ranging between 6 and 91 years) (p<0.001) (Fig. 2.2).

The wild-type strains were classified by analyzing the *B. melitensis rpo*B types. The strain frequencies for these types were 14,3% for *rpo*B biovar 1 and 85,7%, for biovar 3. None of the strains belonged to biovar 2.

2.5 Discussion and Conclusions

In the present study, we intended to shed some light on the still unveiled prevalence scenario of human brucellosis in Portugal, by conducting a retrospective study on about 2700 samples received at the Portuguese NIH over a 7-year period. Overall, serological diagnostic identified 167 (7.2%) positive cases of human brucellosis, of which 61.7% were male and half of the cases were in the age groups between 26-65 years. Preview studies show that in industrialized countries the disease mainly affects men aged between 20 and 45 years, and suggests that the distribution by gender is connected to occupational factor (Corbel, 2006). In fact the people who work with farm animals, especially with cattle, sheep, goats and pigs (e.g., farmers, farm laborers, animal attendants, stockmen, shepherds, sheep shearers, goatherds, pig keepers, veterinarians and inseminators) are at risk through direct contact with infected animals or through exposure to a heavily contaminated environment. Although we found an

irregular pattern of distribution of brucellosis by age groups, the infection rates calculated by molecular techniques revealed that the age groups between 46 and 65 years old are among the ones with the highest rates. This falls within the range of the one described in the ECDC "Annual epidemiological report - Food-and waterborne diseases and zoonose in 2014", in the European Union (EU), reflecting a higher number of cases registered in the age group 45 to 64 years old (Annual epidemiological report, Food-and waterborne diseases and zoonoses. 2014). We also observed a low infection rate in children (5.4%), which is in agreement with data from the European Food Safety Authority, reporting that the vast majority (80%) of the European cases of brucellosis were adults over 25 years. This lower infection rate in children when compared with the one observed in adults likely relies on the low contact of children with the common infection sources, such as infected animals and animal products.

Although the molecular diagnostic was only applied to a subset of samples from 2010 to 2016, from the 259 analyzed samples, 43 (16.6%) were positive for *B. melitensis*. The higher infection rates obtained when using real time PCR when compared with immunological methods are likely due not only to a probable higher sensitivity of the former technique, but also because, according to our experience, PCR is usually requested when the clinician has a strong suspicion of brucellosis (e.g., patients revealing complications associated with the disease). The majority of the PCR positive cases belonged to biovar 3, pointing it as clearly the most common species/biovar involved in the human disease in Portugal. Like other countries, Portugal, applies specific regulations and measures to eradicate the disease, however, regardless of the huge efforts to eliminate it, brucellosis has continued to be an endemic disease where *B.melitensis* biovars 1 and 3 amd *B. abortus* biovars 1 and 3 are the prevailing animals species (Ferreira *et al*., 2013). This is in agreement with the data available for Europe. In fact, species information was provided for 99 of the 332 confirmed cases reported in the EU and Norway between 2008-2012, where 83.8% were reported to be *B. melitensis* (Annual epidemiological report, Food-and waterborne diseases and zoonoses. 2014). Although all clinical cases in Portugal were caused by *B. melitensis*, other *Brucella* species pathogenic to humans have been identified in animals, namely *B. abortus* and *B. suis* (Cristina *et al*., 2015). This underlines the importance to perform the early detection and identification at species' level of the *Brucella* strains obtained from clinical samples (human and animals), which is a critical information to prevent or control the occurrence of outbreaks. For this reason, molecular techniques, such as the real time PCR, particularly when applied to patients with compatible clinical symptoms and negative serological findings, are the most useful approach for laboratory diagnosis due to the rapid and precise identification of the *Brucella* sp. strain present in the clinical sample.

The geographical distribution of brucellosis is constantly changing, with the emergence and reemergence of new outbreaks around the world. Reflecting the social, cultural, and economic policies that describe a changing global society, this pathology has been reflecting this dynamics, making its control and eradication a constant challenge.

In conclusion, despite the control and prevention measures implemented by the national authorities, brucellosis remains a problem in Portugal, with impact in public health and in the economy. This study draws attention to the importance of integrating clinical and laboratory data of human cases in order to increase the efficacy of the response measures, essentially in case of outbreaks. Furthermore, our findings reinforce the need to maintain an active epidemiological surveillance, enabling the early detection of all cases of infection and underlie the need to have a good communication flow between the human and animal Health Ministries, according to the One Health concept, the only valid way to improve the assessment of the actual epidemiological situation of brucellosis and other zoonosis in Portugal.

CHAPTER III

GENOME-SCALE GENETIC RELATEDNESS AMONG BRUCELLA MELITENSIS STRAINS CAUSING HUMAN INFECTIONS IN PORTUGAL

The results presented in this chapter were previously submitted

Ana Pelerito, Alexandra Nunes, Sofia Núncio, João Paulo Gomes. Genome-scale genetic relatedness among *Brucella melitensis* strains causing human infections in Portugal. **Infection, Genetic and Evolution.**

Personal contribution

AP contributed to the design of the study, performed most of the experimental work, interpreted data and wrote the manuscript

3 Genome-scale genetic relatedness among *Brucella melitensis* strains causing human infections in Portugal

3.0Abstract

Brucellosis is an important zoonotic disease that affects both humans and animals. In Portugal, it is an endemic and notifiable disease although brucellosis cases are clearly underreported. To date, laboratory surveillance is based on the traditional MLVA-16 methodology and the associated epidemiological information is scarce. Our goal was to give a step forward in the *Brucella* spp. surveillance in Portugal through the implementation of a whole genome sequencing (WGS) approach. We created a curated species-specific wgMLST scheme enrolling a panel of 2656 targets and used it to perform a retrospective analysis of the genetic relatedness among *B. melitensis* strains causing human infection in Portugal from 2010 to 2018. The strains showed a phylogenetic clustering within genotype II (25 out of 36) and IV (4 out of 36), and shared clades with strains isolated from countries with which Portugal has privileged food trading, tourism and shares eating habits, such as Spain, Italy and Greece. Our results point to the identification of strong associations between B. melitensis strains, likely underlying missed "outbreaks" as 22 out of the 36 strains showed one or multiple genetic linkage with each other's. In fact, the applied gene-by-gene approach grouped these strains into six genetic clusters, each one enrolling putative epidemiological links. Nevertheless, more studies will be mandatory in order to define the appropriate range of cut-offs (probable non-static cut-offs) that best illustrate the association between genetic linkage and epidemiological information and may serve as alerts for the health authorities. Finally, this study constitutes a mark of technological transition for laboratorial surveillance of brucellosis in Portugal, and will unequivocally facilitate the assessment of ongoing and future outbreaks, in order to prevent the transmission spread.

3.1 Introduction

Brucellosis, a disease caused by *Brucella* spp., is one of the world's most widespread zoonoses, with estimated 500 000 new cases annually, and it is the leading cause of economic losses in the production of domestic ruminants (Pappas *et al*., 2006). The frequent sources of human infections are farm animals such as cattle, sheep, goats and pigs but *Brucella* can also infect marine mammals such as dolphins, porpoises, and seals. Humans can contract the disease by contact with infected animals or their products, with unpasteurized milk being the most common source of brucellosis in urban populations (Moreno, 2014). *Brucella* is a gram negative unsporulated and uncapsulated short bacillus that behaves as a facultative intracellular pathogen (Corbel and Brinley, 1984). The genus *Brucella* encloses 12

species, designated based on differences in pathogenicity and host preference such as *B. melitensis* (goats and sheeps), *B. abortus* (cattle and bison), and *B. suis* (swine, hares). *B. melitensis* is the most frequent agent of brucellosis in humans, and it leads to the most severe manifestations of the disease such as undulant fever, joint pain arthritis, endocarditis and meningitis (Whatmore *et al*., 2014; Scholz *et al*., 2010; Scholz *et al*., 2016). The *Brucella* genome contains two circular chromosomes of approximately 2.1 and 1.2Mb, and both share similar GC content, a similar proportion of coding regions and equivalent housekeeping gene distribution (Sankarasubramanian *et al*., 2017; DelVecchio *et al*., 2002). Also, three way genome comparisons of *B. suis*, *B.melitensis* and *B. abortus* sequences, revealed that the majority (>90%) of annotated genes shared 98-100% sequence identity and fewer than 100 genes were identified in only one or two of the three genomes (Ratusha *et al*., 2006). Although prophages and insertion sequences have been reported (Abou Zaki *et al*., 2017; Azam *et al*., 2016; Hammerl *et al*., 2016; Kaden *et al*., 2014), species from *Brucella* genus are considered monomorphic pathogens (Wattam *et al*., 2014).

Once limited genome diversity exists among different *Brucella* species, the analysis of full genome sequences of the different species (and biovars) is of crucial importance, not only to disclose the genetic basis of host preference and virulence differences (as these features must stem from the limited genome diversity), but also for molecular surveillance purposes. Nevertheless, whole genome comparisons and phylogenetic analysis of *Brucella* were only done on a limited scale. Efficient and reliable surveillance programs are essential for detection and control of outbreaks and largely depend on the timely collection and access to epidemiological data and the need of cooperation between different health sectors (*i.e*., human and veterinary) through the exchange of microbiological and associated metadata. In addition, complete epidemiological investigations rely on the availability of standardized and effective molecular typing methods and analysis tools that allow the public health laboratories to identify and trace an outbreak back to its source. Molecular epidemiological studies provide information about genetic grounds and origin of bacterial isolates, but such trace back studies in *Brucella* species can be challenging as they are generally quite conserved. With the technological advances and decreased cost of whole genome sequencing, new methods of pathogen typing, including gene-by-gene comparison using core genome multilocus sequence typing (cgMLST), as well as single-nucleotide polymorphism (SNP) calling based on a reference sequence analysis, are considered to be a suitable and more informative replacement of the gold standard typing schemes (Sankarasubramanian *et al*., 2017). Although the SNP-based analysis may constitute a better option for phylogenetic analyses of conserved genomes (because this approach covers the entire genome, including the intergenic regions (Georgi *et al*., 2017), very recently, efforts to develop cgMLST schemes for *Brucella* have been done (Sankarasubramanian *et al*., 2019; Janowicz *et al*., 2018). One of these schemes involves 2704 genes and is based on a pay-per-use platform (Janowicz *et al*., 2018), whereas the other involves a strikingly lower number of genes (n=164) for differentiating purposes (Sankarasubramanian *et al*., 2019).

In Portugal, human brucellosis is a reportable disease and is among the three most frequent zoonosis (DDO, 2015). This country has a herding tradition, with a high number of people keeping animals at little farmhouses and with a huge tradition of cheese production. Despite both Portuguese reference institutes for human and veterinary diagnosis of brucellosis use the MLVA-16 (i.e., Multiple-Locus Variable number tandem repeat Analysis based on 16 *loci*) methodology as a typing technique in epidemiological studies, there is a lack of communication between human and animal health authorities and the epidemiological link is rarely established. Considering this and the need for a technological transition for surveillance purposes, we developed a wgMLST schema to perform a retrospective analysis of the genetic relatedness among *B. melitensis* strains causing human infections in Portugal. Ultimately, we aimed at identifying potential transmission links that have been missed with the currently implemented surveillance system. This study was based on the collection of *B. melitensis* strains held by the reference laboratory for human Brucellosis at the Portuguese National Institute of Health, which receives all human isolates of *B. melitensis.*

3.2 Materials and Methods

3.2.1 Samples

This study enrolled, all *B. melitensis* strains that were sent to the reference laboratory for human Brucellosis at the Portuguese National Institute of Health during the last nine years, comprehending 37 isolates. Genotyping and demographic data are summarized in Table 3.1. For genomic comparative purposes, it also included 18 strains isolated in Spain, Germany, Hungary and Belgium, which were kindly provided to our lab and that were subjected to all laboratory procedures and analysis (described below). For bioinformatics analysis, all *B. melitensis* genome sequences available at NCBI until January 2019 (n=217) were also included.

All samples were handled in a BLS-3 biocontainment laboratory at the Portuguese National Institute of Health. *Brucella* isolates were cultured on blood agar for 3 to 5 days at 37º C under 5% CO2 and total DNA was extracted from fresh cultures on the NucliSens easyMAG platform (Biomerieux), according to the manufacturer's instructions. All isolates had previously been confirmed as *Brucella* spp. by real time PCR detecting the *Brucella* specific gene *IS711*, *BME* and *Brab* (Pelerito *et al*., 2017).

Table 3.1 - *Brucella melitensis* strains, data of origin, host and year.

Strain	Geographic Region	Host	Year
1P	Unknown	Human	2010
35P	Vila Real	Human	2012
36P	Vila Real	Human	2012
38P	Maia	Human	2012
40P	Vila Real	Human	2012
41P	Vila Real	Human	2012
43P	Vila Nova de Gaia	Human	2012
44P	Unknown	Human	2012
66P	Torres Novas	Human	2012
147P	Cabeceira de Basto	Human	2013

153P	Cabeceira de Basto	Human	2014
166P	Seixo de Ansiães	Human	2011
167P	Vila Franca de Xira	Human	2011
168P	Lourosa	Human	2011
169P	Unknown	Human	2014
177P	Unknown	Human	2014
179P	Baião	Human	2014
180P	Baião	Human	2014
184P	Baião	Human	2014
194P	Lisboa	Human	2015
198P	Lisboa	Human	2015
199P	Loures	Human	2015
200P	Pontinha	Human	2015
209P	Évora	Human	2015
228P	Caldas da Rainha	Human	2016
237P	Coimbra	Human	2016
258P	Vila Nova de Gaia	Human	2016
261P	Unknown	Human	2016
20Pa	Frei Rodrigo	Goat	2002
357Pa	Mafra	Sheep	2004
782Pa	Caldas da Rainha	Goat	2007
804Pa	Fundão	Bovine	2008
463Pa	Vila Viçosa	Sheep	2005
47Pa	Penamacor	Sheep	2001
770Pa	Vila do Conde	Sheep	2007
918Pa	Unknown	Goat	2011

3.2.2 Antimicrobial susceptibility

All isolates were tested for antibiotic resistance to rifampicin (RIF), doxycycline (DOX), streptomycin (STR), gentamicin (GEN), by E- test® (biomerieux, Portugal) according to Clinical and Laboratory Standards Institute (CLSI) guidelines for potential agents of bioterrorism. Briefly, a suspension of bacteria adjusted to 0.5 McFarland units was inoculated on Mueller – Hinton plates supplemented with 5% sheep blood and the gradient strips applied. The plates were incubated at 35 °C±2 °C with 5% CO2 for 48 h before reading. MIC values were interpreted in accordance with the CLSI guidelines (CLSI, The following breakpoints for susceptibilities were used: GEN≤4, STR≤16, DOX≤1. For RIF, CLSI interpretation of *Haemophilus infuenzae* (fastidious bacteria) was used: S≤1, I=2, R≥4. Quality control assays were performed with *Escherichia coli* ATCC #25922 and *Streptococcus pneumoniae* ATCC #49619.

3.2.3 Whole genome sequencing (WGS)

For WGS, high-quality DNA samples (quantified using Qubit, ThermoFisher) were subjected to dual-indexed Nextera XT Illumina library preparation, prior to cluster generation and paired-end sequencing (2×250bp) on a MiSeq Illumina platform (Illumina Inc.) available at the Portuguese NIH, according to the manufacturer's instructions). All genomes were *de novo* assembled using the INNUca v3.1 pipeline (https://github.com/B-UMMI/INNUca), which consists of several integrated modules for reads QA/QC, *de novo* assembly and post-assembly optimization steps. Briefly, after reads' quality analysis (FastQC v0.11.5 - http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and cleaning (Trimmomatic v0.36) (Bolger *et al*., 2014), genomes were assembled with SPAdes 3.10 (Bankevich *et*

al., 2012) and subsequently improved using Pilon v1.18 (Walker *et al*., 2014). Draft genome sizes, mean depth of coverage, number of contigs, and accession numbers are described in Supplementary Table S1.

3.2.4 Implementation of a wgMLST schema for *B. melitensis*

We created a wgMLST schema for *B. melitensis* with chewBBACA v2.0.11 suite (https://github.com/B-UMMI/chewBBACA) (Silva *et al*., 2018) (CreateSchema module; default settings), using all complete genomes of *B. melitensis* available at NCBI (until January 2019) and a training file generated by Prodigal v2.6.3 from the *B. melitensis* M16 reference genome (RefSeq Accession NC_003317 and NC_003318). To curate the schema, allele calling was performed on all complete genomes with default parameters using a BLAST Score Ratio (BSR) threshold of 0.6 in order to remove paralog *loci*. A cgMLST schema was also extracted and allele calling was performed for all genomes of *B. melitensis* available at NCBI until January 2019 (that include 60 complete and 157 draft genomes) as well as for the 55 assemblies (that include sequences of 37 PT strains) performed in our lab, in order to discard genomes yielding less than 95% of called loci. To validate the wgMLST schema, allele calling was performed for the remaining assemblies. The impact of genome quality on allele call was evaluated (Test Genome Quality module) using a maximum number of interactions (-n) of 13 and exclusion thresholds from 0 to a maximum (-t) of 300 with increasing -s values of 5. Considering that the number of present *loci* varied with the inclusion or exclusion of specific genomes, a threshold of 25 was used to select genomes that allow a good discriminatory power for the wgMLST schema creation. The quality of the *loci* panel composing the wgMLST have been assessed using the Schema Evaluation module with default parameters. Basically, loci with high length variability, and annotated as "non-informative paralogous hit (NIPH/NIPHEM)" or "Allele larger/Smaller than length mode (ALM/ASM)" by the chewBBACA Alelle Calling engine in more than 1% of the *B. melitensis* genomes were removed in order to curate the wgMLST schema. Finally, exact and inferred matches were used to construct an allelic profile matrix, where the other allelic classifications were assumed as "missing" *loci*.

3.2.5 Study of genetic relatedness among *B. melitensis* strains isolated in Portugal

Minimum spanning trees (MST) were constructed taking advantage of goeBURST algorithm (Francisco *et al*., 2009) implemented in the PHYLOViZ online web-based tool (Ribeiro-Gonçalves *et al*., 2016), based on 100% shared loci between all strains (i.e., shared-genome MLST). A hierarchical clustering tree were also generated using PHYLOViZ desktop 2.0 (http://www.phyloviz.net/) with distances among strains estimated with Hamming Distance metrics via the single-linkage method. In order to increase the resolution power for cluster analysis within the Portuguese strains, we used PHYLOViZ online 2.0 Beta version (http://online2.phyloviz.net/), which allows maximizing the shared genome in a dynamic manner, i.e., for each sub-set of strains under comparison, the maximum number of shared loci between them is automatically used for tree construction. All allelic distance thresholds used during cluster investigation were expressed as percentages of allele differences (AD), expressed as

the number of allelic differences over the total number of shared loci under comparison. To explore strain sub-sets among our 37 PT strains, a conservative step-by-step approach was performed by applying allelic distance cut-offs ranging from 1 to 0.1% to the initial MST, based on previously described data for cluster investigation in gene-by-gene based surveillance (Llarena *et al*., 2018).

3.3 Data availability

All raw sequence reads used in the present study were deposited in the European Nucleotide Archive under the run accession numbers ERR2938642-ERR2938706.

3.4 Results

All *Brucella* isolates were identified as *B. melitensis* by real time PCR. The obtained MIC values for all tested antibiotics are show in Supplementary Table S3.2. All isolates were susceptible to doxycycline, streptomycin and gentamicin. However, the MIC values for rifampicin ranged from 0.38-32 µg/ml, and according to CLSI breakpoints for slow-growing bacteria (*Haemophilus* sp.), reduced susceptibility (MIC 2-3 µg/ml) in five isolates and probable resistance (MICs≥4 µg/ml) in three strains were demonstrated (CLSI Guideline, 2016). We analyzed the mutational profile of *rpoB* to disclose the genetic basis of resistance to rifampicin but none of the identified SNPs have been linked to this phenotype.

3.4.1 wgMLST to evaluate *B. melitensis* phylogenetic diversity

By using the set of 272 *B. melitensis* genome sequences, we were able to generate a curated species-specific wgMLST scheme that enrolls a panel of 2656 targets (and 17472 alleles) based on the *B. melitensis* 16M reference genome (RefSeq Accession NC_003317 and NC_003318). This wgMLST schema was then applied to investigate phylogenetic relationships between genomes of the 36 PT strains (one was removed from the analysis due to bad quality), to put them in the frame of the worldwide phylogenetic scenario and to disclose potential epidemiological links.

In a first approach, we analyzed the phylogenetic position of PT *B. melitensis* strains in a global tree constructed with WGS data from strains collected worldwide (Supplementary Figure S3.1). As expected, phylogenetic analysis revealed spatial clustering, with five major genotypes being identified (Tan *et al*., 2015). While genotype I comprises strains from the Western Mediterranean Region and Egypt, the broader genotype II harbors strains from the Eastern Mediterranean Region and the Middle and Far East, and genotype III strains from the African continent. On the other hand, genotypes IV and V, which emerged from the same common ancestral derived from genotype III, are assigned to strains from Malta, Portugal and the American Continent. Curiously, despite a few strains cluster in genotype IV clade, the vast majority of the strains isolated in Portugal (25 out of 36) shows up in the clade of genotype II, in particular within sub-genotype IIi (Pisarenko *et al*., 2018). Considering the huge allelic

diversity exhibited by all 271 analyzed strains, a genetic relatedness cut-off of 3% was applied to the hierarchical clustering tree to evaluate potentially linked strain clusters, especially those enrolling the 36 PT strains (Supplementary Figure S3.1). While PT strains from genotype IV do not seem to present any apparent genetic relatedness with strains isolated in other geographic regions around the world, a different scenario was observed for genotype II. Indeed, the 25 PT strains assigned as belonging to genotype II seemed to exhibit a genetic proximity to strains isolated in Spain, Turkey and to two others isolates from Germany (corresponding to two imported cases with unknown origin).

3.4.2 Analysis of genetic relatedness among *B. melitensis* strains isolated in Portugal

In a second approach, a global MST was generated solely for all 36 PT strains (Figure 3.1). Based on the allelic diversity found among the 2191 shared *loci*, we were able to zoom-in the scenario of genotype classification described above. It can be observed that strains within each genotype display considerable fewer allelic differences (between one and 21 for genotype II and between two and 109 for genotype IV) than the ones obtained between genotypes or when compared with strains with unassigned genotype, where distances of more than 1000 allelic differences are observed. For comparative purposes, in parallel we also run the 36 PT strains with the freely available genus-specific cgMLST schema (Sankarasubramanian *et al*., 2019) that uses 164 loci (comprising about ~6% of the *loci* panel used in this study). However, although similar genotype associations were achieved, it revealed less strain discriminatory power, especially within genotype II, with several unrelated strains clustering together (Supplementary Figure S3.2).

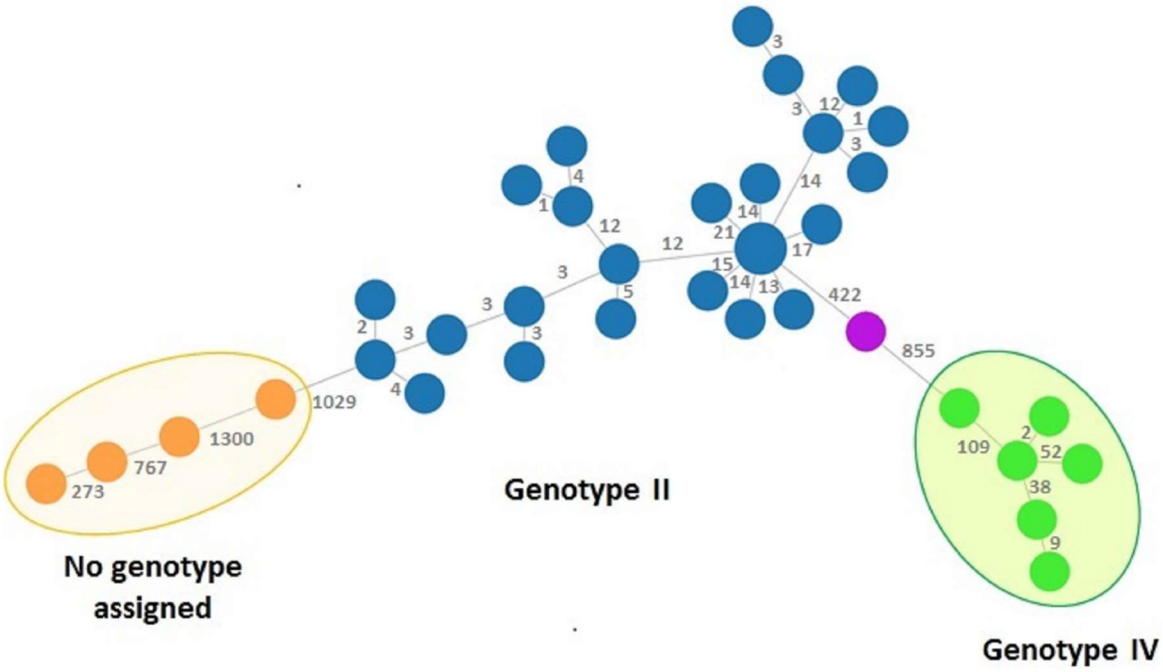


Figure 3.1 – Phylogeny of PT *B. melitensis* strains based on a dynamic gene-by-gene approach using a wgMLST schema with 2656 loci. The Minimum spanning tree (MST) was constructed using the goeBURST algorithm implemented in the PHYLOViZ Online platform, and is based on the allelic diversity found among the 2191 genes shared by 100% of the 36 PT strains. Filled small circles (nodes) represent unique allelic profiles, and are colored based on the assigned genotype according to Tan *et al* (Tan *et al*, 2015).The numbers in grey on the connecting lines represent the allele differences (AD) between strains.

To explore strain sub-sets among our 36PT strains, two additional MST were generated, one for each genotype. Considering that in *Brucella* spp. there is no defined threshold to identify clusters of genetically related strains with high epidemiology congruence, a conservative step-by-step approach was performed by applying allelic distance cut-offs ranging from 1 to 0.1% to the initial MSTs generated for (i) all 36 (ii) genotype II and (iii) genotype IV strains. We firstly selected a threshold of 0.4% (that corresponds to ≤11AD) since it allowed to maximize the number of strain sub-sets identified within each genotype (Figure 3.2). Indeed, after the application of this cut-off to both genotype MSTs, we were able to highlight six genetically related sub-sets of strains, which may theoretically harbor a higher probability to have an epidemiological link. In particular, genotype II strains exhibiting ≤10AD were kept interconnected in four clusters, and strains from genotype IV with ≤11 AD resulted into two potential related clusters (Figure 3.3A). Next, for each identified cluster, a sub-MST was generated in order to maximize the number of shared loci among the strain sub-set (Figure 3.3B), and consequently, to better evaluate the relatedness of strains.

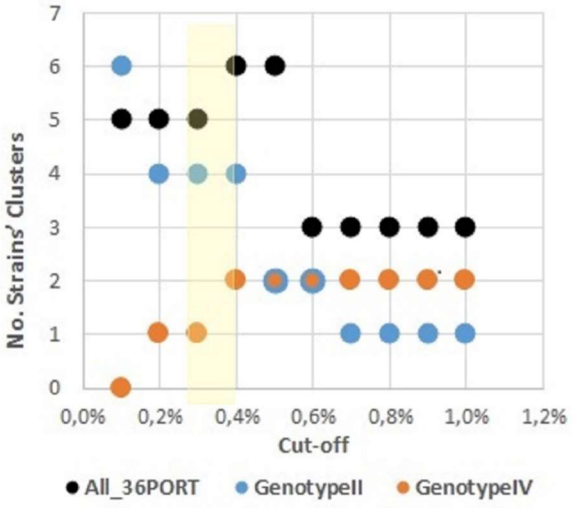


Figure 3.2 – Impact of different allelic distance thresholds on the definition of *B. melitensis* strains' clusters. The number of clusters are shown for allelic distance cut-offs ranging from 0.1 to 1%. This analysis was done both by genotype and by using all strains. For the present dataset of 36PT strains, the cut-off that maximize the number of clusters identified within each cluster is highlighted in yellow.

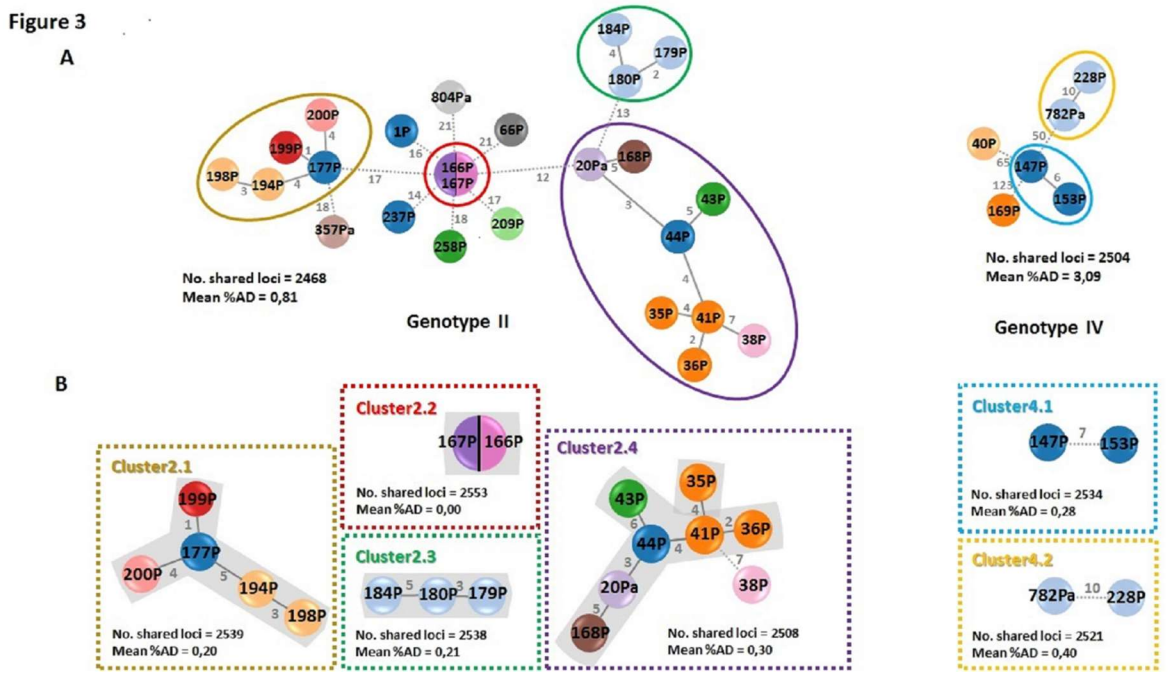


Figure 3.3 – Phylogenetic relationship of PT *B. melitensis* strains by genotype based on a dynamic gene-by-gene approach using a wgMLST schema with 2656 loci. (A) For each genotype identified, the initial MST was constructed based on the allelic diversity found among the shared genes between strains (indicated near each tree). Potential clusters defined for fine-tune analysis are surrounded by colored circles and further detailed in panel B. (B) Sub-MST reconstruction based on the maximum number of shared *loci* (indicated near each tree) between strains forming a putative cluster. For both panels, trees were constructed using the goeBURST algorithm implemented in the PHYLOViZ Online platform. Each filled small circle (node) contains the strain's designation and represents a unique allelic profile. Nodes are colored according to the geographic region where strains were isolated (see Table 1 for details). The numbers in grey on the connecting lines represent the allele differences (AD) between strains. Within each cluster, strains exhibiting strong genetic relatedness are highlighted in grey and connected by solid lines, while strains with borderline genetic relatedness are connected by dashed lines.

Regarding both clusters of genotype IV, despite the inexistence of metadata, we cannot discard a possible epidemiological link between the enrolled strains. Indeed, while for cluster 4.1, strains were isolated from patients of the same northern village with two months of difference (which may be associated with the incubation period of the infection), both strains from cluster 4.2 were collected at a city in the center of Portugal. The later were isolated with a five year distance period (2011 and 2016), where one of them was isolated from a goat whereas the other caused a human infection. Nevertheless, considering the high genetic relatedness and the same isolation local, one cannot discard an epidemiological link. Moreover, a low allelic diversity from two isolations of the same strain five years apart would be congruent with an low evolutionary rate of *Brucella* spp.. For genotype II, strains within the cluster 2.3 are from a confirmed outbreak occurred in 2014 in a small northern region, due to consumption of raw cheese sold in local market (Figure 3B). This outbreak was controlled and it was possible to identify the infected animals (goats) as the source of the infection. No strain was isolated at that time so no genome analysis can be performed. In an opposite scenario, despite both strains from cluster 2.2 are genetically identical (among the loci panel analyzed), they were isolated from patients at geographically distant regions from north and center of Portugal (~300km apart) in 2011, suggesting that their possible linkage may be related with the ingestion of a product of animal origin. Regarding

cluster 2.1, *B. melitensis* strains were isolated from patients within the Lisbon area between 2014 and 2015, but a possible link was never confirm. Finally, for cluster 2.4, the largest cluster analyzed, all strains but 44P (for which no information is available) were isolated in the north of the country, where three of them (35P, 36P, 41P) were from the same city and other three (43P, 168P and 38P) were from neighboring small cities (Supplementary Figure S3). With exception of 20Pa, which was collected from an animal (goat) in 2002, all strains but 168P (isolated in 2011) were from cases of human brucellosis occurred at 2012. No epidemiological information was available, hampering the determination of the potential infection source.

3.5 Discussion

Brucellosis is a zoonosis that is emerging in some regions of the world; in Portugal, it is an endemic and notifiable disease. Although human cases have been reported throughout the country, it is recognized by the Portuguese Health Authorities that brucellosis cases are clearly underreported, which does not allow consistent analysis of risk factors and the proper evaluation of the impact of this disease on public health. Also, the frequent lack of metadata associated with the isolated strains constitutes a hurdle to the epidemiological research, frequently hampering the identification of the infectious source.

The reference laboratory for human Brucellosis at the Portuguese National Institute of Health receives from the hospital laboratories all human isolates of *B. melitensis*, which are typed by MLVA - 16 methodology. Aiming at giving a step forward in the *Brucella* spp. surveillance in Portugal, we created a curated species-specific wgMLST scheme enrolling a panel of 2656 targets (and 17472 alleles) to perform a retrospective analysis of the genetic relatedness among *B. melitensis* strains causing human infection in Portugal. Although two cgMLST schemas were recently developed for *Brucella*, one of them is based on a very small panel of 164 *loci* (Sankarasubramanian *et al*., 2019), and the other runs on a pay-per-use platform (RIDOM SeqSphere) despite enrolling a wider core gene set of 2704 targets (Janowicz A *et al*., 2018).

According to Tan's classification (Tan *et al*., 2015), the isolates from the Portuguese dataset (isolated between 2010 and 2018) essentially clustered in two previously described lineages, namely the East Mediterranean (EM) clade (genotype II) and the Malta and Portugal clade (genotype IV), with few strains falling in unclassified clades (Supplementary Figure S1). The majority of the isolates (25 out of the 36 PT strains) clustered in the Genotype II, in particular within sub-genotype Iii (Pisarenko *et al*., 2018), and seem to reveal a genetic proximity to strains isolated in Spain and Turkey (the two close isolates from Germany are imported cases with unknown origin). Such relatedness with strains from Spain is not surprising considering the border free herding, the common traffic of alimentary products among these countries as well as the tourism and the free circulation of the population. Other six isolates from the present study clustered in the genotype IV, which correlates well with the extremely common circulation of people between countries where brucellosis is endemic, such as Portugal, Italy and Greece,

also sharing similar heating habits. Therefore, the influx of migrations among European countries comes along with raised case counts of an infectious disease.

The implementation of the wgMLST approach allowed us to identify six clusters, where two clusters enroll strains from the genotype IV and four clusters enroll strains from the genotype II. Despite the absence of complete epidemiological information for most of the cases, our results point to the identification of strong associations between some of them, likely underlying missed "outbreaks". In fact, for instance regarding genotype II strains, in the cluster 2.2 two human isolates are genetically "identical", although they were isolated in different geographical locations. This likely discards the contact with an infected animal but suggests a food origin (e.g., cheese) as the highly likely infectious source. For the cluster 2.3 the genetic analysis shows the correlation with epidemiological data, confirming the outbreak occurred in 2014 in a small northern region that had been identified solely based on epidemiological information. Considering the unequivocal close genetic relatedness among all strains from the cluster 2.4, their geographic proximity and the existence of an animal contaminated with likely the "same clone", ten years before these cases of human brucellosis, we can speculate that this clone is endemic in that region for a long time. Thus, it is reasonable to assume that these cases of human brucellosis are likely derived from the consumption of products from contaminated animals of that specific region.

A tricky issue underlying the application of gene-by-gene approaches, such as the wgMLST reported here, concerns the choice of cutoffs to identify putative genetic linkages and this challenge extends to all microorganisms for which genome-scale approaches are being created. For instance, choosing cutoffs that enable zooming-in specific clades of a MST (i.e., enabling a more precise evaluation of the genetic relatedness among the already "most related" strains) modifies its sensitivity, making the exclusion of putative outliers more robust, but may also exclude from the cluster strains with slightly higher genetic differences but with known epi-link. For instance, by applying a threshold of 0.2% (corresponding to ≤6AD) to each sub-MST (Figure 3.3B), we were able to consolidate the strong strains' genetic link within clusters of genotype II, but placed the strain 38P as borderline in cluster 2.4. As no epidemiological information is available for 38P, we cannot assess the accuracy of the chosen threshold for this cluster. Thus, considering that gene-by-gene approaches for WGS-based surveillance of *B. melitensis* are still at the beginning, the choice for the appropriate cut-offs for cluster definition should be a dynamic process and should always be associated with the existing epidemiological data. On this regard, future studies with large datasets and strong epidemiological data will certainly ensure this achievement.

In conclusion, the application of a WGS-based approach for a retrospective evaluation of the genetic relatedness of all *B. melitensis* strains received at the Portuguese reference laboratory between 2010 and 2018 allowed the identification of several highly probable associated cases of brucellosis, where 22 out of the 36 PT strains showed one or multiple genetic linkage with other strains. The implementation of a wgMLST scheme in the reference laboratory constitutes a mark of technological

transition for laboratorial surveillance of brucellosis in this country, and will unequivocally facilitate the assessment of ongoing and future outbreaks in order to prevent the transmission spread. It will allow a better understanding of the epidemiology and dynamics of *Brucella* spp. populations and to gather in depth information, which can be used for source tracing in case of outbreaks within animal holdings, zoonotic or foodborne infections.

CHAPTER IV

EVALUATION OF AN IN SILICO APPROACH FOR MULTIPLE LOCUS VARIABLE NUMBER TANDEM REPEAT ANALYSIS FOR GENETIC CHARACTERIZATION OF BRUCELLA SPP

The results presented in this chapter were previously published

Ana Pelerito[1], Alexandra Nunes[2], Joana Isidro[2], Catarina Silva[3], Ferreira AC[4], Valdezate S[5], Sofia Núncio[1], Enrico Georgi[6], João Paulo Gomes[2*]. Evaluation of an *in silico* approach for Multiple Locus Variable Number Tandem Repeat Analysis for genetic characterization of *Brucella* spp. Scientific Report.

Personal contribution

AP contributed to the design of the study, performed most of the experimental work, interpreted data and wrote the manuscript.

4 Evaluation of an in silico approach for Multiple Locus Variable Number Tandem Repeat Analysis for genetic characterization of *Brucella* spp.

4.0 Abstract

Brucellosis is an important zoonosis that is emerging in some regions of the world, gaining increased relevance with the inclusion of the causing agent *Brucella* spp in the class B bioterrorism group. Until now, Multi-locus VNTR Analysis (MLVA) based on 16 loci has been considered the gold standard for *Brucella* typing. However, although this methodology is laborious, frequently reveals amplification failures and is error-prone in the allele identification. With the rampant release of *Brucella* genomes, the transition from the traditional MLVA to whole-genome sequencing-based typing is inevitable. Nevertheless, in order to avoid a disruptive transition with the loss of massive genetic data obtained throughout decades, it is important to be able to determine *in silico* the MLVA alleles of the nowadays sequenced genomes. On this regard, we aim to evaluate the performance of a Python script that had been previously developed for the rapid *in silico* extraction of the MLVA alleles, by comparing it to the wet-lab MLVA procedure over of 83 strains from different *Brucella* species. The *in silico* approach detected 95.1% of all possible 1328 hits (83 strains x 16 loci) and showed an agreement rate with the wet-lab procedure of up to 84.1%, where major discrepancies are likely due to erroneous interpretations of the gels' DNA patterns underlying the latter. According to our dataset we suggest the use of a minimal depth of coverage of ~50x and a maximum number of ~200 contigs as guiding "boundaries" for the future application of the script.

In conclusion, the evaluated script seems to be a very useful and robust tool for *in silico* extraction of MLVA types of *Brucella* strains, allowing retrospective and prospective molecular epidemiological studies, which are important for maintaining an active epidemiological surveillance of brucellosis.

4.1 Introduction

Brucellosis is one of the world's most important zoonotic diseases causing great damage to husbandry industry and public health (Franc *et al*., 2018). The brucellosis burden specifically on low-income countries has led the World health Organization (WHO) to classify it as one of the world's leading neglected zoonotic diseases (http://www.who.int/neglected_diseases/zoonoses/other_NZDs/en/). However, given the absence of specific signs and symptoms, the disease is commonly under diagnosed (Valdezate *et al*., 2010).

Brucellosis is transmitted to humans through consumption of unpasteurized dairy products or through direct contact with infected animals, placentas or aborted fetuses (Young, 2005). This bacterial disease causes a severely debilitating and disabling illness, with fever, sweating, fatigue, weight loss,

headache, and joint pain persisting for weeks to months. Neurological complications, endocarditis and testicular or bone abscess formation can also occur (Dean *et al*., 2012).

A renewed scientific interest in human brucellosis has been fueled by its recent re-emergence and enhanced surveillance in many areas of the world, and from the inclusion of the causing agent *Brucella* spp. in the group of class B bioterrorism agent (Franco *et al*., 2007). A low infectious dose of 10 to 100 organisms is sufficient to cause an infection and the mechanisms of transmission, through aerosols or food chains, make them easily transmissible to both humans and animals (Tan *et al*., 2015). Therefore, the discrimination between natural outbreaks and/or intentional release of microorganisms may be of crucial importance in the context of the bioterrorism.

Brucella species are characterized by >80% interspecies homology through DNA-DNA hybridization studies and >98% sequence similarity by comparative genomics (Whatmore *et al*., 2006; Kattar *et al*., 2008). In fact, the sequencing of 16S rRNA gene showed a 100% identity between all of the *Brucella* spp. (Georgi *et al*., 2017). Human brucellosis can be caused by various *Brucella* species. The genus currently comprises 12 validly published species, which are genetically highly related to each other, but *Brucella melitensis* is by far the most frequently observed causative agent of human infection (Young, 2005; Georgi *et al*., 2017). On this regard, the knowledge on current major *Brucella* species, biovar and genotype, and their geographic distribution is of great value, especially for tracking back infectious sources and monitoring transmission routes (Pisarenko *et al*, 2018). The simple identification of genus and, in some cases, species by PCR assays, is adequate for purposes of diagnosis of human/animal disease or identification of food contamination but not for the tracing of outbreaks or bioterrorism attacks (Santis *et al*., 2011).

To fulfill the objective of sub-species discrimination, Variable Number Tandem Repeats (VNTR) have been investigated in Multi-locus VNTR Analysis (MLVA) by various scientific groups since 2003. This *Brucella* typing scheme, using 16 VNTRs, has been proven to have the ability to differentiate *Brucella* species, biovar and even the isolates. More importantly, there is an online database of MLVA-16 profiles available to all researchers allowing the comparison of *Brucella* strains at the worldwide scale (Le Fleche *et al*., 2006; Sun *et al*., 2015; Mambres *et al*., 2017). MLVA has become a major molecular typing method to characterize several pathogenic bacterial species, however, this methodology is laborious, time consuming and frequently the amplification of all loci cannot be achieved. Recent implementation of whole-genome single nucleotide polymorphism (SNP)-based typing has led to substantial improvements of both molecular subtyping and phylogenetic analyses in microbiology. The development of core- and whole-genome multilocus sequence typing schemes has been focused on a restrict number of bacterial pathogens, including *Brucella* spp but their application may be tricky (Tan *et al*., 2015; Janowicz *et al*., 2018; Sankarasubramanian *et al*., 2019). In fact, the creation of universal intra or inter species schemes needs to overcome some genetic hurdles such as the existence of paralogous genes, annotation issues, the accessory genome, and nomenclature-associated difficulties. Meanwhile, until whole-genome data is fully established and accepted by the scientific

community for classification/typing purposes in *Brucella*, the *in silico* extraction of the MLVA schemes can be of extreme utility. In fact, not only it overcomes the tremendously laborious laboratory-based MLVA assessment, but it also allows the dynamic cross-comparison with the typing-associated genetic data determined during the last decades. On this regard, a Python script has been recently developed focusing on the *in silico* extraction of *Brucella* MLVA schemes taking advantage of the increasing number of sequenced genomes (Georgi *et al*., 2017). Nevertheless, as no experimental validation of such script was performed, we now aim to evaluate the agreement between experimental (*i.e*., wet lab based) and the developed *in silico* determination of MLVA for strains comprising several *Brucella* species in order to check the validity of such technological transition underlying the genetic characterization of *Brucella*.

4.2 Materials and Methods

4.2.1 Samples

Eighty-three *Brucella* isolates isolated in Portugal, Spain, Germany, Hungary and Belgium (Supplementary Table 4.1) were used in this study. *B. melitensis* strain 16M strain (NC_003317 and NC_003318) was used as reference strain.

All samples were handled in a BLS-3 biocontainment laboratory at the Portuguese National Institute of Health. *Brucella* isolates were cultured on blood agar for 3 to 5 days at 37º C under 5% CO2 and total DNA was extracted from fresh cultures on the NucliSens easyMAG platform (Biomerieux), according to the manufacturer's instructions.

All strains were identified as *Brucella* species by real time PCR, using a previously published assay (Pelerito *et al*., 2017). The molecular methods to identify the infection were performed in a tandem fashion. First, an "in house" real time PCR using hydrolysis probes was used to detect and identify the species of *Brucella* genus. Secondly, for species differentiation, primers and Taqman probes were designed within the BMEII0466 gene for *B. melitensis* and BruAb2_0168 gene for *B. abortus* ((Pelerito *et al*., 2017; Gopaul *et al*., 2008).

4.2.2 MLVA-16 Assay

Single locus amplification of the eight minisatelite loci (panel 1) and eight microsatelite loci (panels 2A and 2B), that constitute the MLVA – 16 assay, was performed as describe by Fletcher *et al*(le Fleche, *et al*., 2006).

PCR reactions were performed in a total volume of 15µl containing 3ng of DNA, 1X PCR reaction buffer, 1U of Taq DNA polymerase (Bioline), 200µM of each dNTP's and 0.3µM of each flanking primers. An initial denaturation step at 96ºC for 5min was followed by 30 cycles of denaturation at 96ºC for 30s, primer annealing at 60ºC for 30s and elongation at 70ºC for 1min. The final extension step was performed at 70ºC for 5min. Amplification products were analysed by 2% (panel 1) and 3%

(panel 2) agarose gel electrophoresis with molecular size markers suitable for the identification of DNA bands ranging from 79 to 914. The total number of repeats at each locus was determined by the correlation with the amplicon size according to the 2013 *Brucella* allele assignment table (version 3.6 available at http://mlva.u-psud,fr) (Le Fleche *et al*., 2006). The reference strain *B. melitensis* 16 M, for which the expected size is known for each VNTR locus, was used as control for alleles assignment.

4.2.3 Whole genome sequencing (WGS)

For each strain, WGS was performed as previously described (Pinto, *et al*., 2018). Briefly, quantification and quality assessment of the purified DNA was performed using the DNA HS Assay Kit (Thermo Fisher Scientific) in the Qubit Fluorometer and agarose gel electrophoresis (0,8%), respectively. High-quality DNA samples were then used to prepare dual-indexed Nextera XT Illumina libraries that were subsequently subjected to cluster generation and paired-end sequencing (2×250bp and 2x300bp) on a MiSeq Illumina platform (Illumina Inc.), according to the manufacturer's instructions.

Reads quality control and bacterial *de novo* assembly were performed using the INNUca v3.1 pipeline (https://github.com/B-UMMI/INNUca), which consists of several integrated modules for reads QA/QC, *de novo* assembly and post-assembly optimization steps. Briefly, after reads' quality analysis (FastQC v0.11.5 - http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and cleaning (Trimmomatic v0.36), genomes were assembled with SPAdes 3.10 (Bankevich *et al*., 2012) and subsequently improved using Pilon v1.18 (Walker *et al*., 2014), with genome coverage being monitored and reported after each processes. In order to evaluate the impact of the "post-assembly polishment" on the assembled genomes and subsequently on the *in silico* MLVA analyses, the SPAdes assemblies were also performed skipping the Pilon step. A final check was also performed. Considering that the *in silico* extraction of the MLVA loci may be influenced by the quality of the assembled genomes another largely used *de novo* assembler, Velvet (Zerbino and Birney, 2008) was applied through VelvetOptimiser v.2.2.5 (https://github.com/tseemann/VelvetOptimiser), for comparative purposes, with and without Pilon. The VelvetOptimiser script was run using trimmed reads for odd k-mer values ranging from 31 to 127 (highest k-mer used in SPAdes), with all program default settings unchanged apart from the minimum output contig size, which was the same as used by SPAdes.

4.2.4 *In silico* MLVA

Bacterial draft genomes were subjected to a Python script for *in silico* extraction of *Brucella* MLVA scheme (with 16 loci) as previously described (Georgi *et al*., 2017). As determining numbers of repeated stretches from WGS data may be error-prone, we carefully checked each locus in respect to the expected total length, internal repeat homogeneity or probability to get collapsed VNTRs during the assembly. All resulting MLVA-16 genotypes were compared to a public database with 2,215 entries of

B. melitensis strains that can be assessed online (http://microbesgenotyping.i2bc.paris-saclay.fr/ (Grissa *et al*., 2008).

4.2.5 Evaluation of agreement between *in silico* and experimental MLVA

To access the accuracy of the *in silico* MLVA approach, we determined the percentage of agreement between experimental genotyping and *in silico* MLVA by calculating the number of identical results (*i.e*., identical called alleles), divided by the total number of hits that were detected simultaneously by both approaches.

Taking into account the well-known error-prone determination of the correct allele by the laboratory approach when some alleles differ by a single repetition (e.g., ~10bp, hardly distinguishable on a gel) we also calculated an adjusted agreement. On this regard, whenever a discordant allele assignment involved a single repetition this was considered a highly-likely matching result.

Finally, for all strains, the performance of the bioinformatic script in extracting all 16 MLVA loci was also evaluated by taking into account the quality of the draft genome generated by the two assemblers (SPAdes and VelvetOptimiser) with and without "post-assembly polishment". Basically, for each condition, both the mean coverage depth and the number of contigs of each draft genome were correlated with the number of extracted loci. Pearson's coefficients (*r*) were measured to see potential linear associations. Nevertheless, as these final evaluations were done as complements of the major strategy, for the sake of clarity, whenever the text refers "*in silico* approach" it refers to the approach that used SPAdes with Pilon.

4.2.6 Data availability

All raw sequence reads used in the present study were deposited in the European Nucleotide Archive under the run accession numbers ERR2938642-ERR2938706 and ERR2993131-ERR2993163 (detailed in Supplementary Table S4.1).

4.3 Results

Our first approach was to perform the experimental MLVA–16 (panel 1, 2A and 2B) on the 83 strains in order to use the obtained data as the basis for comparative purposes with the *in silico* approach. As expected, it was neither possible to amplify all loci in all samples, nor to successfully perform the full *in silico* extraction of the 1328 hits (83 strains x 16 loci) (Fig. 4.1.A). We detected 1260 (94.9%) hits by experimental MLVA, where the alleles for the complete set of the MLVA-16 loci were detected

in 52/83 (62.7%) strains. One, two and ≥3 loci yielded no results for 16/83 (19.3%), 7/83 (8.4%), and 8/83 (9.6%) strains, respectively.

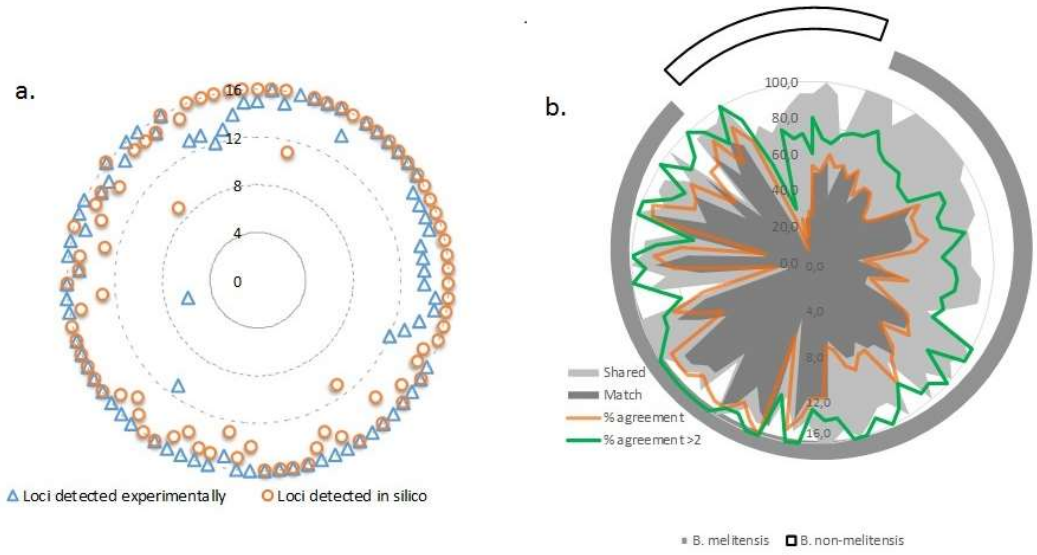


Figure 4.1 – Efficiency of the experimental and *in silico* MLVA approaches in strains' genotyping. Panel A shows the number of MLVA-16 loci detected *per* strain. *In silico* results are relative to the MLVA loci extraction using draft genomes assembled with the INNUca v3.1 pipeline using Pilon. Panel B displays the number of detected loci shared between the two methodologies for each strain as well as those that possess identical alleles. Also, the percentage of agreement between both approaches is shown for each strain. A correction in the number of matching alleles was performed whenever a discordant allele assignment involved a single repetition (see methods for details). The outer circle differentiates the *B. melitensis* strains (dark grey) from the non *B. melitenis* strains (white).

Regarding the *in silico* approach, the alleles for the complete set of 16 loci were detected in 50/83 (60.3%) strains, whereas one, two and ≥3 loci yielded no results for 20/83 (24.1%), 7/83 (8.4%), and 6/83 (7.2%) strains, respectively. Globally, we detected 1263 hits in all 83 strains (95.1%) (Fig. 4.1.A), meaning a similar success rate (simply measured as the number of the alleles called) between the two approaches. However, as shown in panel B of Fig. 4.1, considerable discrepancies were detected, namely when we compared both the number of loci shared between the two methodologies for each strain and those that possess identical alleles. The average of agreement was as low as 68.1% but increased up to 84.2% when the above mentioned adjustment was done (see methods for details). Also, 15 out of the 83 strains belonged to *Brucella* species other than *B. melitensis* and it can be observed that the number of the discrepancies was slightly higher within this group (Fig.4.1.B). Such variations are likely due to the traditional use of *B. melitensis* 16M reference strain as a standard in the experimental MLVA, even when strains from other species are being analysed, leading to some probable erroneous interpretations of the gels' DNA patterns.

When the analysis is performed *per locus*, it can be seen that the discrepancies are more frequent in *loci* that belong to the panel 2A and 2B of the MLVA-16 (Fig. 4.2), which are known to have a higher discriminatory power than the ones of panel 1.

Among all 83 strains, two specific loci (Bruce07 and Bruce21) appear as the most problematic ones in the experimental approach, with amplification failing in 16/83 (19.3%) and 12/83 (14.5%) strains, respectively. Considering a high overlap of strains for which both loci failed, we hypothesized that the amplification failures could rely on particular sequence characteristics of those loci, such as a GC-content higher than the one observed for the remaining loci of the MLVA scheme. However, that presumption was not verified and so we have no reasonable explanation for this. Bruce07 was also the locus exhibiting the highest number of mismatches through the *in silico* MLVA analysis, with a better performance being observed for Bruce21 (with *in silico* extraction only failing in 5/83 (6.02%) strains).

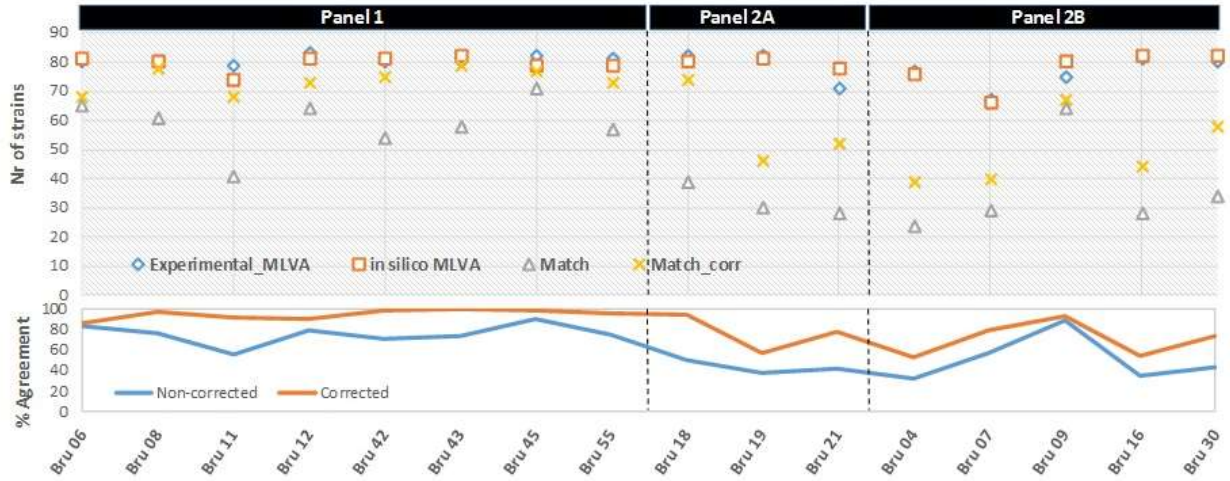


Figure 4.2 – Performance of the experimental and *in silico* MLVA approaches *per* locus. The upper graph represents the number of strains for which it was possible to determine an allele *per* locus as well as the number of strains sharing identical alleles between the two methodologies. The lower graph shows the percentage of agreement *per* locus between both approaches. In both graphs, an adjustment in the number of matching alleles was performed whenever a discordant allele assignment involved a single repetition (see methods for details). For both graphs, the loci are grouped according to the MLVA-16 panel they belong to (i.e., Panel 1, Panel 2A and Panel 2B). *In silico* results are relative to the MLVA loci extraction using draft genomes assembled with the INNUca v3.1 pipeline using Pilon.

Considering the heterogeneous composition of the three loci panels, it was not surprising that the agreement rate between the experimental and the *in silico* approaches was also dependent on the panel under consideration. In fact, considering, for instance, the adjusted values, whereas the mean agreement rate was 84.1% for all 16 loci, it ranged from 73.3% for loci of panel 2 (A and B together) to 95.0% for loci of panel 1.

Taking into account that the Python script for *in silico* extraction of *Brucella* MLVA schemes is applied after the genome assembly, we also inspected the quality of the draft sequences used as input. The influence of the mean depth of coverage and number of contigs on the efficacy of the bioinformatics script is illustrated in Fig.4.3. As expected, a negative linear correlation was observed among the efficacy of the *in silico* MLVA extraction and the number of assembled contigs, with less partitioned

genomes allowing the detection of a higher number of alleles. On the other hand, higher genome mean coverage depth seem to favour the *in silico* extraction of MLVA loci.

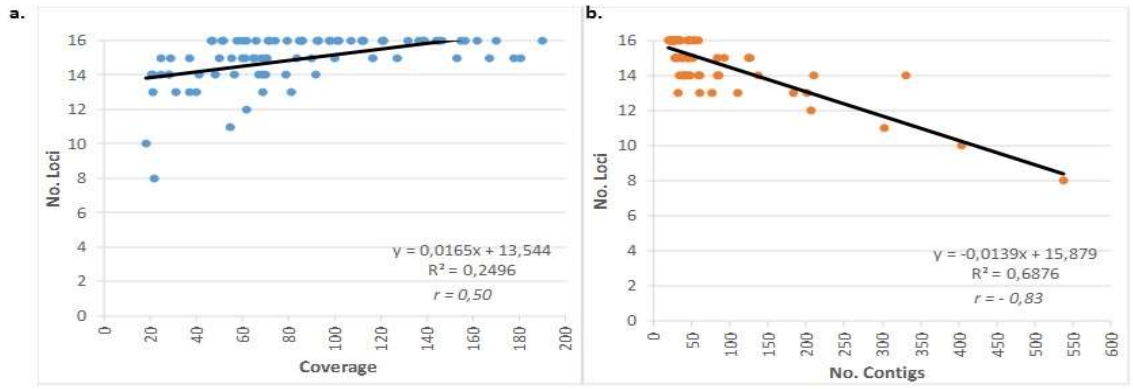


Figure 4.3 - Influence of the mean coverage depth and number of contigs on the efficacy of the *in silico* MLVA extraction. The graphs show the correlation of the efficacy (measured by the number of loci for which an allele was called) of the bioinformatics script with the depth of coverage (left panel) after quality improvement, and with the number of assembled contigs (right panel). For better visualization purposes, one strain exhibiting 3824 contigs was excluded from the graph but not from the calculations. The tendency lines are also shown with the respective equations as well as the Pearson coefficient (*r*). *In silico* results are relative to the MLVA loci extraction using draft genomes assembled with the INNUca v3.1 pipeline using Pilon.

As a final assessment, the performance of the bioinformatics script was also evaluated by using as input, draft genome sequences assembled with a different assembler. Curiously, although no significant differences were observed regarding the number of loci extracted both with and without "post-assembly polishment" (data not shown), for VelvetOptimiser assemblies, the number of detected loci show an unequivocal higher drop for mean depth of coverage <50 as well as for high fragmented genomes than the one obtained for SPAdes assemblies (Figure 4.4).

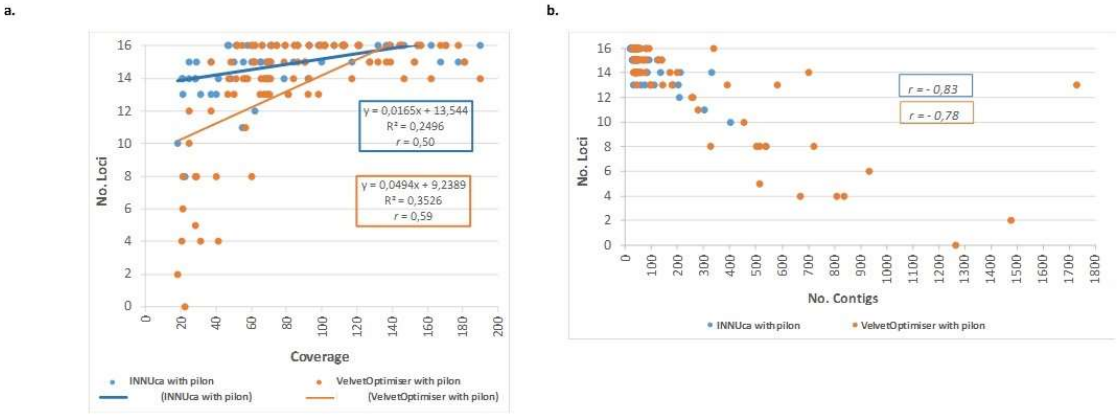


Figure 4.4 - Influence of the quality of the assembled genomes on the efficacy of the *in silico* MLVA loci extraction. The graphs show the correlation of the efficacy (measured by the number of loci for which an allele was called) of the bioinformatic script with the depth of coverage (panel A) and with the number of assembled contigs (panel B) after assembly polishment improvement. Two algorithms (SPAdes and VelvetOptimiser) were used to de novo assembly trimmed reads of each strain (see methods for details). The Pearson coefficients (r) are also shown for both graphs as well as the tendency lines shown with the respective equations (only for panel A).

4.4 Discussion

The control of brucellosis requires an accurate surveillance and the use of high discriminatory methods to characterize outbreak strains and determine the infection source and transmission routes. For many years, multiple typing methods were used for *Brucella* characterization at both species and biovar levels. These relied on host specificity, growth features, biochemical reactions, serotyping and bacteriophage typing, but they lacked discriminatory power (Sun *et al*., 2017). Currently, experimental MLVA is the most widely used approach for outbreak investigations and is still considered the gold standard for *Brucella* typing. The sixteen markers are a combination of moderately variable (minisatellites, panel 1) and highly discriminatory (microsatellites, panel 2) *loci* (Al Dahouk *et al*., 2007). A MLVA typing assay depends on the selection of markers which individually would not provide a relevant clustering. Taken separately, the Tandem Repeat markers are either not informative enough, are too variable or show a high level of homoplasy. As such, the combination of well selected independent loci may be highly discriminatory as previously show for other species (Le Fleche *et al*., 2006).

The MLVA procedure is expensive, experimentally demanding and reveals some accuracy problems regarding the precise determination of the size of the amplified products that are on the basis of the allelic determination. Thus, the final allelic profile may also slightly differ depending on the lab technician evaluation.

On behalf of the unavoidable transition from the classical typing to the WGS-based approaches a Python script was recently developed for the rapid *in silico* extraction of the *Brucella* MLVA alleles (Georgi *et al*., 2017). This will allow that the MLVA types can still be determined in the genomic era, avoiding an undesirable loss of genetic information that has been provided throughout decades by using the gold standard wet lab MLVA-based typing. Our main goal was to evaluate the performance of the developed *in silico* MLVA approach as no experimental validation of such script had been performed so far. Although the general agreement rate (wet-lab versus *in silico*) was low, this was somehow expected due to the well-known gel-associated bias when determining alleles differing by less than 10 bp. When these highly problematic alleles were not considered for the comparison, the agreement rate considerably increased up to 84.1%. Nevertheless, although it is highly likely that most discrepancies are due to the error-prone laboratory-based approach, we cannot discard the existence of some discrepancies due to discrete erroneous allele calling through the *in silico* approach considering the repetitive nature of the sequence stretches that enrol these loci. This is visible when analyzing the discrepancies between MLVA loci panels where loci from panels 2A and 2B (known to have the highest discriminatory power) yielded

more discrepancies than the ones of panel 1. This can be due to the fact that the sequence repetitions contained in the former are typically much smaller (from 3 to 8) than the ones observed in the later (from 12 to 134). Whereas this obviously impacts the proper alleles' distinction in gel electrophoresis, it may also influences the construction of contigs and consequently the *in silico* extraction of MLVA alleles.

Finally, according to the results obtained for the present dataset, we observed that the performance of the *in silico* approach does not seem to be dependent on the post-assembly polishment, but is clearly dependent on the depth of coverage and the degree of assembly fragmentation (where SPAdes performed better) (Supplementary Figure 4.1). Still, a minimal depth of coverage of ~50x and a maximum number of ~200 contigs (a range where both assemblers behaved similarly) seem to constitute guiding "boundaries" for the future application of the script.

In conclusion, the evaluated script seems to be a very useful and robust tool for *in silico* extraction of MLVA types of *Brucella* strains, dealing with a large number of samples in a short time period, and allowing retrospective and prospective molecular epidemiological studies. This allows a continuous and non disruptive transition to a new typing era by putting the newly sequenced strains in the frame of the genetic characterization obtained for thousands of isolates collected worldwide throughout decades. This will certainly be important for public health reference laboratories to maintain an active epidemiological surveillance of brucellosis.

CHAPTER V

GENETIC DIVERSITY OF TYPE IV BRUCELLA SPP. EFFECTORS AMONG B. MELITENSIS STRAINS CIRCULATING IN PORTUGAL

5 Genetic diversity of type IV *Brucella* spp. effectors among *B. melitensis* strains circulating in portugal

5.0 Introduction

One important feature of *Brucella* spp. is the ability to survive and multiply within both phagocytic and non phagocytyic cells. *Brucella* spp. does not produce classical virulence factors, such as exotoxins, cytolisins, exoenzymes, plasmids, fimbria, and drug resistant forms. Instead, major virulence factors include the LPS, the T4SS and the BvrR/BvrS system, which allow interaction with host cell surface, formation of an early and late BCV (*Brucella* containing vacuole) and interaction with endoplasmamtic reticulum (ER) when the bacteria multiply (Ke *et al*.. 2015; Lacerda *et al*., 2013; Myeni *et al*., 2013). The intracellular lifestyle of *Brucella* spp. limits the exposure of these bacteria to the host innate and adaptive immune responses, sequesters the organism from the effects of some antibiotics, and drives the unique features of pathology in infected hosts. *Brucella* spp., like many other intracellular pathogenic bacteria, secretes effector proteins into the host cytoplasm of infected cells in order to circumvent essential functions of the host defense, with the final goal of establishing a long lasting chronic infection (Byndloss *et al*., 2016). Therefore, elucidating the mechanisms involved in their intracellular survival and their ability to evade host immunity is crucial for understanding the pathogenesis of *Brucella* spp., which are frequently used as model organisms to study intracellular bacterial infections.

To restrict long-term protective immunity, *Brucella* spp. first avoid the innate immune response by stealthy entry into host cells. From there, the bacteria control aspects of protein secretion, intracellular trafficking and bacterial replication, ultimately altering the course of the innate and adaptive immune responses (de Barsy *et al*., 2011). Manipulation of the innate immune response seems to be associated to at least three effectors, namely, TcpB/BtpA, BtpB, and VceC (Salcedo *et al*., 2013; de jong *et al*., 2008).

Other effectors such as the BspB contribute to *Brucella* replication by redirecting Golgi-derived vesicles to the *Brucella* Containing Vacuoles (Miller *et al*., 2017). Among the multiple other examples, SepA participates in the early stages of intracellular survival (Dohmer *et al*., 2014), and BspA, BspB and BspF impair host protein secretion (Myeni *et al*., 2013). Putative effector candidates are constantly being identified *in silico* on the basis of several criteria, including shared features with effectors expressed by other bacteria, eukaryotic motifs, GC content, and limited distribution across bacterial genera (Esna Ashari *et al*., 2018; Sankarasubramanian *et al*., 2016, Myeni *et al*., 2013) Nevertheless, the effectors' list is likely far from being complete, and their precise role in the *Brucella* spp biology during the infectious process remains to be elucidated.

In the present ongoing study, we took advantage of the genome sequences of the *Brucella* spp. strains released on the course of the previous chapters, and aimed at evaluating the genetic variability of a set of T4SS effectors among the strains circulating in Portugal.

5.1 Experimental strategy and ongoing results

The whole genome sequence from 38 *Brucella* spp. strains obtained throughout the previous chapters was used to extract the individual sequences from each of 16 effectors, including 5 effectors recently identified by the Salcedo S. laboratory (BAB1_0296, BAB1_1101, BAB1_1533, BAB1_0277 and BAB1_1746) and 11 established effectors (RicA, BspA, CstA, BspB, Bep123, BtpA, SepA, BspF, VceA, VceC, BtpB,). The effectors enrolled in this study have been chosen either because their functions have already been characterized or they have been recently identified using *in silico* screens and confirmed to be translocated into host cells during infection. For example, BtpA and BtpB contain a TIR (Toll/interleukin-1 receptor) domain, BAB1_0296, BAB1_1101, BAB1_1533 and BAB1_1746 have an eukaryotic CAAX motif, BAB1_0277 and BAB2_0691 were detected on the BCV membrane are involving in the formation of the *Brucella* containing vacuole and finally, VceA, Bep123, BspA, BspB, BspF, SepA, CstA their all been described to their functions by many research groups.

The evaluation of their genetic diversity was carried out through an assembly-free strategy using Snippy v3.1 (https://github.com/tseemann/snippy), where reads of each strain are mapped against the respective sequences of *B. melitensis* 16M reference strain (RefSeq Accession NC_003317 and NC_003318). The primary results are illustrated in Figure 5.1.

The most curious finding was for BAB1_1533 gene, a new T4SS effector, as the sequences obtained ranged from 504 to 513 bp, due to the presence of a cytosine – rich region (C) translating a proline-rich region (P), which was variable from 6 to 9 nucleotides, in comparison with the reference strain. In 33,3% (13/38) of the strains this region is present and in some strains a 6 nucleotide deletion (44,7%, 17/38) and a 9 nucleotides deletion (20,4%, 7/38) is observed (Figure 5.2). No correlation with the presence and absence of this region was observed regarding the geographic origin and pathology of the patients and virulence of the strains.

In contrast, for BAB1_0296, as other new effector, no mutations were observed in all strains' sequences of this study. For the BAB1_1101, the third new effector, a single mutation (synonym mutation) was detected in 13% of the strains (5/38).

Regarding the other effectors study, for simplification purposes these can be divided in two groups based on the obtained genetic data. In a first group we observed genes without mutations among the studied strains (BAB1_0678 and BAB1_1279) or solely with synonyms mutations (BAB1_0277, BAB1_1746, BAB1_0279, BAB1_1552 and BAB1_0756). The other group contains the genes coding for the other effectors that revealed exclusively non-synonymous mutations (except for BAB1_1948, which revealed both synonymous and non-synonymous mutations). In general, the observed mutations were random in all strains studied and no correlation was observed regarding the geographic origin and pathology of the patients. Although the results are not conclusive at this stage, we highlight the polymorphism observed for BAB1_1533 due to the heterogeneous presence of the proline rich region. These regions have been

suggested to be associated with the bacterial cell wall being likely required for cell surface expression, but their precise role remains unclear (Areschoug *et al*., 2002). Thus, although in a pure speculative basis, we may hypothesize that the observed heterogeneity may underlie BAB1_1533 expression differences and ultimately, virulence differences, between the strains. Regarding the effectors for which no polymorphism was observed at the protein level (i.e., showing no mutations or only synonymous mutations), we could also speculate that this could be justified by the need to maintain an altered protein for which any mutation would be detrimental for its structure and/or function. For the effectors for which non-synonymous mutations were observed, although the resulting altered proteins may hypothetically reveal phenotypic dissimilarities, this hypothesis would be strengthened if the same alterations are observed in multiple strains, which is not the case for the current dataset.

Nevertheless, it is our aim to enlarge this ongoing polymorphism survey by enrolling all *Brucella* spp. genomes available in the public databases, in order to contribute to the characterization of the cellular function of these effectors, on behalf of a fruitful collaboration with the group of Dra. Suzana Salcedo from CNRS at University of Lyon.

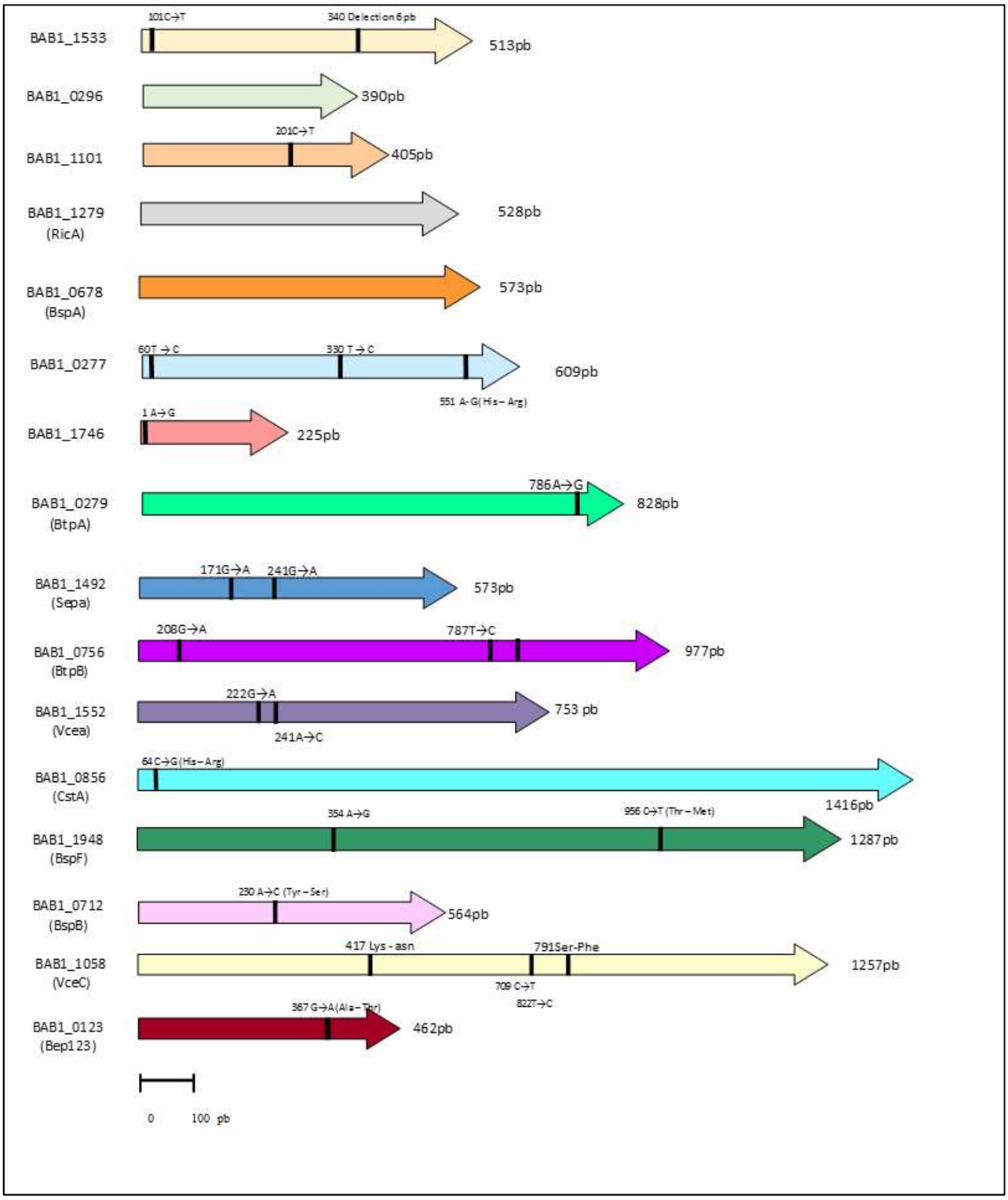


Figure 5.1. Identification of the mutations in all effectors used in this study and. In each gene are label the position in gene sequence that each mutation occur.

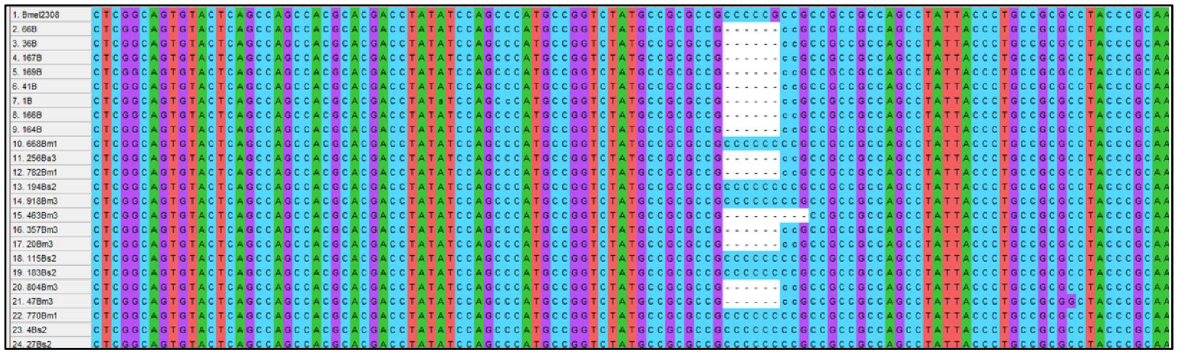


Figure 5.2. Alignment of BAB1_1533 amino acid sequences for strains used in this study and *B. melitensis* 16M (BME1216). The alignment was constructed with MEGA - X

CHAPTER VI

FINAL OVERVIEW, CONCLUDING REMARKS AND FUTURE PERSPECTIVES

6 Final overview, concluding remarks and future perspectives

Brucellosis is an endemic zoonotic disease in low, middle, and high-income countries (including Portugal), that causes devastating losses to the livestock industry including small-scale livestock holders. It places significant burdens on human healthcare systems and limits the economic potential of individuals, communities, and nations where such development is especially important to diminish the prevalence of poverty. The implementation of public policy focused on mitigating the socioeconomic effects of brucellosis in human and animal populations is desperately needed. In Portugal, although brucellosis is not eradicated and is of obligatory, no accurate prevalence data is available. Human cases are reported in all regions of continental Portugal, but only few cases are reported each year, which does not allow consistent analysis of risk factors and the impact on public health, despite the existence of a national program to eradicate brucellosis since 1953 that has been implemented with limited success. According to last data available in 2016 (ECDC, 2016) concerning human disease, Greece had the highest incidence, (1.10 per 100 000 persons), followed by Portugal and Italy (0.48 and 0.35 per 100 000 persons, respectively). These values underline the need to develop further studies that could provide new insights, allowing measures to reduce the impact of brucellosis in Portugal, or at least to prevent and mitigate it. The interdisciplinary "One Health" nature of the effects that brucellosis has indicate that collaboration of veterinary, medical, public health, cultural, economic and social experts is needed to perform a change in the disease burden. Furthermore, the access to epidemiological data and the cooperation between different health sectors (*i.e* human, veterinary and food safety) is recognizably weak. As such, the real time identification of brucellosis outbreaks is hampered as well as the identification of the infectious sources and the interruption of the transmission chains. Moreover, molecular surveillance of brucellosis is based on the traditional MLVA -16 procedure, which may lack sensitivity as only discrete regions of the genome are analyzed for discriminatory purposes.

The reference laboratory of brucellosis at INSA collects all isolates from human brucellosis and has been performing the traditional MLVA - 16 for the last 10 years. Taking advantage of the collection of the *Brucella* isolates at the reference laboratory and on the existence of Illumina sequencing equipment's at INSA, the major goal of this thesis was to contribute to a better knowledge of brucellosis in Portugal through the characterization of the *Brucella* spp. that circulate in this country, and by performing a retrospective analysis of the genetic relatedness among *B. melitensis* strains causing human infection.

We thus started (chapter II) by performing a molecular epidemiology survey of brucellosis infection in Portugal. This constituted an important step in brucellosis surveillance because, so far, the only epidemiological information that could be accomplished was the one taken from the obligatory declaration disease reports. Overall, serological diagnosis identified 167 out of 2313 (7.2%) positive cases of human brucellosis, of which 61.7% were male and half of the cases were in the age groups

between 26-65 years, suggesting that the distribution by gender is associated with the occupational factor. Molecular diagnostics was applied to a subset of 259 samples where 43 (16.6%) were positive for *B. melitensis.* The higher infection rates obtained when using real time PCR in comparison with immunological methods are likely due not only to a probable higher sensitivity of the former technique, but also because, according to our experience, PCR is usually requested when the clinician has a strong suspicion of brucellosis (e.g., patients revealing complications associated with the disease). This study strengthen the importance of integrating clinical and laboratory data of human cases in order to increase the efficacy of the response measures, essentially in the case of outbreaks. Furthermore, our findings reinforce the need to maintain an active epidemiological surveillance, enabling the early detection of all cases of infection and underlie the need to have a good communication flow between the human and animal Health Ministries.

Following the recommendations of the international Health Authorities to perform the surveillance of the infectious disease by using genome-scale approaches, in chapter III, we aimed at implementing a wgMLST schema at INSA, for surveillance of brucellosis. Very recently, efforts to develop cgMLST schemes for *Brucella* have been done by other groups (Sankarasubramanian J *et al*, 2019; Janowicz A *et al*, 2018). However, one of them involves a paid platform system and the other involves a small number of *loci* for discrimination purposes. Therefore, to overcome that hurdle, we created a curated species-specific wgMLST scheme that enrolls a panel of 2656 targets and used it to perform a retrospective analysis of the genetic relatedness among *B. melitensis* strains causing human infection in Portugal. We aimed at identifying potential outbreaks and transmission links that had been missed with the currently implemented surveillance system. We observed that strains showed a phylogenetic clustering within genotype II (25 out of 36) and IV (4 out of 36), and shared clades with strains isolated from countries such as Italy, Spain and Greece, with which Portugal has privileged food trading, tourism and similar eating habits. We have also identified several highly probable associated cases of brucellosis, where 22 out of the 36 PT strains showed one or multiple genetic linkage with other strains. With this approach, it was possible to identify six clusters and despite the absence of complete epidemiological information for most of the cases, our results point to the identification of strong associations between some of them, likely underlying missed "outbreaks". This methodology constitutes a hallmark of technological transition in the brucellosis surveillance in Portugal and will undoubtedly allow a more precise understanding of the epidemiology and dynamics of *Brucella* spp. populations and will unequivocally facilitate the assessment of ongoing and future outbreaks in order to prevent the transmission spread.

Nevertheless, considering that during the last decades the MLVA – 16 procedure has been the typing method, generating a massive amount of genomic information, it is important that the technological transition is not done in an abruptive manner to avoid missing the past genomic data. On this regard, an in house Python script had been developed in order to extract the MLVA schemes directly from genome sequences. However, as no experimental testing had been done with this script, in chapter

IV we aimed to evaluate the performance of this bioinformatics approach for the rapid *in silico* extraction of the MLVA alleles, by comparing it to the wet-lab MLVA procedure over of 83 strains from different *Brucella* species. Globally, the Python script detected 1263 hits in all 83 strains (95.1%), and showed an agreement rate with the wet-lab procedure of up to 84%, where major discrepancies are likely due to the error-prone laboratory-based approach. However, we cannot discard the existence of some discrepancies due to discrete erroneous allele calling through the *in silico* approach considering the repetitive nature of the sequence stretches that enrol these loci. Overall, this in *silico* approach to extract of MLVA-16 genotypes of *Brucella* strains seems to constitute a valid tool for surveillance purposes, enabling the integration of the new acquired genomic data from Brucella strains into the typing data that was collected for decades.

The last research study (chapter V) constitutes an ongoing study aimed at using the available genomic data obtained in the previous chapters to evaluate the genetic polymorphism of several virulence factors of *B. melitensis* strains circulating in Portugal. If significant genetic differences would be observed in some virulence factors for different strains, these data would be used in future studies to check the influence of such mutations in the virulence of *Brucella* spp. We started by analyzing the polymorphism of 16 previously identified Type IV effectors within 38 strains, and observed that only discrete mutations were found. Although a high degree of conservation of a gene is usually associated with an important biological function for which multiple genetic changes may be disadvantageous, these constitute preliminary data and larger datasets must be used.

Globally, we believe the results of the present PhD thesis constitute a step forward in the knowledge of *Brucella* circulating in Portugal. Among others, the major contribution of this dissertation is mainly the development of a new tool that will change the laboratory system applied to the surveillance of brucellosis in our country. In fact, the WGS–based approaches (both the wgMLST and the bioinformatics script) already implemented in the National Reference Laboratory will be thereafter applied for the in-depth genetic characterization of the *Brucella* isolates in a real time manner. Moreover, the implemented wgMLST may be also important for brucellosis surveillance in other countries as the existing platforms are either low informative or are not freely available. It will facilitate both the surveillance of brucellosis in endemic countries and the investigation of the imported cases in countries free for *Brucella*.

The data regarding brucellosis that are synthesized in this PhD thesis suggest that, to reduce or eradicate it, a “Global One Health” approach is essential. The concept “One Health” assumes that human, animal and environmental health are closely intertwined, and that improvement in one of these areas is contingent on the interdependence of all three. Thus, collaboration between professionals across multiple disciplines and sectors is imperative to reach solutions that lead to the mitigation of infectious diseases such as brucellosis.

In the future, it is our intention to strengthen the collaborations with the National Reference Veterinary laboratory and the Food and Economic Security Authority in order to expand the applicability

of these WGS approaches as tools that will better identify brucellosis outbreaks, the infectious sources and stop transmission chains. Ultimately, the impact in individual and public health as well as in the economy will rely on diminishing the morbidity and eventually mortality, and on the reduction of costs associated with treatment of infected people and animals. In this regard, we will pursuit one major goal in the near future, which consists in performing an epidemiological evaluation of human and ruminant brucellosis in areas of infected and vaccinated animals. This will be done essentially in two steps:

i) Construction of an epidemiological database of *B. abortus* and *B. melitensis* infection in humans and animals;

Epidemiological data from the archives and epidemiological questionnaires of the General Directorate of Health will be incorporated into a common database, allowing the characterization of the case's origin as foodborne or occupational disease. The Central Veterinary Services (DGAV) have a software (PISA) which stores data from the compulsory national eradication programs for ruminant brucellosis in Portugal. Epidemiologically relevant data will be also extracted from PISA regarding the surveys made on the infected sheep, goat and bovine herds. In addition, prospective data from epidemiological investigations and laboratory data will be integrated into the database. The expected output will be a database including the last 10 years historical epidemiological and laboratory data on the disease outbreaks and cases in humans and animals.

ii) Analysis of the databases of human and animal brucellosis, as well as the food database, for the investigation of clusters of infection;

Regions with higher infection prevalence and incidence rates in humans and animals will be identified. The intersection of these clusters will be evaluated and the overlapping areas will be specifically defined as areas of high risk for brucellosis and subjected to a thorough epidemiological study. Mass vaccination is an important disease control strategy implemented by the veterinary services in specific epidemiological units considered with a higher risk of brucellosis occurrence. Vaccination could result in a positive or neutral effect in animal or human health. The crossing of the clusters of herds of vaccinated bovine and small ruminants in association with the occurrence of occupational brucellosis in humans will also be done. This would allow studying the evolution of the prevalence and incidence of animal and human brucellosis regarding the implementation of vaccination programs on the identified clusters.

REFERENCES

References

Abou Zaki, N., Salloum, T., Osman, M., Rafei, R., Hamze, M., Tokajian, S. 2017. Typing and comparative genome analysis of *Brucella melitensis* isolated from Lebanon. FEMS Microbiol Lett. 16, 364(19). doi: 10.1093/femsle/fnx199

Al Dahouk, S. Sprague, L.D., Neubauer, H. 2013. New developments in the diagnostic procedures for zoonotic brucellosis in humans. Rev sci tech Off int Epiz. 32(1), 177-188.

Araj, F.G. 2010. Update on laboratory diagnosis of human brucellosis. International Journal of Antimicrobial Agents. 36S, S12-S17.

Areschoug, T., Linse, S.,2 Carlemalm, M. S., Hede, L.O., Lindahl, G. 2002. A Proline-Rich Region with a Highly Periodic Sequence in Streptococcal Protein Adopts the Polyproline II Structure and Is Exposed on the Bacterial Surface. Journal of Bacteriology. 184 (22), 6376–6383

Avijgan, M., Rostamnezhad, M., Jahanbani-Ardakani, H. 2019. Clinical and serological approach to patients with brucellosis: A common diagnostic dilema and worldwide perspective. Microbiol Pathogenesis. 129, 125-130.

Azam, S., Rao, S.B., Jakka, P., NarasimhaRao, V., Bhargavi, B., Gupta, V.K., Radhakrishnan, G. 2016. Genetic Characterization and Comparative Genome Analysis of Brucella melitensis isolates from India. Int J Genomics. doi: 10.1155/2016/3034756.

Baldi, P.C., Giambartolomei, G.H. 2013. Pathogenesis and pathobiology of zoonotic brucellosis in humans. Rev sci tech Off int Epiz. 32(1),117–125.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M, Kulikov, A.S.,Lesin, V.M., Nikolenko, S. I. Pham, S., Prijbelski, A.D., Pyskin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Peyzner, P.A. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 19(5), 455–477.

Barquero – Calvo, E., Chaves – Olarte, E., weiss, D.S., Guzman-Verri, C., Chacon.Díaz, C., Rucavado, A., Moriyón, I., Moreno, E. 2007. *Brucella abortus* uses a stealthy strategy to avoid activation of the innate imune system during the onset of infection. PlosOne. 2, e631.

Bohlin, J., Snipe, L., Cloeckaert, A., Lagesen, K., Ussery, D., Kristoffersen; A.B., Godfroid, J. 2010. Genomic Comparisons of *Brucella* spp. and closely related bacteria using base composital and proteome based methods. BMC Evol Biol. 10,249. doi:10.1186/1471-2148-10249.

Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15), 2114–2120.

Buddle, M.B. 1956. Studies on Brucella ovis (n.bp.), a cause of genital disease of sheep in New Zeland and Australia. J Hyg. 54(3).

Byndloss, M.X., Tsolis, R.M. 2016. Brucella spp. Virulence factors and immunity. Annu. Rev. Anim Biosci. 4, 111-127.

Carvalho, M.S., Barroso, M.R., Pinhal, F., Mota Tavares, F. 1995. Brucelose, alguns aspectos epidemiológicos. Medicina Int, 2 (4), 259-261.

Centers for Disease C. prevention. Suspected brucellosis case prompts investigation of possible bioterrorism – related activity. New Hampshire and Massachusetts, 1999. MMWR Morb Mortl Wkly Rep. 2000; 49(23), 509-12.

CLSI. Guideline M45. (2016). Methods for antimicrobial dilution and disk susceptibility testing of infrequently isolated or fastidious bacteria. (Clinical and Laboratory Standards Institute, Wayne, PA.

Corbel, M.J. 2014. Brucellosis: an overview. Emerg Infect Dis. 1997. 3(2):213-221. rucella papionis sp. nov., isolated from baboons (Papio spp.). Int J Syst Evol Microbiol. 64(Pt 12),4120-41288. doi: 10.1099/ijs.0.065482-0.

Corbel, M.J. 2006. Brucellosis in humans and animals. World Health Organization 2006. Annual epidemiological report, Food-and waterborne diseases and zoonoses. 2014 ECDC SURVEILLANCE REPORT.

D'Anastasio, R., Zipfel, B., Moggi – cecchi, J. 2009. Possible brucellosis in an early homini skeleton from stffontein, South Africa. Plos One.41.

Dean, A. S., Crump, L., Grete, H., Hattendorf, J., Schelling, E., Zinsstag, J. 2012. Clinical Manifestations of Human Brucellosis: A Systematic Review and Meta-Analysis. PLoS Negl Trop Dis. 6(12), e1929. doi:10.1371/journal.pntd.0001929.

Dean, A.S., Crump, L., Greter, H., Schelling, E., Zinsstag, J. 2012. Global Burden of Human Brucellosis: A Systematic Review of Disease Frequency. PLOS Neglected Tropical Disease. 6(10). doi: 10.1371/journal.pntd.0001865.

de Barsy, M., Jamet, A., Filopon, D., Nicolas, C., Laloux, G., Rual, J.F., Muller, A., Twizere, J.C., Nkengfac, B., Vandenhaute, J., Hill, D.E., Salcedo, S.P., Gorvel, J.P., Letesson, J.J., De Bolle, X. 2011. Identification of a Brucella spp. secreted effector specifically interacting with human small GTPase Rab2. Cell Microbiol. 2013(7), 1044-58. doi: 10.1111/j.1462-5822.2011.01601.x.

de Jong, MF., Sun, Y.H., den Hartigh, A.B., van Dijl, J.M., Tsolis, R.M. 2008. Identification of VceA and VceC, two members of the VjbR regulon that are translocated into macrophages by the Brucella type IV secretion system. Mol Microbiol. 70(6), 1378-96. doi:10.1111/j.1365-2958.2008.06487.x.

De Ley, J., Manheim, W., Segers, P., Lievens, A. 1987. Ribosomal ribonucleic acid cistron similarities and taxonomic neighborhood of Brucella and CDC group. Internal J Sist Bacteriol. 37, 35-42.

De Santis, R., Ancora, M., De Massis, F., Ciammaruconi, A., Zill,i K., Di Giannatale, E., Pittiglio, V., Fillo, S., Lista, F. 2013. Molecular strain typing of Brucella abortus isolates from Italy by two VNTR allele sizing technologies. Mol Biotechnol. 2013.55(2), 101-10. doi: 10.1007/s12033-013-9659-3.

DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los T., Ivanova, N.,Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltsman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.J., Haselkorn. R., Kyrpides, N., Overbeek, R. 2002. The genome of the facultative intracellular pathogen *Brucella melitensis*. Proc Natl Acad Sci USA. 99(1), 443-448.

Doenças de Declaração Obrigatória 2012-2015. Direção Geral de Saúde.

Doganay, G.D., Doganay, M. 2013. Brucella as a potential agent of bioterrorism. Recent Pat Antiinfect Drug Discov. 8(1), 27-33.

Döhmer, P.H., Valguarnera, E., Czibener, C., Ugalde, J.E. 2014. Identification of a type IV secretion substrate of Brucella abortus that participates in the early stages of intracellular survival. Cell Microbiol. 16(3), 396-410. doi: 10.1111/cmi.12224. Epub 2013 Nov 3.

Eisenberg, T., Hamann, H.P., Kaim, U., Schlez, K., Seeger, H., Schauerte, N., Melzer, F., Tomaso, H., Scholz, H.C., Koylass, M.S., Whatmore, A.M., Zschöck, M. 2012.Isolation of potentially novel *Brucella* spp. from frogs. Appl Environ Microbiol. May;78(10), 3753-5. doi: 10.1128/AEM.07509-11. Epub 2012 Mar 9.

Eisenberg, T., Riße, K., Schauerte, N., Geiger, C., Blom, J., Scholz, H.C. 2017. Isolation of a novel 'atypical' *Brucella* strain from a bluespotted ribbontail ray (Taeniura lymma).. Antonie Van Leeuwenhoek. Feb;110(2), 221-234. doi: 10.1007/s10482-016-0792-4. Epub 2016 Oct 26.

Esna Ashari, Z., Dasgupta, N., Brayton, K.A., Broschat, S.L. 2018. An optimal set of features for predicting type IV secretion system effector proteins for a subset of species based on a multi-level feature selection approach. PLoS One. 9;13(5), e0197041. doi: 10.1371/journal.pone.0197041.

Ferreira, A.C., Chambel, L., Tenreiro, T., Cardoso ,R., Flor, L., Dias, I.T., Pacheco, T., Garin-Bastuji, B., Le Flèche, P., Vergnaud, G., Tenreiro, R., de Sá M.I. 2012. MLVA16 Typing of Portuguese Human and Animal Brucella melitensis and Brucella abortus isolates. PLos One. 7(8). doi: 10.1371/journal.pone.0042514

Ferreira, A.C., Chambel, L., Corrêa de Sá, I. 2013. Brucella spp. Infection in Portugal: VNTR based typing of *B. melitensis* isolates. Magazine da SPM. 2, 28.03b.

Ficht, T. 2010. Brucella taxonomy and evolution.Future Micobiol. 5(6), 859-866.

Figueiredo, P., Ficht, T.A., Rice-Ficht, A., Rossetti, C.A., Adams. 2015. Pathogenesis and Immunobiology of Brucellosis. Review of Brucella – Host interactions. Infectious Disease Theme Issue. 185(6).

Foster, G., Osterman, B.S., Godfroid, J., Jacques, I., Cloeckaert, A. 2007. *Brucella ceti* sp. nov. and *Brucella pinnipedialis* sp. nov. for *Brucella* strains with cetaceans and seals as their preferred hosts. Int J Syst Evol Microbiol. Nov;57(Pt 11), 2688-93.

Foster, J.T., Beckstrom-Sternberg, S.M., Pearson, T., Beckstrom-Sternberg, J.S., Chain, P.S., Roberto, F.F., Hnath, J., Brettin, T., Keim, P. 2009. Whole-genome-based phylogeny and divergence of the genus Brucella. J Bacteriol. 191(8), 2864–70.

Francisco, A.P., Bugalho, M., Ramirez, M., Carriço, J.A. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics.10, 152.

Franc, K.A., Krecek, R.C., Hasler, B.N. & Arenas-Gamboa, AM. 2018. Brucellosis remains a neglected disease in the developing world: a call for interdisciplinary action. BMC Public Health. 18(125); doi: 10.1186/s12889-017-5016-y.

Franco, M.P., Mulder, M., Gilman, R.H., Smits, H.L. 2007. Human brucellosis. Lancet Infect Dis. 7(12): 775–86.

Franz, D.R., Parrot, C.D., Takafuji, E.T. 1997. The US biological warfare and biological defense programs. In Sidel FR, takafuji ET, Franz DR, editors. Medical Aspects of Chemical and Biological warfare. United of States. Office of the Surgeon General Army. 425-436.

Gad El-Rab, M.O., Kambal, A.M. 1998. Evaluation of a Brucella enzyme immunoassay test (ELISA) in comparison with bacteriological culture and agglutination. J Infect. 36(2), 197-201.

Gee, J.E., BK, D.E., Levett, P.N., Whitney, A.M., Novak, R.T., Popovic, T. 2004. Use of 16s rRNA gene sequencing for rapid confirmatory identification of Brucella isolates. J Clin Microbiol. 42, 3649-3654.

Georgi, E., Walter, M.C., Pfalzgraf, M.T., Northoff, B.H., Holdt, L.M., Scholz, H.C., Zoeller, L., Zangue, S., Antwerpen, M.H. 2017. Whole genome sequencing of Brucella melitensis isolated from 57 patients in Germany reveals high diversity in strains from Middle East. Plos One.

Godfroid, J., Nielsen, K., Saegerman, C. 2010. Diagnosis of brucellosis in livestock and wildlife. Croat Med J. 51(4), 296-305.

Gomez, G., Adams, L.G., Rice-Ficht, A., Ficht, T.A. 2013. Host – *Brucella* interactions and the Brucella genome as tools for subunit antigen discovery and immunization against brucellosis. Front Cell Infect Microbiol. 3, 17. doi:10.3389/fcimb.2013.00017.

Gopaul, K.K., Koylass, M.S., Smith, C.J., Whatmore, A.M. 2008. Rapid Identification of Brucella isolates to the species level by real time PCR based single nucleotide polymorphism (SNP) analysis. BMC Microbiology. 1-14.

Gorvel, J.P. 2014. If you bring an alarm, we will destroy it, said Brucella to the host cell. Virulence. 5, 460-462. doi:10.4161/viru.29092.

Grilló, M.J., Blasco, J.M., Gorvel, J.P., Moriyón, I., Moreno, E. 2012. What have we learned from brucellosis in the mouse model? Vet Res. 43, 29. doi:10.1186/1297-9716-43-29.

Halling SM., Peterson – Burch, B.D., Bricker, B.J., Zuerner, R.L., Qing, Z., Li, L.L., Kapur, V., Alt, D.P., Olsen, S.C. 2005. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. J. Bacteriol. 187(8), 2715-2726.

Hammerl, J.A., Göllner, C., Al Dahouk, S., Nöckler, K., Reetz, J., Hertwig, S. 2016. Analysis of the First Temperate Broad Host Range Brucellaphage (BiPBO1) Isolated from *B. inopinata*. Front Microbiol. 7, 24. doi: 10.3389/fmicb.2016.00024.

Hinic, V., Brodard, I., Thomann, A., Cvetnic, Z., Makaya, P.V., Frey, J., Abril, C. Novel. 2008. identification and differentiation of *Brucella melitensis*, *B. abortus*, *B. suis*, *B. ovis*, *B. canis*, and *B. neotomae* suitable for both conventional and real time PCR systems. Journal of Microbiological Methods.75, 375-378.

Hull, N.C., Schumaker, B.A. 2018. Comparisons of brucellosis between human and veterinary medicine. Infection Ecology and Epidemiology. 8.

Jacobs, M.K. 2004. The history of biologic warfare and bioterrorism. Dematol Clin. 22, 231-246.

Jaý, M., Girault, G., Perrot, L., Taunay, B., Vuilmet, T., Rossignol, F., Pitel, P.H., Picard, E., Ponsart, C., Mick, V. 2018. Phenotypic and Molecular Characterization of Brucella microti-Like Bacteria From a Domestic Marsh Frog (Pelophylax ridibundus). Front Vet Sci. 2018 Nov 15;5, 283. doi:10.3389/fvets.2018.00283.

Jama'ayah, M.Z., Heu, J.Y., Norazah , A. 2011. Seroprevalance of brucellosis among suspected cases in Malaysia. Malays J Pathol. Jun;33(1), 31-4.

Janowicz, A., Massis, F.D., Ancora, M., Cammà, C., Patavino, C., Battisti, A., Prior, K., Harmsen, D., Scholz, H., Zilli, K., Sacchini, L., Giannatle, E.D., Garofolo, G. 2018. Core Genome Multilocus Sequence Typing and Single Nucleotide polymorphism Analysis in the epidemiology of *Brucella melitensis* infections. . J Clin Microbiol. 56, 9.

Kaden, R., Ågren, J., Båverud, V., Hallgren, G., Ferrari, S., Börjesson, J., Lindberg, M., Bäckman, S., Wahab, T. 2014. Brucellosis outbreak in a Swedish kennel in 2013: determination of genetic markers for source tracing. Vet Microbiol. 174(3-4), 523-530. doi: 10.1016/j.vetmic.2014.10.015.

Kay, GI., Sergenat MJ., Giuffra V. 2014. Recovery of a medieval Brucella melitensis genome using shotgun metagenomics. MBio. 5(4).

Kattar, M.M., Jaafar, R.F., Araj, G.F., Le Flèche, P., Matar, G.M., Abi Rached, R., Khalife, S., Vergnaud, G. 2008. Evaluation of a Multilocus Variable Number tandem repeat analysis Scheme for a typing human *Brucella* isolates in a region of Brucellosis endemicity. J Clin Microbiol. 46(12), 3935-3940. doi: 10.1128/JCM.00464-08.

Ke. Y., Wang, Y., Chen, Z. 2015. Type IV secretion system of *Brucella* spp. and its effectors. Front Cell Infect Microbiol. 5,72.

Kim, S. 2015. The interaction between Brucella and the host cell in phagocytosis. In updates on Brucellosis. Chapter 3, pag. 45-60. Edited by Manual Mohammed Baddour, ISBN 978-953-51-2211-1. Publisher: in Tech.

Lacerda, T.L., Salcedo, S.P., Gorvel, J.P. 2013. Brucella T4SS: the VIP pass inside host cells. Curr Opin Microbiol. 45-51. doi: 10.1016/j.mib.2012.11.005.

Lamontagne, J., Béland, M.m Forest, A., Côté-Martin, A., Nassif, N., Tomaki, F., Moriyón, I., Moreno, E., Paramithiotis, E. 2010. Proteomics – based confirmation of protein expression and correction of annotations errors in the Brucella abortus genome. BMC genomics. 11,30. doi:10.1186/1471-2164-11-300.

Lapaque, N., Moriyón I., Moreno, E., Gorvel, J.P. 2005. Brucela lipopolysaccharide acts as a virulence factor. Current Opinion in Microbiology. 8, 60-66.

Le Fleche, P., Jacques, I., Grayon, M., AJ Dahoouk, S., Bouchon, P., Denoeud, F., Nockler, N., Neubauer, H., Guilloteau, L.A., Vergnaud, G. 2006. Evaluation and selection of tandem repeat loci for a Brucella MLVA typing assay. BMC Microbiol. 6, 9.

Letesson, J.J., Lestrate, P., Delrue, R.M., Danese, I., Bellefontaine, F., Fretin, D., Taminiau, B., Tibor, A., Dricot, A., Deschamps, C., Haine, V., Leonard, S., Laurent, T., Mertens, P., Vandenahute, J., De Bolle, X. 2002. Fun stories about Brucella: the "furtive nasty bug". Vet Microbiol. 90(1-4), 317-328.

Liu, W., Jing, Z., Ou, Q., Cui, B., He, Y., Wu, Q. 2012. Complete genome sequence of *Brucella melitensis* biovar 3 strain NI, isolated from an aborted bovine fetus. J Bacteriol. 194(22), 6321.

Llarena, A.K., Ribeiro-Gonçalves, B.F., Nuno Silva, D., Halkilahti, J., Machado, M.P., DaSilva, M.S., *et al.* 2018. INNUENDO: A cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens. EFSA supporting publication. EN-1498, 142.

Lucero, N.E., Escobar, G.I., Ayala, S. M., Jacob, N. 2005. Diagnosis of human brucellosis caused by *Brucella canis*. J Med Microbiol. 54, 457-61.

Mambres, D.H., Boarbi, S., Michel, P., Bouker, N., Escobar-Calle, L., Desqueber, D., Fancello, T., Esbroeck, M.V., Godfroid, J., Fretin, D., Mori, M. 2017. Imported human Brucellosis in Beligium: BIO and molecular typing of bacterial isolates, 1996-2015. Plos One. doi: 10.1371/journal.pone.0174756

Mantur, B.G., Amarnath, S.K., Shinde, R.S. 2007. Review of clinical and laboratory features of human Brucellosis. Indian J Med Microbiol. Jul, 25(3), 188-202.

Marianelli, C., Ciuchini, F., Tarantino, M., Pasquali, P., Adone, R. 2006. Molecular characterization of the rpoB gene in Brucella species: new potential molecular markers for genotyping. Microbes Infect. 8(3),860-865.

Martim-Martim, A.I., Sancho, P., Tejedor, C., fernández – Lago, L., Vizcaíno, N. 2011. Differences in the outer membrane – related properties of the six classical Brucella species. The Veterinary Journal. 189, 103-105.

Marzetti, S., Carranza, C., Roncallo, M., Escobar, G.I., Lucero, N.E. 2013. Recent trends in human *Brucella canis* infection. Comp Immunol Microbiol Infect Dis. 36(1), 55-61. doi:10.1016/j.cimid.2012.09.002.

McDermott, J.D,, Grace, D., Zinsstag, J. 2013. Economics of brucellosis impact and control in low-income countries. Rev Sci Tech;32(1), 249–61.

Miller, C.N., Smith, E.P., Cundiff, J.A., Knodler, L.A., Bailey Blackburn, J., Lupashin, V., Celli, J. 2017. A *Brucella* Type IV Effector Targets the COG Tethering Complex to Remodel Host Secretory Traffic and Promote Intracellular Replication. Cell Host Microbe.13;22(3), 317-329.e7. doi:10.1016/j.chom.2017.07.017.

Moreno, E., Moriyón, I. 2006. The genus Brucella. In falkow S., et al., editors. (ed.), the prokaryotes, vol. 5, p. 315-456 Springer – Verlag, New York, NY.

Moreno, E. 2014. Retrospective and prospective perspectives on zoonotic brucellosis. Front Microbiol. 13(5), 213.

Murray, P.R., Rosenthal, K.S., Kobayashi, G.S., Pfaller, M.A. 2005. Medical Microbiology. 5th ed. Elsevier.

Myeni, S., Child, R., Ng, T.W., Kupko, J.J., Wehrly, T.D., Porcella, S.F., Knodler, L.A., Celli, J. 2013. *Brucella* modulates secretory trafficking via multiple type IV secretion effector proteins. PLoS Pathog. 9(8). doi:10.1371/journal.ppat.1003556.

Orduña, A., A. Almaraz, A. Prado, M. P. Gutiérrez, A. García-Pascual, A. Dueñas, M. Cuervo, R. Abad, B. Hernández, B. Lorenzo, M. A. Bratos, and A. R. Torres. 2000. Evaluation of an immunocapture-agglutination test (Brucellacapt) for serodiagnosis of human brucellosis. J. Clin. Microbiol. 38**:**4000-4005.

Pappas, G., Akritidis, N., Bosilkovski, M., Tsianos, E. 2005. Brucellosis. N Engl J Med. 352 (22), 2325-36.

Pappas, G., Solera, J., Akritidis, N., Tsianos, E. 2005. New approaches to the antibiotic therapy of brucellosis. International journal of antimicrobial agents. 26. doi:101-5. 10.1016/j.ijantimicag.2005.06.001.

Pappas, G., Papadimitriou, P., Akritidis, N., Christou, L., Tsianos, E.V. 2006. The new global map of human brucellosis. Lancet Infect Dis. 6(2), 91–9.

Pappas, G., Papadimitriou, P. 2007. Challenges in **Brucella** bacteraemia. International Journal of Antimicrobial Agents. S29-S31.

Pappas, G. 2010. The changing Brucella ecology: novel reservoirs, new threats. Int J Antimicrob Agents. 36 (Suppl 1):S8-11.

Parkhill, J., Wren, B.W. 2011. Bacterial epidemiology and biology – lessons from genome sequencing. Genome Biology. 12, 230.

Paulsen, I.T., Seshadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L., Brinkac, L.M., Beanam, M.J., Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J., Van Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O., Salzberg, S.L., Hoover, D.L., Lindler, L.E., Halling, S.M., Boyle, S.M., Fraser, C.M. 2002. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc Natl Acad Sci USA. 99(20), 13148-13153.

Pelerito, A., Cordeiro. R., Matos, R., Santos, M.A., Soeiro, S., Santos, J., Manita, C., Rio, C., Santo. M.P., Paixão, E. Nunes, A., Núncio, S. 2017. Human Brucellosis in Portugal – Retrosptective analysis of suspected clinical cases of infection from 2009 to 2016. Plos One. **10**,12(7). doi:10.1371/journal.pone.0179667.

Pisarenko, S.V., Kovalev, D.A., Volynkina, A.S., Ponomarenko, D.G., Rusanova, D.V., Zharinova, N.V., Khachaturova, A.A., Tokareva, L.E., Khvoynova, I.G., Kulichenko, A.N. 2018. Global evolution and phylogeography of *Brucella melitensis* strains. BMC Genomics. 10, 19(1),353. doi:10.1186/s12864-018-4762-2.

Queipo-Ortuño, M.I., Morata, P., Ocón, P., Manchado, P., Colmenero, J.D. 1997. Rapid diagnosis of human brucellosis by peripheral-blood PCR assay. J Clin Microbiol. 35(11), 2927-30.

Rambow-Larsen, A.A., Petersen, E.M., Gourley, C.R., Splitter, G.A. 2009. Brucella regulators: self-control in a hostle environment. Trends Microbiol. 17(8): 371-377. doi:10.1016/j.tim.2009.05.006.

Ratushna, V.G., Sturgill, D.M., Ramamoorthy, S., Reichow, S.A., He, Y., Lathigra, R., Sriranganathan, N., Halling, S.M., Boyle, S.M., Gibas, C.J. 2006. Molecular targets for rapid identification of *Brucella* spp. BMC Microbiol. 22(6), 13.

Ribeiro-Gonçalves, B., Francisco, A.P., Vaz, C., Ramirez, M., Carriço, J.A. 2016. PHYLOViZ Online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. Nucleic Acids Res. 44(W1), W246–W251.

Rich, M., Bannatyne, R.M., Memish, Z.A. 2000. Direct urease test on BACTEC blood cultures: early presumptive diagnosis of brucellosis in an area of endemicity. J Clin Microbiol. 38(4), 1706.

Robichaud, S., Libman, M., Behr, M., Rubin, E. 2004. Prevention of laboratory-acquired brucellosis. Clin Infect Dis. 15;38(12), e119-22.

Roop, R.M., Gaines, J.M., Andreson, E.S., Caswell, C.C., Martin, D.W. 2009. Survival of the fittest:how Brucella strains adapt to their intracellular niche in the host. Med Microbiol Immunol.. 198(4): 221-238. doi:10.1007/s00430-009-0123-8.

Rossetti, C.A., Drake, K.L., Adams, L.G. 2012. Transcriptome analysis of HeLa cells response to Brucella melitensis infection: a molecular approach to understand the role of the mucosal epithelium in the onset of the Brucella pathogenesis. Microbes Infect. 14,756-767.

Ruiz-Ranwez, V., Posadas D. M., Van der Henst, C., Estein, S. M., Arocena, G. M., Abdian, P. L. 2013. BtaE, an adhesin that belongs to the trimeric autotransporter family, is required for full virulence and defines a specific adhesive pole of ***Brucella suis***. Infect. Immun. 81, 996–1007. doi10.1128/IAI.01241-12.

Rust, R. 2006. Brucellosis [Online]. Jun 8. Available from: URL: http://emedicine.medscape.com/article/1164632-overview

Salcedo, S.P., Marchesini, M.I., Degos, C., Terwagne, M., Von Bargen, K., Lepidi, H.,Herrmann, C.K., Santos Lacerda, T.L., Imbert, P.R., Pierre, P., Alexopoulou, L., Letesson, J.J., Comerci, D.J., Gorvel, J.P. 2013. BtpB, a novel Brucella TIR-containing effector protein with immune modulatory functions. Front Cell Infect Microbiol. 8;3,28. doi:10.3389/fcimb.2013.00028.

Sankarasubramanian, J., Vishnu, U.S., Dinakaran, V., Sridhar, J., Gunasekaran, P., Rajendhran, J. 2016. Computational prediction of secretion systems and secretomes of Brucella: identification of

novel type IV effectors and their interaction with the host. Mol Biosyst. 12(1), 178-90. doi: 10.1039/c5mb00607d.

Sankarasubramanian, J., Vshnu, U.S., Gunasekaran, P., Rajendhran, J. 2017. Identification of genetic variations of Brucella spp. through genome – wide association studies. Infection, Genetics and Evolution. 56, 92-98.

Sankarasubramanian, J., Vishnu, U.S., Gunasekaran, P., Rajendhran, J. 2019. Development and evaluation of a core genome multilocus sequence typing (cgMLST) scheme for Brucella spp. Infection, genetics and Evolution.67, 38-43.

Santis, R.D., Ciammaruconi, A., Faggioni, G., Fillo, S., Gentile, B., Di Giannatale, E., Ancora, M., Lista, F. 2011. High throughput MLVA-16 typing for *Brucella* based on the microfluidics technology. BMC Microbiol. 11, 60.

Scholz, H.C., Hubalek, Z., Sedlácek, I., Vergnaud, G., Tomaso, H., Al Dahouk, S., Melzer, F., Kämpfer, P., Neubauer, H., Cloeckaert, A., Maquart, M., Zygmunt, M.S., Whatmore, A.M., Falsen, E., Bahn, P., Göllner, C., Pfeffer, M., Huber, B., Busse, H.J., Nöckler, K. 2008. *Brucella microti* sp. nov., isolated from the common vole *Microtus arvalis*. Int J Syst Evol Microbiol. Feb;58(Pt 2), 375-82. doi: 10.1099/ijs.0.65356-0.

Scholz, H.C., Nöckler, K., Göllner, C., Bahn, P., Vergnaud, G., Tomaso, H., Al Dahouk, S., Kämpfer, P., Cloeckaert, A., Maquart, M., Zygmunt, M.S., Whatmore, A.M., Pfeffer, M., Huber, B., Busse, H.J., De, B.K. 2010. *Brucella inopinata* sp. nov., isolated from a breast implant infection. Int J Syst Evol Microbiol. Apr;60(Pt 4), 801-8. doi: 10.1099/ijs.0.011148-0. Epub 2009 Aug 6.

Scholz, H.C., Revilla-Fernández, S., Al Dahouk, S., Hammerl, J.A., Zygmunt, M.S., Cloeckaert, A., Koylass, M., Whatmore, A.M., Blom, J., Vergnaud, G., Witte, A., Aistleitner, K., Hofer, E. 2016. *Brucella vulpis* sp. nov., isolated from mandibular lymph nodes of red foxes. (Vulpes vulpes). Int J Syst Evol Microbiol. 66(5), 2090-8. doi: 10.1099/ijsem.0.000998. Epub 2016 Feb 29.

Sellem, M.N., Boyle, S.M., Sriranganathan, N. 2008. Brucella: a pathogen without classic virulence genes. Vet Microbiol. 129(1-2), 1-14. doi:10.1016/j.vetmic.2007.11.023.

Seleem, M.N., Boyle, S.M., Sriranganathan, N. 2010. Brucellosis: A re-emerging zoonosis. 2010. Veterinary Microbiology. 140, 392–398.

Shevtsova, E., Shevtsov, A., Mukanov, K., Filipenko, M., Kamalova, D., Sytnik, I., Syzdykov, M., Kuznetsov, A., Akhmetova, A., Zharova, M., Karibaev, T., Tarlykov, P., Ramanculov, E. 2016. Epidemiology of Brucellosis and Genetic Diversity of Brucella abortus in Kazakhstan. PLoS One. 1;11(12), e0167496. doi: 10.1371/journal.pone.0167496.

Silva, M., Machado, M.P., Silva, D.N., Rossi, M., Moran-Gilad, J., Santos, S., Ramirez, M., Carriço, J.A. 2018. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. Microb Genom. 4(3), e000166.

Skalsky, K., Yahav, D., Bishara, J., Pitlik, S., Leibovici, L., Paul, M. 2008. Treatment of human brucellosis: Systematic review and meta-analysis of randomised controlled trials. BMJ (Clinical research ed.). 336. 701-4. 10.1136/bmj.39497.500903.25.

Smith, J.A. 2018. Brucella Lipopolysaccharide and pathogenicity: the core of the matter. Virulence. 9(1), 379-382. doi:10.1080/21505594.2017.1395544.

Solera, J. 2010. Update on brucellosis: therapeutic challenges. Int J Antimicrob Agents. 36 Suppl 1:S18-20. doi: 10.1016/j.ijantimicag.2010.06.015.

Sun, M., Jing, Z., Di, D., Yan, H., Zhang, Z., Xu, Q., Zhang, X., Wang, X., Ni, B., Sun, X., Yan, C., Yang, Z., Tian, L., Li, J., Fan, W. 2017. Multiple Locus Variable – Number tandem – repeat and Single – Nucleotide Polymorphism – Based *Brucella* Typing reveals multiple Lineages in *Brucella melitensis* currently endemic in China. Frontiers in Veterinary Science. 4.215. doi: 10.3389/fvets.2017.00215

Supriya Christopher, B.L., Umapathy, K.L., Ravikumar. 2019. Brucellosis: Review on the Recent Trends in Pathogenicity and Laboratory Diagnosis. Brucellosis: Review on the Recent Trends in Pathogenicity and Laboratory Diagnosis.

Tan, K.K., Tan, Y.C., Chang, L.Y., Lee, K.W., Nore, S.S., Yee, W.Y., Isa, M.N.M., Jafar, F.L., Hoh, C.C., AbuBakar, S. 2015. Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. BMC Genomics. 2015; 16, 93.

Tiller, R.V., Gee, J.E., Frace, M.A., Taylor, T.K., Setubal, J.C., Hoffmaster, A.R., De, B.K. 2010.Characterization of novel Brucella strains originating from wild native rodent species in North Queensland, Australia. Appl Environ Microbiol. Sep;76(17), 5837-45. doi: 10.1128/AEM.00620-10. Epub 2010 Jul 16.

Tonna, I., Tonna, A. 2005. Brucellosis. N Engl J Med. Sep 8,353(10), 1071-2.

Tzaneva, V., Ivanova, S., Georgieva, M., Tasheva, E. 2009. Investigation of the spread of Brucellosis among human and animal populations in southeastern Bulgaria. Eurosurveillance. 14(17).

Valdezate, S., Navarro, A., Villalón, P., Carrasco, G., Saéz-Nieto, J.A. 2010. Epidemiological and Phylogenetic Analysis of Spanish Human *Brucella melitensis* Strains by Multiple Locus Variable Number Tandem Repeat Typing, Hypervariable Octameric Oligonucleotide Fingerprinting, and rpoB Typing. J Clin Microbiol. 48(8), 2734-40. doi: 10.1128/JCM.00533-10.

Verger, J.M., Grimont, F., Grimont, P.A.D. 1985. *Brucella*, a monospecific genus as shown by deoxyribonucleic acid hybridization. IntJ Syst Bacteriol. 35, 292–5.

Vizcaíno, N., Cloeckaert, A. 2012. Biology and Genetics of the Brucella outer membrane, p. 133-161. In I. Lopez-Goni and D. O'Callaghan (ed.), Brucella: Molecular Microbiology and Genomics. Caister Academic Press, Norfolk (GBR).

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE. 9(11).

Wattam, A.R., Foster, J.T., Mane, S.P., Bekxstrom-Stemberg, S.M., Beckstrom-Stemberg, J.M., Dickerman, A.W., Keim, P., Pearson, T., Shukla, M., Ward, D.V., Williams, K.P., Sobral, B.W., Tsolis, R.M., Whatmore, A.M., O'Callaghan, D. 2014. Comparative phylogenomics and evolution of the Brucellae reveal a path to virulence. J Bacteriol 196, 920-930.

Wattam, A.R., Williams, K.P., Snyder, E.E., Almeida Jr, N.F., Shukla, M., Dickerman, A.W. 2009. Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. J Bacteriol. 191(11), 3569–79

Weeks, J.N., Galindo, C.L., Drake, K.L., Adams, G.L., Garner, H.R., Ficht, T.A. 2010. *Brucella melitensis* VjbR and C12-HSL regulons: contributions of the N-dodecanoyl homoserine lactone signaling molecule and LuxR homologue VjbR to gene expression. BMC Microbiol. 10, 167.

Weiming, Y., Meibian, Z., Hua, Z., Xiangjing, G., Yuqing, L. 2018. Emergency response to occupational brucellosis in a pharmaceutical manufacturing enterprise. J Occup Health. 20; 60(5), 404–409. doi: 10.1539/joh.2018-0089-CS.

Weinstein RA., Singh K., Laboratory – acquired infections. Clin Infect Dis. 2009; 49(1):142-147.

Mense, M.G., Borshche,l R.H., Wilhelmsen, C.L. 2004. Pathologic changes associated with brucellosisnexperimentally induced by aerosol exposure in rhesus macaques (macaca mulatta). Am J vet res. 65 (5), 644-652.

Whatmore, A. M., Davison, N., Cloeckaert, A., Al Dahouk, S., Zygmunt, M. S, Brew, S. D., Perrett L. L., Koylass, M. S., Vergnaud G., Quance, C., Scholz, H. C., Dick, E. J., Hubbard, G. N. E. Schlabritz, L. 2014. *Brucella papionis* sp. nov., isolated from baboons (Papio spp.). Int J Syst Evol Microbiol. 1; 64(Pt 12), 4120–4128. doi: 10.1099/ijs.0.065482-0.

Whatmore, A. M. 2009. Current understanding of the genetic diversity of Brucella, an expanding genus of zoonotic pathogens. Infection, Genetics and Evolution. 9, 1168-1184.

Whatmore, A.M., Perret, L.L., MacMillan, A.P. 2007. Characterization of the genetic diversity of *Brucella* by multilocus sequencing. BMC Microbiology.7, 34.

Whatmore, A.M., Shankster, S.I., Perrett, L.L., Murphy, T.J., Brew, S.D., Thirlwall, R.E., Cutler, S.I., MacMillan, A.P. 2006. Identification and characterization of Variable-Number Tandem – Repeat Markers for typing of Brucella spp. J Clin Microbiol. 44, 1982-1993.

Williams D., Gogarten JP., Lapierre P. 2011. Filling the gaps in the genomic landscape. Genome Biol. 11(2), 103.

Wright, A. E., and F. Smith. 1997. On the application of the serum test to the differential diagnosis of typhoid fever and Malta fever.BMJ. 1, 1214-121.

Wright, P.F., Tounkara, K., Lelenta, M., Jeggo, M.H. 1997. International reference standards: antibody standards for the indirect enzyme-linked immunosorbent assay. Rev Sci Tech. 16(3), 824-32.

Wyatt, H.V. 2013. Lessons from the history of brucellosis. Revue Scientific et technique (Internationl Office of Epizootics), 32(1),17-25.

Young, E.J. 1995. An Overview of human brucellosis. Clin Infect Dis.; 21, 283-290.

Young, E.J.2005. *Brucella* species. In: Mandell GL, Bennett JE, Dolin R, eds. Principles and practice of infectious diseases, 6th ed. Philadelphia: Churchill Livingstone. 2669–2672.

Yousefi-Nooraie, R., Mortazhejri, S., Mehrani, M., Sadeghipour, P. 2012. Antibiotics for treating human brucellosis. Cochrane database of systematic reviews (Online). doi:10.1002/14651858.CD007179

http://www.who.int/neglected_diseases/zoonoses/other_NZDs/en/

http://www.who.int/zoonoses/diseases/Brucellosissurveillance.pdf

SUPLEMENTTAL MATERIAL

Suplementtal Material

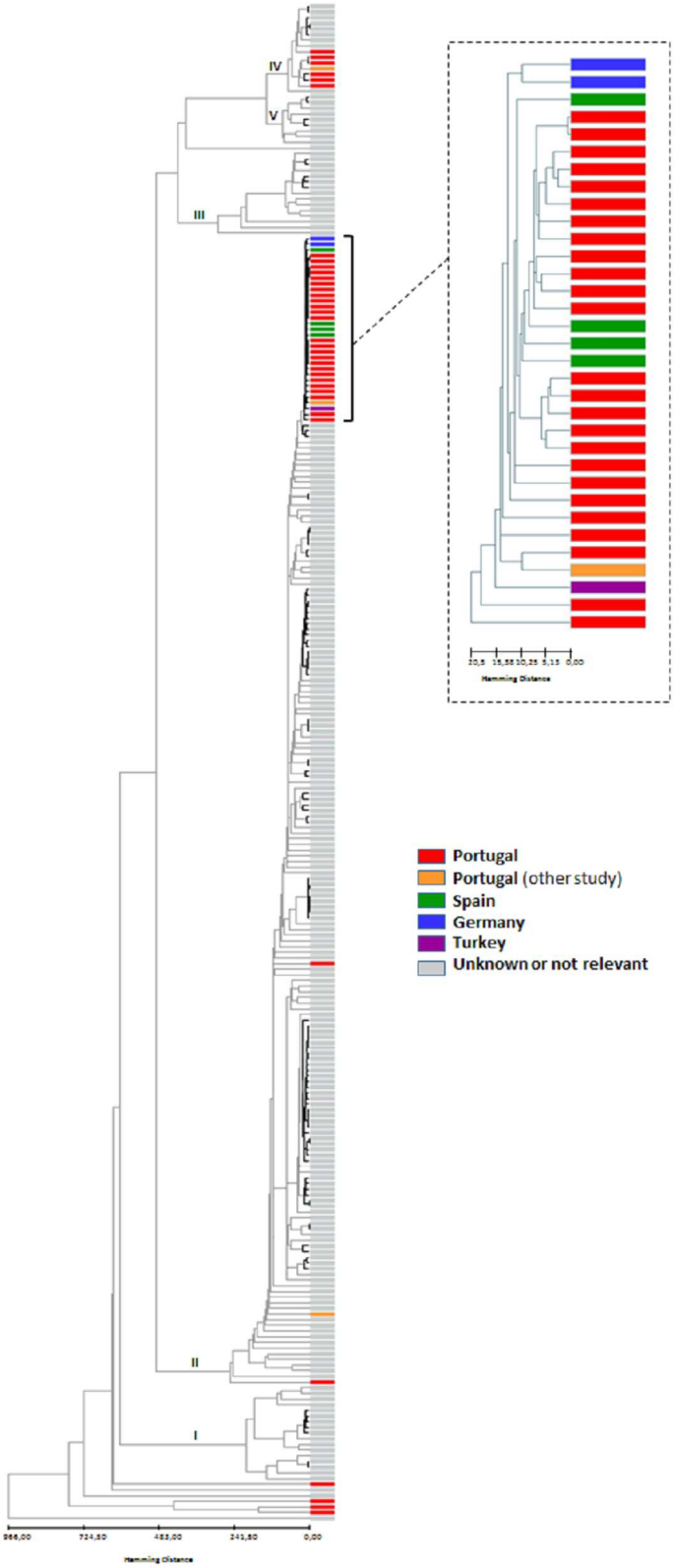
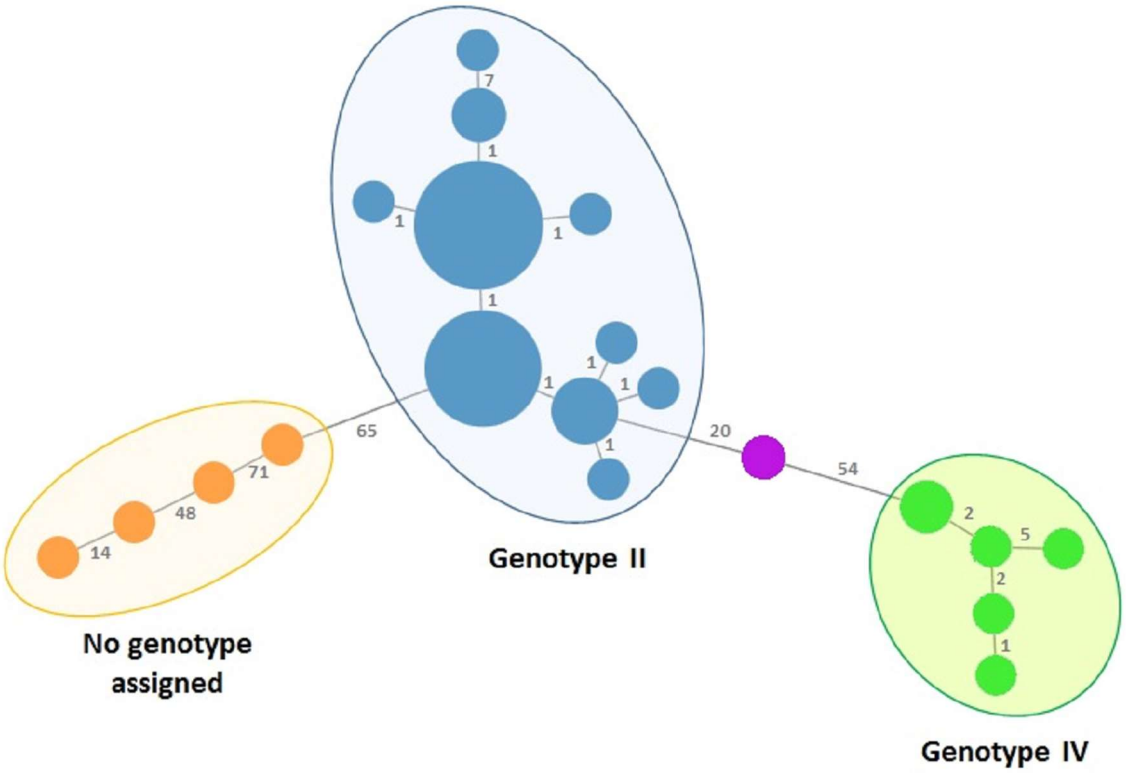


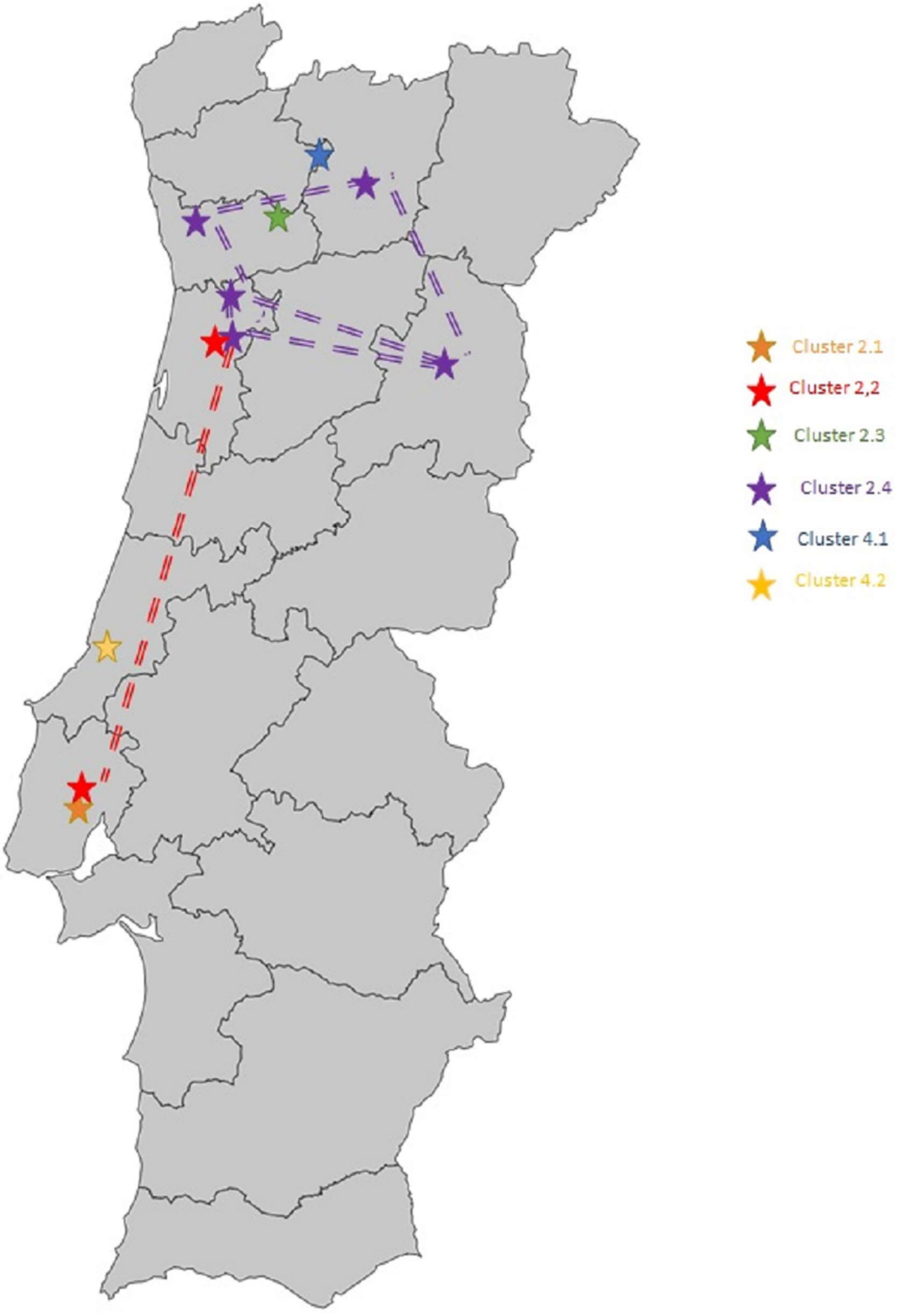
Figure S 3.1 - Hierarchical clustering tree, showing the genetic relationship of the 271 B. melitensis strains collected worldwide, based on a curated species-specific wgMLST scheme that enrolls a panel

of 2656 genes (and 17472 alleles). Distances among strains were estimated with Hamming Distance metrics using the single-linkage method. Branch trees representing clusters linked by a genetic relatedness cut-off of 3% are shown in bold black lines. The PT strains are highlighted in red, while strains genetically related to them appear in different colors (also zoomed in for better visualization), concerning the isolation country. The five major genotypes are also displayed above the tree branches.



Supplementary Figure S2

Figure S 3.2 – Phylogeny of PT *B. melitensis* strains based on a gene-by-gene approach using the genus-specific cgMLST schema with 164 loci. The Minimum spanning tree (MST) was constructed using the goeBURST algorithm implemented in the PHYLOViZ Online platform, and is based on the allelic diversity found among the 164 loci panel (Sankarasubramanian J et al, 2019). Filled small circles (nodes) represent unique allelic profiles. For comparative purposes with the proposed wgMLST scheme of the present study, nodes are colored similarly to Figure 1 and are grouped based on the assigned genotype according to Tan *et al*. (Tan et al, 2015). The numbers in grey on the connecting lines represent the allele differences (AD) between strains.



Supplementary Figure S 3.3 - Geographic location of the isolated 36 PT *B. melitensis* strains. For simplification purposes, the color scheme used to define the clusters is the same as the one presented in Figure 3. Strains belonging to the same putative cluster are connected by the color corresponding to each cluster.

Supplementary Table 3.1 – strain characterization

Strain	Biosample	Run
20Pa	ERS2952753	ERR2938658
47Pa	ERS2952754	ERR2938668
357Pa	ERS2952755	ERR2938653
463Pa	ERS2952756	ERR2938687
770Pa	ERS2952757	ERR2938669
782Pa	ERS2952758	ERR2938670
804Pa	ERS2952759	ERR2938703
918Pa	ERS2952760	ERR2938695
1P	ERS2952761	ERR2938677
35P	ERS2952762	ERR2938664
36P	ERS2952763	ERR2938682
38P	ERS2952764	ERR2938646
41P	ERS2952765	ERR2938690
43P	ERS2952766	ERR2938702
147P	ERS2952767	ERR2938647
153P	ERS2952768	ERR2938674
165P	ERS2952769	ERR2938679
169P	ERS2952770	ERR2938684
177P	ERS2952771	ERR2938659
179P	ERS2952772	ERR2938663
180P	ERS2952773	ERR2938676
184P	ERS2952774	ERR2938700
194P	ERS2952775	ERR2938706
198P	ERS2952776	ERR2938705
199P	ERS2952777	ERR2938667
200P	ERS2952778	ERR2938696
209P	ERS2952779	ERR2938704
228P	ERS2952780	ERR2938693
237P	ERS2952781	ERR2938654
258P	ERS2952782	ERR2938680
261P	ERS2952783	ERR2938642
40P	ERS2952793	ERR2938685
44P	ERS2952794	ERR2938656
66P	ERS2952795	ERR2938649
166P	ERS2952796	ERR2938689
167P	ERS2952797	ERR2938645
168P	ERS2952798	ERR2938651

Supplementary Table S 3.2. - Results of AST testing of *Brucella melitensis*. * CLSI breakpoints for slow-growing bacteria (*Haemophilus* spp.)

Antibiotic agent	MICs Range (µg/ml) of the *B. melitensis* PT strains	CLSI breakpoints for *Brucella* spp. (µg/ml)		
		S ≤	I =	R≥
Rifampicin*	0.38 - 12	1	2	4
Doxycyclin	<0.016-1	1	-	-
Streptomicyn	0.5-4	16	-	-
Gentamicin	0.5-2	4	-	-

Supplementary Table S 4.1 - Strain characterization

Strain	Species	Isolation Country	Collection_date	Host	Biosample	Run
261P	*Brucella melitensis*	Portugal	2016	Homo sapiens	ERS2952783	ERR2938642
213-03E	*Brucella melitensis*	Spain	2003	Homo sapiens	ERS2952739	ERR2938643
170-04E	*Brucella melitensis*	Spain	2004	Homo sapiens	ERS2952736	ERR2938644
167P	*Brucella melitensis*	Portugal	2011	Homo sapiens	ERS2952797	ERR2938645
38P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952764	ERR2938646
147P	*Brucella melitensis*	Portugal	2013	Homo sapiens	ERS2952767	ERR2938647
146-12RK	*Brucella melitensis*	Spain	2012	Unknown	ERS2952750	ERR2938648
66P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952795	ERR2938649
183Pa	*Brucella suis*	Portugal	2009	Boar	ERS2952791	ERR2938650
168P	*Brucella melitensis*	Portugal	2011	Homo sapiens	ERS2952798	ERR2938651
104-13RK	*Brucella melitensis*	Germany	2013	Unknown	ERS2952748	ERR2938652
357Pa	*Brucella melitensis*	Portugal	2004	Sheep	ERS2952755	ERR2938653
237P	*Brucella melitensis*	Portugal	2016	Homo sapiens	ERS2952781	ERR2938654
115Pa	*Brucella suis*	Portugal	2008	Boar	ERS2952790	ERR2938655
44P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952794	ERR2938656
457-06E	*Brucella melitensis*	Spain	2006	Homo sapiens	ERS2952744	ERR2938657
20Pa	*Brucella melitensis*	Portugal	2002	Goat	ERS2952753	ERR2938658
177P	*Brucella melitensis*	Portugal	2014	Homo sapiens	ERS2952771	ERR2938659
183-7RK	*Brucella ovis*	Hungary	2007	Unknown	ERS2952784	ERR2938660
104-12RK	*Brucella melitensis*	Germany	2012	Unknown	ERS2952747	ERR2938661
183-6RK	*Brucella suis*	Hungary	2006	Unknown	ERS2952787	ERR2938662
179P	*Brucella melitensis*	Portugal	2014	Homo sapiens	ERS2952772	ERR2938663
35P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952762	ERR2938664
148-9RK	*Brucella melitensis*	Belgium	2009	Unknown	ERS2952751	ERR2938665
256Pa	*Brucella abortus*	Portugal	2005	Bovine	ERS2952786	ERR2938666
199P	*Brucella melitensis*	Portugal	2015	Homo sapiens	ERS2952777	ERR2938667
47Pa	*Brucella melitensis*	Portugal	2001	Sheep	ERS2952754	ERR2938668
770Pa	*Brucella melitensis*	Portugal	2007	Sheep	ERS2952757	ERR2938669
782Pa	*Brucella melitensis*	Portugal	2007	Goat	ERS2952758	ERR2938670
297-04E	*Brucella melitensis*	Spain	2004	Homo sapiens	ERS2952742	ERR2938671
120-99E	*Brucella melitensis*	Spain	1999	Homo sapiens	ERS2952734	ERR2938672
27Pa	*Brucella suis*	Portugal	2003	Swine	ERS2952789	ERR2938673
153P	*Brucella melitensis*	Portugal	2014	Homo sapiens	ERS2952768	ERR2938674
194Pa	*Brucella suis*	Portugal	2011	Swine	ERS2952792	ERR2938675
180P	*Brucella melitensis*	Portugal	2014	Homo sapiens	ERS2952773	ERR2938676
1P	*Brucella melitensis*	Portugal	2010	Homo sapiens	ERS2952761	ERR2938677
146-10RK	*Brucella melitensis*	Spain	2010	Unknown	ERS2952749	ERR2938678
165P	*Brucella melitensis*	Portugal	2011	Homo sapiens	ERS2952769	ERR2938679
258P	*Brucella melitensis*	Portugal	2016	Homo sapiens	ERS2952782	ERR2938680
104-11RK	*Brucella melitensis*	Germany	2011	Unknown	ERS2952746	ERR2938681
36P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952763	ERR2938682
44-07E	*Brucella melitensis*	Spain	2007	Homo sapiens	ERS2952743	ERR2938683
169P	*Brucella melitensis*	Portugal	2014	Homo sapiens	ERS2952770	ERR2938684
40P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952793	ERR2938685
146-11RK	*Brucella abortus*	Spain	2011	Unknown	ERS2952785	ERR2938686
463Pa	*Brucella melitensis*	Portugal	2005	Sheep	ERS2952756	ERR2938687
723-07E	*Brucella melitensis*	Spain	2007	Homo sapiens	ERS2952745	ERR2938688
166P	*Brucella melitensis*	Portugal	2011	Homo sapiens	ERS2952796	ERR2938689
41P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952765	ERR2938690
238-04E	*Brucella melitensis*	Spain	2004	Homo sapiens	ERS2952741	ERR2938691
204-01E	*Brucella melitensis*	Spain	2001	Homo sapiens	ERS2952738	ERR2938692
228P	*Brucella melitensis*	Portugal	2016	Homo sapiens	ERS2952780	ERR2938693
228-03E	*Brucella melitensis*	Spain	2003	Homo sapiens	ERS2952740	ERR2938694
918Pa	*Brucella melitensis*	Portugal	2011	Goat	ERS2952760	ERR2938695
200P	*Brucella melitensis*	Portugal	2015	Homo sapiens	ERS2952778	ERR2938696
4Pa	*Brucella suis*	Portugal	2000	Swine	ERS2952788	ERR2938697
167-00E	*Brucella melitensis*	Spain	2000	Homo sapiens	ERS2952735	ERR2938698
194-00E	*Brucella melitensis*	Spain	2000	Homo sapiens	ERS2952737	ERR2938699
184P	*Brucella melitensis*	Portugal	2014	Homo sapiens	ERS2952774	ERR2938700
183-4RK	*Brucella melitensis*	Hungary	2004	Unknown	ERS2952752	ERR2938701
43P	*Brucella melitensis*	Portugal	2012	Homo sapiens	ERS2952766	ERR2938702
804Pa	*Brucella melitensis*	Portugal	2008	Bovine	ERS2952759	ERR2938703
209P	*Brucella melitensis*	Portugal	2015	Homo sapiens	ERS2952779	ERR2938704
198P	*Brucella melitensis*	Portugal	2015	Homo sapiens	ERS2952776	ERR2938705
194P	*Brucella melitensis*	Portugal	2015	Homo sapiens	ERS2952775	ERR2938706
MLVA01	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983828	ERR2993143
MLVA02	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983829	ERR2993144
MLVA03	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983830	ERR2993162
MLVA04	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983831	ERR2993153
MLVA05	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983832	ERR2993139
MLVA06	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983833	ERR2993160
MLVA07	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983834	ERR2993142
MLVA08	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983835	ERR2993136
MLVA09	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983836	ERR2993147
MLVA10	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983837	ERR2993149
MLVA11	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983838	ERR2993150
MLVA12	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983839	ERR2993161
MLVA13	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983840	ERR2993154

Strain	Species	Isolation Country	Collection_date	Host	Biosample	Run
MLVA14	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983841	ERR2993133
MLVA16	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983842	ERR2993159
MLVA17	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983843	ERR2993141
MLVA18	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983844	ERR2993132
MLVA19	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983845	ERR2993146
MLVA20	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983846	ERR2993135
MLVA22	*Brucella spp*	Germany	Unknown	Homo sapiens	ERS2983847	ERR2993145
MLVA23	*Brucella spp*	Germany	Unknown	Homo sapiens	ERS2983848	ERR2993138
MLVA24	*Brucella spp*	Germany	Unknown	Homo sapiens	ERS2983849	ERR2993152
MLVA25	*Brucella spp*	Germany	Unknown	Homo sapiens	ERS2983850	ERR2993155
MLVA28	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983851	ERR2993137
MLVA29	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983852	ERR2993163
MLVA30	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983853	ERR2993151
MLVA31	*Brucella spp*	Germany	Unknown	Homo sapiens	ERS2983854	ERR2993140
MLVA32	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983855	ERR2993157
MLVA33	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983856	ERR2993148
MLVA34	*Brucella spp*	Germany	Unknown	Homo sapiens	ERS2983857	ERR2993158
MLVA35	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983858	ERR2993134
MLVA36	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983859	ERR2993156
MLVA37	*Brucella melitensis*	Germany	Unknown	Homo sapiens	ERS2983860	ERR2993131