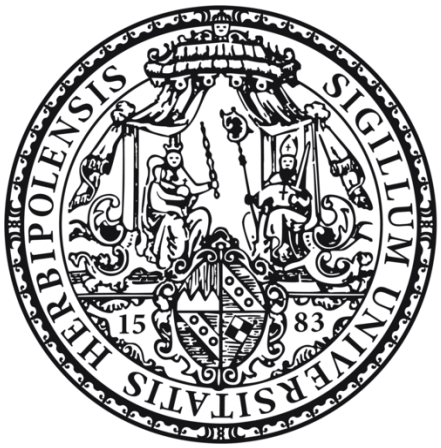


Genomic and internet based analysis of *Coxiella burnetii*

Genomische und Internet-basierte Analyse von *Coxiella burnetii*

Doctoral thesis for a doctoral degree

at the Graduate School of Life Sciences,

Julius-Maximilians-University Würzburg,

Section Infection and Immunity

submitted by

Akinyemi Mandela Fasemore

from

Lagos, Nigeria

Würzburg 2022

Submitted on: ... ...

Office stamp

Members of the *Promotionskomitee*:

Chairperson: **Prof. Jörg Schultz**

Primary Supervisor: **Prof. Dr. Thomas Dandekar**

Supervisor (Second): **Prof. Dr. Konrad Förstner**

Supervisor (Third): **PD Dr. Dimitrios Frangoulidis**

Date of Public Defence: ... ...

Date of Receipt of Certificates: ... ...

Affidavit

I hereby confirm that my thesis entitled "Genomic and internet based analysis of *Coxiella burnetii*" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and /or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation "Genomische und Internet-basierte Analyse von *Coxiella burnetii*" eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum Unterschrift

SUMMARY

Coxiella burnetii, a Gram negative obligate intracellular bacterium, is the causative agent of Q fever. It has a world wide distribution and has been documented to be capable of causing infections in several domestic animals, livestock species, and human beings. Outbreaks of Q fever are still being observed in livestock across animal farms in Europe, and primary transmission to humans still occurs especially in animal handlers. Public health authorities in some countries like Germany are required by law to report human acute cases denoting the significance of the challenge posed by *C. burnetii* to public health.

In this thesis, I have developed a platform alongside methods to address the challenges of genomic analyses of *C. burnetii* for typing purposes. Identification of *C. burnetii* isolates is an important task in the laboratory as well as in the clinics and genotyping is a reliable method to identify and characterize known and novel isolates. Therefore, I designed and implemented several methods to facilitate the genotyping analyses of *C. burnetii* genomes *in silico* via a web platform. As genotyping is a data intensive process, I also included additional features such as visualization methods and databases for interpretation and storage of obtained results. I also developed a method to profile the resistome of *C. burnetii* isolates using a machine learning approach. Data about antibiotic resistance in *C. burnetii* are scarce majorly due to its lifestyle and the difficulty of cultivation in laboratory media. Alternative methods that rely on homology identification of resistance genes are also inefficient in *C. burnetii*, hence, I opted for a novel approach that has been shown to be promising in other bacteria species. The applied method relied on an artificial neural network as well as amino acid composition of position specific scoring matrix profile for feature extraction. The resulting model achieved an accuracy of ≈ 0.96 on test data and the overall performance was significantly higher in comparison to existing models. Finally, I analyzed two new *C. burnetii* isolates obtained from an outbreak in Germany, I compared the genome to the RSA 493 reference isolate and found extensive deletions across the genome landscape.

This work has provided a new digital infrastructure to analyze and characterize *C. burnetii* genomes that was not in existence before and it has also made a significant contribution to the existing information about antibiotic resistance genes in *C. burnetii*.

ZUSAMMENFASSUNG

Coxiella burnetii, ein Gram-negatives, obligat intrazelluläres Bakterium, ist der Erreger des Q-Fiebers. Er hat eine weltweite Verbreitung und ist nachweislich in der Lage, Infektionen bei verschiedenen Haustieren, Nutztieren und Menschen zu verursachen. Ausbrüche von Q-Fieber werden immer noch in Tierbeständen in Europa beobachtet, und die Primärübertragung auf den Menschen erfolgt nach wie vor allem durch Kontakt mit entsprechenden Tieren und ihren Ausscheidungen. Das öffentliche Gesundheitssystem in einigen Ländern wie Deutschland hat eine Meldepflicht für akute Fälle beim Menschen festgelegt, was die Bedeutung des Erregers bzw. seiner ausgelösten Erkrankung für die öffentliche Gesundheit verdeutlicht. In dieser Doktorarbeit habe ich eine Plattform neben weiteren Methoden entwickelt, um die Herausforderungen der Genomanalyse von *C. burnetii* für Genotypisierungsverfahren zu adressieren. Die Identifizierung von *C. burnetii*-Isolaten erfüllt eine wichtige Funktion im Labor sowie in den Krankenhäusern, und die Genotypisierung ist eine verlässliche Methode, um bekannte und neue Isolate zu identifizieren und zu charakterisieren. Daher habe ich mehrere Methoden konzipiert und implementiert, um die Analyse zur Genotypisierung von *C. burnetii*-Genomen in silico über eine Web-Plattform zu erleichtern. Da die Genotypisierung ein datenintensiver Prozess ist, habe ich ebenfalls zusätzliche Features wie Visualisierungsmethoden und Datenbanken zur Interpretation und Speicherung der erhaltenen Ergebnisse mitaufgenommen. Ferner habe ich eine Methode zur Erstellung des Resistomprofils von *C. burnetii*-Isolaten unter Verwendung eines Ansatzes des maschinellen Lernens entwickelt. Daten über Resistenzfaktoren bei *C. burnetii* sind rar, was hauptsächlich auf die obligat intrazelluläre Lebensweise der *Coxiellen* und die Schwierigkeiten bei der Kultivierung in Labormedien zurückzuführen ist. Alternative Methoden, die auf der Identifizierung der Homologie von Resistenzgenen basieren, sind bei *C. burnetii* ebenfalls ineffizient. Aus diesem Grund entschied ich mich für einen neuen Ansatz, der sich bereits bei anderen Bakterienspezies als vielversprechend erwiesen hat. Die verwendete Methode basiert auf einem artifiziellen neuronalen Netzwerk sowie auf der Aminosäurezusammensetzung des positionsspezifischen Matrixprofils zur Extraktion von Features. Das daraus resultierende Modell erzielte eine Genauigkeit von $\approx$ 0,96 bei den Testdaten und die Gesamtleistung war signifikant höher im Vergleich zu den bereits vorhandenen Methoden. Schließlich analysierte ich zwei neue *C. burnetii*-Isolate, die von einem Q-Fieberausbruch in Deutschland stammten. Ich verglich das Genom mit dem RSA 493 Referenz Isolat und fand extensive Deletionen über das Genom sequenz. Mit dieser Arbeit wird eine neue digitale Infrastruktur zu Analyse von *C. burnetii*- Genomen bereitgestellt, die es vorher noch nicht gab. Zudem liefert diese Arbeit einen wichtigen Beitrag zu den bereits vorhandenen Informationen über Antibiotikaresistenzgene bei in *C. burnetii*.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

API	Application Programming Interface
UI	User Interface
MST	Multispacer sequence typing
MLVA	Multiple-Locus Variable Number Tandem Repeat Analysis
LCV	Large Cell Variant
SCV	Small Cell Variant
PFGE	Pulsed-field gel electrophoresis
PCR	Polymerase chain reaction
CDC	Centers for Disease Control and Prevention
AIDS	Acquired Immune Deficiency Syndrome
DNA	Deoxyribonucleic acid
NGS	Next generation sequencing
NCBI	National Center for Biotechnology Information
melt-MAMA	Melt Analysis of Mismatch Amplification Mutation Assays
SRA	Sequence Read Archive
CDS	Coding Sequence
SDS-PAGE	Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis
GG	Genomic Groups
RFLP	Restriction Fragment Length Polymorphisms
rRNA	Ribosomal Ribonucleic Acid
rDNA	Ribosomal Deoxyribonucleic acid
CARD	Comprehensive Antibiotic resistance database
ARDB	Antibiotic Resistance Genes Database
ARG	Antibiotic Resistance Genes
COFRADIC	Combined Fractional Diagonal Chromatography

WGA Whole Genome Amplification

HTML Hypertext Markup Language

CSS Cascading style sheet

CGI Common Gateway Interface

MVC Model View Controller

HTTP Hyper Text Transfer Protocol

WSGI Web Server Gateway Interface

SNP Single Nucleotide Polymorphism

Part I

INTRODUCTION

1 INTRODUCTION

1.1 ORIGINS AND CLASSIFICATION OF *Coxiella burnetii*

In the middle of summer in the year 1935, over 800 employees of a large meat works in Brisbane, Australia had been diagnosed with an atypical fever condition. The fever was described as "a continued one with a duration between seven to twenty-four days". The Director of Health and Medical services for Queensland was puzzled and he sent a doctor named E.H. Derrick to investigate the matter. The mystery of this illness was further confounded when the standard procedures of diagnosis back then, blood culture tests and agglutination tests, failed to detect any disease [40].

Since they could not match the symptoms to any known disease, it was apparent that this was a novel type of clinical entity and as such a name was needed. There were suggestions like "abattoir fever" and "Queensland rickettsial fever", however they were considered to imply a negative representation of Queensland cattle industry. They called the disease Q fever, the "Q" was an acronym for "query" since the agent was not known [100].

The pathological agent of this disease would remain a mystery until 1937, when Dr F. M. Burnet and another colleague Mavis Freeman discovered rickettsial bodies in the spleen of infected mice. These mice were inoculated with an emulsion of the liver from an infected guinea-pig that was sent by Dr. E. H. Derrick. Unfortunately, they hypothesized incorrectly (with the benefit of hindsight) that the agent was a virus since it could pass through a pore/filter membrane [24].

Earlier in the spring of 1935 also, two American bacteriologists, G. E. Davis and H. R. Cox, reported to have recovered a filter-passing infectious agent from a group of 200 *Dermacentor andersoni* ticks collected near Nine Mile creek in Montana, USA. These ticks were subdivided into four groups of 50 individuals each and were fed with the blood of a guinea pig through a feeding capsule that was clipped to the belly of the guinea pig. During 3 weeks all the guinea pigs died of unknown causes and as such prompted an investigation into the cause of death. They discovered the agent was of rickettsial origins and could pass through a filter. They obtained samples from Queensland and found that the agent of the Australian Q fever and their own agent which they called "X virus" might be closely related. They also reported that one laboratory technician that worked on maintenance of the guinea pigs in Montana also felt ill showing clinical features that were similar to that of the abattoir workers in Queensland, suggesting the relationship of this infection to Q fever. The only difference was that they could not reproduce the infection in monkeys using X virus as reported by Burnet and Freeman with the agent from Q fever. Nevertheless, they argued that the mere ability of these agents to pass through a filter cannot

solely justify its classification as a virus and suggested it should be classified as a "bartonella" [34].

Eventually, a species name *Rickettsia diaporica* was proposed by the team of bacteriologists working at the Rocky mountain laboratory, suggesting a member of the Rickettsiae genus that can pass through pores. E. H Derrick proposed another name which was *Rickettsia burnetti*, with the objective to recognize the foundational works of Dr. F. M. Burnet on the Q fever agent. Research work on the agent was intensified in the following years and it was eventually discovered that is was significantly different from other Rickettsia members prompting the suggestion of Coxiella as a new genera in 1948 by Cornelius B. Philip and a recognition of the works of H.R Cox [100].

Although initially hypothesized to belong to the *Rickettsiaceae* family, *C. burnetii* is more closely related to *Legionella spp* and has been taxonomically classified to belong to the gamma subdivision of Proteobacteria under the family Coxiellaceae [63].

1.2 MICROBIOLOGY OF *C. burnetii*

C. burnetii is a coccobacillus, 0.2 to 0.4 µm wide and 0.4 to 1 µm long, that cannot be stained with the Gram technique in clinical samples or cell culture. It can however be stained with a special staining technique called Gimenez staining. Nevertheless, it has been classified as Gram negative bacteria because it possesses an envelope that is similar to the ones found in Gram negative bacteria [53] [44] [25].

It is an obligate intracellular pathogen that is capable of replicating in a broad range of hosts including monkeys, guinea-pigs, mice, fish, ticks, humans and livestock [24] [63]. It has been shown to be highly resistant to adverse physical conditions of high acidity, high temperature and chemical disruptions [63].

C. burnetii can be observed in two forms as a result of its two phase developmental cycle, the exponentially replicating large cell variant form (LCV), and the stationary non replicating small cell variant form (SCV) [102] [44] [142]. Both forms can be separated *in vitro* based on their densities, when subjected to centrifugation in cesium chloride [63].

The SCV are smaller in size compared to the LCV and are characterized by a condensed chromatin, that appears as a thick dark spot under an electron microscope unlike the LCV that have a dispersed chromatin and of a lesser shade under the electron microscope as shown in Figure 1.1.

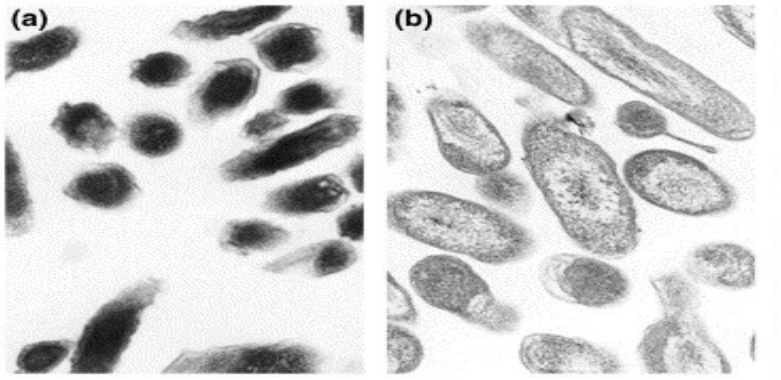


Figure 1.1: Electron micrograph of SCV (a) and LCV (b).
Original image is sourced from [63]

A brief summary of the pathogenesis of *C. burnetii* involves entry into the host cells typically via alveolar openings [108], although both SCV and LCV forms have been shown to be infective *in vivo*, it is generally considered that the infectious agent is in the SCV form [108] [102]. The SCV gets engulfed passively into a phagolysosome-like chamber called a parasitophorous vesicle (PV) [108]. Interactions between the PV and the host machinery, autophagosomes and endolysosomes, ensures it is provided with the necessary nutrients and the SCV transforms into the more metabolic active LCV. The LCV replicates intracellularly with a generation time of roughly eleven hours [30]. After about five to six days, the LCVs transform reversely to SCV and are released after the lysis of the host cell [108] [102].

In 2003, Seshadri *et al*. published the first genome of a *C. burnetii* strain, the above mentioned tick isolate from the Nine Mile creek [148]. Before then, Williams *et al*. using Pulsed-field gel electrophoresis (PFGE) and Polymerase chain reaction (PCR) were able to describe the physical and genetic map of *C. burnetii* [179]. Both studies established the groundwork for genetic investigations in *C. burnetii*. The genomes of *C. burnetii* strains are roughly 2 Mb in size with an approximate coding capacity of 90 percent and the average GC content is approximately 42.5 percent [108]. All *C. burnetii* isolates contains one of the four autonomously replicating plasmid types (QpH1, QpRS, QpDV and QpDG) identified so far [78] [181] [88] [108]. Plasmidless strains with plasmid homologues to QpH1 and QpR inserted into the chromosome have also been reported [180] .

1.3 THE ROLES OF *C. burnetii* IN THE EPIDEMIOLOGY OF Q FEVER

The first two subsections were intended to give a condensed exposè to the history and microbiology of *C. burnetii*, the aetiological agent of Q fever. In this subsection I will try to elucidate the relationship between the agent and the disease in an epidemiological context. The aim is to highlight the public health significance of this bacteria.

Q fever is a zoonosis that is distributed worldwide, the implication of the leading statement is that the reservoir for the aetiological agent is extensive and global [68]. Several important reservoirs have been identified up to date including cats [67], livestock such as cattle, goats and sheep [3] and wild rodents [175]

Infection in humans is often as a result of exposure to and inhalation of contaminated aerosols from bio materials such as placentas and amniotic fluid of infected livestock [68]. Birth products usually contain a high number of infective particles compared to excretory products [124]. Other sources such as consumption of unpasteurized milk from infected livestock [45], sexual transmission and direct transmission via intradermal inoculation and blood transfusion[135] have also been documented. Sharon *et al*. also reported a case of suspected transmission via vaginally excreted infectious placenta between two pregnant women [7]. The ability of *C. burnetii* to engage different vehicles of transmission can be attributed to its stability in the environment (even for extended time periods) and its ability to be highly resistant to severe conditions such as osmotic, mechanical and chemical stress especially in the SCV phase [44]. These features as well as its high infectivity (infectious dose less than 10 bacteria) significantly boost the ability of *C. burnetii* to establish an infection in a wide range of host and in minimal time leading to outbreaks, and also make it's suitable as a potential grade B biological weapon agent according to the CDC [44]. It's also of note that *C. burnetii* can persist asymptomatically in humans for a life time period, however, the infection may be reactivated by an immune deficiency such as AIDS [133], pregnancy, cardiac valvular abnormality, a vascular aneurysm or prosthesis, hemodialysis [90] [68].

Two types of Q fever infections have been documented in humans, acute Q fever and chronic Q fever [133]. The acute case is characterized by symptoms such as febrile, illness, headaches and pneumonia. It is usually self-limiting and has been reported to manifest in 40% of Q fever cases [106]. The chronic course (up to 2% of acute cases) is characterized by a persistent infection accompanied usually by endocarditis, which can be life-threatening [79].

The difference observed in the clinical manifestation of Q fever was initially attributed to genomic differences in the offending strain. First, a correlation with plasmid types was suggested [181] [140], then it was hypothesized that it was a result of the deletion of the *ada*A gene in certain strains [188], however both hypotheses have been refuted [47] [157] [161]. Raoult *et al*., [44] attributed the observed clinical differences to host factors such as age, immunity, pregnancy and genetic factors.

1.4 GENOMIC ANALYSIS OF *C. burnetii*

Genomic analysis can be defined as the identification, measurement or comparison of genomic features such as DNA sequence, structural variation, gene expression, or regulatory and functional element annotation at a genomic scale. Methods for genomic analysis typically require high-throughput sequencing or

microarray hybridization and bioinformatics [191]. This definition highlights in clear terms the multi objectives of genomic analysis. Microarray hybridization techniques, although not obsolete, are gradually becoming relegated [99].

Genomic analysis of *C. burnetii* like every other bacteria is usually objective oriented, the applied methods and tools are determined mostly by the aim of the analysis. The scope of this project covers three segments of *C. burnetii* genomic analysis: Analysis of whole genomic sequences, genomic analysis for typing purposes, genomic analysis for the identification of resistance genes.

1.4.1 *Analysis of whole genomic sequences*

The first analysis of *C. burnetii* sequences dates back to 1985, when Samuel and colleagues used restriction enzymes to compare the homology between several disease isolates and the QpH1 plasmid and also compared the QpH1 plasmid to the QpRS plasmid [140]. One year after the Chain termination method was commercialized, Thiele *et al.* applied this to sequence the entire QpH1 plasmid describing its size and the number of open reading frames present within the sequence [162].

Since the earlier studies, which were based on restriction enzymes and DNA patterns, genomic analysis of *C. burnetii* has undergone a major shift in paradigm. The two major drivers of this shift are improvement in sequencing technology and *C. burnetii* culturing methods. The introduction of a new generation of pyrosequencing technology in 2004 by 454 Life Sciences [118] and the subsequent application to the human genome in 2008 [173], completely revolutionized the field of genome sequencing with a higher throughput and less error rate. Although the costs were initially exorbitant, it has been falling since [176]. This has led to an increase in the total number of sequenced isolates deposited at the genome database of the National Center for Biotechnology Information (NCBI). The plot in Figure 1.2 shows the cumulative total number of *C. burnetii* genomes deposited per year from 2007 to 2022, in total, there are 83 genomes deposited as of 2022-02-12 and NCBI classifies the genomes into several assembly nomenclature types as shown in Figure 1.3.

DNA material for sequencing experiments is usually obtained from cultures of *C. burnetii* which were grown in a laboratory under axenic conditions after isolation from disease or non-disease sources. *C. burnetii* is an obligate intracellular bacteria that has to be grown under biosafety level 3 (BSL-3) conditions, this increases the level of handling complexity[70] [148] [64]. *C. burnetii* can be grown on eukaryotic cells like Buffalo Green Monkey cells with essential media and nutrient supplements [48] [174]. Hemsley *et al.*, 2019 introduced a different method where they isolated DNA directly from tissue samples using magnetic beads [64] without prior cultivation.

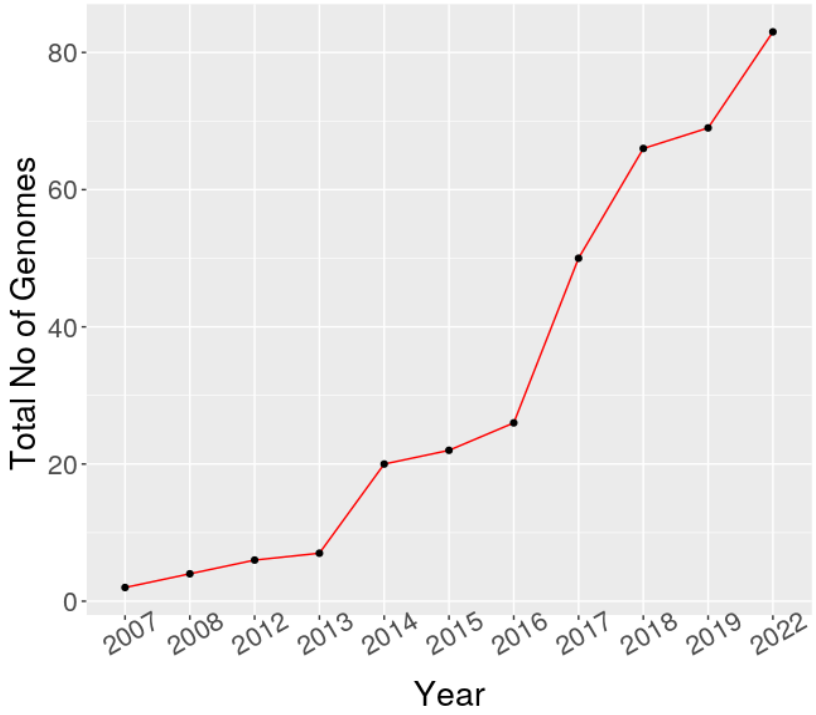


Figure 1.2: *C. burnetii* genomes deposited per year in NCBI database

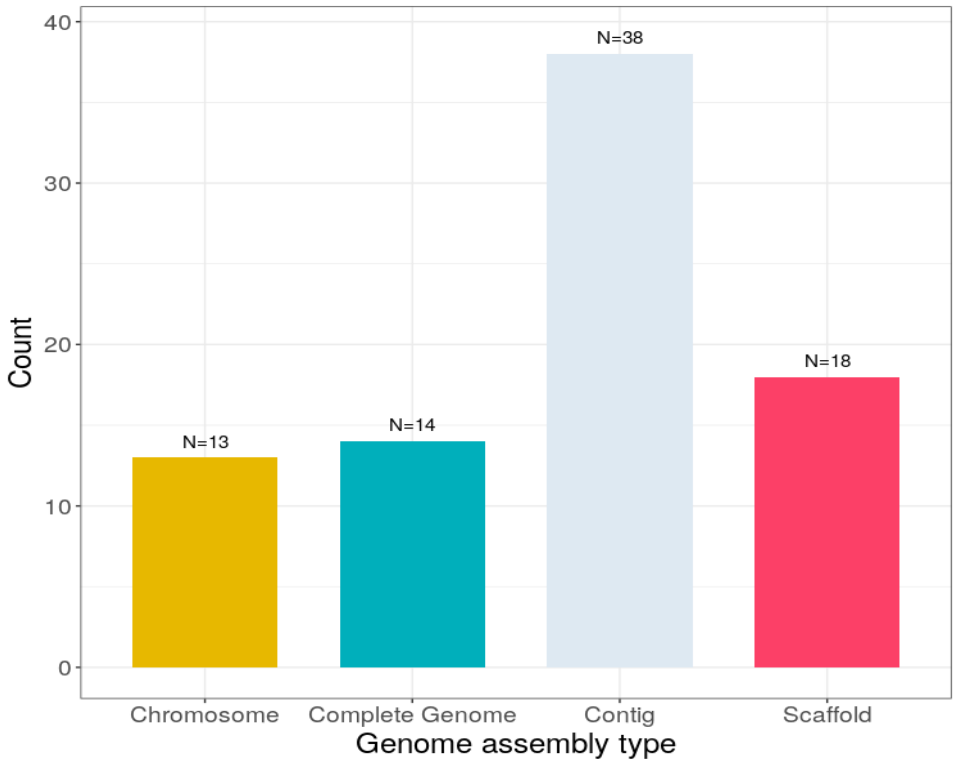


Figure 1.3: *C. burnetii* genomes assembly types on NCBI

Confirmation of *C. burnetii* DNA sequence is done by targeting the *com1* gene [114] or targeting the IS1111 transposon-like repetitive region [143]. PCR protocols like multiplex-PCR assays [122] [48], melt-MAMA [70] have also been

used to rapidly amplify *C. burnetii* DNA materials from various sources, or detect certain genomic features [48].

In order to obtain whole genomic sequences, the extracted DNA will have to be sequenced and then assembled. The major sequencing platforms are Illumina, Pacific Biosciences (PacBio), Ion Torrent and Oxford Nanopore [129]. The plot in Figure 1.4 shows the usage frequency of the different sequencing platforms for *C. burnetii* experiments

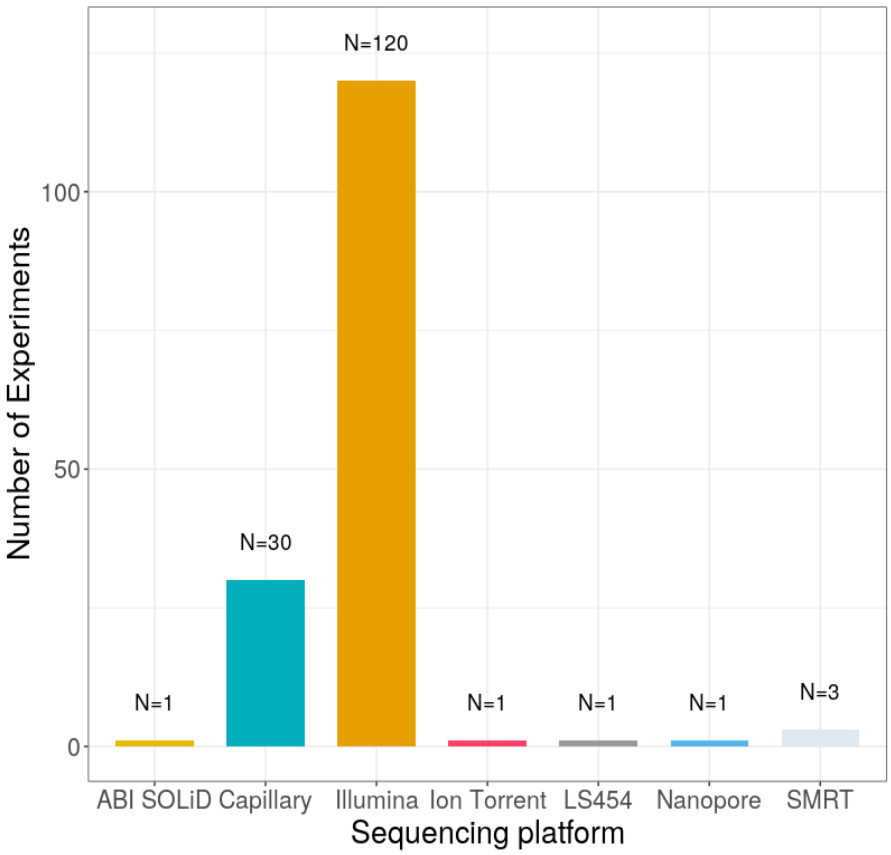


Figure 1.4: The usage of sequencing platforms for *C. burnetii* sequencing experiments on NCBI SRA [89]

After sequencing, the obtained reads will have to be assembled. The sequencing platform determines the nature of the reads. Sequence assembly has been described as the problem of reconstructing a string entity from its set of k-mers [13]. The major challenges include ensuring assembly quality, high computer memory requirement and execution time [27]. Several solutions for overcoming this are offered in the numerous assemblers including but not limited to SPAdes [13], Jira [27], Velvet [186]. Two different approaches for sequence assembly exist. The first is de-novo genome assembly, this approach involves assembling short reads in the absence of a template [182]. The second approach is comparative genome assembly, this method involves assembling short reads based on a template reference sequence [123].

For *C. burnetii* sequence assembly, both approaches have been applied [174] [148]. Several repetitions of the assembly process will usually be carried out to find the optimal parameters so as to obtain as few contigs as possible with

the least amount of errors [174]. Parameters such as N50, GC Content and GC skew can be used as indicators to estimate assembly coherence [57]. GC content for *C. burnetii* is usually between 40-45% [138]

Identifications of genes and putative open reading frames that encodes proteins (CDS) usually is the first step in *C. burnetii* genomic analysis once a whole sequence is available. Several prediction tools exist for this process, Prodigal [71], Glimmer [37], Prokka [145], GeneMarkS [20] e.t.c. Validation of predicted protein can be done by searching against the non-redundant database from NCBI [127] or the Uniprot database [9], predicted sequences that do not meet a certain length or judged to contain erroneous mutations may be filtered in this step [148]. Furthermore, functional proteins such as Signal peptides and putative membrane protein domains are predicted with SIGNALP [16] and TOPPRED [28]

After identification of putative genes, Structural RNAs like rRNA and tRNAs are identified. Tools like RNAHmmer [85], Barrnap [146] and tRNAscan-SE [95] can be used for this purpose.

Prodigal [71], Glimmer [37], Prokka [145], GeneMarkS [20], RNAHmmer [85] and tRNAscan-SE [95] have all been applied to predict genomic features in *C. burnetii* [174] [148].

The earliest classification of *C. burnetii* isolates into groups was carried out by Hendrix *et al.*, [65] based on restriction endonuclease digestion of chromosomal DNA followed by separation on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). They subdivided *C. burnetii* isolates into 6 genomic groups (GG I- GG VI). They showed the correlation of the groups with plasmid type as presented in Table 1.1. This classification is still relevant for current *C. burnetii* literature. It has been updated to include added information as discussed below.

Hornstra *et al.*, 2011 [70] extended the information of these groups with MST genotypes, they showed that GG I group contains three main MST groups 26, 16 and 17 and GG III is occupied by two main MST groups 20 and 19. as detailed in Table 1.1.

Several other works have extended these six groups with meta information. Beare *et al.*, 2006 [15] proposed that aside from the Nine-Mile reference strain, GG group I isolates can be found across the world whereas GG II isolates are mostly implicated in European Q fever outbreaks [164].

Table 1.1: Genomic grouping of *C. burnetii*

Genomic group	Plasmid type	Region	MST group	Disease status
I	QpH1	Worldwide	26, 16, 17	
II	QpH1	Mostly European	34, 24, 32, 12, 11, 13, 14, 15, 33, 22, 23, 29, 25, 18	
III	QpH1	European, American	20, 19	
IV	QpRS	African, Russian, American	8, 9, 10, 27, 28, 31, 5, 6, 7, 1, 3, 2, 4, 30	
VI	QpDG	American	60	

1.4.2 *Genomic analysis of C. burnetii for typing purposes*

Genetic distinction between *C. burnetii* strains is epidemiologically important for investigating the source of an outbreak [101]. Several features in the genome of *C. burnetii* strains such as repeat sequences, insertion sequences, spacer regions, genes and plasmids have been exploited to discriminate among closely related *C. burnetii* strains with success. Massung et al., 2012 categorized the characterization methods for *C. burnetii* under two main headings "Typing methods Used before 2005" and "High resolution Post-2005 Typing" methods [101].

1.4.2.1 *Historical methods for genotyping*

The earlier methods were based on PCR detection of single genome targets like mucZ, icd, com1 and 16s/23s rRNAs, plasmid types and restriction fragment length polymorphisms (RFLP) analyzed with SDS-PAGE. They were generally challenged by inter and intra laboratory reproducibility and often discriminatory power was not convincing.

SINGLE GENE TARGET - MUCZ, ICD, COM1

The com1 gene was the first genomic target that was used to characterize *C. burnetii* isolates. 21 isolates from different human and animal sources were characterized under four groups [187]. Sekeyová *et al.*, also applied this method to characterize 37 isolates, they classify them under five groups confirming

earlier results from Zhang *et al*. [147]. mucZ is a 715 bp gene, Sekeyová *et al*. characterized four positions where mutations had occurred and exploited this information to discriminate 31 *C. burnetii* isolates under four groups [147]

Nguyen and Hirai, 2006 exploited the isocitrate dehydrogenase (icd) gene to discriminate 19 *C. burnetii* isolates into three groups using PCR-restriction fragment length polymorphism [113].

16S/23S RRNAS

rRNA regions have not been particular successful in discriminating between *C. burnetii* strains as shown by two separate studies: Stein *et al*. 1993 and Stein *et al*. 1997. The first study used a 1,418 bp fraction of the 16S-rRNA to differentiate between six isolates of *C. burnetii*, achieving only a slight variation which failed to produce any epidemiological or clinical correlation. The second study exploited the internal transcribed 16S–23S rDNA spacer (ITS) region to evaluate 22 different *C. burnetii* isolates. The obtained results were comparable to the first suggesting the inability of this approach to discriminate *C. burnetii* strains [158] [159] [101].

RFLP/PFGE

Restriction fragment length polymorphism analysis using SDS-PAGE was successfully used to classify *C. burnetii* strains into six groups (I-VI) in 1991 by Hendrix and colleagues, after they digested total chromosomal DNA with *Eco* RI and *Bam* HI restriction endonuclease [65]. This classification correlated with that of plasmid typing, other restriction enzymes such as *Not* I and SFiI [62] [78] also have been applied to distinguish larger collections of *C. burnetii* isolates [101].

PLASMID TYPE

C. burnetii strains carry five different plasmid types: four extra chromosomal plasmid types (QpH1, QpRS, QpDV, and QpDG) and strains that have plasmids integrated within its chromosome. [163] [168] [66] [151]. Based on this, *C. burnetii* strains have been classified into 5 genomic groups [101].

1.4.2.2 *Modern methods for genotyping*

The availability of cheaper sequencing technologies, falling low cost of computing power, and the invention of more efficient culture techniques has led to the development of a new set of highly discriminative, cost efficient and high resolution genotyping systems to identify and characterize *C. burnetii* isolates [101][185].

IS1111

The IS1111 element is an insertion sequence that is found in multiple copies in the genome of several bacterial genera including *Coxiella*. The idea to use these sequences to differentiate between *Coxiella* species was first proposed by Hoover *et al*., in 1992 [69]. The diversity and uniqueness of the adjacent regions

to the various copies of these elements in *Coxiella* species was first exploited as a typing scheme by Denison *et al.*, 2007 [39]. Bleichert *et al.*, 2018 (unpublished) extended the 10 previously defined IS1111 positions with 22 novel positions.

MLVA

Multi Locus VNTR Analysis (MLVA) is a molecular genotyping method that exploits variability in the copy numbers of Tandem Repeats [26]. It's widely used to perform molecular typing of pathogenic species e.g *Mycobacterium spp* for epidemiological purposes. Its origins can be traced to the field of forensic science where it was used for DNA fingerprinting of samples from humans [26].

Repeat units used for MLVA in *C. burnetii* can be classified under two categories based on the length of their base pairs, minisatellites, these are usually above 9 base pairs, and microsatellites, these are usually between 6 to 7 base pairs [11].

Currently there are 17 markers for MLVA typing based on the RSA 493 genome [11].

The combination of these markers(panels) have been used in various MLVA genotyping projects as highlighted in Table 1.2

Table 1.2: Table of MLVA markers

Marker name			
Arricau	Frangoulidis	Svraka	Unit
Cbu0033_ms01_17bp_4U_248bp	ms01		4
Cbu0448_ms03_12bp_7U_229bp	ms03		7
Cbu0988_ms07_126bp_8U_1112bp			5-8
Cbu1316_ms12_126bp_8U_1074bp			4-9
Cbu1941_ms20_18bp_15U_402bp	ms20		9
Cbu1963_ms21_12bp_6U_210bp	ms21		6
Cbu1980_ms22_11bp_6U_246bp	ms22		6-7
Cbu0831_ms26_9bp_4U_127bp	ms26	Cox3	4-5
Cbu1351_ms30_18bp_6U_215bp	ms30		6
Cbu1941_ms36_9bp_4U_447bp			4
Cbu0197_ms23_7bp_8U_157bp	ms23	Cox6	8-9
Cbu0259_ms24_7bp_27U_344bp	ms24	Cox4	27-30
Cbu0838_ms27_6bp_4U_276bp	ms27	Cox2	6-7
Cbu0839_ms28_6bp_6U_276bp	ms28	Cox5	6-7
Cbu1418_ms31_7bp_5U_182bp	ms31	Cox7	6-7
Cbu1435_ms33_7bp_9U_262bp	ms33		9
Cbu1471_ms34_6bp_5U_210bp	ms34		6

MST

Multispacer sequence typing (MST), is based on sequencing of intergenic regions in *C. burnetii* based on the assumption that these regions are under lower selection pressure [55]. The MST genotyping method utilizes 10 spacers with an average of 13 alleles each between 300 - 700bp [55]. Analysis of results can be done through a website (http://ifr48.timone.univ-mrs.fr) where all the allele information and profiles are presented.

*ada*A GENE

The acute disease antigen A (*ada*A), is a 28kDa protein that was first identified in isolates from patients with acute Q fever [165]. Initially proposed as a diagnostic marker for acute Q fever by Zhang *et al*. [188], It was later refuted by a Spanish group that showed that certain strains from acute Q fever patients were *ada*A gene negative [74].

Frangoulidis *et al*., 2013 [47], showed the importance of this region as a genotyping feature for *C. burnetii* strains. They described three different *ada*A gene variants (*ada*A, *ada*A_{SNP} and *ada*A_{rep}) and two main deletion types (Q154 and Q212). They also reported a correlation between plasmid type and *ada*A genotype.

1.4.3 *Genomic analysis of C. burnetii for antibiotic resistance genes*

The manifestation of *C. burnetii* infection can be categorized under two clinical states, acute Q fever which is usually a self limiting infection characterized by mild symptoms and chronic Q fever which is characterized by a persistent infection that is usually life threatening and has been shown to be fatal in 10% of cases [44] [68].

Antibiotics are usually the first line of treatment, the Center for Disease and Control (CDC) recommends a regimen of second generation tetracycline, doxycycline, for two weeks (acute) or several months/up to years (chronic) depending on the clinical nature of the infection. In recent years, three cases of doxycycline resistant infections have been reported, two from human patients and one from goat [131] [138]. This has necessitated an interest in the *C. burnetii* genome for answers to the question of antibiotic resistance in clinical isolates.

The advent of NGS, offers new methods to look comprehensively at bacteria genomes for cues that can be used to explain certain phenotypes such as antibiotic resistance. This method is particularly relevant to *C. burnetii* because the difficulty to cultivate this organism on axenic mediums prevents the use of conventional antibiotic susceptibility methods [23]. PCR detection methods have been applied as an alternative solution in *C. burnetii* genomes. Spyridaki and colleagues investigated quinolone-resistance in 12 *C. burnetii* isolates, confirming a point mutation in the resistant isolates [156]. Another study also exploited real-time PCR to test antibiotic susceptibility of *C. burnetii* cells to chloramphenicol, rifampin, tetracycline as well as 3 other antibiotics, reporting that PCR is a viable method to test antibiotic susceptibilities in *C. burnetii* [23].

The availability of whole genomic sequences provides the opportunity for in-depth analysis of genomic sequences for resistance genes that could assist in clinical decisions [169]. The bottle neck to achieving this is the development of computational methods that are simple enough to be implementable in several species and also complex enough to command confidence during interpretation. Several approaches have been described and applied up till date with the most common being the "best hit" approach. The best hit approach relies on a database of known antibiotic resistance gene (ARG), ideally curated and well annotated. ARGs are identified or predicted based on the best hits of sequence searches against existing ARG databases [10].

Several well annotated and curated databases are available for this purpose, with the most popular being Comprehensive Antibiotic resistance database (CARD) [75] and Antibiotic Resistance Genes Database (ARDB) [92] [51]. Investigation using this method has not been successful for *C. burnetii* (data not published), however, it has been applied in the genomes of other bacterial species as well as metagenomic samples [82] [98] [184] [31].

The proteome of *C. burnetii* has been profiled under stress to three kinds of antibiotics, levofloxacin [72], tetracycline [172] and doxycycline [189] with the aim to document mechanisms of resistance. Vranakis et al. analyzed the genome of the Q212 strain using the methionine-COFRADIC procedure, where they label peptides from treated and untreated samples with slightly different chemical compounds and analyzed them using Liquid chromatography–mass spectrometry (LC-MS). They reported that 5 proteins were up-regulated upon treatment with doxycycline and 19 proteins were down-regulated, annotation of the proteins using Go-terms revealed cellular functions which were not directly associated with tetracycline resistance [172].

Zuñiga-Navarrete et al. also profiled the genome of a single *C. burnetii* isolate, RSA 331, and reported 15 proteins that were significantly altered in the presence of a doxycycline antibiotic, these sets of proteins were reported to be involved in cellular functions such as protein synthesis, cell division, maintenance of membrane integrity and pH homeostasis [189]. I et al. like the previously discussed studies also based their investigation on a single genome, Nine Mile (RSA493), they documented the differences between a levofloxacin resistant strain and a levofloxacin susceptible strain using the COFRADIC procedure [50]. They reported 2 proteins as being up-regulated in the resistant strain but found no direct evidence for the resistance state [72].

Musso et al. showed the importance of single nucleotide polymorphism in deciphering resistance phenotype in *C. burnetii*. A single SNP, Gly in place of Glu at position 87 in the GyrA gene region of *C. burnetii*, was determined to confer resistance to quinolone. They reported that this observation was only apparent for in vitro-selected high-level-resistant isolates [112].

1.5 CHALLENGES IN GENOMIC ANALYSIS OF *C. burnetii*

Challenges in genomic analysis can be grouped under two classes, wet lab challenges and dry lab challenges. Wet lab challenges are associated with laboratory procedures. Historically, the most significant wet lab challenge for genomic analysis was growing *C. burnetii* isolates outside the host cell [117]. The development of a complex axenic medium as well as its subsequent improvements has helped address this challenge, *C. burnetii* isolates can now be grown outside the host cell [117] [141] [116]. Another significant wet lab challenge is obtaining a sufficient amount of genetic material (DNA) from isolates for downstream analysis. Upstream cultivation in axenic medium only generates a limited amount of DNA, 10^9 copies/µl , which is further reduced inadvertently during the process of cell inactivation through damage. Whole genome amplification (WGA) methods such as rolling circle amplification and multiple displacement amplification have been applied in *C. burnetii* with good results [36] [35] [84] [29] although having limitations due to error rates.

This thesis focuses on the dry lab challenges associated with *C. burnetii* genome analysis. Since *Coxiella* is a pathogen, the major dry lab challenges we sought to tackle were differentiation methods among strains, isolate discovery, resistance patterns and genotyping data management.

Several methods for genotyping in *C. burnetii* have been described earlier in previous section. *in silico* implementation of all the genotyping methods are currently unavailable except for MST. *in silico* implementation of these methods will provide the opportunity to rapidly classify *C. burnetii* isolates especially in outbreak situations, and also provide an avenue for rapid data sharing especially when it's all done via a web based platform. We sought to implement such a solution for all the genotyping methods available for *C. burnetii* isolates

Isolate discovery and comparison is an essential feature of an epidemiological investigation. Historical *C. burnetii* isolates are dispersed over the literature space and collecting such data is almost an herculean task. The solution for this will be a database where historical isolates can be queried and new isolates submitted. This is already implemented for other pathogens like *S. aureus* [77]. We sought to address this challenge by providing an online resource with query and submission features also for *Coxiella*.

Investigation of resistance patterns in *C. burnetii* using the "best-hit" method has been unsuccessful (data not published). The "best-hit" method is characterized by high rate of false-negative predictions and this has been highlighted in literature [10], we sought a superior approach that overcomes the flaw of the "best-hit" method by not being completely dependent on the stringency of sequence homology, but more extensive and can account for variability in antibiotic resistance genes. Machine learning approaches such as Artificial Neural Network provide the privilege to profile sequences with higher accuracy and probability methods to estimate the accuracy of the prediction. This approach is completely new to *Coxiella* sequences, however, it has been applied in other bacteria species like *P. aeruginosa Escherichia coli* and metagenomic sequence [80]

[139] [109]. Also data from literature shows that investigation into resistance patterns in *C. burnetii* were often based on a single isolate in most cases, making comparison among several *C. burnetii* isolates impossible. We include several isolates in our resistance analysis in order to overcome this challenge.

A central repository to manage *Coxiella* typing data is still absent, although resources such as the MST and MLVA groups provides access to historical genotyping data, they fail to address several aspects of genotyping data management such as bulk uploading, access to other genotyping methods, dynamic phylogenetic tree plot and visualization of metadata. We attempt to solve this challenge by developing an umbrella online platform, where multiple genotyping data sets and approaches are housed. This is a novel approach and the advantage of this is that it offers a rapid way to use multiple genotyping methods to investigate new isolates.

1.6 WEB APPLICATIONS DEVELOPMENT

In 1991, Tim Berners-Lee invented the World Wide Web at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. The purpose of his invention was to facilitate easy document access and exchange between the databases that were available on the CERN network [18].

Web applications are applications that are accessed via web resources such as web browsers, over a network. Web applications reside on web servers unlike computer applications that are stored locally on the operating system of a device [5] [177].

Extensive reach of the World wide web has ensured that Web applications are being used in every aspect of human services such as commerce, health, education, finance, entertainment and governance. [167] [94] [137]

Historically computer applications were developed via a precompiled client program and server-side code. Updates on the server-side code required a complementary update on the client program, adding to the cost of maintenance and complexity [177]. Web applications are agnostic to this approach and hence did not suffer from the drawbacks. They are developed using standard web formats such as Hypertext Markup Language (HTML) [19] and JavaScript. Web applications are a recent human invention, however the quick adaptation and wide applications has enabled it to enjoy a front seat in technological advancement. Figure 1.6 highlights the most notable years in the history of web applications with the exclusion of HTML which was introduced in 1993 [19]

Development of web applications that meet the needs of numerous users, require a comprehensive approach that captures every requirement correctly [167]. Building such a complex application from scratch requires a significant amount of work and it's subjected to errors that might be serious. This can be avoided by using a web application framework [91].

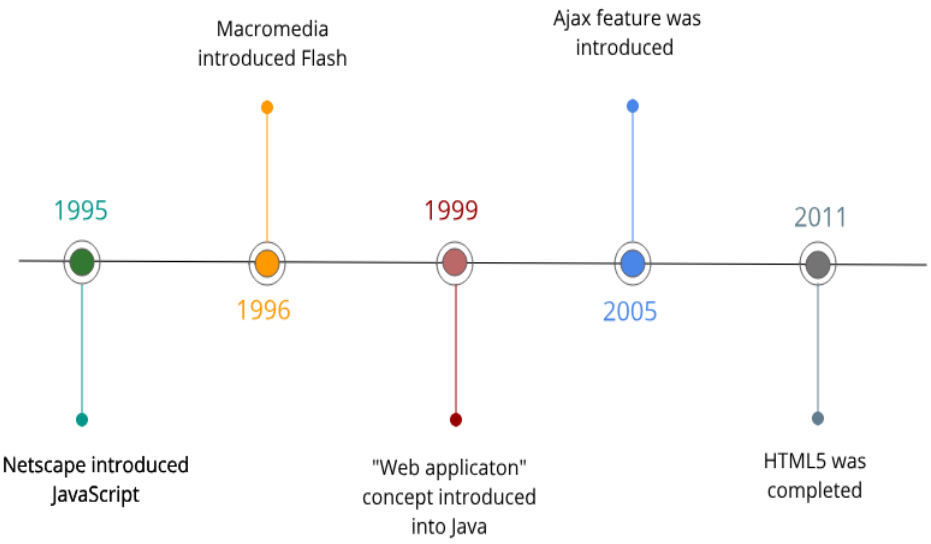


Figure 1.5: Historical timeline of the creation of web development resources [177]

Web application frameworks are software that provide a conventional way to build and deploy web applications on the world wide web using a series of foundations that can be customized by an application developer [178] [91] [76]. In 1993, the Common gateway interface (CGI) was introduced, the aim was to be able to generate dynamic content on a web page, however this implementation was fraught with adverse effect on the server load [178]. This paved the way for web specific languages like PHP, which could interact dynamically with web contents without the pitfalls of CGI [59] albeit with an aggregation of several libraries to carry out specific tasks. This was the foundation for the development of Full-stack web application frameworks, which essentially are based on a single programming language and contain a cohesive software stack that can be used to handle specific tasks in a web application [178].

The most common web framework architecture is the Model-View-Controller (MVC) architecture. First described in 1979 by Trygve Reenskaug [136], It separates the data model (Knowledge) from the view (Visual representation of the knowledge) based on actions that serve as inputs to the controller [178] [136]. There are over 100 web application frameworks presently available in different programming languages [192]. Most are open source and are community maintained, some are proprietary and the use case also differs in some stacks. Table 1.3 highlights the popular frameworks.

Table 1.3: Table of available modern Web application frameworks

Programming Language	Framework name
Python	Bottle, BlueBream, Cherrypy, CubicWeb, Django, Flask, Grok, jam.py, Google app engine, **Pyramid**, Tornado, Zope, Web2py, Zope 2
JavaScript	Angular, Backbone.js, angular.js, Express.js, Vue.js, Meteor, KnockoutJS, Polymer, React.js, Sails.js, Cappucino
C++	CppCMS, Drogon, Poco, Tntnet, Wt
Haskell	Snap, Yesod
Scala	Lift, Play, Scalatra
PHP	CakePhP, CodeIgniter, Fat-Free, FuelPHP, Laravel, Kajona, Kohana, Li3, Nette Framework,Silex, Prado, Phalcon, Silverstripe, Yii
Ruby	Ruby on Rails, Camping, Padrino, PureMVC, Sinatra
Java	Apache OFBiz, Apache Struts, Apache Tapestry, Apache Wicket, vraptor, vaadin, spring, spark, Play, OpenXava, jWt, jVx, Grails, JavaServer Faces, OpenLaszlo,
Perl	Mojolicious, Maypole, Mason, Dancer, Catalyst

1.6.1 *Pyramid web application framework*

The choice of selecting a web application framework from the numerous available options is a difficult one for an application developer. However, it can be simplified by certain steps such as considering only frameworks available in a specific programming language or suitable for certain tasks.

Pyramid is a Python based web application framework that was first published as the Pylons web framework in 2005 [126]. It is a light framework that provides a developer with several customizable solutions for templating,

database and security [2]. This makes it very suitable for web projects in the life sciences, since Python libraries for solving computational challenges in life sciences are numerous and readily available and the Pyramid framework is highly customizable to create a fine solution for a specific task.

Pyramid applications run on a web server called *waitress* [190]. After project creation, running the command "pserve" with the configuration file in the project directory as an argument will start a running instance of a Pyramid web application.

The two most important files in a pyramid project are the configuration file (.ini) and the __init__.py file. The configuration file (referred to as .ini file subsequently) holds the configuration information for connecting to the web server, There are two types of .ini file, the development.ini and the production.ini. Both files are similar except that the production.ini does not contain information for debugging and as such recommended in a production environment [153]. The __init__.py file contains the actual code and is used to initialize the application. A typical __init__.py file is shown below

Listing 1.1: __init__.py file

```
from pyramid.config import Configurator

def main(global_config, **settings):
        config = Configurator(settings=settings)
        config.add_static_view('static', 'static', cache_max_age
           =3600)
        config.add_route('home', '/')
        config.add_route('page', '/page')
        config.scan('views')
        return config.make_wsgi_app()
```

The application lifecycle starts with the "pserve" command. This command is responsible for 2 main functions, first it locates the application entry points and configures the logging method using the .ini file, secondly it creates a Web Server Gateway Interface (WSGI) application using a library called *pasteDeploy*. To create the WSGI instance, it uses an entry point which is specified under the [app:main] section in the .ini file [154].

Listing 1.2: A typical Pyramid config file

```
        [app:main]

        use = egg:AwesomePyramidProject
```

```
        pyramid.reload_templates = true
        pyramid.debug_authorization = false
        pyramid.debug_notfound = false
        pyramid.debug_routematch = false
        pyramid.default_locale_name = en
        pyramid.includes =
        pyramid_debugtoolbar

        [loggers]

        keys = root, AwesomePyramidProject

        [logger_root]
        level = INFO
        handlers = console

        [logger_AwesomePyramidProject]
        level = DEBUG
        handlers =
        qualname = myproject
```

This entry point is usually the location of the __init__.py file, this file is used to create a Configuration instance, which contains the application's registry containing the configuration information. Next the make_wsgi_app function of the Configuration instance is called which returns a router instance that serves as a connector between the application server and web server, hence creating the WSGI application instance [154].

"pserve" uses the created WSGI instance and the make_server function which relies on port number configuration in the .ini files to create the final server that can then accept requests [154]

Part II

MATERIALS AND METHODS

2

MATERIALS AND METHODS

2.1 *C. burnetii* TYPING DATABASE IMPLEMENTATION

2.1.1 *Platform architecture*

The platform (referred to as CoxBase subsequently) architecture can be grouped in two parts, the front end and the back as illustrated in the figure below.

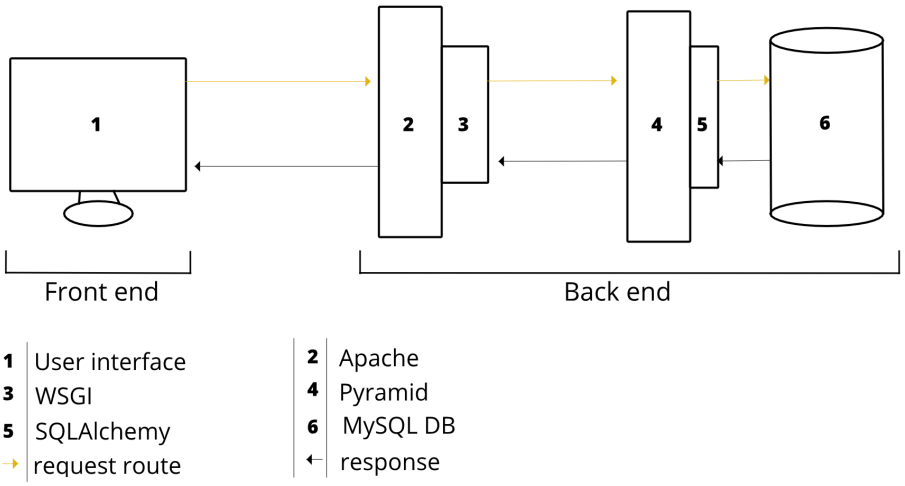


Figure 2.1: CoxBase architecture

2.1.1.1 *Front End*

The main component of the front end is the web user interface (UI). It allows users to interact with the platform in a web browser. The non-opinionated structure of the internet ensures any browser can be used, however, experience might deviate a little. The UI on the CoxBase platform is designed for several uses cases, I highlight all the use cases below:

GENOTYPING DATA QUERY: Genotyping data for *C. burnetii* genotyping methods are formatted differently. In order to present a uniform and intuitive method to query this data, we have used a table query method, where the table headers are the marker name and the column is an input field for the user. Figure 2.2, Figure 2.3 and Figure 2.4 depicts the query User Interface (UI) for the Multiple-Locus Variable Number Tandem Repeat Analysis (MLVA) genotyping and Multispacer sequence typing (MST) methods used on CoxBase. This was implemented with pure HTML and designed with CSS.

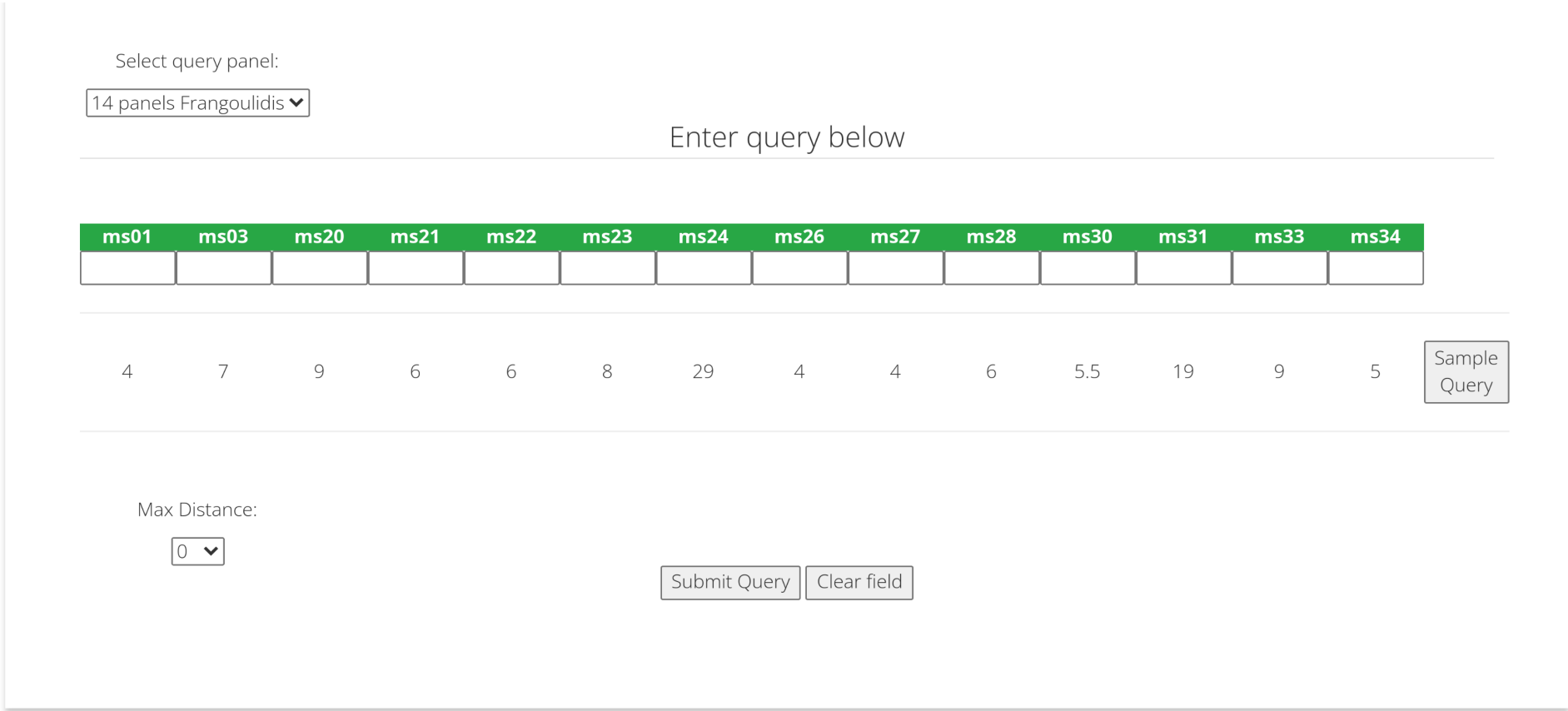


Figure 2.2: The 14 panel MLVA genotyping data query UI

PRIMER SEQUENCE QUERY: Primers are essential for genotyping in *C. burnetii*. A multi selection approach was implemented for the different query method in order to ensure that multiple methods can be combined for a single query. An external library Select2 (https://select2.org/) was utilized to aggregate multiple selections.

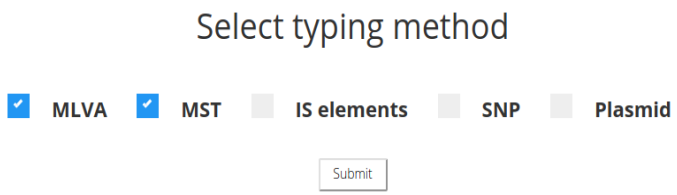


Figure 2.5: Primer query UI

SEQUENCE ANALYSIS: Five *in silico* typing methods for *C. burnetii* were implemented on CoxBase. A single intuitive interface was implemented across the five methods to ensure consistency for the users. Two input methods are available, the first input method is via a text area (A small space where sequences to be analyzed can be pasted), the second method is via file input, where a file containing sequences to be analyzed can be selected from the user's computer. The latter is the recommended method.

PROGRAMMING LANGUAGE USED FOR THE FRONT END
Javascript

EXTERNAL LIBRARIES
Bootstrap (https://getbootstrap.com/)
Select2 (https://select2.org/)
Leaflet (https://leafletjs.com/)
Multiselectjs (http://loudev.com/#home)
Chartjs (https://www.chartjs.org/)

2.1.1.2 *Back End*

The back end can be summarized as every component of the system that resides on the computing server. There are three main components in this category for the CoxBase platform. They are Apache server [46], Python[170] web framework (Pyramid) and MySQL database.

The Apache Server [46] is an open source, secure, efficient and extensible HTTP server. Originally launched in 1995, it is the most popular web server

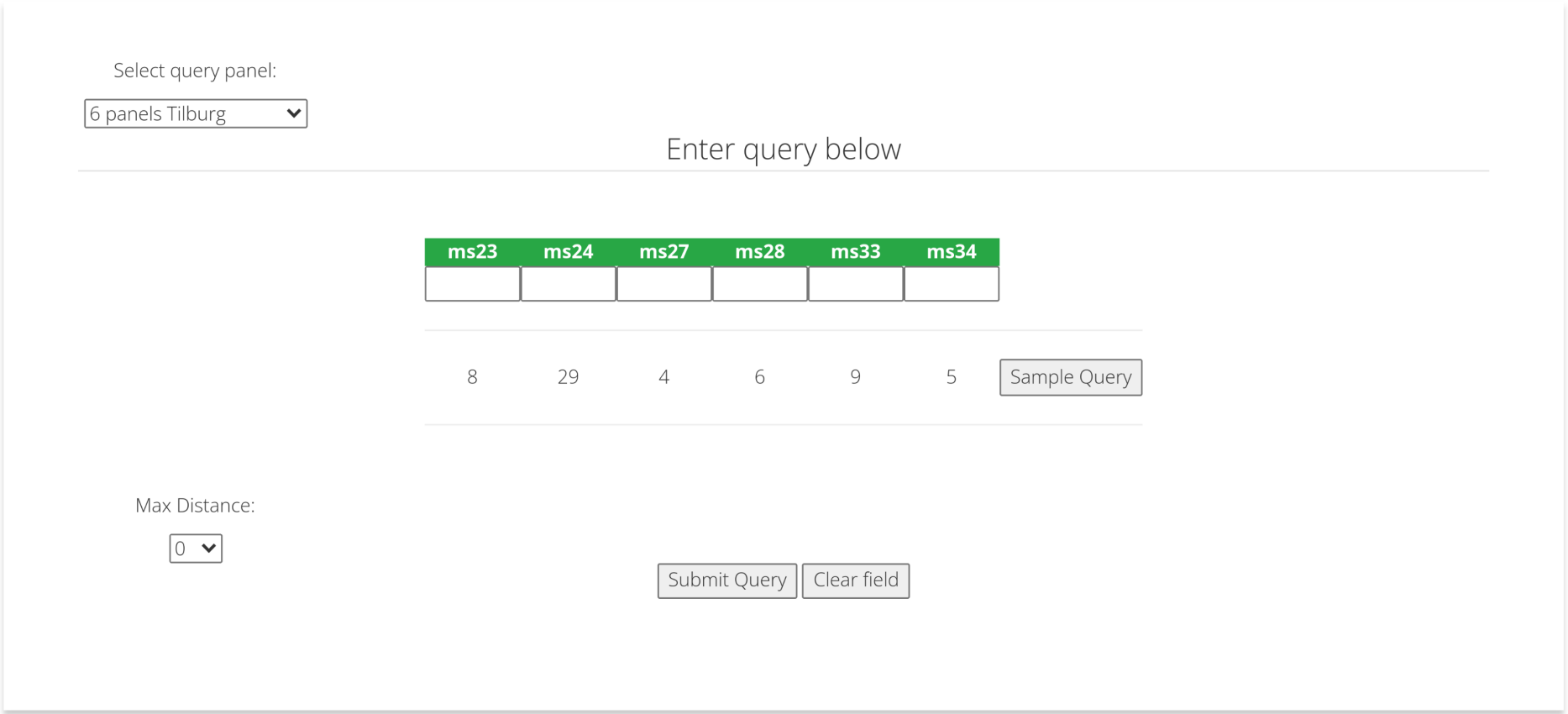


Figure 2.3: The 6 panel MLVA genotyping data query UI

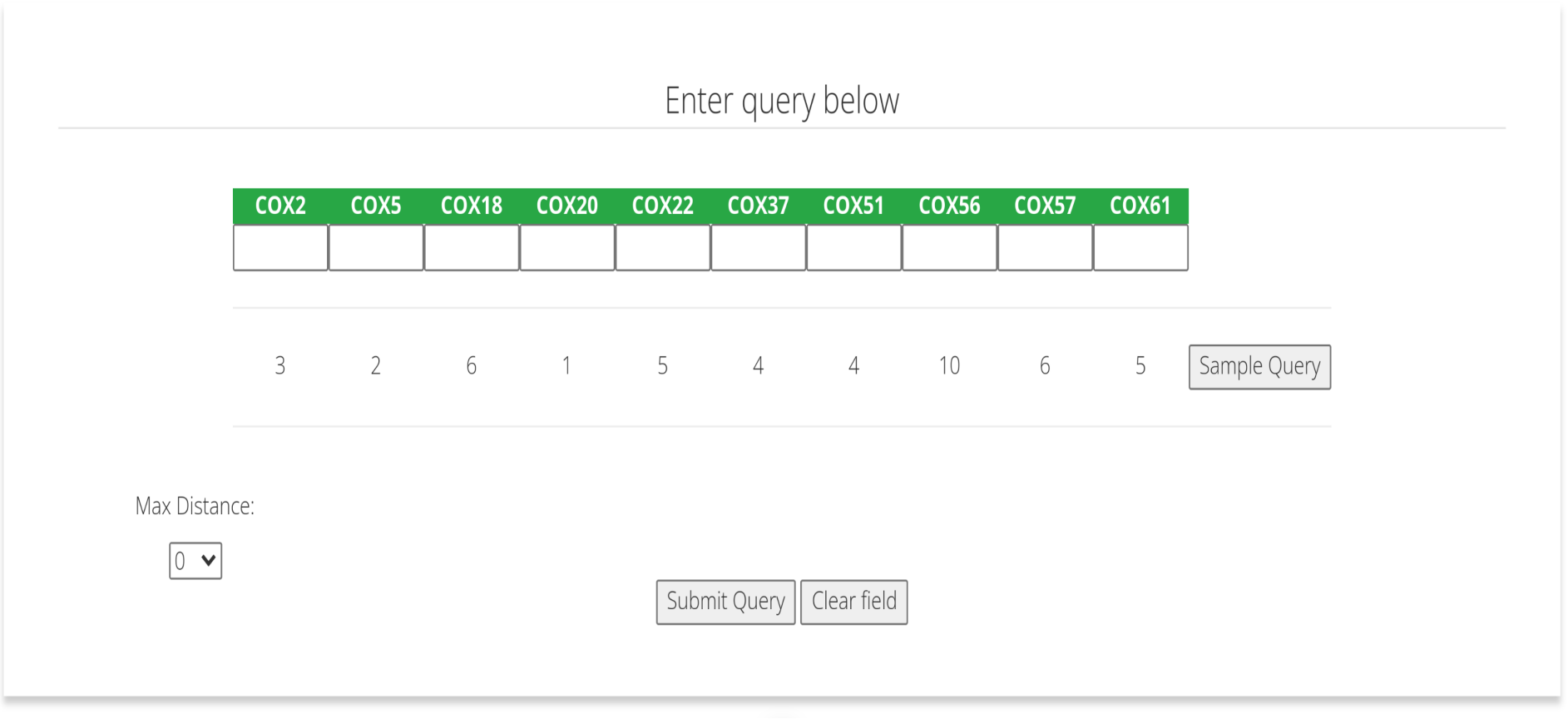


Figure 2.4: MST query UI

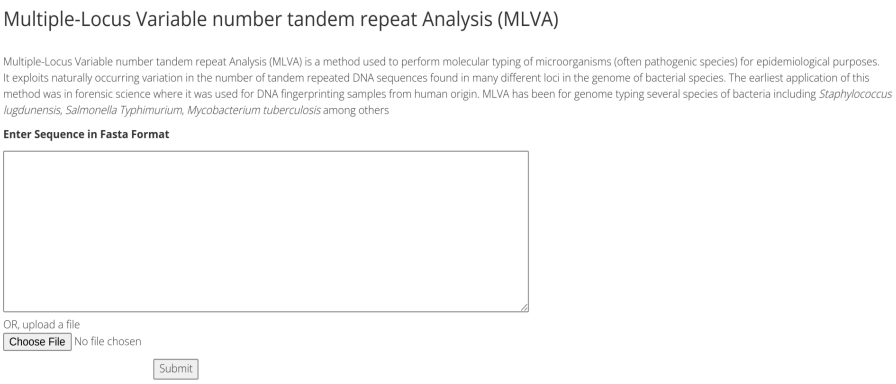


Figure 2.6: The sequence upload interface for MLVA genotyping on CoxBase

on the internet [46]. The Apache server is used to handle and process requests on the CoxBase platform. It is extended by a Web Server Gateway Interface (WSGI), that forwards every request to the Python [170] web application.

The Python web application was built using the Pyramid framework [2]. The purpose behind the selection of this framework was that it is non-opinionated, making it easier to customize features to specific use cases. The framework provided the environment for computation, querying the database for requested resources and subsequently passing the result to the server for the end user. The database is a MySQL Enterprise Edition.

2.1.2 *Database Architecture*

CoxBase implements a 3-tier architecture which separates the complexity of the application from the user to the actual data source into 3 tiers. The user is at the presentation tier, the application tier is driven by SQLAlchemy. SQLAlchemy is a Python SQL Toolkit [14], that provides a form of abstraction between the Python application and the MySQL database with InnoDB engine (data tier).

The relational model was utilized to store data in the database as it allows to model relationship between several entities as we expected will be the case for this platform.

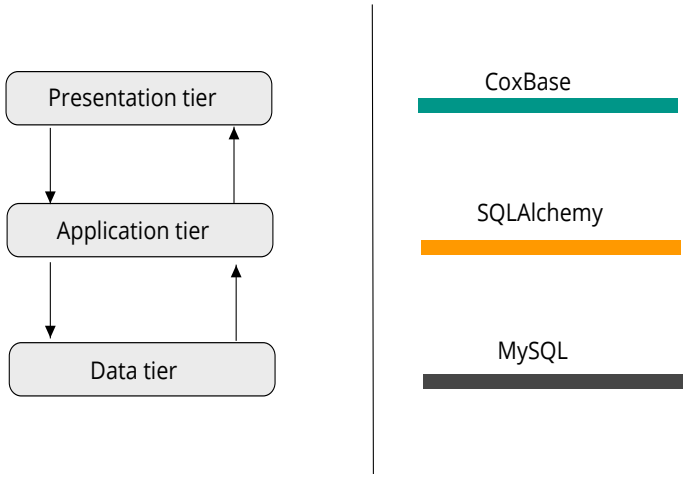


Figure 2.7: 3-tier architecture implementation on CoxBase

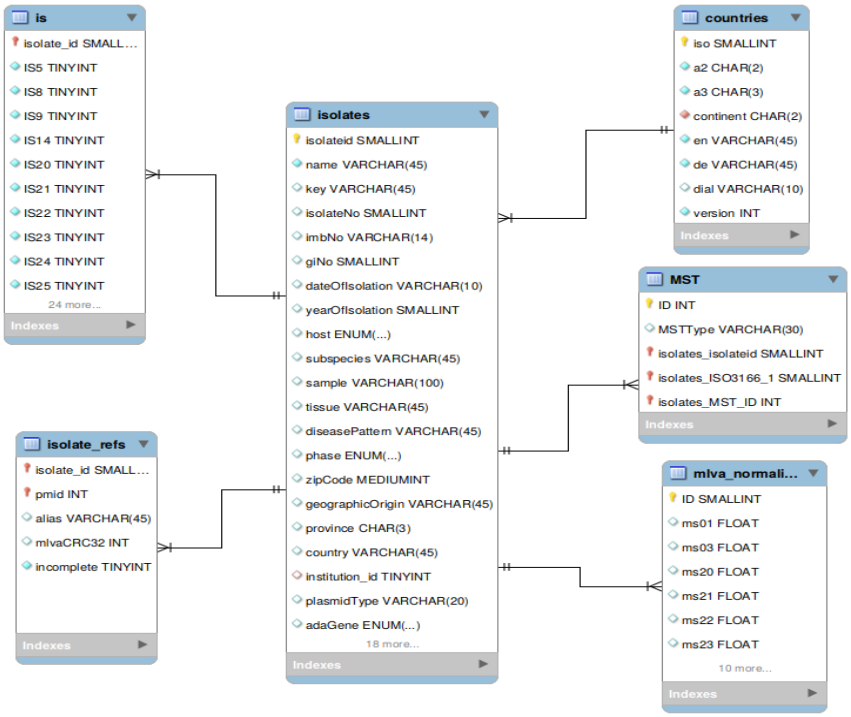


Figure 2.8: Query feature schema on CoxBase

The database consists of 31 tables that hold raw data with different mapping cardinalities. All tables and their descriptions are listed in appendix A.1. The tables can be sub grouped into schemas based on the type of feature on CoxBase

that utilizes its data. At the center of CoxBase features is isolates discovery. There are six tables housing data involved with this feature. At the center of the query schema [Figure 2.1.2] is the "isolate" table, which contains 38 attributes. The "isolateid" which is the primary key on the "isolate" table references the "isolate_id" attribute of the "is" and "isolate_ref" tables. The "mlva_normalized" and "MST" table both references the "mlva genotype" and "MST" attribute in the "isolate" table. This implementation makes it possible to query all the isolates with a particular genotype using the "isolate" table or querying the typing tables and afterwards searching for isolates with the the genotypes in the typing query.

The analysis schema consists of nine tables. Four tables manage MLVA typing data, one table is used to manage all submissions and the remaining 4 tables are used to manage, MST, adA, IS1111 and SNP typing data. Every table in the analysis schema has a unique ID column which is used as the primary key from a source table and a foreign key on a relation table. This approach ensures atomicity while maintaining completeness among each typing data.

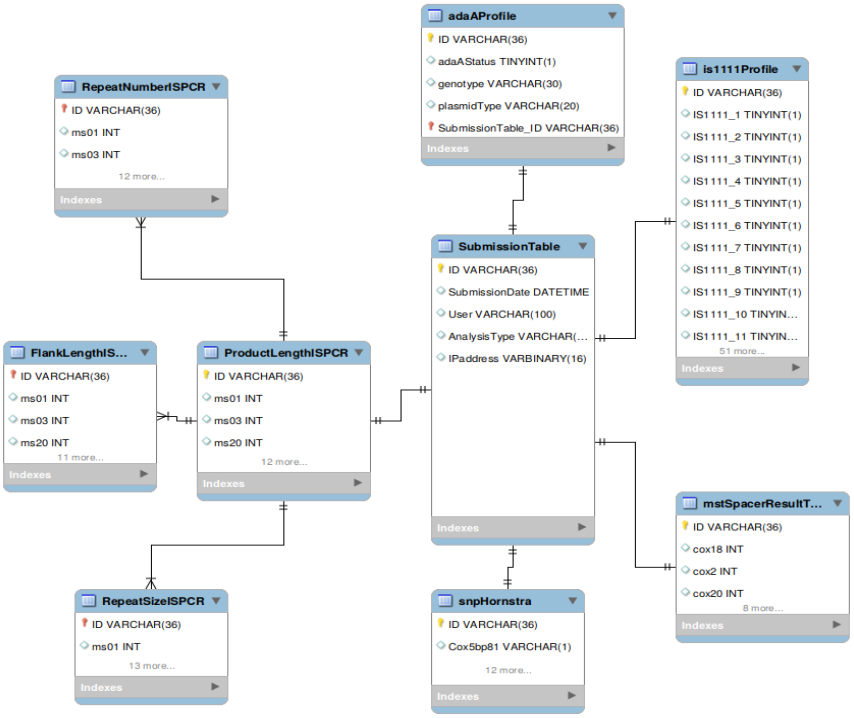


Figure 2.9: Analysis feature schema on CoxBase

The profile submission schema is made up of three tables. The first table manages MLVA profile submissions and the second table manages MST profile submission. The third table is a reference table for the two former tables and it manages the metadata associated with submitted typing data. The first two tables serve as place holders for the curation process. After curation the entry is transferred to the respective typing table.

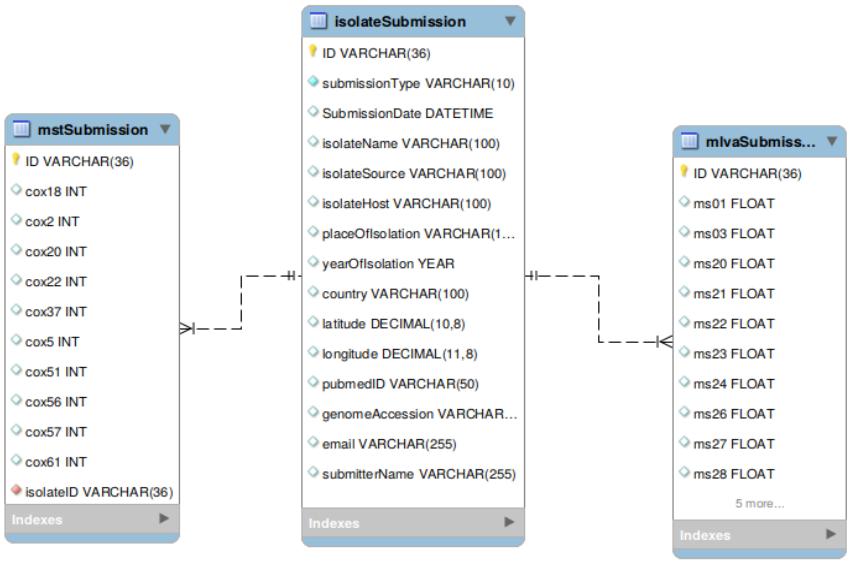


Figure 2.10: Submission schema on CoxBase

2.1.3 *Web application framework construction*

The Pyramid application was implemented in Python on a Ubuntu Linux 18.04 LTS operating system. To facilitate strict adherence to best software practices the application was installed via the pyramid-cookiecutter-starter library [128].

The next step was to establish a means of connection to the database by "creating a session". In Pyramid application, there are two approaches to this, scoped (non-global) sessions and global sessions. For this project, we implemented a global session approach which was added explicitly to the request object via a Python class instance. The advantage of this implementation is that it is quite easy to manipulate data based on the current request object, makes the view code shorter and cleaner and also it makes it easier to carry out functional tests with a live database session. The procedure for achieving this is to add the session object (essentially just a python class) as an argument to the "request_factory" method of the pyramid configurator instance.

After establishing a connection between application code and the database, the next step was to write "views callables". View callables are python objects that accept a request and return a response [126]. In Pyramid, view callables are implemented via a view configurator class. The view configurator associates the view callable with the view configuration, essentially mapping the request url to the appropriate response [126].

Several helper methods were also written to facilitate the computation and transformation of data. Specific details are provided under implementation methods.

2.2 GENOTYPING METHODS IMPLEMENTATION

2.2.1 *MLVA*

The MLVA *in silico* typing method was implemented in Python. The amplicon sequence of 14 MLVA markers was extracted with the e-PCR tool [144] based on primers from Frangoulidis *et al* [48]. For every marker, its known flanking length was subtracted from the determined amplicon sequence length. The obtained value was then divided by its repeat size to determine the repeat number as shown in the equation below. Table 2.1 contains all the data used for *in silico* implication of MLVA typing method

$$RN = (AL - FL)/RS$$

where RN = repeat number, AL = amplicon length, FL = flanking length and RS = repeat size.

2.2.2 *MST*

For the MST implementation, the USEARCH tool [43] was used to extract amplicon sequences from uploaded *C. burnetii* sequences based on primer sequences from Glazunova *et al* [55]. Afterwards, the extracted amplicon for each spacer primer was globally aligned to a library of spacer alleles in order to determine its allele type. Sequences without an allele type are classified as novel. The combination of MST spacer type (MST profile) can then be further used to query the database for the MST group.

2.2.3 *IS1111*

The IS1111 genotyping is a binary detection method, it explores the presence or absence of insertion elements at specific positions in the genome of *C. burnetii*. For the implementation, uploaded *C. burnetii* sequences are investigated for these positions using the e-PCR tool based on primer from Bleichert & Hanczaruk 2012 (unpublished). The primer library is an extension of the IS1111 primers published by Denison *et al* [39].

2.2.4 *adaA and plasmid type*

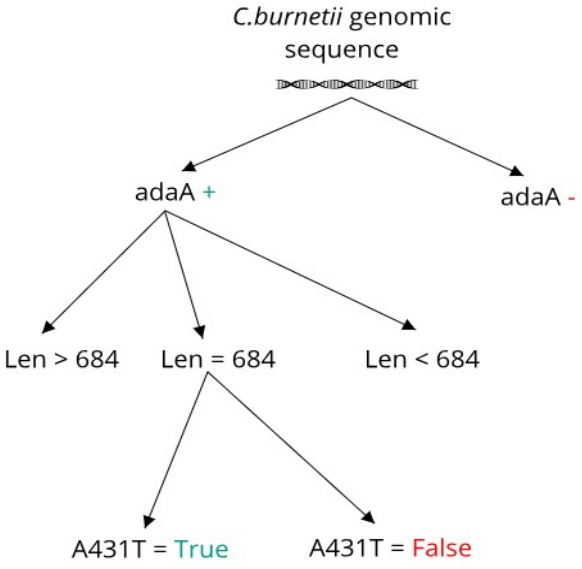
The *ada*A typing implementation relies on plasmids from [47]. They reported five variants of the *ada*A gene which we have used as a schema to classify *C. burnetii* isolates on CoxBase [47]. The first step is to extract the *ada*A amplicon sequence, this is done with the USEARCH tool. This step differentiates between *ada*A positive and *ada*A negative strains. The next step is to annotate the *ada*A type of the amplicon sequence. In order to type the extracted *ada*A amplicon sequence, we first compared the length to the standard nucleotide length of the *ada*A gene, which is 684 nucleotides. *C. burnetii* strains with *ada*A sequences

Table 2.1: Table of MLVA markers, their primer sequences, flanking length and repeat size

Marker name	Flanking length	Repeat size	Primer sense	Primer sequence
ms01	176	16	forward	GCCCTTGTCATCTTGCGG
			reverse	TCAAGTATTAATGAGCGTCG
ms03	142	12	forward	TGTCGATAAATCGGGAAACTT
			reverse	ACTGGGAAAAGGAGAAAAAGA
ms20	96	33	forward	CTGAAACCAGTCTTCCCTCAAC
			reverse	CTTTATCTTGGCCTCGCCCTTC
ms21	137	12	forward	AGCATCTGCCTTCTCAAGTTTC
			reverse	TGGGAGGTAGAAGAAAAGATGG
ms22	174	11	forward	GGGGTTTGAACATAGCAATACC
			reverse	CAATATCTCTTTCTCCCGCATT
ms23	91	7	forward	GGACAAAAATCAATAGCCCGTA
			reverse	GAAAACAGAGTTGTGTGGCTTC
ms24	126	7	forward	ATGAAGAAAGGATGGAGGGACT
			reverse	GATAGCCTGGACAGAGGACAGT
ms26	97	9	forward	GCAATCCAGTTGGAAAGAA
			reverse	ATTGAAGTAATCCATCGATGATT
ms27	249	6	forward	TTTTGAGTAAAGGCAACCCAAT
			reverse	CAAACGTCGCACTAACTCTACG
ms28	107	6	forward	AATGGAGTTTGTTAGCAAAGAAA
			reverse	AAAGACAAGCAAAACGATAAAAA
ms30	199	18	forward	ATTTCCTCGACATCAACGTCTT
			reverse	AGTCGATTTGGAAACGGATAAA
ms31	99	7	forward	ACAGGCCGGTATTCTAACC
			reverse	CCTCAGCACCCATTCAG
ms33	193	7	forward	TAGGCAGAGGACAGAGGACAGT
			reverse	ATGGATTTAGCCAGCGATAAAA
ms34	168	6	forward	TGACTATCAGCGACTCGAAGAA
			reverse	TCGTGCGTTAGTGTGCTTATCT

less than the standard length are typed as incomplete *ada*A genotypes, strains with sequences more than the standard length are typed as insertion genotypes. Strains with the standard length are evaluated for the nucleotide at position 431, if it's a thymine then it's typed as SNP variant, if it's an adenine then it's

typed as the wildtype variant. Figure 2.11 shows the workflow of our *ada*A *in silico* implementation.



adaA -	Deletion genotype
Len > 684	Insertion genotype
Len < 684	Incomplete genotype
A431T = True	SNP genotype
A431T = False	Wildtype

Figure 2.11: *ada*A genotyping workflow

For the *in silico* plasmid type detection, we implemented a detection based on *C. burnetii* plasmid primers shown in table 2.2.

Table 2.2: *C. burnetii* plasmid primers

Plasmid name	Primer sense	Primer sequence
QpRS	forward	CTCGTACCCAAAGACTATGAATATATCC
	reverse	CACATTGGGTATCGTACTGTCCCT
QpH1	forward	TGACAAATAGAATTTCTTCATTTTGATG
	reverse	GCTTATTTTCTTCCTCGAATCTATGAAT
QpDG	forward	TGGCGAGGTGTTCGGTATGAG
	reverse	CTTAGCGATTTATGGTTCCGTC
QpDV	forward	CTTATTTCAAAGAGTTCCTGCTAG
	reverse	CGCAACCGGCTGTTGTGC

2.2.5 *SNP typing*

We implemented the SNP typing method based on the findings of Hornstra *et al* [70]. They evaluated SNP at 12 positions in the genome of *C. burnetii*. Based on SNP type at these positions they classified *C. burnetii* isolates under three classes. To implement this *in silico*, we detected amplicons using the primers published by them. Afterwards, the type of base in each of the positions were evaluated and the result was stored in a database.

2.3 PREDICTION OF THE RESISTOME IN *Coxiella* ISOLATES

2.3.1 *Used genomes and pre-processing methods*

We downloaded sixty-one *C. burnetii* genomes from the RefSeq database [127]. In order to predict protein coding sequences from the FASTA files, we used the Prokka tool [145].

2.3.2 *Artificial Neural Network model implementation*

2.3.2.1 *Feature Characterization*

Feature characterization is the process of encoding data into numerical vectors. We applied two methods to convert the genomic sequence data into numerical vectors. In the first method, we generated a feature vector based on composition, distribution and transition (CTD) properties of protein using the Pybiomed library [41], In the second method, we generated a 400 dimension vector, based on Position Specific Scoring Matrix (PSSM) profiles and amino acid composition. The PSSM was derived from a PSI-BLAST of the query protein against a database of 4335 manually curated antibiotic resistance gene protein sequences from Comprehensive Antibiotic Resistance Database (CARD) [75] and Antibiotic resistance gene database (ARDB) [92].

2.3.2.2 *Training and test library*

The training and test library comprised 14,974 antibiotic resistant genes that came from three sources: CARD, ARDB and UNIPROT [8]. The ARGs are distributed under 40 categories as shown in figure 2.12. Curation and annotation of the ARGSs was done by Arango-agoty *et al* [10].

2.3.2.3 *Neural network*

The Artificial neural network was made up of five layers, four hidden layers and one output layer. Three dropout layers were inserted in between hidden layers. The main purpose of this is to randomly "drop out" layer outputs during training, this helps regularize the network and avoid overfitting [12]. To complete the architecture we implemented a softmax activation function on

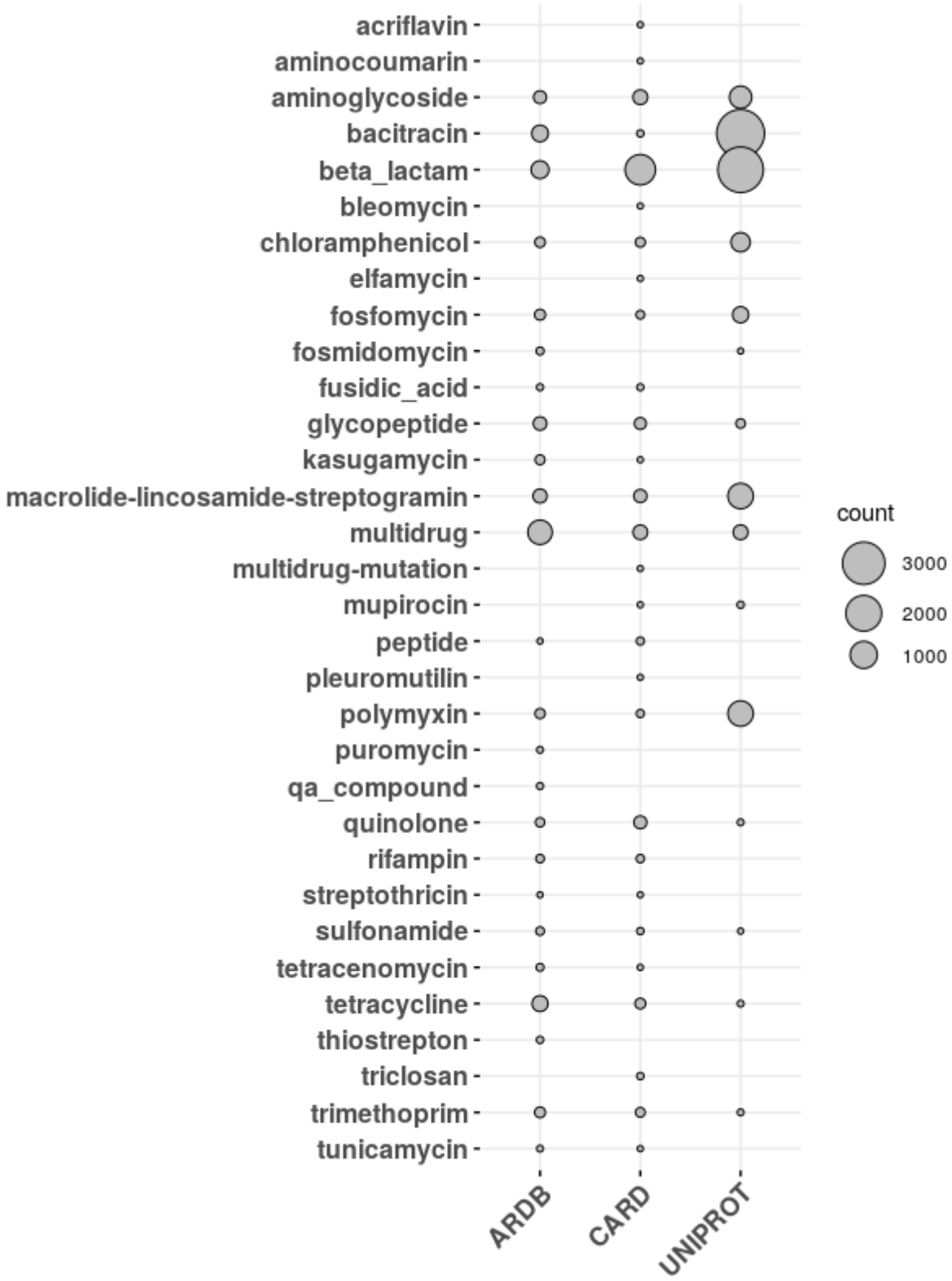


Figure 2.12: Distribution of antibiotic resistant genes in our training and test library and proportion from each database

the output layer to compute the probability of each prediction. The network was initially implemented with the noLearn Python package [115], which is a wrapper for Lasagne, which is also a wrapper for Theano [105], It was finally implemented with tensorflow [1] due to portability issues with noLearn.

2.4 GENOMIC ANALYSIS OF NOVEL STRAINS OF *C. burnetii*

2.4.1 *Isolation, sequencing and genome assembly*

The isolation and sequencing was done by Mathias Walter, and according to him: the strains were isolated from swabs taken from the vagina of a sheep and the placenta of a goat from Q fever diseased flocks. DNA was isolated using NucleoSpin-Kit from Machery-Nagel (Düren, Germany) following established and recommended protocols. Genomic library was prepared using the NEBNext Ultra II FS DNA Library Preparation Kit. Sequencing was carried out on the Illumina Miseq sequencer using the paired-end protocol. Genome assembly was performed using SPAdes [13], and afterwards the sequences were sent to me for further analysis. I carried out quality assessment of the assembly with Quast [58] and used Mauve [32] to reorder misassembled contigs.

2.4.2 *Gene prediction, sequence comparison and phylogenetic analysis*

Coding sequences (CDS) were predicted with GlimmerS [37] and Prodigal [71]. Non-coding tRNA and rRNA analysis were done with ARAGORN [87] and Barrnap [146] respectively. Further annotation was carried out via Prokka [145]: Gene names were ascertained via similarity search to the Swiss-Prot database [22], NCBI Bacterial Antimicrobial Resistance Reference Gene Database and, a database of transposable elements [49]. CDS that were not annotated from these three databases were further subjected to protein profile similarity search of the Rfam [56], TIGRFAM [60] and HAMAP [121] database with HMMER [42].

Sequence comparison was carried out using average nucleotide identity (ANI) via Pyani [125] between the novel isolates and a library of 60 *C. burnetii* genomes downloaded from RefSeq. ANI measures the level of genomic similarity at the nucleotide level based on the coding region of compared genomes. Phylogenetic analysis was carried out using the CoxBase platform.

Part III

RESULTS

3

RESULTS

3.1 COXBASE PLATFORM

The platform can be accessed at this url https://coxbase.q-gaps.de

3.1.1 *Genotyping application examples*

The CoxBase platform was tested with 50 *C. burnetii* genomic sequences obtained from the RefSeq Database (List of genomes in appendix A.2). Eleven of the genomes were complete chromosome assembly, thirteen were classified as incomplete chromosome assembly, fifteen were contigs and eleven were scaffolds. Several of the genomes were of unknown genotypes (40/50) based on the popular *C. burnetii* typing schemes or were not published if known. The genome sequences were downloaded in FASTA format and subsequently typed using our implementation. The obtained results are discussed below:

3.1.1.1 *MlVA typing*

The fifty *C. burnetii* genomes were successfully categorized under 4 main groups based on our *in silico* MLVA typing implementation. It was observed that certain MLVA markers were less likely to produce an amplicon *in silico* compared to others and as such, a measure called "Marker effectiveness" was established. This simply means the probability that a marker will produce an amplicon *in silico*. The markers ms03, ms21, ms22, ms27, ms31 and, ms34 were observed to be the most reliable MLVA markers using this measure.Markers that do not produce an amplicon *in silico* were however still reported, as exemplified in Figure 3.2. Following a successful *in silico* genotyping, features to compare the obtained MLVA typing result with MLVA profiles of existing isolates are provided on CoxBase. The first feature is based on the dissimilarity distance matrix using the Shrivers distance [150]. For the implementation, the obtained MLVA profile was added as a row to a table containing rows of known MLVA profiles, a dissimilarity matrix was then calculated from entries in the entire table using the Shriver algorithm which was implemented in Python. The resulting matrix was then visualized as a tree using an implementation of the PhyD3 library [81] as shown in Figure 3.3. The second feature is based on SQL queries to find existing MLVA profiles with the exact profile of the obtained MLVA result or the most similar MLVA profile to it. The implementation is described in the flow chart in Figure 3.4

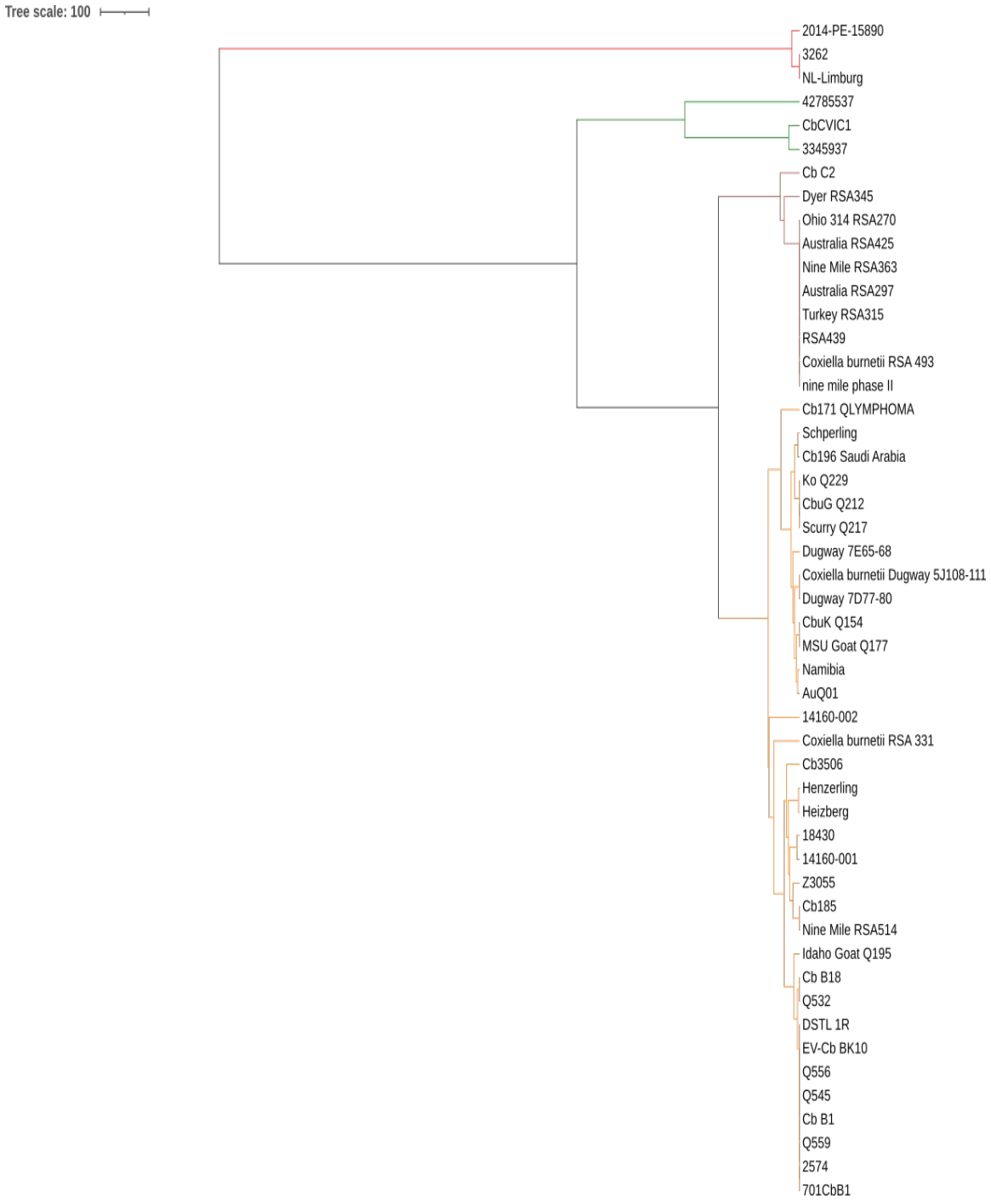


Figure 3.1: Unrooted dendrogram of typed *Coxiella* genomes based on MLVA profiles obtained from CoxBase

	ms01	ms03	ms20	ms21	ms22	ms23	ms24	ms26	ms27	ms28	ms30	ms31	ms33	ms34
Product length	248	215	Not detected	222	246	Not detected	Not detected	Not detected	270	162	306	136	248	192
Flank length	176	142	Not detected	137	174	Not detected	Not detected	Not detected	249	107	199	99	193	168
Repeat size	16	12	Not detected	12	11	Not detected	Not detected	Not detected	6	6	18	7	7	6
Repeat number	**4**	**6**	**Not detected**	**7**	**6**	**Not detected**	**Not detected**	**Not detected**	**3**	**9**	**5**	**5**	**7**	**4**

Figure 3.2: MLVA typing result of the Q321 genome done on CoxBase. MLVA markers ms20, ms23, ms24 and ms26 did not produce any amplicon *in silico*
Table Details: First column denotes calculated properties for each marker and the first row shows the corresponding marker names.
*Not detected: There was no amplicon found for the marker

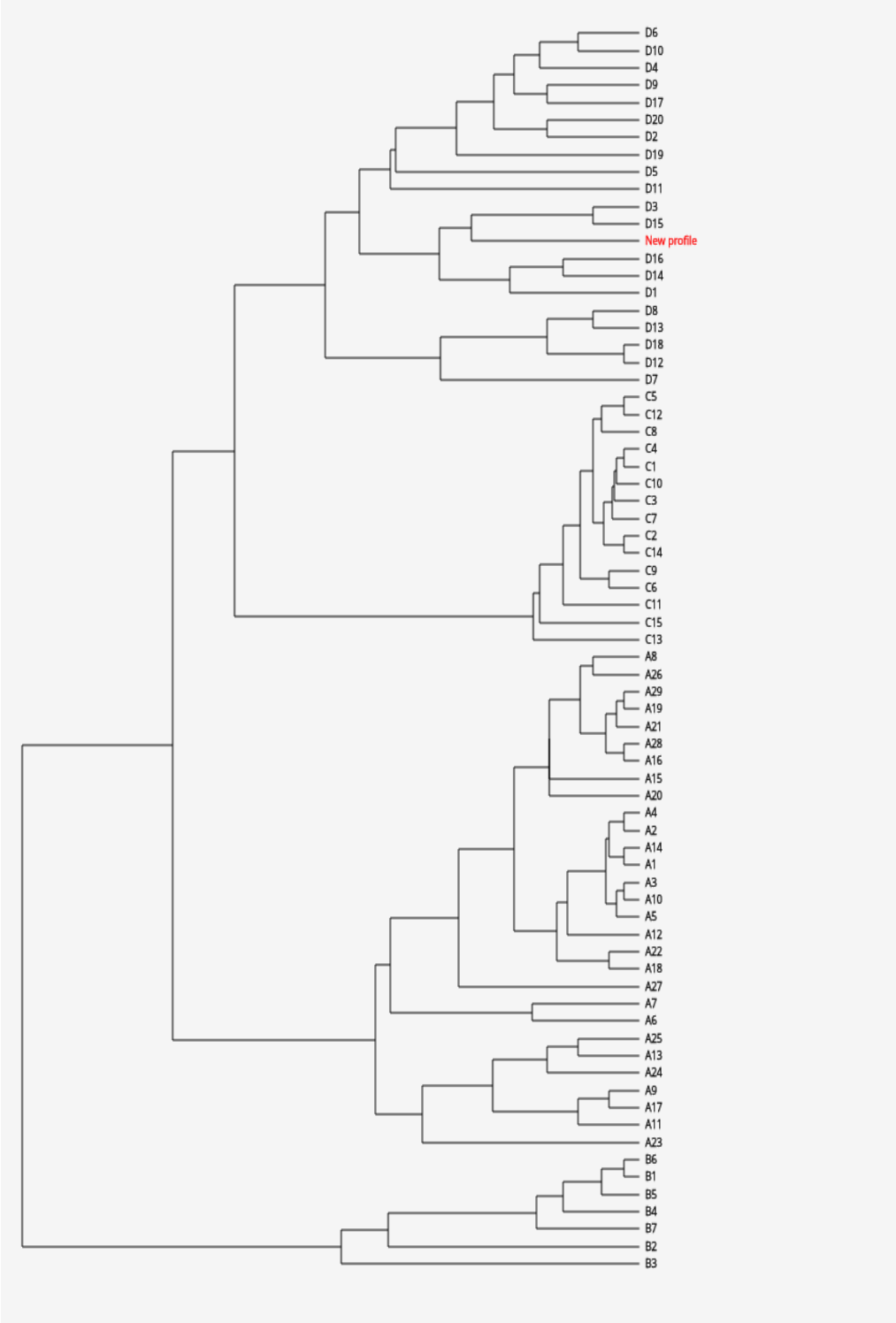


Figure 3.3: Visualization of MLVA genotyping on CoxBase.
*Highlighted node indicates the new profile that was submitted by the user

3.1.1.2 *MST typing*

All fifty genomes were successfully genotyped into their MST groups with our MST *in silico* typing implementation except for cases where the sequence quality was poor. Descriptive analysis revealed that the most frequently ob-

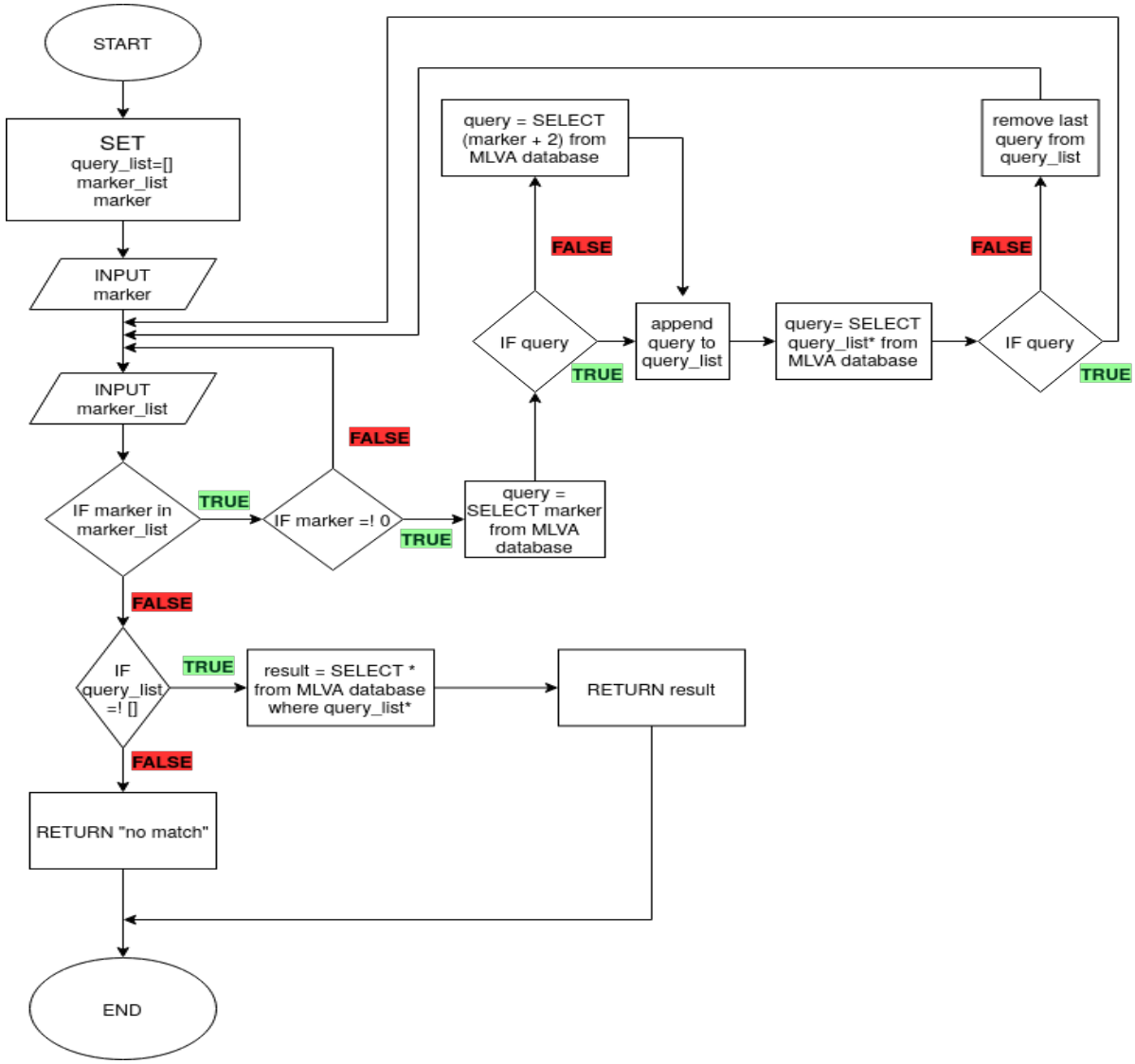


Figure 3.4: Flow chart of SQL based MLVA comparison on CoxBase
Legend:
Oval - Start/End
Rectangle - Process
Parallelogram - Input/Output
Diamond - Decision
Arrows - Connector

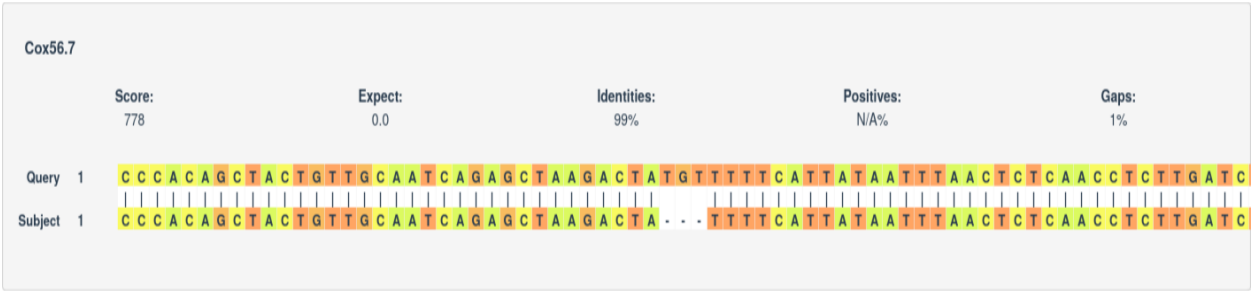
served MST genotype was MST group 61 (13/50), followed by MST 16 (11/50). Our MST typing implementation extends features from the MST resource in France (https://ifr48.timone.univ-mrs.fr/mst/coxiella_burnetii/). With our implementation, the obtained MST profile can be used directly to query for the MST group. Furthermore, we observed that often spacers that are reported as new spacers could be as a result of noise in the sequence, therefore we implemented a visualization feature for new spacers based on the BlasterJS library [21]. Figure 3.5 highlights our implementation.

cox2	cox5	cox18	cox20	cox22	cox37	cox51	cox56	cox57	cox61
10	10	11	10	6	4	12	new cox56	10	10

(a)

Description	Max score	Total score	Query cover	E value	Identities
Cox56.7	778	778.0	100%	0.0	99%
Cox56.9	773	773.0	100%	0.0	98%
Cox56.2	773	773.0	100%	0.0	98%
Cox56.8	767	767.0	100%	0.0	98%
Cox56.1	763	763.0	100%	0.0	98%
Cox56.4	761	761.0	100%	0.0	98%
Cox56.3	760	760.0	100%	0.0	97%
Cox56.5	758	758.0	100%	0.0	98%
Cox56.12	756	756.0	100%	0.0	98%
Cox56.11	756	756.0	100%	0.0	97%
Cox56.10	752	752.0	100%	0.0	98%
Cox56.15	713	713.0	100%	0.0	96%

(b)

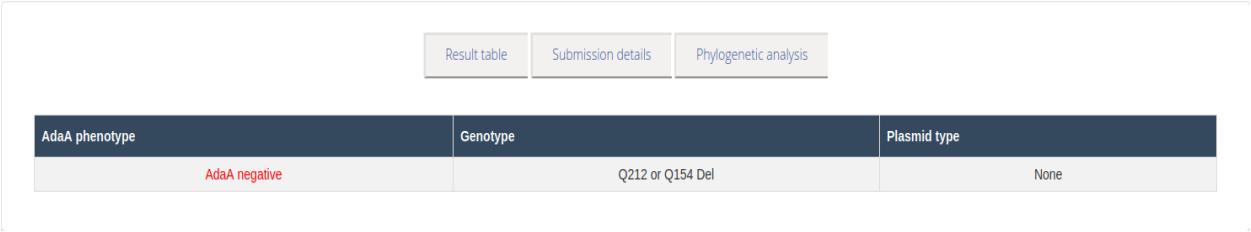


(c)

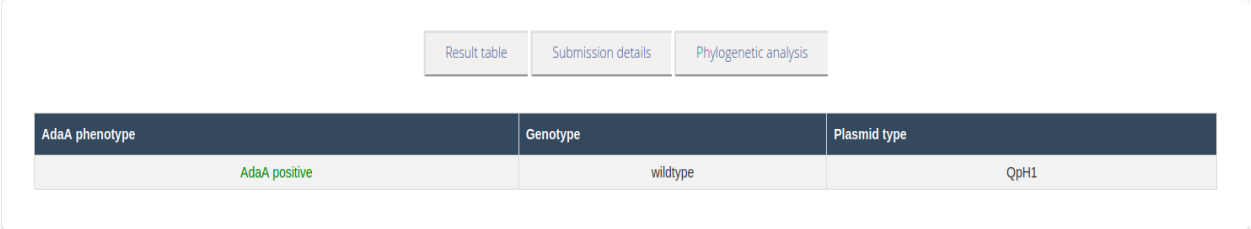
Figure 3.5: (a). Result of *in silico* MST typing of *C. burnetii* strain Dugway 7D77-80, showing a new sequence type for Cox56. (b). Table of alignment of all Cox56 alleles with the new Cox 56 sequence. (c). Visualization of the alignment with the highest score reveals multiple deletions

3.1.1.3 *adahttps://coxbase.q-gaps.deA and Plasmid typing*

The *ada*A and Plasmid typing results revealed that most of the analyzed *C. burnetii* isolates were *ada*A positive (41/50). Majority of the positive strains were of the wildtype genotype (34/41). The *ada*A phenotype typing result on CoxBase was color coded, and reported using a table for simplicity purposes as shown in Figure 3.6. Twenty-three out of the fifty analyzed *C. burnetii* genomes were of known plasmid types. We were able to confirm the known plasmid type and also obtained the plasmid type of the other genomes with our implementation. QpH1 was the most predominant plasmid type (31/43) in the genomes we analyzed, seven genomes were plasmidless, and of the remaining genomes three had the QpDG plasmid, three had the QpRS plasmid and one the QpDV plasmid.



(a)



(b)

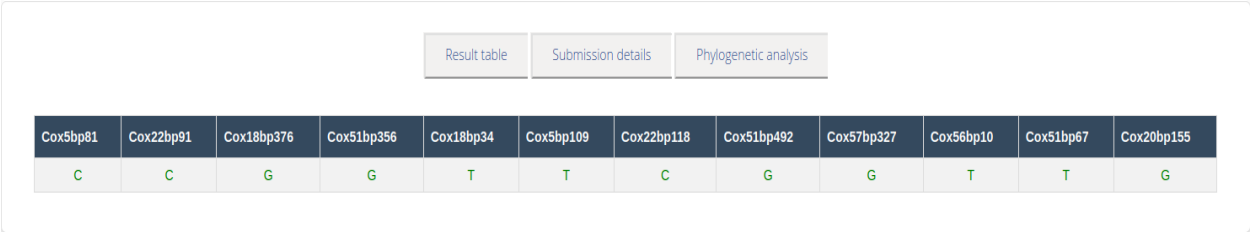
Figure 3.6: (a). *ada*A and plasmid typing of *C. burnetii* strain CBUG_212. (b). *ada*A and plasmid typing of *C. burnetii* strain RSA 493

3.1.1.4 *SNP typing*

The SNP typing implementation reports SNPs at 12 positions within the *C. burnetii* genome. Our implementation only reports the nucleotide base at these positions as shown in Figure 3.7.

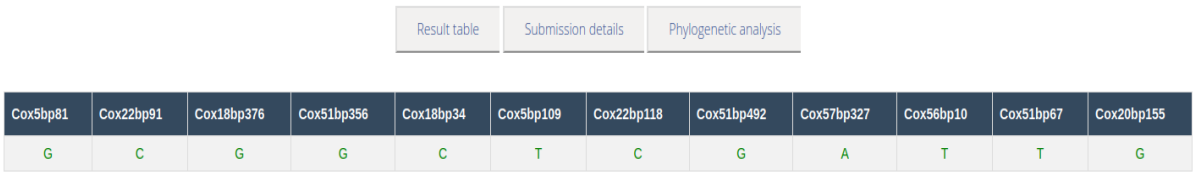
3.1.1.5 *IS1111 typing*

Detection of IS1111 localizations is a binary task. The result is either positive or negative unlike the other genotyping methods. Therefore, we reported the result using a visual approach of green cross for positive detections and a red dash for negative detections. The results are also compiled in a tabular form and we provided a feature for it to be saved as a comma separated value (CSV) file.



Cox5bp81	Cox22bp91	Cox18bp376	Cox51bp356	Cox18bp34	Cox5bp109	Cox22bp118	Cox51bp492	Cox57bp327	Cox56bp10	Cox51bp67	Cox20bp155
C	C	G	G	T	T	C	G	G	T	T	G

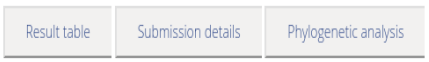
(a)



Cox5bp81	Cox22bp91	Cox18bp376	Cox51bp356	Cox18bp34	Cox5bp109	Cox22bp118	Cox51bp492	Cox57bp327	Cox56bp10	Cox51bp67	Cox20bp155
G	C	G	G	C	T	C	G	A	T	T	G

(b)

Figure 3.7: (a). SNP typing of *C. burnetii* strain CBUG_212. (b). SNP typing typing of *C. burnetii* strain RSA 493



IS element	state	IS element	state	IS element	state	IS element	state	IS element	state
IS1111-1	+	IS1111-2	+	IS1111-3	+	IS1111-4	+	IS1111-5	+
IS1111-6	+	IS1111-7	+	IS1111-8	+	IS1111-9	+	IS1111-10	+
IS1111-11	+	IS1111-12	+	IS1111-13	+	IS1111-14	+	IS1111-15	+
IS1111-16	+	IS1111-17	+	IS1111-18	+	IS1111-19	+	IS1111-20	+
IS1111-21	-	IS1111-22	-	IS1111-23	-	IS1111-24	-	IS1111-25	-
IS1111-26	-	IS1111-27	-	IS1111-28	-	IS1111-29	-	IS1111-30	-
IS1111-30	-	IS1111-31	-	IS1111-32	-	IS1111-34	-	IS1111-35	-
IS1111-36	-	IS1111-37	-	IS1111-38	-	IS1111-39	-	IS1111-40	-
IS1111-41	-	IS1111-42	-	IS1111-43	-	IS1111-44	-	IS1111-45	-
IS1111-46	-	IS1111-47	-	IS1111-48	-	IS1111-49	-	IS1111-50	-
IS1111-51	-	IS1111-53	-	IS1111-54	-	IS1111-55	-	IS1111-56	-
IS1111-57	-	IS1111-58	-	IS1111-59	-	IS1111-60	-	IS1111-61	-
IS1111-84	-								

Figure 3.8: IS1111 genotyping *C. burnetii* strain RSA 493 showing the 20 IS1111 transposase reported by [148]

3.1.2 *Database query application examples*

3.1.2.1 *MLVA and MST database query*

As discussed earlier in section 2.1.1.1, the input field for MLVA and MST query on the platform were implemented in tables and fieldsets. MLVA can accept floating point inputs but the MST implementation cannot.

The MLVA database contains 71 MLVA groups that belong to over 400 isolates. Queries are implemented on the MLVA groups to improve computational time. Features to define how close a match should be to the query data is also implemented with a maximum distance of five mismatches and a minimum distance of 0 i.e a perfect match.

The MST database contains 67 entries, Queries are also implemented with a feature to address closeness of the result to the query as described above for MLVA query.

The screenshot below describes the MLVA profile query for a *C. burnetii* strain isolated from a goat in Germany.

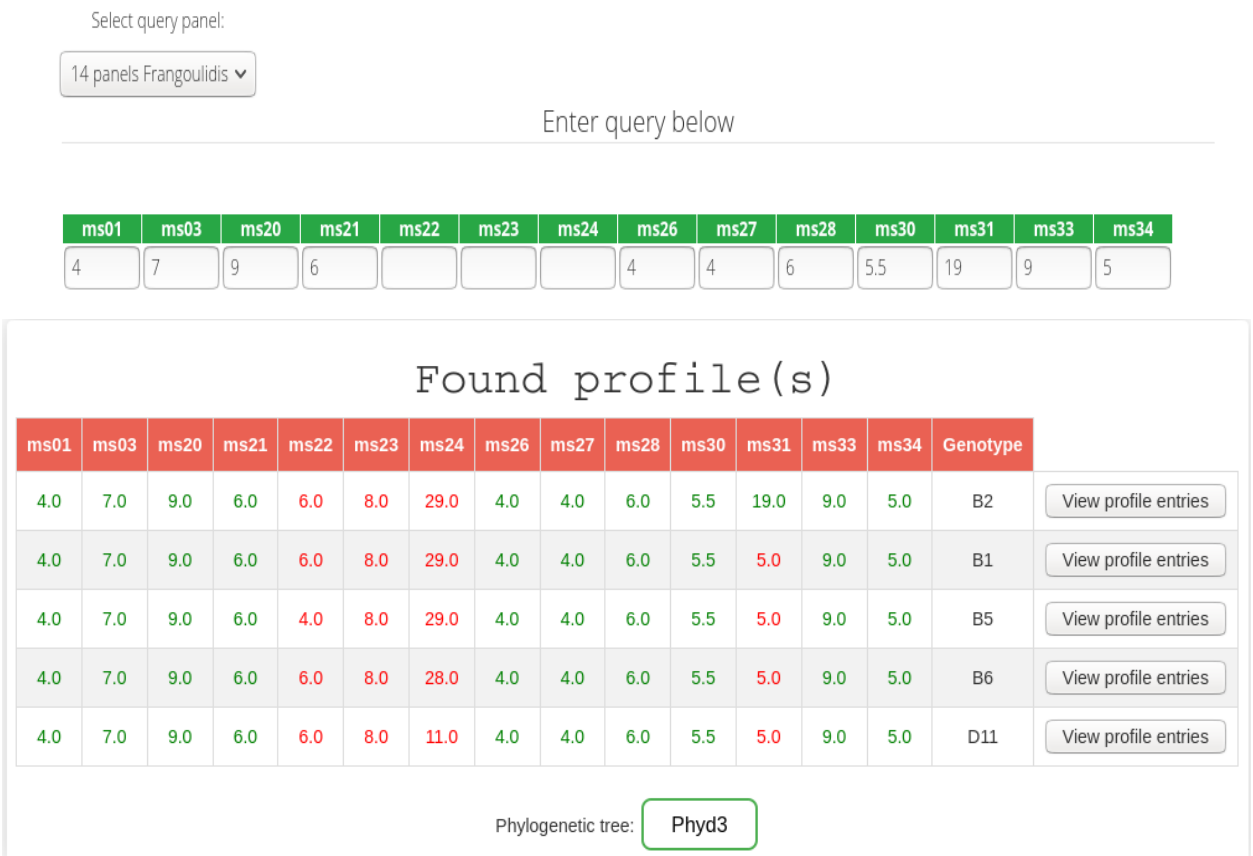


Figure 3.9: MLVA results of *C. burnetii* isolate from a goat in Germany

The results of the query are color-coded to aid visual comprehension. Markers that are equal with the query valued are green and misses are colored red. Also the returned result is sorted in order of similarity to the query.

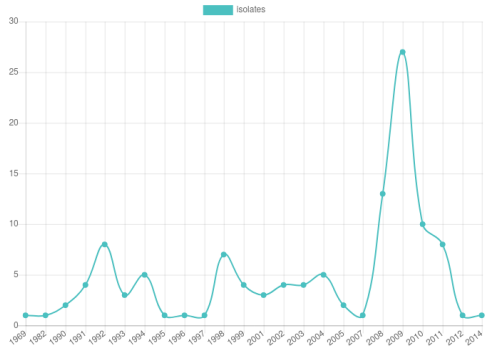
3.1.2.2 *Primer sequences*

Primer sequences are essential to PCR analysis of *C. burnetii* genomes. Therefore, we have included a query interface for primers for five genotyping (MLVA, MLST, Plasmid, IS1111 and *ada*A) methods on the CoxBase platform.

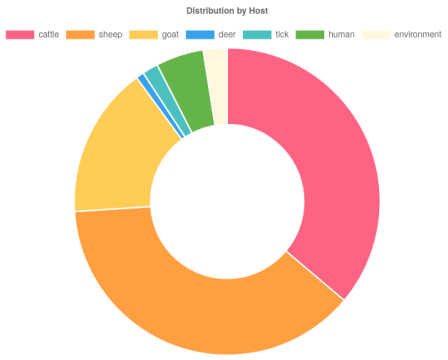
3.1.3 *Visualization*

3.1.3.1 *Metadata*

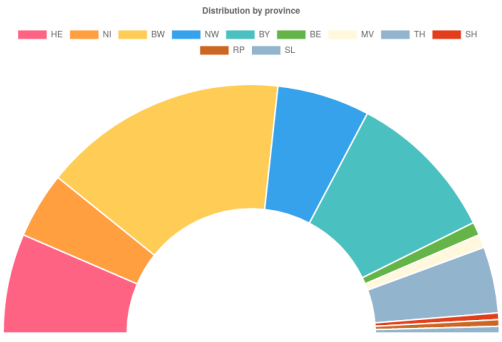
To visualize isolate metadata, we used four categories: year of isolation, MLVA genotype, location of isolation, and host. This was implemented for countries with a sufficient number of isolates (total isolate count is greater than 10). Figure 3.10 shows the metadata visualization of isolates from Germany.



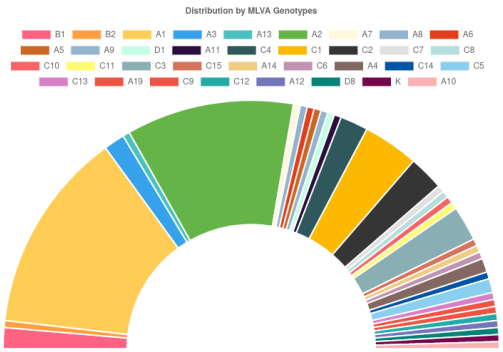
(a) Line plot of number of isolates against the year of isolation showing a peak in year 2010, which was the peak of Q fever outbreak in the Netherlands.



(b) Distribution plot of the host of German isolates, showing that cattle and sheep are the most common hosts.



(c) Distribution plot of the provinces of German isolates, it reveals that the majority of the isolates are from southern Germany.



(d) Genotype distribution plot reveal that A2 and C1 are abundantly represented in German isolates.

Figure 3.10: Metadata visualization on CoxBase

3.1.4 *Genome browser*

We implemented a Genome viewer on the CoxBase platform that was based on the NCBI genome viewer tool [132]. Users can access the genomes of 18 reference *C. burnetii* strains. The genome can be investigated/analyzed using features such as sequence search, sequence download, sequence markup and annotation of various genomic features.

3.1.5 *Submissions*

We implemented a submission interface that allows users to submit isolate metadata as well as MLVA and MST profiles to the database. The purpose of this is to encourage users who want to share their data to do so and also additional data will help to enrich the database.

3.2 PREDICTION OF THE RESISTOME IN *Coxiella* ISOLATES

3.2.1 *Model evaluation*

We evaluate the ANN models with a 0.2 split of the train-test library. This was chosen to ensure maximum training set representativesness. scikit-learn metrics [120] was used to determine the accuracy, precision and recall values as documented in Table 3.1 and Table 3.2. For the nolearn implementation, comparisons between the two models reveal that the PSSM-COMP model had a higher level of accuracy after 40 epochs of training. For the Tensorflow implementation we introduce a new hyper-parameter, we selected a batch size of 40, and trained with early stopping based on validation loss.

Table 3.1: Table of evaluated parameter for nolearn implementation

	Model	
Parameter	CTD	PSSM-COMP
Accuracy	0.62	0.95
Weighted precision	0.43	0.94
Weighted recall	0.62	0.95
Weighted F1-score	0.49	0.94

Table 3.2: Table of evaluated parameter for tensorflow implementation

	Model	
Parameter	CTD	PSSM-COMP
Accuracy	0.92	0.92
Weighted precision	0.92	0.89
Weighted recall	0.92	0.92
Weighted F1-score	0.91	0.90

3.2.2 *Model calibration*

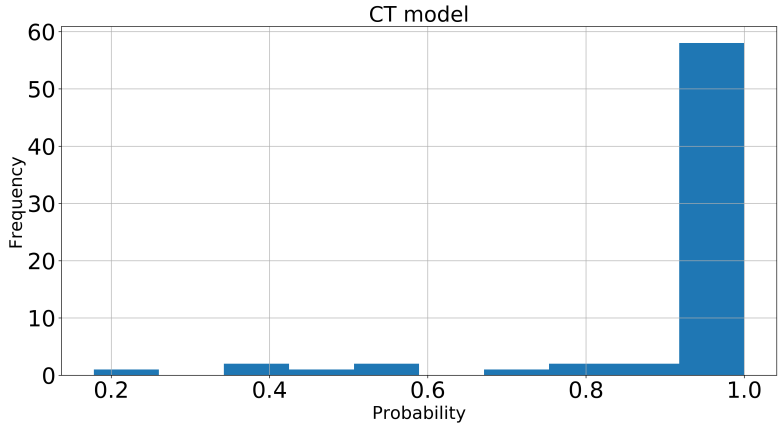
Since the aim of our model was prediction of ARGs from *Coxiella* genomes, we sought to calibrate the models based on accuracy as well as confidence using known datasets. We calibrated three of the four models on two datasets (We did not use the CTD model from the nolearn implementation due to low accuracy).

3.2.2.1 *Berglund dataset*

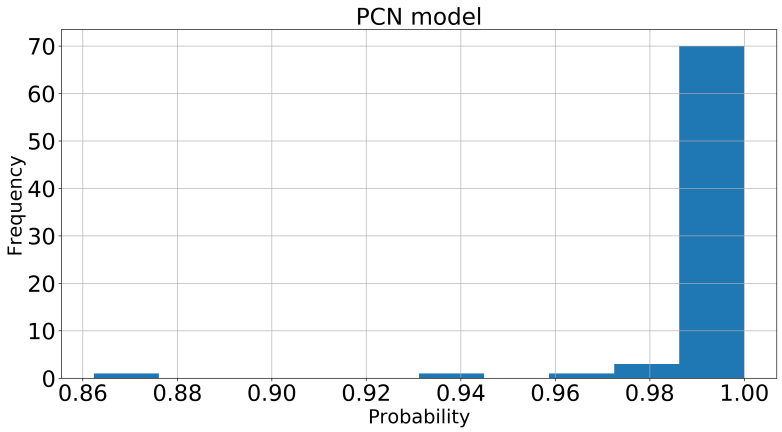
This dataset consists of 76 experimentally validated beta-lactamase amino acid sequences obtained from Berglund et al. [17]. The most successful model was the PSSM-COMP model implemented with tensorflow, thereafter referred to as PCT model. It classified all the 76 sequences as beta-lactam sequences. The PSSM-COMP model implemented with nolearn, thereafter referred to as PCN model, classified 75 out of the 76 amino acid sequences as beta-lactams, The CTD model implemented with tensorflow, thereafter referred to as CT model, classified 69 sequences as beta-lactams and 7 sequences were categorized incorrectly. Since our implementation output is a probabilistic distribution as a result of the softmax activation, we plotted a histogram of the classification output's probability for instances where the classification was correct in order to visualize the confidence of the model as seen in Figure 3.11.

3.2.2.2 *MEGARes dataset*

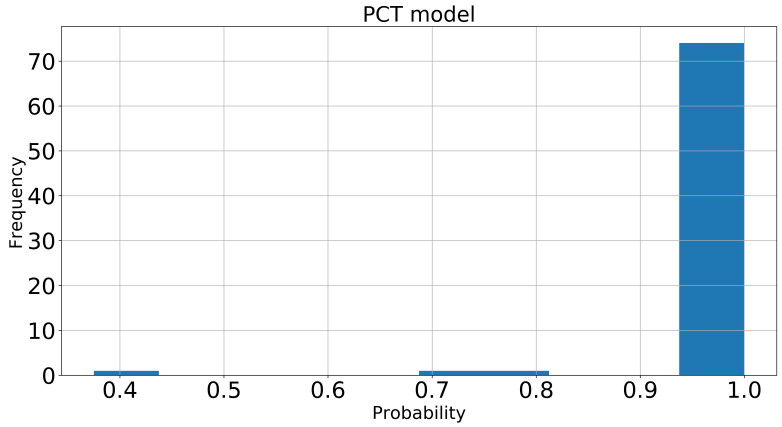
The second dataset contained 7126 antibiotic resistance genes, obtained from the MEGARres database [86]. For the PSSM-COMP based models only 6514 sequences were evaluated. The PCN model categorized 5380/6514 ARG sequences correctly, the PCT model categorized 4979/6514 sequences correctly while the CT model categorized 5205/7126 sequences correctly. However, when we binned the prediction based on the confidence of the model using reported probabilities, we observed a significantly higher accuracy at probabilities that were greater than or equal to 0.98. Thus, we chose this probability as the threshold filter for profiling the *Coxiella* resistome.



(a) The CT model categorized roughly 60 sequences with high confidence.



(b) The PCN model categorized roughly 70 sequences with high confidence.



(c) The PCT model categorized roughly 73 sequences with high confidence.

Figure 3.11: Visualization of different model confidence based on the Berglund dataset

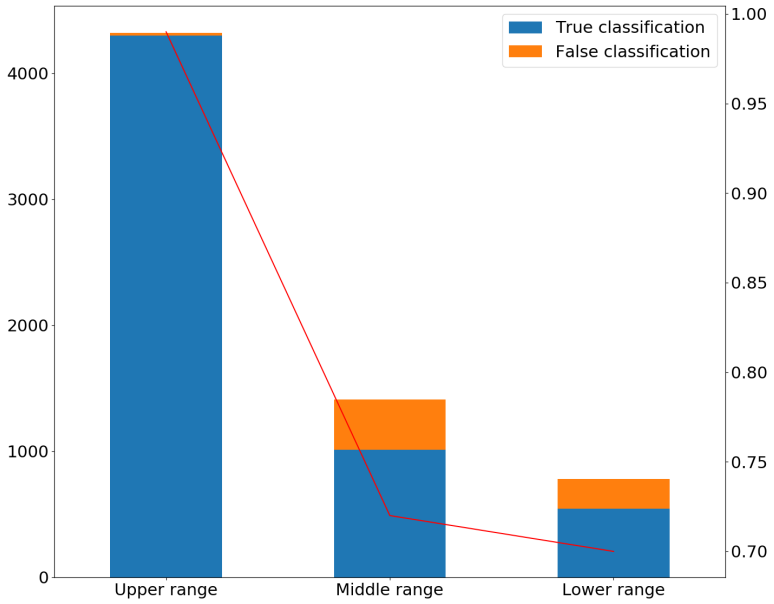


Figure 3.12: Assessment of the distribution of true and false classification between bins of probabilities reveals that the bins with higher probabilities contains higher numbers of true classification. Upper range bin = p $\geq$ 0.98, middle range bin = 0.98 <p >0.5, lower range bin = p $\leq$ 0.5

3.2.3 *Prediction of antibiotic resistance in C. burnetii*

We predicted ARG sequences under 5 categories in *C. burnetii*; Multidrug, beta-lactam, macrolide- lincosamide-streptogramin, polymyxin and bacitracin. Multidrug was the most predicted category and all the analyzed genomes had a predicted ARG sequence in multiple categories. It is worthy of note that strain cb109, which was isolated from a doxycycline resistant patient, was the only strain that had ARG sequences in all classes. Figure 3.13 shows the number of predicted sequences in different ARG categories.

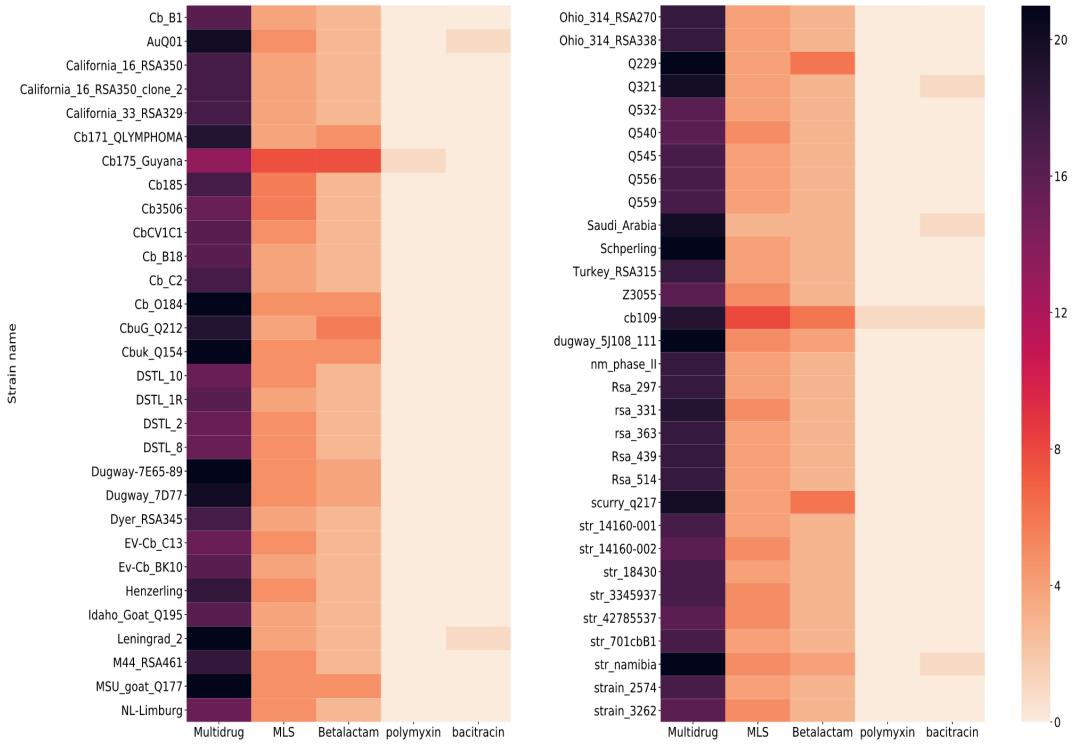


Figure 3.13: Total number of ARG sequences detected in the evaluated isolates of *Coxiella burnetii* depicted in a heatmap. Most of the predicted ARG sequences belong to the multidrug category.

3.2.4 *Predicted ARG in the genome of a doxycycline resistant C. burnetii strain.*

We profiled the genome of the cb109 strain with our ANN model. We predicted 19 ARGs under the multidrug category, six sequences under beta-lactam category, eight under the MLS category and one sequence was predicted as a polymyxin ARG. In comparison to other analyzed genomes, the CB109 genome had the most number of ARG sequences. Most of the sequences classified under the Multidrug category were annotated to be efflux proteins. The largest of this was acriflavine resistance protein, it had a length of 1022 amino acids and was annotated as a homolog of the AcrB/AcrD/AcrF family of protein sequences that has been reported to be involved in multidrug resistance [61].

Subunits of known multidrug proteins such as Bcr/CflA family efflux transporter, Multidrug resistance transporter A and D and Acriflavin resistance plasma membrane protein were also predicted in the genome of this isolate.

Homology annotation of the six amino acid sequences that were predicted as beta-lactam ARGs revealed that half of them had degenerate alignments with known beta-lactam resistance genes from the *Serratia* genus. This observation had earlier been discussed when the genome of the strain Nine Mile phase I RSA493 was published [148]. The remaining sequences were annotated as cell wall peptidases, which have been documented to be capable of binding penicillin [155] [73].

3.2.5 *Comparison of the resistome of acute and chronic Q fever strains*

We selected the genome of Scurry and Q212 as representatives of chronic Q fever isolates and Nine mile and Henzerling genome as representatives of acute Q fever isolates. We observed both Scurry and Q212 had the same number of detected ARGs, 30, and was slightly different from the number of predicted ARGs in the Nine mile, 25, and the Henzerling genome, 27.

Most of the predicted multidrug ARGs from the genomes of both the chronic and acute Q fever isolates were proteins involved in efflux systems. We observed that a few of the predicted sequences were subunits or had produced a degenerated alignment and others aligned to complete protein sequences.

In Table 3.3, we describe the observed state of the predicted ARG genes for both acute and chronic *C. burnetii* isolates.

Table 3.3: Predicted multidrug ARGs in acute and chronic isolates

Predicted ARG	Detection			
	Henzerling	493	Scurry	Q212
Multidrug resistance transporter Bcr family	Yes^{*d}	Yes	Yes	Yes
Bcr/CflA family efflux transporter	Yes	Yes	Yes	Yes
Acriflavin resistance periplasmic protein	Yes	Yes	Yes	Yes
Acriflavin resistance plasma membrane protein	Yes	Yes	Yes	Yes

Membrane fusion protein (MFP) family protein	Yes	Yes	No	No
Type I secretion outer membrane protein	Yes	Yes	Yes	Yes
Type I secretion outer membrane protein II	Yes	Yes	Yes	Yes^{*d}
Protein translocase subunit SecD	Yes	Yes	Yes	Yes
Protein translocase subunit SecF	Yes	Yes	Yes	Yes
Transporter, MFS superfamily	Yes	Yes	Yes^{*i}	Yes^{*i}
Multidrug resistance protein A	Yes	Yes	Yes	Yes
Multidrug resistance protein B	Yes^{*dn}	Yes^{*dn}	No	No

^{*d} contains a deletion ^{*i} incomplete subunit ^{*dn} degenerate sequence

Literature search of the predicted multidrug ARG revealed that most have been documented to play some role in multidrug resistance across various pathogens. For example, in *Escherichia coli* and *Bacillus subtilis*, accessory factor, SecDF, which was predicted to be a multidrug ARGs in *C. burnetii*, has been documented to be important export proteins. Also in *Staphylococcus aureus*, it has been shown to be capable of mediating resistance in [130]. We also predicted Type I secretion outer membrane protein as a multidrug ARG sequence. This has been documented as an important component of efflux systems implicated in multidrug efflux [149].

Based on the predicted multidrug resistome between the chronic isolates and the acute isolates, we observed no observable difference that could explain the difference in the clinical nature of the isolates.

We observed that the predicted beta-lactam ARG genes in both acute and chronic isolates had degenerated alignments with known beta-lactam sequences which corroborates Seshadri *et al* [148].

Antibiotics such as Macrolides, Lincosamides, and Streptogramins have distinct molecular structures, but they all interact with the peptidyl transferase (PTC) active site of the ribosome [111]. In both chronic and acute isolates, four ATP-Binding proteins were predicted to have MLS ARGs. Resistance to MLS drugs has been shown to be conferred by ATPases in both Gram positive and Gram negative bacteria [111] [96] [152] [6]. Clindamycin (lincosamide) in conjunction with doxycycline has been proven to lower mortality in chronic Q fever cases [134], and numerous newer macrolides have been demonstrated to modestly improve the number of febrile days in acute Q fever patients when compared to beta-lactamases but not as well as doxycycline [52].

3.3 GENOMIC ANALYSIS OF NOVEL STRAINS OF *C. burnetii*

3.3.1 *Genome features*

Genomic assembly of the isolate from sheep (TiHoQ-1091), resulted in 43 contigs and a genome length of 2039178 bp while the genomic assembly of the isolate from goat (TiHoQ-2219) resulted in 70 contigs and a length of 2000714 bp. Both genomes have a G+C content of 42.52 %.

The values of predicted genomic features are given in the table below: ´

Table 3.5: Predicted genomic features

	TiHoQ-1019	TiHoQ-2219
Total genes	2116	2055
Protein coding	2067	2006
RNA genes	49	49
Genes with predicted function	1381	1361
Genes assigned hypothetical function	686	645
Genes assigned COGs	816	810
Genes with signal peptides	167	166
Genes with Pfam domain	246	240

3.3.2 *Plasmid analysis*

The QpH1 plasmid was predicted as the plasmid species in both genomes using the CoxBase platform. We extracted the plasmid sequence using a custom Python script that appends contigs that were mapped to the QpH1 plasmid.

The QpH1 has been described as a virulence factor that aids in the colonization of bone marrow-derived murine macrophages [97]. We confirmed the presence of gene CBUA0037 and CBUA0038 in both genomes, these genes have been described as essential for plasmid maintenance [97], suggesting the plasmid is active in both isolates. We used Sibelia [107] to compare the plasmids to the QpH1 plasmid of RSA 493, we observed several Single Nucleotide Polymorphisms as well as deletions in several predicted genes. A total of 24 synteny blocks were generated in the TiHoQ-1019 and TiHoQ-2219 which aligned to 22 gene positions as seen in Figure 3.17 and Figure 3.18 respectively. We also compared the plasmids to the QpH1 with Mauve [33]. We could confirm our observation from the Sibelia anaylis that shows genomic rearrangement of a large block of genes that are co-linear between the TiHoQ-1019 plasmid and the QpH1 plasmid. We observed more rearrangements between the NM1 QpH1 plasmid and the TiHoQ-2219 plasmid.

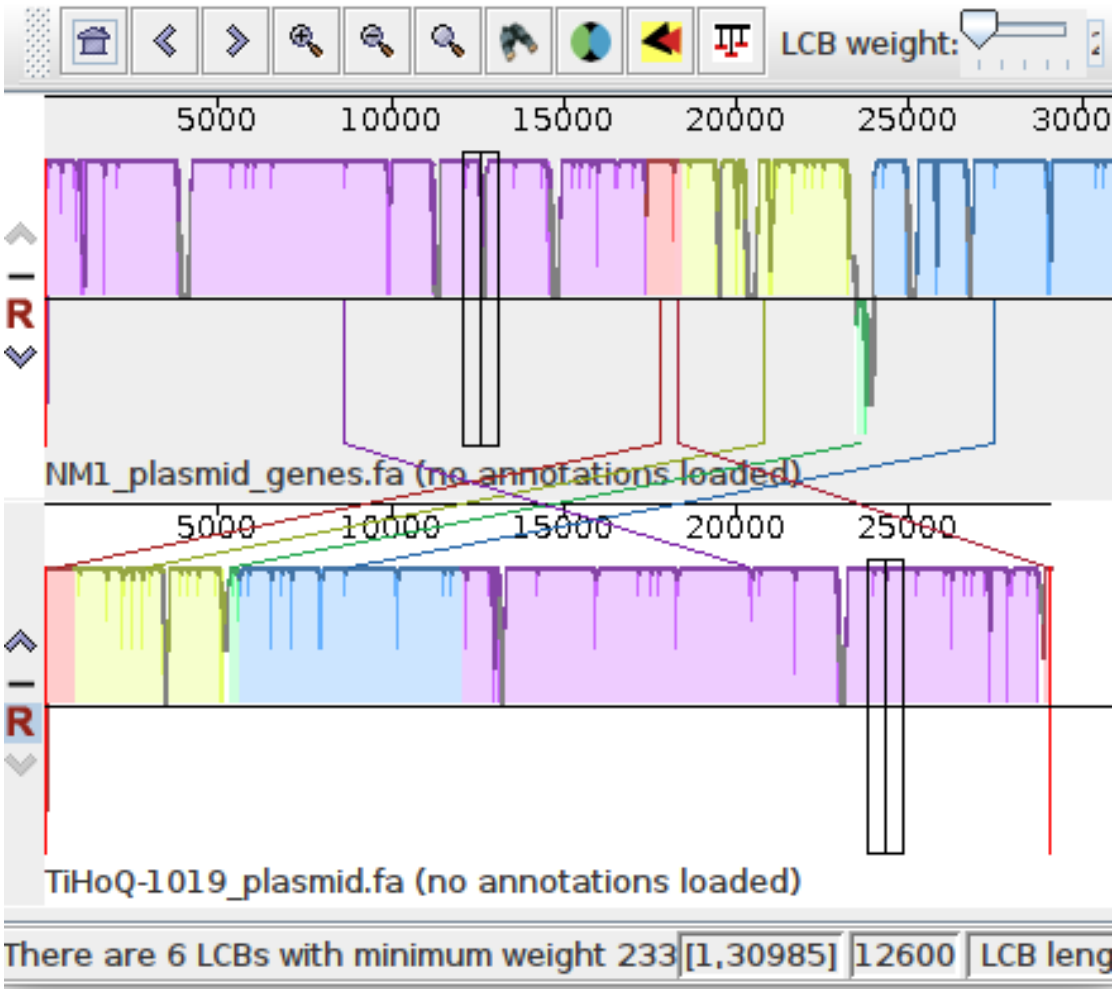


Figure 3.14: Screenshot of Mauve GUI visualization of the alignment between the QpH1 plasmid and the TiHoQ-1019 plasmid. LCBs are colored boxes and crossing lines shows positions where rearrangements have occured

We annotated the predicted genes with Prokka, 37 genes were successfully annotated (30 as hypothetical proteins and 7 with biological functions) in both

the TiHoQ-1019 and TiHoQ-2219 plasmids. The table of predicted genes is listed here Section A.3 in the appendix section.

The QpH1 plasmid has been reported to be enriched in genes encoding Dot/Icm substrates and six hypothetical proteins have been identified CpeA (CBUA0006), CpeB (CBUA0013), CpeC (CBUA0014), CpeD (CBUA0015), CpeE (CBUA0016), and CpeF (CBUA0023) [171]. We could confirm the existence of all the genes encoding Dot/Icm substrates in TiHoQ-1019, as shown in Figure 3.15. However, in the TiHoQ-2219 isolate, the CpeF (CBUA0023) gene had a major deletion of over 100 amino acid residues - only 119 aa residues of the 233 residues in CpeF was present, this is shown in Figure 3.16 .

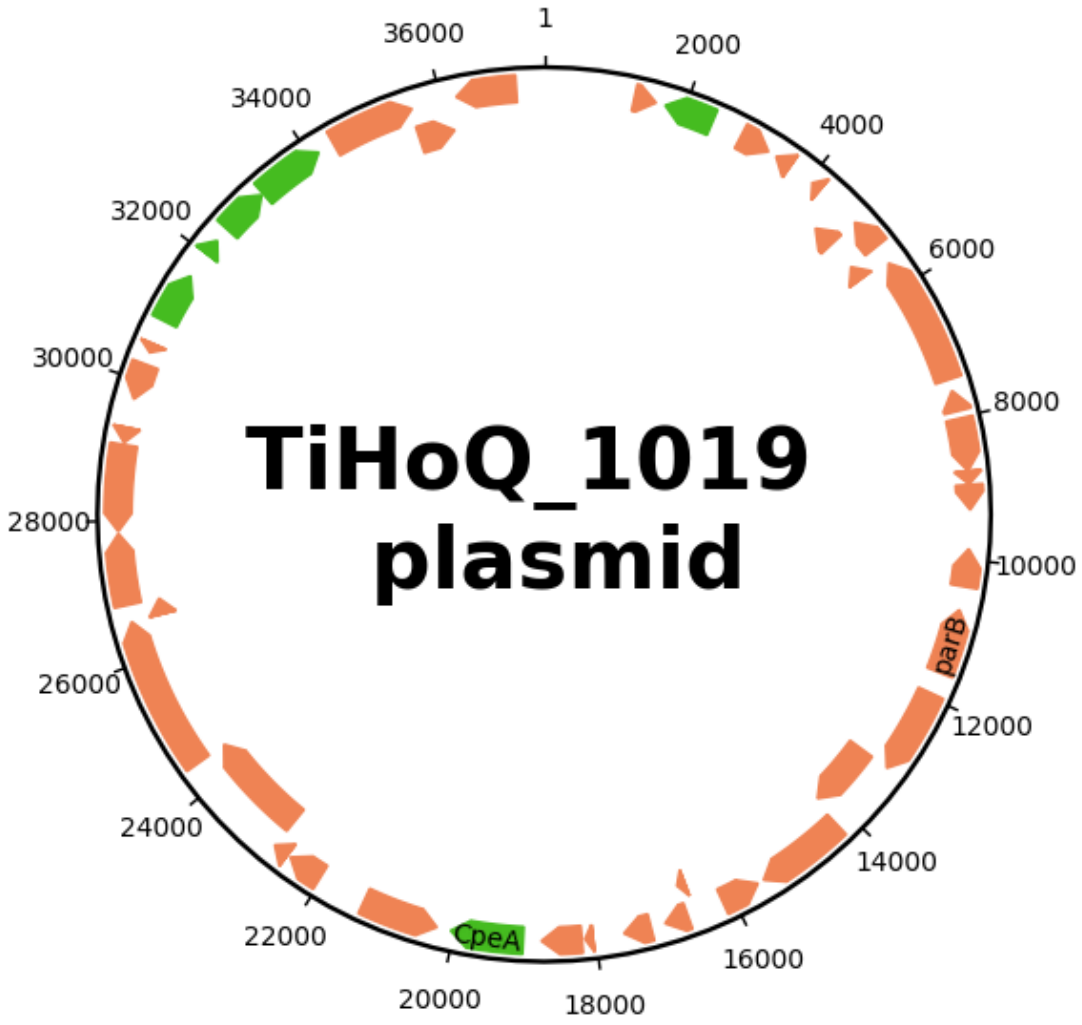


Figure 3.15: Plasmid map of the TiHoQ-1019 plasmid. Highlighted locus in green are locations of genes encoding Dot/Icm substrates.

3.3.3 *Chromosome analysis*

We predicted open reading frames with Glimmer then compared the predicted ORFs to that of ORFs from RSA 493 to estimate the genetic similarities and

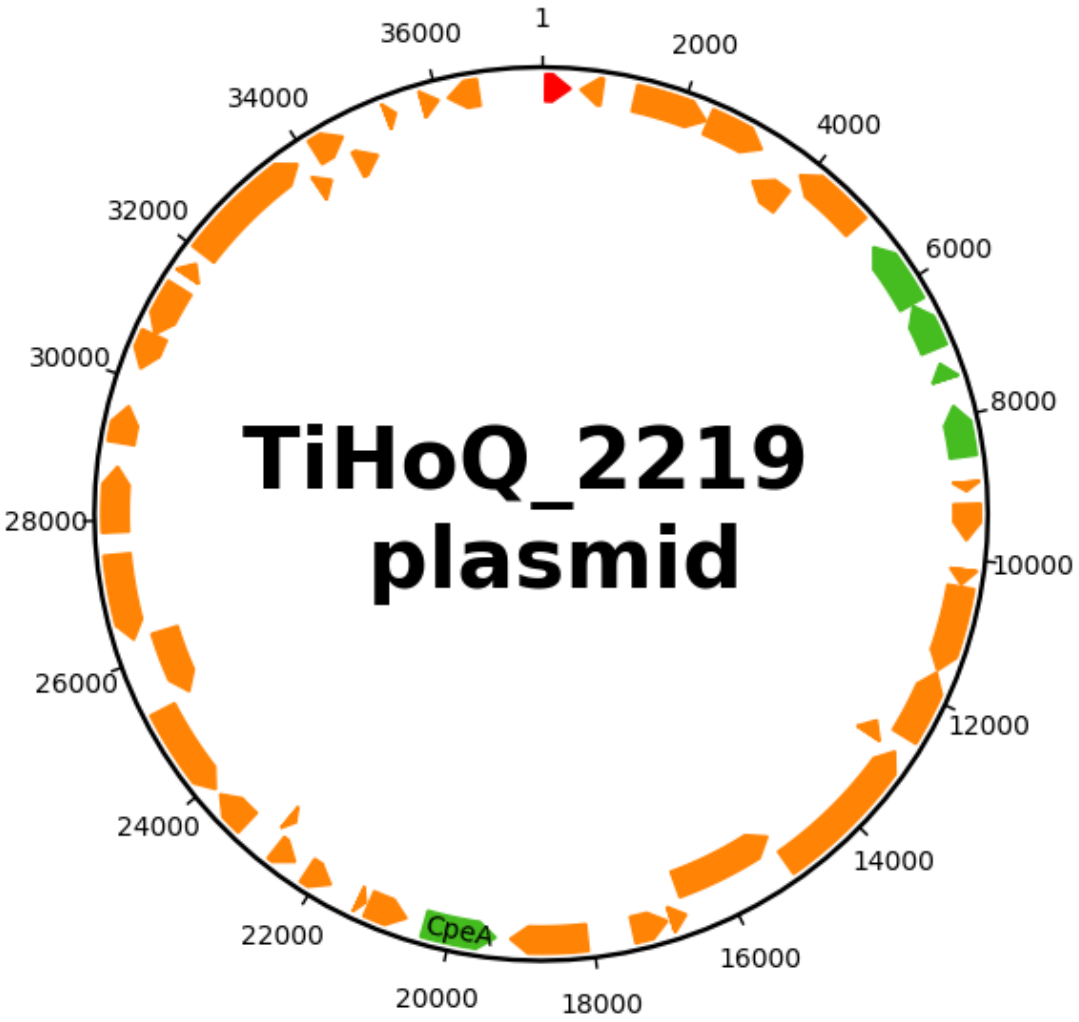


Figure 3.16: Plasmid map of the TiHoQ-1019 plasmid. Highlighted locus in green are locations of genes encoding Dot/Icm substrates. Highlighted in red is the incomplete CpeF fragment

variations between the two isolates. 1336 synteny blocks were generated between both genomes and RSA 493. There were 1047 synteny blocks with complete alignment shared between the TiHoQ-2219 genome and the NM1 strain of which over 803 had some form of SNP compared to the NM1 genome. 524 SNP occurred at start positions and due to the degeneracy of start codons can be effectively ignored. We observed SNP at 279 genes which occurred at 442 locations.

The genome of *C. burnetii* isolate, Nine Mile phase I RSA493 published in 2003 contained 28 insertion sequence elements; 20 belonging to the IS1111 family, 5 to the IS30 and 3 to the ISAs1 family [148]. We sought to identify the location of insertion elements in both genomes using BLASTN search of all the IS genes that were found in the Nine Mile phase I RSA493 genome. In the TiHoQ-1019 genome, we could only identify 2 intact IS families (IS30 and ISAs1) of the 3 families and the IS1111 family was incomplete. The TiHoQ-2219 genomes had 3 intact IS families (IS1111, IS30 and ISAs1), however only a single copy was intact and the other occurrences were incomplete.

The genome of a significant number of *C. burnetii* isolates have been annotated to contains pseudogenes (genes that carry inactivating substitutions, insertions or deletions in comparison to their coding ancestral source): Nine Mile phase I RSA493 was reported as having 83 of these genes with over representation in pathways that are no longer relevant to the obligate intracellular lifestyle of *C. burnetii* [148]. We investigated the occurrence of pseudo genes in the two isolates using pseudofinder [160]. We annotated 471 pseudogenes candidates in both TiHoQ-1019 and TiHoQ-2219. To filter out false positives, we carried out a blastx analysis of the annotated pseudogenes sequences against a dataset of proteins form Uniprot that was extracted using the search term "Coxiella". We only selected candidates that had a coverage of over 50% sequence identity to the complete length of their nearest matches and a evalue above 0.001 [93]. Using this method we obtained 400 and 401 pseudogene candidates for TiHoQ-1019 and TiHoQ-2219 candidates respectively. The percentage identity of all the alignments was between 26 % to 100 %. It has been reported that statistically significant homologs can share less than 20 % percentage identity [119]. We also filtered out pseudogenes with perfect homology to proteins with known functions, which we denoted as false positives. In total we had 326 and 324 pseudogene candidates for TiHoQ-1019 and TiHoQ-2219 after excluding the false positive candidates.

C. burnetii is pathogenic microbe that is restricted to an obligate intracellular lifestyle, therefore to successfully colonize a host it must carry a battery of virulence genes that will aid to subvert the host defense mechanisms. We profiled the analyzed genomes for known virulence factors in *C. burnetii* using of homology based on annotation with Prokka. We could identify proteins that have been described to aid in adhesion such Ankyrin repeats and Thyroglobulin type-1 repeat protein [148], intact Dot/Icm machinery for protein transport [110], several multidrug ABC transporters and other efflux transporters (efflux_EmrB: drug resistance MFS transporter, drug:H+ antiporter-2) essential for detoxification [83].

3.3.4 *Phylogenetic analysis*

We analyzed the genomes for relatedness to other *C. burnetii* genomes downloaded from RefSeq using the average nucleotide identity tool, Pyani [125], which uses the nucmer [38] for alignment purposes. We observed that the isolate TiHoQ-1091 was most similar to strain 2574 as shown in Figure 3.19, which is an isolate from Netherlands that was isolated from cattle. The isolate TiHoQ-2219 was similar to the strain Q556 based on the tree generated by the ParSNP aligner [166]. The Q556 isolate was isolated from a cattle in France.

3.3.5 *Genotyping*

Both isolates were genotyped as belonging to the MST group 20 and *ada*A positive. MLVA genotyping revealed that both TiHoQ-2219 and TiHoQ-1019 belonged to the C cluster based on the scheme used by Frangoulidis et al. [48]. IS1111 Genotyping reveal that TiHoQ-2219 lacked the presence of two transposases IS1111-16 and IS1111-32 which were present in TiHoQ-1019. SNP typing could not discriminate between both isolates as they have the same SNP profile based on the Hornstra et al. typing scheme [70]. All genotyping analysis was done on the CoxBase platform.

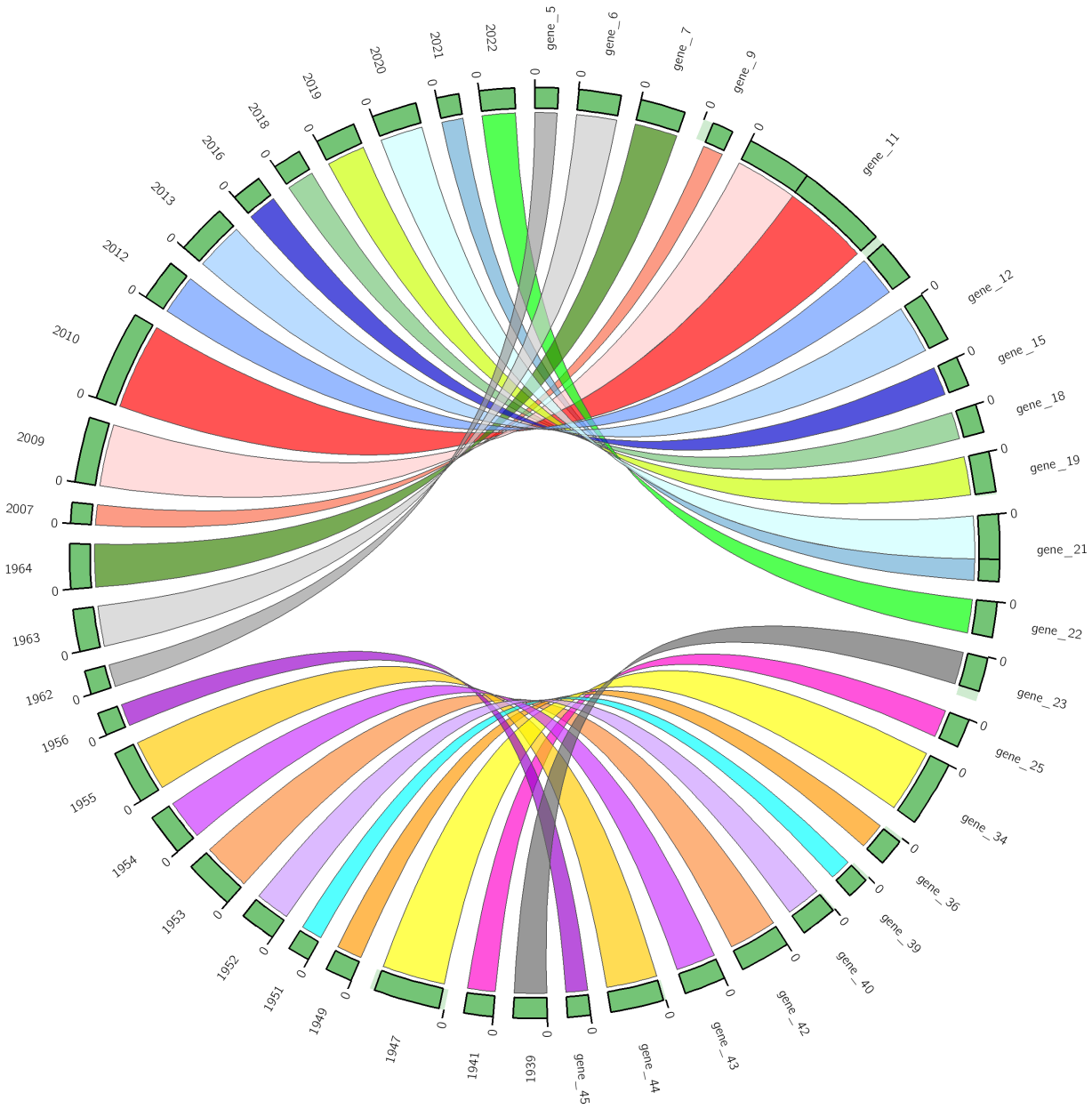


Figure 3.17: Circos visualization of synteny blocks alignment between TiHoQ-1019 and QpH1 plasmid from RSA 493. 22 gene positions were designated as region of synteny between the two plasmids.

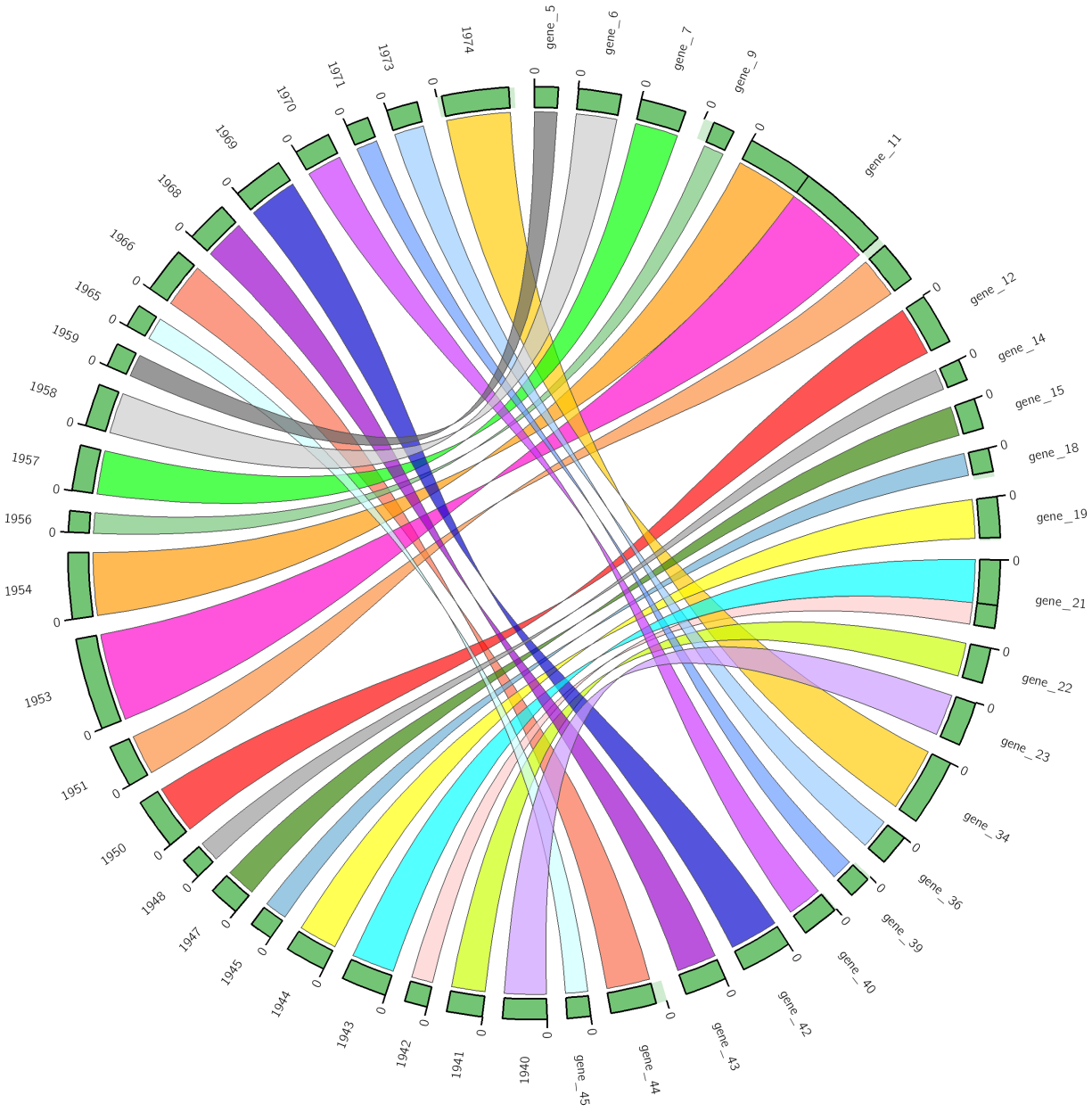


Figure 3.18: Circos visualization of synteny blocks alignment between TiHoQ-2219 and QpH1 plasmid from RSA 493.

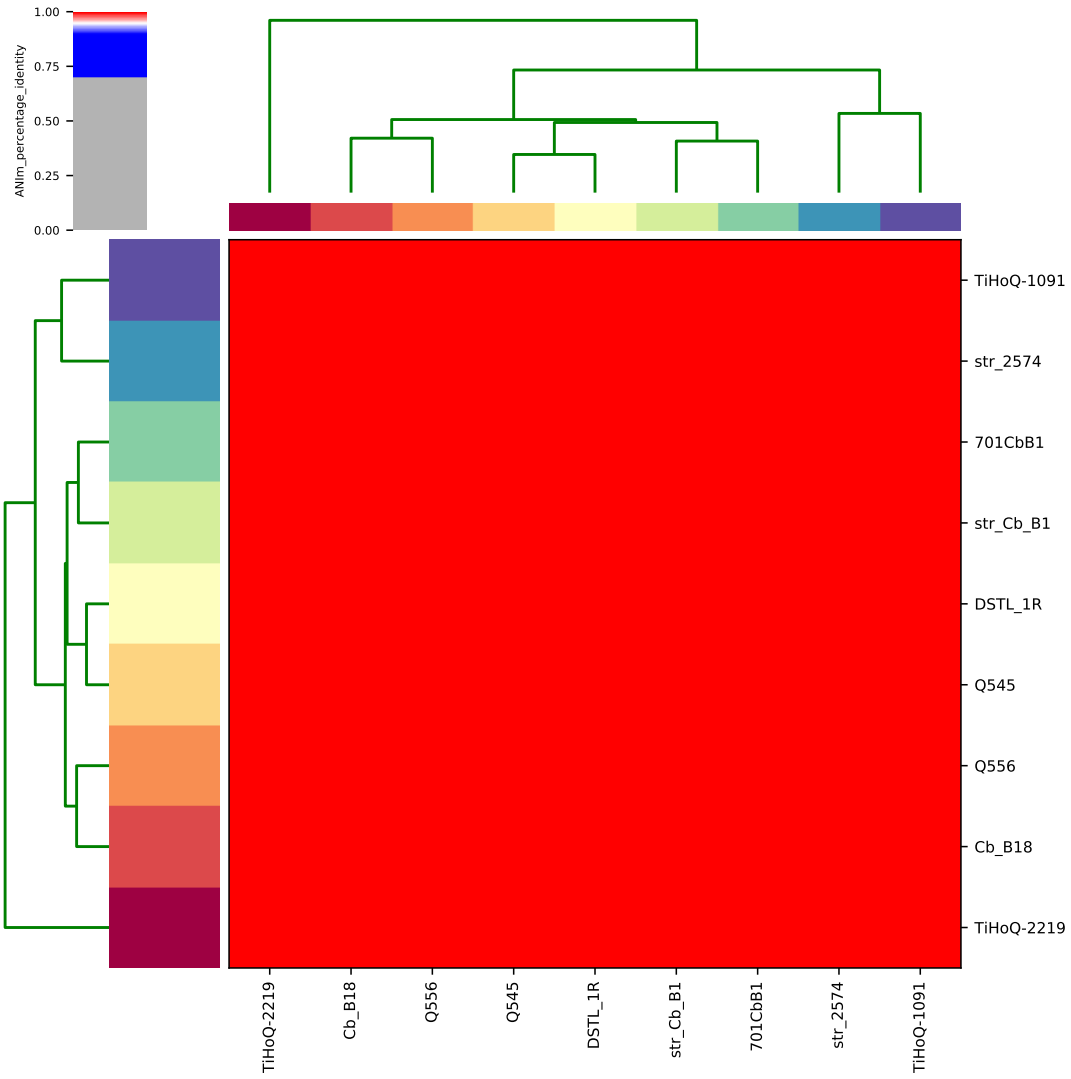


Figure 3.19: Average nucleotide identity plot of analyzed genomes shows that isolate TiHoQ-1019 is most closely related to strain 2574

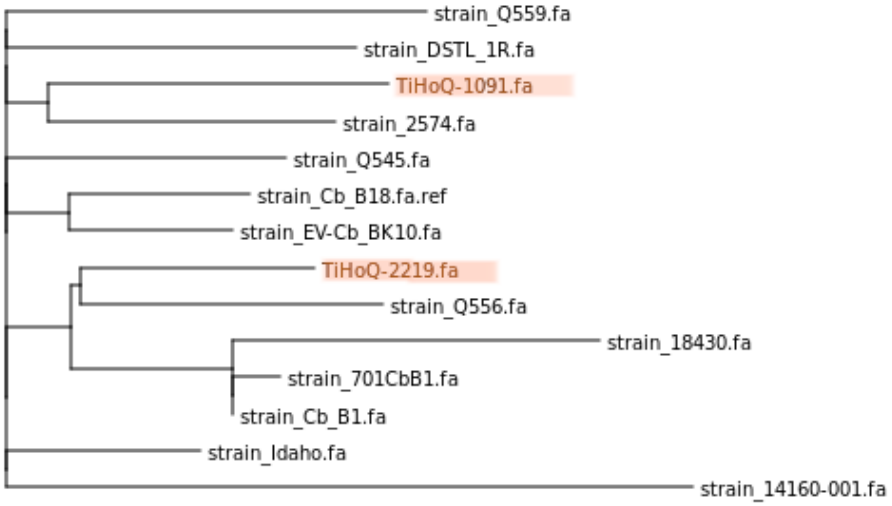


Figure 3.20: Unrooted phylogenetic tree highlighting the position of the two novel isolates. TiHoQ-1019 is confirmed as most closely related to strain 2574 and TiHoQ-2219 as closely related to strain Q556.

Part IV

DISCUSSION

4

DISCUSSION

4.1 COXBASE PLATFORM AS AN EPIDEMIOLOGICAL TOOL FOR ANALYZING AND GENOTYPING *C. burnetii* GENOMES

Genotyping of microbial agents, especially pathogenic ones, is important for the identification of novel strains as well as control and comparison of clinically relevant isolates. Therefore, genomic tools and software to facilitate this process should be readily available. Platforms that serve this purpose are fairly common in fields of economically important microorganisms but are often neglected in less "economically interesting" species.

C. burnetii is an economically interesting species because of the impact an outbreak could have on livestock and human health. Nonetheless, a stable platform that is easy to access and updated, to analyze *C. burnetii* isolates is not readily available. The main focus of this thesis was to implement such a platform and that aim has been realized with the accompanying introduction of several features that can aid the genotyping process. Techniques to distinguish between *C. burnetii* isolates is an "old" scientific task that has refused to be laid to rest with several genetic features such as plasmid, mucZ, com1, icd, and 16S/23S rRNA being early actors. The rise of high throughput sequencing prepared the stage for new participants such as interspacer sequences, repeat region, SNP and *ada*A gene. The details of these techniques are documented in the Introductory section. Our project titled "CoxBase", draw from these techniques to create a unified platform with the ability to combine or carry out a genotyping task with one or more of these techniques.

The evolution of different methods to distinguish between *C. burnetii* isolates is also a reflection of the difficulty to achieve a consensus typing scheme for this agent as most of the early methods lacked sufficient discriminatory power and also had reproducibility challenges. The newer methods also suffered intrinsic setbacks such as cost of materials as well as harmonization challenges. The optimal genotyping method in essence is a combination of several methods. In a wet lab setting this will be a cumbersome and expensive task. The advent of web technologies has made it possible to simulate these processes *in silico* and this is one of the strengths of the CoxBase platform.

The multiple genotyping approach is a data generating process and as such the obtained data will need to be stored in a place and also in a condition where it can be retrieved later and further analyzed. We have included such a feature into the platform as well. Results of genotyping analysis are saved in the database for future retrieval. Inadvertently, this also makes sharing it easier, which is also an element of reproducible science.

Visualization methods are an important part of data organization as well as sharing. We have included several features to visualize genomic data such as

genome browsers, to summarize data such as dynamic plots of metadata, geolocation maps for spatial comprehension of data as well as a central distribution map that displays all the isolates in the database in their respective areas of isolation. These features are intended to enable the quick understanding of the isolate data that are stored inside the database. We have also included features to make phylogenetic plots dynamically from data from the database as well as data obtained from a new analysis.

The CoxBase platform is not the first attempt at a web accessible platform that aims to provide a genotyping and database service for *C. burnetii*, existing platforms such as the MLVA platform (https://microbesgenotyping.i2bc.paris-saclay.fr/) and MST platform (https://ifr48.timone.univ-mrs.fr/mst/coxiella_burnetii/strains.html) focus on a single genotyping method. CoxBase distinguishes itself as the first multi genotyping and database platform. Unlike the first attempts the analysis option is not exclusive to a genotyping method and it also provides more accompanied features compared to the older genotyping platforms.

The novel features that we have included on the CoxBase platform include:

- Interactive distribution map of *C. burnetii* isolates in our database: The purpose of this feature is to summarize the data in our database at a country level. This should hopefully serve as a quick feature that can be use to access information about *C. burnetii* isolates from a specific country. This feature is unique to the CoxBase platform.

- *in silico* genotyping of whole genomes using SNP method from Hornstra et al. [70]: This genotyping feature is unique to the CoxBase platform. Although the method is not as popular as the MLVA and MST method, it can be used as a secondary or confirmatory genotyping method.

- *in silico* genotyping of whole genomes using the IS1111 primers: This genotyping feature is also unique to the CoxBase platform. It is a relatively new genotyping method and it has been reported to have high discriminatory power. It can also be used as a secondary genotyping method.

- *in silico* genotyping of whole genomes using the adaA primers: This genotyping feature was included as a secondary genotyping method and it is also unique to the CoxBase platform.

- Historical query of genotyping results: Only the CoxBase platform offers this feature in comparison to other *C. burnetii* genotyping platforms. It's aimed at improving data portability, reproducibility and findability.

- Multi field aggregation query of MLVA genotyping data: This implementation makes it possible for researchers to fine tune their query for genotyping data using a combination of metadata fields. The aggregation method is also only offered on the CoxBase platform in comparison to the other *C. burnetii* genotyping platforms.

- Dynamic phylogenetic tree plots of MLVA profiles based on Shriver distance: The Shriver distance is an appropriate metric for estimating genetic distance between repeats of *C. burnetii* origin because it accounts for larger number of alleles and higher levels of heterozygosity than classic genetic distance measures [150]. We have introduced a feature to create dynamic phylogenetic trees that can be used to estimate relatedness of novel MLVA profiles to historical profiles.

- Reporting of novel MST alleles with Visualization of alignment and query of MST groups: In contrast to the MST platform (https://ifr48.timone.univ-mrs.fr/mst/coxiella_burnetii/strains.html), we have introduced querying features for MST groups based on spacer alleles as well as visualization of novel allele alignments in order for researchers to be able to distinguish between a true novel sequence and a false positive which can be easily spotted due to the appearance of gaps or deletions or bad nucleotides (N) in the alignment.

- Integration of selected CoxBase genomes with a genome viewer: The CoxBase platform is the only *C. burnetii* genotyping platform that is integrated with a genome browser, the purpose of which is to facilitate the visualization and development of genotyping markers.

- Provision of a unique classification of the MLVA 6 panel genotype: The MLVA 6 panel unlike the MLVA 14 panel introduced by Frangoulidis et al. [48] has no universally accepted genotype for the unique profiles. We have introduced a provisional genotype based on integers to discriminate between unique profiles.

4.2 A MACHINE LEARNING APPROACH TO PROFILING THE RESISTOME IN *C. burnetii*

The detection of new antibiotic resistance genes (ARG) is a phenomenon that in the best case scenario should be in the laboratory and not in the clinics - because then it could be too late and life-threatening. This statement implies that this should be a routine task, sort of a surveillance approach to profile isolates for genes that could reduce the efficacy of antibiotics. Methods to identify ARGs rely on the "best hit" approach and the drawback of this I have already discussed in previous chapters. This calls for the discovery of new approaches to profile ARGs.

Our attempt at this challenge was to design a machine learning model based on an artificial neural network that can predict ARGs from protein sequences. Several authors have applied a similar methodology to characterize ARGs from both metagenomic and genomic data as discussed in previous chapters. The uniqueness of our approach as compared to other authors is in the use of PSSM composition method to encode protein features that were used for the prediction. Although the method has been used to predict other genomic features such as Type IV effectors [183] DNA-binding proteins [4]. This is

the first time to my knowledge that this method was used to encode protein sequences for ARG prediction.

We chose to use a resistome approach because it gives a global snapshot of antibiotic resistance genes that are present in the analyzed genomes and thus a better chance at comparison of multiple isolates as opposed to scanning the genomes for individual antibiotic resistance genes.

We designed two models with different methods used for feature encoding. The better performing model was the one that uses the PSSM-Comp algorithm for encoding of amino acid sequences. We hypothesize that the one reason for this might be the size of the feature vector. The feature vector for the better performing model was 400 dimension in comparison to the less performing model with 136 dimension and as such it might model the data better compared to the later.

We observed that the predicted number of ARGs in the doxycycline resistant isolate was higher in comparison to the reference non-resistant clinical isolates. This data fit into the clinical observation, however we couldn't find any supporting evidence in the genome that could explain the resistance pattern observed in the resistant isolate. This is not an unusual phenomenon, moreover it was reported by the author when the genome was published that they couldn't substantiate any reasonable findings to explain the resistant phenotype within the genome [138].

Data about molecular investigation into antibiotic resistance in *C. burnetii* are few due to reasons previously discussed in Chapter 1. Our model predicted ARGs mostly in the multi-drug category, which were annotated as majorly proteins involved in efflux mechanisms. This was in comparison with reports from Zuñiga-Navarrete et al., where they showed that transport system proteins are important for defense against reactive oxygen species from antibiotic exposure [189].

Our results also echoes what is already documented in the literature concerning resistance to beta-lactams in *C. burnetii* isolates for example, all the detected beta-lactam genes were degenerated in comparison to functional beta-lactam genes [148].

To our knowledge this is the first project to profile resistance genes in *C. burnetii* at a global genomic level using a machine learning approach.

4.3 GENOMIC ANALYSIS OF *Coxiella* ISOLATES

The observed statistics for the genomic features of both isolates are in line with what was reported by Seshadri et al. and Walter et al. for the *C. burnetii* isolate from Namibia and Nine Mile, USA respectively [148] [174]. The DNA GC content is almost identical to the Namibia isolate however, the genomes sizes were bigger compared to the Nine mile isolate [148] [174].

The plasmid on both isolates is the QpH1 plasmid. This plasmid is also carried by the reference strain RSA 493. However, the plasmid from the two isolates carried several SNPs. We annotated 37 genes in the QpH1 plasmid in

the two isolates, this was slightly fewer than the QpH1 plasmid sequence with 39 genes (accession number: AE016829.2) submitted by Seshadri et al. and the QpH1 plasmid sequence with 48 genes (accession number: NC_002118.1) submitted by Thiele et al. [162]. The discrepancies we hypothesize might be as a result of the differences in the analysis workflow that was used as both Seshadri et al. and Thiele et al. used different gene prediction workflows and it is also worthy of note that Thiele et al. reported seven pseudo genes.

Whole genome comparative analysis with average nucleotide analysis (ANI) suggests that the isolate TiHoQ-1019 is closer to strain 2574 than the Nine mile RSA 493 strain. We could confirm this with the percentage of aligned bases to the reference genome which was 99.875% in the strain 2574 and 98.151% than in the NM 1 genome. The isolate TiHoQ-2219 was also closer to the strain 2574 than the NM 1 genome. This suggests indeed that the two isolates are European isolates.

Genotyping of the two isolates was carried out via CoxBase. We could affirm that both strains belonged to the MST 20. This group consist predominantly of isolates from France and Spain according to data from the MST data collection [54]. This also lends credibility to the hypothesis that the isolates are European and were not imported from a different continent. The MLVA genotyping results indicate that both isolates belong to MLVA genotype C, which consists of mostly European isolates, this also suggests that the isolates are of European origin. Although, one could argue that most of the *C. burnetii* isolates present in the different databases are European, hence, it is very likely that any observed interpretation might be due to chance. This is true and suggests that our observations should be taken with care.

We hope that the new isolates can add to the body of knowledge that is already present on *C. burnetii* genomes.

4.4 CONCLUSION

In this thesis, I developed an open web-based platform with genome analysis capabilities and as such succeeded in laying the foundation for a new approach to internet-based genotyping of *C. burnetii* isolates. We utilized mostly open-source tools in order to facilitate ease of collaborations as well as continuity of the project when the funding period is over. The implemented features we hope will facilitate quick identification, comparison, and discovery of novel and historical *C. burnetii* isolates. Statistics about the usage of the platform within the time frame 2020-01 to 2022-03 is shown in the Figure 4.1 and Figure 4.2. During this period, the most frequently used genotyping method is the MLVA genotyping method followed by the MST genotyping method.

I have also contributed to the existing data on resistance about *C. burnetii* isolates via the development of a machine learning model that is capable of profiling whole genomes. The investigation into the resistome of several *C. burnetii* genomes, affirmed information that was already known and also shed lights on new paths such as the finding about MLS resistance as well polymyxin

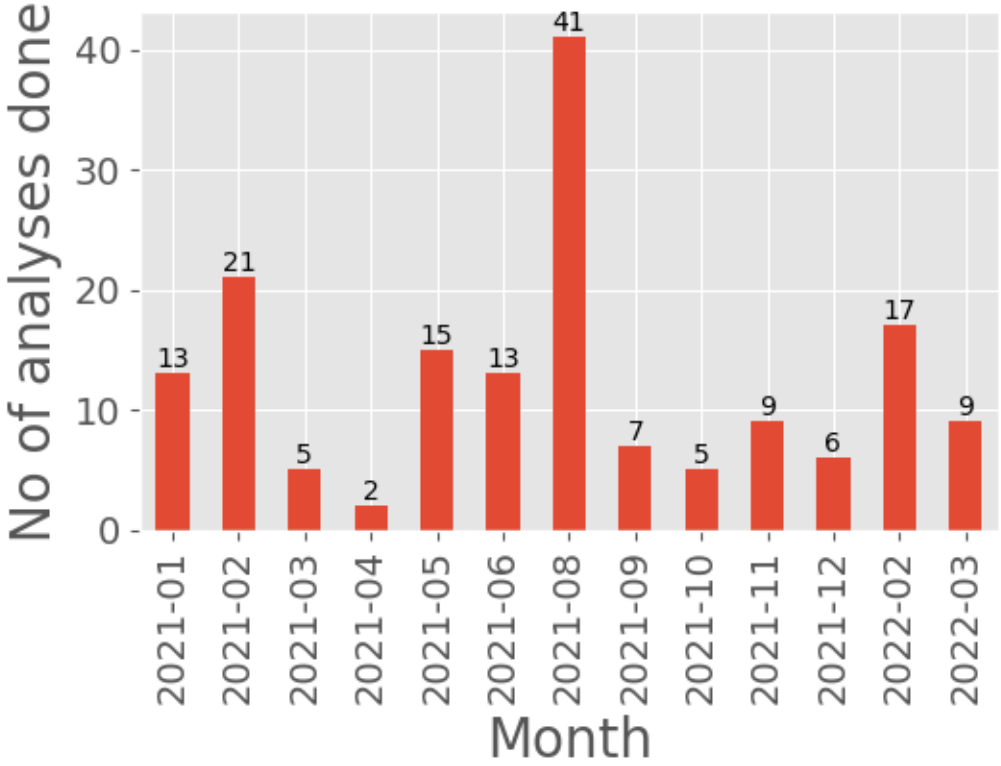


Figure 4.1: Usage stats on the total number of analyses done per month on CoxBase from 2021-01 to 2022-03

resistance. This hypothesis will definitely require a wet laboratory experiment for confirmation, nevertheless our approach suggests this is the right direction in the quest for methods to profile resistance in a bacteria species that is difficult to profile using existing methods such as *C. burnetii*.

Finally, Our analysis of the two novel *C. burnetii* isolates revealed that they are not closely related to the NM 1 genome but to isolates from the European continent.

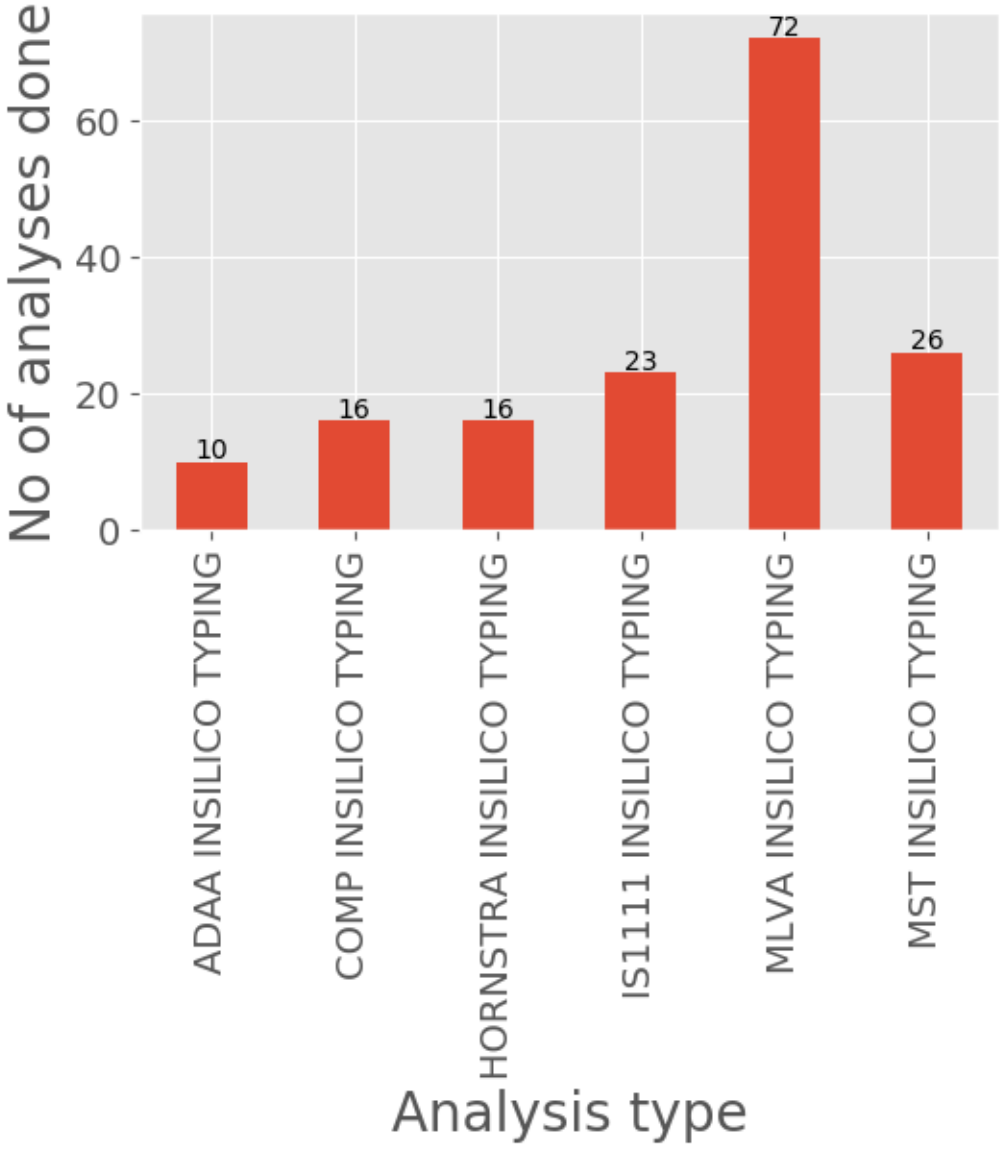


Figure 4.2: Usage stat on the total number of analysis done per genotyping feature on CoxBase from 2021-01 to 2022-03

Part V

OUTLOOK

5

OUTLOOK

5.1 OUTLOOK

This thesis has led to the development of tools to assist the internet based genome typing of *C. burnetii*. The maintenance of these tools as well as the introduction of new isolate into the database is important in order to keep the platform updated. As outlook, we will like to update the database regularly and maintain the underlining software. The platform is currently running on de.NBI infrastructure which is public and government funded for projects across Germany. We hope to collaborate with ZB MED – Information Centre for Life Sciences, Cologne, Germany, to deploy the project application into a Kubernetes cluster when the life time of the project runs out.

Also, we would like to extend the application using container technology. The reason for this is to ensure ease of portability of the entire application in case we need to switch deployment servers. For this we would like to use Docker [104] as it is the most popular container technology, Linux based and very easy to use and port configurations.

We are also working on the evaluation of predicted antibiotic resistance genes using the progenomes [103] database as well as a possible wet lab collaboration. The progenomes database was selected because it contains high quality functional annotation of antibiotic resistance genes based on Hidden Markov Model predictions observed from the integration of the ResFams [51] and CARD [75] databases.

Finally, we would also like to compile the manuscript for the methods and results of the global resistome prediction.

BIBLIOGRAPHY

[1] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[2] Inc Agendaless Consulting. *The Pyramid Web Framework v1.10.4*. URL: https://trypyramid.com/ (visited on 08/13/2020).

[3] Jørgen S Agerholm. "Coxiella burnetii associated reproductive disorders in domestic animals-a critical review." In: *Acta Veterinaria Scandinavica* 55.1 (Feb. 18, 2013), p. 13. ISSN: 0044-605X. DOI: 10.1186/1751-0147-55-13. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3577508/ (visited on 03/03/2020).

[4] Shandar Ahmad and Akinori Sarai. "PSSM-based prediction of DNA binding sites in proteins." In: *BMC bioinformatics* 6 (2005), p. 33.

[5] Sabah Al-Fedaghi. "Developing Web Applications." In: *International Journal of Software Engineering and Its Applications* 5 (May 2011). DOI: 10.1007/978-1-4302-3531-6_12.

[6] J. Allignet, V. Loncle, and N. el Sohl. "Sequence of a staphylococcal plasmid gene, vga, encoding a putative ATP-binding protein involved in resistance to virginiamycin A-like antibiotics." eng. In: *Gene* 117.1 (Aug. 1992), pp. 45–51. ISSN: 0378-1119. DOI: 10.1016/0378-1119(92)90488-b.

[7] Sharon Amit, Shiri Shinar, Ora Halutz, Yafit Atiya-Nasagi, and Michael Giladi. "Suspected Person-to-Person Transmission of Q Fever Among Hospitalized Pregnant Women." In: *Clinical Infectious Diseases* 58.11 (June 1, 2014), e146–e147. ISSN: 1058-4838. DOI: 10.1093/cid/ciu151. URL: https://academic.oup.com/cid/article/58/11/e146/2895430 (visited on 03/04/2020).

[8] Rolf Apweiler et al. "UniProt: the Universal Protein knowledgebase." eng. In: *Nucleic Acids Research* 32.Database issue (Jan. 2004), pp. D115–119. ISSN: 1362-4962. DOI: 10.1093/nar/gkh131.

[9] Rolf Apweiler et al. "Activities at the Universal Protein Resource (UniProt)." In: *Nucleic Acids Research* 42 (2014), pp. D191 –D198.

[10] Gustavo Arango-Argoty, Emily Garner, Amy Pruden, Lenwood S. Heath, Peter Vikesland, and Liqing Zhang. "DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data." In: *Microbiome* 6.1 (Feb. 2018), p. 23. ISSN: 2049-2618. DOI: 10.1186/s40168-018-0401-z. URL: https://doi.org/10.1186/s40168-018-0401-z (visited on 08/26/2020).

[11] Nathalie Arricau-Bouvery, Yolande Hauck, Awatef Bejaoui, Dimitrios Frangoulidis, Christelle C. Bodier, Armel Souriau, Hermann Henry Bernard Meyer, Heinrich Karl Johann Neubauer, Annie Rodolakis, and Gilles Vergnaud. "Molecular characterization of Coxiella burnetii isolates by infrequent restriction site-PCR and MLVA typing." In: *BMC Microbiology* 6 (2005), pp. 38 –38.

[12] Pierre Baldi and Peter Sadowski. "The Dropout Learning Algorithm." In: *Artificial intelligence* 210 (May 2014), pp. 78–122. ISSN: 0004-3702. DOI: 10.1016/j.artint.2014.02.004. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3996711/ (visited on 08/27/2020).

[13] Anton Bankevich et al. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." In: *Journal of computational biology : a journal of computational molecular cell biology* 19 5 (2012), pp. 455–77.

[14] Michael Bayer. *SQLAlchemy: The Python SQL Toolkit and Object Relational Mapper*. URL: https://www.sqlalchemy.org/ (visited on 08/14/2020).

[15] Paul A Beare, James E Samuel, Dale Howe, Kimmo Virtaneva, Stephen F Porcella, and Robert A Heinzen. "Genetic diversity of the Q fever agent, Coxiella burnetii, assessed by microarray-based whole-genome comparisons." In: *Journal of bacteriology* 188.7 (2006), pp. 2309–24.

[16] Jannick Dyrløv Bendtsen, Henrik Nielsen, Gunnar von Heijne, and Søren Brunak. "Improved prediction of signal peptides: SignalP 3.0." In: *Journal of molecular biology* 340 4 (2004), pp. 783–95.

[17] Fanny Berglund, Nachiket P. Marathe, Tobias Österlund, Johan Bengtsson-Palme, Stathis Kotsakis, Carl-Fredrik Flach, D. G. Joakim Larsson, and Erik Kristiansson. "Identification of 76 novel B1 metallo - β- lactamases through large-scale screening of genomic and metagenomic data." eng. In: *Microbiome* 5.1 (2017), p. 134. ISSN: 2049-2618. DOI: 10.1186/s40168-017-0353-8.

[18] Tim Berners-Lee. *Information Management: A Proposal*. https://www.w3.org/History/1989/proposal.html. [Online; accessed 17-August-2020]. 1989.

[19] Tim Berners-Lee and Dan Connolly. "Hypertext Markup Language (HTML): A Representation of Textual Information and MetaInformation for Retrieval and Interchange." In: 2003.

[20] John Besemer, Alexandre Lomsadze, and Mark Borodovsky. "GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions." In: *Nucleic acids research* 29 12 (2001), pp. 2607–18.

[21] Aitor Blanco-Míguez, Florentino Fdez-Riverola, Borja Sánchez, and Anália Lourenço. "BlasterJS: A novel interactive JavaScript visualisation component for BLAST alignment results." en. In: *PLOS ONE* 13.10 (Oct. 2018), e0205286. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0205286. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205286 (visited on 02/18/2020).

[22] B. Boeckmann. "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003." In: *Nucleic Acids Research* 31.1 (2003), pp. 365–370. DOI: 10.1093/nar/gkg095. URL: https://doi.org/10.1093%2Fnar%2Fgkg095.

[23] Robert E. Brennan and James E. Samuel. "Evaluation of Coxiella burnetii antibiotic susceptibilities by real-time PCR assay." eng. In: *Journal of Clinical Microbiology* 41.5 (May 2003), pp. 1869–1874. ISSN: 0095-1137. DOI: 10.1128/jcm.41.5.1869-1874.2003.

[24] F. M. Burnet and Mavis Freeman. "Experimental Studies on the Virus of "Q" Fever." In: *Reviews of Infectious Diseases* 5.4 (July 1, 1983), pp. 800–808. ISSN: 0162-0886. DOI: 10.1093/clinids/5.4.800. URL: https://academic.oup.com/cid/article/5/4/800/415223 (visited on 02/24/2020).

[25] P R Burton, J Stueckemann, and D Paretsky. "Electron microscopy studies of the limiting layers of the rickettsia Coxiella burneti." In: *Journal of Bacteriology* 122.1 (Apr. 1975), pp. 316–324. ISSN: 0021-9193. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC235672/ (visited on 02/27/2020).

[26] Ana Victoria Bustamante and Andrea Mariel Sanso. "Multiple-Locus Variable-Number of Tandem-Repeats Analysis as Subtyping Technique for Food-Borne Pathogens." In: 2018.

[27] Te-Chin Chu, Chen-Hua Lu, Tsunglin Liu, Greg C. Lee, Wen-Hsiung Li, and Arthur Chun-Chieh Shih. "Assembler for de novo assembly of large genomes." In: *Proceedings of the National Academy of Sciences* 110 (2013), E3417 –E3424.

[28] Manuel G. Claros and Gunnar von Heijne. "TopPred II: an improved software for membrane protein structure predictions." In: *Computer applications in the biosciences : CABIOS* 10 6 (1994), pp. 685–6.

[29] Jill Hager Cocking et al. "Selective whole genome amplification and sequencing of Coxiella burnetii directly from environmental samples." en. In: *Genomics* 112.2 (Mar. 2020), pp. 1872–1878. ISSN: 0888-7543. DOI: 10.1016/j.ygeno.2019.10.022. URL: http://www.sciencedirect.com/science/article/pii/S0888754319302058 (visited on 11/16/2020).

[30] Sherry A. Coleman, Elizabeth R. Fischer, Dale Howe, David J. Mead, and Robert A. Heinzen. "Temporal analysis of Coxiella burnetii morphological differentiation." In: *Journal of Bacteriology* 186.21 (Nov. 2004), pp. 7344–7352. ISSN: 0021-9193. DOI: 10.1128/JB.186.21.7344-7352.2004.

[31] Longzhu Cui, Hui min Neoh, Mitsutaka Shoji, and Keiichi Hiramatsu. "Contribution of vraSR and graSR Point Mutations to Vancomycin Resistance in Vancomycin-Intermediate Staphylococcus aureus." In: *Antimicrobial Agents and Chemotherapy* 53.3 (2009), pp. 1231–1234. DOI: 10.1128/aac.01173-08. URL: https://doi.org/10.1128%2Faac.01173-08.

[32] Aaron C.E. Darling, Bob Mau, Frederick R. Blattner, and Nicole T. Perna. "Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements." In: *Genome Research* 14.7 (July 2004), pp. 1394–1403. ISSN: 1088-9051. DOI: 10.1101/gr.2289704. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC442156/ (visited on 11/25/2020).

[33] Aaron CE Darling, Bob Mau, Frederick R Blattner, and Nicole T Perna. "Mauve: multiple alignment of conserved genomic sequence with rearrangements." In: *Genome research* 14.7 (2004), pp. 1394–1403.

[34] Gordon E. Davis, Herald R. Cox, R. R. Parker, and R. E. Dyer. "A Filter-Passing Infectious Agent Isolated from Ticks." In: *Public Health Reports (1896-1970)* 53.52 (1938), pp. 2259–2282. ISSN: 0094-6214. DOI: 10.2307/4582746. URL: https://www.jstor.org/stable/4582746 (visited on 02/24/2020).

[35] Frank B. Dean, John R. Nelson, Theresa L. Giesler, and Roger S. Lasken. "Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification." In: *Genome Research* 11.6 (June 2001), pp. 1095–1099. ISSN: 1088-9051. DOI: 10.1101/gr.180501. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC311129/ (visited on 11/16/2020).

[36] Frank B. Dean et al. "Comprehensive human genome amplification using multiple displacement amplification." en. In: *Proceedings of the National Academy of Sciences* 99.8 (Apr. 2002). Publisher: National Academy of Sciences Section: Biological Sciences, pp. 5261–5266. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.082089499. URL: https://www.pnas.org/content/99/8/5261 (visited on 11/16/2020).

[37] Arthur L. Delcher, Kirsten A. Bratke, Edwin C. Powers, and Steven Salzberg. "Identifying bacterial genes and endosymbiont DNA with Glimmer." In: *Bioinformatics* 23 6 (2007), pp. 673–9.

[38] Arthur L Delcher, Adam Phillippy, Jane Carlton, and Steven L Salzberg. "Fast algorithms for large-scale genome alignment and comparison." In: *Nucleic acids research* 30.11 (2002), pp. 2478–83.

[39] Amy M. Denison, Herbert A. Thompson, and Robert F. Massung. "IS1111 insertion sequences of Coxiella burnetii: characterization and use for repetitive element PCR-based differentiation of Coxiella burnetii isolates." In: *BMC Microbiology* 7 (2007), pp. 91 –91.

[40] E. H. Derrick. ""Q" Fever, a New Fever Entity: Clinical Features, Diagnosis and Laboratory Investigation." In: *Reviews of Infectious Diseases* 5.4 (July 1, 1983), pp. 790–800. ISSN: 0162-0886. DOI: 10.1093/clinids/5.4.790. URL: https://academic.oup.com/cid/article/5/4/790/415211 (visited on 02/24/2020).

[41] Jie Dong, Zhi-Jiang Yao, Lin Zhang, Feijun Luo, Qinlu Lin, Ai-Ping Lu, Alex F. Chen, and Dong-Sheng Cao. "PyBioMed: a python library for various molecular representations of chemicals, proteins and DNAs and their interactions." In: *Journal of Cheminformatics* 10.1 (Mar. 2018), p. 16. ISSN: 1758-2946. DOI: 10.1186/s13321-018-0270-2. URL: https://doi.org/10.1186/s13321-018-0270-2 (visited on 08/26/2020).

[42] SEAN R. EDDY. "A NEW GENERATION OF HOMOLOGY SEARCH TOOLS BASED ON PROBABILISTIC INFERENCE." In: *Genome Informatics 2009*. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO., 2009. DOI: 10.1142/9781848165632_0019. URL: https://doi.org/10.1142%2F9781848165632_0019.

[43] Robert C. Edgar. "Search and clustering orders of magnitude faster than BLAST." en. In: *Bioinformatics* 26.19 (Oct. 2010), pp. 2460–2461. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq461. URL: https://academic.oup.com/bioinformatics/article/26/19/2460/230188 (visited on 01/27/2020).

[44] Carole Eldin, Cléa Mélenotte, Oleg Mediannikov, Eric Ghigo, Matthieu Million, Sophie Edouard, Jean-Louis Mege, Max Maurin, and Didier Raoult. "From Q Fever to Coxiella burnetii Infection: a Paradigm Change." In: *Clinical Microbiology Reviews* 30.1 (Jan. 1, 2017), pp. 115–190. ISSN: 0893-8512, 1098-6618. DOI: 10.1128/CMR.00045-16. URL: https://cmr.asm.org/content/30/1/115 (visited on 02/27/2020).

[45] D. B. Fishbein and D. Raoult. "A cluster of Coxiella burnetii infections associated with exposure to vaccinated goats and their unpasteurized dairy products." In: *The American Journal of Tropical Medicine and Hygiene* 47.1 (July 1992), pp. 35–40. ISSN: 0002-9637. DOI: 10.4269/ajtmh.1992.47.35.

[46] The Apache Software Foundation. *Apache HTTP Server*. URL: https://httpd.apache.org/ (visited on 08/13/2020).

[47] Dimitrios Frangoulidis et al. "Microevolution of the Chromosomal Region of Acute Disease Antigen A (adaA) in the Query (Q) Fever Agent Coxiella burnetii." en. In: *PLOS ONE* 8.1 (Jan. 2013), e53440. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0053440. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0053440 (visited on 12/26/2019).

[48] Dimitrios Frangoulidis et al. "Molecular analysis of Coxiella burnetii in Germany reveals evolution of unique clonal clusters." In: *International journal of medical microbiology : IJMM* 304 7 (2014), pp. 868–76.

[49] Y. Gelfand, A. Rodriguez, and G. Benson. "TRDB–The Tandem Repeats Database." In: *Nucleic Acids Research* 35.Database (2007), pp. D80–D87. DOI: 10.1093/nar/gkl1013. URL: https://doi.org/10.1093%2Fnar%2Fgkl1013.

[50] Kris Gevaert, Jozef Van Damme, Marc Goethals, Grégoire R. Thomas, Bart Hoorelbeke, Hans Demol, Lennart Martens, Magda Puype, An Staes, and Joël Vandekerckhove. "Chromatographic Isolation of Methionine-containing Peptides for Gel-free Proteome Analysis." In: *Molecular & Cellular Proteomics* 1.11 (2002), pp. 896–903. DOI: 10.1074/mcp.m200061-mcp200. URL: https://doi.org/10.1074%2Fmcp.m200061-mcp200.

[51] Molly K Gibson, Kevin J Forsberg, and Gautam Dantas. "Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology." In: *The ISME Journal* 9.1 (2014), pp. 207–216. DOI: 10.1038/ismej.2014.106. URL: https://doi.org/10.1038%2Fismej.2014.106.

[52] Achilleas Gikas, Diamantis P. Kofteridis, Andreas Manios, John Pediaditis, and Yiannis Tselentis. "Newer Macrolides as Empiric Treatment for Acute Q Fever Infection." en. In: *Antimicrobial Agents and Chemotherapy* 45.12 (Dec. 2001). Publisher: American Society for Microbiology Journals Section: CLINICAL THERAPEUTICS, pp. 3644–3646. ISSN: 0066-4804, 1098-6596. DOI: 10.1128/AAC.45.12.3644-3646.2001. URL: https://aac.asm.org/content/45/12/3644 (visited on 09/29/2020).

[53] Domingo F. Giménez. "Staining Rickettsiae in Yolk-Sac Cultures." In: *Stain Technology* 39.3 (Jan. 1, 1964), pp. 135–140. ISSN: 0038-9153. DOI: 10.3109/10520296409061219. URL: https://doi.org/10.3109/10520296409061219 (visited on 02/27/2020).

[54] O. Glazunova, V. Roux, O. Freylikman, Z. Sekeyova, G. Fournous, J. Tyczka, N. Tokarevich, E. Kovacova, T.J. Marrie, and D Raoult. *Multi Spacers Typing - Coxiella Burnetii*. URL: https://ifr48.timone.univ-mrs.fr/mst/coxiella_burnetii/strains.html (visited on 03/28/2022).

[55] Olga Glazunova, Véronique Roux, Olga Freylikman, Zuzana Sekeyova, Ghislain Fournous, Judith Tyczka, Nikolai Tokarevich, Elena Kovávcová, T J Marrie, and Didier Raoult. "Coxiella burnetii Genotyping." In: *Emerging Infectious Diseases* 11 (2005), pp. 1211 –1217.

[56] S. Griffiths-Jones. "Rfam: annotating non-coding RNAs in complete genomes." In: *Nucleic Acids Research* 33.Database issue (2004), pp. D121–D124. DOI: 10.1093/nar/gki081. URL: https://doi.org/10.1093%2Fnar%2Fgki081.

[57] Alexey A. Gurevich, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. "QUAST: quality assessment tool for genome assemblies." In: *Bioinformatics* 29 8 (2013), pp. 1072–5.

[58] Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. "QUAST: quality assessment tool for genome assemblies." eng. In: *Bioinformatics (Oxford, England)* 29.8 (Apr. 2013), pp. 1072–1075. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt086.

[59] Andi Gutmans, Stig Bakken, and Derick Rethans. *PHP 5 Power Programming (Bruce Perens' Open Source Series)*. Prentice Hall PTR, 2004.

[60] D. H. Haft. "TIGRFAMs: a protein family resource for the functional identification of proteins." In: *Nucleic Acids Research* 29.1 (2001), pp. 41–43. DOI: 10.1093/nar/29.1.41. URL: https://doi.org/10.1093%2Fnar%2F29.1.41.

[61] Katsuhiko Hayashi, Ryosuke Nakashima, Keisuke Sakurai, Kimie Kitagawa, Seiji Yamasaki, Kunihiko Nishino, and Akihito Yamaguchi. "AcrB-AcrA Fusion Proteins That Act as Multidrug Efflux Transporters." In: *Journal of Bacteriology* 198.2 (2016). Ed. by R. L. Gourse, pp. 332–342. ISSN: 0021-9193. DOI: 10.1128/JB.00587-15. eprint: https://jb.asm.org/content/198/2/332.full.pdf. URL: https://jb.asm.org/content/198/2/332.

[62] R. Heinzen, G. L. Stiegler, L. L. Whiting, S. A. Schmitt, L. P. Mallavia, and M. E. Frazier. "Use of pulsed field gel electrophoresis to differentiate Coxiella burnetii strains." eng. In: *Annals of the New York Academy of Sciences* 590 (1990), pp. 504–513. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.1990.tb42260.x.

[63] Robert A Heinzen, Ted Hackstadt, and James E Samuel. "Developmental biology of Coxiella burnetii." In: *Trends in Microbiology* 7.4 (Apr. 1, 1999). ISSN: 0966-842X. DOI: 10.1016/S0966-842X(99)01475-4. URL: http://www.sciencedirect.com/science/article/pii/S0966842X99014754 (visited on 02/27/2020).

[64] Claudia M. Hemsley, Paul O'Neill, Angela E. Essex-Lopresti, Isobel H. Norville, Tim Atkins, and Richard W. Titball. "Extensive genome analysis of Coxiella burnetii reveals limited evolution within genomic groups." In: *BMC Genomics* 20 (2019).

[65] Laura R. Hendrix, James E. Samuel, and Louis P. Mallavia. "Differentiation of *Coxiella burnetii* isolates by analysis of restriction-endonuclease-digested DNA separated by SDS-PAGE." en. In: *Journal of General Microbiology* 137.2 (Feb. 1991). bibtex*:Hendrix1991, pp. 269–276. ISSN: 1350-0872, 1465-2080. DOI: 10.1099/00221287-137-2-269. URL: http://mic.sgmjournals.org/content/137/2/269 (visited on 12/09/2014).

[66] Christian Jäger Hermann Willems Mirko Ritter and Danielle Thiele. "Plasmid-homologous sequences in the chromosome of plasmidless Coxiella burnetii Scurry Q217." In: *Journal of bacteriology* 179 10 (1997), pp. 3293–7.

[67] D. Higgins and T. J. Marrie. "Seroepidemiology of Q fever among cats in New Brunswick and Prince Edward Island." In: *Annals of the New York Academy of Sciences* 590 (1990), pp. 271–274. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.1990.tb42231.x.

[68] Hamidreza Honarmand. *Q Fever: An Old but Still a Poorly Understood Disease*. 2012. DOI: https://doi.org/10.1155/2012/131932. URL: https://www.hindawi.com/journals/ipid/2012/131932/ (visited on 03/02/2020).

[69] Timothy A. Hoover, Michael H. Vodkin, and John C. Williams. "A Coxiella burnetti repeated DNA element resembling a bacterial insertion sequence." In: *Journal of bacteriology* 174 17 (1992), pp. 5540–8.

[70] Heidie M. Hornstra et al. "Rapid Typing of Coxiella burnetii." In: *PLoS ONE* 6.11 (Nov. 2011). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0026201. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3206805/ (visited on 06/17/2019).

[71] Doug Hyatt, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. "Prodigal: prokaryotic gene recognition and translation initiation site identification." In: *BMC Bioinformatics* 11 (2009), pp. 119 –119.

[72] Vranakis I, De Bock Pj, Papadioti A, Tselentis Y, Gevaert K, Tsiotis G, and Psaroulaki A. *Identification of potentially involved proteins in levofloxacin resistance mechanisms in Coxiella burnetii*. en. ISSN: 1535-3907 Issue: 2 Publisher: J Proteome Res Volume: 10. Feb. 2011. DOI: 10.1021/pr100906v. URL: https://pubmed.ncbi.nlm.nih.gov/21070068/ (visited on 09/16/2020).

[73] F Ishino, W Park, S Tomioka, S Tamaki, I Takase, K Kunugita, H Matsuzawa, S Asoh, T Ohta, and B G Spratt. "Peptidoglycan synthetic activities in membranes of Escherichia coli caused by overproduction of penicillin-binding protein 2 and rodA protein." In: *The Journal of biological chemistry* 261.15 (1986), pp. 7024–31.

[74] Isabel Jado et al. "Molecular method for the characterization of Coxiella burnetii from clinical and environmental samples: variability of genotypes in Spain." In: *BMC Microbiology* 12 (2011), pp. 91 –91.

[75] Baofeng Jia et al. "CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database." eng. In: *Nucleic Acids Research* 45.D1 (2017), pp. D566–D573. ISSN: 1362-4962. DOI: 10.1093/nar/gkw1004.

[76] Ralph E. Johnson. "Frameworks equals components plus patterns." In: *Commun. ACM* 40 (Sept. 1997), pp. 39–42. DOI: 10.1145/262793.262799.

[77] Keith A. Jolley, James E. Bray, and Martin C. J. Maiden. "Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications." eng. In: *Wellcome Open Research* 3 (2018), p. 124. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.14826.1.

[78] C. Jäger, H. Willems, D. Thiele, and G. Baljer. "Molecular characterization of *Coxiella burnetii* isolates." In: *Epidemiology and Infection* 120.2 (Mar. 1998). ISSN: 09502688. DOI: 10.1017/S0950268897008510. URL: http://www.journals.cambridge.org/abstract_S0950268897008510 (visited on 08/22/2018).

[79] L. M. Kampschreur et al. "Chronic Q Fever in the Netherlands 5 Years after the Start of the Q Fever Epidemic: Results from the Dutch Chronic Q Fever Database." In: *Journal of Clinical Microbiology* 52.5 (May 1, 2014), pp. 1637–1643. ISSN: 0095-1137. DOI: 10.1128/JCM.03221-13. URL: http://jcm.asm.org/cgi/doi/10.1128/JCM.03221-13 (visited on 01/23/2020).

[80] Ariane Khaledi et al. "Predicting antimicrobial resistance in Pseudomonas aeruginosa with machine learning-enabled molecular diagnostics." In: *EMBO Molecular Medicine* 12.3 (2020), e10264. DOI: 10.15252/emmm.201910264. eprint: https://www.embopress.org/doi/pdf/10.15252/emmm.201910264. URL: https://www.embopress.org/doi/abs/10.15252/emmm.201910264.

[81] Lukasz Kreft, Alexander Botzki, Frederik Coppens, Klaas Vandepoele, and Michiel Van Bel. "PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization." In: *Bioinformatics* 33.18 (May 2017), pp. 2946–2947. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx324. eprint: https://academic.oup.com/bioinformatics/article-pdf/33/18/2946/25164096/btx324.pdf. URL: https://doi.org/10.1093/bioinformatics/btx324.

[82] Erik Kristiansson, Jerker Fick, Anders Janzon, Roman Grabic, Carolin Rutgersson, Birgitta Weijdegård, Hanna Söderström, and D. G. Joakim Larsson. "Pyrosequencing of Antibiotic-Contaminated River Sediments Reveals High Levels of Resistance and Gene Transfer Elements." In: *PLoS ONE* 6.2 (2011). Ed. by Francisco Rodriguez-Valera, e17038. DOI: 10.1371/journal.pone.0017038. URL: https://doi.org/10.1371%2Fjournal.pone.0017038.

[83] Priyank Kumar, Dong-Mei Zhang, Kurt Degenhardt, and Zhe-Sheng Chen. "Autophagy and transporter-based multi-drug resistance." In: *Cells* 1.3 (2012), pp. 558–75. DOI: 10.3390/cells1030558.

[84] Sanjay Kumar, Shefali Raj Gangoliya, Mustapha Berri, Annie Rodolakis, and Syed Imteyaz Alam. "Whole genome amplification of the obligate intracellular pathogen Coxiella burnetii using multiple displacement amplification." eng. In: *Journal of Microbiological Methods* 95.3 (Dec. 2013), pp. 368–372. ISSN: 1872-8359.

[85] Karin Lagesen, Peter Hallin, Einar Andreas Rødland, Hans-Henrik Stærfeldt, Torbjørn Rognes, and David Wayne Ussery. "RNAmmer: consistent and rapid annotation of ribosomal RNA genes." In: *Nucleic Acids Research* 35 (2007), pp. 3100 –3108.

[86] Steven M. Lakin et al. "MEGARes: an antimicrobial resistance database for high throughput sequencing." eng. In: *Nucleic Acids Research* 45.D1 (2017), pp. D574–D580. ISSN: 1362-4962. DOI: 10.1093/nar/gkw1009.

[87] D. Laslett. "ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences." In: *Nucleic Acids Research* 32.1 (2004), pp. 11–16. DOI: 10.1093/nar/gkh152. URL: https://doi.org/10.1093%2Fnar%2Fgkh152.

[88] S. Lautenschläger, H. Willems, C. Jäger, and G. Baljer. "Sequencing and characterization of the cryptic plasmid QpRS from *Coxiella burnetii*." In: *Plasmid* 44.1 (July 2000), pp. 85–88. ISSN: 0147619X. DOI: 10.1006/plas.2000.1470. URL: http://linkinghub.elsevier.com/retrieve/pii/S0147619X00914707 (visited on 09/11/2018).

[89] Rasko Leinonen, Hideaki Sugawara, and Martin Shumway. "The Sequence Read Archive." In: *Nucleic Acids Research* 39 (2011), pp. D19 – D21.

[90] F. Leonetti, D. Raoult, B. Dussol, P. Brunet, and Y. Berland. "Chronic Q fever in hemodialysis patients." In: *Nephron* 67.2 (1994), pp. 231–233. ISSN: 1660-8151. DOI: 10.1159/000187934.

[91] Jing Li and Takeshi Chusho. "A Web Application Framework for End-User-Initiative Development with a Visual Tool." In: 2012.

[92] Bo Liu and Mihai Pop. "ARDB–Antibiotic Resistance Genes Database." eng. In: *Nucleic Acids Research* 37.Database issue (Jan. 2009), pp. D443–447. ISSN: 1362-4962. DOI: 10.1093/nar/gkn656.

[93] Yang Liu, Paul M. Harrison, Victor Kunin, and Mark Gerstein. "Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes." In: *Genome Biology* 5.9 (2004), R64. ISSN: 1474-760X. DOI: 10.1186/gb-2004-5-9-r64. URL: https://doi.org/10.1186/gb-2004-5-9-r64.

[94] Kyle Loudon. *Developing Large Web Applications: Producing Code That Can Grow and Thrive*. " O'Reilly Media, Inc.", 2010.

[95] Todd M. Lowe and Sean R. Eddy. "tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence." In: *Nucleic acids research* 25 5 (1997), pp. 955–64.

[96] Jacek Lubelski, Wil N. Konings, and Arnold J. M. Driessen. "Distribution and Physiology of ABC-Type Transporters Contributing to Multidrug Resistance in Bacteria." In: *Microbiology and Molecular Biology Reviews : MMBR* 71.3 (Sept. 2007), pp. 463–476. ISSN: 1092-2172. DOI: 10.1128/MMBR.00001-07. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2168643/ (visited on 09/29/2020).

[97] Shengdong Luo et al. "The Coxiella burnetii QpH1 Plasmid Is a Virulence Factor for Colonizing Bone Marrow-Derived Murine Macrophages." In: *Journal of Bacteriology* 203.9 (2021), e00588–20. DOI: 10.1128/JB.00588-20.

[98] Liping Ma, Bing Li, and Tong Zhang. "Abundant rifampin resistance genes and significant correlations of antibiotic resistance genes and plasmids in various environments revealed by metagenomic analysis." In: *Applied Microbiology and Biotechnology* 98.11 (2014), pp. 5195–5204. DOI: 10.1007/s00253-014-5511-3. URL: https://doi.org/10.1007%2Fs00253-014-5511-3.

[99] Kirk J. Mantione, Richard M. Kream, Hana Kuzelova, Radek Ptacek, Jiri Raboch, Joshua Samuel, and G B Stefano. "Comparing Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq." In: *Medical Science Monitor Basic Research* 20 (2014), pp. 138 –141.

[100] T. J. Marrie. ""Historical aspects of Q fever" in Q Fever. Volume I." In: *The Disease* 35 (1990).

[101] Robert F. Massung, Sally J. Cutler, and Dimitrios Frangoulidis. "Molecular typing of Coxiella burnetii (Q fever)." In: *Advances in experimental medicine and biology* 984 (2012), pp. 381–96.

[102] T F McCaul and J C Williams. "Developmental cycle of Coxiella burnetii: structure and morphogenesis of vegetative and sporogenic differentiations." In: *Journal of Bacteriology* 147.3 (Sept. 1981), pp. 1063–1076. ISSN: 0021-9193. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC216147/ (visited on 02/28/2020).

[103] Daniel R Mende et al. "proGenomes2: an improved database for accurate and consistent habitat, taxonomic and functional annotations of prokaryotic genomes." In: *Nucleic Acids Research* 48.D1 (Oct. 2019), pp. D621–D625. ISSN: 0305-1048. DOI: 10.1093/nar/gkz1002. eprint: https://academic.oup.com/nar/article-pdf/48/D1/D621/31697347/gkz1002.pdf. URL: https://doi.org/10.1093/nar/gkz1002.

[104] Dirk Merkel. "Docker: lightweight linux containers for consistent development and deployment." In: *Linux journal* 2014.239 (2014), p. 2.

[105] Bart van Merriënboer, Dzmitry Bahdanau, Vincent Dumoulin, Dmitriy Serdyuk, David Warde-Farley, Jan Chorowski, and Yoshua Bengio. "Blocks and Fuel: Frameworks for deep learning." In: *arXiv:1506.00619 [cs, stat]* (June 2015). arXiv: 1506.00619. URL: http://arxiv.org/abs/1506.00619 (visited on 08/27/2020).

[106] Katja Mertens, Claudia Gerlach, Heinrich Neubauer, and Klaus Henning. "Q fever – An Update." In: *Current Clinical Microbiology Reports* 4.1 (Mar. 1, 2017), pp. 61–70. ISSN: 2196-5471. DOI: 10.1007/s40588-017-0059-5. URL: https://doi.org/10.1007/s40588-017-0059-5 (visited on 03/04/2020).

[107] Ilya Minkin, Anand Patel, Mikhail Kolmogorov, Nikolay Vyahhi, and Son Pham. "Sibelia: A Scalable and Comprehensive Synteny Block Generation Tool for Closely Related Microbial Genomes." In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013, pp. 215–229. DOI:

10.1007/978-3-642-40453-5_17. URL: https://doi.org/10.1007%2F978-3-642-40453-5_17.

[108] Michael F Minnick and Rahul Raghavan. "Genetics of Coxiella burnetii: on the path of specialization." In: *Future microbiology* 6.11 (Nov. 2011). ISSN: 1746-0913. DOI: 10.2217/fmb.11.116. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104754/ (visited on 02/28/2020).

[109] Danesh Moradigaravand, Martin Palm, Anne Farewell, Ville Mustonen, Jonas Warringer, and Leopold Parts. "Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data." In: *PLOS Computational Biology* 14.12 (Dec. 2018), pp. 1–17. DOI: 10.1371/journal.pcbi.1006258. URL: https://doi.org/10.1371/journal.pcbi.1006258.

[110] Irina Morozova, Xiaoyan Qu, Shundi Shi, Gifty Asamani, Joseph E. Greenberg, Howard A. Shuman, and James J. Russo. "Comparative sequence analysis of the icm/dot genes in Legionella." In: *Plasmid* 51.2 (2004), pp. 127–147. ISSN: 0147-619X. DOI: https://doi.org/10.1016/j.plasmid.2003.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0147619X04000022.

[111] Victoriia Murina, Marje Kasari, Vasili Hauryliuk, and Gemma C Atkinson. "Antibiotic resistance ABCF proteins reset the peptidyl transferase centre of the ribosome to counter translational arrest." In: *Nucleic Acids Research* 46.7 (Apr. 2018), pp. 3753–3763. ISSN: 0305-1048. DOI: 10.1093/nar/gky050. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5909423/ (visited on 09/29/2020).

[112] D Musso, M Drancourt, S Osscini, and D Raoult. "Sequence of quinolone resistance-determining region of gyrA gene for clinical isolates and for an in vitro-selected quinolone-resistant strain of Coxiella burnetii." In: *Antimicrobial Agents and Chemotherapy* 40.4 (1996), pp. 870–873. DOI: 10.1128/aac.40.4.870. URL: https://doi.org/10.1128%2Faac.40.4.870.

[113] Sa Van Nguyen and Katsuya Hirai. "Differentiation of Coxiella burnetii isolates by sequence determination and PCR-restriction fragment length polymorphism analysis of isocitrate dehydrogenase gene." en. In: *FEMS Microbiology Letters* 180.2 (1999). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1 6968.1999.tb08803.x, pp. 249–254. ISSN: 1574-6968. DOI: 10.1111/j.1574-6968.1999.tb08803.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1574-6968.1999.tb08803.x (visited on 11/02/2020).

[114] Isobel H. Norville, M. Gill Hartley, Eriel Martínez, Franck Cantet, Matteo Bonazzi, and Timothy P Atkins. "Galleria mellonella as an alternative model of Coxiella burnetii infection." In: *Microbiology* 160 Pt 6 (2014), pp. 1175–81.

[115] Daniel Nouri. *dnouri/nolearn*. original-date: 2012-11-22T21:41:47Z. Aug. 2020. URL: https://github.com/dnouri/nolearn (visited on 08/27/2020).

[116] Anders Omsland, Paul A. Beare, Joshua Hill, Diane C. Cockrell, Dale Howe, Bryan Hansen, James E. Samuel, and Robert A. Heinzen. "Isolation from Animal Tissue and Genetic Transformation of Coxiella burnetii Are Facilitated by an Improved Axenic Growth Medium." In: *Applied and Environmental Microbiology* 77.11 (June 2011), pp. 3720–3725. ISSN: 0099-2240. DOI: 10.1128/AEM.02826-10. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3127619/ (visited on 11/16/2020).

[117] Anders Omsland, Diane C. Cockrell, Dale Howe, Elizabeth R. Fischer, Kimmo Virtaneva, Daniel E. Sturdevant, Stephen F. Porcella, and Robert A. Heinzen. "Host cell-free growth of the Q fever bacterium Coxiella burnetii." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.11 (Mar. 2009), pp. 4430–4434. ISSN: 0027-8424. DOI: 10.1073/pnas.0812074106. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2657411/ (visited on 11/16/2020).

[118] Kristin Patrick. "454 life sciences: illuminating the future of genome sequencing and personalized medicine." In: *The Yale journal of biology and medicine* 80.4 (2007), pp. 191–4.

[119] William R Pearson. "An introduction to sequence similarity ("homology") searching." In: *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* Chapter 3 (2013), Unit3.1. DOI: 10.1002/0471250953.bi0301s42.

[120] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[121] Ivo Pedruzzi et al. "HAMAP in 2015: updates to the protein family classification and annotation system." In: *Nucleic Acids Research* 43.D1 (2014), pp. D1064–D1070. DOI: 10.1093/nar/gku1002. URL: https://doi.org/10.1093%2Fnar%2Fgku1002.

[122] Alvaro Piñero, Jesús Félix Barandika, Ana Luisa García-Pérez, and Ana Hurtado. "Genetic diversity and variation over time of Coxiella burnetii genotypes in dairy cattle and the farm environment." In: *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 31 (2015), pp. 231–5.

[123] Mihai Pop, Adam M. Phillippy, Arthur L. Delcher, and Steven Salzberg. "Comparative genome assembly." In: *Briefings in bioinformatics* 5 3 (2004), pp. 237–48.

[124] Sarah Rebecca Porter, Guy Czaplicki, Jacques Mainil, Raphaël Guattéo, and Claude Saegerman. "Q Fever: current state of knowledge and perspectives of research of a neglected zoonosis." In: *International Journal of Microbiology* 2011 (2011), p. 248418. ISSN: 1687-9198. DOI: 10.1155/2011/248418.

[125] Leighton Pritchard, Rachel H. Glover, Sonia Humphris, John G. Elphinstone, and Ian K. Toth. "Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens." In: *Anal. Methods* 8 (1 2016), pp. 12–24. DOI: 10.1039/C5AY02550H. URL: http://dx.doi.org/10.1039/C5AY02550H.

[126] The Pylon Project. *About the Pylons web framework*. URL: https://pylonsproject.org/about-pylons-framework.html (visited on 08/13/2020).

[127] Kim D. Pruitt, Tatiana Tatusova, and Donna R. Maglott. "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." In: *Nucleic acids research* 33 Database issue (2007), pp. D501–4.

[128] Pylons. *pyramid-cookiecutter-starter*. 2020. URL: https://github.com/Pylons/pyramid-cookiecutter-starter (visited on 11/23/2020).

[129] Michael A. Quail, Miriam W. Smith, Paul G. Coupland, Thomas Dan Otto, Simon R. Harris, Thomas R. Connor, Anna Bertoni, Harold P Swerdlow, and Yong Gu. "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers." In: *BMC Genomics* 13 (2012), pp. 341 –341.

[130] Chantal Quiblier, Annelies S. Zinkernagel, Reto A. Schuepbach, Brigitte Berger-Bächi, and Maria M. Senn. "Contribution of SecDF to Staphylococcus aureus resistance and expression of virulence factors." eng. In: *BMC microbiology* 11 (Apr. 2011). ISSN: 1471-2180. DOI: 10.1186/1471-2180-11-72.

[131] JEAN-MARC ROLAIN, FRÉDÉRIC LAMBERT, and DIDIER RAOULT. "Activity of Telithromycin against Thirteen New Isolates of C. burnetii Including Three Resistant to Doxycycline." In: *Annals of the New York Academy of Sciences* 1063.1 (2005), pp. 252–256. DOI: 10.1196/annals.1355.039. eprint: https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1196/annals.1355.039. URL: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1196/annals.1355.039.

[132] Sanjida H. Rangwala et al. "Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV)." In: *Genome Research* 31.1 (2020), pp. 159–169. DOI: 10.1101/gr.266932.120. URL: https://doi.org/10.1101%2Fgr.266932.120.

[133] D. Raoult, P. Brouqui, B. Marchou, and J.-A. Gastaut. "Acute and Chronic Q Fever in Patients with Cancer." In: *Clinical Infectious Diseases* 14.1 (Jan. 1, 1992), pp. 127–130. ISSN: 1058-4838. DOI: 10.1093/clinids/14.1.127. URL: https://academic.oup.com/cid/article/14/1/127/354158 (visited on 03/03/2020).

[134] Didier Raoult, Pierre Houpikian, Hervé Tissot Dupont, Jean Marc Riss, J. Arditi-Djiane, and Philippe Brouqui. "Treatment of Q Fever Endocarditis: Comparison of 2 Regimens Containing Doxycycline and Ofloxacin or Hydroxychloroquine." en. In: *Archives of Internal Medicine* 159.2 (Jan. 1999). Publisher: American Medical Association, pp. 167–173. ISSN: 0003-9926. DOI: 10.1001/archinte.159.2.167. URL: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/414865 (visited on 09/29/2020).

[135] Didier Raoult and Thomas Marrie. "Q Fever." In: *Clinical Infectious Diseases* 20.3 (Mar. 1, 1995), pp. 489–496. ISSN: 1058-4838. DOI: 10.1093/clinids/20.3.489. URL: https://academic.oup.com/cid/article/20/3/489/321561 (visited on 03/03/2020).

[136] Trygve Reenskaug. "The model-view-controller (mvc) its past and present." In: *University of Oslo Draft* (2003).

[137] G. Rossi. *Web engineering: modelling and implementing web applications*. Vol. 12. Springer-Verlag New York Inc, 2008.

[138] Laetitia Rouli, Jean-Marc Rolain, Adil El Filali, Catherine Robert, and Didier Raoult. "Genome Sequence of Coxiella burnetii 109, a Doxycycline-Resistant Clinical Isolate." In: *Journal of Bacteriology* 194.24 (2012), pp. 6939–6939. ISSN: 0021-9193. DOI: 10.1128/JB.01856-12. eprint: https://jb.asm.org/content/194/24/6939.full.pdf. URL: https://jb.asm.org/content/194/24/6939.

[139] Yasmin Neves Vieira Sabino, Mateus Ferreira Santana, Linda Boniface Oyama, Fernanda Godoy Santos, Ana Júlia Silva Moreira, Sharon Ann Huws, and Hilário Cuquetto Mantovani. "Characterization of antibiotic resistance genes in the species of the rumen microbiota." en. In: *Nature Communications* 10.1 (Nov. 2019). Number: 1 Publisher: Nature Publishing Group, p. 5252. ISSN: 2041-1723. DOI: 10.1038/s41467-019-13118-0. URL: https://www.nature.com/articles/s41467-019-13118-0 (visited on 09/21/2020).

[140] J E Samuel, M E Frazier, and L P Mallavia. "Correlation of plasmid type and disease caused by Coxiella burnetii." In: *Infection and Immunity* 49.3 (Sept. 1985), pp. 775–779. ISSN: 0019-9567. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC261272/ (visited on 03/11/2020).

[141] Savannah E. Sanchez, Eduardo Vallejo-Esquerra, and Anders Omsland. "Use of Axenic Culture Tools to Study Coxiella burnetii." eng. In: *Current Protocols in Microbiology* 50.1 (2018), e52. ISSN: 1934-8533. DOI: 10.1002/cpmc.52.

[142] Kelsi M. Sandoz, David L. Popham, Paul A. Beare, Daniel E. Sturdevant, Bryan Hansen, Vinod Nair, and Robert A. Heinzen. "Transcriptional Profiling of Coxiella burnetii Reveals Extensive Cell Wall Remodeling in the Small Cell Variant Developmental Form." In: *PloS One* 11.2 (2016), e0149957. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0149957.

[143] Franciska M. Schets, Lianne de Heer, and Ana Maria de Roda Husman. "Coxiella burnetii in sewage water at sewage water treatment plants in a Q fever epidemic area." In: *International journal of hygiene and environmental health* 216 6 (2013), pp. 698–702.

[144] Gregory D. Schuler. "Sequence Mapping by Electronic PCR." In: *Genome Research* 7.5 (Apr. 1997), pp. 541–550. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC310656/.

[145] Torsten Seemann. "Prokka: rapid prokaryotic genome annotation." In: *Bioinformatics* 30 14 (2014), pp. 2068–9.

[146] Torsten Seemann and Tim Booth. *Bacterial/Archaeal Ribosomal RNA Predictor*. 2020. URL: https://github.com/tseemann/barrnap (visited on 06/23/2020).

[147] Zuzana Sekeyová, Véronique Roux, and Didier Raoult. "Intraspecies diversity of Coxiella burnetii as revealed by com1 and mucZ sequence comparison." en. In: *FEMS Microbiology Letters* 180.1 (Nov. 1999). Publisher: Oxford Academic, pp. 61–67. ISSN: 0378-1097. DOI: 10.1111/j.1574-6968.1999.tb08778.x. URL: https://academic.oup.com/femsle/article/180/1/61/537112 (visited on 11/02/2020).

[148] Rekha Seshadri et al. "Complete genome sequence of the Q-fever pathogen Coxiella burnetii." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.9 (Apr. 29, 2003). ISSN: 0027-8424. DOI: 10.1073/pnas.0931379100. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC154366/ (visited on 04/04/2019).

[149] Andrew Sharff, Cristina Fanutti, Jiye Shi, Chris Calladine, and Ben Luisi. "The role of the TolC family in protein transport and multidrug efflux." In: *European Journal of Biochemistry* 268.19 (2001), pp. 5011–5026. DOI: 10.1046/j.0014-2956.2001.02442.x.

[150] M D Shriver, L Jin, E Boerwinkle, R Deka, R E Ferrell, and R Chakraborty. "A novel measure of genetic distance for highly polymorphic tandem repeat loci." In: *Molecular Biology and Evolution* 12.5 (Sept. 1995), pp. 914–920. ISSN: 0737-4038. DOI: 10.1093/oxfordjournals.molbev.a040268. eprint: https://academic.oup.com/mbe/article-pdf/12/5/914/11166272/20shri.pdf. URL: https://doi.org/10.1093/oxfordjournals.molbev.a040268.

[151] Hermann Willems Silke Lautenschläger, Christian Jäger, and Georg Baljer. "Sequencing and characterization of the cryptic plasmid QpRS from Coxiella burnetii." In: *Plasmid* 44 1 (2000), pp. 85–8.

[152] Kavindra V. Singh, George M. Weinstock, and Barbara E. Murray. "An Enterococcus faecalis ABC homologue (Lsa) is required for the resistance of this species to clindamycin and quinupristin-dalfopristin." eng. In: *Antimicrobial Agents and Chemotherapy* 46.6 (June 2002), pp. 1845–1850. ISSN: 0066-4804. DOI: 10.1128/aac.46.6.1845-1850.2002.

[153] Jesse Smith. *An Introduction to Python Web Development Using the Pyramid Framework, Part 3*. URL: https://www.informit.com/articles/article.aspx?p=2161678 (visited on 08/19/2020).

[154] Jesse Smith. *An Introduction to Python Web Development Using the Pyramid Framework, Part 4*. URL: https://www.informit.com/articles/article.aspx?p=2161679 (visited on 08/19/2020).

[155] B G Spratt. "Distinct penicillin binding proteins involved in the division, elongation, and shape of Escherichia coli K12." In: *Proceedings of the National Academy of Sciences of the United States of America* 72.8 (1975), pp. 2999–3003.

[156] Ioanna Spyridaki, Anna Psaroulaki, Ana Aransay, Efstathia Scoulica, and Yannis Tselentis. "Diagnosis of quinolone-resistant Coxiella burnetii strains by PCR-RFLP." In: *Journal of Clinical Laboratory Analysis* 14.2 (2000). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291098-2825%282000%2914%3A2%3C59%3A%3AAID-JCLA4%3E3.0.CO%3B2-P, pp. 59–63. ISSN: 1098-2825. DOI: https://doi.org/10.1002/(SICI)1098-2825(2000)14:2<59::AID-JCLA4>3.0.CO;2-P. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291098-2825%282000%2914%3A2%3C59%3A%3AAID-JCLA4%3E3.0.CO%3B2-P (visited on 11/09/2020).

[157] A. Stein and D. Raoult. "Lack of pathotype specific gene in human Coxiella burnetii isolates." In: *Microbial Pathogenesis* 15.3 (Sept. 1993), pp. 177–185. ISSN: 0882-4010. DOI: 10.1006/mpat.1993.1068. URL: http://www.sciencedirect.com/science/article/pii/S0882401083710685 (visited on 03/11/2020).

[158] A. Stein, N. A. Saunders, A. G. Taylor, and D. Raoult. "Phylogenic homogeneity of Coxiella burnetii strains as determinated by 16S ribosomal RNA sequencing." In: *FEMS microbiology letters* 113.3 (Nov. 1993), pp. 339–344. ISSN: 0378-1097. DOI: 10.1111/j.1574-6968.1993.tb06537.x.

[159] Andreas Stein, Danuta Kruszewska, Joanny Gouvernet, and Didier Raoult. "Study of the 16S–23S ribosomal DNA internal spacer of Coxiella burnetii." en. In: *European Journal of Epidemiology* 13.4 (June 1997), pp. 471–475. ISSN: 1573-7284. DOI: 10.1023/A:1007389315808. URL: https://doi.org/10.1023/A:1007389315808 (visited on 11/02/2020).

[160] Mitchell J. Syberg-Olsen, Arkadiy I. Garber, Patrick J. Keeling, John P. McCutcheon, and Filip Husnik. "Pseudofinder: detection of pseudogenes in prokaryotic genomes." In: *bioRxiv* (2021). DOI: 10.1101/2021.10.07.463580. eprint: https://www.biorxiv.org/content/early/2021/10/08/2021.10.07.463580.full.pdf.

[161] D. Thiele and H. Willems. "Is plasmid based differentiation of Coxiella burnetii in 'acute' and 'chronic' isolates still valid?" eng. In: *European Journal of Epidemiology* 10.4 (Aug. 1994), pp. 427–434. ISSN: 0393-2990. DOI: 10.1007/bf01719667.

[162] D. Thiele, H. Willems, M. Haas, and H. Krauss. "Analysis of the entire nucleotide sequence of the cryptic plasmid QpH1 fromCoxiella burnetii." en. In: *European Journal of Epidemiology* 10.4 (Aug. 1994), pp. 413–420. ISSN: 1573-7284. DOI: 10.1007/BF01719665. URL: https://doi.org/10.1007/BF01719665 (visited on 11/04/2020).

[163] Danielle Thiele and Hermann Willems. "Is plasmid based differentiation of Coxiella burnetii in acute and chronic isolates still valid?" In: *European Journal of Epidemiology* 10 (2005), pp. 427–434.

[164] Jeroen J H C Tilburg, Hendrik-Jan I J Roest, Sylvain Buffet, Marrigje H Nabuurs-Franssen, Alphons M Horrevorts, Didier Raoult, and Corné H W Klaassen. "Epidemic genotype of Coxiella burnetii among goats, sheep, and humans in the Netherlands." In: *Emerging infectious diseases* 18.5 (2012), pp. 887–9. DOI: 10.3201/eid1805.111907.

[165] Ho To, Akitoyo Hotta, Guo Quan Zhang, Sa Van Nguyen, Motohiko Ogawa, Tsuyoshi Yamaguchi, Hideto Fukushi, Ken ichi Amano, and Katsuya Hirai. "Antigenic characteristics of polypeptides of Coxiella burnetii isolates." In: *Microbiology and immunology* 42 2 (1998), pp. 81–5.

[166] Todd J Treangen, Brian D Ondov, Sergey Koren, and Adam M Phillippy. "The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes." In: *Genome Biology* 15.11 (2014), p. 524. DOI: 10.1186/preaccept-2573980311437212. URL: https://doi.org/10.1186%2Fpreaccept-2573980311437212.

[167] L. Uden. "Design process for Web applications." In: *IEEE MultiMedia* 9.4 (2002), pp. 47–55.

[168] Danka Valkova and Jan Kazar. "A new plasmid (QpDV) common to Coxiella burnetii isolates associated with acute and chronic Q fever." In: *FEMS microbiology letters* 125 2-3 (1995), pp. 275–80.

[169] Pieter-Jan Van Camp, David B. Haslam, and Aleksey Porollo. "Prediction of Antimicrobial Resistance in Gram-Negative Bacteria From Whole-Genome Sequencing Data." In: *Frontiers in Microbiology* 11 (2020), p. 1013. ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.01013. URL: https://www.frontiersin.org/article/10.3389/fmicb.2020.01013.

[170] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[171] Daniel E. Voth, Paul A. Beare, Dale Howe, Uma M. Sharma, Georgios Samoilis, Diane C. Cockrell, Anders Omsland, and Robert A. Heinzen. "The Coxiella burnetii Cryptic Plasmid Is Enriched in Genes Encoding Type IV Secretion System Substrates." In: *Journal of Bacteriology* 193.7 (2011), pp. 1493–1503. DOI: 10.1128/JB.01359-10.

[172] Iosif Vranakis, Pieter-Jan De Bock, Anastasia Papadioti, Yannis Tselentis, Kris Gevaert, Georgios Tsiotis, and Anna Psaroulaki. "Quantitative Proteome Profiling of C. burnetii under Tetracycline Stress Conditions." In: *PLOS ONE* 7.3 (2012). Publisher: Public Library of Science, pp. 1–8. DOI: 10.1371/journal.pone.0033599. URL: https://doi.org/10.1371/journal.pone.0033599.

[173] Meredith Wadman. "James Watson's genome sequenced at high speed." In: *Nature* 452.7189 (2008), p. 788. DOI: 10.1038/452788b.

[174] Mathias C. Walter, Caroline Öhrman, Kerstin Myrtennäs, Andreas Sjödin, Mona Byström, Pär Larsson, Anna Macellaro, Mats Forsman, and Dimitrios Frangoulidis. "Genome sequence of Coxiella burnetii strain Namibia." In: *Standards in Genomic Sciences* 9 (2014), pp. 22 –22.

[175] J. P. Webster, G. Lloyd, and D. W. Macdonald. "Q fever (Coxiella burnetii) reservoir in wild brown rat (Rattus norvegicus) populations in the UK." In: *Parasitology* 110 (Pt 1) (Jan. 1995), pp. 31–35. ISSN: 0031-1820. DOI: 10.1017/s0031182000081014.

[176] KA Wetterstrand. "DNA Sequencing Costs: Data." In: *Available at: https://www.genome.go genomics/fact-sheets/DNA-Sequencing-Costs-Data* (2022), Accessed [25 July 2022].

[177] Wikipedia contributors. *Web application — Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Web_application&oldid=971114308. [Online; accessed 17-August-2020]. 2020.

[178] Wikipedia contributors. *Web framework — Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Web_framework&oldid=970559664. [Online; accessed 17-August-2020]. 2020.

[179] H. Willems, Cornelie Jäger, and Georg Baljer. "Physical and Genetic Map of the Obligate Intracellular Bacterium Coxiella burnetii." In: *Journal of Bacteriology* 180.15 (Aug. 1998). ISSN: 0021-9193. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107364/ (visited on 03/02/2020).

[180] H. Willems, M. Ritter, C. Jäger, and D. Thiele. "Plasmid-homologous sequences in the chromosome of plasmidless Coxiella burnetii Scurry Q217." In: *Journal of Bacteriology* 179.10 (May 1997). ISSN: 0021-9193. DOI: 10.1128/jb.179.10.3293-3297.1997.

[181] H. Willems, D. Thiele, and H. Krauss. "Plasmid based differentiation and detection of *Coxiella burnetii* in clinical samples." In: *European Journal of Epidemiology* 9.4 (July 1993), pp. 411–418. ISSN: 0392-2990, 1573-7284. DOI: 10.1007/BF00157399. URL: http://link.springer.com/10.1007/BF00157399 (visited on 12/10/2018).

[182] Yinlong Xie et al. "SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads." In: *Bioinformatics* 30 12 (2013), pp. 1660–6.

[183] Yi Xiong, Qiankun Wang, Junchen Yang, Xiaolei Zhu, and Dong-Qing Wei. "PredT4SE-Stack: Prediction of Bacterial Type IV Secreted Effectors From Protein Sequences Using a Stacked Ensemble Method." In: *Frontiers in Microbiology* 9 (2018). DOI: 10.3389/fmicb.2018.02571. URL: https://doi.org/10.3389%2Ffmicb.2018.02571.

[184] Ying Yang, Bing Li, Feng Ju, and Tong Zhang. "Exploring Variation of Antibiotic Resistance Genes in Activated Sludge over a Four-Year Period through a Metagenomic Approach." In: *Environmental Science & Technology* 47.18 (2013), pp. 10197–10205. DOI: 10.1021/es4017365. URL: https://doi.org/10.1021%2Fes4017365.

[185] Bing Yu and Changbin Zhang. "In silico PCR analysis." In: *Methods in molecular biology* 760 (2011), pp. 91–107.

[186] Daniel R. Zerbino and Ewan Birney. "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." In: *Genome research* 18 5 (2008), pp. 821–9.

[187] Guo Quan Zhang, Ho To, Tsuyoshi Yamaguchi, Hideto Fukushi, and Katsuya Hirai. "Differentiation of Coxiella burnetii by Sequence Analysis of the Gene (com1) Encoding a 27-kDa Outer Membrane Protein." en. In: *Microbiology and Immunology* 41.11 (1997), pp. 871–877. ISSN: 1348-0421. DOI: 10.1111/j.1348-0421.1997.tb01943.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1348-0421.1997.tb01943.x (visited on 11/02/2020).

[188] Guoquan Zhang, Ho To, Kasi E. Russell, Laura R. Hendrix, Tsuyoshi Yamaguchi, Hideto Fukushi, Katsuya Hirai, and James E. Samuel. "Identification and characterization of an immunodominant 28-kilodalton Coxiella burnetii outer membrane protein specific to isolates associated with acute disease." eng. In: *Infection and Immunity* 73.3 (Mar. 2005), pp. 1561–1567. ISSN: 0019-9567. DOI: 10.1128/IAI.73.3.1561-1567.2005.

[189] Fernando Zuñiga-Navarrete, Gabriela Flores-Ramirez, Maksym Danchenko, Oldrich Benada, Anton Skriba, and Ludovit Skultety. "Proteomic analysis revealed the survival strategy of *Coxiella burnetii* to doxycycline exposure." In: *Journal of Proteomics* 208 (Sept. 2019), p. 103479. ISSN: 18743919. DOI: 10/gf6cjz. URL: https://linkinghub.elsevier.com/retrieve/pii/S1874391919302519 (visited on 09/27/2019).

[190] Zope foundation and contributors. *Waitress*. URL: https://pypi.org/project/waitress/ (visited on 08/19/2020).

[191] nature.com. *Genomic analysis*. 2020. URL: https://www.nature.com/subjects/genomic-analysis (visited on 06/23/2020).

[192] webtechsurvey.com. "Web Application Frameworks." In: *Available at: https://webtechsurvey.com/technology-type/web-application-frameworks/technologies* (2022), Accessed [25 July 2022].

Part VI

APPENDIX

APPENDIX

A.1 COXBASE TABLES SCHEMA

Listing A.1: Description of all tables in the CoxBase database

```
<?xml version="1.0"?>
<mysqldump xmlns:xsi="http://www.w3.org/2001/XMLSchema-
    instance">
<database name="MLVA">
    <table_structure name="FlankLengthISPCR">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="ms01" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms03" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms20" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms21" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms22" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms23" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms24" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms26" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms27" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms28" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms30" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms31" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms33" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms34" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <key Table="FlankLengthISPCR" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            307" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="FlankLengthISPCR" Engine="InnoDB"
             Version="10" Row_format="Dynamic" Rows="318"
             Avg_row_length="206" Data_length="65536"
```

```
        Max_data_length="0" Index_length="0"
        Data_free="0" Create_time="2022-02-05 05
        :26:21" Update_time="2022-03-07 10:17:06"
        Collation="latin1_swedish_ci" Create_options=
        "" Comment="" />
</table_structure>
<table_structure name="MLVAProfile">
    <field Field="ID" Type="int(11)" Null="NO" Key="
        PRI" Extra="auto_increment" Comment="" />
    <field Field="PanelType" Type="varchar(30)" Null
        ="NO" Key="" Extra="" Comment="" />
    <field Field="ms01" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms03" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms20" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms21" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms22" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms23" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms24" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms26" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms27" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms28" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms30" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms31" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms33" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms34" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <key Table="MLVAProfile" Non_unique="0" Key_name
        ="PRIMARY" Seq_in_index="1" Column_name="ID"
        Collation="A" Cardinality="43" Null=""
        Index_type="BTREE" Comment="" Index_comment="
        " />
    <options Name="MLVAProfile" Engine="InnoDB"
        Version="10" Row_format="Dynamic" Rows="43"
        Avg_row_length="381" Data_length="16384"
        Max_data_length="0" Index_length="0"
        Data_free="0" Auto_increment="44" Create_time
        ="2022-02-05 05:26:21" Collation="
        latin1_swedish_ci" Create_options="" Comment=
        "" />
</table_structure>
<table_structure name="MST">
```

```
        <field Field="ID" Type="int(11)" Null="NO" Key="
            PRI" Extra="auto_increment" Comment="" />
        <field Field="MSTType" Type="varchar(30)" Null="
            YES" Key="" Extra="" Comment="" />
        <key Table="MST" Non_unique="0" Key_name="
            PRIMARY" Seq_in_index="1" Column_name="ID"
            Collation="A" Cardinality="66" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <options Name="MST" Engine="InnoDB" Version="10"
             Row_format="Dynamic" Rows="66"
            Avg_row_length="248" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Auto_increment="67" Create_time
            ="2022-02-05 05:26:21" Collation="
            latin1_swedish_ci" Create_options="" Comment=
            "" />
    </table_structure>
    <table_structure name="NewickTable">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="nwk" Type="text" Null="NO" Key=""
            Extra="" Comment="" />
        <key Table="NewickTable" Non_unique="0" Key_name
            ="PRIMARY" Seq_in_index="1" Column_name="ID"
            Collation="A" Cardinality="121" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <options Name="NewickTable" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="124"
            Avg_row_length="2114" Data_length="262144"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-02-11 11:16:33"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
    </table_structure>
    <table_structure name="Plasmid">
        <field Field="ID" Type="int(11)" Null="NO" Key="
            PRI" Extra="auto_increment" Comment="" />
        <field Field="PlasmidType" Type="varchar(15)"
            Null="YES" Key="" Extra="" Comment="" />
        <key Table="Plasmid" Non_unique="0" Key_name="
            PRIMARY" Seq_in_index="1" Column_name="ID"
            Collation="A" Cardinality="6" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <options Name="Plasmid" Engine="InnoDB" Version=
            "10" Row_format="Dynamic" Rows="6"
            Avg_row_length="2730" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Auto_increment="7" Create_time=
            "2022-02-05 05:26:21" Collation="
            latin1_swedish_ci" Create_options="" Comment=
            "" />
```

```
    </table_structure>
    <table_structure name="ProductLengthISPCR">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="ms01" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
64      <field Field="ms03" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms20" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms21" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms22" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms23" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
69      <field Field="ms24" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms26" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms27" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms28" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms30" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
74      <field Field="ms31" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms33" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms34" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <key Table="ProductLengthISPCR" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            309" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="ProductLengthISPCR" Engine="
            InnoDB" Version="10" Row_format="Dynamic"
            Rows="320" Avg_row_length="204" Data_length="
            65536" Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-03-07 10:17:06"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
79  </table_structure>
    <table_structure name="RepeatNumberISPCR">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="ms01" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms03" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
84      <field Field="ms20" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
```

```
        <field Field="ms21" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms22" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms23" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms24" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
89      <field Field="ms26" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms27" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms28" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms30" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms31" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
94      <field Field="ms33" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms34" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <key Table="RepeatNumberISPCR" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            307" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="RepeatNumberISPCR" Engine="InnoDB
            " Version="10" Row_format="Dynamic" Rows="318
            " Avg_row_length="206" Data_length="65536"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-03-07 10:17:06"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
    </table_structure>
99  <table_structure name="RepeatSizeISPCR">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="ms01" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms03" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms20" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
104     <field Field="ms21" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms22" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms23" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms24" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="ms26" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
```

```
109         <field Field="ms27" Type="int(11)" Null="YES"
                Key="" Extra="" Comment="" />
            <field Field="ms28" Type="int(11)" Null="YES"
                Key="" Extra="" Comment="" />
            <field Field="ms30" Type="int(11)" Null="YES"
                Key="" Extra="" Comment="" />
            <field Field="ms31" Type="int(11)" Null="YES"
                Key="" Extra="" Comment="" />
            <field Field="ms33" Type="int(11)" Null="YES"
                Key="" Extra="" Comment="" />
114         <field Field="ms34" Type="int(11)" Null="YES"
                Key="" Extra="" Comment="" />
            <key Table="RepeatSizeISPCR" Non_unique="0"
                Key_name="PRIMARY" Seq_in_index="1"
                Column_name="ID" Collation="A" Cardinality="
                307" Null="" Index_type="BTREE" Comment=""
                Index_comment="" />
            <options Name="RepeatSizeISPCR" Engine="InnoDB"
                Version="10" Row_format="Dynamic" Rows="318"
                Avg_row_length="206" Data_length="65536"
                Max_data_length="0" Index_length="0"
                Data_free="0" Create_time="2022-02-05 05
                :26:21" Update_time="2022-03-07 10:17:06"
                Collation="latin1_swedish_ci" Create_options=
                "" Comment="" />
        </table_structure>
        <table_structure name="SampleMetadata">
119         <field Field="ID" Type="int(11)" Null="NO" Key="
                PRI" Extra="auto_increment" Comment="" />
            <field Field="SampleStrain" Type="varchar(50)"
                Null="YES" Key="" Extra="" Comment="" />
            <field Field="SampleYear" Type="year(4)" Null="
                YES" Key="" Extra="" Comment="" />
            <field Field="SampleHost" Type="varchar(30)"
                Null="YES" Key="" Extra="" Comment="" />
            <field Field="SampleSource" Type="varchar(100)"
                Null="YES" Key="" Extra="" Comment="" />
124         <field Field="SampleCountry" Type="varchar(30)"
                Null="YES" Key="" Extra="" Comment="" />
            <field Field="CountryProvince" Type="varchar
                (100)" Null="YES" Key="" Extra="" Comment=""
                />
            <field Field="Latitude" Type="decimal(10,8)"
                Null="YES" Key="" Extra="" Comment="" />
            <field Field="Longitude" Type="decimal(11,8)"
                Null="YES" Key="" Extra="" Comment="" />
            <field Field="PubmedID" Type="int(11)" Null="YES
                " Key="" Extra="" Comment="" />
129         <field Field="PlasmidID" Type="int(11)" Null="NO
                " Key="MUL" Extra="" Comment="" />
            <field Field="MSTID" Type="int(11)" Null="NO"
                Key="MUL" Extra="" Comment="" />
            <field Field="TypingID" Type="int(11)" Null="NO"
                 Key="MUL" Extra="" Comment="" />
```

```
        <field Field="MLVAID" Type="int(11)" Null="NO"
            Key="MUL" Extra="" Comment="" />
        <key Table="SampleMetadata" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            117" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <key Table="SampleMetadata" Non_unique="1"
            Key_name="PlasmidID" Seq_in_index="1"
            Column_name="PlasmidID" Collation="A"
            Cardinality="6" Null="" Index_type="BTREE"
            Comment="" Index_comment="" />
        <key Table="SampleMetadata" Non_unique="1"
            Key_name="MSTID" Seq_in_index="1" Column_name
            ="MSTID" Collation="A" Cardinality="1" Null="
            " Index_type="BTREE" Comment="" Index_comment
            ="" />
        <key Table="SampleMetadata" Non_unique="1"
            Key_name="TypingID" Seq_in_index="1"
            Column_name="TypingID" Collation="A"
            Cardinality="43" Null="" Index_type="BTREE"
            Comment="" Index_comment="" />
        <key Table="SampleMetadata" Non_unique="1"
            Key_name="MLVAID" Seq_in_index="1"
            Column_name="MLVAID" Collation="A"
            Cardinality="43" Null="" Index_type="BTREE"
            Comment="" Index_comment="" />
        <options Name="SampleMetadata" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="117"
            Avg_row_length="140" Data_length="16384"
            Max_data_length="0" Index_length="65536"
            Data_free="0" Auto_increment="118"
            Create_time="2022-02-05 05:26:21" Collation="
            latin1_swedish_ci" Create_options="" Comment=
            "" />
    </table_structure>
    <table_structure name="SubmissionTable">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="SubmissionDate" Type="datetime"
            Null="YES" Key="" Default="CURRENT_TIMESTAMP"
             Extra="" Comment="" />
        <field Field="User" Type="varchar(100)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="AnalysisType" Type="varchar(25)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="IPaddress" Type="varbinary(16)"
            Null="YES" Key="" Extra="" Comment="" />
        <key Table="SubmissionTable" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            632" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="SubmissionTable" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="652"
```

```
            Avg_row_length="150" Data_length="98304"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-03-07 10:17:06"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
    </table_structure>
    <table_structure name="TypingMetadata">
        <field Field="ID" Type="int(11)" Null="NO" Key="
            PRI" Extra="auto_increment" Comment="" />
        <field Field="ClusterType" Type="varchar(5)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="Genotype" Type="varchar(10)" Null=
            "YES" Key="" Extra="" Comment="" />
        <key Table="TypingMetadata" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            43" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="TypingMetadata" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="43"
            Avg_row_length="381" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Auto_increment="44" Create_time
            ="2022-02-05 05:26:21" Collation="
            latin1_swedish_ci" Create_options="" Comment=
            "" />
    </table_structure>
    <table_structure name="UserTable">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="email" Type="varchar(255)" Null="
            NO" Key="" Extra="" Comment="" />
        <field Field="username" Type="varchar(100)" Null
            ="NO" Key="" Extra="" Comment="" />
        <field Field="password" Type="varchar(60)" Null=
            "NO" Key="" Extra="" Comment="" />
        <field Field="group" Type="varchar(100)" Null="
            YES" Key="" Extra="" Comment="" />
        <key Table="UserTable" Non_unique="0" Key_name="
            PRIMARY" Seq_in_index="1" Column_name="ID"
            Collation="A" Cardinality="3" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <options Name="UserTable" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="3"
            Avg_row_length="5461" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Collation="latin1_swedish_ci"
            Create_options="" Comment="" />
    </table_structure>
    <table_structure name="adaAProfile">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
```

```
        <field Field="adaAStatus" Type="tinyint(1)" Null
            ="YES" Key="" Extra="" Comment="" />
        <field Field="genotype" Type="varchar(30)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="plasmidType" Type="varchar(20)"
            Null="YES" Key="" Extra="" Comment="" />
        <key Table="adaAProfile" Non_unique="0" Key_name
            ="PRIMARY" Seq_in_index="1" Column_name="ID"
            Collation="A" Cardinality="118" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <options Name="adaAProfile" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="124"
            Avg_row_length="132" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-02-15 09:49:30"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
    </table_structure>
    <table_structure name="is1111Profile">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="IS1111_1" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_2" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_3" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_4" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_5" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_6" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_7" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_8" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_9" Type="tinyint(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_10" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_11" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_12" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_13" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_14" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_15" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_16" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
```

```
<field Field="IS1111_17" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_18" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_19" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_20" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_21" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_22" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_23" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_24" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_25" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_26" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_27" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_28" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_29" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_30" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_31" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_32" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_33" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_34" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_35" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_36" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_37" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_38" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_39" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_40" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_41" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_42" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="IS1111_43" Type="tinyint(1)" Null=
    "YES" Key="" Extra="" Comment="" />
```

```
        <field Field="IS1111_44" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_45" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_46" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_47" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_48" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_49" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_50" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_51" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_53" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_54" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_55" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_56" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_57" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_58" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_59" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_60" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_61" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="IS1111_84" Type="tinyint(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <key Table="is1111Profile" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            138" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="is1111Profile" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="147"
            Avg_row_length="334" Data_length="49152"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-02-20 09:20:25"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
    </table_structure>
    <table_structure name="isolateSubmission">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="submissionType" Type="varchar(10)"
             Null="NO" Key="" Extra="" Comment="" />
```

```
        <field Field="SubmissionDate" Type="datetime"
            Null="YES" Key="" Default="CURRENT_TIMESTAMP"
             Extra="" Comment="" />
        <field Field="isolateName" Type="varchar(100)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="isolateSource" Type="varchar(100)"
             Null="YES" Key="" Extra="" Comment="" />
        <field Field="isolateHost" Type="varchar(100)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="placeOfIsolation" Type="varchar
            (100)" Null="YES" Key="" Extra="" Comment=""
            />
        <field Field="yearOfIsolation" Type="year(4)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="country" Type="varchar(100)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="latitude" Type="decimal(10,8)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="longitude" Type="decimal(11,8)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="pubmedID" Type="varchar(50)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="genomeAccession" Type="varchar(50)
            " Null="YES" Key="" Extra="" Comment="" />
        <field Field="email" Type="varchar(255)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="submitterName" Type="varchar(255)"
             Null="YES" Key="" Extra="" Comment="" />
        <key Table="isolateSubmission" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            37" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="isolateSubmission" Engine="InnoDB
            " Version="10" Row_format="Dynamic" Rows="37"
             Avg_row_length="442" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Collation="latin1_swedish_ci"
            Create_options="" Comment="" />
    </table_structure>
    <table_structure name="isolate_refs">
        <field Field="isolate_id" Type="smallint(5)
            unsigned" Null="NO" Key="PRI" Extra=""
            Comment="" />
        <field Field="pmid" Type="int(11)" Null="NO" Key
            ="PRI" Extra="" Comment="PubMed ID" />
        <field Field="alias" Type="varchar(45)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="mlvaCRC32" Type="int(10) unsigned"
             Null="YES" Key="" Extra="" Comment="" />
        <field Field="incomplete" Type="tinyint(1)
            unsigned" Null="NO" Key="" Default="0" Extra=
            "" Comment="" />
```

```
264         <key Table="isolate_refs" Non_unique="0"
                Key_name="PRIMARY" Seq_in_index="1"
                Column_name="isolate_id" Collation="A"
                Cardinality="455" Null="" Index_type="BTREE"
                Comment="" Index_comment="" />
            <key Table="isolate_refs" Non_unique="0"
                Key_name="PRIMARY" Seq_in_index="2"
                Column_name="pmid" Collation="A" Cardinality=
                "659" Null="" Index_type="BTREE" Comment=""
                Index_comment="" />
            <key Table="isolate_refs" Non_unique="1"
                Key_name="IDX_refs" Seq_in_index="1"
                Column_name="pmid" Collation="A" Cardinality=
                "31" Null="" Index_type="BTREE" Comment=""
                Index_comment="" />
            <key Table="isolate_refs" Non_unique="1"
                Key_name="IDX_refs" Seq_in_index="2"
                Column_name="isolate_id" Collation="A"
                Cardinality="659" Null="" Index_type="BTREE"
                Comment="" Index_comment="" />
            <options Name="isolate_refs" Engine="InnoDB"
                Version="10" Row_format="Dynamic" Rows="659"
                Avg_row_length="99" Data_length="65536"
                Max_data_length="0" Index_length="16384"
                Data_free="0" Create_time="2022-02-05 05
                :26:21" Collation="latin1_general_ci"
                Create_options="row_format=DYNAMIC" Comment="
                " />
269     </table_structure>
        <table_structure name="isolates">
            <field Field="isolateid" Type="smallint(5)
                unsigned" Null="NO" Key="PRI" Extra="
                auto_increment" Comment="" />
            <field Field="name" Type="varchar(45)" Null="NO"
                 Key="UNI" Extra="" Comment="" />
            <field Field="key" Type="varchar(45)" Null="YES"
                 Key="UNI" Extra="" Comment="" />
274         <field Field="isolateNo" Type="smallint(5)
                unsigned" Null="YES" Key="" Extra="" Comment=
                "" />
            <field Field="imbNo" Type="varchar(14)" Null="
                YES" Key="" Extra="" Comment="" />
            <field Field="giNo" Type="smallint(5) unsigned"
                Null="YES" Key="" Extra="" Comment="" />
            <field Field="dateOfIsolation" Type="varchar(10)
                " Null="YES" Key="" Extra="" Comment="" />
            <field Field="yearOfIsolation" Type="smallint(5)
                 unsigned" Null="YES" Key="" Extra="" Comment
                ="" />
279         <field Field="host" Type="enum('cattle','deer','
                goat','human','mouse','rodent','sheep','tick
                ','environment','other')" Null="YES" Key=""
                Extra="" Comment="" />
            <field Field="subspecies" Type="varchar(45)"
                Null="YES" Key="" Extra="" Comment="" />
```

```
<field Field="sample" Type="varchar(100)" Null="
    YES" Key="" Extra="" Comment="" />
<field Field="tissue" Type="varchar(45)" Null="
    YES" Key="" Extra="" Comment="" />
<field Field="diseasePattern" Type="varchar(45)"
     Null="YES" Key="" Extra="" Comment="" />
<field Field="phase" Type="enum('I','II','I/II')
    " Null="YES" Key="" Extra="" Comment="" />
<field Field="zipCode" Type="mediumint(8)
    unsigned" Null="YES" Key="" Extra="" Comment=
    "" />
<field Field="geographicOrigin" Type="varchar
    (45)" Null="YES" Key="" Extra="" Comment="" /
    >
<field Field="province" Type="char(3)" Null="YES
    " Key="" Extra="" Comment="" />
<field Field="country" Type="varchar(45)" Null="
    YES" Key="" Extra="" Comment="" />
<field Field="institution_id" Type="tinyint(3)
    unsigned" Null="YES" Key="MUL" Extra=""
    Comment="" />
<field Field="plasmidType" Type="varchar(20)"
    Null="YES" Key="" Extra="" Comment="" />
<field Field="adaGene" Type="enum('neg.','pos
    .','pos.*','pos.S','Q154-del','Q212-del','pos
    .?')" Null="YES" Key="" Extra="" Comment="" /
    >
<field Field="mlvaGenotype" Type="char(3)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="comment" Type="varchar(255)" Null=
    "YES" Key="" Extra="" Comment="" />
<field Field="restrictionGroup" Type="varchar
    (10)" Null="YES" Key="" Extra="" Comment="" /
    >
<field Field="mstGroup" Type="tinyint(3)
    unsigned" Null="YES" Key="MUL" Extra=""
    Comment="" />
<field Field="isGenotype" Type="int(10) unsigned
    " Null="YES" Key="" Extra="" Comment="" />
<field Field="isGroup" Type="char(1)" Null="YES"
     Key="" Extra="" Comment="" />
<field Field="mlvaCrc32" Type="int(10) unsigned"
     Null="YES" Key="" Extra="" Comment="" />
<field Field="genotype" Type="int(10) unsigned"
    Null="YES" Key="MUL" Extra="" Comment="" />
<field Field="ISO3166_1" Type="smallint(5)
    unsigned" Null="YES" Key="MUL" Extra=""
    Comment="" />
<field Field="isRef" Type="tinyint(1) unsigned"
    Null="NO" Key="" Default="0" Extra="" Comment
    "" />
<field Field="chronic" Type="tinyint(1) unsigned
    " Null="NO" Key="" Default="0" Extra=""
    Comment="" />
```

```
<field Field="exclude" Type="tinyint(1) unsigned
    " Null="NO" Key="" Default="0" Extra=""
    Comment="" />
<field Field="invalid" Type="tinyint(1) unsigned
    " Null="NO" Key="" Default="0" Extra=""
    Comment="" />
<field Field="neighbour" Type="tinyint(1)
    unsigned" Null="NO" Key="" Default="0" Extra=
    "" Comment="" />
<field Field="snp16" Type="smallint(5) unsigned"
     Null="YES" Key="MUL" Extra="" Comment="" />
<field Field="snp23" Type="smallint(5) unsigned"
     Null="YES" Key="MUL" Extra="" Comment="" />
<field Field="dbname" Type="varchar(45)" Null="
    YES" Key="" Extra="" Comment="" />
<key Table="isolates" Non_unique="0" Key_name="
    PRIMARY" Seq_in_index="1" Column_name="
    isolateid" Collation="A" Cardinality="491"
    Null="" Index_type="BTREE" Comment=""
    Index_comment="" />
<key Table="isolates" Non_unique="0" Key_name="
    IDX_name" Seq_in_index="1" Column_name="name"
     Collation="A" Cardinality="491" Null=""
    Index_type="BTREE" Comment="" Index_comment="
    " />
<key Table="isolates" Non_unique="0" Key_name="
    IDX_key" Seq_in_index="1" Column_name="key"
    Collation="A" Cardinality="164" Null="YES"
    Index_type="BTREE" Comment="" Index_comment="
    " />
<key Table="isolates" Non_unique="1" Key_name="
    IDX_mst" Seq_in_index="1" Column_name="
    mstGroup" Collation="A" Cardinality="21" Null
    ="YES" Index_type="BTREE" Comment=""
    Index_comment="" />
<key Table="isolates" Non_unique="1" Key_name="
    IDX_genotype" Seq_in_index="1" Column_name="
    genotype" Collation="A" Cardinality="79" Null
    ="YES" Index_type="BTREE" Comment=""
    Index_comment="" />
<key Table="isolates" Non_unique="1" Key_name="
    IDX_iso" Seq_in_index="1" Column_name="
    ISO3166_1" Collation="A" Cardinality="37"
    Null="YES" Index_type="BTREE" Comment=""
    Index_comment="" />
<key Table="isolates" Non_unique="1" Key_name="
    FK_isolates_snp23" Seq_in_index="1"
    Column_name="snp23" Collation="A" Cardinality
    ="29" Null="YES" Index_type="BTREE" Comment="
    " Index_comment="" />
<key Table="isolates" Non_unique="1" Key_name="
    FK_isolates_snp16" Seq_in_index="1"
    Column_name="snp16" Collation="A" Cardinality
    ="17" Null="YES" Index_type="BTREE" Comment="
    " Index_comment="" />
```

```
        <key Table="isolates" Non_unique="1" Key_name="
            FK_isolates_institutions" Seq_in_index="1"
            Column_name="institution_id" Collation="A"
            Cardinality="11" Null="YES" Index_type="BTREE
            " Comment="" Index_comment="" />
        <options Name="isolates" Engine="InnoDB" Version
            ="10" Row_format="Dynamic" Rows="491"
            Avg_row_length="166" Data_length="81920"
            Max_data_length="0" Index_length="131072"
            Data_free="0" Auto_increment="571"
            Create_time="2022-02-05 05:26:21" Collation="
            utf8_general_ci" Create_options="row_format=
            DYNAMIC" Comment="" />
    </table_structure>
    <table_structure name="mlvaALL">
        <field Field="isolate_id" Type="smallint(5)
            unsigned" Null="NO" Key="PRI" Extra=""
            Comment="" />
        <field Field="ms01" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms03" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms20" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms21" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms22" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms23" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms24" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms26" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms27" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms28" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms30" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms31" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms33" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms34" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="crc32" Type="int(10) unsigned"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="cnt" Type="tinyint(3) unsigned"
            Null="YES" Key="MUL" Extra="" Comment="" />
        <field Field="ngt" Type="varchar(3)" Null="YES"
            Key="MUL" Extra="" Comment="" />
        <field Field="ngtS" Type="char(3)" Null="YES"
            Key="" Extra="" Comment="" />
```

```
        <field Field="clustNo" Type="tinyint(3) unsigned
            " Null="YES" Key="" Extra="" Comment="" />
        <key Table="mlvaALL" Non_unique="0" Key_name="
            PRIMARY" Seq_in_index="1" Column_name="
            isolate_id" Collation="A" Cardinality="194"
            Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <key Table="mlvaALL" Non_unique="1" Key_name="
            IDX_cnt" Seq_in_index="1" Column_name="cnt"
            Collation="A" Cardinality="3" Null="YES"
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <key Table="mlvaALL" Non_unique="1" Key_name="
            IDX_ngt" Seq_in_index="1" Column_name="ngt"
            Collation="A" Cardinality="72" Null="YES"
            Index_type="BTREE" Comment="" Index_comment="
            " />
344     <options Name="mlvaALL" Engine="InnoDB" Version=
            "10" Row_format="Dynamic" Rows="194"
            Avg_row_length="253" Data_length="49152"
            Max_data_length="0" Index_length="32768"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Collation="latin1_general_ci"
            Create_options="row_format=DYNAMIC" Comment="
            " />
    </table_structure>
    <table_structure name="mlvaSubmission">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="ms01" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
349     <field Field="ms03" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms20" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms21" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms22" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms23" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
354     <field Field="ms24" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms26" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms27" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms28" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms30" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
359     <field Field="ms31" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms33" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
```

```
        <field Field="ms34" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="isolateID" Type="varchar(36)" Null
            ="NO" Key="MUL" Extra="" Comment="" />
        <key Table="mlvaSubmission" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            19" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
364     <key Table="mlvaSubmission" Non_unique="1"
            Key_name="isolateID" Seq_in_index="1"
            Column_name="isolateID" Collation="A"
            Cardinality="19" Null="" Index_type="BTREE"
            Comment="" Index_comment="" />
        <options Name="mlvaSubmission" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="19"
            Avg_row_length="862" Data_length="16384"
            Max_data_length="0" Index_length="16384"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Collation="latin1_swedish_ci"
            Create_options="" Comment="" />
    </table_structure>
    <table_structure name="mlva_normalized">
        <field Field="ID" Type="smallint(6)" Null="NO"
            Key="PRI" Extra="auto_increment" Comment="" /
            >
369     <field Field="ms01" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms03" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms20" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms21" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms22" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
374     <field Field="ms23" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms24" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms26" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms27" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms28" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
379     <field Field="ms30" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms31" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms33" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
        <field Field="ms34" Type="float" Null="YES" Key=
            "" Extra="" Comment="" />
```

```
        <field Field="ngt" Type="varchar(3)" Null="YES"
            Key="" Extra="" Comment="" />
384     <key Table="mlva_normalized" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            71" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="mlva_normalized" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="71"
            Avg_row_length="230" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Auto_increment="72" Create_time
            ="2022-02-05 05:26:21" Collation="
            latin1_swedish_ci" Create_options="" Comment=
            "" />
    </table_structure>
    <table_structure name="mstSpacerResultTable">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
389     <field Field="cox18" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox2" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox20" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox22" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox37" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
394     <field Field="cox5" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox51" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox56" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox57" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox61" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
399     <key Table="mstSpacerResultTable" Non_unique="0"
             Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            198" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="mstSpacerResultTable" Engine="
            InnoDB" Version="10" Row_format="Dynamic"
            Rows="207" Avg_row_length="237" Data_length="
            49152" Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-02-15 09:49:30"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
    </table_structure>
    <table_structure name="mstSubmission">
```

```
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="cox18" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox2" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox20" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox22" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox37" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox5" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox51" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox56" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox57" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="cox61" Type="int(11)" Null="YES"
            Key="" Extra="" Comment="" />
        <field Field="isolateID" Type="varchar(36)" Null
            ="NO" Key="MUL" Extra="" Comment="" />
        <key Table="mstSubmission" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            15" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <key Table="mstSubmission" Non_unique="1"
            Key_name="isolateID" Seq_in_index="1"
            Column_name="isolateID" Collation="A"
            Cardinality="15" Null="" Index_type="BTREE"
            Comment="" Index_comment="" />
        <options Name="mstSubmission" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="15"
            Avg_row_length="1092" Data_length="16384"
            Max_data_length="0" Index_length="16384"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Collation="latin1_swedish_ci"
            Create_options="" Comment="" />
    </table_structure>
    <table_structure name="mstgroups">
        <field Field="groupid" Type="tinyint(3) unsigned
            " Null="NO" Key="PRI" Extra="auto_increment"
            Comment="" />
        <field Field="COX2" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX5" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX18" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX20" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
```

```
        <field Field="COX22" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX37" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX51" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX56" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX57" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <field Field="COX61" Type="tinyint(3) unsigned"
            Null="NO" Key="" Extra="" Comment="" />
        <key Table="mstgroups" Non_unique="0" Key_name="
            PRIMARY" Seq_in_index="1" Column_name="
            groupid" Collation="A" Cardinality="40" Null=
            "" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="mstgroups" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="40"
            Avg_row_length="409" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Auto_increment="41" Create_time
            ="2022-02-05 05:26:21" Collation="
            latin1_general_ci" Create_options="" Comment=
            "" />
    </table_structure>

    <table_structure name="primer">
        <field Field="id" Type="smallint(5) unsigned"
            Null="NO" Key="PRI" Extra="auto_increment"
            Comment="" />
        <field Field="name" Type="varchar(45)" Null="NO"
             Key="UNI" Extra="" Comment="" />
        <field Field="marker_id" Type="smallint(5)
            unsigned" Null="YES" Key="MUL" Extra=""
            Comment="" />
        <field Field="forward" Type="varchar(45)" Null="
            NO" Key="" Extra="" Comment="" />
        <field Field="reverse" Type="varchar(45)" Null="
            NO" Key="" Extra="" Comment="" />
        <field Field="probe" Type="varchar(255)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="probe2" Type="varchar(60)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="range_min" Type="smallint(5)
            unsigned" Null="YES" Key="" Extra="" Comment=
            "" />
        <field Field="range_max" Type="smallint(5)
            unsigned" Null="YES" Key="" Extra="" Comment=
            "" />
        <field Field="forward_code" Type="varchar(45)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="reverse_code" Type="varchar(45)"
            Null="YES" Key="" Extra="" Comment="" />
```

```
        <field Field="probe_code" Type="varchar(45)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="probe2_code" Type="varchar(45)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="description" Type="varchar(255)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="pmid" Type="int(11)" Null="YES"
            Key="MUL" Extra="" Comment="PubMed ID" />
        <field Field="forward_lot_number" Type="int(10)
            unsigned" Null="YES" Key="" Extra="" Comment=
            "" />
        <field Field="reverse_lot_number" Type="int(10)
            unsigned" Null="YES" Key="" Extra="" Comment=
            "" />
        <key Table="primer" Non_unique="0" Key_name="
            PRIMARY" Seq_in_index="1" Column_name="id"
            Collation="A" Cardinality="171" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <key Table="primer" Non_unique="0" Key_name="
            IDX_name" Seq_in_index="1" Column_name="name"
             Collation="A" Cardinality="171" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <key Table="primer" Non_unique="1" Key_name="
            FK_primer_pubmed" Seq_in_index="1"
            Column_name="pmid" Collation="A" Cardinality=
            "18" Null="YES" Index_type="BTREE" Comment=""
             Index_comment="" />
        <key Table="primer" Non_unique="1" Key_name="
            FK_primer_marker" Seq_in_index="1"
            Column_name="marker_id" Collation="A"
            Cardinality="91" Null="YES" Index_type="BTREE
            " Comment="" Index_comment="" />
        <options Name="primer" Engine="InnoDB" Version="
            10" Row_format="Dynamic" Rows="171"
            Avg_row_length="287" Data_length="49152"
            Max_data_length="0" Index_length="49152"
            Data_free="0" Auto_increment="285"
            Create_time="2022-02-05 05:26:21" Collation="
            latin1_general_ci" Create_options="row_format
            =DYNAMIC" Comment="" />
    </table_structure>
    <table_structure name="snpHornstra">
        <field Field="ID" Type="varchar(36)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <field Field="Cox5bp81" Type="varchar(1)" Null="
            YES" Key="" Extra="" Comment="" />
        <field Field="Cox22bp91" Type="varchar(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="Cox18bp376" Type="varchar(1)" Null
            ="YES" Key="" Extra="" Comment="" />
        <field Field="Cox51bp356" Type="varchar(1)" Null
            ="YES" Key="" Extra="" Comment="" />
```

```
        <field Field="Cox18bp34" Type="varchar(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="Cox5bp109" Type="varchar(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="Cox22bp118" Type="varchar(1)" Null
            ="YES" Key="" Extra="" Comment="" />
        <field Field="Cox51bp492" Type="varchar(1)" Null
            ="YES" Key="" Extra="" Comment="" />
        <field Field="Cox57bp327" Type="varchar(1)" Null
            ="YES" Key="" Extra="" Comment="" />
        <field Field="Cox56bp10" Type="varchar(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="Cox51bp67" Type="varchar(1)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="Cox20bp155" Type="varchar(1)" Null
            ="YES" Key="" Extra="" Comment="" />
        <key Table="snpHornstra" Non_unique="0" Key_name
            ="PRIMARY" Seq_in_index="1" Column_name="ID"
            Collation="A" Cardinality="53" Null=""
            Index_type="BTREE" Comment="" Index_comment="
            " />
        <options Name="snpHornstra" Engine="InnoDB"
            Version="10" Row_format="Dynamic" Rows="58"
            Avg_row_length="282" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Update_time="2022-02-16 10:44:35"
            Collation="latin1_swedish_ci" Create_options=
            "" Comment="" />
    </table_structure>
    </table_structure>
    <table_structure name="tilburg_geolocation">
        <field Field="ID" Type="smallint(5) unsigned"
            Null="NO" Key="PRI" Extra="auto_increment"
            Comment="" />
        <field Field="name" Type="varchar(50)" Null="YES
            " Key="" Extra="" Comment="" />
        <field Field="mlvaGenotype" Type="smallint(5)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="country" Type="varchar(100)" Null=
            "YES" Key="" Extra="" Comment="" />
        <field Field="longitude" Type="decimal(11,8)"
            Null="YES" Key="" Extra="" Comment="" />
        <field Field="latitude" Type="decimal(10,8)"
            Null="YES" Key="" Extra="" Comment="" />
        <key Table="tilburg_geolocation" Non_unique="0"
            Key_name="PRIMARY" Seq_in_index="1"
            Column_name="ID" Collation="A" Cardinality="
            233" Null="" Index_type="BTREE" Comment=""
            Index_comment="" />
        <options Name="tilburg_geolocation" Engine="
            InnoDB" Version="10" Row_format="Dynamic"
            Rows="233" Avg_row_length="70" Data_length="
            16384" Max_data_length="0" Index_length="0"
            Data_free="0" Auto_increment="237"
```

```
        Create_time="2022-02-05 05:26:21" Collation="
        latin1_swedish_ci" Create_options="" Comment=
        "" />
</table_structure>
<table_structure name="tilburg_isolates">
    <field Field="ID" Type="smallint(5)" Null="NO"
        Key="PRI" Extra="" Comment="" />
    <field Field="name" Type="varchar(100)" Null="
        YES" Key="" Extra="" Comment="" />
    <field Field="host" Type="varchar(100)" Null="
        YES" Key="" Extra="" Comment="" />
    <field Field="source" Type="varchar(100)" Null="
        YES" Key="" Extra="" Comment="" />
    <field Field="country" Type="varchar(100)" Null=
        "YES" Key="" Extra="" Comment="" />
    <field Field="pmid" Type="varchar(100)" Null="
        YES" Key="" Extra="" Comment="" />
    <field Field="profile_ID" Type="smallint(5)"
        Null="YES" Key="" Extra="" Comment="" />
    <key Table="tilburg_isolates" Non_unique="0"
        Key_name="PRIMARY" Seq_in_index="1"
        Column_name="ID" Collation="A" Cardinality="
        236" Null="" Index_type="BTREE" Comment=""
        Index_comment="" />
    <options Name="tilburg_isolates" Engine="InnoDB"
         Version="10" Row_format="Dynamic" Rows="236"
         Avg_row_length="69" Data_length="16384"
        Max_data_length="0" Index_length="0"
        Data_free="0" Create_time="2022-02-05 05
        :26:21" Collation="latin1_swedish_ci"
        Create_options="" Comment="" />
</table_structure>
<table_structure name="tilburg_profile">
    <field Field="ID" Type="smallint(6)" Null="NO"
        Key="PRI" Extra="auto_increment" Comment="" /
        >
    <field Field="ms23" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms24" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms27" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms28" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms33" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <field Field="ms34" Type="float" Null="YES" Key=
        "" Extra="" Comment="" />
    <key Table="tilburg_profile" Non_unique="0"
        Key_name="PRIMARY" Seq_in_index="1"
        Column_name="ID" Collation="A" Cardinality="
        90" Null="" Index_type="BTREE" Comment=""
        Index_comment="" />
    <options Name="tilburg_profile" Engine="InnoDB"
        Version="10" Row_format="Dynamic" Rows="90"
```

```
            Avg_row_length="182" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Auto_increment="91" Create_time
            ="2022-02-05 05:26:21" Collation="
            latin1_swedish_ci" Create_options="" Comment=
            "" />
    </table_structure>
    <table_structure name="tmp_reg">
        <field Field="pattern" Type="char(7)" Null="NO"
            Key="PRI" Extra="" Comment="" />
        <key Table="tmp_reg" Non_unique="0" Key_name="
            PRIMARY" Seq_in_index="1" Column_name="
            pattern" Collation="A" Cardinality="7" Null="
            " Index_type="BTREE" Comment="" Index_comment
            ="" />
        <options Name="tmp_reg" Engine="InnoDB" Version=
            "10" Row_format="Dynamic" Rows="7"
            Avg_row_length="2340" Data_length="16384"
            Max_data_length="0" Index_length="0"
            Data_free="0" Create_time="2022-02-05 05
            :26:21" Collation="latin1_general_ci"
            Create_options="" Comment="" />
    </table_structure>
</database>
</mysqldump>
```

A.2 *C. burnetii* GENOMES ANALYZED WITH COXBASE

Table A.1 - Table A.2 contains the description of all *C. burnetii* genomes that were analyzed with the CoxBase platform.

ISOLATE NAME	LEVEL	REFSEQ
Coxiella burnetii RSA 493	Complete chromosome	NC_002971.4/NC_004704.2
Coxiella burnetii Dugway 5J108-111	Complete chromosome	NC_009727.1/NC_009726.1
3262	Complete chromosome	NZ_CP013667.1/NZ_CP013668.1
2014-PE-15890	Complete chromosome	NZ_CP032542.1
CbuK_Q154	Complete chromosome	NC_011528.1/NC_011526.1
MSU Goat Q177	Complete chromosome	NZ_CP018150.1/NC_010258.1
Coxiella burnetii RSA 331	Complete chromosome	NC_010117.1/NC_010115.1
CbuG_Q212	Complete chromosome	NC_011527.1
Z3055	Complete chromosome	NZ_LK937696.1
nine mile phase II	Complete chromosome	NZ_CP035112.1/NZ_CP035111.1
RSA439	Complete chromosome	NZ_CP018005.1/NZ_CP018006.1
18430	chromosome	NZ_CP014557.1/NZ_CP014558.1
2574	chromosome	NZ_CP014555.1/NZ_CP040060.1
701CbB1	chromosome	NZ_CP014553.1/NZ_CP014554.1
14160-001	chromosome	NZ_CP014551.1/NZ_CP014552.1
Scurry_Q217	chromosome	NZ_CP014565.1
Henzerling	chromosome	NZ_CP014559.1/NZ_CP014560.1
Heizberg	chromosome	NZ_CP014561.1/CP014562.1
Namibia	chromosome	NZ_CP007555.1/NZ_CP007556.1
CbCVIC1	chromosome	NZ_CP014549.1/NZ_CP014550.1
42785537	chromosome	NZ_CP014548.1/NZ_CP014547.1
3345937	chromosome	NZ_CP014354.1/NZ_CP014355.1
14160-002	chromosome	NZ_CP014836.1/NZ_CP014837.1
Cb_C2	contig	GCF_000612785.1
Cb171_QLYMPHOMA	contig	GCF_000826165.1

Table A.1: List of analysed *C. burnetii* genomes

ISOLATE NAME	LEVEL	REFSEQ
Cb196 Saudi Arabia	contig	GCF_000820465.1
NL-Limburg	contig	GCF_000967075.1
Cb185	scaffold	
Dugway 7D77-80	scaffold	GCF_002247545.1
Dugway 7E65-68	scaffold	GCF_002247155.1
Q559	contig	GCF_002896835.1
Cb_B1	contig	GCF_000613025.1
Cb_B18	contig	GCF_000723305.1
Nine Mile RSA514	scaffold	GCF_002924395.1
Idaho Goat Q195	scaffold	GCF_002247185.1
Q532	contig	GCF_002896735.1
Q545	contig	GCF_002896755.1
Q556	contig	GCF_002896775.1
Dyer RSA345	scaffold	GCF_002247265.1
Turkey RSA315	scaffold	GCF_002247205.1
Australia RSA297	scaffold	GCF_002924305.1
Nine Mile RSA363	scaffold	GCF_002924345.1
Australia RSA425	scaffold	GCF_002924425.1
EV-Cb_BK10	contig	GCF_000723245.1
Ohio 314 RSA270	scaffold	GCF_002247285.1
DSTL_1R	contig	GCF_003849965.1
Ko_Q229	contig	GCF_002247225.1
Cb3506	contig	GCF_002896795.1
AuQ01	contig	NZ_JPVV01000067.1

Table A.2: List of analysed *C. burnetii* genomes

A.3 GENOMIC ANALYSIS OF NOVEL *C. burnetii* GENOMES

Table A.3 - Table A.4 shows the genes that were annotated from the TiHoQ-1019 plasmid with Prokka.

Table A.5 - Table A.6 shows the genes that were annotated from the TiHoQ-2219 plasmid with Prokka.

LOCUS_TAG	FTYPE	LENGTH_BP	GENE	EC_NUMBER	COG	PRODUCT
JDKBAGCM_00001	CDS	312				hypothetical protein
JDKBAGCM_00002	CDS	702				hypothetical protein
JDKBAGCM_00003	CDS	444				hypothetical protein
JDKBAGCM_00004	CDS	228				hypothetical protein
JDKBAGCM_00005	CDS	342				hypothetical protein
JDKBAGCM_00006	CDS	366				hypothetical protein
JDKBAGCM_00007	CDS	234				hypothetical protein
JDKBAGCM_00008	CDS	1839	traI			Multifunctional conjugation protein TraI
JDKBAGCM_00009	CDS	297				hypothetical protein
JDKBAGCM_00010	CDS	654	cpdA	3.1.4.53	COG1409	3′,5′-cyclic adenosine monophosphate phosphodiesterase CpdA
JDKBAGCM_00011	CDS	354				hypothetical protein
JDKBAGCM_00012	CDS	534	yrdA		COG0663	Protein YrdA
JDKBAGCM_00013	CDS	870	parB		COG1475	putative chromosome-partitioning protein ParB
JDKBAGCM_00014	CDS	1221				hypothetical protein
JDKBAGCM_00015	CDS	1005	noc			Nucleoid occlusion protein
JDKBAGCM_00016	CDS	1290				hypothetical protein

Table A.3: Table(A) of annotated genes from TiHoQ-1019 plasmid

LOCUS_TAG	FTYPE	LENGTH_BP	GENE	EC_NUM	COG	PRODUCT
JDKBAGCM_00017	CDS	558				hypothetical protein
JDKBAGCM_00018	CDS	339				hypothetical protein
JDKBAGCM_00019	CDS	459				hypothetical protein
JDKBAGCM_00020	CDS	570	xerC_1			Tyrosine recombinase XerC
JDKBAGCM_00021	CDS	1020				hypothetical protein
JDKBAGCM_00022	CDS	1086				hypothetical protein
JDKBAGCM_00023	CDS	507				hypothetical protein
JDKBAGCM_00024	CDS	228				hypothetical protein
JDKBAGCM_00025	CDS	1581				hypothetical protein
JDKBAGCM_00026	CDS	2199				hypothetical protein
JDKBAGCM_00027	CDS	1011				hypothetical protein
JDKBAGCM_00028	CDS	1221	xerC_2			Tyrosine recombinase XerC
JDKBAGCM_00029	CDS	216				hypothetical protein
JDKBAGCM_00030	CDS	465				hypothetical protein
JDKBAGCM_00031	CDS	732				hypothetical protein
JDKBAGCM_00032	CDS	234				hypothetical protein
JDKBAGCM_00033	CDS	684				hypothetical protein
JDKBAGCM_00034	CDS	972				hypothetical protein
JDKBAGCM_00035	CDS	1077				hypothetical protein
JDKBAGCM_00036	CDS	528				hypothetical protein
JDKBAGCM_00037	CDS	852				hypothetical protein

Table A.4: Table(B) of annotated genes from TiHoQ-1019 plasmid

LOCUS_TAG	FTYPE	LENGTH_BP	GENE	EC_NUMBER	COG	PRODUCT
OHJEEAEA_00001	CDS	312				hypothetical protein
OHJEEAEA_00002	CDS	1053				hypothetical protein
OHJEEAEA_00003	CDS	852				hypothetical protein
OHJEEAEA_00004	CDS	528				hypothetical protein
OHJEEAEA_00005	CDS	1077				hypothetical protein
OHJEEAEA_00006	CDS	972				hypothetical protein
OHJEEAEA_00007	CDS	234				hypothetical protein
OHJEEAEA_00008	CDS	732				hypothetical protein
OHJEEAEA_00009	CDS	516				hypothetical protein
OHJEEAEA_00010	CDS	216				hypothetical protein
OHJEEAEA_00011	CDS	1221	xerC_1			Tyrosine recombinase XerC
OHJEEAEA_00012	CDS	1011				hypothetical protein
OHJEEAEA_00013	CDS	2199				hypothetical protein
OHJEEAEA_00014	CDS	1581				hypothetical protein
OHJEEAEA_00015	CDS	228				hypothetical protein
OHJEEAEA_00016	CDS	507				hypothetical protein
OHJEEAEA_00017	CDS	1086				hypothetical protein
OHJEEAEA_00018	CDS	1020				hypothetical protein
OHJEEAEA_00017	CDS	1086				hypothetical protein
OHJEEAEA_00018	CDS	1020				hypothetical protein
OHJEEAEA_00019	CDS	570	xerC_2			Tyrosine recombinase XerC
OHJEEAEA_00020	CDS	459				hypothetical protein

Table A.5: Table(A) of annotated genes from TiHoQ-2219 plasmid

LOCUS_TAG	FTYPE	LENGTH_BP	GENE	EC_NUMBER	COG	PRODUCT
OHJEEAEA_00021	CDS	339				hypothetical protein
OHJEEAEA_00022	CDS	558				hypothetical protein
OHJEEAEA_00023	CDS	924				hypothetical protein
OHJEEAEA_00024	CDS	1005	noc			Nucleoid occlusion protein
OHJEEAEA_00025	CDS	1221				hypothetical protein
OHJEEAEA_00026	CDS	870	parB		COG1475	putative chromosome-partitioning protein ParB
OHJEEAEA_00027	CDS	534	yrdA		COG0663	Protein YrdA
OHJEEAEA_00028	CDS	354				hypothetical protein
OHJEEAEA_00029	CDS	747	cpdA	3.1.4.53	COG1409	3′,5′-cyclic adenosine monophosphate phosphodiesterase CpdA
OHJEEAEA_00030	CDS	1839	traI			Multifunctional conjugation protein TraI
OHJEEAEA_00031	CDS	234				hypothetical protein
OHJEEAEA_00032	CDS	366				hypothetical protein
OHJEEAEA_00033	CDS	342				hypothetical protein
OHJEEAEA_00034	CDS	228				hypothetical protein
OHJEEAEA_00035	CDS	444				hypothetical protein

Table A.6: Table(B) of annotated genes from TiHoQ-2219 plasmid

PUBLICATIONS

Sabrina Summer, Anna Smirnova, Alessandro Gabriele, Ursula Toth, **Akinyemi Mandela Fasemore**, Konrad U Förstner, Lauriane Kuhn, Johana Chicher, Philippe Hammann, Goran Mitulović, Nina Entelis, Ivan Tarassov, Walter Rossmanith, Alexandre Smirnov. (2020). YBEY is an essential biogenesis factor for mitochondrial ribosomes. *Nucleic Acids Research*, 48(17), pp 9762–9786.

Fasemore, A. M., Helbich, A., Walter, M. C., Dandekar, T., Vergnaud, G., Förstner, K. U., and Frangoulidis, D. (2021). CoxBase: an Online Platform for Epidemiological Surveillance, Visualization, Analysis, and Typing of Coxiella burnetii Genomic Sequences. *mSystems*, 6(6).

Frangoulidis, D., Walter, M.C., **Fasemore, A.M.**, Cutler, S.J. (2022). Coxiella burnetii . In: de Filippis, I. (eds) Molecular Typing in Bacterial Infections, Volume II. *Springer*, Cham.

Rabea Müller, **Akinyemi Mandela Fasemore**, Muhammad Elhossary, Konrad U. Förstner. (2021). *Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop*, PMLR 170:68-72

Best among mankind is
he who bring benefits
to others.

— Al-Tabarani

ACKNOWLEDGMENTS

I owe the success of this project to many individuals who have supported me during the course of my PhD. I am grateful for all the support and assistance.

I would like to thank Prof. Konrad Förstner for his pivotal role as my primary supervisor. I appreciate his trust and faith in me for recommending me for this project to PD Dr. Dimitrios Frangoulidis and the extensive support he provided for me during the course of the project. I am also grateful for the numerous opportunities he gave me as a graduate student to work on other interesting side projects. I am very grateful to Prof. Thomas Dandekar for his insightful feedback during the committee meetings, his contributions to all the manuscripts as well as his willingness to provide administrative support during the course of the project. I am very grateful to PD Dr. Dimitrios Frangoulidis for the opportunities he gave me to work on several interesting side projects during the course of this project, the funding for the project and also his willingness to make sure that I am very comfortable whenever I am in Munich.

I owe my gratitude to all the colleagues that helped me and taught me so much during the course of this project, most importantly Dr. Andrea Helbich (thank you for your kind support), the Data Science and Services group and Mathias Walter.

I am very grateful to my parents, Mr. and Mrs. Fasemore, my siblings and Mr. Bayo Sokunbi for their prayers, unending love, and guidance. My gratitude to my beautiful wife and daughter, Maryam and Maya, is immense. I am most grateful for the two of you and I appreciate your sincere trust, patience and outstanding support.

Finally, I am grateful to God for his mercy, protection, kindness, and love to my family.

CURRICULUM VITAE

Education

08/2018 – 09/2022
PhD student at the lab of Prof. Förstner, University of Würzburg, Würzburg, Germany.
Thesis title: Genomic and internet based analysis of *Coxiella burnetii*

10/2016 – 04/2018
MSc FOKUS Life Sciences, University of Würzburg, Germany.
Interdisciplinary Master in Molecular Life Sciences. Admitted to MSc/PhD fast track in April 2018.
Thesis title: Comparative analysis of expression patterns and regulatory motifs for several bacteria based on RNA-Seq data.

09/2009 – 04/2014
BSc. in Microbiology, Lagos State University, Lagos, Nigeria.
Thesis title: Bacteriological Examination of Ear Swabs from Asymptomatic Undergraduates in Lagos State University.

08/2011 – 09/2017
West African Senior School Certificate Examination, Federal Government College, Idoani, Ondo State, Nigeria.

Software Competencies

Frameworks
Flask, Django, Pyramid, React, Bootstrap, Jekyll, PySpark, Express, WordPress

Languages
Python, Bash, Javascript, R

Cloud technologies
AWS, Server provisioning with Ansible, Docker, Kubernetes

Database
MySql, MongoDB, Firebase, SQLAlchemy

Machine Learning
TensorFlow, Keras, Sci-Kit Learn

Systems & Networking
Gitlab, Github, Continous Integration, Unix/Linux systems

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić.