National Defence University
Series 1: Research Publications No. 13

USE OF MOLECULAR METHODS IN BIOTHREAT PREPAREDNESS

Katja Koskela

KATJA KOSKELA

USE OF MOLECULAR METHODS IN BIOTHREAT PREPAREDNESS

Research and Development Department

Centre for Military Medicine

and

Department of Bacteriology and Immunology

University of Helsinki

and

Doctoral Programme in Microbiology and Biotechnology

University of Helsinki

ACADEMIC DISSERTATION

To be presented for public examination, with the permission of the Faculty of Medicine, in Lecture Hall 5, University Main Building, (Fabianinkatu 33)

on 5th of May 2017, at 12 o´clock.

HELSINKI 2017

NATIONAL DEFENCE UNIVERSITY
SERIES 1: RESEARCH PUBLICATIONS NO. 13

MAANPUOLUSTUSKORKEAKOULU
JULKAISUSARJA 1: TUTKIMUKSIA NRO 13

USE OF MOLECULAR METHODS IN BIOTHREAT PREPAREDNESS

KATJA KOSKELA

NATIONAL DEFENCE UNIVERSITY
HELSINKI 2017

Katja Koskela: *Use of Molecular Methods in Biothreat Preparedness*
National Defence University
Series 1: Research Publications No. 13
Doctoral dissertation

Supervisors: Professor Simo Nikkari, MD, PhD
Research and Development Department
Centre for Military Medicine
Helsinki, Finland

Professor Mikael Skurnik, PhD
Department of Bacteriology and Immunology
Medicum
University of Helsinki
Helsinki, Finland

Reviewers: Docent Kaisu Rantakokko-Jalava, MD, PhD
Turku University Hospital, Turku, Finland

Docent Susanna Lukinmaa-Åberg, PhD
Päijät-Häme Central Hospital, Lahti, Finland

Custos: Professor Mikael Skurnik, PhD
University of Helsinki

Opponent: Florent Sebbane, PhD
INSERM Research Director
Center for Infection and Immunity of Lille
University of Lille, Institut Pasteur de Lille
Lille, France

Most recent publications in pdf-format: http://www.doria.fi

ISBN 978-951-25-2902-5 (Pbk.)
ISBN 978-951-25-2903-2 (PDF)
ISSN 2342-9992 (Print)
ISSN 2343-0001 (Web)

Maanpuolustuskorkeakoulu
National Defence University

Juvenes Print
Tampere 2017

“The more I learn, the more I realize how much I don't know”

Albert Einstein

TABLE OF CONTENTS

ABSTRACT

A biological threat is an epidemic or its threat caused by a microbe or biological material of a magnitude that would overwhelm healthcare services due to the contagiousness or wide distribution of infections. A biological threat can be naturally occurring, such as the West African Ebola epidemic of 2014-2016, or the consequence of an intentional release of a microbe or toxin.

The aim of this thesis was to develop and use molecular methods in order to reliably and rapidly identify potential biological threat agents. The focus was on the detection and typing of biological threat agents, whether they are naturally occurring or intentionally released. Different molecular methods were used: polymerase chain reaction (PCR) to detect and differentiate pathogenic from non-pathogenic bacterial strains, 16S ribosomal RNA (rRNA) gene sequencing to investigate polymicrobial samples, and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) spacer comparison for bacterial strain typing.

Cholera, a disease caused by *Vibrio cholerae* bacterium, is a major public health problem worldwide and a potential bioterrorism agent, according to the Centers for Disease Control and Prevention (CDC). In this thesis, an accurate PCR-based method was developed to detect *V. cholerae* strains: one assay for pathogenic strains and another for all *V. cholerae* strains. In addition, three different PCR platforms were compared. The PCR assays proved to be suitable for the reliable identification and differentiation of *V. cholerae* strains. The PCR platforms gave identical results, which indicate that the assays can be transferred between the platforms while maintaining sufficient sensitivity and specificity.

Two 16S rRNA gene-based detection methods, using Sanger sequencing or pyrosequencing, were employed to study the presence of bacterial residues in carotid artery tissue samples and in livers of splenomegalic voles. The objectives were to observe the utility of the two methods and compare their performance. Both methods were found to be convenient approaches to detect and identify bacterial species present in different matrices and thus could be employed when investigating polymicrobial samples. In addition, the two methods gave similar results which emphasises the reliability of the methods and their results.

The *Yersinia* genus includes three human pathogens; *Y. pestis*, the causal agent of plague and a potential biothreat agent, as well as *Y. enterocolitica* and *Y. pseudotuberculosis*, which commonly cause self-limiting enteritis. Due to the high level of DNA similarity between *Y. pestis* and *Y. pseudotuberculosis*, typing of *Yersinia* species and distinguishing them from each other has been challenging. Here, CRISPR spacers were used for typing *Yersinia pseudotuberculosis* complex strains. This method proved to be a promising tool, although the large diversity of different spacer sequences hindered

the clustering of different strains. In addition, CRISPR data of *Y. pseudotuberculosis* and *Y. pestis* were compared to examine phylogenetic relationships, but surprising lack of shared spacers limited any further resolution from being made.

In this thesis, molecular methods were developed and used to detect, identify, and type potential biological threat agents. PCR assays developed can be transferred to a field-deployable instrument and thus employed close to the patient, for example, during epidemics. PCR results were ready within a few hours, enabling a rapid and appropriate medical response. The 16S rRNA gene-based methods can be utilized in detection of biological agents, which are challenging or laborious to identify using traditional methods. The CRISPR-based sequencing method can be used for typing different strains of *Y. pseudotuberculosis*, if a comprehensive reference database is made available.

DNA sequencing and recently next-generation sequencing have become powerful tools to identify and type biological agents. Sequencing methods can also be utilized in epidemiological investigations and source tracking. Different molecular methods have evolved recently and detection has become fast and more reliable. Rapid detection of microbes enables swift medical countermeasures, and accurate identification and typing methods facilitate the ability to distinguish a natural outbreak from an intentional release.

TIIVISTELMÄ (SUMMARY IN FINNISH)

Mikrobin tai biologisen materiaalin aiheuttamaa epidemiaa tai sen uhkaa kutsutaan biologiseksi uhkaksi, silloin kun tauti ei tartuntavaaransa vuoksi ole yhteiskunnan normaaliresurssein hoidettavissa tai kun kyseessä on laaja epidemia, jonka hallitsemiseen tavanomaiset resurssit eivät riitä. Biologinen uhka voi olla luonnollinen, kuten esimerkiksi Länsi-Afrikan laaja ebolaepidemia vuosina 2014–2016, tai tahallinen mikrobin tai toksiinin levittäminen.

Väitöskirjan tavoitteena oli kehittää ja hyödyntää molekyylibiologisia menetelmiä biologisten uhka-agenssien nopeaa ja luotettavaa tunnistamista varten. Työn tarkoituksena oli tunnistaa ja tyypittää biologisia uhka-agensseja, olivat ne sitten luonnollisia ja tahallisesti levitettyjä. Työssä hyödynnettiin polymeraasiketjureaktiota (PCR) bakteerien tunnistamisessa sekä tautia aiheuttavien ja vaarattomien bakteerikantojen erottelussa toisistaan. Lisäksi käytettiin hyväksi 16S ribomaalisen RNA (rRNA) -geenin sekvensointia tutkittaessa polymikrobisia näytteitä, ja Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) -geenialueiden vertailua bakteerien tyypityksessä.

Kolera, *Vibrio cholerae* -bakteerin aiheuttama tauti, on maailmanlaajuisesti suuri kansanterveydellinen ongelma. Lisäksi Yhdysvaltain tartuntatautivirasto CDC (Centers for Disease Control and Prevention) on luokitellut *V. cholerae* -bakteerin yhdeksi potentiaaliseksi biouhka-agenssiksi. Työssä kehitettiin kahteen polymeraasiketjureaktioon perustuva *V. cholerae* -bakteerin tunnistusmenetelmä. Toisen menetelmän avulla pystytään tunnistamaan kaikki *V. cholerae* -kannat ja toinen havaitsee vain tautia aiheuttavat kannat. Lisäksi työssä vertailtiin kolmea eri PCR-laitetta. Työssä kehitetyt tunnistusmenetelmät osoittautuivat käyttökelpoisiksi ja luotettaviksi tunnistettaessa ja eroteltaessa *V. cholerae* -kantoja. Käytetyt PCR-laitteet antoivat samanlaiset tulokset, mikä mahdollisti tunnistusmenetelmän siirtämisen laitteiden välillä ilman, että sillä oli vaikutusta menetelmän herkkyyteen tai spesifisyyteen.

Kahta erillistä 16S rRNA -geeniin perustuvaa menetelmää yhdistettynä Sanger- ja pyrosekvensointiin hyödynnettiin tutkittaessa bakteerijäämiä kaulavaltimokudoksesta ja myyrien maksanäytteistä. Tavoitteena oli tutkia menetelmien käytettävyyttää ja vertailla menetelmiä toisiinsa. Käytetyt menetelmät osoittautuivat soveltuviksi bakteerilajien havaitsemiseen erilaisista näytematriiseista ja menetelmiä pystytään hyödyntämään tutkittaessa polymikrobisia näytteitä. Lisäksi käytetyt kaksi eri menetelmää tunnistivat samoja bakteerisukuja samoista näytteistä, mikä lisää menetelmien ja tulosten luotettavuutta.

Yersinia-bakterien sukuun kuuluu kolme ihmisille tautia aiheuttavaa lajia; *Y. pestis*, ruton aiheuttaja ja potentiaalinen biouhkabakteeri sekä *Y. enterocolitica* ja *Y. pseudotuberculosis*, jotka aiheuttavat suolistotulehduksia. Koska *Y. pestis* ja *Y. pseudotuberculosis* -bakteerien genomit ovat hyvin samanlaisia, kantojen tyypittäminen ja erottaminen toisistaan on haastavaa. Tässä työssä CRISPR spacer -geenialueita hyödynnettiin *Yersinia*-suvun kantojen tyypittämisessä. Tyypitysmenetelmä osoittautui lupaavaksi työkaluksi, vaikka toisistaan eroavien spacer-sekvenssien laaja kirjo vaikeutti kantojen ryhmittämistä ja vertailua. Lisäksi *Y. pestis* ja *Y. pseudotuberculosis* -kantojen fylogeneettistä suhdetta tutkittiin vertailemalla kantojen spacer-sekvenssejä. Yllättäen lajeilla oli hyvin vähän yhteisiä spacereita. Näin ollen tutkimus ei tuonut lisätietoa kantojen fylogeneettisistä suhteista.

Väitöskirjassa hyödynnettiin molekyylibiologisia menetelmiä potentiaalisten biologisten uhka-agenssien osoittamisessa, tunnistamisessa ja tyypittämisessä. Kehitetty PCR-menetelmä pystyttiin siirtämään kenttäkelpoiselle PCR-laitteelle, mikä tekee mahdolliseksi laitteen käyttämisen lähellä potilasta esimerkiksi epidemian aikana. Tulokset olivat valmiina muutamissa tunneissa, mikä mahdollistaa nopeat lääkinnälliset toimenpiteet. Lisäksi bakteerien 16S rRNA-geenialueen sekvensointiin perustuvia menetelmiä pystytään hyödyntämään sellaisten biologisten agenssien seulonnassa ja tunnistamisessa, joiden identifiointi tavanomaisilla menetelmillä, esimerkiksi viljelemällä, olisi haastavaa tai työlästä. CRISPR-geenialueeseen perustuvaa menetelmää pystyttäisiin hyödyntämään bakteerien tyypityksessä, mikäli laaja referenssitietokanta olisi käytettävissä.

DNA:n sekvensointi ja viime vuosina varsinkin uuden sukupolven sekvensointimenetelmät ovat osoittautuneet käyttökelpoisiksi työkaluiksi biologisten agenssien tunnistamisessa ja tyypittämisessä. Sekvensointimenetelmiä pystytään hyödyntämään myös epidemiologisissa tutkimuksissa ja selvitettäessä taudinaiheuttajan alkuperää. Erilaiset molekyylibiologiset menetelmät ovat kehittyneet valtavasti viimevuosina ja taudinaiheuttajien tunnistamisesta on tullut nopeaa ja luotettavaa. Nopea tunnistaminen luo perustan lääkinnällisten vastatoimien aloittamiselle. Tarkka tunnistaminen ja tyypittäminen antavat myös mahdollisuuden erottaa tahallinen levitys luonnollisesta epidemiasta.

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals (I-IV)

I **Koskela KA, Matero P, Blatny JM, Fykse EM, Olsen JS, Nuotio LO, Nikkari S.** A multiplatform real-time polymerase chain reaction detection assay for *Vibrio cholera*e. Diagn Microbiol Infect Dis. 2009 Nov, 65(3):339-44.

II **Renko J, Koskela KA, Lepp PW, Oksala N, Levula M, Lehtimäki T, Solakivi T, Kunnas T, Nikkari S, Nikkari ST.** Bacterial DNA signatures in carotid atherosclerosis represent both commensals and pathogens of skin origin. Eur J Dermatol. 2013 Jan-Feb, 23(1):53-8.

III **Koskela KA, Kalin-Mänttäri L, Hemmilä H, Smura T, Kinnunen PM, Niemimaa J, Henttonen H, Nikkari S.** Metagenomic Evaluation of Bacteria from Voles. Vector Borne Zoonotic Dis. 2017 Feb, 17(2): 123-33.

IV **Koskela KA, Mattinen L, Kalin-Mänttäri L, Vergnaud G, Gorgé O, Nikkari S, Skurnik M.** Generation of a CRISPR database for *Yersinia pseudotuberculosis* complex and role of CRISPR-based immunity in conjugation. Environ Microbiol. 2015 Nov, 17(11):4306–21.

ABBREVIATIONS

16S rRNA	16S ribosomal ribonucleic acid
B	Biological
BC	Before Christ
bp	Base pair
BSL	Biosafety level
BTWC	Biological and Toxin Weapons Convention
BW	Biological weapon
BWC	Biological Weapons Convention
C	Chemical
Cas	CRISPR associated genes
CDC	Centers for Disease Control and Prevention
cfu	Colony forming unit
Clm	Chloramphenicol
CRISPR	Clustered regularly interspaced short palindromic repeats
Dap	Diaminopimelic acid
DNA	Deoxyribonucleic acid
DOD	US Department of Defense
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen = German Collection of Microorganism and cultures
ELISA	Enzyme linked immonoabsorbent assay
FFI	Forsvarets forskningsinstitutt = Norwegian Defence Research Establishment
FBI	Federal Bureau of Investigation
HHS	Department of Health and Human Services
IPC	Internal positive control
Kan	Kanamycin
MALDI-TOF-MS	Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry
MERS-Cov	Middle East respiratory syndrome coronavirus
MLVA	Multi locus variable number tandem repeat analysis
MLST	Multilocus sequence typing
Nal	Nalidixic acid
NATO	North Atlantic Treaty Organisation
NGS	Next Generation Sequencing
NTC	No template control
PCR	Polymerase chain reaction
PFU	Plaque forming unit
PPE	Personal protective equipment
RDP	Ribosomal Database Project

RNA	Ribonucleic acid
SARS	Severe acute respiratory syndrome
SNP	Single nucleotide polymorphism
ST	Sequence type
ssp.	Subspecies
UN	United Nations
US	United States
USAMRIID	The United States Army Medical Research Institute of Infectious Diseases
USDA	United States Department of Agriculture
VEE	Venezuelan equine encephalitis
VNTR	Variable number tandem repeats
WGS	Whole genome sequencing
WHO	World Health Organization

1. INTRODUCTION

In 2001, the world learned that a few letters containing powder could create panic within a population. People became aware of biological weapons and of their use for hostile purposes. While researchers were aware of the threat, it nevertheless came as a surprise and no one was prepared for this kind of event [1, 2].

Many states have joined the Biological and Toxin Weapons Convention (BTWC) (also known as the Biological Weapons Convention (BWC)), which forbids the development and production of biological agents or toxins that can be used as biological weapons. However, the BTWC allows states to use biological agents and toxins for peaceful purposes. Every year, states send an up-to-date Confidence Building Measures declaration to the United Nations (UN), in which they report the facilities they have and what kind of research they conduct. Currently, the BTWC has 173 states parties and nine signatory states. Despite the fact that many states follow the BWC, terrorist groups and other hostile people operate outside of these restrictions.

Biological warfare is defined as the use of a biological agent (e.g., bacteria, virus or toxin) to kill or incapacitate humans, animals or plants. Intentional release of a biological agent is considered a bioterrorism attack. Biocrime is "the use of a biological agent to kill or make ill a single individual or small group of individuals" [3]. The agents can be spread through air, water or food or using vehicles, for example explosives. Biological warfare agents are classified by the Centres for Disease Control and Prevention (CDC) based on their severity and ability to spread.

A biological threat can also be naturally occurring, e.g., the West African Ebola outbreak in 2014–2016 and cholera outbreaks in Zimbabwe 2008–2009 [4]. Furthermore, pathogens evolve and occasionally pose significant health threats, e.g., coronaviruses (MERS-Cov and SARS) [5] and avian influenza virus [6]. The spectrum of different biological risks is presented in Figure 1.

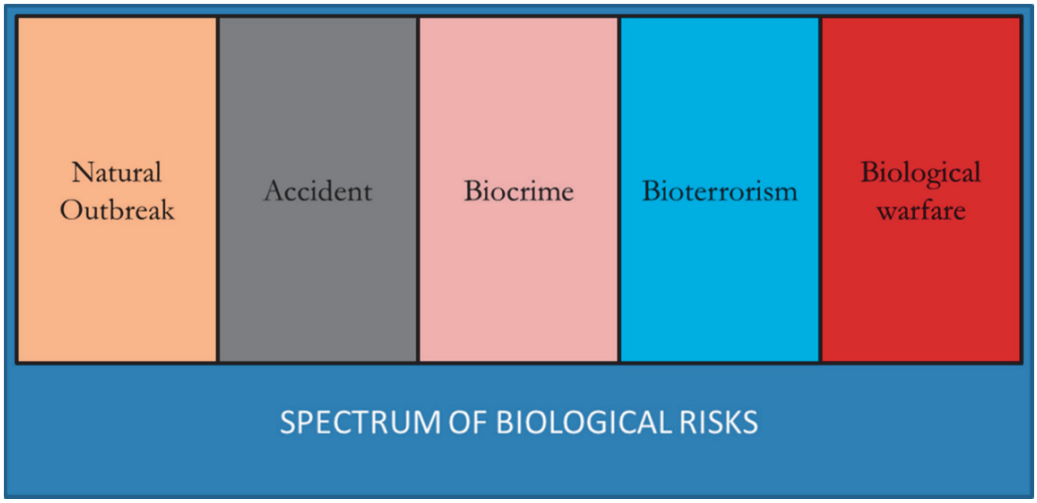


Figure 1: Spectrum of potential biological risks (modified from *Science Needs for Microbial Forensics, Developing Initial International Research Priorities* [2014]).

The focus of this thesis concerns the detection and typing of biological threat agents, whether they are naturally occurring or consequence of intentional release.

2. REVIEW OF THE LITERATURE

2.1. History

In this section, a short historical overview on bioterrorism and biological warfare is given. However, caution should be used with regard to the earliest bioterrorism attacks because they lack reliable scientific data.

2.1.1. Early use of biological warfare

Use of biological agents to intentionally harm the enemy has been reported as early as the 14th century BC. In those accounts, the Hittites sent diseased rams, perhaps infected with tularemia, to harm their enemies. In the 4th century BC, biological agents were spread using infected arrows [7]. Plague has been used as a biological warfare agent several times to eliminate enemies [8]. For example, in 1346, plague victims were catapulted over the city walls of Caffa, known as Feodosia in modern-day Crimea. Similarly, during the plague pandemics of the 14th century, the bodies of dead soldiers were catapulted into enemy ranks in Carolstein [9]. A similar tactic was used in a battle between Russia and Sweden in 1710 [8]. Polluting wells and water sources with dead animals has also been a common way to weaken the enemy throughout history [9]. Smallpox has also been used as a biological weapon. As an example, British forces distributed blankets containing smallpox to Native Americans in the 18th century. Because of this, a large smallpox outbreak within local tribes occurred [3, 9].

2.1.2. Biological weapon (BW) programs

In 1925, the Geneva Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases and Bacteriological Methods of Warfare was adopted and entered into force in 1928 [10]. However, following its ratification, at least, Belgium, Canada, France, Great Britain, Italy, the Netherlands, Germany, Poland, Japan, and the Soviet Union established biological weapon (BW) programs. The United States of America (US) also had a BW program until the country ratified the Geneva Protocol in 1975 [9, 11].

The US BW program was set up in 1942. A laboratory research facility was built in Maryland (Fort Detrick), as well as several other facilities and testing locations around the country [7]. This program remained active until 1972. Also the former Soviet Union established an extensive BW program called "Biopreparat Program". Both countries had large offensive programs and tried to weaponize biological agents which could be used to attack humans, crops and livestock [12].

Japan also had an extensive and systematic BW program. Under its official name of "Army Epidemic Prevention Research", Unit 731 was established in 1932. This unit, for example, inoculated prisoners-of-war with different biological agents to study their effects. They also developed bombs which could be used to spread biological agents. [7]. Japan also used biological weapons several times in China during World War II, e.g., in 1940, *Y. pestis* bacteria were spread by mixing rice and wheat with infected fleas in Chuhsien. This episode killed 21 people. A few weeks later, planes dropped *Y. pestis* bacteria on the city of Ningpo, killing 99 people in 34 days. In 1941, plague attacks in Suiyuan, Ninghsia and Hunan provinces caused serious epidemics [8].

2.1.3. Biological and Toxin Weapons Convention (BTWC)

During the late 1960s, it became obvious that the Geneva Protocol failed to prevent the proliferation and use of biological weapons. In response, the use of biological warfare agents was prohibited by the Biological and Toxin Weapons Convention (BTWC) established in 1972 and entered into force in 1975 [9]. However, the convention did not include an inspection mechanism and it remained easy to conceal BW programs within national biotechnology programs. Thus BTWC failed to completely stop the development and manufacture of biological weapons. For example, the former Soviet Union "Biopreparat Program" remained active until 1992 [13].

An inspection mechanism has been suggested as a way to improve BTWC, because the only way to currently monitor the action is through the annual reports volunteered to the UN by cooperative countries. In contrast, the Chemical Weapons Convention includes routine on-site inspections of industrial and chemical weapons-related facilities [14].

2.1.4. Biological warfare in the 20th century

Bioterrorism attacks have continued to feature in human history and, in 1984, the Bhagwan Shree Rajneesh cult contaminated salad bar products with *Salmonella enterica* serovar Typhimurium in Oregon. A total of 14 restaurants were involved in the attack and 751 fell ill. It was not until 1986 that this was recognized as bioterrorism because this kind of outbreak could also happen naturally [7]. This example demonstrates the challenge faced trying to determine if an outbreak is natural or deliberate.

Another example is the Aum Shinrikyo sect, whose members tried to attack Tokyo many times during 1990–1995 using anthrax spores (i.e., highly resistant, dormant bacterial structures) and botulinum toxin. Fortunately, they only obtained non-pathogenic strains and had difficulties producing aerosols [7, 15].

2.1.5. Amerithrax and powder letters

In 2001, letters containing athrax spores were posted to US media and government officials resulting in the deaths of five people [3]. This event, now known as "Amerithrax" was a wakeup call for many countries to prepare themselves for bioterrorism attacks [16]. This attack exposed at least 10,000 people to anthrax spores and all recieved the antibiotic treatment. Extensive forensic investigations were conducted over seven years and decontamination of offices and other buildings took years to complete. The person allegedly responsible was an anthrax researcher, Bruce Ivins, who worked at the Fort Detrick biodefence laboratory in Maryland. He committed suicide in 2008 before facing trial [17]. After the US events, letters containing white powder also appeared in Finland. These letters did not contain anthrax spores.

Occasionally, letters containing harmless powder are sent to create panic and disrupt normal activity or a government or civilian agency. Typically, these letters contain, for example, flour and/or uncommon biological agents [18, 19]. When investigating powder letters, it is of paramount importance to determine if the letter contains pathogenic agents or not.

2.1.6. Laboratory accidents

Conducting research with BW agents can lead to serious accidents if risk assessments and precautions are not taken into account. This is especially crucial when working with spore-forming bacteria.

Probably the best know BW accident happened in 1979 in Sverdlovs (now Ekaterinburg, Russia) where anthrax spores were accidently released from a military microbiology facility, contaminating a large area around the BW research center [20, 21]. At that time, Soviet authorities claimed the epidemic was caused by contaminated meat and only in 1992, Russia admitted that "military developments were the cause" [22]. The number of victims remains unknown [7].

Another recent example is the case where the US Department of Defense (DOD) accidently sent live anthrax spores to 183 laboratories in seven countries in 2015 [23]. This was because the method used to inactivate the spores was insufficient. Also, the procedure to confirm that samples did not contain live agents was not done properly, in that the incubation period to culture the spores was too short.

2.2. Potential and probable biothreats today

2.2.1. Select agents and toxins

Select agents are bacteria, viruses and toxins that can be used as biological weapons and which "pose a severe threat to public health and safety" [24]. The CDC has a list which currently contains 65 agents and toxins that pose a threat to humans, animals or plants. These agents have been divided into categories A, B and C, depending on their severity (Table 1). Agents posing the greatest threat belong to the category A and include those that are easily transmitted, have high death rates and have a major public health impact. Other organizations and countries have their own lists of potential biological agents, e.g., North Atlantic Treaty Organization (NATO) and Australian group. In addition, the US Departments of Health and Human Services (HHS) and Agriculture (USDA) have established the Select Agents and Toxin List. Twelve potential biological agents also known as the "dirty dozen", include the following agents: *Bacillus anthracis*, *Y. pestis, Francisella tularensis, Brucella* spp., *Burkholderia pseudomallei, Coxiella burnetii*, Variola virus, Venezuelan equine encephalitis virus, Marburg and Ebola virus, Botulinum toxin, Ricin, and Staphylococcal enterotoxin B (SEB) [25]. However, when talking about the threat posed by bioterrorism agents it is good to keep in mind that only a few are relatively easy to prepare and disperse [26].

Highly pathogenic agents must be handled in Biosafety level (BSL) 3 or 4 laboratories. BSL-3 and BSL-4 laboratory requirements are defined by the WHO Laboratory Biosafety Manual. There are also national regulations, for instance, defining the biosafety level where a certain agent can be handled. High containment laboratory facilities and protocols can prevent the accidental release of highly pathogenic agents to the environment. Safety precautions and structural requirements can also prevent the theft of pathogens and protect workers from exposure.

In the next section, some agents that have been used in bioterrorism attacks or are otherwise considered potential bioterrorism agents are described in more detail.

Table 1: Category A and B agents of the CDC classification

		Agent	Disease
Category A	**Bacteria**	*Bacillus anthracis*	anthrax
		Clostridium botulinum toxin	botulism
		Yersinia pestis	plague
		Variola major and *other related pox viruses*	smallpox
		Francisella tularensis	tularemia
	Viruses	Viral hemorrhagic fevers	
		Arenaviruses: Junin	
		Machupo	
		Guanarito	
		Chapare	
		Lassa	
		Lujo	
		Bunyaviruses: Hantaviruses causing Hanta Pulmonary syndrome	
		Rift Valley Fever	
		Crimean Congo Hemorrhagic Fever	
		Flaviviruses: Dengue	
		Filoviruses: Ebola	
		Marburg	
Category B	**Bacteria**	*Burkholderia pseudomallei*	melioidosis
		Coxiella burnetii	Q-fever
		Brucella species	brucellosis
		Burkholderia mallei	glanders
		Chlamydia psittaci	psittacosis
		Rickettsia prowazekii	typhus fever
	Toxins	*Ricin toxin (Ricinus communis)*	
		Epsilon toxin (Clostridium perfringens)	
		Staphylococcus enterotoxin B (SEB)	
	Food- and waterborne pathogens		
	Bacteria	Diarrheagenic *E.coli*	
		Pathogenic Vibrios	
		Shigella species	
		Salmonella	
		Listeria monocytogenes	
		Campylobacter jejuni	
		Yersinia enterocolitica	
	Viruses	Caliciviruses	
		Hepatitis A	
	Protozoa	*Cryptosporidium parvum*	
		Cyclospora cayatanensis	
		Giardia lamblia	
		Entamoeba histolytica	
		Toxoplasma gondii	
		Naegleria fowleri	
		Balamuthia mandrillaris	
	Fungi	Microsporidia	
	Mosquito-borne encephalitis viruses	West Nile virus (WNV)	
		LaCrosse encephalitis (LACV)	
		California encephalitis	
		Venezuelan equine encephalitis (VEE)	
		Eastern equine encephalitis (EEE)	
		Western equine encephalitis (WEE)	
		Japanese encephalitis virus (JE)	
		St. Louis encephalitis virus (SLEV)	

2.2.2. Anthrax – *Bacillus anthracis* bacteria

Bacillus anthracis is a Gram-positive, spore forming, non-motile bacterium and is the etiological agent of anthrax [27]. Anthrax is a zoonotic disease and is endemic in many parts of the world [28]. *B. anthracis* forms spores that are very stable and remain viable for decades. These spores can be aerosolized, which makes spreading the disease easier. High mortality rate, transmission by aerosol, and persistence in the environment make this BW agent one of the most threatening [29]. There are three different pathways for *B. anthracis* to infect humans: through skin contact (cutaneous anthrax), ingestion (gastrointestinal anthrax) or inhalation (pulmonary anthrax). Pulmonary anthrax is the most severe form [30].

Besides being a potential BW agent, *B. anthracis* is also a public health problem. The natural hosts for *B. anthracis* are herbivores, and humans can contract the disease from infected animals or animal products [31]. Vaccination has lowered the incidence of the disease within livestock, but in some parts of Asia and Africa vaccination programs are rather sporadic and the problem still exists [32]. This explains occasional natural outbreaks of anthrax in different parts of the world [33-36].

2.2.3. Botulism – *Clostridium botulinum* bacteria

Clostridium botulinum is an anaerobic, spore-forming bacterium and is commonly present in the environment (soil, dust and aquatic sediments) [37]. This bacterium produces botulinum toxin, which is one of the most poisonous substances known. In humans, the toxin causes botulism a neurological disease that is fatal without treatment [38].

Because botulinum toxin is so lethal and easy to produce, it is considered to be a potential biothreat agent and is category A agent according to the CDC. However, it is not the most optimal choice for a biological weapon because it does not spread from human-to-human, it degrades rapidly in the environment and an antitoxin is available [38, 39].

2.2.4. Plague – *Yersinia pestis* bacteria

The first plague pandemic, also known as the Justinian plague, started in Egypt (A.D. 541). The pandemic spread from Africa to Europe and Asia, killing 50–60% of the human population. The second pandemic (Black Death) in 14th century killed 20–30 million people in Europe. In 1855, the third pandemic started in China and killed more than 12 million people there and in India [40].

In 1894, Alexandre Yersin isolated *Yersinia pestis*, the causal agent of plague [41]. *Y. pestis* is a Gram-negative coccobacillus and can cause three different forms of plague; bubonic, pneumonic and septicemic. Pneumonic plague is the most severe form with a mortality rate approaching 100% without appropriate antibiotic treatment [42]. With treatment, the mortality rate can still be around 50%. *Y. pestis* is highly infectious and as few as 10 bacteria can cause the disease. The high mortality rate, spreading via aerosols, and a low infectious dose are the reasons why *Y. pestis* is considered a potential bioterrorism agent [43, 44].

Rodents are the main animal reservoir for *Y. pestis*. Humans can contract the disease from infected animals or through vectors, i.e., a flea bite [45]. Human-to-human transmission is possible through saliva or mucus droplets emitted from a person suffering from pneumonic plague [44]. Lately, plague has been categorized as a re-emerging disease and it has caused huge public health problems, especially in African countries. During a five-year period from 2010 to 2015, almost 3,300 human cases were reported globally, with the majority of them in Africa [46, 47].

2.2.4.1 *Yersinia pseudotuberculosis*

Yersinia pseudotuberculosis is a Gram-negative, non-spore-forming rod or coccobacilli and a member of the genus *Yersinia*. The *Yersinia* genus includes three human pathogens; *Y. pestis, Y. enterocolitica* and *Y. pseudotuberculosis* [42]. *Y. enterocolitica* and *Y. pseudotuberculosis* commonly cause self-limiting enteritis [48]. Earlier studies, suggest that *Y. pestis* emerged approximately 1,500–6,400 years ago from a *Y. pseudotuberculosis* clone [49, 50] and these two strains are ~97% similar at the DNA level [51]. Due to this reason, the typing of *Yersinia* species has been challenging. While several approaches have been applied to typing *Yersinia* species [52-55], discrimination of isolates remains a challenge.

Y. pseudotuberculosis occurs naturally worldwide and can infect a variety of domestic and wild animals, such as rodents, wild birds, deers and goats [56-59]. The bacterium causes acute gastroenteritis and mesenteric lymphadenitis in humans, with the main symptoms being abdominal pain and fever. Human infections occur after consuming contaminated food products. Typically, *Y. pseudotuberculosis* causes self-limited infection, but occasionally patients are hospitalized and post-infection complications such as reactive arthritis and erythema nodosum are common [60, 61].

Y. pseudotuberculosis can cause food and water-borne outbreaks. During 2000–2015 eight outbreaks of *Y. pseudotuberculosis* were reported in Finland, and in many cases the source has been contaminated carrots, iceberg lettuce or raw milk [60-63]. The incidence rate of *Y. pseudotuberculosis* in Finland during the past ten years is presented in Figure 2. Differences among years are mainly due to the presence of individual epi-

demics. For example, the high peak in 2006 was caused by two separate epidemics that took place that year.

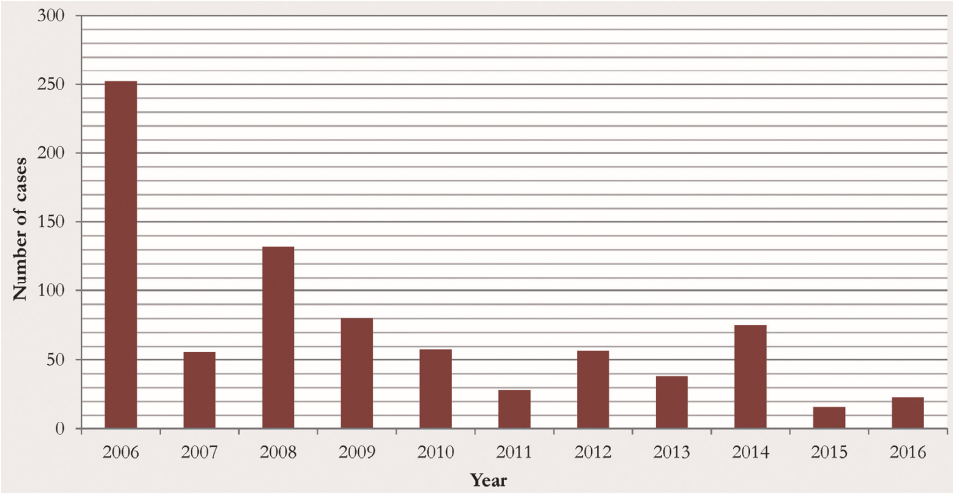


Figure 2: Human infectious of *Y. pseudotuberculosis* in Finland (years 2006-2016). Source: Finnish Infectious Diseases Register, National Institute for Health and Welfare.

2.2.5. Smallpox – Variola major virus

The Variola major virus is a member of the genus *Orthopoxvirus* and causes the severe disease known as smallpox. Besides being a highly lethal disease (with a mortally rate of around 30%), it is also highly contagious [64]. It has caused massive epidemics throughout the last 300 years and perhaps 300–500 million deaths in the 20th century alone. Thanks to global vaccination campaigns, smallpox was successfully eradicated in 1980 [65].

Following its eradication, a decision was made to destroy all smallpox stocks worldwide. Only the US and Russia were allowed to store stocks of smallpox in restricted locations. There has been discussion if these remaining stocks should be destroyed [66].

If smallpox is used as a biological weapon, it will be a serious threat due to its high mortality rate and because an increasing proportion of the human population is unvaccinated. The disease is highly infectious and could rapidly cause a pandemic [67].

2.2.6. Tularemia – *Francisella tularensis* bacteria

The *F. tularensis* bacteria cause tularemia, a zoonotic disease which can be lethal to humans and animals. It is a highly virulent bacterium and as few as ten bacterial cells can cause a fatal disease [68].

F. tularensis occurs throughout North America and Eurasia and can cause local outbreaks in endemic areas. Depending on infection route, *F. tularensis* can cause ulceroglandular, glandular, oropharyngeal, gastrointestinal or pneumonic tularemia [69]. The species is divided into four subspecies: *tularensis* (type A), *holartica* (type B)*, novisida* and *mediasita* [70]. *F. tularensis* subsp. *tularensis,* also called type A strain, is the most virulent with a mortality rate of 5–6% without treatment. Mortality of the pneumonic form of type B is less than 0.5% [69]. Type A strain occurs only in North America while type B is present throughout the Northern Hemisphere, including Finland [70]. No confirmed reports of human-to-human transmission exist. The other two subspecies (*novisida* and *mediasita*) rarely cause disease in human [71].

2.2.7. Ebola virus

The highly virulent Ebola virus causes hemorrhagic fever and currently comprises five different strains known as Zaire, Sudan, Taï Forest, Bundibugyo and Reston. Zaire and Sudan ebolaviruses are the most virulent [72].

Ebola virus can spread directly from infected animals (e.g., fruit bats and chimpanzees) or by handling bush meat. Direct transmission among humans can occur through infected blood, body fluids, or direct skin contact. It is critically important to use approved personal protective equipment (PPE) (e.g., clothing, goggles and cloves) when treating Ebola patients[73, 74].

Ebola has great potential as a bioterrorism agent due to its ability to cause severe hemorrhagic fever with a high mortality rate that can reach 90% [75]. At the moment, there is no specific vaccine against Ebola, but many potential candidates are being developed [72].

The Ebola outbreak in West Africa during 2014–2016 was the largest and longlasting Ebola epidemic to date, with a death toll of over 11,000. It was the first time that an Ebola outbreak occurred in West Africa, and demonstrated the lack of preparedness for these events in developing countries. International response and coordination must be improved so that similar events are prevented in the future, e.g., increase zoonotic surveillance and support public health systems [76].

2.2.8. Brucellosis – *Brucella* species bacteria

The genus *Brucella* can be divided into classical species of *B. abortus, B. melitensis, B. canis, B. ceti, B. ovis, B. neotomae, B. pinnipedialis* and *B. papionis* and atypical species *B. microti* and *B. inopinata.* New *Brucella* isolates have been identified but these have not yet been formally described and classified [77]. Of the classical *Brucella* species, *B. abortus*, *B. melitensis, B. suis* and *B. canis* are known human pathogens [78].

Brucella is a worldwide zoonotic disease that causes abortion in livestock such as sheep, goats, cows and pigs. It can infect humans through unpasteurized animal products (e.g., cheese or milk) or through direct contact with infected animal tissue [79].

From the bioterrorism perspective, *Brucella* is highly infectious; 10–100 bacterial cells can cause the disease. It is highly resistant in nature but does not form spores [78].

2.2.9. Q-fever – *Coxiella burnetii* bacteria

The causal agent of Q-fever is *Coxiella burnetii*, a small (0.2-1.0 μm) coccobacillus. Cultivation of the bacterium is challenging and the first cell-free medium was not reported until 2009 [80]. *C. burnetii* is resistant to many disinfectants, e.g., formaldehyde vapour in low humidity is ineffective [81].

Q-fever can manifest an acute or chronic disease. Clinical outcomes can be asymptomatic or severe, typically causing fever, pneumonia or even death. The mortality rate varies from 0.5 to 1.5%. The infectious dose is very low (1-10 colony forming units [cfu]), and symptoms normally present 10 to 90 days after inhalation. Chronic Q-fever can occur many months or even years after the infection [81].

The US and former Soviet Union examined *C. burnetii* in their offensive biological weapon programs. If this agent is used in a biological attack, the most likely delivery route would be via aerosol. The WHO estimated in 1970 that, if 50kg of *C. burnetii* is aerosolized over an urban area with half a million inhabitants, it would cause 150 deaths as well as 9,000 chronic and 125,000 acute Q-fever cases [82].

Natural Q-fever infections occur worldwide, making it difficult to distinguish natural and deliberate outbreaks. Recently, *C. burnetii* has caused large outbreaks in livestock of the Netherlands and, in November 2009, Dutch authorities reported 2,293 human cases and six deaths. They began vaccinating and culling livestock (mainly goats and sheep) in 2010. After that, acute human cases decreased significantly [83]. During outbreaks, it has been shown that *C. burnetii* can spread long distances and remain resistant in the environment for long time periods [84, 85].

2.2.10. Ricin toxin – *Ricinus communis* plant

Ricin toxin occurs naturally in *Ricinus communis*, a tropical plant cultivated for castor oil production. Ricin toxin is a by-product released during the processing of castor beans [86].

Natural exposure to ricin is very rare, even though there are cases where people have ingested castor beans, intentionally or accidentally [87]. Nevertheless, people can

be exposed to ricin via ingestion, inhalation or injection, with the latter two being the most lethal routes. Ricin toxin blocks protein synthesis and causes multiorgan toxicity in humans. Severity depends on the exposure route, the amount and purity of the toxin and size of toxin particles [88].

Ricin has been used in the recent history, e.g., letters containing ricin were sent in the US in 2003 [89].

2.2.11. *Vibrio cholerae*

Vibrio cholerae is a Gram-negative, motile rod or comma-shaped bacterium that causes cholera. The bacterium is transmitted via the faecal-oral route or through exposure to contaminated water or food. Cholera is an unusually severe diarrhoeal disease that can kill healthy adults through massive electrolyte loss and dehydration within a few hours of symptoms presenting [90, 91].

Cholera is endemic in Africa, Asia, the Middle East and South America and continues to be a huge public health problem especially in countries were sanitary conditions are poor [92]. In 2013, 47 countries reported a total of 129,064 cholera cases and 2,102 deaths to WHO. Apart from being a burden to global public health, *V. cholerae* is also a potential agent to be used in a bioterrorism attack. Therefore, *V. cholerae* is classified as a category B bioterrorism agent by the US CDC.

Of the two most pathogenic *V. cholerae* serogroups (O1 and O139), strains belonging to group O1 have caused most of the cholera epidemics. These two serogroups harbor the cholera toxin (*ctx)* and toxin co-regulated pilus (*tcp)* genes, which together regulate the expression of cholera toxin. This toxin causes the main symptoms of cholera [93]. Other *V. cholerae* serogroups (also called non-O1 *Vibrio cholerae*) can sometimes cause bacteremia, especially in immunocompromised patients [94].

Rehydration is the main treatment for cholera patients. In addition, effective antibiotic therapy may be applied, but therapy should be based on local antimicrobial resistance [95]. There are also two WHO prequalified oral cholera vaccines available, but if a large epidemic were to occur, there are not enough vaccines available to meet the need [96]

2.3. Methods to detect and type biological agents

The most common methods used to identify bacterial strains are presented in this section. Also different strain typing and subtyping methods are discussed in the text.

The principle of typing is to discriminate and identify different bacterial isolates within the same species. Typing methods provide high-resolution discrimination of closely-related isolates and separate avirulent and virulent strains. Accurate and rapid analyses can minimize the response time, help to prevent the spread of an outbreak and enable to start appropriate medication. Typing bacterial strains can also be used to track the source and determine whether the outbreak is natural or deliberate.

2.3.1. Culture

Culture is the oldest method to identify bacterial strains. Louis Pasteur used urine and meat extracts to grow bacteria. Robert Koch on the other hand started using solid culture media. First, potato pieces were used as culture media before agar was discovered. Solid cultures enabled making pure cultures from bacterial strains [97]. The nutrient composition of the culture broth or media is crucial to bacterial growth. Furthermore, incubation time, atmosphere, and temperature can vary depending on bacterial strain [98]. Culture remains an important method in clinical microbiology but it can take days to grow the strains and resolution of similar strains can be poor. As an example, a potential biothreat agent *F. tularensis* is slow-growing and it can take several days to obtain a result [99]. It has also been said that only about 1% of bacteria can be cultivated using traditional culture techniques, so the majority cannot be cultured or are very difficult to grow [100]. For example, *C. burnetii* is challenging to cultivate because it requires a specific technique and conditions to grow [101].

2.3.2. Lateral flow devices

Lateral flow devices can be used to detect the presence or absence of substances in a sample. Various chemical components (e.g., drugs, pesticides) and biological agents (e.g., bacteria, viruses, and toxins) can be detected using this technology [102]. One broadly known application of this technology is the home pregnancy test. The principle of these tests is simple; if the sample contains the target analyte, it binds to specific antibodies, conjugated with colored or fluorescent particles. This component is then visualized in the test line zone when it reacts with specific biological components (mostly antibodies). In every device there is also a control line wich indicates the proper conditions in the test [103].

While lateral flow devices are convenient for screening in field conditions, they are also used in basic clinical laboratories. For rapid detection, lateral flow devices can provide an answer within minutes but these tests lack specificity and sensitivity [104]. Thus, it is important to check the limits of the particular unit used and compare similar devices from different manufacturers. These devices can still give a good preminary result in the event of a bioterrorism attack. Because tests are easy-to-use and fast, first responders or reconnaissance teams can perform the test at the scene

and begin appropriate treatment, if needed. While helpful, it remains important to confirm the result using more reliable methods.

2.3.3. Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry

In recent years, matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOFMS) has become a new diagnostic tool to identify microbes. It has been used for the detection, identification, and typing of strains, for antibiotic resistance studies, and for epidemiological studies [105]. Several publications involving highly pathogenic agents and toxins such as, *Brucella* spp., *B. anthracis*, *Y. pestis*, *F. tularensis* and *C. botulinum* relied on MALDI-TOFMS for their identification and typing [106-109]. For example, Lista and colleagues [108] established accurate and rapid identification method for *Brucella* species. They showed that it was possible to distinguish closely-relative isolates but it required building a custom-made reference library containing 17 different *Brucella* variants. In this study, 152 different *Brucella* pure cultures were investigated. MALDI-TOFMS has also been used to distinguish different *Botulinum* neurotoxins and detect anthrax toxin [109].

2.3.4. Polymerase chain reaction

Polymerase chain reaction (PCR) was first discovered in the 1980s by Kary Mullis. PCR is a technique to amplify a specific DNA template with the aid of a thermostable DNA polymerase and primers specific to the target gene region [110]. The PCR reaction is based on a thermal cycle of a DNA denaturation step, a primer annealing step and an extension step (Figure 3). Typically, 25-40 cycles are performed in a single PCR run. Using PCR, billions of copies can be made from the original DNA template. After a PCR run, amplicons are visualised using agarose-gel electrophoresis and a fluorescent stain. Over the past 20 years, PCR technology has developed to the modern standard of real-time PCR where reaction success is monitored during thermal cycling with the aid of a fluorescent probe (Figure 4).

One advantages of real-time PCR over conventional PCR is multiplexing, where the simultaneous amplification of several different targets can be performed in a single reaction. Real-time PCR is less laborious, faster and since there is no need to open the reaction tubes after the PCR run, contamination of the laboratory and pure reagents can more easily be prevented.

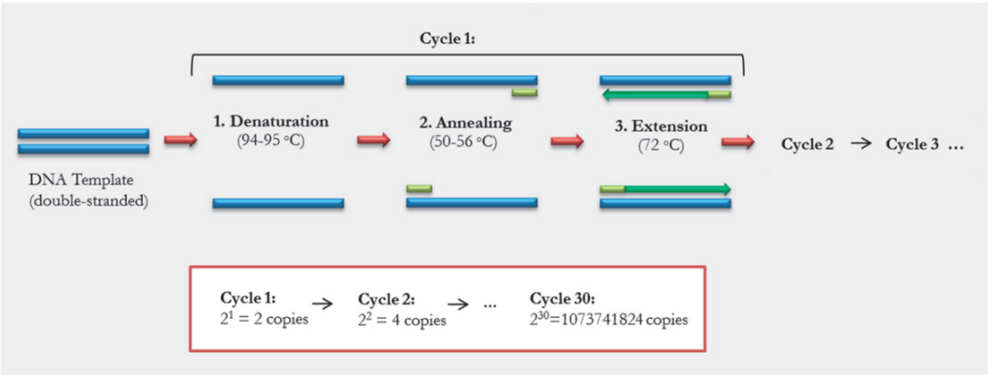


Figure 3: PCR reaction workflow. PCR is based on thermal cycles. In the denaturation step the double-stranded DNA is separated into two single strands. At the annealing temperature primers selectively bind to the complementary target on the denatured DNA strand (light green boxes) and in extension step the polymerase enzyme constructs the DNA from nucleotide bases present in the reaction mixture.

PCR is a very sensitive method. Thus it is advisable to perform all PCR steps in separate rooms. In addition, reagents should be divided into aliquots, disposable gloves used and changed frequently and pipettes with disposable filter tips used in all steps. In a pre-PCR clean room (where PCR mixtures are prepared), it is wise to wear separate PPE.

A possible drawback of PCR is the potential for false positive and false negative results. A false positive can occur if closely related neighbor species are amplified during the PCR reaction. A false negative on the other hand, is normally due to the low sensitivity of the assay. Sensitivity is crucial especially with samples containing a small amount of the agent. Another reason for a false negative result is the presence of inhibitors in the PCR reaction. PCR inhibitors are inhibiting substances, which may be present in the sample, and can affect the sensitivity of the assay [111]. To overcome the drawbacks of PCR, it is important to validate the assay. This includes specificity testing (power to discriminate closely-related species), sensitivity testing (determining the detection limit of the assay) and ubiquity testing (checking the assay's ability to detect all target strains). It is also crucial to test that the sample matrix does not interfere with the PCR reaction, e.g., when analysing clinical samples. It is advisable to use an internal positive control (IPC) to make sure that inhibitors do not affect the final results.

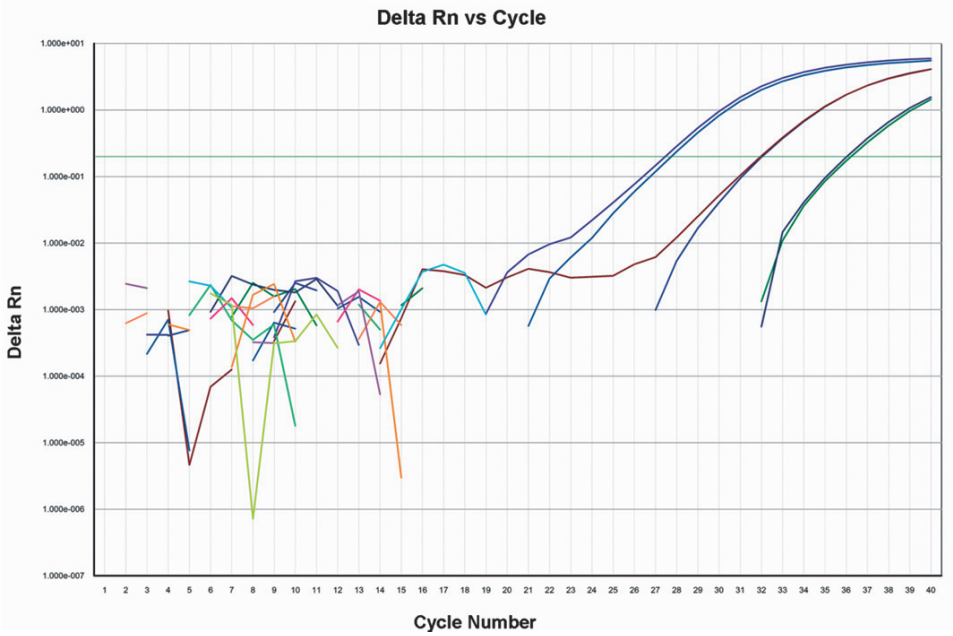


Figure 4: Real-time PCR amplification chart. If the sample contains the specific target gene, the amplification curve is obtained when the reaction includes fluorescent probe.

PCR has been used in many applications when detecting potential biological warfare agents and other biological threats [112-115]. PCR can be used for typing bacterial strains; however, this usually requires multiple primers and probes. Variable number tandem repeat (VNTR) and multilocus VNTR analyses (MLVA) are examples of broadly used PCR typing methods. In multilocus sequence typing (MLST), PCR is combined with sequencing and the sequence types (ST) are determined from the different bacterial isolates.

PCR can also be utilized in single nucleotide polymorphism (SNP) typing, where single base-pair differences can be detected in genomes. SNP has become an essential typing tool especially when typing genetically monomorphic pathogens [116]. SNP typing is widely used in epidemiological and forensic investigations [117, 118]. In addition, PCR can be used to amplify CRISPR gene regions for typing bacterial species.

2.3.5. 16S rRNA gene-based detection methods

Carl Woese discovered the usefulness of small-subunit rRNA (16S rRNA) gene sequences for the identification and classifications of bacteria [119]. Broad-range PCR method (Br-PCR) is based on using a universal primer pair to amplify the bacterial 16S rRNA gene or a part of it [120].

The 16S rRNA gene is present in all bacteria, thus methods based on this gene are useful for screening and detecting bacterial strains from diverse samples. They are especially useful when investigating those that cannot be cultured [121] or when investigating samples containing only small amounts of bacterial DNA. The 16S rRNA PCR and sequencing methods have been widely used for studying different sample types, e.g., clinical samples, environmental samples, and biothreat agents [122-126].

2.3.6. Clustered regularly interspaced short palindromic repeats (CRISPR)

CRISPR elements are present in half of all bacterial species. During invasion (by phages or plasmids), bacteria may capture a DNA sequence originating from the invading phage or plasmid into the CRISPR locus. This sequence, called a spacer, is integrated between direct repeats (DR), i.e., an array of repetitive sequences. In other words, the CRISPR locus is composed of repeats of sequence reads called DRs, and in between the DRs are unique spacers. CRISPR-Cas (CRISPR-associated genes) system encodes a so-called RNA-mediated immune system which can destroy invading plasmids or bacteriophages if bacteria have earlier integrated a spacer specific for this invading element (Figure 5) [127, 128].

Using PCR and sequencing techniques, CRISPR arrays can be amplified and sequenced. CRISPR gene regions and particularly spacer arrangements have been used for typing bacterial species.

Recently, the CRISPR-Cas system (and especially one Cas protein [Cas 9]), has been employed in genome editing. It has been challenging to edit the genome in a precise position at the target gene but CRISPR-Cas 9 has made this possible. Nowadays, this technique has been widely adopted and it has, for example, been successfully used to perform single point mutations (insertions and deletions), larger deletions and genomic rearrangements [129]. Despite its success, there has been speculation and concern about this new technique because it can be used to modify bacterial genomes and thus creating a huge biological threat. The US Intelligence Community stated in the Worldwide Threat Assessment report (published in February 2016) that genome editing may increase the risk of the creation of harmful biological agents (https://www.dni.gov/files/documents/SASC_Unclassified_2016_ATA_SFR_FINAL.pdf). In summary, this technology has many good applications, but there is also a possibility to misuse it for illegal purposes.

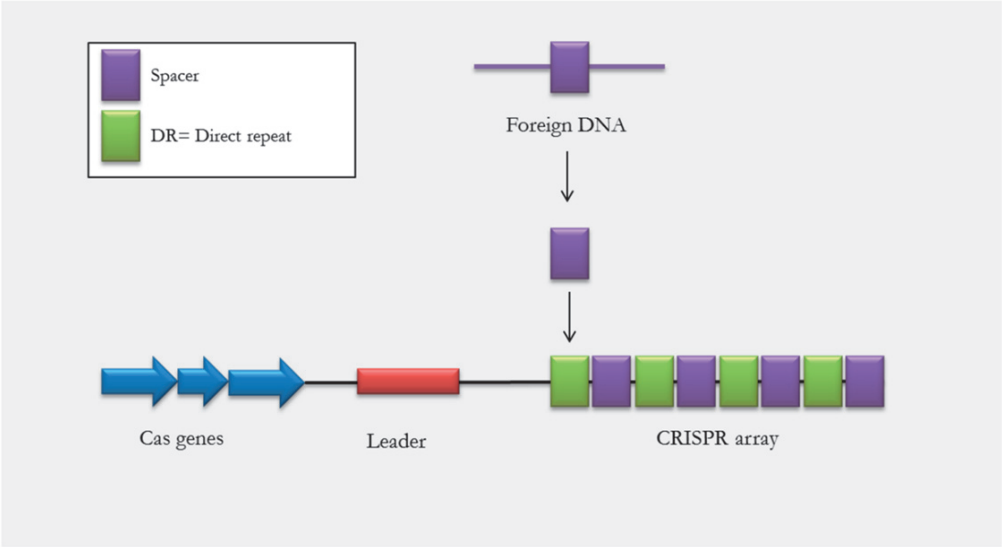


Figure 5: Acquisition of the CRISPR array - the invading foreign DNA is integrated as a spacer (normally 30–45 bp long) to the CRISPR array.

2.3.7. Sequencing and next-generation sequencing (NGS)

First generation sequencing (ie., Sanger sequencing) and especially next-generation sequencing (NGS, also known as second generation sequencing) have become strong alternatives for detection and typing of biological agents [130]. The advantage of NGS, when compared with Sanger sequencing, is its ability to generate millions of reads in a single run [131,132, 133]. The first NGS instrument was the 454 pyrosequencing device, which uses sequencing by synthesis approach and can produce approximately 700 bp long read lengths. Another NGS platform, which can produce relatively long read lengths, is Ion Torrent (400 bp) [131]. Short-read platforms, such as, Illumina MiSeq and HiSeq, are nowadays widely used. All of these sequencing platforms have a number of applications within clinical microbiology and epidemiology [134, 135].

However, relatively advanced bioinformatic skills are required to assemble the sequence reads correctly, correctly analyse NGS data, and to interpret the results [136]. Different NGS approaches are now becoming used in epidemiology [137, 138] and in the future, NGS technology will probably be widely used in routine clinical diagnostics. This technology will become more attractive as operating costs fall and improved sequencing platforms are launched.

The problem with second generation sequencing platforms is that they normally generate short sequence reads. The recently launched PacBio sequencer (also called

a third generation sequencer) has overcome this obstacle and can generate long sequence reads. The biggest disadvantage of this instrument is the high cost [139]. There is another new sequencer (MiniON) available on the market, which generates long sequence reads [140]. It is a small, hand-held instrument so it could be used in field and closer to the patient. At the moment, one significant limitation when using this instrument is its high error rate. However, improvements have been made and will continue to address this issue [141, 142].

2.3.8. Forensic investigations and typing strains using NGS

In cases where the source of the epidemic is unknown, it is good to know which strains are endemic in which countries (or areas) and what type of strains are present in different parts of the world. If the pathogen does not appear naturally in the country, the agent has likely been deliberately released. Studying and typing strains from different parts of the country and, if possible, worldwide is one approach to track the source of the epidemic.

In 2001, a comprehensive forensic investigation was conducted in the US in response to Amerithrax. The investigation included numerous physical, chemical, and genetic analyses. At that time, whole genome sequencing (WGS) was not commonly used in forensics and sequencing methods were limited at the time. During the investigation, anthrax spore genomes were compared to the *B. anthracis* Ames ancestor genome [143]. The Federal Bureau of Investigation (FBI) also collected a comprehensive Ames strain collection, altogether 1,070 isolates from domestic and international laboratories. Only eight isolates from that collection contained the four genetic mutations that were found in the strains present in the anthrax letters. These eight identical strains were related to the RMR-1029 strain and all of these eight strains were originally from the United States Army Medical Research Institute of Infectious Disease (USAMRIID). This was the same place where Dr. Ivans, allegedly responsible for this episode, had worked [17].

These WGS analyses also revealed that it was possible to identify single nucleotide changes from whole genomes. Researchers were able to detect forensic markers (i.e., fingerprints) by comparing whole-genome sequences [143].

Currently, there are many techniques based on NGS, which can be used for typing bacterial strains. WGS is one approach in addition to targeted amplification. In the latter method, only genes or parts of the gene regions are sequenced [144]. NGS data can also be used to identify SNPs and investigate, for example, the genetic diversity among bacterial isolates [116].

3. AIMS OF THE STUDY

The aim of this study was to create and investigate methods in order to detect and type biological agents.

The specific aims of this study were:

I. To develop a real-time PCR assay which can detect *V. cholerae* strains, one method for the toxigenic strains and a second for other *V. cholerae* strains, and to compare the performance of three different real-time PCR intruments.

II. To evaluate the utility of 16S rRNA gene-based methods for studying clinical samples and to investigate uncertainty in these methods.

III. To utilize the CRISPR method for typing *Y. pseudotuberculosis* strains and to study the congruence between *Y. pseudotuberculosis* and *Y. pestis* CRISPR data to potentially establish a phylogenetic relationship.

4. MATERIALS AND METHODS

Materials and methods are described in publications I-IV. An overview of the materials and methods are presented in Tables 2, 3 and 4.

4.1. Samples and data used in this study

4.1.1. Ethical permission

The ethics committee of Tampere University Hospital approved the publication II; moreover, all patients provided an informed consent. No ethical permission was needed for publication III as the Finnish Act on the Use of Animals for Experimental Purposes (62/2006) and the Finnish Animal Experiment Board (16th May, 2007) do not classify snap-trapping as an animal experiment.

4.1.2. Human samples

A total of eight carotid endarterectomy samples were studied from patients with symptomatic carotid stenosis (II).

4.1.3. Bacterial samples

A total of 87 bacterial strains were investigated in publication I. Seventy *Vibrio* strains were obtained from the Norwegian Defence Research Establishment (FFI); 63 of these were specified to be *V. cholerae*. Four *Serratia marcescens* strains were obtained from the German Collection of Microorganism and Cultures (DSMZ, Braunschweig, Germany). Additionally, one *V. cholerae* strain was provided by the National Institute for Health and Welfare (Helsinki, Finland). Twelve closely related or otherwise interesting bacterial DNA samples were included in the publication I. In publication IV, 90 *Y. pseudotuberculosis* complex strains (including 76 *Y. pseudotuberculosis*, 10 *Y. similis* and 4 *Y. wautersii* strains) were analysed here.

4.1.4. Vole samples

In publication III, 61 vole samples were studied: including, 21 field voles (*Microtus agrestis*), 37 tundra voles (*Microtus oeconomus*) and three bank voles (*Myodes glareolus*). Voles with enlarged spleens were selected for this publication because that may indicate a potential bacterial infection.

Table 2: Materials used in the thesis.

Material type		Amount of samples	Original publication
Bacteria samples	*Vibrio cholerae*	63	I
	Vibrio alginolytis	2	I
	Vibrio fischeri	1	I
	Vibrio fluvialis	1	I
	Vibrio mimicus	2	I
	Vibrio metschnikovii	1	I
	Vibrio parahaemolyticus	1	I
	Bacillus anthracis	2	I
	Bacillus cereus	1	I
	Brucella melitensis	1	I
	Campylobacter jejunii	1	I
	Campylobacter upsaliensis	1	I
	Escherichia coli (ETEC)	1	I
	Fransicella tularensis	1	I
	Listonella anguillarum	1	I
	Salmonella typhimurium	1	I
	Serratia marcescens	4	I
	Yersinia enterocolitica	1	I
	Yersinia pestis	1	I
Internal positive control (IPC)	TaqMan Exogenous Internal Positive Control		I
Negative control	DNase RNase free water		I
Positive control	*Vibrio cholerae* DNA		I
Human samples	Carotid artery tissue	8	II
Extraction control			II
Enhanced negative controls			II
Vole samples	*Microtus agrestis* (field vole)	21	III
	Microtus oeconomus (tundra vole)	37	III
	Myodes glareolus (bank vole)	3	III
Extraction controls			III
Negative PCR controls			III
Bacteria samples	*Yersinia pseudotuberculosis*	76	IV
	Yersinia similis	10	IV
	Yersinia wautersii	4	IV
Sequence data	*Yersinia pseudotuberculosis*	44	IV
	Yersinia pestis	201	IV

4.1.5. Sequence data

In publication IV, besides our own sequence data, CRISPR sequences from 40 *Y. pseudotuberculosis* and 195 *Y. pestis* strains from earlier analyses were included (Vergnaud and Gorgé, unpublished). Also CRISPR sequences were extracted from complete whole genome sequences which were publicly available (four *Y. pseudotuberculosis* and six *Y. pestis* strains). In total, 335 *Y. pseudotuberculosis* complex strains were investigated.

4.2. Bacterial culture and nucleic acid extraction

Bacteria were cultured as described in the publications (I, IV). Briefly, the *Vibrio* spp. and *S. marcescens* strains were propagated in appropriate cultivation media. Total DNA was extracted from the *Vibrio* spp. cultures using the MagNA Pure LC instrument (Roche Molecular Systems, Basel, Switzerland) and from the *Serratia* strains using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). After the extraction, DNA concentrations and purity were measured with the NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, US).

Y. pseudotuberculosis, *Y. similis* and *Y. wautersii* strains were cultivated in lysogeny broth (LB) and DNA was extracted using the JetFlex DNA isolation kit (Genomed GmdH, Löhne, Germany). Conjugation experiments were performed using *Yersinia*-selective CIN -agar. When necessary, appropriate antibiotics or selective agents were added to the agar (e.g., chloramphenicol [Clm], nalidixic acid [Nal], kanamycin [Kan], and diaminopimelic acid [Dap]).

In publication II, the carotid artery paraffin-embedded sections were treated and digested with proteinase K before using in PCR reactions.

DNA from 61 vole liver samples was extracted using the Wizard genomic DNA Purification Kit (Promega, Madison, US) following the manufacturer's protocol for animal tissue. DNA concentrations were determined with the NanoDrop® ND-1000 Spectrophotometer (publication III).

4.3. PCR Methods

Both conventional PCR and real-time PCR were used in our studies. Conventional PCR was used in publications II, III and IV and real-time PCR in publication I.

In publication I, two PCR assays were designed using Primer Express Software (Thermo Fisher Scientific, Wilmington, US). Target genes for *V. cholerae* assays were selected based on a literature search. One assay was design to detect the pathogenic

strains utilizing the *ctx* gene and the other to identify all *V. cholerae* strains using the *toxR* gene as a target. Ubiquity testing was done using 61 *V. cholerae* strains. The specificity of these assays was tested by studying closely related bacterial species (including, for example, non-cholera *Vibrio* species, *Listonella* and *Campylobacter*). Also *S. marcescens,* along with LT-toxin producing *E. coli,* were studied because it was noted based on a BLAST search, that they shared similarities with the primers used.

In publications II and III, previously published primers targeting the 16S rRNA gene region were used. In publication IV, published and designed primer pairs targeting *Yersinia* CRISPR regions were utilized. Primers sequences used in the thesis are presented in Table 3.

Table 3: Primers used in the thesis.

Publication	Target	Primer	Nucleotide sequence (5'-3')	Reference
I	*toxR*	*toxR_F*	TGGCATCGTTAGGGTTAGCAA	I
		toxR_R	CATTCACAGCCCTGAAGTTTCA	I
		toxR_Probe	FAM-CGTAAGGTTATGTTTTCC-MGBNFQ	I
	ctxA	*ctxA_F*	ACTCACTCTGTCCTCTTGGCATAA	I
		ctxA_R	GCAGATTCTAGACCTCCTGATGAAAT	I
		ctxA_Probe	FAM-ACCACCTGACTGCTT-MGBNFQ	I
II+III	*16S rRNA*	*fD1mod*	AGAGTTTGATCYTGGYTYAG	Kotilainen *et al.,*1998
		16S1RR-1	CTTTACGCCCARTRAWTCCG	Wilbrink *et al.,* 1998
III	*16S rRNA*	*pA*	AGAGTTTGATCMTGGCTCAG*	Lane DJ 1991
		*pD**	GTATTACCGCGGCTGCTG*	Edwards *et al.,* 1989
IV	*YP1*	*YP1_F*	AATTTTGCTCCCCAAATAGCAT	Le Flèche *et al.*, 2001
		YP1_R	TTTTCCCCATTAGCGAAATAAGTA	Pourcel *et al.,* 2004
	YP2	*YP2_F*	ATATCCTGCTTACCGAGGGT	Pourcel *et al.,* 2005
		YP2_R	AATCAGCCACGCTCTGTCTA	Pourcel *et al.,* 2005
		CRISP_YP2_F	GAACCCTAAAAACGAAGCTATG	IV
		CRISP_YP2_R	CCCGGCAGATTGTCATTAC	IV
	YP3	*YP3_F*	GCCAAGGGATTAGTGAGTTAA	Pourcel *et al.,* 2005
		YP3_R	TTTACGCATTTTGCGCCATTG	Pourcel *et al.,* 2005
	pYptb32953	*pIP32-F*	GGCGAATTCTTTACCGTTCCCTCAAATCC	IV
		pIP32-R	GGCGAATTCTGCCGCTAACTTCTATGCAA	IV
		pIP32-F2	GTGGGACGCGCTAAAGATTA	IV
		pIP32-R2	CCGTCTAGCGGTAATTTGGA	IV

* + barcode sequence

4.4. PCR controls

Different PCR controls were used to check that the PCR results were reliable. In every PCR run, a negative and a positive control were used; negative controls being DNase and RNase free water (Sigma-Aldrich, Saint Lous, Missouri, USA) or negative extraction control and the positive control being the target DNA of the assay.

Internal positive control (IPC) (TaqMan® Exogenous Internal positive control, Thermo Fisher Scientific, Wilmington, US) was used in the real-time PCR runs to check that inhibitors do not disturb the PCR reaction (Table 2). Unfortunately, the RAZOR instrument can detect only one wavelength and thus does not allow the use of IPC.

4.5. Instruments

In publication I, all three real-time PCR instruments employed the TaqMan chemistry, and identical primers and probes were used in each. The instruments used were the Applied Biosystems (ABi) 7300, the ABi 7900HT Fast (Thermo Fisher Scientific, Wilmington, US) and the RAZOR Instrument (BioFire Defens, Salt Lake City, US). A conventional PCR instrument (DNA Engine (PTC-200) Peltier Thermal Cycler (Bio-Rad Life Sciences, Hercules, US)) was used in publications II, III and IV.

Table 4: Methods, instruments and programs used in the thesis.

Method	Instrument or program	Publication
Real-time PCR	Applied Biosystems® 7300 Real-Time PCR Systems (Thermo Fisher Scientific)	I
	Applied Biosystems® 7900HT Real-Time PCR Systems (Thermo Fisher Scientific)	I
	Razor® Instrument (BioFire Defence)	I
DNA extraction	MagNA Pure LC instrument (Roche)	I
Convertional PCR	DNA Engine (PTC-200) Peltier Thermal Cycler (Bio-Rad Life Science)	II, III, IV
Sanger sequencing	3730xlDNA Analyser or PRISM 3100 Genetic Analyzer or 3100xl Capillary Sequencer Analyzer (Applied Biosystems)	II, III, IV
Pyrosequencing	454 GS FLX (Roche)	III
Programs		
Real-time PCR	Primer ExpressSoftware	I
	7500 System Sequence Detection Software	I
	RAZOR Software	I
Sanger sequencing	ARB Software	II
	Chromas	II
	RDP Classifier	III
	Sequencher	III, IV
	Staden Package	IV
Pyrosequencing	Mothur	III
	RDP Classifier	III

4.6. Methods based on the 16S rRNA gene

4.6.1. Broad-range PCR

In publication II, eight patient samples were analysed with conventional broad-range PCR (Br-PCR). This method was also used in publication III to study vole liver samples.

Briefly, part of the 16S rRNA gene was amplified with PCR, the amplicon was then cloned into a vector and the vector transferred to *Escherichia coli* using the TOPO® TA Cloning® Kit (Invitrogen Corporation, Carlsbad, US). Next, colony PCR was performed from the *E. coli* colonies. PCR products of the correct size (approximately 550bp; in *E. coli* 8F-575R) were purified and sequenced at the University of Helsinki (Sequencing Core Facility) using the 3100xl Capillary Sequence Analyzer (Thermo Fisher Scientific, Wilmington, US). Sequence data were analysed with Se-

quencher 5.1 (Gene Codes Corporation, Ann Arbor, US) and sequence homology was determined with the BLAST tool provided by GenBank (Nucleotide collection [nr/nt]). In publication III, sequence data was also analysed using the Ribosomal Database Project (RDP) Classifier.

The reliability of Br-PCR was evaluated by creating contaminant sequence libraries using only DNase and RNase free water or extraction controls as a template. Cloning, sequencing, and analysing the sequence data was done as described earlier.

In publication II, ten individual colonies were sequenced from each patient sample, as well as 34 colonies from the enhanced negative controls. Sequence-types sharing over 99% similarity with contaminant sequences (originating from negative controls) were omitted from the final results.

In publication III, two to thirteen resulting amplicons from each vole sample clone library were sequenced. In addition, two commercial PCR kits (AmpliTaq Gold® DNA Polymerase LD (Thermo Fisher Scientific, Wilmington, US) and DyNAzyme I DNA Polymerase (Thermo Fisher Scientific, Wilmington, US)) were used in the PCR reactions to create a contaminant library consisting of 529 contaminant sequence-types. In this publication, an additional 27 contaminant clones originating from the three different DNA extraction controls were sequenced.

4.6.2. Pyrosequencing (454 Sequencing)

In publication III, the presence of bacterial DNA in 61 vole liver samples was studied employing two methods based on the 16S rRNA gene: conventional Br-PCR and pyrosequencing. Also, pooled DNA extraction control was analysed using pyrosequencing.

Conventional Br-PCR was performed as described above. In pyrosequencing, DNA was amplified using barcoded primers in PCR reactions which cover the V1-V3 region (in *E. coli* 27F-518R) (Table 3). After purification of the PCR products, twelve samples were sequenced in a single run following the 454-GS FLX titanium protocol (Roche, Branford, US). Pyrosequencing and part of the sequence data analysis was conducted using the RDP Classifier at the Institute of Biotechnology at the University of Helsinki (UH). The more accurate data analysis was completed at the Department of Virology (UH) using Mothur according to the work-flow described by Scloss and colleagues [145].

4.7. CRISPR-based typing

Three previously recognized *Yersinia pestis* CRISPR loci (YP1, YP2 and YP3) [146, 147] were used to study *Y. pseudotuberculosis* complex strains. PCR reactions were performed as described in publication IV. Sequencing was performed at the University of Helsinki (Sequencing Core Facility) using the 3100xl Capillary Sequence Analyzer. Sequence data were analysed using Sequencher 5.1 or the Staden Package [148] and sequences were further studied with the CRISPRFinder (http://crispr.u-psud.fr/Server/CRISPRfinder.php) and BLAST search tools.

5. RESULTS

5.1. *Vibrio cholerae* detection (I)

In publication I, two real-time PCR assays were established to detect *V. cholerae* strains. The specificity was 100% with all three instruments used; all 63 *V. cholerae* strains were detected with the designed *toxR* assay, while only the toxigenic *V. cholerae* strains gave a positive signal when using the *ctxA* assay. In addition, none of the specificity panel strains, including closely related bacterial species, were amplified.

The limit of detection for all of the instruments, when using pure DNA as a template was 100 fg per reaction (roughly 20 bacterial cells). The duration of the PCR runs and volumes of PCR mixtures needed for two parallel reactions with the ABi 7300, the ABi 7900HT Fast, and the RAZOR instruments were 100, 35, and 43 minutes, and 50, 20 and, 300µl, respectively. All in all, the developed assays were proven to be specific and fast for detecting *V. cholerae* stains

5.2. Utility of the 16S rRNA method (II, III)

The aims of the two publications were to observe the utility of 16S rRNA gene-based methods for studying bacterial samples and to evaluate the reliability of these methods.

In publication II, eight patient samples (carotid artery tissue) were studied using Br-PCR. From three patient samples all ten sequence-types were omitted because they either shared a high similarity with negative control sequences or were plastid/non-bacterial sequences. The remaining 23 sequence-types of five patients represented mainly *Micrococcus, Dolosigranulum, Corynebacterium, Bradyrhizobium, Brachybacterium, Propionibacterium, Rhodococcus*, and *Stenotrophomonas*. The main contaminant sequence-types present in the DNA extraction control were *Corynebacterium*, *Actinomyces*, *Streptococcus mitis*, *Microbacterium backeri* and *Rothia*.

In publication III, the splenomegalic voles investigated carried microbes belonging to their normal flora and a variety of pathogens or potential pathogens (Table 5). The main bacterial findings, using pyrosequencing, belonged to the families Francisellaceae, Bartonellaceae, Anaplasmataceae, Lachnospiraceae, Ruminococcaceae, Porphyromonadaceae, and Mycoplasmataceae. The same bacterial sequence-types were also detected when using the more conventional Br-PCR method, but a wider range of bacterial species were identified with pyrosequencing (Table 5).

In publication III, three different contaminant libraries were established. Using Br-PCR, the primary contaminant sequence-types detected from three negative extraction controls were *Propionibacterium acnes, Acidovorax, Dechloromonas, Escherihia coli* and *Bacteroides*. The other contaminant library using conventional Br-PCR was created using two different PCR kits and the main contaminant sequences found were uncultured *Burkholderia, Lactococcus, Phyllobacterium*, *Schlegelella, Sulfurospirillum*, uncultured Ruminococcaceae and *Propionibacterium.* When investigating the pooled negative extraction control using pyrosequencing, the main bacterial families identified were Pseudomonadaceae, Enterobacteriaceae, Bacteroidaceae and Enterococcaceae.

Table 5: Main Br-PCR and pyrosequencing results of vole liver samples.

Vole ID	Br-PCR findings	Pyrosequencing findings [1]
1 [2]	Lachnospiraceae, Ruminococcaceae	Lachnospiraceae, Porphyromonadaceae, Ruminococcaceae, Cytophagaceae, Coriobacteriaceae, Prevotellaceae
2	Lachnospiraceae, Ruminococcaceae	Lachnospiraceae, Ruminococcaceae, Porphyromonadaceae, Prevotellaceae, Cytophagaceae
3	Lachnospiraceae, *Lactobacillus*, Coriobacteriaceae	Lachnospiraceae, Ruminococcaceae, Porphyromonadaceae, Prevotellaceae, Cytophagaceae
4	Lachnospiraceae, Ruminococcaceae, *Helicobacter*	Lachnospiraceae, Ruminococcaceae, Porphyromonadaceae, Cytophagaceae
5	Lachnospiraceae	Lachnospiraceae, Ruminococcaceae, Porphyromonadaceae, Cytophagaceae, Mycoplasmataceae
6	*Mycoplasma, Pseudomonas*	Mycoplasmataceae, Ruminococcaceae, Campylobacteraceae, Prevotellaceae
7	*Mycoplasma,* Ruminococcaceae	Mycoplasmataceae, Campylobacteraceae, Propionibacteriaceae
8	*Mycoplasma*	Mycoplasmataceae, Prevotellaceae
9	Unclassified [2]	Mycoplasmataceae, Carnobacteriaceae
10	*Bartonella, Mycoplasma*	Bartonellaceae, Mycoplasmataceae
11	*Bartonella*	Bartonellaceae, Mycoplasmataceae
12	*Bartonella*	Bartonellaceae, Comamonadaceae
13	*Bartonella*	Bartonellaceae, Ruminococcaceae
14	*Corynebacterium*	Bartonellaceae, Ruminococcaceae
15	*Bartonella*	Bartonellaceae
16	*Bartonella, Anaplasma, Acinetobacter*	Bartonellaceae
17	*Bartonella, Anaplasma*	Bartonellaceae, Anaplasmataceaea
18	*Anaplasma*	Anaplasmataceae
19	*Francisella, Anaplasma, Mycoplasma, Lactobacillus*	Francisellaceae, Anaplasmataceae, Ruminococcaceae
20	*Francisella*	Francisellaceae
21	*Francisella*	Francisellaceae
22	*Bartonella*	Corynebacteriaceae, Moraxellaceae
23	Unclassified [3]	Unclassified bacteria

1) Sequence-types which exceed the relative abundance of 5% in the sample
2) Control vole; no splenomegaly
3) With BLAST search tool: *Mycoplasma* [93 %]

5.3. *Yersinia pseudotuberculosis* typing with CRISPR method (IV)

In publication IV, *Y. pseudotuberculosis* complex strains (including *Y. pseudotuberculosis, Y. similis* and *Y. wautersii*) using CRISPRs elements (YP1, YP2 and YP3) were typed and compared to *Y. pestis* CRISPR sequences.

There were some difficulties in sequencing the whole CRISPR locus and amplifying YP1, YP2 or YP3 loci from all of the strains. Some of the sequences were as long as 3000 bp and it was not possible to sequence those in a single run using Sanger sequencing. To overcome this, internal primers were designed for the longer fragments.

Of the 90 strains studied, 60 could be amplified using the YP1 PCR; no PCR product was obtained for 20 *Y. pseudotuberculosis* or ten *Y. similis* strains. With YP2 PCR, 61 strains were amplified and 19 failed to yield PCR product. In addition, 85 strains were amplified with the YP3 PCR. However, due to difficulties experienced in sequencing, only partial sequences or no sequence at all were obtained from some of the strains.

The sequence data from 335 *Y. pseudotuberculosis* complex strains (including those earlier sequenced by Vergnaud and Gorgé and others available in public database) were analysed. Altogether, more than 6,000 spacer sequences representing 1,902 different spacers were obtained. It was surprising how much the spacer composition varied among different strains, apart from over 1,000 spacers being unique to a single strain. When the spacers present in *Y. pseudotuberculosis, Y. similis,* and *Y. wautersii* strains were aligned manually, some strain-specific patterns were observed. As an example, three spacer arrangement patterns are displayed in Figure 6, where clear similarities between the spacer compositions of ST42 and ST43 strains can be seen. The amount of different spacers in strains also varied considerably. For instance, the YP1 locus could contain two (e.g., strain Y80) to forty spacers (e.g., IP32884) (Figure 6).

When comparing *Y. pseudotuberculosis, Y. similis,* and *Y. wautersii* spacers to *Y. pestis* spacers, very little overlap was observed and only 33 spacers showed a significant similarity to those of *Y. pestis*. Also, strains of *Y. pestis* contained significantly fewer spacers than those of *Y. pseudotuberculosis*.

A BLAST search was performed to determine the possible origins of the obtained spacers, but about 90% failed to provide any significant matches. However, some spacers had similarities with plasmids and bacteriophages. For example, 31 spacers shared a similarity with the cryptic 27,702 bp plasmid of *Y. pseudotuberculosis* strain IP32953 and twenty of the spacers had similarities to different bacteriophages. In

addition, 40 spacers were >88% similar to strain 8081 of *Y. enterocolitica* and a single *Y. pseudotuberculosis* strain could carry several of the foregoing spacers.

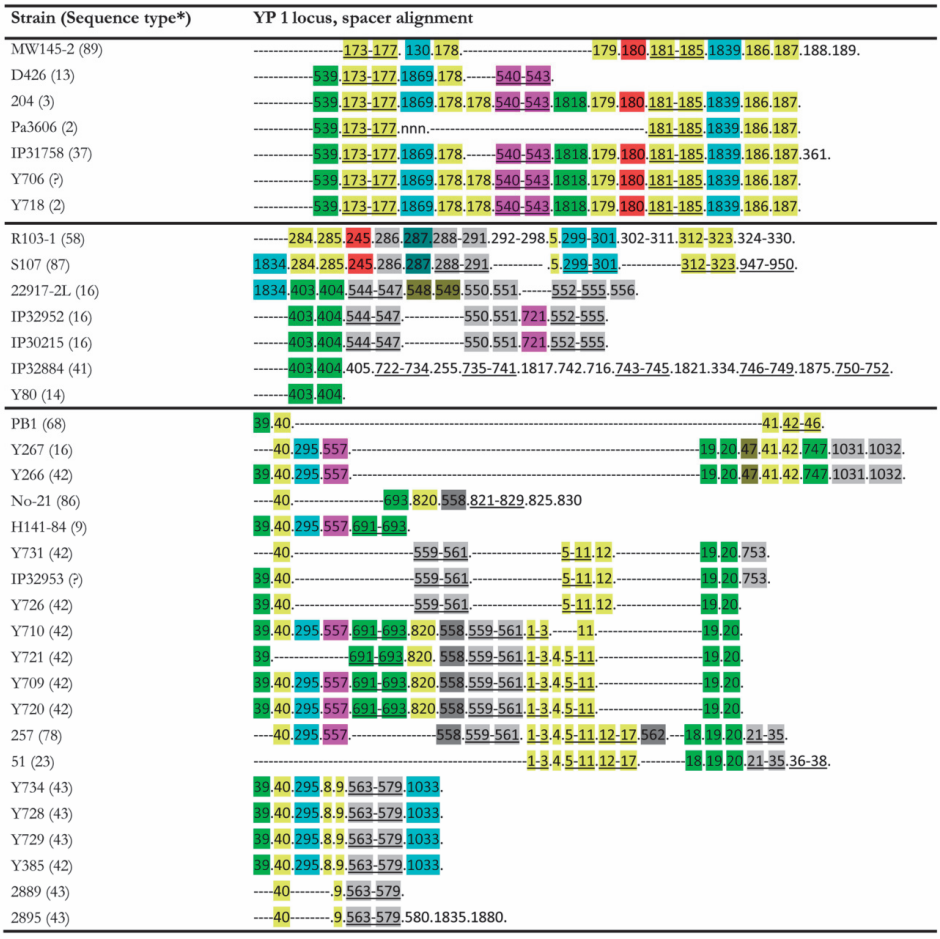


Figure 6: CRISPR YP1 spacer alignments of selected *Y. pseudotuberculosis* strains. Each number in the figure highlighted with colour represents different spacers. Gaps have been introduced between spacers to maximize the alignment. *Multilocus sequence type (ST) according to MLST Databases at the ERI, University College Cork (http://mlst.ucc.ie/mlst/dbs/Ypseudotuberculosis/GetTableInfo_html).

6. DISCUSSION

6.1. Detecting biological agents using molecular methods

The real-time PCR assays developed here proved to be specific and reliable tools for detecting *V. cholerae* strains and, moreover, distinguishing the pathogenic strains from non-pathogenic ones. Other groups have also published real-time PCR assays to detect and discriminate *Vibrio* strains [149-151]. Currently, detection of *Vibrio* strains is mainly conducted using different PCR methods, but NGS approaches have also been used [152].

PCR and real-time PCR have been used successfully to detect other biothreat agents as well. Many research groups have published PCR methods to identify potential biothreat agents, such as *F. tularensis*, *B. anthracis*, *Y. pestis, Brucella* spp. and Ebola virus [112, 114, 115, 153, 154]. Likewise, multiplex real-time PCR detection methods have been set up to identify biothreat agents. For instance, Woubit and colleagues [155] developed a real-time PCR assay using twelve primer pairs to detect *F. tularensis* ssp. *tularensis*, *V. cholerae, Y. pestis, E.coli* O157:H7*, Shigella dysentriae* and *Salmonella typhi* in a single run. Also Janse and colleagues [113] designed a multiplex PCR to simultaneously detect *B. anthracis*, *F. tularensis,* and *Y. pestis* using nine primer pairs.

In this thesis, three real-time PCR instruments were used. The robust and field-deployable RAZOR instrument delivered as reliable results as the other tested instruments. However, the disadvantage of the RAZOR instrument is that it required six-to-fifteen times more PCR mixture (300 µl for two reactions) than the ABi 7300 instrument (50 µl for two reactions) or the ABi FAST instrument (20 µl per two reactions), respectively, which increases the costs considerably. Currently, there are other field-deployable PCR instruments available and their reaction volume can be as little as 10 µl. These kinds of field-deployable solutions can be used in field or close to the patient, which could be beneficial during epidemics and outbreaks [156].

No differences were seen in the detection limits of the PCR instruments. The DNA used in this study was extracted from pure cultures and the quality and concentration were measured. Thus, the detection limit represents the lowest amount of purified target DNA that can be detected in the PCR reaction itself and is affected by, for example, the number of cycles and mastermix used in the reaction. The detection limit of a PCR reaction can also vary, for instance, depending on the number of copies of the target gene in the genome. If the concentration of the target DNA in the purified sample is very low, a higher volume of template in the reaction may improve the detection limit. However, the sample may also contain inhibitors, and in those cases detection is conversely improved by diluting the template.

Shipley and colleagues [157] compared two extraction methods for studying select biothreat agents (*B. anthracis, Y. pestis*, vaccinia virus and VEEV). They used two different matrices (buffer and blood) and performed DNA extraction of serially-diluted samples. Extraction efficiency was evaluated with real-time PCR. Samples with 10 CFU/ml of bacteria (*B. anthracis* and *Y. pestis)* or 500 PFU/ml and 1000 PFU/ml of vaccinia virus and VEEV, gave positive signals, respectively. As their study also show, the detection limit of a given target can vary depending on the sample matrix. To obtain good-quality DNA from a complex sample, appropriate sample preparation and DNA extraction methods must be used. If the DNA quality is poor, it can impair the detection limit or even result in a false negative for positive samples. With challenging sample matrices such as soil, blood or unkown powder, it is necessary to employ an appropriate and efficient DNA extraction method. Depending on the matrix, it can be essential to perform additional degradation and purification steps to obtain pure enough DNA.

Reliable and rapid detection are key elements of biothreat preparedness. Early detection and identification of an agent is critical for the management of an outbreak, whether natural or intentional. DNA extraction and real-time PCR are becoming faster and it is now possible to obtain results within hours, which enables swift medical countermeasures. In this thesis, results were ready within 43 minutes with the RAZOR instrument. The ABi 7900HT Fast instrument performs even faster (35 mins), but it is not field-deployable and must be installed either in a stationary or mobile laboratory setting. After this paper was published in 2009, novel PCR reagents and instruments have made it possible to run the PCR even faster.

There are also simple automated real-time PCR instruments available that perform sample preparation and PCR reaction (for example GeneXpert, Cepheid, Sunnyvale, US). In the future, these kinds of easy-to-use and rapid instruments will likely become more common, but at the moment, the price per reaction is still rather high when compared to in-house real-time PCR methods. However, establishing a new in-house method requires more accurate validation and testing, which means more resources (e.g. work hours and money) before the test is ready-to-use. Currently, commercial PCR kits are available to detect different pathogens, and these have made it easier to get reliable assays without the need for extensive validation and testing.

There are many molecular methods available to detect biological agents. One is PCR-based methods, but partial and whole genome sequencing is also becoming common, especially with pathogens that are difficult to identify with conventional methods or challenging to cultivate. Other alternative amplification tehnologies are available, such as isothermal amplification, but PCR is still the most widely-used for detection and identification of pathogens.

6.2. Detecting biothreat agents with 16S rRNA gene-based methods

In this thesis, the 16S rRNA gene-based detection methods were used to study the bacterial DNA present in carotid artery tissue samples and livers of splenomegalic voles. The conventional Br-PCR method was used in both the studies, whereas pyrosequencing was applied only on vole samples. The aims were to evaluate the utility of these methods (i.e., Br-PCR and pyrosequencing) and compare them.

The sequence-types present in carotid artery samples (*Micrococcus, Dolosigranulum, Corynebacterium, Bradyrhizobium, Brachybacterium, Propionibacterium, Rhodococcus*, and *Stenotrophomonas)* discovered using the Br-PCR method, may play a role in the etiology of atherosclerosis. At that time, only 10 colonies per sample were sequenced. To obtain more detailed information about these samples, more colonies should be sequenced or amplicon sequencing or other NGS methods utilized. The main clinically-relevant bacterial sequence-types detected from vole samples with Br-PCR and pyrosequencing (*Mycoplasma, Bartonella, Anaplasma,* and *Fransicella*) have already been found in voles [126, 158-161]. However, this was the first time that this kind of metagenomic study on voles was carried out in Finland. Splenomegalic voles were studied because spleen enlargement may indicate a bacteremia. Other causes may include a parasitic infection or a hormonal imbalance.

A wider range of bacterial species was identified using pyrosequencing. This is because more sequence-types (112 to 24,246 sequences) were obtained than with the more conventional Br-PCR method (two to 15). In addition, the Br-PCR is more laborious and time consuming. Nevertheless, both 16S rRNA sequencing methods gave similar results. Both were able to detect the bacterial sequence-types present in different matrices, thus they could be used for studying unknown, unculturable and polymicrobial samples. Earlier studies support this conclusion. 16S rRNA gene-based methods have been successfully used when studying clinical, environmental and biological threat agents [122-126].

In this thesis, only a part of the 16S rRNA gene was utilized: in the more conventional Br-PCR method approximately a 550bp long region (in *E. coli* 8F-575R) and with pyrosequencing the V1-V3 region (in *E. coli* 27F-518R). However, different regions or even the entire gene (about 1550bp) can be analysed. As expected, when using the whole gene, differentiation of the organism is more accurate and, for example, enables *B. anthracis* to be distinguished from other spore-forming *Bacillus* strains [162]. However, sometimes it is not necessary to sequence the whole gene to obtain informative results [163]. Furthermore, selection of 16S rRNA regions and primers can affect the results; particularly when using short sequence reads [164].

16S rRNA gene-based methods can be applied in various situations and offer an alternative to more conventional techiques. For example, the identification of some rare dangerous pathogens (e.g., *Francisella*, *Burkholderia* and *Brucella*) can be challenging for a typical clinical laboratory due to their rare occurrence. During an investigation of powder letters, 16S rRNA gene-based methods can yied information about the bacterial diversity in the sample and consequently reveal if it contains pathogenic agents. Another appealing feature of these methods is that uncommon, unexpected, or unculturable bacteria can be detected and profiled even when the total bacterial load in the sample is low.

Despite being very useful in many applications, there are drawbacks to using 16S rRNA gene-based methods. The fact is that plastic ware [165, 166], polymerases, and PCR reagents are contaminated by exogenous DNA [167-171]. Reagents may become contaminated as many PCR enzymes originate from bacteria. Nucleic acid residues may contaminate reagents, plastic-ware, or water during manufacturing. In addition, nucleic acid extraction kits are another important source of exogenous DNA [172, 173].

Exogenous DNA may be problematic, especially with samples containing few bacteria [168, 174], as it can lead to false positive results if insufficient precautions are taken. Therefore, when using 16S rRNA gene-based methods, it is essential to recognize possible contaminants. This can be done by using water and/or a negative DNA extraction control as PCR template. For example, in a few model studies, a contaminant clone library has been built enabling the elimination of contaminant sequences from the final results [175, 176]. Unfortunately, there are still a lot of studies neglecting this and, as a consequence, creating a risk of publishing untrue findings. In numerous studies, even the most common contaminating sequences have not been eliminated. Nowadays, NGS has made it possible to sequence amplicons without the need to clone them. Nevertheless, the contamination issue exists when PCR products are sequenced using NGS [177] and this must be taken into account when using modern technologies.

During an epidemiological outbreak or biological attack, it is crucial to first determine the bacterial species and then the specific strain involved. These kinds of broad bacterial detection methods can help determine which bacterial species are present in the sample. Consequently, more accurate detection and typing methods can be utilized to investigate the sample more thoroughly if necessary.

6.3. Typing biological agents

In the event of an outbreak or biological attack, typing of bacterial isolates is crucial as was seen in the 2001 case of anthrax letters being sent to media and government offices in the US. There are many typing methods used to characterize bacterial isolates, such as pulse-field gel electrophoresis (PFGE), VNTR, MLVA, MLST. In recent years, NGS approaches have also become valuable tools for epidemiological and forensic investigations. It seems likely they will become more widely used for routine diagnostics in the near future.

There have been different approaches applied to type *Yersinia* species, but typing remains challenging due to the close relationships among species. Previous studies have shown that CRISPR typing is a promising tool for the differentiation of bacterial isolates. For example, Bachmann and colleagues [178] used CRISPR regions to type *Salmonella enterica* serovars. *Staphylococcus aureus* has also been typed based on CRISPR spacer sequences [179]. In this thesis, CRISPR sequences were used for typing *Y. pseudotuberculosis* complex strains because this method has earlier proven to be useful with *Y. pestis*. Furthermore, phylogenetic relationships among *Y. pseudotuberculosis* and *Y. pestis* strains were studied by comparing CRISPR spacers.

Due to the high diversity of spacers in *Y. pseudotuberculosis* complex strains, the CRISPR method can, in fact, be useful for typing *Y. pseudotuberculosis* strains. However, it would require building an extensive reference database. In this thesis, when *Y. pseudotuberculosis, Y. similis,* and *Y. wautersii* spacers were compared to those of *Y. pestis*, only a few spacers showed similarity to *Y. pestis* and, surprisingly, *Y. pestis* carry fewer spacers than *Y. pseudotuberculosis*. This may be due to different living conditions experienced by these species, or that *Y. pestis* does not integrate spacers so readily as *Y. pseudotuberculosis*. Unfortunately, results obtained in this thesis, the comparison of CRISPR spacers did not provide any futher resolution of the phylogenetic relationships among strains.

For the long CRISPR sequences (up to 3,000 bp), internal primers were designed to obtain the whole sequence. However, this solution did not always result in full-length CRIPSR sequence. CRISRP arrays are composed of approximately 30 bp repeats: direct repeats (DRs), which are entirely or almost identical with each other, and unique spacers. Thus, internal primers had to be designed within the short spacer region. As such, in some cases it was challenging to design optimal primers and thus the primers did not always work as planned.

Our research group has used NGS to sequence *Y. pseudotuberculosis* strains and it has proven to be a practical method to fill in the CRISPR sequencing gaps. In our recent study, more than one hundred *Y. pseudotuberculosis* serotype O1 strains were sequenced. Preliminary analyses reveal clear CRISPR patterns and clustering of the

strains has been possible (unpublished). Other researchers have also used NGS data and filtered the CRISPR regions for data analyses [178, 180]. NGS-based CRISPR typing does not require long hands-on-time and published genome sequences can be explored, but it does require computing expertise and advanced bioinformatic skills.

Overall, different typing methods have evolved dramatically over recent years. For example, NGS combined with SNP analysis has made it possible to distinguish and cluster isolates based on which country or region they can be found in. These kinds of analyses are essential in tracking the source of a natural outbreak or origin of agents used in a bioterrorism attack.

7. CONCLUDING REMARKS

Cholera continues to be a significant threat to human health, not only as a naturally-occurring disease but also as a possible bioweapon. Fast and reliable diagnostic methods are invaluable, especially in the case of intentional release in geographical areas or drinking water supplies, where presence of the bacteria is not expected.

In this study, PCR assays for the identification of *V. cholerae* were developed, which can readily be transferred among different PCR platforms while maintaining sufficient sensitivity and specificity. Developed assays could also be transfered to a field-deployable instrument and employed by first responders or used in a mobile laboratory. Conventional and real-time PCR have proven to be suitable methods for the detection of different biological agents.

Two methods, conventional Br-PCR and pyrosequencing of the 16S rRNA gene were found to be convenient approaches for the detection of bacterial DNA present in different matrices and which could be used to investigate for example powder letters or unknown bacterial samples. In addition, the two different methods gave similar results which increase the reliability of the results.

Typing of bacterial strains is very important when the origin of the bacterial source is unknown. There are different approaches to type bacterial strains, of which the CRISPR sequence-based typing method is a relatively new one employed in this thesis. According to the experience gained from studying *Y. pseudotuberculosis*, it can be concluded that this method can be used for *Y. pseudotuberculosis* typing, but it would require building a comprehensive reference database.

In recent years, different NGS-based typing approaches have been developed. These methods are becoming valuable tools when investigating and typing biothreat agents. NGS approaches can also be utilized in epidemiological investigation and source tracking.

Nowadays we are more prepared to detect biological agents but there is still a lot of work to do. Biological preparedness and co-operation among national and international agencies are critical. Different molecular methods to detect and type biological agents have evolved during recent years to become faster and more reliable.

8. ACKNOWLEDGEMENTS

This study was carried out at the Centre for Military Medicine, Research and Development Department. I have been very fortunate to do this work in such an encouraging atmosphere.

Foremost, I would like to express my gratitude to my supervisor Professor Simo Nikkari. I thank you for encouraging me to start this project and for the effort you have put into this thesis. I sincerely thank my other supervisor, Professor Mikael Skurnik for guidance and support. You have always found time to give me invaluable advice and guidance throughout these years.

I would like to express my sincere gratitude to the reviewers Kaisu Rantakokko-Jalava and Susanna Lukinmaa-Åberg for their valuable and constructive comments in reviewing and improving the thesis. I also thank Florent Sebbane, who has kindly agreed to be the opponent in the public examination of the thesis.

I also want to thank all the co-authors; I could not have done this without you. Your input to the publications and your help and support during this process has been crucial.

I have been privileged to have the opportunity to work with an acknowledged and wonderful working team. It has been a joy to work with you in a friendly and supportive work environment. Especially, I am deeply grateful to Heidi and Laura for your guidance, encouragement and friendship. I warmly thank Kari I, Anna, Elina, Liina, Markos, Heli and Kari K for being wonderful workmates and for your advices and support. I also want to thank all my former fellow workmates in the Centre for Military Medicine for the fruitful cooperation and for sharing the memorable daily working life with me during these years.

I wish to express my warmest gratitude to my family and friends. You have brought joy and laughter to my life and shared the ups and downs with me. It is always good to spend some time with you to clear the mind and taking thoughts out off work.

My parents and siblings are also warmly thanked for always supporting and believing me. Finally, I want to express my heartfelt thanks to Jari for your flexibility, love, and support, and for reminding me that there is a life outside work.

Helsinki, April 2017

Katja Koskela

9. REFERENCES

1. Block S. The Growing Threat of Biological Weapons. American Scientist 2001;89:28-37.

2. Bellamy RJ, Freedman AR. Bioterrorism. QJM 2001;94(4):227-34.

3. Jansen HJ, Breeveld FJ, Stijnis C, Grobusch MP. Biological warfare, bioterrorism, and biocrime. Clinical Microbiology and Infection 2014;20(6):488-96.

4. Chirisa I, Nyamadzawo L, Bandauko E, Mutsindikwa N. The 2008/2009 cholera outbreak in Harare, Zimbabwe: case of failure in urban environmental health and planning. Rev Environ Health 2015;30(2):117-24.

5. Al-Tawfiq JA, Omrani AS, Memish ZA. Middle East respiratory syndrome coronavirus: current situation and travel-associated concerns. Front Med 2016;

6. Lai S, Qin Y, Cowling BJ, Ren X, Wardrop NA, Gilbert M et al. Global epidemiology of avian influenza A H5N1 virus infection in humans, 1997-2015: a systematic review of individual case data. Lancet Infect Dis 2016;

7. Barras V, Greub G. History of biological warfare and bioterrorism. Clinical Microbiology and Infection 2014;20(6):497-502.

8. Ligon BL. Plague: A Review of its History and Potential as a Biological Weapon. Seminars in Pediatric Infectious Diseases 2006;17(3):161-70.

9. Riedel S. Biological warfare and bioterrorism: a historical review. Proc (Bayl Univ Med Cent) 2004;17(4):400-6.

10. Kenyon IR, Gutschmidt K, Cosivi O. Legal aspects and international assistance related to the deliberate use of chemicals to cause harm. Toxicology 2005;214(3):249-55.

11. Roffey R, Tegnell A, Elgh F. Biological warfare in a historical perspective. Clin Microbiol Infect 2002;8(8):450-4.

12. Frischknecht F. The history of biological warfare. Human experimentation, modern nightmares and lone madmen in the twentieth century. EMBO Rep 2003;4 Spec No:S47-52.

13. Domaradskij IV, Orent LW. Achievements of the soviet biological weapons programme and implications for the future. Rev - Off Int Epizoot 2006;25(1):153-61.

14. Uzumcu A. The Chemical Weapons Convention--disarmament, science and technology. Anal Bioanal Chem 2014;406(21):5071-3.

15. Keim P, Smith KL, Keys C, Takahashi H, Kurata T, Kaufmann A. Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan. J Clin Microbiol 2001;39(12):4566-7.

16. Jernigan DB, Raghunathan PL, Bell BP, Brechner R, Bresnitz EA, Butler JC et al. Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. Emerg Infect Dis 2002;8(10):1019-28.

17. Committee on Review of the Scientific Approaches Used During the FBI's Investigation of the 2001 *Bacillus Anthracis* Mailings; National Research Council. Review of the Scientific Approaches Used during the FBI's Investigation of the 2001 Anthrax Letters. 2011; by the National Academy of Sciences, Washington (DC)

18. Crighton T, Hoile R, Coleman NV. Comparison of quantitative PCR and culture-based methods for evaluating dispersal of *Bacillus thuringiensis* endospores at a bioterrorism hoax crime scene. Forensic Sci Int 2012;219(1-3):88-95.

19. Dybwad M, van der Laaken AL, Blatny JM, Paauw A. Rapid identification of *Bacillus anthracis* spores in suspicious powder samples by using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS). Appl Environ Microbiol 2013;79(17):5372-83.

20. Abramova FA, Grinberg LM, Yampolskaya OV, Walker DH. Pathology of inhalational anthrax in 42 cases from the Sverdlovsk outbreak of 1979. Proc Natl Acad Sci U S A 1993;90(6):2291-4.

21. Jackson PJ, Hugh-Jones ME, Adair DM, Green G, Hill KK, Kuske CR et al. PCR analysis of tissue samples from the 1979 Sverdlovsk anthrax victims: the presence of multiple *Bacillus anthracis* strains in different victims. Proc Natl Acad Sci U S A 1998;95(3):1224-9.

22. Meselson M, Guillemin J, Hugh-Jones M, Langmuir A, Popova I, Shelokov A et al. The Sverdlovsk anthrax outbreak of 1979. Science (New York, N.Y.) 1994;266(5188):1202-8.

23. Anonymous . Secret service. Nature 2015;523(7562):501-2.

24. Centers for Disease Control and Prevention (CDC), Department of Health and Human Services (HHS). Possession, use, and transfer of select agents and toxins; biennial review. Final rule. Fed Regist 2012;77(194):61083-115.

25. Szinicz L. History of chemical and biological warfare agents. Toxicology 2005;214(3):167-81.

26. Ryan CP. Zoonoses likely to be used in bioterrorism. Public Health Rep 2008;123(3):276-81.

27. Moayeri M, Leppla SH, Vrentas C, Pomerantsev AP, Liu S. Anthrax Pathogenesis. Annu Rev Microbiol 2015;69:185-208.

28. Pile JC, Malone JD, Eitzen EM, Friedlander AM. Anthrax as a potential biological warfare agent. Arch Intern Med 1998;158(5):429-34.

29. Goel AK. Anthrax: A disease of biowarfare and public health importance. World journal of clinical cases 2015;3(1):20-33.

30. Dixon TC, Meselson M, Guillemin J, Hanna PC. Anthrax. N Engl J Med 1999;341(11):815-26.

31. Sweeney DA, Hicks CW, Cui X, Li Y, Eichacker PQ. Anthrax infection. Am J Respir Crit Care Med 2011;184(12):1333-41.

32. Shafazand S, Doyle R, Ruoss S, Weinacker A, Raffin TA. Inhalational anthrax: epidemiology, diagnosis, and management. Chest 1999;116(5):1369-76.

33. Chakraborty A, Khan SU, Hasnat MA, Parveen S, Islam MS, Mikolon A et al. Anthrax outbreaks in Bangladesh, 2009-2010. Am J Trop Med Hyg 2012;86(4):703-10.

34. Navdarashvili A, Doker TJ, Geleishvili M, Haberling DL, Kharod GA, Rush TH et al. Human anthrax outbreak associated with livestock exposure: Georgia, 2012. Epidemiol Infect 2016;144(1):76-87.

35. Orlos Z, Kardos L, Misak O, Lenart B, Lampe Z, Kovacs I et al. Outbreak of anthrax in adults and adolescents: a review of 9 cases in a regional teaching hospital in East Hungary. Clin Microbiol Infect 2016;

36. Reddy R, Parasadini G, Rao P, Uthappa CK, Murhekar MV. Outbreak of cutaneous anthrax in Musalimadugu village, Chittoor district, Andhra Pradesh, India, July-August 2011. J Infect Dev Ctries 2012;6(10):695-9.

37. Espelund M, Klaveness D. Botulism outbreaks in natural environments - an update. Front Microbiol 2014;5:287.

38. Zhang JC, Sun L, Nie QH. Botulism, where are we now? Clin Toxicol (Phila) 2010;48(9):867-79.

39. Arnon SS, Schechter R, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS et al. Botulinum toxin as a biological weapon: medical and public health management. JAMA 2001;285(8):1059-70.

40. Perry RD, Fetherston JD. *Yersinia pestis*--etiologic agent of plague. Clin Microbiol Rev 1997;10(1):35-66.

41. Butler T. Plague history: Yersin's discovery of the causative bacterium in 1894 enabled, in the subsequent century, scientific progress in understanding the disease

and the development of treatments and vaccines. Clinical Microbiology and Infection 2014;20(3):202-9.

42. Pechous RD, Sivaraman V, Stasulli NM, Goldman WE. Pneumonic Plague: The Darker Side of *Yersinia pestis*. Trends Microbiol 2016;24(3):190-7.

43. Riedel S. Plague: from natural disease to bioterrorism. Proc (Bayl Univ Med Cent) 2005;18(2):116-24.

44. Raoult D, Mouffok N, Bitam I, Piarroux R, Drancourt M. Plague: history and contemporary analysis. J Infect 2013;66(1):18-26.

45. Stenseth NC, Atshabar BB, Begon M, Belmain SR, Bertherat E, Carniel E et al. Plague: past, present, and future. PLoS Med 2008;5(1):e3.

46. Anonymous . Plague around the world, 2010-2015. Wkly Epidemiol Rec 2016;91(8):89-93.

47. Bitam I, Dittmar K, Parola P, Whiting MF, Raoult D. Fleas and flea-borne diseases. Int J Infect Dis 2010;14(8):e667-76.

48. Atkinson S, Williams P. *Yersinia* virulence factors - a sophisticated arsenal for combating host defences. F1000Res 2016;5:10.12688/f1000research.8466.1. eCollection 2016.

49. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 1999;96(24):14043-8.

50. Harbeck M, Seifert L, Hansch S, Wagner DM, Birdsell D, Parise KL et al. *Yersinia pestis* DNA from skeletal remains from the 6(th) century AD reveals insights into Justinianic Plague. PLoS Pathog 2013;9(5):e1003349.

51. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM et al. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 2004;101(38):13826-31.

52. Ch'ng SL, Octavia S, Xia Q, Duong A, Tanaka MM, Fukushima H et al. Population structure and evolution of pathogenicity of *Yersinia pseudotuberculosis*. Appl Environ Microbiol 2011;77(3):768-75.

53. Laukkanen-Ninios R, Didelot X, Jolley KA, Morelli G, Sangal V, Kristo P et al. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. Environ Microbiol 2011;13(12):3114-27.

54. Souza RA, Pitondo-Silva A, Falcao DP, Falcao JP. Evaluation of four molecular typing methodologies as tools for determining taxonomy relations and for identifying species among *Yersinia* isolates. J Microbiol Methods 2010;82(2):141-50.

55. Voskressenskaya E, Leclercq A, Tseneva G, Carniel E. Evaluation of ribotyping as a tool for molecular typing of *Yersinia pseudotuberculosis* strains of worldwide origin. J Clin Microbiol 2005;43(12):6155-60.

56. Zhang S, Zhang Z, Liu S, Bingham W, Wilson F. Fatal yersiniosis in farmed deer caused by *Yersinia pseudotuberculosis* serotype O:3 encoding a mannosyltransferase-like protein WbyK. J Vet Diagn Invest 2008;20(3):356-9.

57. Gaydos JK, Zabek E, Raverty S. *Yersinia pseudotuberculosis* septicemia in a beaver from Washington State. J Wildl Dis 2009;45(4):1182-6.

58. Giannitti F, Barr BC, Brito BP, Uzal FA, Villanueva M, Anderson M. Yersinia pseudotuberculosis infections in goats and other animals diagnosed at the California Animal Health and Food Safety Laboratory System: 1990-2012. J Vet Diagn Invest 2014;26(1):88-95.

59. Wallner-Pendleton E, Cooper G. Several outbreaks of *Yersinia pseudotuberculosis* in California turkey flocks. Avian Dis 1983;27(2):524-6.

60. Pärn T, Hallanvuo S, Salmenlinna S, Pihlajasaari A, Heikkinen S, Telkki-Nykänen H et al. Outbreak of *Yersinia pseudotuberculosis* O:1 infection associated with raw milk consumption, Finland, spring 2014. Euro Surveill 2015;20(40):10.2807/1560,7917.ES.2015.20.40.30033.

61. Jalava K, Hakkinen M, Valkonen M, Nakari UM, Palo T, Hallanvuo S et al. An outbreak of gastrointestinal illness and erythema nodosum from grated carrots contaminated with *Yersinia pseudotuberculosis*. J Infect Dis 2006;194(9):1209-16.

62. Kangas S, Takkinen J, Hakkinen M, Nakari UM, Johansson T, Henttonen H et al. *Yersinia pseudotuberculosis* O:1 traced to raw carrots, Finland. Emerg Infect Dis 2008;14(12):1959-61.

63. Nuorti JP, Niskanen T, Hallanvuo S, Mikkola J, Kela E, Hatakka M et al. A widespread outbreak of *Yersinia pseudotuberculosis* O:3 infection from iceberg lettuce. J Infect Dis 2004;189(5):766-74.

64. Babkin IV, Babkina IN. The origin of the variola virus. Viruses 2015;7(3):1100-12.

65. Theves C, Biagini P, Crubezy E. The rediscovery of smallpox. Clin Microbiol Infect 2014;20(3):210-8.

66. Kaiser J. Eradication. Pressure growing to set a date to destroy remaining smallpox stocks. Science 2011;331(6016)

67. Henderson DA, Inglesby TV, Bartlett JG, Ascher MS, Eitzen E, Jahrling PB et al. Smallpox as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. JAMA 1999;281(22):2127-37.

68. Petersen JM, Schriefer ME. Tularemia: emergence/re-emergence. Vet Res 2005;36(3):455-67.

69. Ellis J, Oyston PC, Green M, Titball RW. Tularemia. Clin Microbiol Rev 2002;15(4):631-46.

70. Sjöstedt A. Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. Ann N Y Acad Sci 2007;1105:1-29.

71. Dai S, Mohapatra NP, Schlesinger LS, Gunn JS. Regulation of *Francisella tularensis* virulence. Front Microbiol 2011;1:144.

72. Na W, Park N, Yeom M, Song D. Ebola outbreak in Western Africa 2014: what is going on with Ebola virus? Clin Exp Vaccine Res 2015;4(1):17-22.

73. Weber DJ, Rutala WA, Fischer WA, Kanamori H, Sickbert-Bennett EE. Emerging infectious diseases: Focus on infection control issues for novel coronaviruses (Severe Acute Respiratory Syndrome-CoV and Middle East Respiratory Syndrome-CoV), hemorrhagic fever viruses (Lassa and Ebola), and highly pathogenic avian influenza viruses, A(H5N1) and A(H7N9). Am J Infect Control 2016;44:e91-e100.

74. Rewar S, Mirdha D. Transmission of ebola virus disease: an overview. Annals of global health 2014;80(6):444-51.

75. Borio L, Inglesby T, Peters CJ, Schmaljohn AL, Hughes JM, Jahrling PB et al. Hemorrhagic fever viruses as biological weapons: medical and public health management. JAMA 2002;287(18):2391-405.

76. Jacobsen KH, Aguirre AA, Bailey CL, Baranova AV, Crooks AT, Croitoru A et al. Lessons from the Ebola Outbreak: Action Items for Emerging Infectious Disease Preparedness and Response. Ecohealth 2016;13(1):200-12.

77. Scholz HC, Revilla-Fernández S, Al Dahouk S, Hammerl JA, Zygmunt MS, Cloeckaert A et al. *Brucella vulpis* sp. nov., a novel *Brucella* species isolated from mandibular lymph nodes of red foxes (*Vulpes vulpes*) in Austria. Int J Syst Evol Microbiol 2016;

78. Pappas G, Panagopoulou P, Christou L, Akritidis N. *Brucella* as a biological weapon. Cell Mol Life Sci 2006;63(19-20):2229-36.

79. Seleem MN, Boyle SM, Sriranganathan N. Brucellosis: A re-emerging zoonosis. Vet Microbiol 2010;140(3–4):392-8.

80. Omsland A, Cockrell DC, Howe D, Fischer ER, Virtaneva K, Sturdevant DE et al. Host cell-free growth of the Q fever bacterium *Coxiella burnetii*. Proc Natl Acad Sci U S A 2009;106(11):4430-4.

81. Oyston PC, Davies C. Q fever: the neglected biothreat agent. J Med Microbiol 2011;60(Pt 1):9-21.

82. WHO. Health Aspects of Chemical and Biological Weapons, 1st ed. Geneva: World Health Organization 1970;

83. Limonard GJ, Peters JB, Besselink R, Groot CA, Dekhuijzen PN, Vercoulen JH et al. Persistence of impaired health status of Q fever patients 4 years after the first Dutch outbreak. Epidemiol Infect 2015;:1-6.

84. Tissot-Dupont H, Amadei MA, Nezri M, Raoult D. Wind in November, Q fever in December. Emerg Infect Dis 2004;10(7):1264-9.

85. Burnett MW. Q Fever. J Spec Oper Med 2015;15(2):109-11.

86. Anderson PD. Bioterrorism: toxins as weapons. J Pharm Pract 2012;25(2):121-9.

87. Thornton SL, Darracq M, Lo J, Cantrell FL. Castor bean seed ingestions: a state-wide poison control system's experience. Clin Toxicol (Phila) 2014;52(4):265-8.

88. Griffiths GD. Understanding ricin from a defensive viewpoint. Toxins (Basel) 2011;3(11):1373-92.

89. Centers for Disease Control and Prevention (CDC). Investigation of a ricin-containing envelope at a postal facility--South Carolina, 2003. MMWR Morb Mortal Wkly Rep 2003;52(46):1129-31.

90. Sack DA, Sack RB, Nair GB, Siddique AK. Cholera. Lancet 2004;363(9404):223-33.

91. Mandal S, Mandal MD, Pal NK. Cholera: a great global concern. Asian Pacific Journal of Tropical Medicine 2011;4(7):573-80.

92. Verma R, Khanna P, Chawla S. Cholera vaccine: new preventive tool for endemic countries. Hum Vaccin Immunother 2012;8(5):682-4.

93. Singh DV, Matte MH, Matte GR, Jiang S, Sabeena F, Shukla BN et al. Molecular analysis of *Vibrio cholerae* O1, O139, non-O1, and non-O139 strains: clonal relationships between clinical and environmental isolates. Appl Environ Microbiol 2001;67(2):910-21.

94. Anderson AM, Varkey JB, Petti CA, Liddle RA, Frothingham R, Woods CW. Non-O1 *Vibrio cholerae* septicemia: case report, discussion of literature, and relevance to bioterrorism. Diagn Microbiol Infect Dis 2004;49(4):295-7.

95. Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB. Cholera. Lancet 2012;379(9835):2466-76.

96. Deen J, von Seidlein L, Luquero FJ, Troeger C, Reyburn R, Lopez AL et al. The scenario approach for countries considering the addition of oral cholera vaccination in cholera preparedness and control plans. Lancet Infect Dis 2016;16(1):125-9.

97. Blevins SM, Bronze MS. Robert Koch and the 'golden age' of bacteriology. International Journal of Infectious Diseases 2010;14(9):e744-51.

98. Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015;28(1):208-36.

99. Haristoy X, Lozniewski A, Tram C, Simeon D, Bevanger L, Lion C. *Francisella tularensis* bacteremia. J Clin Microbiol 2003;41(6):2774-6.

100. Vartoukian SR, Palmer RM, Wade WG. Strategies for culture of 'unculturable' bacteria. FEMS Microbiol Lett 2010;309(1):1-7.

101. Mori M, Mertens K, Cutler SJ, Santos AS. Critical Aspects for Detection of *Coxiella burnetii*. Vector Borne Zoonotic Dis 2017;17(1):33-41.

102. Ngom B, Guo Y, Wang X, Bi D. Development and application of lateral flow test strip technology for detection of infectious agents and chemical contaminants: a review. Anal Bioanal Chem 2010;397(3):1113-35.

103. Koczula KM, Gallotta A. Lateral flow assays. Essays Biochem 2016;60(1):111-20.

104. Mirski T, Bartoszcze M, Bielawska-Drozd A, Cieslik P, Michalski AJ, Niemcewicz M et al. Review of methods used for identification of biothreat agents in environmental protection and human health aspects. Ann Agric Environ Med 2014;21(2):224-34.

105. Singhal N, Kumar M, Kanaujia PK, Virdi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol 2015;6:791.

106. Karagoz S, Kilic S, Berk E, Uzel A, Celebi B, Comoglu S et al. *Francisella tularensis* bacteremia: report of two cases and review of the literature. New Microbiol 2013;36(3):315-23.

107. Lasch P, Wahab T, Weil S, Palyi B, Tomaso H, Zange S et al. Identification of Highly Pathogenic Microorganisms by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: Results of an Interlaboratory Ring Trial. J Clin Microbiol 2015;53(8):2632-40.

108. Lista F, Reubsaet FA, De Santis R, Parchen RR, de Jong AL, Kieboom J et al. Reliable identification at the species level of *Brucella* isolates with MALDI-TOF-MS. BMC Microbiol 2011;11:267,2180-11-267.

109. Kalb SR, Boyer AE, Barr JR. Mass Spectrometric Detection of Bacterial Protein Toxins and Their Enzymatic Activity. Toxins (Basel) 2015;7(9):3497-511.

110. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. 1986. Biotechnology 1992;24:17-27.

111. Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors - occurrence, properties and removal. J Appl Microbiol 2012;113(5):1014-26.

112. Janzen TW, Thomas MC, Goji N, Shields MJ, Hahn KR, Amoako KK. Rapid detection method for *Bacillus anthracis* using a combination of multiplexed real-time PCR and pyrosequencing and its application for food biodefense. J Food Prot 2015;78(2):355-61.

113. Janse I, Hamidjaja RA, Bok JM, van Rotterdam BJ. Reliable detection of *Bacillus anthracis, Francisella tularensis* and *Yersinia pestis* by using multiplex qPCR including internal controls for nucleic acid extraction and amplification. BMC Microbiol 2010;10:314,2180-10-314.

114. Matero P, Hemmilä H, Tomaso H, Piiparinen H, Rantakokko-Jalava K, Nuotio L et al. Rapid field detection assays for *Bacillus anthracis, Brucella* spp., *Francisella tularensis* and *Yersinia pestis*. Clin Microbiol Infect 2011;17(1):34-43.

115. Mölsa M, Hemmilä H, Katz A, Niemimaa J, Forbes KM, Huitu O et al. Monitoring biothreat agents (*Francisella tularensis, Bacillus anthracis* and *Yersinia pestis*) with a portable real-time PCR instrument. J Microbiol Methods 2015;115:89-93.

116. Girault G, Blouin Y, Vergnaud G, Derzelle S. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. BMC Genomics 2014;15:288,2164-15-288.

117. Sobrino B, Brión M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int 2005;154(2–3):181-94.

118. Silaghi C, Pfeffer M, Kiefer D, Kiefer M, Obiegala A. *Bartonella*, Rodents, Fleas and Ticks: a Molecular Field Study on Host-Vector-Pathogen Associations in Saxony, Eastern Germany. Microb Ecol 2016;

119. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 1977;74(11):5088-90.

120. Chen K, Neimark H, Rumore P, Steinman CR. Broad range DNA probes for detecting and amplifying eubacterial nucleic acids. FEMS Microbiol Lett 1989;48(1):19-24.

121. Lleo MM, Ghidini V, Tafi MC, Castellani F, Trento I, Boaretti M. Detecting the presence of bacterial DNA by PCR can be useful in diagnosing culture-negative

cases of infection, especially in patients with suspected infection and antibiotic therapy. FEMS Microbiol Lett 2014;354(2):153-60.

122. Ruppitsch W, Stoger A, Indra A, Grif K, Schabereiter-Gurtner C, Hirschl A et al. Suitability of partial 16S ribosomal RNA gene sequence analysis for the identification of dangerous bacterial pathogens. J Appl Microbiol 2007;102(3):852-9.

123. Wilbrink B, van der Heijden IM, Schouls LM, van Embden JD, Hazes JM, Breedveld FC et al. Detection of bacterial DNA in joint samples from patients with undifferentiated arthritis and reactive arthritis, using polymerase chain reaction with universal 16S ribosomal RNA primers. Arthritis & Rheumatism 1998;41(3):535-43.

124. Gee JE, Sacchi CT, Glass MB, De BK, Weyant RS, Levett PN et al. Use of 16S rRNA gene sequencing for rapid identification and differentiation of *Burkholderia pseudomallei* and *B. mallei*. J Clin Microbiol 2003;41(10):4647-54.

125. Mahmoudi N, Robeson MS 2nd, Castro HF, Fortney JL, Techtmann SM, Joyner DC et al. Microbial community composition and diversity in Caspian Sea sediments. FEMS Microbiology Ecology (Wiley) 2015;91(1):1-11.

126. Razzauti M, Galan M, Bernard M, Maman S, Klopp C, Charbonnel N et al. A Comparison between Transcriptome Sequencing and 16S Metagenomics for Detection of Bacterial Pathogens in Wildlife. PLoS Negl Trop Dis 2015;9(8):e0003929.

127. Mojica FJ, Rodriguez-Valera F. The discovery of CRISPR in archaea and bacteria. FEBS J 2016;

128. Shariat N, Dudley EG. CRISPRs: molecular signatures used for pathogen subtyping. Appl Environ Microbiol 2014;80(2):430-9.

129. Choi KR, Lee SY. CRISPR technologies for bacterial systems: Current achievements and future directions. Biotechnol Adv

130. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-Generation Sequencing for Infectious Disease Diagnosis and Management: A Report of the Association for Molecular Pathology. J Mol Diagn 2015;17(6):623-34.

131. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17(6):333-51.

132. Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. The next-generation sequencing technology and application. Protein Cell 2010;1(6):520-36.

133. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. Genomics 2009;93(2):105-11.

134. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, Garcia-Cobos S et al. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol 2017;243:16-24.

135. Deng X, den Bakker HC, Hendriksen RS. Genomic Epidemiology: Whole-Genome-Sequencing-Powered Surveillance and Outbreak Investigation of Food-borne Bacterial Pathogens. Annu Rev Food Sci Technol 2016;7:353-74.

136. Oliver GR, Hart SN, Klee EW. Bioinformatics for clinical next generation sequencing. Clin Chem 2015;61(1):124-35.

137. Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D et al. High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. J Hosp Infect 2010;75(1):37-41.

138. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM et al. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. BMJ Open 2012;2(3):10.1136/bmjopen,2012-001124. Print 2012.

139. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics 2015;13(5):278-89.

140. Mikheyev AS, Tin MM. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour 2014;14(6):1097-102.

141. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. Nat Methods 2015;12(4):351-6.

142. Madoui MA, Engelen S, Cruaud C, Belser C, Bertrand L, Alberti A et al. Genome assembly using Nanopore-guided long and error-free DNA reads. BMC Genomics 2015;16:327,015-1519-z.

143. Rasko DA, Worsham PL, Abshire TG, Stanley ST, Bannan JD, Wilson MR et al. *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. Proc Natl Acad Sci U S A 2011;108(12):5027-32.

144. Gardner SN, Frey KG, Redden CL, Thissen JB, Allen JE, Allred AF et al. Targeted amplification for enhanced detection of biothreat agents by next-generation sequencing. BMC Res Notes 2015;8:682,015-1530-0.

145. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 2011;6(12):e27310.

146. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 2005;151(Pt 3):653-63.

147. Le Fleche P, Hauck Y, Onteniente L, Prieur A, Denoeud F, Ramisse V et al. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC Microbiol 2001;1:2.

148. Staden R. The Staden sequence analysis package. Mol Biotechnol 1996;5(3):233-41.

149. Bliem R, Schauer S, Plicka H, Obwaller A, Sommer R, Steinrigl A et al. A novel triplex quantitative PCR strategy for quantification of toxigenic and nontoxigenic *Vibrio cholerae* in aquatic environments. Appl Environ Microbiol 2015;81(9):3077-85.

150. Huang J, Zhu Y, Wen H, Zhang J, Huang S, Niu J et al. Quadruplex real-time PCR assay for detection and identification of *Vibrio cholerae* O1 and O139 strains and determination of their toxigenic potential. Appl Environ Microbiol 2009;75(22):6981-5.

151. Tall A, Teillon A, Boisset C, Delesmont R, Touron-Bodilis A, Hervio-Heath D. Real-time PCR optimization to identify environmental *Vibrio spp.* strains. J Appl Microbiol 2012;113(2):361-72.

152. Robins WP, Mekalanos JJ. Genomic science in understanding cholera outbreaks and evolution of *Vibrio cholerae* as a human pathogen. Curr Top Microbiol Immunol 2014;379:211-29.

153. Wu Q, McFee WE, Goldstein T, Tiller RV, Schwacke L. Real-time PCR assays for detection of *Brucella* spp. and the identification of genotype ST27 in bottlenose dolphins (*Tursiops truncatus*). J Microbiol Methods 2014;100:99-104.

154. Liu L, Sun Y, Kargbo B, Zhang C, Feng H, Lu H et al. Detection of Zaire Ebola virus by real-time reverse transcription-polymerase chain reaction, Sierra Leone, 2014. J Virol Methods 2015;222:62-5.

155. Woubit A, Yehualaeshet T, Habtemariam T, Samuel T. Novel genomic tools for specific and real-time detection of biothreat and frequently encountered foodborne pathogens. J Food Prot 2012;75(4):660-70.

156. Mölsa M, Koskela KA, Rönkkö E, Ikonen N, Ziegler T, Nikkari S. Detection of influenza A viruses with a portable real-time PCR instrument. J Virol Methods 2012;181(2):188-91.

157. Shipley MA, Koehler JW, Kulesh DA, Minogue TD. Comparison of nucleic acid extraction platforms for detection of select biothreat agents for use in clinical resource limited settings. J Microbiol Methods 2012;91(1):179-83.

158. Buffet JP, Marsot M, Vaumourin E, Gasqui P, Masseglia S, Marcheteau E et al. Co-infection of *Borrelia afzelii* and *Bartonella* spp. in bank voles from a suburban forest. Comp Immunol Microbiol Infect Dis 2012;35(6):583-9.

159. Rossow H, Sissonen S, Koskela KA, Kinnunen PM, Hemmilä H, Niemimaa J et al. Detection of *Francisella tularensis* in voles in Finland. Vector Borne Zoonotic Dis 2014;14(3):193-8.

160. Brown DR, Talkington DF, Thacker WL, Brown MB, Dillehay DL, Tully JG. *Mycoplasma microti* sp. nov., isolated from the respiratory tract of prairie voles (*Microtus ochrogaster*). Int J Syst Evol Microbiol 2001;51(Pt 2):409-12.

161. Kallio ER, Begon M, Birtles RJ, Bown KJ, Koskela E, Mappes T et al. First report of *Anaplasma phagocytophilum* and *Babesia microti* in rodents in Finland. Vector Borne Zoonotic Dis 2014;14(6):389-93.

162. Sacchi CT, Whitney AM, Mayer LW, Morey R, Steigerwalt A, Boras A et al. Sequencing of 16S rRNA gene: a rapid tool for identification of *Bacillus anthracis*. Emerg Infect Dis 2002;8(10):1117-23.

163. Clarridge JE,3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 2004;17(4):840,62, table of contents.

164. Ghyselinck J, Pfeiffer S, Heylen K, Sessitsch A, De Vos P. The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. PLoS One 2013;8(8):e71360.

165. Gill P, Rowlands D, Tully G, Bastisch I, Staples T, Scott P. Manufacturer contamination of disposable plastic-ware and other reagents—An agreed position statement by ENFSI, SWGDAM and BSAG. Forensic Science International: Genetics 2010;4(4):269-70.

166. Schmidt T, Hummel S, Herrmann B. Evidence of contamination in PCR laboratory disposables. Naturwissenschaften 1995;82(9):423-31.

167. Böttger EC. Frequent contamination of Taq polymerase with DNA. Clin Chem 1990;36(6):1258-9.

168. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarski EB, Fox AJ. Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. J Clin Microbiol 2000;38(5):1747-52.

169. Fredricks DN, Relman DA. Improved amplification of microbial DNA from blood cultures by removal of the PCR inhibitor sodium polyanetholesulfonate. J Clin Microbiol 1998;36(10):2810-6.

170. Newsome T, Li BJ, Zou N, Lo SC. Presence of bacterial phage-like DNA sequences in commercial Taq DNA polymerase reagents. J Clin Microbiol 2004;42(5):2264-7.

171. Rand KH, Houck H. Taq polymerase contains bacterial DNA of unknown origin. Mol Cell Probes 1990;4(6):445-50.

172. Evans GE, Murdoch DR, Anderson TP, Potter HC, George PM, Chambers ST. Contamination of Qiagen DNA extraction kits with *Legionella* DNA. J Clin Microbiol 2003;41(7):3452-3.

173. Peters RP, Mohammadi T, Vandenbroucke-Grauls CM, Danner SA, van Agtmael MA, Savelkoul PH. Detection of bacterial DNA in blood samples from febrile patients: underestimated infection or emerging contamination? FEMS Immunol Med Microbiol 2004;42(2):249-53.

174. Meier A, Persing DH, Finken M, Böttger EC. Elimination of contaminating DNA within polymerase chain reaction reagents: implications for a general approach to detection of uncultured pathogens. J Clin Microbiol 1993;31(3):646-52.

175. Moissl C, Osman S, La Duc MT, Dekas A, Brodie E, DeSantis T et al. Molecular bacterial community analysis of clean rooms where spacecraft are assembled. FEMS Microbiol Ecol 2007;61(3):509-21.

176. Renko J, Lepp PW, Oksala N, Nikkari S, Nikkari ST. Bacterial signatures in atherosclerotic lesions represent human commensals and pathogens. Atherosclerosis 2008;201(1):192-7.

177. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 2014;12:87,014-0087-z.

178. Bachmann NL, Petty NK, Ben Zakour NL, Szubert JM, Savill J, Beatson SA. Genome analysis and CRISPR typing of *Salmonella enterica* serovar Virchow. BMC Genomics 2014;15:389,2164-15-389.

179. Yang S, Liu J, Shao F, Wang P, Duan G, Yang H. Analysis of the features of 45 identified CRISPR loci in 32 *Staphylococcus aureus*. Biochem Biophys Res Commun 2015;464(3):894-900.

180. Rahmatabadi SS, Nezafat N, Negahdaripour M, Hajighahramani N, Morowvat MH, Ghasemi Y. Studying the features of 57 confirmed CRISPR loci in 29 strains of *Escherichia coli*. J Basic Microbiol 2016;56(6):645-53.

ORIGINAL PUBLICATIONS

I

Available online at www.sciencedirect.com
ScienceDirect
Diagnostic Microbiology and Infectious Disease 65 (2009) 339–344
DIAGNOSTIC MICROBIOLOGY AND INFECTIOUS DISEASE
www.elsevier.com/locate/diagmicrobio

A multiplatform real-time polymerase chain reaction detection assay for *Vibrio cholerae*

Katja A. Koskela[a], Pirjo Matero[a], Janet M. Blatny[b], Else M. Fykse[b], Jaran Strand Olsen[b], Lasse O. Nuotio[a], Simo Nikkari[a,*]

[a]*Centres for Military Medicine and for Biological Threat Preparedness, FI-00290 Helsinki, Finland*
[b]*Norwegian Defence Research Establishment, PO Box 25, N-2027 Kjeller, Norway*

Received 30 January 2009; accepted 7 July 2009

Abstract

We report a multiplatform real-time polymerase chain reaction methodology based on genes encoding for the regulatory toxR activator and enterotoxin A protein to determine enterotoxigenic *Vibrio cholerae* types from other vibrios. This assay, which was successfully validated on a collection of 87 bacterial strains, including 63 representatives of *V. cholerae* and 8 noncholera vibrios provides a rapid tool for detection and identification of cholera.

Keywords: Real-time PCR; Vibrio cholera; toxR; ctxA; Field detection

Cholera remains a global threat to public health, especially in developing countries (World Health Organization, 2008). Furthermore, *Vibrio cholerae* is classified as a potential category B bioterrorism agent by the Centers for Disease Control and Prevention. There is, accordingly, a need to develop fast and reliable methods to detect and identify the agent.

V. cholerae expresses several virulence factors; the most important of these are the attachment fimbriae, referred to as toxin-coregulated pili (tcp), and the highly potent enterotoxin (ctx). Only 2 serotypes, O1 and O139, of *V. cholerae* are associated with epidemic or pandemic cholera. Other serotypes can occasionally cause small-scale outbreaks of diarrhea, but those strains very rarely harbor the *tcp* and *ctx* genes (Bhattacharya et al., 2006; Sack et al., 2004). The cholera toxin transcriptional activator (toxR) protein, encoded by *toxR*, is the primary regulator of the *ctx-tcp* operon (Rivera et al., 2001). In contrast to the *tcp* and *ctx* genes, *toxR* is commonly present also in nonpathogenic environmental non-O1 and non-O139 *V. cholerae* isolates (Singh, 2001). Therefore, in this study, we initially targeted regulatory toxR gene sequences to detect all *V. cholerae* strains, followed by distinction of enterotoxigenic (ctx+) types with another polymerase chain reaction (PCR) assay. We tested performance of these assays on the Applied Biosystems' 7300 and 7900HT Fast (ABI, Foster City, CA), as well as on the Idaho Technology's RAZOR (Salt Lake City, UT) field-deployable instrument.

All bacterial strains used in this study and their sources are listed in Table 1. The primers and MGB probes shown in Table 2 were designed using Primer Express 3.0 software (ABI). In addition to the oligonucleotides, the 25-μL reaction mixture for real-time PCR performed with the ABI 7300 instrument contained 12.5-μL 2× TaqMan Universal PCR Master Mix (ABI), 2.5-μL template, 2.5-μL 10× Exo Internal positive control (IPC) Mix, and 0.5-μL 50× Exo IPC DNA (ABI). The 10-μL reaction mixture for the ABI 7900HT Fast instrument contained 5-μL 2× TaqMan Fast Universal PCR master mix (ABI), 1-μL template, 1-μL 10× Exo IPC Mix, and 0.2-μL 50× Exo IPC DNA. The 100 μL reaction mixture for the RAZOR instrument contained 50-μL Premix Ex Taq master mix (Takara Bio, Otsu, Japan) and 10-μL template. With the RAZOR, simultaneous IPC testing was not possible because the apparatus monitors only 1 wavelength (475 ± 20 nm).

* Corresponding author. Tel.: +358-9-1812-5821; fax: +358-9-1812-5799.
E-mail address: simo.nikkari@mil.fi (S. Nikkari).

0732-8893/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.diagmicrobio.2009.07.009

Table 1
Species, strains, and sources of bacteria and the presence of the target genes as determined by the developed PCR assays

Bacterium	Code and serotype	Donator	Source	toxR	ctxA
V. cholerae	VC 2	1	NK	+	−
V. cholerae	VC 4	1	NK	+	−
V. cholerae	VC 5	1	NK	+	−
V. cholerae	VC 6	1	NK	+	−
V. cholerae	VC 14	1	Marinara mix, Thailand	+	−
V. cholerae	VC 18	1	Shrimp, Thailand	+	−
V. cholerae	VC 21	1	Raw scampi	+	−
V. cholerae	VC 24	1	Raw scampi	+	−
V. cholerae	VC 27	1	Scampi, Bangladesh	+	−
V. cholerae	VC 37	1	NK	+	−
V. cholerae	VC 45	1	NK	+	−
V. cholerae	VC 46	1	NK	+	−
V. cholerae	VC 47	1	NK	+	−
V. cholerae	VC 48	1	Shrimp, India	+	−
V. cholerae	VC 61	1	NK	+	−
V. cholerae	VC 69	1	NK	+	−
V. cholerae	VC 84	1	Scampi	+	−
V. cholerae	VC 86	1	Shrimp	+	−
V. cholerae	VC 110	1	Mussel, Norway	+	−
V. cholerae	VC 117	1	Mussel, Norway	+	−
V. cholerae	VC 134	1	Mussel, Norway	+	−
V. cholerae	VC 161	1	Mussel, Norway	+	−
V. cholerae	VC 177	1	Mussel, Norway	+	−
V. cholerae	VC 198	1	Mussel, Norway	+	−
V. cholerae	VC 211	1	Mussel, Norway	+	−
V. cholerae	VC 216	1	Barracuda, India	+	−
V. cholerae	VC 229	2	Pus	+	−
V. cholerae	VC 230	2	Feces	+	−
V. cholerae	VC 232	2	Pus	+	−
V. cholerae	VC 233	2	Feces	+	−
V. cholerae	VC 234	2	Feces	+	−
V. cholerae	VC 235	2	Feces	+	−
V. cholerae	VC 236	2	Feces	+	−
V. cholerae	VC 237	2	Feces	+	−
V. cholerae	VC 238	2	Feces	+	−
V. cholerae	VC 239	2	Feces	+	−
V. cholerae	VC 240	2	Feces	+	−
V. cholerae	VC 241	2	Feces	+	−
V. cholerae	VC 242, O1, Ogawa El Tor	2	Feces	+	+
V. cholerae	VC 243, O1, Ogawa El Tor	2	Feces	+	+
V. cholerae	VC 246	1	Mussel, Norway	+	−
V. cholerae	VC 250	1	Oyster, Norway	+	−
V. cholerae	VC 285	1	Mussel, Norway	+	−
V. cholerae	VC 286	1	Mussel, Norway	+	−
V. cholerae	VC 293	1	Mussel, Norway	+	−
V. cholerae	VC 328	1	Mussel, Norway	+	−
V. cholerae	VC 329	1	Mussel, Norway	+	−
V. cholerae	VC 330	1	Mussel, Norway	+	−
V. cholerae	VC 344	1	Mussel, Norway	+	−
V. cholerae	VC 347	1	Scampi, Bangladesh	+	−
V. cholerae	VC 348	1	Scampi, Bangladesh	+	−
V. cholerae	VC 354	1	Shrimp	+	−
V. cholerae	VC 358	1	Mussel, Norway	+	−
V. cholerae	VC 359	1	Mussel, Norway	+	−
V. cholerae	VC 503	1	Water, Norway	+	−
V. cholerae	VC 504	1	Water, Norway	+	−
V. cholerae	VC 518	1	Water, Norway	+	−
V. cholerae	VC 552	1	Water, Norway	+	−
V. cholerae	VC 553	1	Water, Norway	+	−
V. cholerae	VC 570	1	Water, Norway	+	−
V. cholerae	VC 571	1	Water, Norway	+	−
V. cholerae	VC 572	1	Water, Norway	+	−
V. cholerae	VC 614	1	Water, Norway	+	−
Other *V.* spp.					
V. alginolyticus	VA 054	1	NK	−	−
V. alginolyticus	VA 647	1	NK	−	−
V. fischeri	HAMBI 2941	3	HAMBI	−	−
V. fluvialis	VF 062	1	NK	−	−
V. mimicus	VM 052	1	NK	−	−
V. mimicus	VM 345	1	NK	−	−
V. metschnikovii	VM 116	1	NK	−	−
V. parahaemolyticus	VP 160	1	NK	−	−
Other bacterial species					
B. anthracis	ATCC 4229	4	ATCC	−	−
B. anthracis	Sterne 7702	4	NK	−	−
B. cereus	HAMBI 250, ATCC 10987	3	HAMBI	−	−
B. melitensis	681	5	Human, Finland	−	−
C. jejuni	E1 2702/1/04	4	NK	−	−
C. upsaliensis	GNS2897	4	NK	−	−
E. coli (ETEC)	RH 4266	6	Feces	−	−
F. tularensis	LVS, ATCC 29684	7	Vole, Russia	−	−

Table 1 (*continued*)

Bacterium	Code and serotype	Donator	Source	toxR	ctxA
L. anguillarum	2271/1	4	Salmon, Finland	–	–
S. typhimurium	ATCC 13311	8	ATCC	–	–
S. marcescens	DMS 1608	9	DSMZ	–	–
S. marcescens	DMS 30121, ATCC 13880	9	DSMZ	–	–
S. marcescens	DMS 30126	9	DSMZ	–	–
S. marcescens	DMS 46342, ATCC 27117	9	DSMZ	–	–
Y. enterocolitica	20373/79	8	NK	–	–
Y. pestis	EV76-c	8	NK	–	–

The bacteria were cultivated using standard laboratory procedures, followed by total DNA extraction with MagNA Pure technology (Roche, Basel, Switzerland). The PCR thermocycling parameters with the ABI 7300 instrument were 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C, whereas with the ABI 7900HT Fast instrument, they were 20 s at 95 °C, followed by 40 cycles of 1 s at 95 °C and 20 s at 60 °C. These parameters for the RAZOR instrument were as follows: 10 s at 94 °C, followed by 55 cycles of 5 s at 94 °C and 30 s at 60 °C. The duration of the real-time PCR assays with the ABI 7300, the ABI 7900HT Fast, and RAZOR instruments was 100, 35, and 43 min, respectively. Identical specificity results were obtained with all 3 PCR instruments.

NK = not known, ETEC = *Enterotoxigenic E. coli*.

1. Norwegian Veterinary School of Science, Oslo, Norway.
2. Department of Food-borne Infections, Norwegian Institute of Public Health, Oslo, Norway.
3. HAMBI-collection (Department of Applied Chemistry and Microbiology, Faculty of Agriculture and Forestry, University of Helsinki).
4. The Laboratory Strain Collection of the Finnish Food Safety Authority (EVIRA, Helsinki/Oulu, Finland).
5. Turku University Central Hospital, Turku, Finland.
6. National Institute for Health and Welfare, Helsinki, Finland.
7. Swedish Defence Research Agency (FOI, Umeå, Sweden).
8. Department of Bacteriology and Immunology, University of Helsinki, Helsinki, Finland.
9. German Collection of Microorganism and Cultures (DSMZ, Braunschweig, Germany).

The sensitivity of both developed PCR assays on all 3 platforms was at least 100 fg per reaction when purified *V. cholerae* (VC 243) DNA was used as template. This equals to total genome weight of roughly 20 bacterial cells. The sensitivity of the PCR assays to detect *V. cholerae* bacteria in spiked saline, as well as in brackish seawater samples, was at least 1800 colony-forming units (CFU)/mL for the ABI instruments and 180 CFU/mL for the RAZOR field instrument (Fig. 1). These sensitivities correspond to approximately 2 CFU/reaction with the ABI 7900HT Fast and RAZOR instruments and to 5 CFU/reaction with the ABI 7300. The average numbers of *V. cholerae* in endemic areas of Bangladesh and Peruvian coastal waters have been reported as 5×10^3 and 6×10^5 cells/mL, respectively (Brayton et al., 1987; Franco et al., 1997). Our sensitivity exceeds these expected levels in natural waters 3- to 3000-fold.

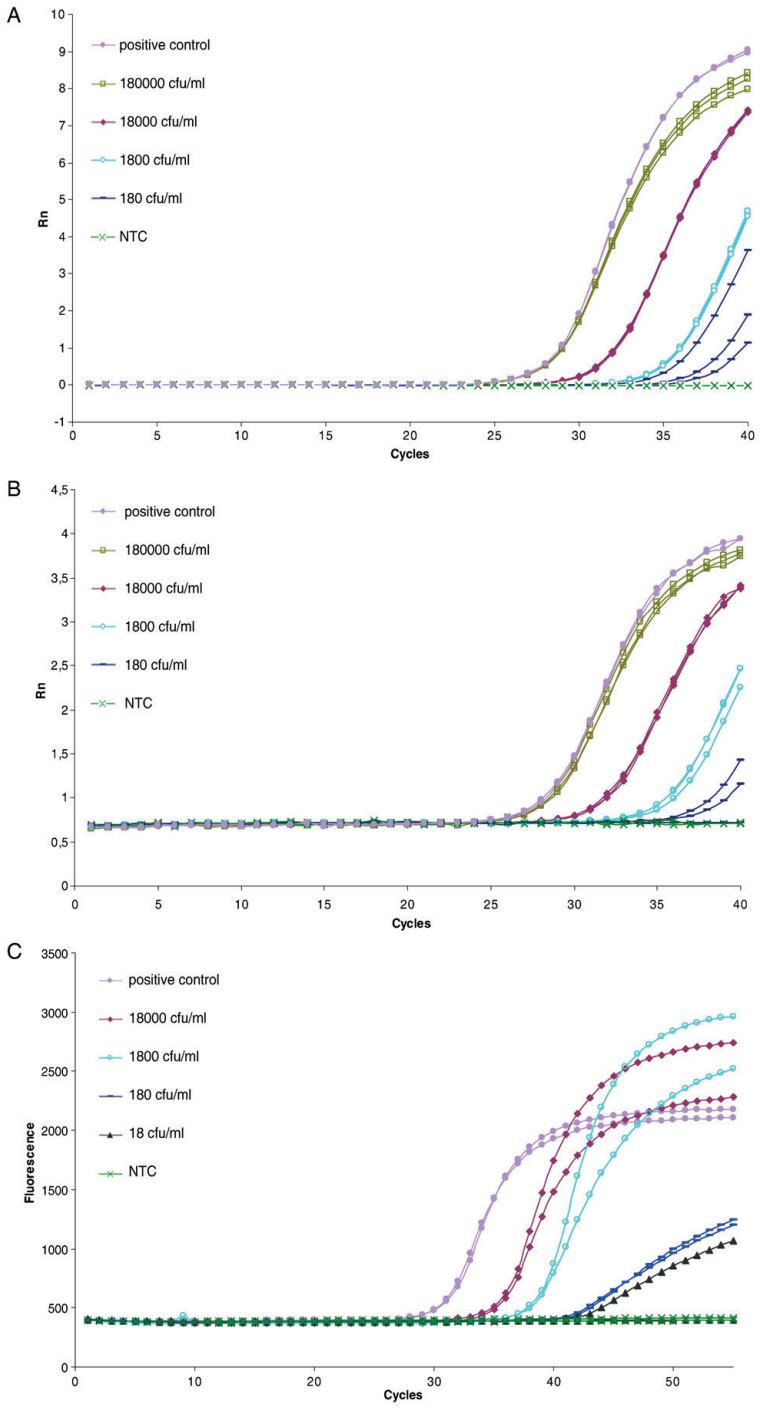
Lipp et al. (2003) reported successful detection of 0.4 *V. cholerae* cells/mL directly from coastal waters and plankton by concentrating the sample before conventional multiplex end-point PCR analysis. The technical sensitivity of our PCR assay is comparable with other earlier reports. Lyon (2001) performed TaqMan PCR with the ABI 7700 instrument to amplify DNA purified from dilutions of *V. cholera* cells with lowest detection limits of 7.3 and 8.2 CFU per reaction. Gubala and Proll (2006) used heat-lysed *V. cholerae* cells as template, and their detection limit was 5 CFU per reaction with the Smart Cycler. In addition, Chomvarin et al. (2007) reported sensitivity of 100 fg per reaction while detecting purified *V. cholerae* DNA in a duplex PCR requiring further agarose gel analysis.

Our results show no significant differences in final assay sensitivities between the 3 platforms while using the described parameters. However, differences were observed in the average crossing threshold/crossing point (ct/cp) values. When 100 fg of purified *V. cholerae* DNA was used as template in 10 parallel reactions, the average ct/cp values were 36.4, 36.5, and 41.6 with the toxR assay and 36.3, 36.3, and 39.3 with the ctxA assay, with the ABI 7300, ABI 7900HT Fast, and RAZOR instruments, respectively. The ct values of the ABI instruments were consistent with each other, whereas the cp of the RAZOR was higher. The ABI algorithms compare signal intensity with a predetermined threshold level, whereas the RAZOR seeks another point where the sample fluorescence exceeds that of the background. In short, the algorithms in ABI instruments can be used for quantitative analysis, whereas the RAZOR is

Table 2
Oligonucleotides used in the *V. cholerae* PCR assays

Target gene	Primer	Nucleotide sequence	Size of amplicon (bp)	Optimal concentrations (nmol/L)	
				ABI instruments	RAZOR field instrument
toxR	Forward	5′-TGGCATCGTTAGGGTTAGCAA-3′	68	300	300
	Reverse	5′-CATTCACAGCCCTGAAGTTTCA-3′		900	900
	Probe	5′-FAM-CGTAAGGTTATGTTTTCC-MGBNFQ-3′		250	250
ctxA	Forward	5′-ACTCACTCTGTCCTCTTGGCATAA-3′	67	300	900
	Reverse	5′-GCAGATTCTAGACCTCCTGATGAAAT-3′		300	900
	Probe	5′-FAM-ACCACCTGACTGCTT-MGBNFQ-3′		250	250

Primer concentrations for PCR were optimized in a matrix format as described earlier (Skottman et al., 2007).


A
positive control
180000 cfu/ml
18000 cfu/ml
1800 cfu/ml
180 cfu/ml
NTC
Rn
Cycles
B
positive control
180000 cfu/ml
18000 cfu/ml
1800 cfu/ml
180 cfu/ml
NTC
Rn
Cycles
C
positive control
18000 cfu/ml
1800 cfu/ml
180 cfu/ml
18 cfu/ml
NTC
Fluorescence
Cycles

primarily designed to qualitatively detect the presence or absence of target DNA.

All of the 63 *V. cholerae* strains gave an unambiguously positive signal with the toxR assay, whereas the ctxA assay was reactive only with the pathogenic El Tor O1 strains (VC 242 and VC 243, Table 1). The specificity of the assays was verified by analyzing phylogenetically closely related bacterial species, including noncholera vibrios, *Listonella*, *Campylobacter*, as well as other water-borne pathogens and selected biothreat agents. In addition, we tested 4 *Serratia marcescens* as well as an heat-labile enterotoxin (LT) toxin-producing *Escherichia coli* strain because these species had been implicated during our initial oligonucleotide design to share similarities with our target genes. However, none of the above specificity control strains gave a positive signal in the developed assays as tested on all 3 platforms, whereas the IPCs associated with these runs were reactive in the ABI tests indicating absence of PCR inhibitors.

Use of the same target genes, *ctxA* and *toxR*, as in the current study was recently reported, and the duration of that isothermal real-time nucleic acid sequence-based amplification assay was 95 min (Fykse et al., 2007). Other earlier studies based on real-time PCR detection of *V. cholerae* have used a variety of technology platforms with inherent time requirements for thermocycling. The SmartCycler II system (Cepheid, Sunnyvale, CA) (Blackstone et al., 2007), SmartCycler with SYBR green dye (Cepheid, Sunnyvale, CA) (Gubala, 2006), SYBR Green LightCycler PCR assay (Roche, Basel, Switzerland) (Fukushima et al., 2003), and ABI Prism 7700 sequence detection system (Lyon, 2001) required approximately 40 min, over an hour, 1 h, and 100 min, respectively, for assay completion.

Although we have deployed the tungsten-halogen lamp-based 29-kg ABI 7300 instrument in a mobile laboratory (data not shown), use of the rapidly (35 min) performing 82-kg ABI 7900HT Fast with its fragile laser excitation technology is limited to a stationary laboratory. In contrast, the RAZOR instrument has been originally designed for field use, and its lightweight (4.1 kg) and small size allow it to be carried by 1 person.

The high sensitivity and specificity of PCR, ease of transfer between platforms using the TaqMan technology, and especially the possibility to employ a portable field instrument make these new assays feasible alternatives for cholera diagnostics.

Acknowledgments

The authors thank L.M. Rørvik, Department of Food Safety and Infection Biology, Norwegian School of Veterinary Sciences, Oslo, Norway and J. Lassen, Department of Food-borne Infections, Norwegian Institute of Public Health, Oslo, Norway for kindly providing *Vibrio* strains. T. Pasanen and E. Tarkka from Helsinki District Central Hospital Laboratory Diagnostics (HUSLAB), Helsinki, Finland are acknowledged for culturing the *Vibrio* strains. Provision of the LT-producing *Escherichia coli* strain by A. Siitonen from the National Institute for Health and Welfare, Helsinki, Finland and of the *Listonella anguillarum*, *Campylobacter upsaliensis*, *Campylobacter jejuni*, and *Bacillus anthracis* strains by the Finnish Food Safety Authority Evira, Helsinki/Oulu, Finland are also gratefully acknowledged. The *Brucella melitensis* strain was kindly provided by K. Rantakokko-Jalava from the Turku University Central Hospital, Turku, Finland; the *Francisella tularensis* strain by M. Forsman from the Swedish Defence Research Agency, Umeå, Sweden; and the *Salmonella typhimurium*, *Yersinia enterocolitica*, and *Yersinia pestis* strains by M. Skurnik from the Department of Bacteriology and Immunology, University of Helsinki, Helsinki, Finland. The authors also thank H. Hemmilä for DNA purification of the *Vibrio* strains. A Siitonen and S. Lukinmaa are acknowledged for critical reading of the manuscript.

References

Bhattacharya T, Chatterjee S, Maiti D, Bhadra RK, Takeda Y, Nair GB, Nandy RK (2006) Molecular analysis of the *rstR* and *orfU* genes of the CTX prophages integrated in the small chromosomes of environmental *Vibrio cholerae* non-O1, non-O139 strains. *Environ Microbiol* 8:526–534.

Blackstone GM, Nordstrom JL, Bowen MD, Meyer RF, Imbro P, DePaola A (2007) Use of a real time PCR assay for detection of the *ctxA* gene of *Vibrio cholerae* in an environmental survey of Mobile Bay. *J Microbiol Methods* 68:254–259.

Brayton PR, Tamplin ML, Huq A, Colwell RR (1987) Enumeration of *Vibrio cholerae* O1 in Bangladesh waters by fluorescent-antibody direct viable count. *Appl Environ Microbiol* 53:2862–2865.

Chomvarin C, Namwat W, Wongwajana S, Alam M, Thaew-Nonngiew K, Sinchaturus A, Engchanil C (2007) Application of duplex-PCR in rapid and reliable detection of toxigenic *Vibrio cholerae* in water samples in Thailand. *J Gen Appl Microbiol* 53:229–237.

Franco AA, Fix AD, Prada A, Paredes E, Palomino JC, Wright AC, Johnson JA, McCarter R, Guerra H, Morris Jr JG (1997) Cholera in Lima, Peru,

Fig. 1. Representative amplification plots from the developed real-time PCR ctxA assay to detect *V. cholerae* (VC 243) from spiked brackish seawater collected from the Gulf of Finland. Cells were separated from serial dilutions of *V. cholerae* in autoclaved seawater by centrifugation, followed by proteinase K (Roche) treatment and heat inactivation before PCR analysis with the (A) ABI 7300, (B) ABI 7900HT Fast, and (C) RAZOR instruments, as performed in 3, 3, and 2 replicates, respectively. According to these amplification plots, all replicates from suspensions 180 to 180 000 CFU/mL were detected with the ABI 7300 instrument (A). In this representative run with the ABI 7900HT Fast instrument, only 2 of 3 replicates were positive from the last serial dilution containing 180 CFU/mL, whereas, all higher concentrations showed reactivity in all 3 parallel runs (B). The RAZOR instrument displayed reactivity in 1 of 2 samples containing 18 CFU/mL of *V. cholerae* as template, and all other higher concentrations were positive (C). The final detection limit based on testing of 10 replicates with the ABI instruments was 1800 CFU/mL. Based on 8 reactions with the RAZOR instrument, the final limit of detection was determined as 180 CFU/mL. Data from the other runs are not shown. No template controls were prepared similarly as the spiked seawater samples, but without addition of bacteria.

correlates with prior isolation of *Vibrio cholerae* from the environment. *Am J Epidemiol* 146:1067–1075.

Fukushima H, Tsunomori Y, Seki R (2003) Duplex real-time SYBR green PCR assays for detection of 17 species of food- or waterborne pathogens in stools. *J Clin Microbiol* 41:5134–5146.

Fykse EM, Skogan G, Davies W, Olsen JS, Blatny JM (2007) Detection of *Vibrio cholerae* by real-time nucleic acid sequence-based amplification. *Appl Environ Microbiol* 73:1457–1466.

Gubala AJ (2006) Multiplex real-time PCR detection of *Vibrio cholerae*. *J Microbiol Methods* 65:278–293.

Gubala AJ, Proll DF (2006) Molecular-beacon multiplex real-time PCR assay for detection of *Vibrio cholerae*. *Appl Environ Microbiol* 72:6424–6428.

Lipp EK, Rivera IN, Gil AI, Espeland EM, Choopun N, Louis VR, Russek-Cohen E, Huq A, Colwell RR (2003) Direct detection of *Vibrio cholerae* and *ctxA* in Peruvian coastal water and plankton by PCR. *Appl Environ Microbiol* 69:3676–3680.

Lyon WJ (2001) TaqMan PCR for detection of *Vibrio cholerae* O1, O139, non-O1, and non-O139 in pure cultures, raw oysters, and synthetic seawater. *Appl Environ Microbiol* 67:4685–4693.

Rivera IN, Chun J, Huq A, Sack RB, Colwell RR (2001) Genotypes associated with virulence in environmental isolates of *Vibrio cholerae*. *Appl Environ Microbiol* 67:2421–2429.

Sack DA, Sack RB, Nair GB, Siddique AK (2004) Cholera. *Lancet* 363:223–233.

Singh DV (2001) Molecular analysis of *Vibrio cholerae* O1, O139, non-O1, and non-O139 strains: clonal relationships between clinical and environmental isolates. *Appl Environ Microbiol* 67:910–921.

Skottman T, Piiparinen H, Hyytiainen H, Myllys V, Skurnik M, Nikkari S (2007) Simultaneous real-time PCR detection of *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis*. *Eur J Clin Microbiol Infect Dis* 26:207–211.

World Health Organization (2008) Cholera, fact sheets no 107. Geneva: WHO.

II

Eur J Dermatol 2013; 23(1): 53-8

Jaana RENKO[1, a]
Katja A. KOSKELA[2, a]
Paul W. LEPP[3]
Niku OKSALA[4]
Mari LEVULA[5]
Terho LEHTIMÄKI[5]
Tiina SOLAKIVI[1]
Tarja KUNNAS[1]
Simo NIKKARI[2, 6]
Seppo T. NIKKARI[1]

[1] Department of Medical Biochemistry,
[2] CB Defence and Environmental Health Centre,
Centre for Military Medicine,
Helsinki, Finland
[3] Department of Biology,
Minot State University,
Minot, ND, USA
[4] Department of Surgery,
[5] Department of Clinical Chemistry,
University of Tampere School of Medicine,
33014 Tampere, Finland
[6] Department of Medical Microbiology and Immunology,
University of Turku,
Turku, Finland
[a] The first two authors share equal authorship

Reprints: S. Nikkari
<simo.nikkari@mil.fi>

Article accepted on 10/9/2012

Bacterial DNA signatures in carotid atherosclerosis represent both commensals and pathogens of skin origin

Infectious agents have been suggested to be involved in atherosclerosis. By using a novel subtraction broad-range PCR approach, we defined bacterial DNA signatures in surgically removed sterile carotid artery endarterectomy plaques of patients with carotid atherosclerosis. Eighty partial bacterial 16S rDNA nucleotide sequences from eight patients were studied. Furthermore, 34 clones representing 21 bacterial sequence-types from the reagents used for DNA extraction and PCR amplification were determined. After subtraction of these potential methodological contaminants, 23 bacterial sequence-types were considered as clinically relevant findings. The most prominent phylum, *Actinobacteria,* accounted for 74 % of these relevant sequences. Furthermore, according to the Human Microbiome project database, interestingly, nearly all (94%) of the sequences were associated with the human skin microbiome.

Key words: atherosclerosis, bacterial DNA, PCR, chronic inflammatory disease, *Actinobacteria*, skin microbiome

Endothelial dysfunction and the consequential inflammatory response are now generally accepted as essential mechanisms of atherogenesis. Infectious agents have been studied extensively as a possible cause of this vascular disease. In particular, microbial agents such as *Chlamydophila pneumoniae*, *Helicobacter pylori*, cytomegalovirus, herpes simplex virus and bacteria involved in dental infections have been implicated in the initiation and progression of atherosclerosis [1]. However, confirmation that infections caused by specific pathogens play causal roles in atherosclerosis has not been established. In fact, recent studies suggest a hypothesis of multiple bacterial colonization in arterial lesions [2-4].

We used the broad-range bacterial 16S rDNA PCR methodology (Br-PCR) to study which bacterial signatures are found in surgically sterile atherosclerotic plaque samples of patients with symptomatic carotid stenosis. Potentially this methodology enables identification of all known bacterial species, as well as detection of previously uncharacterized bacteria. Cloning and nucleotide sequencing of the PCR products, followed by subtraction analysis of potentially non-relevant bacterial sequences seen in enhanced control reagents, were used to reveal the clinically relevant sequence-types that were further compared with entries in the Human Microbiome project database.

Materials and methods

Carotid endarterectomies were performed on eight patients (A-H) with hemodynamically significant (>70%) and symptomatic atherosclerotic plaque (4 men and 4 women; age 59-83 years,). None of the patients was known to have any clinical signs of infection in the days before the surgery. After surgical removal, the carotid artery tissues (intima and inner media) were immediately soaked in RNALater solution (Ambion Inc., Austin, USA). For histochemistry, one section of each specimen was fixed in 4% paraformaldehyde in PBS, embedded in paraffin and cut and stained with hematoxylin-eosin. According to the morphological findings, all samples were from advanced atherosclerotic plaques. All patients considered for the study gave written informed consent before surgery, and the study was approved by the ethics committee of Tampere University Hospital.

For DNA analysis, carotid artery plaque samples adjacent to the paraffin-embedded sections were prepared as described earlier [4]. Success in DNA extraction was verified by the ability to amplify a 290-bp human β-actin gene sequence by PCR. Br-PCR was performed as described earlier by using primers 16S1RR-1 and fD1mod (corresponding to positions 8-27 and 575-556 in the *E. coli* 16S

doi:10.1684/ejd.2012.1908

To cite this article: Renko J, Koskela KA, Lepp PW, Oksala N, Levula M, Lehtimäki T, Solakivi T, Kunnas T, Nikkari S, Nikkari ST. Bacterial DNA signatures in carotid atherosclerosis represent both commensals and pathogens of skin origin. *Eur J Dermatol* 2013; 23(1): 53-8 doi:10.1684/ejd.2012.1908

rRNA gene) [5]. To assess potential background sequences from the reagents used for DNA extraction and PCR amplification, two types of negative controls were used. The DNA extraction control contained all other reagents except the carotid samples. In addition to these standard PCR controls, enhanced negative controls were created by amplification for 12 additional PCR cycles [4]. Cloning of the PCR products was performed using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA). Individual colonies were amplified directly by PCR using the M13 forward (-20) and reverse priming sites. Ten resulting amplicons of the expected size from each biopsy clone library were sequenced with the automated ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) using BigDye Terminator Cycle sequencing chemistry (Applied Biosystems). This was followed by manual editing and alignment using the Chromas 2.31 (Technelysium, Eden Prairie, MN) and ClustalW sequence analysis softwares. The BLAST search tool (National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/BLAST/) was used for identification and comparison of the obtained sequences to previously published sequence-types. Sequence-types were further compared to entries in the Human Microbiome Project (HMP) database (http://www.hmpdacc.org/resources/blast.php). One aim of HMP is to define the microbial communities found at different sites on the human body, including skin, blood and nasal passages, as well as the gastrointestinal and urogenital tracks. During our study this reference genome database contained 4439, 2624, 2970, 1363, 1725 entries from each site, respectively.

In the literature, there is no universal agreement on the degree of sequence divergence acceptable within species or genus. Suggested cut-off values based on whole 16S rDNA sequences are 97% sequence homology for members of same genus and 99-100% similarity for members of same species [6, 7]. This was used as a guideline in assigning a genus or species name to a partial 16S rDNA sequence in our present study.

Bacterial sequences of positive patient samples and the enhanced negative controls were included in the systematic phylogenetic analysis. Initial alignment of amplified sequences was performed as published earlier [4]. Briefly, the automated 16S rRNA sequence aligner of the ARB software was run against a database of 102,134 complete and partial rRNA sequences followed by manual alignment of ambiguously and incorrectly aligned positions on the basis of conserved primary sequence and secondary structure. By using a maximum-likelihood algorithm, the phylogenetic associations were determined from 495 masked positions. A least-squares fit of Jukes-Cantor corrected evolutionary distances and maximum parsimony algorithms were used to confirm these associations.

Results

From each of the eight carotid endarterectomy sample libraries derived from subjects A – H, ten clones were sequenced. These represented 37 different sequences. Also, 34 clones from the enhanced control clone libraries were sequenced, representing 21 different sequences. A total of 14 sequences from the patient samples (from subjects C, H and E) that shared over 99% similarity with sequences from the enhanced controls were omitted from the patient analysis, as their potential origin as methodological contaminants could not be ruled out. The omitted sequences included all sequence-types from subject H. Furthermore, only non-bacterial mitochondrial and plastid sequences were identified in samples from subjects A and D. After this subtraction analysis, the remaining 23 bacterial sequences from five patients belonged to three phyla and eight genera (*table 1*). The most prominent phylum in the carotid specimens was *Actinobacteria*, accounting for 74% (17/23) of clones from five clinical specimens. *Proteobacteria* and *Firmicutes* were the second and third most prevalent phyla composing 17% (4/23) and 9% (2/23) of the clones. Members of eight genera comprised 96% (22/23) of the carotid clones analyzed. The genera were: *Micrococcus* (52%; 12/23), *Dolosigranulum* (9%; 2/23), *Corynebacterium* (9%; 2/23), *Bradyrhizobium* (9%; 2/23), *Brachybacterium* (4%; 1/23), *Propionibacterium* (4%; 1/23), *Rhodococcus* (4%; 1/23), and *Stenotrophomonas* (4%; 1/23) (*table 1*). Based on the phylogenetic analysis of the partial 16S rDNA sequences, a total of nine species were identified from the endarterectomy samples (*table 1*). Only one of these sequences did not match with any previously published sequence of a known bacterial species.

In summary, nine bacterial species-specific sequence-types and one sequence-type representing a putatively new species (uncultured bacterium GQ007349) were detected in the five carotid specimens (*table 1*). The 21 potential non-relevant sequence-types representing 14 species that were present both in the enhanced control PCR products and patient samples are listed in *table 2*. Phylogenetic relationships among the atherosclerotic associated bacterial 16S rDNA sequences are shown in *figure 1*.

When the 23 relevant sequence-types were analyzed against entries in the Human Microbiome Project (HMP) database [8], 16 (70%) were found to have relevance (>97% similarity) with published entries and nearly all of those (94%, 15/16) were of skin origin. The results of this comparison are shown in *table 3*.

Discussion

In the present study we investigated carotid artery atherosclerotic plaques to determine the possible presence of bacterial DNA signatures in these samples. Earlier, we studied bacterial diversity from lesions of coronary artery disease, [3] abdominal atherosclerosis [4] and temporal arteritis [9] and found the presence of both commensal and known human pathogenic bacteria in these atherosclerotic sites. However, we had not previously been able to determine a predominant site in the human body as the source of these sequences.

The bacterial sequence-types that we identified in five carotid artery specimens after subtraction analysis represented three phyla, eight genera, and nine species. Eight of the species determined in our study were closely related to previously described human pathogens. Additionally, one sequence-type was related to bacteria without any published clinical association, and one was potentially a new species.

Table 1. Bacterial sequence-types detected from the carotid artery plaques, and characteristics of these carotid atherosclerosis cases.

Subject	Age/Sex*	Sequence no.	Phyla	Bacterial sequence type	Similarity to sequences submitted to GenBank	GenBank accession number
B	59/F	B2, B5, B9	Actinobacteria	*Micrococcus luteus*	99.8% (503/504 bp)	AF057289
		B1,B3,B10	Actinobacteria	*Micrococcus sp.*	99.2%-99.6% (500/504 bp – 502/504 bp)	FJ015031
		B6, B7, B8	Actinobacteria	*Micrococcus sp.*	99.8%-100% (503/504 bp – 504/504 bp)	EU446199
C	69/M	C2	Actinobacteria	*Propionibacterium acnes*	99.8% (505/506 bp)	AB097215
		C3	Proteobacteria	*Stenotrophomonas maltophilia*	100% (528/528 bp)	AJ293470
E	83/M	E1	Actinobacteria	*Brachybacterium muris*	99.4% (502/505 bp)	NR_024571
		E3, E5	Actinobacteria	*Corynebacterium mucifaciens*	100% (502/502 bp)	AF537599
		E4, E9	Proteobacteria	*Bradyrhizobium elkanii*	99.8% (472/473 bp)	FJ025104
		E6	Actinobacteria	*Micrococcus sp.*	99.8% (503/504 bp)	EU446199
		E7	Actinobacteria	*Micrococcus luteus*	99.0% (499/504 bp)	AJ717368
F	78/M	F4, F6	Firmicutes	*Dolosigranulum pigrum*	99.6%-99.8% (530/532 bp – 521/532 bp)	GU395995
		F7	Actinobacteria	*Micrococcus luteus*	99.4% (501/504 bp)	AJ536198
G	73/M	G1	Actinobacteria	*Rhodococcus erythropolis*	99.8% (499/500 bp)	EF491951
		G7	Proteobacteria	Uncultured bacterium	99.6% (526/528 bp)	GQ007349

**M = male, F = female*

Although micrococci (subject B, E and F) are common environmental bacteria and primarily non-pathogenic commensals of human microflora, *Micrococcus luteus* (subjects B, E, and F) is an opportunistic pathogen that has been reported as a causative agent of septic shock and prosthetic valve endocarditis [10, 11]. *Propionibacterium acnes* (subject C) is also a commensal of human skin and an opportunistic pathogen. It has been associated with several cases of pacemaker-associated infective endocarditis [12-14]. The presence of both *P. acnes* and *Bradyrhizobium elkanii* (subject E) nucleic acid sequences have previously been detected from abdominal aortic aneurysms by direct 16S rDNA PCR amplification [15], as in our current study. Furthermore, *P. acnes* has been isolated from a mycotic aortic aneurysm by bacterial culture [16], whereas *Bradyrhizobia* are environmental plant root nodule bacteria that have not been isolated in man by classic methodologies [15]. *Dolosigranulum pigrum* (subject F) is a rare but emerging gram-positive opportunistic pathogen that has been isolated from patients with nosocomial pneumonia and ventilator-associated pneumonia [17, 18]. Bloodstream infection caused by *Rhodococcus erythropolis* (subject G) in a cancer patient was recently reported, adding the published cases of human infection with *R. erythropolis* to a total of five. Identification of *R. erythropolis* by Baba and co-workers was performed by 16S rDNA sequencing of isolates grown in aerobic blood culture bottles [19]. *Stenotrophomonas maltophilia* (subject C) is a pathogen that is associated with serious nosocomial infections such as bloodstream infection and pneumonia, especially in debilitated patients [20-22]. Furthermore, *S. maltophilia* can cause foreign body (pacemaker, prosthetic valve and central venous catheter) related endocarditis [23-25]. *Corynebacterium mucifaciens* (AF537599, subject E) has frequently been isolated from human blood cultures and in other clinical material [26]. *Brachybacterium muris* (subject E) has been isolated from the liver of a laboratory mouse strain [27] and has not previously been detected in man.

Only one of the 23 (4%) bacterial sequence-types found in this study was previously uncharacterized. Because little is known about its closest relatives, the biological and clinical significance of this putative pathogen is unclear. However, this sequence-type (GQ007349) has previously been described as a member of the human skin microbiome [28].

It is evident that in the current study Br-PCR detected DNA of bacteria commonly considered to be of less clinical importance. Sequence-types of the more widely studied atherogenic species such as *Chlamydophila pneumoniae* and *Helicobacter pylori* [1] were not observed. Earlier studies performed with species-specific primers have reported the presence of *C. pneumoniae* DNA in samples of atherosclerotic arteries [29], which indicates that, when using specific DNA primers, the sensitivity of the

Table 2. Potential non-relevant contaminants that were present in the DNA extraction control PCR products.

Bacterial sequence-type	GenBank accession number*
Uncultured bacterium	FJ893810, FJ893882, GQ029598, GQ03810, GQ111622, EU344712 [93.2%], EU768130 [94.6%]
Achromobacter xylosoxidans subsp. xylosoxidans	AF225979
Actinobaculum sp.	AY207066 [95.5%]
Actinomyces dentalis	NR_025633
Actinomyces odontolyticus	GQ131411
Actinomyces naeslundii	AJ234055
Actinomyces sp.	AF385521
Corynebacterium vitarumen	X84680
Corynebacterium durum	AF537593
Corynebacterium sp.	FJ269041
Lautropia mirabilis	X97652
Microbacterium barkeri	NR_026164 [97.3%]
Mobiluncus curtisii	AJ576084 [97.3%]
Rothia sp.	AJ131122
Streptococcus mitis	GU045389

**Accession number of the nearest 16S rDNA sequences is given for each strain. Percentage of 16S rDNA sequence similarity is indicated in brackets, given only for values <98%*

PCR may be higher. Therefore, we can not rule out the possible presence of these pathogens in our carotid artery plaque samples.

Reagents used in DNA extraction and PCR may contain bacterial degradation products or traces of DNA and under certain experimental conditions these background causing sequences may become detectable after PCR amplification. Already in 1998, Tanner *et al.* reported bacterial contaminant sequences in PCR reagents [30]. These contaminants have been speculated to derive from various

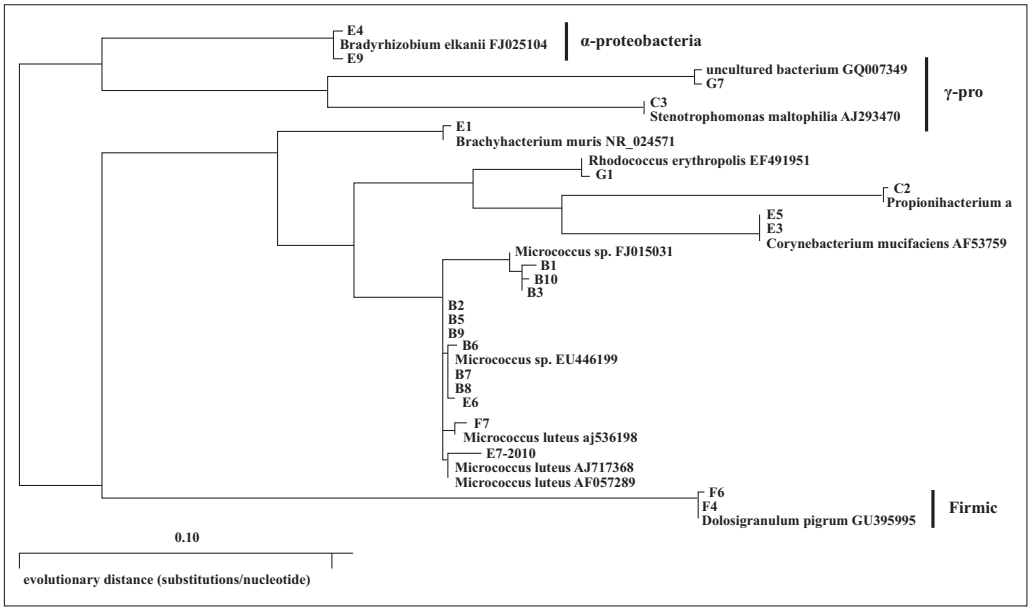


Figure 1. Phylogenetic relationships of atherosclerosis associated bacterial 16S rDNA sequences.
Phylogenetic relationships were inferred from 495 masked positions using a maximum-likelihood algorithm. Subject identification and sequence number are marked at the end of the lines. Names of the nearest reference sequences are given.

Table 3. Comparison of sequence-types from carotid artery plaques to entries in the Human Microbiome Project (HMP) sequence database. In the table, only sequence similarities exceeding 97% are shown.

	GenBank	Human Microbiome Project			
Sequence no.	**Bacterial sequence type**	**Bacterial sequence type**	**Similarity to sequences in Human Microbiome project**	**Accession number**	**source**
B2, B5, B9	*Micrococcus luteus*	*Micrococcus luteus*	99.8% (503/504 bp)	ADCD01000049	skin
B1,B3,B10	*Micrococcus sp.*	*Micrococcus luteus*	97.2%-97.6% (490/504 bp-492/504 bp)	ADCD01000049	skin
B6, B7, B8	*Micrococcus sp.*	*Micrococcus luteus*	99.4%-99.6% (501/504 bp-502/504 bp)	ADCD01000049	skin
C2	*Propionibacterium acnes*	*Propionibacterium acnes*	99.8% (505/506 bp)	ADWC01000017	skin
C3	*Stenotrophomonas maltophilia*	no relevant similarity found			
E1	*Brachybacterium muris*	no relevant similarity found			
E3, E5	*Corynebacterium mucifaciens*	*Corynebacterium genitalium*	97.0% (491/506 bp)	CM000961	urogenital
E4, E9	*Bradyrhizobium elkanii*	no relevant similarity found			
E6	*Micrococcus sp.*	*Micrococcus luteus*	99.4% (501/504 bp)	ADCD01000049	skin
E7	*Micrococcus luteus*	*Micrococcus luteus*	99.0% (499/504 bp)	ADCD01000049	skin
F4, F6	*Dolosigranulum pigrum*	no relevant similarity found			
F7	*Micrococcus luteus*	*Micrococcus luteus*	98.8% (498/504 bp)	ADCD01000049	skin
G1	*Rhodococcus erythropolis*	*Rhodococcus erythropolis*	98.4% (494/502 bp)	ACNO01000030	skin
G7	Uncultured bacterium	no relevant similarity found			

components of the PCR reagents, including oligonucleotides, water, or other buffers used in PCR [31-33]. Attempts to reduce the amount of contaminant DNA in PCR reagents have been published. Unfortunately, these methods have been observed to reduce amplification efficiency of the PCR reaction [34, 35]. In our study the AmpliTaq polymerase LD (Low DNA) enzyme was used to minimize potential bacterial contaminants deriving from the polymerase enzyme. According to the manufacturer, this enzyme is highly purified to significantly reduce bacterial DNA sequences commonly present in recombinant protein preparations (Applied Biosystems, Foster City, CA). To rule out these potentially non-relevant sequence-types we generated clone libraries from enhanced negative controls and subtracted these bacterial sequence-types from those derived from the clinical samples [4]. These DNA extraction controls contained all other reagents except the carotid specimens during DNA extraction. The controls were optimal for our study design as, due to ethical and technical constraints, it would have been impossible to obtain standardized samples from healthy subjects, particularly as atherosclerosis starts early on in life. Sixteen of the 34 contaminant sequence-types shared significant similarity (>97%) with entries in the HMP database. In huge contrast to findings with the relevant sequence-types, nearly all of which were related to skin-derived sequence-types, only one of these 16 defined contaminant sequence-types was skin-related in the HMP database and all other sequence-types were linked with other body sites, e.g. the urogenital and oral tracks, as well as blood. However, we speculate that, as in our current study, these sequences may represent technical reagent contaminants in the HMP database as well.

Profiles of bacterial 16S rDNA sequences varied between individuals. According to Ott and his colleagues [2] this could point to additional host mechanisms of infection control that have to be considered in the pathophysiology of atherosclerosis. The most prominent phylum in the carotid specimens was *Actinobacteria*, accounting for 74% of cloned sequences from five clinical specimens. When we compared the sequence-types detected in our current study with those from our previous study on atherosclerotic lesions [4], some congruence can be seen. In both studies *Actinobacteria* (*Corynebacteria* and *Micrococci*) predominate.

Based on our current findings, *Actinobacteria* may be one of the etiologic agents in the initiation and progression of atherosclerosis. Although *Actinobacteria* may cause disease, as outlined above, they are normal commensals of human microflora and constitute more than a third of the healthy human microbiota. Interestingly, when comparing the 23 representative sequence-types to those published in the HMP database, we discovered that, of those sequence-types with adequately matching similarity to database entries (>97%), nearly all (15/16, 94%) belonged to the skin microbiome. Additionally, the one uncharacterized

sequence-type (GQ007349) was also recently recognized to be of skin origin. However, this sequence-type was not listed in the HMP project database.
The skin is the largest human organ and chronic inflammatory diseases present with a myriad of cutaneous manifestations. Conversely, the skin microbiome may have a role in the etiology of atherosclerosis. Whether the bacteria enter the carotid plaques as whole viable bacteria or e.g. phagocytosed bacterial compounds and fragments of DNA remains unsolved. ■

Disclosure. *Financial support: This study was supported by grants from Competitive research funding of the Pirkanmaa Hospital District and the Academy of Finland. Conflict of interest: none.*

References

1. Prasad A, Zhu J, Halcox JP, Waclawiw MA, Epstein SE, Quyyumi AA. Predisposition to atherosclerosis by infections: role of endothelial dysfunction. *Circulation* 2002; 106: 184-90.
2. Ott SJ, El Mokhtari NE, Musfeldt M, *et al*. Detection of diverse bacterial signatures in atherosclerotic lesions of patients with coronary heart disease. *Circulation* 2006; 113: 929-37.
3. Lehtiniemi J, Karhunen PJ, Goebeler S, Nikkari S, Nikkari ST. Identification of different bacterial DNAs in human coronary arteries. *Eur J Clin Invest* 2005; 35: 13-6.
4. Renko J, Lepp PW, Oksala N, Nikkari S, Nikkari ST. Bacterial signatures in atherosclerotic lesions represent human commensals and pathogens. *Atherosclerosis* 2008; 201: 192-7.
5. Nikkari S, Lopez FA, Lepp PW. Broad-range bacterial detection and the analysis of unexplained death and critical illness. *Emerg Infect Dis* 2002; 8: 188-94.
6. Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* 2000; 38: 3623-30.
7. Janda JM, Abbott SL. Bacterial identification for publication: when is enough enough? *J Clin Microbiol* 2002; 40: 1887-91.
8. Peterson J, Garges S, Giovanni M. The NIH Human Microbiome Project. *Genome Res* 2009; 19: 2317-23.
9. Renko J, Kalela A, Karhunen PJ. Do temporal arteritis lesions contain bacterial DNA? *Eur J Clin Invest* 2003; 33: 657-61.
10. Albertson D, Natsios GA, Gleckman R. Septic shock with *Micrococcus luteus*. *Arch Intern Med* 1978; 138: 487-8.
11. Seifert H, Kaltheuner M, Perdreau-Remington F. *Micrococcus luteus* endocarditis: case report and review of the literature. *Zentralbl Bakteriol* 1995; 282: 431-5.
12. Zedtwitz-Liebenstein K, Gabriel H, Graninger W. Pacemaker endocarditis due to *Propionibacterium acnes*. *Infection* 2003; 31: 184-5.
13. Chakour M, Revel F, Godreuil C, Plotton C, Aubry A, Koeck JL. Infectious endocarditis due to *Propionibacterium acnes* on a mechanical heart valve and cardiac stimulator electrode. *Presse Med* 2002; 31: 1414.
14. Chua AG, Ding J, Schoch PE, Cunha BA. Pacemaker-induced endocarditis due to *Propionibacterium acnes*. *Clin Infect Dis* 1998; 27: 1541-2.
15. Marques da Silva R, Caugant DA, Eribe ER, *et al*. Bacterial diversity in aortic aneurysms determined by 16S ribosomal RNA gene analysis. *J Vasc Surg* 2006; 44: 1055-60.
16. Brook I, Frazier EH. Aerobic and anaerobic microbiology of mycotic aortic aneurysm. *Clin Infect Dis* 1999; 28: 928-9.
17. Hoedemaekers A, Schulin T, Tonk B, Melchers WJ, Sturm PD. Ventilator-associated pneumonia caused by *Dolosigranulum pigrum*. *J Clin Microbiol* 2006; 44: 3461-2.
18. Lécuyer H, Audibert J, Bobigny A, *et al*. *Dolosigranulum pigrum* causing nosocomial pneumonia and septicemia. *J Clin Microbiol* 2007; 45: 3474-5.
19. Baba H, Nada T, Ohkusu K, Ezaki T, Hasegawa Y, Paterson DL. First case of bloodstream infection caused by *Rhodococcus erythropolis*. *J Clin Microbiol* 2009; 47: 2667-9.
20. Aisenberg G, Rolston KV, Dickey BF, Kontoyiannis DP, Raad II, Safdar A. *Stenotrophomonas maltophilia* pneumonia in cancer patients without traditional risk factors for infection, 1997-2004. *Eur J Clin Microbiol Infect Dis* 2007; 26: 13-20.
21. Lai CH, Chi CY, Chen HP, *et al*. Clinical characteristics and prognostic factors of patients with *Stenotrophomonas maltophilia* bacteremia. *J Microbiol Immunol Infect* 2004; 37: 350-8.
22. Muder RR, Harris AP, Muller S, *et al*. Bacteremia due to *Stenotrophomonas (Xanthomonas) maltophilia*: a prospective, multicenter study of 91 episodes. *Clin Infect Dis* 1996; 22: 508-12.
23. Takigawa M, Noda T, Kurita T, *et al*. Extremely late pacemaker-infective endocarditis due to *Stenotrophomonas maltophilia*. *Cardiology* 2008; 110: 226-9.
24. Bayle S, Rovery C, Sbragia P, Raoult D, Brouqui P. *Stenotrophomonas maltophilia* prosthetic valve endocarditis: a case report. *J Med Case Reports* 2008; 2: 174.
25. Crum NF, Utz GC, Wallace MR. Stenotrophomonas maltophilia endocarditis. *Scand J Infect Dis* 2002; 34: 925-7.
26. Bernard KA, Munro C, Wiebe D, Ongsansoy E. Characteristics of rare or recently described *Corynebacterium* species recovered from human clinical material in Canada. *J Clin Microbiol* 2002; 40: 4375-81.
27. Buczolits S, Schumann P, Weidler G, Radax C, Busse HJ. *Brachybacterium muris* sp. nov., isolated from the liver of a laboratory mouse strain. *Int J Syst Evol Microbiol* 2003; 53: 1955-60.
28. Grice EA, Kong HH, Conlan S, *et al*. Topographical and temporal diversity of the human skin microbiome. *Science* 2009; 324: 1190-2.
29. Campbell LA, O'Brien ER, Cappuccio AL, *et al*. Detection of *Chlamydia pneumoniae* TWAR in human coronary atherectomy tissues. *J Infect Dis* 1995; 172: 585-8.
30. Tanner MA, Goebel BM, Dojka MA, Pace NR. Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl Environ Microbiol*. 1998; 64: 3110-3.
31. Böttger EC. Frequent contamination of Taq polymerase with DNA. *Clin Chem* 1990; 36: 1258-9.
32. Hughes MS, Beck LA, Skuce RA. Identification and elimination of DNA sequences in *Taq* DNA polymerase. *J Clin Microbiol* 1994; 32: 2007-8.
33. Goto M, Ando S, Hachisuka Y, Yoneyama T. Contamination of diverse nifH and nifH-like DNA into commercial PCR primers. *FEMS Microbiol Lett* 2005 M; 246: 33-8.
34. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarski EB, Fox AJ. Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. *J Clin Microbiol* 2000; 38: 1747-52.
35. Philipp S, Huemer HP, Irschick EU, Gassner C. Obstacles of Multiplex Real-Time PCR for Bacterial 16S rDNA: Primer Specifity and DNA Decontamination of Taq Polymerase. *Transfus Med Hemother* 2010; 37: 21-8.

III

VECTOR-BORNE AND ZOONOTIC DISEASES
Volume 17, Number 2, 2017

DOI: 10.1089/vbz.2016.1969

Metagenomic Evaluation of Bacteria from Voles

Katja A. Koskela,[1] Laura Kalin-Mänttäri,[1] Heidi Hemmilä,[1] Teemu Smura,[2] Paula M. Kinnunen,[1,3,*] Jukka Niemimaa,[4] Heikki Henttonen,[4] and Simo Nikkari[1]

Abstract

Voles (Arvicolinae, Rodentia) are known carriers of zoonotic bacteria such as *Bartonella* spp. and *Francisella tularensis*. However, apart from *F. tularensis*, the bacterial microbiome of voles has not previously been determined in Finland and rarely elsewhere. Therefore, we studied liver samples from 61 voles using 16S ribosomal RNA gene PCR analysis, followed by Sanger sequencing. Twenty-three of these samples were also studied with tag-encoded pyrosequencing. The samples originated from 21 field voles (*Microtus agrestis*), 37 tundra voles (*Microtus oeconomus*), and 3 bank voles (*Myodes glareolus*). With the more conventional 16S rDNA PCR analysis, 90 (33%) of the recovered 269 sequence types could be identified to genus level, including *Bartonella*, *Francisella*, *Mycoplasma*, *Anaplasma*, and *Acinetobacter* in 31, 15, 9, 9, and 9 sequences, respectively. Seventy-five (28%) matched best with sequences of uncultured bacteria, of which 40/75 could be classified to the order Clostridiales and, more specifically, to families Lachnospiraceae and Ruminococcaceae. Pyrosequencing from 23 samples revealed comparable and similar results: clinically relevant bacterial families such as Mycoplasmataceae, Bartonellaceae, Anaplasmataceae, and Francisellaceae were recognized. These analyses revealed significant bacterial diversity in vole livers, consisting of distinct and constant sequence patterns reflecting bacteria found in the intestinal gut, but including some known zoonotic pathogens as well. The molecular bacterial sequence types determined with the two different techniques shared major similarities and verified remarkable congruency between the methods.

Keywords: diagnostics, rodents, PCR, zoonotic

Introduction

Voles are carriers of zoonotic bacteria such as *Francisella tularensis* (Rossow et al. 2014c), *Bartonella* spp. (Buffet et al. 2012, 2013), *Anaplasma* spp. (Kallio et al. 2014), and *Mycoplasma* spp. (Brown et al. 2001). In Finland, tularemia (caused by *F. tularensis*) is an endemic disease, and voles are considered to play a role as amplification hosts preceding human epidemics (Rossow et al. 2014a).

16S rRNA gene sequencing has provided a strong alternative to traditional culture-based identification of bacteria (Weisburg et al. 1991) in clinical microbiology (Salipante et al. 2013) and medicine (Winglee et al. 2014), as well as in extensive projects such as the Human Microbiome Project (Human Microbiome Project Consortium 2012).

Broad-range 16S rDNA PCR (Br-PCR) is based on universal primers detecting conserved regions in the chromosome coding 16S rRNA genes (Schmidt and Relman 1994), followed by bacterial identification by Sanger sequencing of intervening variable and hypervariable DNA regions. Alternatively, 16S rRNA gene sequences can be retrieved by using tag-encoded primers targeting hypervariable regions, followed by the use of next-generation sequencing technology (*e.g.*, pyrosequencing) that generates hundreds of thousands of sequences in a single run (Shendure and Ji 2008). Both techniques enable not only identification of previously known bacterial species but also discovery of DNA sequences of previously uncharacterized bacteria. Recently, microbial populations have been studied using next-generation techniques to investigate, for example, ticks in Japan (Qiu et al. 2014) and voles in France (Razzauti et al. 2015).

The main objective of this work was to determine the vole microbiome present in livers and to identify potential zoonotic pathogens of vole communities in Finland. Other

[1]Centre for Military Medicine, Helsinki, Finland.
[2]Department of Virology, University of Helsinki, Helsinki, Finland.
[3]Defence Command Finland, Plans and Policy Division, Helsinki, Finland.
[4]Natural Resources Institute Finland, Vantaa, Finland.
*Current affiliation: Finnish Food Safety Authority Evira, Helsinki, Finland.

objectives were to compare the two screening methods, conventional Br-PCR and tag pyrosequencing, and to evaluate their suitability for investigating complex, polymicrobial bacterial DNA samples.

Materials and Methods

Sample collection

Altogether, 60 snap-trapped voles with splenomegaly originating from three locations in Finnish Lapland (Kolari, Pisavaara, and Pyhätunturi) and two locations in Central Finland (Konnevesi and Pieksämäki) were studied. Altogether, 21 field voles (*Microtus agrestis*), 36 tundra voles (*Microtus oeconomus*), and three bank voles (*Myodes glareolus*) were included (Table 1). In addition, one tundra vole from Lapland without splenomegaly served as a control. Three voles from Central Finland had been previously identified positive for *F. tularensis* by PCR (Rossow et al. 2014c). Liver was chosen as the optimal organ for screening based on our earlier studies (Rossow et al. 2014c).

Traps were set in the evening and collected soon after sunrise. During the trappings, either in September in Lapland or early May or October in Central Finland, temperatures remained between 0°C and 10°C. Therefore, voles caught in the morning were fresh. They were dissected or frozen immediately.

Ethics statement

No ethical permit was needed because the Finnish Act on the Use of Animals for Experimental Purposes (62/2006) and the Finnish Animal Experiment Board (16th May, 2007) do not classify snap-trapping as an animal experiment. However, a permit (23/5713/2001) for capturing protected species (as bycatch) was granted by the Finnish Ministry of the Environment; none of our target species belonged to the protected ones.

DNA extraction

DNA was extracted from about 20 mg of liver tissue with the Wizard Genomic DNA Purification Kit (Promega). Additionally, three negative controls containing nuclease-free water (W4502; Sigma-Aldrich) instead of sample were used to screen for bacterial DNA contamination in extraction reagents.

Conventional Br-PCR analysis

DNA from 61 vole samples was used for amplification of a partial bacterial 16S rRNA gene sequence (~550 bp, 8F-575R) with universal primers, fD1mod (Kotilainen et al. 1998) and 16S1RR-B (Wilbrink et al. 1998). To assess potential contaminant sequences from extraction and PCR reagents, two different contamination libraries were established: First, for investigating exogenous bacterial DNA in extraction reagents, three negative extraction controls were analyzed. Second, to study the presence of contaminating DNA in PCR reagents, nuclease-free water was used as template in no template control (NTC) reactions, utilizing two different reaction mixtures. To avoid false-positive results in PCR, careful preventive measures were taken (Kwok and Higuchi 1989, Lo and Chan 2006). (Materials and Methods section in the Supplementary Data; Supplementary Data are available online at www.liebertpub.com/vbz).

Br-PCR amplification reaction mixture in the volume of 50 μL contained Gene Amp® PCR Buffer (Life Technologies [LT]), 200 μM of each dNTP (LT), 0.4 μM of each primer, 4 mM $MgCl_2$ (LT), 2 U AmpliTaq Gold® DNA Polymerase LD (LT), 5 μL template, and nuclease-free water (W4502; Sigma-Aldrich). PCRs were performed using the DNA Engine (PTC-200) Peltier Thermal Cycler (Bio-Rad Life Sciences). Reaction conditions were 3 min at 94°C, followed by 30 cycles of amplification at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s, and the final extension step at 72°C for 10 min.

PCR products of the correct size were cloned to *Escherichia coli* vectors using the TOPO TA Cloning Kit according to the manufacturer's instructions (Invitrogen Corporation). Randomly selected individual colonies were amplified directly by PCR using the M13 priming sites. Amplicons of the expected size were sequenced using Applied Biosystems Dye Terminator (v.3.1) sequencing kit (LT) and reactions were run on 3100xl Capillary Sequence Analyzer (LT) at the University of Helsinki (Sequencing Core Facility). Sequences were edited with the Sequencher 5.1 program (Gene Codes Corporation).

In addition to BLAST analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi), sequence types were classified using the Ribosomal Database Project (RDP) naïve Bayesian Classifier, Version 2.6 (Wang et al. 2007), with 80% bootstrap threshold value. Chimeric sequences were detected with the DECIPHER tool (Wright et al. 2012). Sequences sharing

Table 1. Number of Voles by Species and Trapping Locality

Trapping location (trapping year)	*Number of samples per vole species*			*Total no.*
	Microtus agrestis, *field vole*	Microtus oeconomus, *tundra vole*	Myodes glareolus, *bank vole*	
Central Finland				
Konnevesi, 2009	10[a]			10
Pieksämäki, 2008	5			5
Lapland[b]				
Kolari, 2011	4	17[c]		21
Pisavaara, 2011		11	1	12
Pyhätunturi, 2011	2	9	2	13
Total no.	21	37	3	61

[a]Three voles highly positive for *Francisella tularensis.*
[b]Located in northern Finland.
[c]No splenomegaly in one vole.

over 99% similarity with contaminant sequences found from controls were omitted (Renko et al. 2013).

Tag pyrosequencing

Liver DNA samples from 22 voles with and one without splenomegaly were used for tag-encoded pyrosequencing analysis of the 16S rRNA gene (V1-V3 region, 27F-518R), as described previously (Hanski et al. 2012), except 35–45 cycles were used in PCR runs. Additionally, three negative extraction controls pooled together were studied. Sequencing of PCR products using the 454 platform was done at Institute of Biotechnology, University of Helsinki (Finland). The sample-specific sequences were uploaded into the RDP Classifier (release 11.2) (Wang et al. 2007) to identify the bacterial classes and genera, with 80% as the threshold value. Sequences present in the pooled negative extraction control were omitted from the final results. Samples were considered positive for a specific bacterial sequence type when the number of sequence reads exceeded 10.

Further sequence data analyzing was done using mothur (Schloss et al. 2009) according to workflow described by Schloss et al. (2011). Before analyzing the sequences, the tag, primer sequences, and low-quality sequences (*i.e.*, ambiguous nucleotides, homopolymers longer than eight nucleotides, average quality score less than 35) were removed. Sequences were aligned to Greengenes reference database (DeSantis et al. 2006); chimeras were detected using UCHIME algorithm (Edgar et al. 2011). Chimeric sequences and sequences present in the pooled negative extraction control were removed from the dataset. Cutoff of 0.03 was used for the clustering of operational taxonomic units (OTUs). Alpha diversity was estimated using Shannon and inverse Simpson indexes. The Yue and Clayton measure of dissimilarity was used for dendrogram construction and for the Principal Coordinate Analysis (PCoA).

Results

With conventional Br-PCR, 408 randomly selected clone sequences from the 61 vole liver samples were sequenced. Fourteen sequences of poor quality were excluded from the final results. In addition, one plastid sequence and one sequence deciphered as a chimera were removed from the final results. Furthermore, 12 and 96 of the sequences represented the host species, bank vole and field vole, respectively (Supplementary Table S1). These sequences and a total of 15 sequences that shared over 99% similarity with sequences from the contaminant libraries were omitted from the final results. Of the remaining 269 sequence types, 33% (90/269) were identifiable at least to genus level, including *Bartonella*, *Francisella*, *Mycoplasma*, *Anaplasma* (including *Ehrlichia phagocytophila*) (Dumler et al. 2001), and *Acinetobacter* in 31, 15, 9, 9, and 9 sequences, respectively (Table 2). Furthermore, 28% (75/269) matched best with uncultured bacteria. Sequence analysis with the RDP Classifier revealed that over half (40/75) of those were classified to the order Clostridiales and, more specifically, to families Lachnospiraceae and Ruminococcaceae (Supplementary Table S1). Interestingly, a noteworthy group of sequences comprising 33% (89/269) of all the data shared 89–93% similarity with *Mycoplasma* spp. and could only be classified as bacteria using RDP.

Altogether, 23 vole liver-derived DNA samples and a pooled extraction control sample were also analyzed using 16S rRNA gene pyrosequencing. Contaminant sequence types present in the pooled extraction control were omitted from the final results (Supplementary Table S2). Majority of the remaining reads represented Francisellaceae (28%) and Bartonellaceae (17%) (Fig. 1). Other common families were Anaplasmataceae (9%), Lachnospiraceae (7%), Ruminococcaceae (6%), Porphyromonadaceae (6%), and Mycoplasmataceae (4%). In addition, 11% of the sequence reads belonged to an artificial unclassified taxon comprising unidentifiable sequences. The amount of sequence reads obtained varied from 3580 reads to over 24 000 reads per sample (Fig. 2a). The relevant sequence types detected from the vole samples are presented in Figure 2b. In general, Br-PCR and pyrosequencing gave comparable results. This can be seen in Table 3, where the most common findings are compared. In 17 samples of 23, the most abundant findings were in congruence.

Using conventional Br-PCR, 27 cloned sequences from three negative extraction controls were obtained and these represented 11 sequence types (Supplementary Table S3). BLAST analysis revealed that 12 of the sequence types had 100% identity with *Propionibacterium acnes*. To determine DNA remnants present in PCR reagents, altogether 529 sequences from 19 NTCs were analyzed using conventional Br-PCR. Sequences were cloned and sequenced also from NTCs from which no amplification product was seen in the agarose gel (Supplementary Fig. S1). The main contaminant sequence types identified were *Lactococcus lactis*, uncultured *Burkholderiaceae* bacterium, *Phyllobacterium* sp., *Schlegelella* sp./*Leptotrix* sp., uncultured *Sulfurospirillum* sp., uncultured Ruminococcaceae bacterium, *Altererythrobacter* sp., *Propionibacterium* sp., and *Clostridium* sp. (Supplementary Table S4). With 454 pyrosequencing, the main bacterial findings in the pooled negative extraction control belonged to genera *Halomonas*, *Shewanella*, *Pseudomonas*, *Citrobacter*, *Bacteroides*, and *Enterococcus* (Supplementary Table S2).

Shannon and inverse Simpson diversity indexes for each sample are shown in Table 4. There were no statistically significant differences between bacterial diversities of different vole species. Heatmap of the relative abundance of bacterial taxons is shown in Supplementary Figure S2. No clear-cut differences between bacterial communities of *M. agrestis* and *M. oeconomus* were observed in the heatmap.

The Yue and Clayton measure of dissimilarity between the bacterial community structures was estimated and the resulting distance matrix was used for dendrogram construction (Supplementary Fig. S3) and for the PCoA. While parsimony and unweighted UniFrac methods suggested no statistically significant clustering in respect to vole species or the geographic region of the sampled voles, the weighted UniFrac method (that takes the branch lengths of the dendrogram into account) suggested differences between the clustering ($p < 0.001$).

PCoA plot is shown in Figure 3. Analysis of molecular variance (AMOVA) was conducted to determine if the spatial separation in the PCoA plot was statistically significant, that is, voles of the same species or voles sampled from the same location shared similar communities and similar abundance levels. There were no statistically significant differences between the bacterial communities of different vole species. However, AMOVA suggested differences when comparing voles captured in Pisavaara and Kolari.

Table 2. Taxonomic Distribution of Bacterial Sequence Types Detected from Vole Liver Tissue with Conventional Br-PCR Utilizing BLAST Search Tool

Phylum	*Class*	*Order*	*Family*	*Genus*	*Number of 16S rRNA gene sequences*	*Number of voles*
Proteobacteria						
	Alphaproteobacteria					
		Rhizobiales	Bartonellaceae	*Bartonella*	31	9
		Rickettsiales	Anaplasmataceae	*Anaplasma*	9	4
				Uncultured *alphaproteobacterium*	2	2
	Betaproteobacteria					
		Burkholderiales	Comamonadaceae	*Comamonas*	2	1
			Ralstoniaceae	*Ralstonia*	2	2
	Gammaproteobacteria					
		Thiotrichales	Francisellaceae	*Francisella*	15[a]	3
		Pseudomonadales	Moraxellaceae	*Acinetobacter*	9	6
			Pseudomonadaceae	*Pseudomonas*	2	2
		Enterobacteriales	Enterobacteriaceae	*Buttiauxella*	2	1
		Oceanospirillales	Halomonadaceae	Uncultured *Chromohalobacter*	2	2
	Epsilonproteobacteria					
		Campylobacterales	Helicobacteraceae	*Helicobacter*	1	1
Firmicutes						
	Mollicutes					
		Mycoplasmatales	Mycoplasmataceae	*Mycoplasma*	9	5
	Clostridia					
		Clostridiales	Ruminococcaceae	Uncultured *Ruminococcus*	1	1
			Eubacteriaceae	Uncultured *Eubacterium*	1	1
				Uncultured *Firmicutes* bacterium	9[b]	7
Actinobacteria						
	Actinobacteria					
		Actinomycetales	Corynebacteriaceae	Uncultured *Corynebacterium*	1	1
			Micrococcaceae	*Micrococcus*	1	1
			Brevibacteriaceae	*Brevibacterium*	1	1
			Dermabacteraceae	*Brachybacterium*	1	1

Undetermined and nonbacterial sequence types[c]	*Number of 16S rRNA gene sequences*	*Number of voles*
Uncultured bacterium	39[b]	15
Uncultured organism	2	1
Closest match: *Mycoplasma* spp. (89–93%)	89	17
Closest match: uncultured bacterium (76–96%)	36[b]	20
Closest match: uncultured *Firmicutes* bacterium (95%)	2	1

[a]Sequences derived from voles highly positive for *F. tularensis* (Rossow et al. 2014c).
[b]Six sequences derived from the control vole without splenomegaly: two uncultured *Firmicutes* bacteria, two uncultured bacteria, two closest match: uncultured bacteria (94–95%).
[c]Percentage of sequence similarity to previously published sequences is indicated in brackets, given only for values <97%.

Finally, nonparametric T-test (White et al. 2009) was used to determine whether there are differently represented OTUs between vole species or between *M. oeconomus* voles captured in Pisavaara and Kolari. Between the vole species, this analysis suggested differences in the abundance of *Francisella* ($p = 0.0439$), unclassified *Firmicutes* ($p = 0.0409$), Clostridiales ($p = 0.0010$), and Lachnospiraceae ($p = 0.0048$). Between *M. oeconomus* voles captured in Pisavaara and Kolari, the analysis suggested different abundance levels for *Oscillibacter* ($p = 0.0343$) and *Mycoplasma* ($p = 0.0060$).

Discussion

We identified bacterial sequence types in vole liver samples using two metagenomic techniques targeting 16S rRNA genes: conventional Br-PCR and tag pyrosequencing. We also compared the two methods and studied their suitability for investigating complex, polymicrobial bacterial DNA samples.

Sixty-one vole samples were investigated using the more conventional Br-PCR analysis. The main clinically relevant

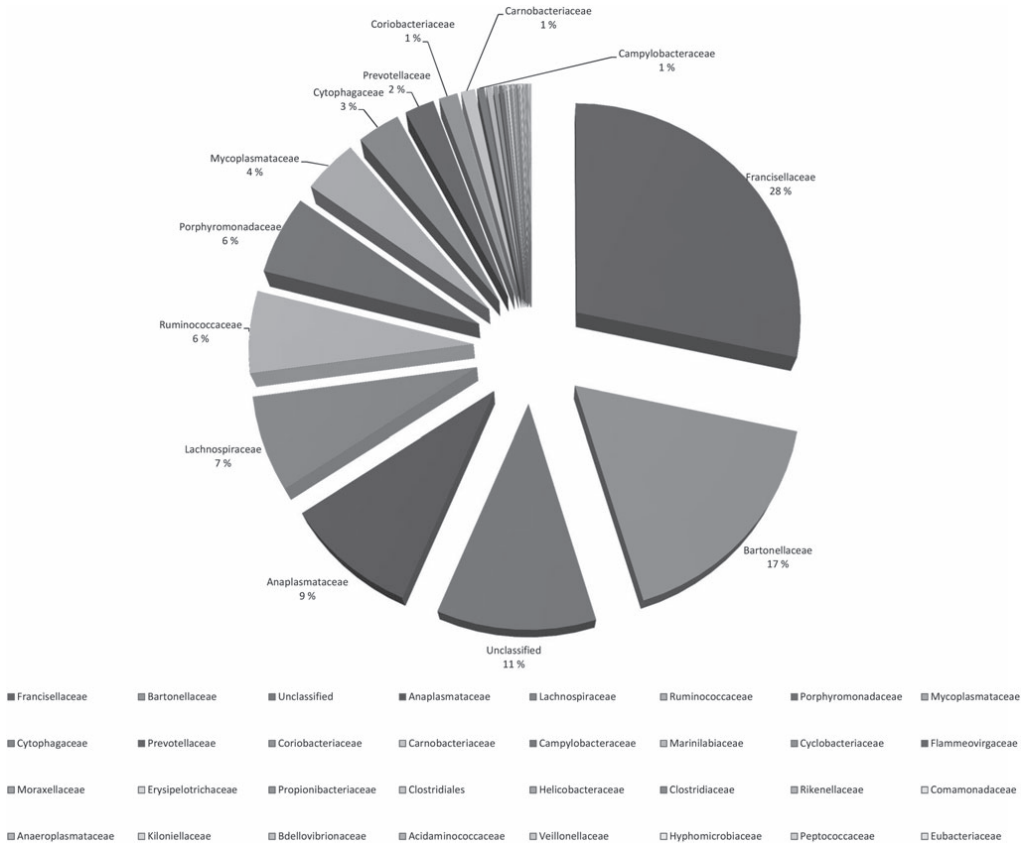


FIG. 1. Relative percentage of bacterial families from 23 vole DNA samples based on the number of sequence reads with pyrosequencing. Contaminant sequences (*Halomonas*, *Shewanella*, and contaminant sequences based on our negative extraction control) and sequences in which the total number of reads per bacterial family was fewer than one hundred are omitted. Only bacterial families with read percentages >1% are indicated.

bacterial sequences detected belonged to the genera *Bartonella*, *Francisella*, *Mycoplasma*, *Anaplasma*, and *Acinetobacter.* Based on these findings, 23 samples were selected for pyrosequencing. These included one sample from the control vole without splenomegaly and three samples previously identified positive for *F. tularensis* (Rossow et al. 2014c), as well as an extraction control.

Using pyrosequencing, the bacterial composition of four samples lacked any clinically relevant pathogens, containing mainly members of the families Ruminococcaceae, Porphyromonadaceae, and *Lachnospiraceae*, in addition to an artificial unclassified taxon (Fig. 2b and Supplementary Table S1). This bacterial pattern was very similar to the control vole's microflora and was perceived as intestinal normal flora. Sixteen samples, however, had a completely distinctive bacterial profile, with families such as Mycoplasmataceae, Bartonellaceae, Anaplasmataceae, or Francisellaceae being highly predominant. In the two remaining samples, Bartonellaceae was present in one sample according to the conventional Br-PCR, but pyrosequencing revealed mainly sequence types belonging to Corynebacteriaceae and Moraxellaceae families. In addition, vole TUL25, which had sequence types sharing 89–93% similarity with *Mycoplasma* spp., with conventional Br-PCR, gave an unclear bacterial profile with pyrosequencing and hence almost 90% of the reads belonged to the unclassified group. The results of three *Francisella*-positive samples corresponded with the findings of Rossow et al. (2014c): sequences of the highly positive voles, TUL32 and TUL33, were dominated by Francisellaceae, whereas one vole (TUL37), exhibiting lower positivity in *F. tularensis*-specific PCR, also contained, for example, Anaplasmataceae besides Francisellaceae. Overall, pyrosequencing analysis of 21 liver samples gave congruent results to those of the Br-PCR (Table 3).

Mycoplasma spp., more specifically *M. moatsii*- and rarely *M. penetrans*-specific sequence types, were found moderately often in voles from both Central Finland and Lapland. *Mycoplasma* infection is a leading cause of pneumonia

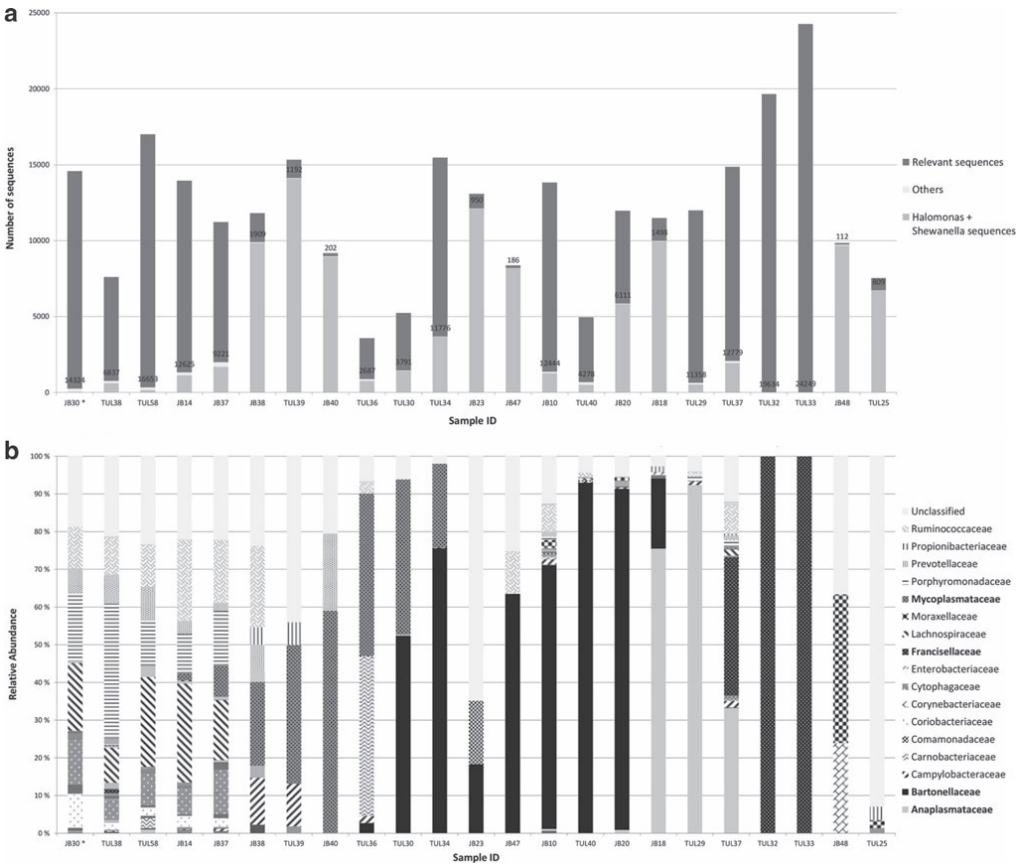


FIG. 2. Sequence types detected from vole liver tissue with pyrosequencing. **(a)** Total number of reads per sample, including relevant sequences, contaminant sequences (*Halomonas* and *Shewanella*), and other sequences consisting of other contaminant sequences based on the negative extraction control, as well as sequences in which the number of reads was fewer than 10 per sample. The number of relevant sequence types varied from 112 to 24,249 reads per sample. **(b)** Relative abundance of relevant sequence types represented at family level. *Control vole (no splenomegaly visible).

worldwide in humans (Ishiguro et al. 2013) and other species, including laboratory rodents (Lindsey and Cassell 1973, Sun et al. 2013). Wild rodents commonly carry *Mycoplasma* spp. (Sashida et al. 2013, Sumithra et al. 2013), but reports on clinical disease are almost nonexistent. *M. moatsii* may colonize intestines of wild rats (Giebel et al. 1990). More detailed studies on *Mycoplasma* in wild rodents are certainly needed.

Bartonella spp. were commonly seen in the voles. The animal reservoir for bartonellae is large, including rodents such as rats, mice, and voles (Krauss et al. 2003, Buffet et al. 2012, 2013, Hayman et al. 2013, Silaghi et al. 2016). Human bartonellosis in Finland is rare, but may be life-threatening (Jalava et al. 1995). Our findings included both undetermined (spp.) sequences and sequence types representing known species, such as *Bartonella rochalimae*, *Bartonella taylorii*, *Bartonella grahamii*, and *Bartonella henselae*. All these have been found in wild rodents (Krauss et al. 2003); the first three (or their close relatives) are common findings in bank and field voles (Buffet et al. 2012, 2013).

Numerous sequences belonged to Anaplasmataceae and *Anaplasma.* Within this genus, *Anaplasma phagocytophilum* infections are well known in the bank vole, field vole, common vole (*Microtus arvalis*), and common shrew (*Sorex araneus*), but these small mammals harbor a subtype different from those causing tick-transmitted granulocytic anaplasmosis of veterinary and medical importance (Brown et al. 2001, Majazki et al. 2013). *A. phagocytophilum* is common in bank voles in Finland (Kallio et al. 2014).

Among several other mammals, wild rodents host zoonotic *F. tularensis* (Rossow et al. 2014b, 2014c), which was present also in this study. Furthermore, DNA belonging to Corynebacteriaceae family was detected from the liver samples. At least two species, *Corynebacterium pseudotuberculosis*

Table 3. Comparison of Bacterial Findings Using Br-PCR and Pyrosequencing Methods, Identified by Ribosomal Database Project Classifier

Vole ID	*Conventional Br-PCR findings*[a]	*Pyrosequencing findings*[b]
JB30[c]	Lachnospiraceae, Ruminococcaceae	Lachnospiraceae (18.0%), Porphyromonadaceae (18.0%), Ruminococcaceae (11.0%), Cytophagaceae (12.1%), Coriobacteriaceae (9.0%), Prevotellaceae (6.2%)
TUL38	Lachnospiraceae, Ruminococcaceae	Porphyromonadaceae (35.8%), Ruminococcaceae (9.8%), Lachnospiraceae (9.3%), Prevotellaceae (7.1%), Cytophagaceae (5.3%)
TUL58	Lachnospiraceae, *Lactobacillus*, Coriobacteriaceae	Lachnospiraceae (23.8%), Ruminococcaceae (11.1%), Porphyromonadaceae (12.1%), Prevotellaceae (8.2%), Cytophagaceae (8.6%)
JB14	Lachnospiraceae, Ruminococcaceae, *Helicobacter*	Lachnospiraceae (26.2%), Ruminococcaceae (21.4%), Porphyromonadaceae (10.3%), Cytophagaceae (6.7%)
JB37	Lachnospiraceae	Ruminococcaceae (16.5%), Lachnospiraceae (15.9%), Porphyromonadaceae (14.4%), Cytophagaceae (11.8%), Mycoplasmataceae (8.3%)
JB38	*Mycoplasma*, *Pseudomonas*	Mycoplasmataceae (22.2%), Ruminococcaceae (21.7%), Campylobacteraceae (12.5%), Prevotellaceae (7.1%)
TUL39	*Mycoplasma*, Ruminococcaceae	Mycoplasmataceae (36.8%), Campylobacteraceae (11.3%), Propionibacteriaceae (6.1%)
JB40	*Mycoplasma*	Mycoplasmataceae (59.0%), Prevotellaceae (20.5%)
TUL36	Unclassified bacteria[d]	Mycoplasmataceae (43.1%), Carnobacteriaceae (42.5%)
TUL30	*Bartonella*, *Mycoplasma*	Bartonellaceae (52.4%), Mycoplasmataceae (41.1%)
TUL34	*Bartonella*	Bartonellaceae (75.7%), Mycoplasmataceae (22.2%)
JB23	*Bartonella*	Bartonellaceae (18.3%), Comamonadaceae (16.8%)
JB47	*Bartonella*	Bartonellaceae (63.5%), Ruminococcaceae (11.3%)
JB10	*Corynebacterium*	Bartonellaceae (69.9%), Ruminococcaceae (7.0%)
TUL40	*Bartonella*	Bartonellaceae (93.0%)
JB20	*Bartonella*, *Anaplasma*, *Acinetobacter*	Bartonellaceae (90.4%)
JB18	*Bartonella*, *Anaplasma*	Anaplasmataceae (75.5%), Bartonellaceae (18.6%)
TUL29	*Anaplasma*	Anaplasmataceae (92.4%)
TUL37	*Francisella*, *Anaplasma*, *Mycoplasma*, *Lactobacillus*	Francisellaceae (36.7%), Anaplasmataceae (33.1%), Ruminococcaceae (7.6%)
TUL32	*Francisella*	Francisellaceae (100%)
TUL33	*Francisella*	Francisellaceae (100%)
JB48	*Bartonella*	Moraxellaceae (39.4%), Corynebacteriaceae (23.9%)
TUL25	Unclassified bacteria[d]	Unclassified bacteria

[a]Includes findings with assignment confidence of $\geq 97\%$.
[b]Includes findings of sequence types, which exceed the relative abundance of 5% in the sample. Percentage is indicated in *brackets*.
[c]Control vole; no splenomegaly.
[d]Results with BLAST search tool: *Mycoplasma* sp. (93%).

and *Corynebacterium ulcerans*, colonize or infect ruminants and horses and occasionally humans (Krauss et al. 2003). However, the most notable human infection is caused by *C. diphtheria* that is considered nearly exclusively as a human pathogen (Sing et al. 2016).

Many of our findings reflect the normal intestinal flora of other rodent species. We found bacterial DNA representing members of the order Clostridiales, some of which (*e.g.*, *Clostridium* spp.) are well-known parts of normal murine and probably also voles' intestinal flora (Salzman et al. 2002). More specifically, representatives from families Lachnospiraceae and Ruminococcaceae were common. More experimental studies are needed to determine whether our findings reflect nonviable bacterial remnants in the liver or leakage of the intestine walls as part of early decomposition.

Family Moraxellaceae and, more specifically, *Acinetobacter* were also present in our vole samples. Acinetobacteria are common in nature and occur in normal flora of mice (Pedron et al. 2012) and humans (Bergogne-Berezin et al. 1996). However, every now and then these opportunistic bacteria cause nosocomial epidemics, which may lead to lethal outcomes resulting, for example, from pneumonia. Several fatal infections have also been described in military personnel [reviewed in O'Shea (2012)].

The composition of the intestinal commensal microflora of wild voles is poorly known. However, that of laboratory mice has been studied decades ago by cultivation (Dubos et al. 1965, Schaedler et al. 1965) and later by molecular and microdissection methods (Salzman et al. 2002, Nava et al. 2011, Pedron et al. 2012). According to these studies, murine microflora includes obligate anaerobes, such as *Bacteroides* spp. and *Clostridium* spp, facultative anaerobes, such as *Lactobacillus* spp., *Enterococcus* spp., and *Enterobacillus* spp., as well as aerobic *Acinetobacter* spp. In the mouse colon, bacteria belonging to the phylum *Firmicutes* are enriched, Lachnospiraceae and Ruminococcaceae being the predominant families (Nava et al. 2011). In addition to *Firmicutes*, bacteria from phylum *Bacteroidetes* contribute to the luminal bacterial contents in mice and humans (Ley et al. 2005, Garner et al. 2009, Nava et al. 2011). In murine crypta, *Acinetobacter*

Table 4. Inverse Simpson and Shannon Diversity Indexes for Each Vole Sample

Vole ID	*Number of seqs*	*Coverage*	*Sobs*	*Inv Simpson*	*Inv Simpson_lci*	*Inv Simpson_hci*	*Shannon*	*Shannon_lci*	*Shannon_hci*
M. agrestis-Kolari-JB47	196	0.94898	24	5.031596	4.308735	6.045893	2.060098	1.877782	2.242415
M. agrestis-Kolari-JB48	163	0.883436	42	22.152685	18.350143	27.94309	3.271317	3.124267	3.418366
M. agrestis-Konnevesi-TUL25	235	0.940426	38	11.923244	9.958942	14.852812	2.889178	2.740537	3.037818
M. agrestis-Konnevesi-TUL29	7301	0.997809	51	1.084337	1.074389	1.09447	0.296553	0.267212	0.325894
M. agrestis-Konnevesi-TUL30	1225	0.964898	86	1.555836	1.474931	1.646133	1.264418	1.143053	1.385783
M. agrestis-Konnevesi-TUL32	19,366	0.999742	7	1.001034	1.000396	1.001671	0.005198	0.00228	0.008115
M. agrestis-Konnevesi-TUL33	23,598	0.999746	10	1.001018	1.000445	1.001591	0.005452	0.002649	0.008256
M. agrestis-Konnevesi-TUL34	8439	0.997512	38	1.056439	1.049035	1.063948	0.192404	0.170742	0.214066
M. agrestis-Konnevesi-TUL36	241	0.941909	39	8.763636	6.91328	11.966498	2.857274	2.687942	3.026606
M. agrestis-Konnevesi-TUL37	8634	0.995367	137	2.851895	2.78899	2.917704	1.755014	1.71436	1.795667
M. agrestis-Konnevesi-TUL38	2221	0.936065	359	55.552526	50.360425	61.938292	4.868758	4.806312	4.931205
M. agrestis-Pieksämäki-TUL39	962	0.97817	92	29.800851	27.080017	33.1295	3.797899	3.72818	3.867618
M. agrestis-Pieksämäki-TUL40	2731	0.980227	94	1.155055	1.131996	1.179073	0.544098	0.478564	0.609633
M. agrestis-Pieksämäki-TUL58	6890	0.938607	925	53.263711	50.070867	56.891481	5.250534	5.205474	5.295593
M. oeconomus-Kolari-JB37	2721	0.935685	503	93.968157	85.200048	104.747985	5.376409	5.323114	5.429705
M. oeconomus-Kolari-JB38	2153	0.990246	88	12.684122	12.003375	13.446725	3.070901	3.012539	3.129263
M. oeconomus-Kolari-JB40	292	0.938356	43	14.933568	12.584067	18.361796	3.073928	2.949389	3.198467
M. oeconomus-Pisavaara-JB14	5651	0.931163	969	152.318786	143.114026	162.788994	5.879128	5.839977	5.91828
M. oeconomus-Pisavaara-JB18	1027	0.991237	25	2.160917	2.034111	2.304584	1.214994	1.128263	1.301724
M. oeconomus-Pisavaara-JB20	5606	0.996432	66	1.200618	1.18166	1.220194	0.59598	0.552348	0.639612
M. oeconomus-Pisavaara-JB23	222	0.963964	31	16.836651	14.318574	20.429379	3.013612	2.901857	3.125366
M. oeconomus-Pyhätunturi-JB10	10,003	0.997201	104	1.601494	1.570382	1.633863	1.24906	1.208947	1.289173
Control vole-JB30	3571	0.877345	752	69.684333	63.776785	76.798018	5.360887	5.303846	5.417929

Sobs, number of observed OTUs; Inv Simpson, Inverse Simpson's diversity index, with lower (lci) and upper (hci) 95% confidence intervals; Shannon, Shannon index, with lower (lci) and upper (hci) 95% confidence intervals.

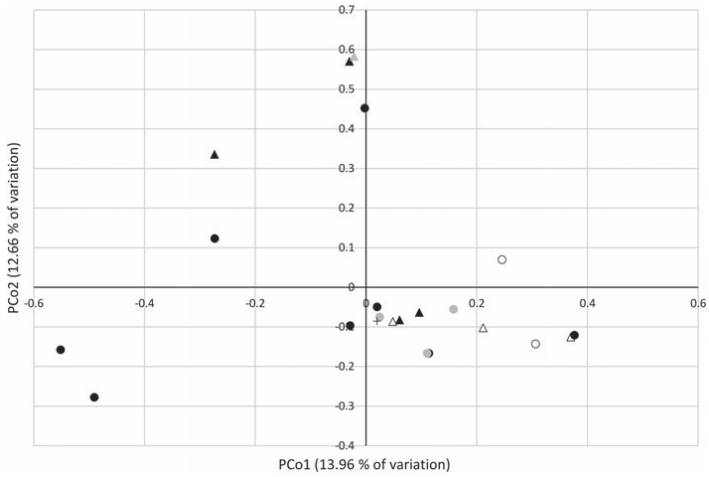


FIG. 3. Principal Coordinates Analysis (PCoA). *Microtus agrestis* samples are indicated with *circles*, *Microtus oeconomus* with *triangles* and *M. oeconomus* control vole with a *cross*. The collection sites are indicated as follows: Kolari *empty circle/triangle*; Konnevesi *black circle*; Pisavaara *black triangle*; Pieksamäki *gray circle*; Pyhätunturi *gray triangle*. Note that the two principle coordinated explain only 26.62% of the total variation.

spp. is common (Pedron et al. 2012). These previous findings from mice reflected those seen in our study on voles.

While the OTU-based analysis of the pyrosequencing data suggested some differences between the bacterial communities of different vole species and between some of the locations, these results should be interpreted cautiously. The AMOVA suggested differences between voles captured in Pisavaara compared with those captured in Kolari. This may be due to relatively high proportion of Bartonellaceae in the *M. oeconomus* voles captured in Pisavaara and the relatively high proportions of *Mycoplasmataceae* in the voles captured in Kolari. We purposely studied splenomegalic voles as an indicator of potential infection; our results may not represent the average normal flora. The nonpathogenic findings from the sole vole with visibly normal spleen suggest this as well.

Razzauti et al. (2015) studied voles collected from France using next-generation sequencing techniques and detected 45 potential zoonotic bacterial genera from spleen samples. These included, for example, *Bartonella*, *Rickettsia*, *Borrelia*, *Neoehrlichia*, *Anaplasma*, and *Francisella*. Several same bacterial genera were also discovered from our voles. The differences may be due to geographical location, the organs studied, and techniques used, as well as our selection based on splenomegaly.

We found pyrosequencing useful in providing wide-ranging and vast information on the bacterial composition of vole liver tissue. Tagged pyrosequencing is less laborious, resulting in less hands-on time compared with the conventional Br-PCR method as 16S rDNA PCR products can be characterized directly without cloning. Furthermore, since barcoded PCR primers allow dozens of different samples to be analyzed in the same pyrosequencing run, microbial communities can be characterized at a fast rate. Overall, pyrosequencing proved to be a favorable method for characterization of bacterial flora, especially if the number of samples is large. On the other hand, conventional Br-PCR is useful when studying fewer samples and when high-technology instrumentation and expertise are unavailable.

To obtain sufficient amount of PCR product for pyrosequencing, we had to increase the number of PCR cycles from 35 to 45. This raised the amount of contaminant sequences in the final results. In seven samples, for example, the amount of Halomonadaceae and Shewanellaceae reads exceeded even 8000 (Fig. 1a). Previous studies have similarly shown increasing proportion of contaminant sequences and their dominance in samples containing low bacterial copy numbers (Salter et al. 2014). In our study, this was especially evident with three samples, where the relevant sequences numbered only hundreds.

The DECIPHER tool detected one chimera from our Br-PCR data. In addition, it was unable to classify several sequences. According to Wright et al. (2012), indecipherable sequences are often either chimeric or not 16S. Furthermore, altogether 108 sequences shared 78–100% similarity with vole sequences (field and bank vole), thus not being of bacterial 16S rRNA origin (Supplementary Table S1).

We were able to identify 33% of the Br-PCR sequence types to genus level, which leaves a substantial portion of undetermined sequences. These probably represent previously uncharacterized bacteria occurring in vole communities in Finland. Of special interest is a group of sequence types sharing some similarity with *Mycoplasma* spp. The significance of these findings needs clarification with genus-specific identification methods.

According to this study, contaminating exogenous DNA is common in extraction and PCR reagents and it originates from a variety of bacteria (Supplementary Tables S2–S4). Some of the bacterial species, for example, *Pseudomonas*, *Micrococcus*, *Sphingomonas*, *Methylobacterium*, *Acidovorax*, and *Phyllobacterium*, have been reported earlier (Maiwald et al. 1994, Tanner et al. 1998, Barton et al. 2006), but some novel sequence types were also found. Different PCR reagents and extraction kits seem to contain different contaminants.

Interestingly, many of the contaminant sequence types (such as *Phyllobacterium* sp., *Sphingomonas* sp., *Variovorax* sp., and *Pseudomonas* sp.) have been reported to exist in several body sites based on a method similar to this study (Supplementary Fig. S4) and published in the Human Microbiome Project (www.hmpdacc.org/HM16STR/). These findings should be confirmed by detection methods unaffected

by potential amplification of contaminant DNA in the reagents (Salter et al. 2014, Aho et al. 2015).

Conclusions

This study revealed significant bacterial diversity in vole liver samples, consisting of known pathogens and reflecting that of intestinal flora as well. Overall, the findings were independent from the trapping location. The voles commonly carried potential or verified zoonotic bacterial pathogens and therefore they may play a role in spreading them.

The two methods used gave comparable results. Pyrosequencing needs less hands-on time, being more suitable with a large sample number. On the other hand, the Br-PCR method is ideal when studying fewer samples. In addition, our results remind us of the need to acknowledge the possible presence of exogenous DNA in molecular reagents. Ignorance may lead to false conclusions.

Nucleotide sequence accession numbers

The sequences reported in this article were deposited in NCBI GenBank with the accession numbers KT961130–KT961324.

Acknowledgments

The authors thank Heidi Rossow for sample preparation. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. This work is part of the European Biodefence project EDA B-1325-ESM4-GP and has been partially funded by EU grant FP7-261504 EDENext and cataloged as EDENext 442 (www.edenext.eu).

Author Disclosure Statement

No competing financial interests exist.

References

Aho VT, Pereira PA, Haahtela T, Pawankar R, et al. The microbiome of the human lower airways: A next generation sequencing perspective. World Allergy Organ J 2015; 8:23.

Barton HA, Taylor NM, Lubbers BR, Pemberton AC. DNA extraction from low-biomass carbonate rock: An improved method with reduced contamination and the low-biomass contaminant database. J Microbiol Methods 2006; 66:21–31.

Bergogne-Berezin E, Towner KJ. *Acinetobacter* spp. as hosocomial pathogens: microbiological, clinical, and epidemiological features. Clin Microbiol Rev 1996; 9:148–165.

Brown DR, Talkington DF, Thacker WL, Brown MB, et al. *Mycoplasma microti* sp. nov., isolated from the respiratory tract of prairie voles (*Microtus ochrogaster*). Int J Syst Evol Microbiol 2001; 51:409–412.

Buffet JP, Kosoy M, Vayssier-Taussat M. Natural history of *Bartonella*-infecting rodents in light of new knowledge on genomics, diversity and evolution. Future Microbiol 2013; 8: 1117–1128.

Buffet JP, Marsot M, Vaumourin E, Gasqui P, et al. Co-infection of *Borrelia afzelii* and *Bartonella* spp. in bank voles from a suburban forest. Comp Immunol Microbiol Infect Dis 2012; 35:583–589.

Buffet JP, Pisanu B, Brisse S, Roussel S, et al. Deciphering bartonella diversity, recombination, and host specificity in a rodent community. PLoS One 2013; 8:e68956.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006; 72:5069–5072.

Dubos R, Schaedler RW, Costello R, Hoet P. Indigenous, normal, and autochthonous flora of the gastrointestinal tract. J Exp Med 1965; 122:67–76.

Dumler JS, Barbet AF, Bekker CP, Dasch GA, et al. Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: Unification of some species of Ehrlichia with Anaplasma, Cowdria with Ehrlichia and Ehrlichia with Neorickettsia, descriptions of six new species combinations and designation of Ehrlichia equi and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. Int J Syst Evol Microbiol 2001; 51:2145–2165.

Edgar RC, Haas BJ, Clemente JC, Quince C, et al. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011; 27:2194–2200.

Garner CD, Antonopoulos DA, Wagner B, Duhamel GE, et al. Perturbation of the small intestine microbial ecology by streptomycin alters pathology in a Salmonella enterica serovar typhimurium murine model of infection. Infect Immun 2009; 77:2691–2702.

Giebel J, Binder A, Kirchhoff H. Isolation of *Mycoplasma moatsii* from the intestine of wild Norway rats (*Rattus norvegicus*). Vet Microbiol 1990; 22:23–29.

Hanski I, von Hertzen L, Fyhrquist N, Koskinen K, et al. Environmental biodiversity, human microbiota, and allergy are interrelated. Proc Natl Acad Sci U S A 2012; 109:8334–8339.

Hayman DT, McDonald KD, Kosoy MY. Evolutionary history of rat-borne *Bartonella*: The importance of commensal rats in the dissemination of bacterial infections globally. Ecol Evol 2013; 3:3195–3203.

Human Microbiome Project Consortium. A framework for human microbiome research. Nature 2012; 486:215–221.

Ishiguro T, Takayanagi N, Yamaguchi S, Yamakawa H, et al. Etiology and factors contributing to the severity and mortality of community-acquired pneumonia. Intern Med 2013; 52: 317–324.

Jalava J, Kotilainen P, Nikkari S, Skurnik M, et al. Use of the polymerase chain reaction and DNA sequencing for detection of *Bartonella quintana* in the aortic valve of a patient with culture-negative infective endocarditis. Clin Infect Dis 1995; 21: 891–896.

Kallio ER, Begon M, Birtles RJ, Bown KJ, et al. First report of *Anaplasma phagocytophilum* and *Babesia microti* in rodents in Finland. Vector Borne Zoonotic Dis 2014; 14:389–393.

Kotilainen P, Jalava J, Meurman O, Lehtonen OP, et al. Diagnosis of meningococcal meningitis by broad-range bacterial PCR with cerebrospinal fluid. J Clin Microbiol 1998; 36:2205–2209.

Krauss H, Weber A, Appel M, Enders B, et al. *Zoonoses. Infectious Diseases Transmissible from Animals to Humans*, 3rd ed. Washington, DC: ASM Press, American Society for Microbiology; 2003.

Kwok S, Higuchi R. Avoiding false positives with PCR. Nature 1989; 339:237–238.

Ley RE, Backhed F, Turnbaugh P, Lozupone CA, et al. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A 2005; 102:11070–11075.

Lindsey JR, Cassell H. Experimental *Mycoplasma pulmonis* infection in pathogen-free mice. Models for studying mycoplasmosis of the respiratory tract. Am J Pathol 1973; 72:63–90.

Lo YM, Chan KC. Setting up a polymerase chain reaction laboratory. Methods Mol Biol 2006; 336:11–18.

Maiwald M, Ditton HJ, Sonntag HG, von Knebel Doeberitz M. Characterization of contaminating DNA in Taq polymerase which occurs during amplification with a primer set for *Legionella* 5S ribosomal RNA. Mol Cell Probes 1994; 8: 11–14.

Majazki J, Wuppenhorst N, Hartelt K, Birtles R, et al. *Anaplasma phagocytophilum* strains from voles and shrews exhibit specific ankA gene sequences. BMC Vet Res 2013; 9:235.

Nava GM, Friedrichsen HJ, Stappenbeck TS. Spatial organization of intestinal microbiota in the mouse ascending colon. ISME J 2011; 5:627–638.

O'Shea MK. *Acinetobacter* in modern warfare. Int J Antimicrob Agents 2012; 39:363–375.

Pedron T, Mulet C, Dauga C, Frangeul L, et al. A crypt-specific core microbiota resides in the mouse colon. MBio 2012; 3:10.

Qiu Y, Nakao R, Ohnuma A, Kawamori F, et al. Microbial population analysis of the salivary glands of ticks; a possible strategy for the surveillance of bacterial pathogens. PLoS One 2014; 9:e103961.

Razzauti M, Galan M, Bernard M, Maman S, et al. A comparison between transcriptome sequencing and 16S metagenomics for detection of bacterial pathogens in wildlife. PLoS Negl Trop Dis 2015; 9:e0003929.

Renko J, Koskela KA, Lepp PW, Oksala N, et al. Bacterial DNA signatures in carotid atherosclerosis represent both commensals and pathogens of skin origin. Eur J Dermatol 2013; 23:53–58.

Rossow H, Forbes KM, Tarkka E, Kinnunen PM, et al. Experimental infection of voles with *Francisella tularensis* indicates their amplification role in tularemia outbreaks. PLoS One 2014a; 9:e108864.

Rossow H, Ollgren J, Klemets P, Pietarinen I, et al. Risk factors for pneumonic and ulceroglandular tularaemia in Finland: A population-based case-control study. Epidemiol Infect 2014b; 142:2207–2216.

Rossow H, Sissonen S, Koskela KA, Kinnunen PM, et al. Detection of *Francisella tularensis* in voles in Finland. Vector Borne Zoonotic Dis 2014c; 14:193–198.

Salipante SJ, Sengupta DJ, Rosenthal C, Costa G, et al. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. PLoS One 2013; 8:e65226.

Salter SJ, Cox MJ, Turek EM, Calus ST, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 2014; 12:87.

Salzman NH, de Jong H, Paterson Y, Harmsen HJ, et al. Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. Microbiology 2002; 148:3651–3660.

Sashida H, Sasaoka F, Suzuki J, Watanabe Y, et al. Detection of hemotropic mycoplasmas in free-living brown sewer rats (*Rattus norvegicus*). J Vet Med Sci 2013; 75:979–983.

Schaedler RW, Dubos R, Costello R. The development of the bacterial flora in the gastrointestinal tract of mice. J Exp Med 1965; 122:59–66.

Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 2011; 6:e27310.

Schloss PD, Westcott SL, Ryabin T, Hall JR, et al. Introducing mothur: Open-source, platform-independent, community--supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009; 75:7537–7541.

Schmidt TM, Relman DA. Phylogenetic identification of uncultured pathogens using ribosomal RNA sequences. Methods Enzymol 1994; 235:205–222.

Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008; 26:1135–1145.

Silaghi C, Pfeffer M, Kiefer D, Kiefer M, et al. *Bartonella*, rodents, fleas and ticks: A molecular field study on host-vector-pathogen associations in Saxony, Eastern Germany. Microb Ecol 2016; 72:965–974.

Sing A, Konrad R, Meinel DM, Mauder N, et al. *Corynebacterium diphtheriae* in a free-roaming red fox: Case report and historical review on diphtheria in animals. Infection 2016; 44:441.

Sumithra TG, Chaturvedi VK, Susan C, Siju SJ, et al. Mycoplasmosis in wildlife: A review. Eur J Wildl Res 2013; 59: 769–781.

Sun X, Jones HP, Dobbs N, Bodhankar S, et al. Dendritic cells are the major antigen presenting cells in inflammatory lesions of murine *Mycoplasma* respiratory disease. PLoS One 2013; 8:e55984.

Tanner MA, Goebel BM, Dojka MA, Pace NR. Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. Appl Environ Microbiol 1998; 64:3110–3113.

Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007; 73: 5261–5267.

Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 1991; 173:697–703.

White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. PLoS Comput Biol 2009; 5:e1000352.

Wilbrink B, van der Heijden IM, Schouls LM, van Embden JD, et al. Detection of bacterial DNA in joint samples from patients with undifferentiated arthritis and reactive arthritis, using polymerase chain reaction with universal 16S ribosomal RNA primers. Arthritis Rheum 1998; 41:535–543.

Winglee K, Eloe-Fadrosh E, Gupta S, Guo H, et al. Aerosol *Mycobacterium tuberculosis* infection causes rapid loss of diversity in gut microbiota. PLoS One 2014; 9:e97048.

Wright ES, Yilmaz LS, Noguera DR. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl Environ Microbiol 2012; 78:717–725.

Address correspondence to:
Simo Nikkari
Centre for Military Medicine
Tukholmankatu 8A
Helsinki FI-00290
Finland

E-mail: simo.nikkari@mil.fi

IV

environmental microbiology

Environmental Microbiology (2015) **17**(11), 4306–4321 doi:10.1111/1462-2920.12816

Generation of a CRISPR database for *Yersinia pseudotuberculosis* complex and role of CRISPR-based immunity in conjugation

Katja A. Koskela,[1] Laura Mattinen,[2] Laura Kalin-Mänttäri,[1,2] Gilles Vergnaud,[3,4,5] Olivier Gorgé,[3,4,6] Simo Nikkari[1] and Mikael Skurnik[2,7]*

[1]*Research and Development Department, Centre for Military Medicine, Helsinki, Finland.*
[2]*Department of Bacteriology and Immunology, Haartman Institute and Research Programs Unit, Immunobiology, University of Helsinki, PO Box 21, 00014 Helsinki, Finland.*
[3]*Univ Paris-Sud, Institut de Génétique et Microbiologie, UMR8621, Orsay, France.*
[4]*CNRS, Orsay, France.*
[5]*ENSTA ParisTech, Palaiseau, France.*
[6]*DGA/MNRBC, Vert le Petit, France.*
[7]*Helsinki University Central Hospital Laboratory Diagnostics, Helsinki, Finland.*

Summary

The clustered regularly interspaced short palindromic repeat – CRISPR-associated genes (CRISPR-Cas) system is used by bacteria and archaea against invading conjugative plasmids or bacteriophages. Central to this immunity system are genomic CRISPR loci that contain fragments of invading DNA. These are maintained as spacers in the CRISPR loci between direct repeats and the spacer composition in any bacterium reflects its evolutionary history. We analysed the CRISPR locus sequences of 335 *Yersinia pseudotuberculosis* complex strains. Altogether 1902 different spacer sequences were identified and these were used to generate a database for the spacer sequences. Only ~10% of the spacer sequences found matching sequences. In addition, surprisingly few spacers were shared by *Yersinia pestis* and *Y. pseudotuberculosis* strains. Interestingly, 32 different protospacers were present in the conjugative plasmid pYptb32953. The corresponding spacers were identified from 35 different *Y. pseudotuberculosis* strains indicating that these strains had encountered pYptb32953 earlier. In conjugation experiments, pYptb32953-specific spacers generally prevented conjugation with spacer-positive and spacer-free strains. However, some strains with one to four spacers were invaded by pYptb32953 and some spacer-free strains were fully resistant. Also some spacer-positive strains were intermediate resistant to conjugation. This suggests that one or more other defence systems are determining conjugation efficiency independent of the CRISPR-Cas system.

Received 30 January, 2015; accepted 11 February, 2015. *For correspondence. E-mail mikael.skurnik@helsinki.fi; Tel. (+358) 9 191 26464; Fax: (+358) 9 191 26382.

Introduction

Yersinia pseudotuberculosis is a Gram-negative bacterium that causes disease in humans and animals. In humans, *Y. pseudotuberculosis* is a cause of food-borne associated illness with symptoms of fever and abdominal pain, and sometimes diarrhoea. In animals, it causes tuberculosis-like disease (Aleksic *et al*., 1995; Naktin and Beavis, 1999; Tauxe, 2004). *Yersinia pestis*, the bacterium responsible for plague, evolved from its *Y. pseudotuberculosis* ancestor approximately 1500–6400 years ago (Achtman *et al*., 1999; Morelli *et al*., 2010; Cui *et al*., 2013; Harbeck *et al*., 2013). In a multilocus sequence typing (MLST) study, *Y. pseudotuberculosis*, *Y. pestis* (representing a single ST), the recently described *Yersinia similis* (Sprague *et al*., 2008) and a number of distinct strains, called the Korean group and recently named as *Yersinia wautersii* (Savin *et al*., 2014), were collectively named as a *Y pseudotuberculosis* complex (Laukkanen-Ninios *et al*., 2011). Due to their close evolutionary relationship, *Y. pseudotuberculosis* and *Y. pestis* are very similar and share ≥ 97% nucleotide sequence identity for most of the chromosomal genes depending on the *Y. pseudotuberculosis* strain in question. *Yersinia pseudotuberculosis* is commonly typed serologically based on the lipopolysaccharide O-antigen. Some of the 15 known serotypes are divided into subtypes (O:1a, O:1b, O:1c, O:2a, O:2b, O:2c, O:4a, O:4b, O:5a, O:5b) resulting in a total of 21 serotypes (Bogdanovich *et al*., 2003). *Yersinia pestis* does not express O-antigen due to

pseudogenes in the O-antigen biosynthetic genes; however, comparison of the *Y. pestis* O-antigen gene cluster sequence with those of different *Y. pseudotuberculosis* serotype gene clusters suggested that *Y. pestis* evolved from a *Y. pseudotuberculosis* serotype O:1b strain (Skurnik *et al*., 2000).

Therefore, differentiating and typing of these two species has been challenging (Chauvaux *et al*., 2011). For instance, an earlier study has suggested ribotyping as one potential typing method, but even with this method differentiation was not accurate (Voskressenskaya *et al*., 2005). Additionally, MLST (Ch'ng *et al*., 2011; Laukkanen-Ninios *et al*., 2011), 16S rRNA gene sequencing and pulsed-field gel electrophoresis (Souza *et al*., 2010) have been used for typing of *Y. pseudotuberculosis*. Some of these methods can identify and differentiate *Yersinia* species, but still typing of *Y. pseudotuberculosis* is challenging.

The clustered regularly interspaced short palindromic repeat – CRISPR-associated genes (CRISPR-Cas) system is a RNA-based immune system that regulates invasions of plasmids and viruses in bacteria and archaea. The functional mechanisms of CRISPR and its whole biological significance are still not fully known (Bolotin *et al*., 2005; Pourcel *et al*., 2005; Barrangou *et al*., 2007; Hale *et al*., 2009; Garneau *et al*., 2010; Makarova *et al*., 2011; Sorek *et al*., 2013). CRISPRs are constructed from a chain of 21 to 47 bp repeated sequences [called direct repeats (DR)] and in between DRs are unique spacer sequences. These spacers represent foreign DNA originating predominantly from bacteriophages and plasmids. A leader sequence is located at the 5′-end of the CRISPR and usually the *cas* genes are located upstream of the leader of one of the CRISPR loci (Bolotin *et al*., 2005; Pourcel *et al*., 2005; Karginov and Hannon, 2010; Sontheimer and Marraffini, 2010). The three main types of CRISPR-Cas systems differ in the composition of *cas* genes and in the mechanisms of CRISPR RNA processing and interference (Makarova *et al*., 2011; 2013; Wiedenheft *et al*., 2012). *Yersinia* contain the subtype I-F CRISPR-Cas system (Haft *et al*., 2005; Makarova *et al*., 2011) and the *cas* genes are located upstream of the most ancestral spacers (Fig. 1) of one of the three CRISPR loci present in *Yersinia*. The CRISPR locus and the *cas* genes have the same transcription direction.

When a prokaryote comes into contact with foreign DNA, the host may integrate a fragment of this DNA, known as a protospacer, into the CRISPR locus as a new spacer. Earlier studies show that approximately 45% of bacteria and nearly all of archaea contain a CRISPR-Cas system (Grissa *et al*., 2007a; Pourcel and Drevet, 2013). The new spacers are acquired at the leader proximal end, such that leader distal spacers are older, thus often shared between more isolates (Pourcel *et al*., 2005; Barrangou *et al*., 2007). Due to their high diversity, the CRISPR sequences have been used for typing (Shariat and Dudley, 2014), for example, for *Mycobacterium tuberculosis* (Kamerbeek *et al*., 1997), *Campylobacter* (Schouls *et al*., 2003), *Streptococcus thermophilus* (Horvath *et al*., 2008), *Escherichia coli* (Diez-Villasenor *et al*., 2010; Touchon *et al*., 2011; Delannoy *et al*., 2012a,b), *Salmonella enterica* (Liu *et al*., 2011; Fabre *et al*., 2012) and also for *Erwinia amylovora* (Rezzonico *et al*., 2011; McGhee and Sundin, 2012). In *Yersinia*, there are three loci named as YP1, YP2 and YP3 of which the YP1 locus was initially used as a variable number tandem repeat marker (Le Flèche *et al*., 2001; Pourcel *et al*., 2005).

In this study, our aim was to generate a comprehensive *Y. pseudotuberculosis* complex database of CRISPR spacers, to use the database to distinguish between strains, and to compare these results with the 90 sequence types (ST) defined in the recent MLST study (Laukkanen-Ninios *et al*., 2011).

Results

The YP1 locus was amplified from 60 of the 90 Skurnik laboratory *Y. pseudotuberculosis* complex strains. Twenty *Y. pseudotuberculosis* strains and all 10 *Y. similis* strains yielded no polymerase chain reaction (PCR) product (Table S1). From the amplified YP1 locus fragments, five could be only partially sequenced and no sequence was obtained from 11 PCR products. The YP2 locus was amplified and sequenced from 61 strains, 19 strains yielded no PCR products and the PCR products of four strains could not be sequenced; in addition, six strains yielded non-CRISPR sequences. The YP3 locus was amplified and sequenced from 81 strains. Five strains yielded no PCR products, two strains were partly sequenced and the PCR product from one strain could not be sequenced; in addition, one strain yielded non-CRISPR sequence. We did not push to optimize the PCR-based sequencing approach as whole genome sequencing is a present day viable alternative. Typically highest numbers of spacers were found from the YP1 and YP3 loci (up to 50 different), while very few were in the YP2 locus.

Analysis of the YP1, YP2 and YP3 CRISPR loci in 335 Y. pseudotuberculosis *complex strains*

The above sequence data were complemented with the CRISPR loci sequences of 40 *Y. pseudotuberculosis* strains and 195 *Y. pestis* strains (Table S1) from earlier investigations [(Pourcel *et al*., 2005; Cui *et al*., 2008) and Vergnaud G and Gorgé O, unpublished]. In addition,

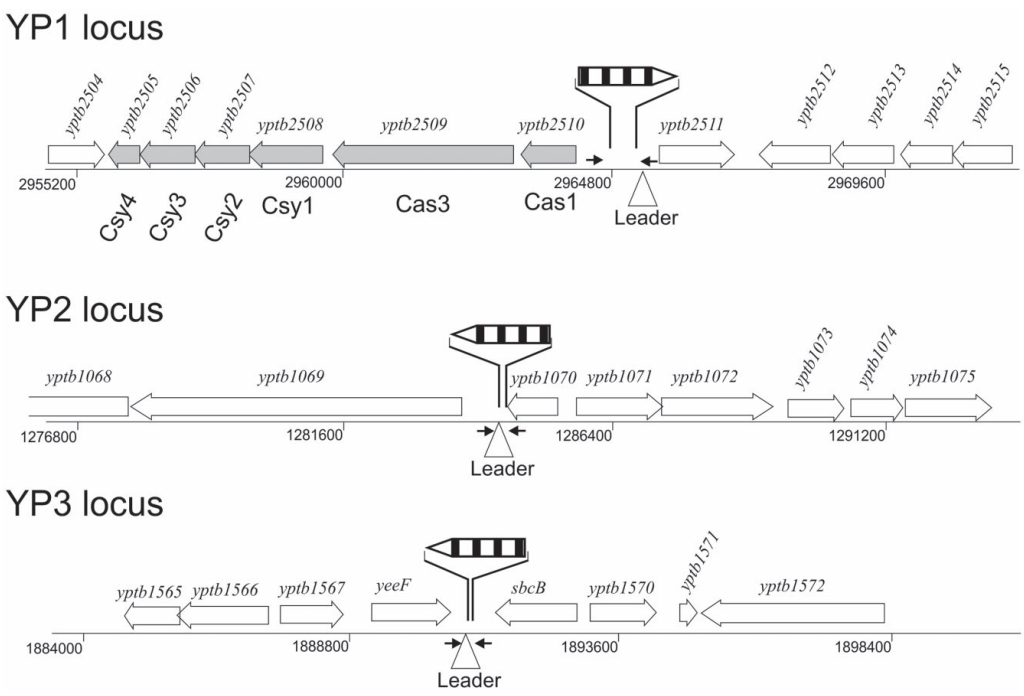


Fig. 1. The locations of the YP1, YP2 and YP3 CRISPR loci in the genome of *Y. pseudotuberculosis* strain IP32953 (Accession No. BX936398). The CRISPR associated *cas* and *csy* genes are shown with grey shading and the variable CRISPR repeat sequences as striped arrows. The new spacers are added between the leader sequence and the arrowhead of the CRISPR repeat sequences. The locations of the leader sequence and of the primers used to amplify the loci are indicated by a triangle and black arrows respectively.

we extracted the CRISPR loci sequences from published complete genome sequences of four *Y. pseudotuberculosis* and six *Y. pestis* strains (Table S1). Then, the sequence data of altogether 335 *Y. pseudotuberculosis* complex strains were analysed by the CRISPRFinder tool. The DR consensus sequence identified from these sequences was identical to that of *Y. pestis* (Cui *et al.*, 2008), i.e. 5'-TTTCTAAGCTACCTGTGCGGCAGTGAAC-3'. Similar to Cui and colleagues, we identified a number of modified DRs with differences to consensus DR in various positions of the CRISPR loci (Table S3).

Altogether more than 6000 spacers with 1902 different spacer sequences were identified among the analysed sequence data (Table S4). The numbering in Table S4 is used to distinguish the spacers. Surprisingly little overlap of spacer distribution between the strains was noticed. One thousand one hundred and fifty-three spacers were unique to single strains (shown in Table S4), 311 were present in two strains, 143 in three strains, 77 in four strains, *c.* 56 in five strains, *c.* 47 in six strains and 25 in seven strains. Those that were shared in ≥ 8 *Y. pseudotuberculosis* and *Y. similis* strains are shown in Table S5. Since these spacers did not give any significant hits in BLASTN search (Table S5), we at present have no clues of their origins except for spacer #7 that had similarity to *E. coli* plasmid sequence. This spacer was present in eight strains. The most common spacers (#1074, #1149, #507, #40) were always found close to the most ancestral, i.e. the leader distal end of the CRISPR loci.

To visualize possible evolutionary relationships between the strains based on the organization of the spacer sequences, the *Y. pseudotuberculosis*, *Y. similis* and *Y. wautersii* strain-specific spacer patterns were manually aligned, and the alignments are shown in Table S6. If all spacers present in the most recent common ancestor of *Y. pseudotuberculosis* complex had been subsequently maintained, present day strains should have the same root (most ancient) spacer. This is not the case; instead, there were several root spacers both in the YP1 and YP3 loci. The most likely explanation is that older spacers were randomly lost. Furthermore, gaps had to be introduced to the spacer patterns to maximize their alignment. When the

spacer pattern alignments were used for grouping of the strains, we found that the phylogenies of YP1 and YP3 loci seem not to be congruent. In fact, the alignments indicated that the spacers had accumulated independently to these main storage loci. Table S6A and B shows the strains sorted based on YP1 and YP3 alignments respectively. As an example of this, the YP1 and YP3 alleles of selected strains are shown in Table 1. For instance, the strains carrying YP1 alleles with root spacers 539.173.177.-- could carry YP3 allele root spacers 507.1238.-- or 1149.1332.-- or 1149.1199.-- (Table 1 top). Conversely, strains carrying any of these YP3 allele root spacers could carry two or more very different YP1 allele root spacers (Table 1, bottom). These data strongly suggest that horizontal gene transfer (HGT) and recombination between the CRISPR loci has occurred within the *Y. pseudotuberculosis* complex; however, with the present data, we cannot evaluate the full extent of such mosaicisms. Within the strain groups presented in Table S6, one can observe plenty of examples of possible recombination events and reassortments leading to deletions of spacer(s). For example, in the spacer 539.173-rooted YP1 group (Table 1, top), the spacer block 539.173 . . . 187 is present in six strains; however, it is not identical in them as in some strains spacers from the middle are missing (e.g. MW145-2, that strain is also missing the most ancient spacer #539). Strain Toyama60, on the other hand, has gained spacer block 539.173–177, but not as the most ancient one but the block has recombined after the ancient spacers 104.801.802 (Table 1). On the other hand, in the spacer 1149-rooted YP3 alignments (Table 1, bottom), similar events can also be easily tracked. The clonal evolution of CRISPR loci observed in *Y. pestis* (Cui *et al*., 2008) may be an exception reminiscent of the situation observed with *Mycobacterium canettii* and the *Mycobacterium tuberculosis* complex (Blouin *et al*., 2014).

Among the 84 *Y. pseudotuberculosis* strains with YP1 sequences, the most prevalent ancient or root spacer in the YP1 locus was #39 (present in 14 strains), followed by spacers #103, #40, #539, #403, #76 and #581 (present in 13, 9, 6, 4, 3 and 3 strains, respectively). Eight different ancient spacers were shared by two strains and 15 strains had unique ancient spacers. Sequence information for the YP1 locus was not obtained for 50 of the 124 *Y. pseudotuberculosis* and 10 *Y. similis* strains.

The YP2 locus of the strains carried generally one or two spacers, with only three exceptions in which the locus carried six or eight spacers (Tables S6 and S7). By sequence comparison, a couple of different repeat variants and altogether 17 different spacers were detected in the YP2 locus. No spacers were present in the *Y. similis* YP2 locus. Figure 2 and Table S7 show the alignments of the *Y. pseudotuberculosis* YP2 locus sequences and their comparison to the *Y. pestis* CO92 YP2 locus. Here we exploited for the grouping of strains the CRISPR 5'- and 3'-flanking sequences obtained from the YP2 PCR fragments (Fig. 2 and Table S7). Comparison of the YP2 3'-flanks revealed the presence or absence from the strains of five distinct sequence elements that we named as 3'A to 3'E (Fig. 2 and Table S7). Interestingly, all five 3' elements are present only in the *Y. pestis* and in six *Y. pseudotuberculosis* strains (Fig. 2). Most other strains were missing the 731–736 bp 3′-E and the 78 bp 3′-B element. The 3′-E element of strain CO92 includes the whole *ypo2574* gene encoding a putative membrane protein of the DUF1440 protein family. The 32 bp 3'D element was present in all strains. The 66 bp 3′-A and the 25 bp 3′-C elements were missing only from the *Y. similis* strains. Absence of the 3′-A element that is the YP2 locus CRISPR leader sequence might explain why *Y. similis* does not carry any spacers in the YP2 locus.

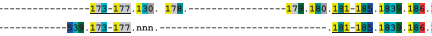
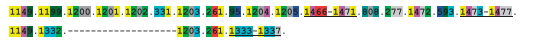
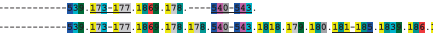
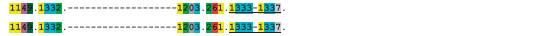
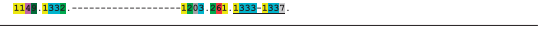
The YP3-locus spacer comparison is presented in Table S6B. Based on the identity of the most ancestral spacer, the strains could be grouped into > 10 groups. Spacers #507, #1149 and #1111 define the largest groups with 38, 31 and 12 strains, respectively. The other spacer groups #511, #1132, #1156, #1199, #1589, #1616, #1622 and #1853 included two to five strains each. The remaining strains had either a sporadic most ancestral spacer (seven strains) or we did not get a PCR product or sequence from the locus (23 strains).

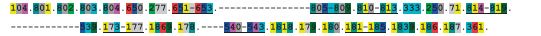
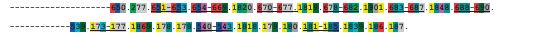
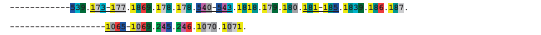
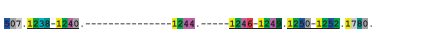
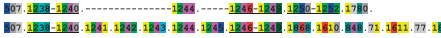
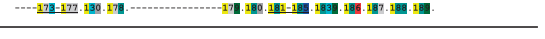
Interestingly, 72 spacers were present in two different CRISPR loci. In one instance, this peculiarity was observed in a single strain. Spacer #808 was found in both the YP1 and YP3 locus of strain No-151. Duplications occurred, for example spacer #257 was found in strains MW101-1 and TE-93081 as a tandem repeat duplicate in the YP1 locus. Another example is the spacer pair #1348.1349 that is present twice in strain DC356-2. Also spacer #1 is present twice in strain BB1152 (Table S6).

Relationships between spacer-based grouping and STs

We next wanted to find out if the spacer-based grouping was in line with the MLST study (Laukkanen-Ninios *et al*., 2011). Comparison of our results to the MLST minimal spanning tree of Laukkanen-Ninios and colleagues (2011) revealed that the CRISPR spacer-based grouping is not in synchrony with the MLST typing; at best, weak correlation could be detected (Figs S1 and S2). However, in all spacer-based groups, closest CRISPR types tended to belong to closely related STs. As an example, one can take the YP3 spacer subgroup of #507–1350-rooted strains (Table S6B and Fig. S2) that grouped in the MLST analysis with a maximum cross-link distance of 5

Table 1. YP1 and YP3 spacer alignments of selected *Y. pseudotuberculosis* strains.[a].

Strain	YP1 locus	YP3 locus
Toyama60		
MW145-2		
Pa3606		
IP31758		
D426		
Y718		

Strain	YP3 locus	YP1 locus
RD20		
No-151		
IP31758		
Toyama60		
D426		
Gifu-liver		
Y718		
Y732		
WP-931110		
Pa3606		
MW145-2		

a. nnn, the middle part of the strain Pa3606 YP1 locus PCR fragment could not be completely sequenced.

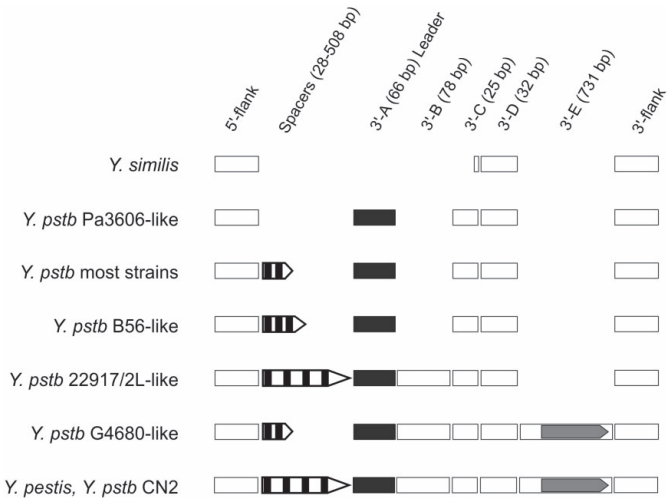


Fig. 2. Schematic organization of the YP2 locus elements used for grouping of the strains. The list of strains and the respective sequences are given in Table S7. The leader sequence element 3′-A is indicated by the black box, the spacer-containing sequence by the striped arrow, the other sequence elements by open boxes. The open reading frame within element 3′-E is indicated by the grey arrow.

(Laukkanen-Ninios *et al*., 2011). Thus, the CRISPR loci are highly more differentiating than MLST as we found among the *Y. pseudotuberculosis* complex strains (excluding the *Y. pestis* strains) no CRISPR-identical strains. On the other hand, same ST strains tend to carry same spacers, for example, three of the four ST16 strains share 12 of the total of 17 different YP1 spacers (Table S6A). A similar situation could also be seen with the YP1 spacers of ST42, ST43 and ST19 strains and with YP3 spacers of ST42, ST43, ST9, ST48 and ST14 strains (Table S6B). Clearly, more complete CRISPR loci sequences of strains representing individual STs will be needed to get better picture of intra-ST CRISPR evolution, as previously done for *Y. pestis* which represents a single ST within the *Y. pseudotuberculosis* complex (Cui *et al*., 2008; Riehm *et al*., 2012).

Origin of spacers

BLASTN searches revealed that a number of spacer sequences showed similarity to various plasmid and bacteriophage sequences (Tables S8 and S9). It was interesting to notice that a few spacers (e.g. #585, #283, #82, #1206, #1154 and #1001) present in *Yersinia* species had matches to plasmids. Spacer #82 shows 97% identity (one base pair difference) to *Y. enterocolitica* plasmid pYE854 and *Y. pseudotuberculosis* IP31758 59 kb plasmid. In Table S9, spacers that have similarities with different bacteriophage sequences are shown. There was good identity to e.g. *Enterobacteria*, *Erwinia*, *Escherichia*, *Salmonella* and *Burkholderia* phages. In Table S10, spacer sequences were compared with whole genome sequences. The bacterial species earlier seen in Table S9 (bacteriophage hits) can also be seen in this table. Spacer #1697 appears in many bacterial species simply because this spacer occurs in the highly conserved 16S ribosomal RNA gene. In many instances, the spacers were located in prophage-like elements similar to *Yersinia*-specific spacers (see below).

Yersinia*-specific spacers.* Some spacers were present in the genomes or plasmids of other *Yersinia* species. Altogether 40 spacers showed significant similarity to sequences present in *Y. enterocolitica* 8081 genome (Table 2). The 8081 genome carries four prophage-like elements (Thomson *et al*., 2006), and 38 of the 40 spacer sequences were located within two of them, ϕYE98 (22 spacers) and ϕYE250 (16 spacers). A few strains carried two 8081 prophage-specific spacers.

Most spacers identified from *Y. pestis* were not shared with *Y. pseudotuberculosis*. The ones that are shared are shown in Tables S11 and S12. Table S13 is the conversion table for spacer nomenclature from previous studies to the present database (Cui *et al*., 2008; Riehm *et al*., 2012). Specifically, in the YP1 locus, spacers #403–405 were shared by a few *Y. pseudotuberculosis* strains, while in the YP3 locus spacer #507 seems to be present in all *Y. pestis* strains and is also common in *Y. pseudotuberculosis* (Tables S6 and S11). Interestingly, spacers #257, #1901 and #1902 are identical except for the one and two extra Gs present in the two latter ones respectively (Table S4). Spacer #1901 is very common in the YP2 locus of *Y. pestis* strains while spacer #257 is present in YP1 locus of 11 *Y. pseudotuberculosis* strains, and in two of the strains it is present in two copies. Furthermore, #257 appears in the YP3 locus in seven *Y. pseudotuberculosis* strains and

Table 2. Spacers of *Y. pseudotuberculosis* complex with matching sequences (> 88% identity in BLASTN search) to genomic sequence of *Y. enterocolitica* 8081 (GenBank Accession No. AM286415).

Spacer	Strain	YP locus	Location in 8081 genome	BLASTN search identity %
Spacers in prophage ϕYE98 (location in 8081 genome: 981223–1011295)				
1779	WP-931108	YP3	984218	88 (28/32)
1191	**79136**	YP3	984296	88 (28/32)
986	**WP-931110**	YP1	986574	97 (31/32)
1193	**79136**	YP3	987079	91 (29/32)
566	**2889, 2895, Y385, Y728, Y729**	YP1	987179	97 (31/32)
988	**WP-931110**	YP1	987234	97 (31/32)
244	PC95-219-1	YP1	988311	91 (29/32)
699	**H892-36-87**	YP1	991747	91 (29/32)
711	**H892-36-87**	YP1	991941	91 (29/32)
1471	IP31758, No-151, OK10700, **R30**	YP3	992073	100 (32/32)
1076	22917-2L, IP32952	YP2	993118	91 (29/32)
1203	8727-7, D426,D54,Gifu-liver	YP3	994228	94 (30/32)
1748	**R30**	YP3	994348	91 (29/32)
55	**BB28, Y74**	YP1	995024	94 (30/32)
86	**F-401-1,** Wla658, **WP-931110**	YP1	1001481	97 (31/32)
50	**BB28, Y74**	YP1	1004019	88 (28/32)
1629	**MW48, R103-2, R626R**	YP3	1004236	88 (28/32)
123	**H-3831**	YP1	1009835	100 (32/32)
79	**F-401-1**	YP1	1010599	91 (29/32)
647	**CN2**	YP1	1010603	91 (29/32)
1317	**CN3**-5end	YP3	1010802	88 (28/32)
846	PT245	YP1	1010919	94 (30/32)
Spacers in prophage ϕYE200 (location in 8081 genome: 1991720–2007210)				
885	**R104-2**	YP1	1993391	88 (28/32)
Spacers in prophage ϕYE250 (location in 8081 genome: 2503099–2554665)				
1315	**CN3**-5end	YP3	2527852	88 (28/32)
1681	Pa8728, d54	YP3	2533319	94 (30/32)
1270	B56, **No-21**	YP3	2533871	97 (31/32)
1256	AZ960106-1	YP3	2534586	100 (32/32)
1662	**MW48, R103-2, R626R**	YP3	2535837	94 (30/32)
1412	H146-84K, **R111, YPIII**	YP3	2538051	88 (28/32)
225	OK5466	YP1	2538088	94 (30/32)
570	**2889, 2895, Y385, 728, Y729**	YP1	2538983	97 (31/32)
1295	**CN2, R104-2**	YP3	2539577	100 (32/32)
121	**H-3831**	YP1	2540279	94 (30/32)
889	**R104-2**	YP1	2542146	100 (32/32)
1457	**H892-36-87**	YP3	2544550	88 (28/32)
918	**R111, YPIII**	YP1	2548392	94 (30/32)
821	**No-21**	YP1	2550610	91 (29/32)
695	**H892-36-87**	YP1	2552642	100 (32/32)
1234	93422-5end, **CN3**-5end	YP3	2552879	91 (29/32)
Other location *Ye2993 gltK*				
120	G5431	YP1	3264360	88 (28/32)

Strains in bold carry several spacers similar to *Y. enterocolitica* sequence. All spacers are 32 nt long, except #846 that is 33 nt long.

twice in one of the strains. As mentioned before, *Y. pseudotuberculosis* strains rarely contained more than two spacers in YP2 loci. In contrast, *Y. pestis* YP2 locus usually carries three to six spacers or more.

Finally, to extend the spacer comparisons between *Y. pseudotuberculosis* and *Y. pestis*, all the spacer sequences present in *Y. pseudotuberculosis*, *Y. similis* and *Y. wautersii* were used to search the *Y. pestis* genomes. Table S12 lists the 33 spacer sequences identified. The table shows seven hits to *Y. pestis* CRISPR elements, but also 14 hits to *Y. pestis* prophages.

pYptb32953. Altogether 32 unique spacers for the 27 702 bp cryptic plasmid pYptb32953 of *Y. pseudotuberculosis* IP32953 (Chain *et al.*, 2004) were identified in 34 strains. The distribution of the spacers along the plasmid sequence is shown in Table 3. No significant distribution bias can be detected. A majority (22) of the spacers map to the forward (+) and 10 map to the reverse (–) strand of the plasmid (Table 3). Spacer #1362 has two matches in the plasmid, one is a 100% match to nt 16931 (–) strand and the other a 31/32 (97%) match to nt 23321 (+) strand. Altogether 30 different strains had a spacer sequence

Table 3. The spacers specific for protospacers in pYptb32953, the cryptic 27 702 bp plasmid of *Y. pseudotuberculosis* strain IP32953 (Accession No. BX936400.1).

Spacer	Present in strains	Locus	Serotype	ST[a]	Location in plasmid (strand)	Protospacer sequence in pYptb32953 ± 6 nt. The GG PAM motif is grey shaded.	Identity % (32 nt)	Conjugation resistance type (Table 4)
234	OK5466	YP1	O:5b	73	3697 (+)	tagtct/gatgggtctcaaaatacgcaccaaagggaacg/**GG**aaaa	100	Intermediate
790	J51	YP1	O:13	47	5093 (+)	gatttc/aacgaaaaaaacgccggtaatgcgtcgattgt/**GG**ggac	100	Non-resistant
1285	BB28	YP3	O:2b	53	5312 (+)	gttaaa/agtggggaacctaccggatggaatccgtttcg/**G**ctgaa	100	Non-resistant
287	R103-1	YP1	O:3	58	5432 (+)	acccga/caggaaacccgcctcagtgacgccgttgatgc/t**G**ttat	100	Non-resistant
	S107	YP1	O:5b	87				--
1178	79136	YP3	O:1b	88	5469 (+)	ctgtta/tgttggggcttgaccacgccagccgtgaccac/**GG**Tatt	100	Non-resistant
789	J51	YP1	O:13	47	5751 (+)	tactgg/atggacggcgtttagtgctgtttattgatgag/**G**tctgg	100	Non-resistant
1450	H-3831	YP3	O:4a	48	7359 (+)	agtctt/ttcgagtcaatcatgggaagactatctttatt/**GG**cagc	100	Fully resistant
1167	774	YP3	O:4a	32	8228 (–)	gtgagc/gcggttaaataccccccgcattagtaatgaa/**GG**tgat	88	Fully resistant
464	Y722	YP1	unknown	19	9676 (–)	gcgagt/ttatcggggtcggttgcatcactaatgacatt/**GG**aaat	100	--
	Vlassel	YP1	O:3	57				Fully resistant
	IP32802	YP1	unknown	19				--
	Y716	YP1	unknown	19				--
	H1960003	YP1	unknown	unknown				--
784	J51	YP1	O:13	47	10080 (–)	gaaaac/accaaggtagtgacataaccggcgcgagcatt/aattac	100	Non-resistant
1454	H-3831	YP3	O:4a	48	10082 (+)	ttggtg/ttttcggatgatgaggcggttattagcactga/**GG**tggg	100	Fully resistant
1264	B56	YP3	R	9	10680 (+)	tgatcc/agcatattaacccgacgatggtattacgctat/**GG**cgat	100	Fully resistant
	No.21	YP3	O:1a	86				Intermediate
556	22917/2L	YP1	O:5a	16	12868 (+)	cagcat/gagaactatgtgcatctgttttatccgtcaga/**GG**gtgg	100	Intermediate
1820	Gifu-liver	YP1	O:1b	22	12869 (+)	agcatg/agaactatgtgcatctgttttatccgtcagag/**GG**tgga	94	Intermediate
1632	R103-2 (*Y. sim*)	YP3	O:5b	45	13031 (+)	gttcag/gtcatgatgggttaatcacgtcaatgcacgct/**GG**cacc	100	Fully resistant
	MW48 (*Y. sim*)	YP3	O:9	80				Fully resistant
	R626R (*Y. sim*)	YP3	O:9	83				Non-resistant
1528	R103-1	YP3	O:3	58	13276 (+)	aaatga/aacatttaattagaccatgttgggtggctgtc/**GG**tttg	100	Non-resistant
	J92	YP3	O:13	82				Non-resistant
	S107	YP3	O:5b	87				--
134	H-3831	YP1	O:4a	48	13374 (+)	acaccg/acgacggttaacagcaccttgccccagtggca/**GG**aaaa	100	Fully resistant
1449	H-3831	YP3	O:4a	48	13453 (+)	catggt/caaagcgcatcacggatttttcaggggggataac/**G**Gcagc	100	Fully resistant
1444	H-3831	YP3	O:4a	48	13545 (+)	gttgcg/agtaacagcagctcggcatggtttaacaccgc/**GG**catg	94	Fully resistant
760	J51	YP1	O:13	47	13640 (+)	ttgttt/aacctgcaatacggttgatgtttatttgtctc/**GG**tcag	100	Non-resistant
133	H-3831	YP1	O:4a	48	14452 (–)	cacaga/ttttatttggtgatatttgaattgatcggcaa/**GG**cgta	100	Fully resistant
1154	H892-36/87	YP3	O:1a	12	15362 (+)	agttca/caaacaacattaaataatgctaataattatac/t**G**attc	100	Non-resistant
	CN3	YP3	O:14 (R)	17				Fully resistant
	2889	YP3	O:1b	43				--
	2895	YP3	O:1b	43				--
	Y734	YP3	unknown	43				--
1361	DD110	YP3	O:6	11	15405 (+)	caagag/tgagtaacattacaaatgtagctattgaagag/**G**ctaat	100	Fully resistant
	Pa8728	YP3	R	60				Intermediate
1579	KP1244-2B	YP3	O:2c	56	16774 (–)	tttttt/cattacctcattgatactcggaacttcatcaa/**GG**cagt	97	Fully resistant
1362	DD110	YP3	O:6	11	16931 (–)	cgttcg/cgggggtggcttgttggcctccccgcttcact/**GG**cttt	100	Fully resistant
1586	MW101-1	YP3	O:4b	28	17275 (–)	ggggat/atatccccccaaattaaccgcccactggggtg/**GG**cttt	100	Non-resistant
	PC95-219-1	YP3	R	84				Non-resistant
1585	MW101-1	YP3	O:4b	28	17545 (+)	cggagt/ggcgattgccgttgattggggtaactgcaagt/tatcgc	94	Non-resistant
	PC95-219-1	YP3	R	84				Non-resistant
218	WP-931109	YP1	O:15	22	21892 (–)	tgctct/tttcctcggtctggctggcggtgactgtcgcc/**GG**ttgg	97	Intermediate
	OK5466	YP1	O:5b	73				Intermediate
1362	DD110	YP3	O:6	11	23321 (+)	gttctc/gggggtggcttgtcggcctccccgcttcactg/**G**ctttt	97	Fully resistant
283	Pa8728	YP1	R	60	24168 (–)	agtggg/tgatgtgcagcatgaaagctatattgcctcat/**GG**ctta	100	Intermediate
232	OK5466	YP1	O:5b	73	24547 (–)	ctgtgt/ttaatgtccagcaaatagacggcttgccactag/a**G**aca	94	Intermediate
1407	No.21	YP3	O:1a	86	27085 (+)	tggggc/cgtagtgctgtttaaccgtttttgtggtcag/**GG**tatg	100	Intermediate
	H141-84	YP3	unknown	9				--
	Y717	YP3	unknown	9				--

a. ST, Multilocus sequence type according to MLST Databases at the Unversity of Warwick, Warwick Medical School (http://mlst.ucc.ie/mlst/dbs/Ypseudotuberculosis/GetTableInfo_html) (Laukkanen-Ninios *et al*., 2011).
Strains OK5466, J51, R103-1, S107, H-3831, No.21, DD110, Pa8728, MW101-1, and PC95-219-1 carry several different pYptb32953 spacers.

with 100% identity to the plasmid pYptb32953 sequence, and additionally, a few spacers with some mismatches were identified. To see whether the spacer-carrying strains would reject the pYptb32953 plasmid and the spacers-free strains would accept it in conjugation experiments, we tagged the plasmid with a *cat* gene (see *Experimental procedures*). We first demonstrated that pYptb32953 is indeed a self-conjugative plasmid as

Table 4. Influence of pYptb32953-specific spacers on pYptb32953::*cat* conjugation and plasmid mobilization restriction by *Y. pseudotuberculosis* and *Y. similis* strains.

		pYptb32953::*cat* conjugations			Mobilization frequency (%)	
Strain	Spacer no	I	II	III	pTM100-waaF	pTM100-CRISPR
Non-resistant group (> 0.2)						
PB1	–	0.6	0.5	1	~7	~2
Wla658	–	1.67	4.5	1.88		
DC356-2	–	1	0.82	0.9		
257	1115	1	1.67	2.5		
J51	760.784.789.790	0.35	0.33	1		
R103-1	287.1528	0.55	4.5	2.5	5–10	1
J92	1528	0.3	0.7	0.71		
79136	1178	2.33	2.33	2.8		
R626R (*Y. similis*)	1632	0.83	0.67	1	1	0.1
MW101-1	1585.1586	0.2	0.5	0.53		
PC95-219-1	1586	0.33	0.4	0.5		
H892-36/87	1154	1.67	1	1		
BB28	1285	1	1.25	1.33		
Intermediate resistant group (0.01–0.2)						
Gifu-Liver	1820	0.044	0.008	0.04		
22917/2L	556	0.1	0.08	0.013		
No.21	1264.1407	0.033	0.2	0.022		
Pa8728	283.1361	0.13	0.2	0.083		
WP-931109	218	0.13	0.2	0.13		
OK5466	218.232.234	0.067	0.05	0.17		
Fully resistant group (0–0.0099)						
YPIII	–	0	0.0063	0		
Vlassel	464	0.031	0.002	0.00067		
H-3831	133.134.1444.1449.1450.1454	0	0.004	0.004	1	< 0.01
774	1167	0	0	0.001		
B56	1264	0.004	0.01	0.033		
R103-2 (*Y. similis*)	1632	0	0	0	1–5	0.1–0.5
MW48 (*Y. similis*)	1632	0.00005	0	0	0.1–1	0
WP-931205	1010	0	0	0		
Toyama-60	1775	0	0	0		
CN3	1154	0	0.0002	0		
DD110	1361.1362	0.048	0.01	0.0078		
KP1244-2B	1579	0.00017	0.00011	0.000083		

Presented are for pYptb32953::*cat* conjugations transconjugant/recipient ratios of three parallel mating experiments. Variations between the ratios of the three matings are due to inaccuracy in the serial dilution drop method used for measuring bacterial concentrations in mating mixtures (see Fig. 3). Mobilization of plasmids pTM100-waaF and pTM100-CRISPR was performed with six strains, highlighted in grey, and the mobilization frequencies are given as transconjugant/recipient percentages.

predicted based on its annotated sequence (Chain *et al.*, 2004). The pYptb32953::*cat* transferred itself efficiently from IP32953 to *E. coli* strain PM191NaR (data not shown). The conjugation frequencies to 31 different *Y. pseudotuberculosis* strains were determined (Table 4). Examples of the conjugation experiments are shown in Fig. 3. Among the strains, we observed three levels of restriction to pYptb32953::*cat* conjugation: (i) non-resistant group to which the plasmid transferred without any apparent restriction and under the experimental conditions used 30–100% of recipients were transformed, (ii) intermediate resistant group where 1–20% of recipients were transformed and (iii) fully resistant group with less than 0.01% transformants. While 11 of 12 among the fully resistant strains carried a pYptb32953-specific spacer, 10 spacer-carrying strains were among the 13 strains in the non-resistant group. Among the four spacer-free strains tested, three were non-resistant and one, YPII, was fully resistant (Table 4).

To find out whether the resistance differences in the spacer-carrying strains could be explained by the presence or absence of the type-IF specific PAM motif GG at the 3′-end of the protospacer (Mojica *et al.*, 2009; Wiedenheft *et al.*, 2011; Cady *et al.*, 2012), the pYptb32953 spacer-flanking sequences were analysed (Table 3). Altogether 23 of the 32 protospacers were flanked by the GG PAM motif and there was no correlation between the presence or absence of the PAM motif and resistance. For example, the non-resistant strain J51 carries four spacers and two of the protospacers carry the GG PAM motif. Spacer #1632 containing the PAM motif is present in three *Y. similis* strains. One of the strains is

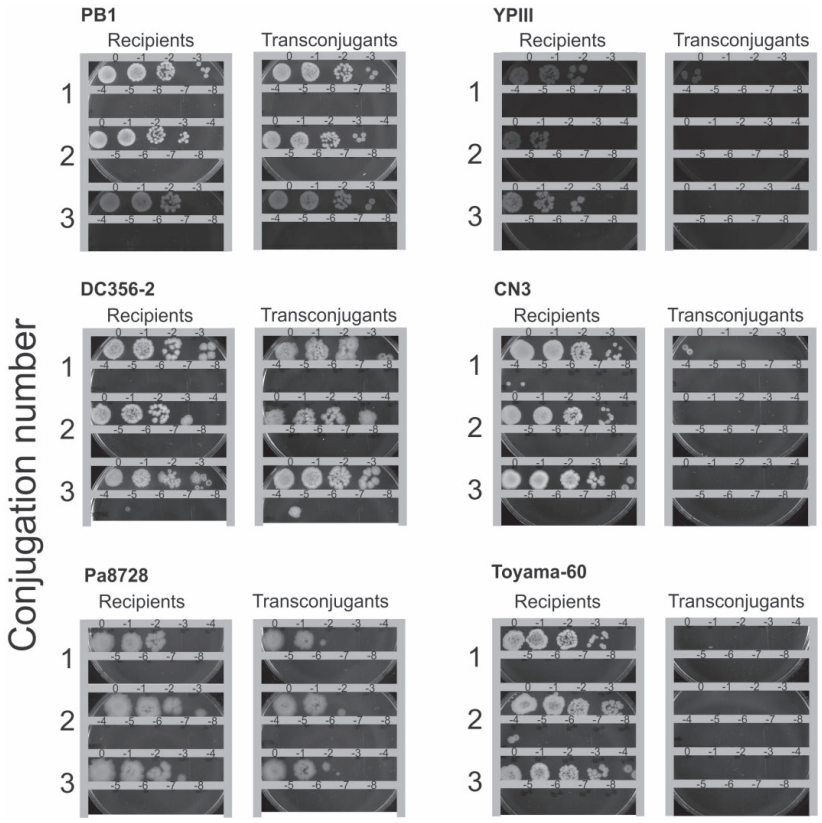


Fig. 3. Conjugation experiments showing the serial 10-fold dilution plating of 5 µl drops on selective CIN-agar plates without or with Clm allowing the growth from the mating mixture of all the recipient bacteria or only the transconjugant bacteria respectively. PB1 and DC356-2 represent non-resistant strains, Pa8728 represents intermediate resistant strains, and YPIII, CN3 and Toyama-60, the fully resistant strains.

non-resistant and two are fully resistant. In addition, there are fully resistant strains that carry a spacer missing the PAM motif (Table 3).

To find out whether the CRISPR-Cas system in the *Y. pseudotuberculosis* strains is functional, we constructed a pair of plasmids based on mobilizable plasmid pTM100 (Michiels and Cornelis, 1991). pTM100-CRISPR carries a 909 bp fragment of pYptb32953 (nucleotides 13002–13910) that contains six protospacers present as spacers in eight of the strains (Table 4). pYM100-waaF was used as a spacer-free control plasmid. The plasmids were mobilized into a set of six spacer-carrying and spacer-free strains representing the non- and fully-resistant groups. A functional CRISPR-Cas system should restrict the mobilization of pTM100-CRIPSR but not that of pTM100-waaF into a spacer-carrying strain, and there should not be any differences in mobilization of either plasmid into a spacer-free strain. The results presented in Table 4 demonstrate that mobilization frequency of pTM100-CRIPSR to all five spacer-carrying strains was significantly lower than that of pTM100-waaF while no difference could be seen with spacer-free strain PB1. These results demonstrated that the CRISPR-Cas system in *Y. pseudotuberculosis* is functional.

As a single nucleotide change in the protospacer sequence in a phage genome may allow the phage to escape the CRISPR immunity (Levin *et al.*, 2013), we also checked for this possibility as some of the spacers had one to four mismatches with the protospacers (88–97% identity over 32 nt, Table 3). Also this seemed not to correlate with the resistance as spacer #1167 in strain 774 had four mismatches with the protospacer but the strain was fully resistant. Also #1579 in strain KP1244-2B had one mismatch but the strain was fully resistant.

Discussion

The three CRISPR loci of 335 *Y. pseudotuberculosis* complex strains were analysed. Altogether 1902 different spacers were found, and surprisingly, little overlap between the strains was observed. In spite of this, we noticed some correlation between the *Y. pseudotuberculosis* STs and CRISPR spacers. To visualize evolutionary relationships between the strains, we aligned the spacer profiles of the strains based on both the YP1 and YP3 spacers, but these alignments showed no congruence. This is a strong argument for the influence of HGT in shaping the genomes of *Y. pseudotuberculosis* and that specifically influences the YP1 and YP3 loci. This is supported by analogous reassortment of CRISPR loci in *Sulfolobus islandicus* (Held *et al*., 2013) and in *E. coli* (Almendros *et al*., 2014). On the other hand, it also reflects the facts (i) that we intentionally selected the strains to represent as divergent collection of *Y. pseudotuberculosis* complex strains as possible, and (ii) that the number of the strains included in the study was still relatively small. Therefore, to draw meaningful evolutionary conclusions, CRISPR sequence data from larger number of strains is needed. The alignments using the YP2 loci demonstrated the high similarity between the *Y. similis* sequences and their distinct separation from other *Y. pseudotuberculosis* complex species (Table S7).

Earlier studies suggested that *Y. pseudotuberculosis* ST43 is the closest relative to *Y. pestis* (Laukkanen-Ninios *et al*., 2011; Riehm *et al*., 2012). Interestingly, the spacers of the six ST43 strains investigated here shared almost no spacers with *Y. pestis* (Table S11). It will be interesting to investigate more ST43 strains. The most ancestral *Y. pestis* YP1 and YP3 spacers are observed in *Y. pseudotuberculosis* ST14, ST16, ST41 and ST87 (Table S6) in a similar position but these STs are not close neighbours to *Y. pestis* or to each other (Laukkanen-Ninios *et al*., 2011). This indicates that these spacers were acquired well before *Y. pestis* speciation and were subsequently lost in most *Y. pseudotuberculosis* lineages or that CRISPR loci may be transferred horizontally.

Another peculiarity in our dataset was the observation that in some cases, a spacer was found from two different loci in one strain. Furthermore, certain spacers were shared between strains but occurred at different positions or even in different loci. This may be due to the fact that different strains have been invaded by the protospacer-carrying DNA in separate occasions.

Previous studies have shown that the CRISPR variability may be used for typing bacterial species, even though the CRISPR sequence diversity was not as wide as in the *Y. pseudotuberculosis* complex (Riehm *et al*., 2012). For example, Fabre and others concluded that the CRISPR spacer content in *Salmonella* correlated with MLST and serotyping results, and they indicated that CRISPR analysis may be a powerful tool for molecular typing of *Salmonella* isolates (Fabre *et al*., 2012; Shariat *et al*., 2013a,b). It was also shown that *E. coli* CRISPR typing combined with MLST analysis could differentiate strains from a single clonal group (Touchon *et al*., 2011). CRISPR has also proven to be a good typing tool for the clonal *Y. pestis* and hypothetical evolutionary models have been created based on the CRISPR spacer arrays (Cui *et al*., 2008). This has to be treated with utmost care as our present results revealed very big differences between the *Y. pseudotuberculosis* strains and indications that evolution within *Y. pseudotuberculosis* might not be clonal. This method may be very useful for forensic applications; however, this would require an extensive reference collection. We show here that each ST would need to be considered almost as a single entity, as previously done for ST90 (*Y. pestis*).

The most common spacers had significant similarities mainly with *Yersinia* species. Comparison of spacer with plasmid sequences indicated one notable plasmid, pYptb32953, the cryptic 27 702 kb plasmid of *Y. pseudotuberculosis* strain IP32953 which had significant similarities with 32 spacers.

When the pYptb32953-specific spacers were identified from the 31 strains, we set out to test whether the presence or absence of the spacers would influence the conjugation frequency of pYptb32953::*cat* to a strain. As spacer-negative control strains, we selected four strains. Our hypothesis was that the plasmid would transform the spacer-negative strains but not the spacer-positive ones. The results shown in Table 4 were unexpected and demonstrated that bacteria are versatile. In addition, the finding that pYptb32953::*cat* conjugated efficiently to 10 spacer-positive strains raised the possibility that the CRISPR-Cas system in these strains would not be functional. The mobilization experiments carried out with the pTM100-waaF/pTM100-CRISPR plasmids, however, clearly demonstrated that the CRISPR-Cas system is functional also in these strains. Interestingly, we observed that the CRISPR-Cas system-based resistance was not 100% tight but could reduce the mobilization frequency to *c*. 10% of the spacer-free mobilization. We can make important conclusions from the results. First, strain YPIII that lacks any pYptb32953-specific spacers was fully resistant to pYptb32953::*cat* transformation. The strain likely carries an efficient restriction modification system or lacks a receptor for the pYptb32953 conjugation apparatus. Second, the presence of a group of spacer-positive strains that showed intermediate resistance to pYptb32953::*cat* transformation shows that the resistance can be leaky. We can speculate that one or more other defence systems in addition to the CRISPR-Cas system are required to achieve full conjugation resistance. Thus,

it is likely that these systems are not present in the non-resistant and intermediate resistant groups. Further studies are warranted to elucidate the molecular mechanisms behind these phenomena.

When comparing the *Y. pseudotuberculosis*, *Y. similis* and *Y. wautersii* spacer sequences to the *Y. pestis* genomic sequences (Table S12), altogether 33 spacers showed significant similarity. Seven of the spacers were spacers in the CRISPR loci of *Y. pestis*, and of the remaining 26 spacers, 14 had hits in the prophage sequence in the *Y. pestis* genomes. Fewer spacer sequence hits were observed with *Y. pestis* strains 91001 and Antiqua reflecting the fact that they are missing certain prophages present in other *Y. pestis* strains. That has been described also earlier (Song *et al*., 2004; Chain *et al*., 2006).

We faced some difficulties in both PCR amplification and sequencing some of the CRISPR loci, and to overcome this whole genome sequencing will be used in future. Whole genome sequencing will also open new possibilities to distinguish *Y. pseudotuberculosis* strains from each other.

Final conclusions

Our results suggest that *Y. pestis* after divergence from *Y. pseudotuberculosis* has lived protected or secluded life, and it has not encountered many foreign transforming DNAs at least when measured with numbers of CRISPR spacers. Apparently, there are rare instances in the *Y. pestis* life cycle where it is exposed to other bacteria or bacteriophages. This is realistic as the infected tissues in rodents and/or humans after fleabite are practically sterile; however, while we do not fully understand the life style of *Y. pestis* in the environment, the microbiota of the flea might be a likely source of foreign DNA. Therefore, *Y. pestis* has a relatively low number of spacers compared with *Y. pseudotuberculosis.* The latter, on the contrary, is widely spread in nature and seems to have been highly exposed to various insulting genetic elements and this is visible in the high number of spacers present in a single strain, for example the strain YPIII has altogether 75 spacers.

Experimental procedures

Bacterial isolates

A total of 76 *Y. pseudotuberculosis*, 10 *Y. similis* and 4 *Y. wautersii* isolates from the Skurnik laboratory strain collection were analysed in this study (Table S1). These isolates were selected to cover the largest possible geographic area, host range and to represent as many of the 21 serotypes as possible. Altogether 83 of the 90 STs (Laukkanen-Ninios *et al*., 2011) were represented, each with a single isolate except for ST3, ST14 and ST43 that were represented by three, four and three isolates, respectively. In addition, sequence data of 40 *Y. pseudotuberculosis* strains and 195 *Y. pestis* strains from earlier investigations were included [(Pourcel *et al*., 2005; Cui *et al*., 2008) and Vergnaud G and Gorgé O, unpublished] as well as CRISPR loci sequences of published complete genome sequences of four *Y. pseudotuberculosis* and six *Y. pestis* strains (Table S1).

Culture conditions

The bacterial strains were grown in lysogeny broth (LB) (Bertani, 2004): the *Yersinia* strains at 20–22°C unless otherwise mentioned and *E. coli* strains at 37°C. LB agar (LA) plates used for solid cultures were prepared by supplementing LB with 1.5% Bacto Agar. *Yersinia* selective CIN-agar plates (Oxoid) were used in conjugation experiments. When required, appropriate antibiotics were added as follows: chloramphenicol (Clm) 20 $\mu g\ ml^{-1}$, nalidixic acid (Nal) 100 $\mu g\ ml^{-1}$, kanamycin (Kan) 100 $\mu g\ ml^{-1}$ and diaminopimelic acid (Dap) 0.3 mM.

DNA extraction

Genomic DNA was isolated using the JetFlex DNA isolation kit (GENOMED GmdH, Löhne, Germany).

Sequencing of the CRISPR loci

The three CRISPR loci of *Y. pseudotuberculosis* complex were targeted by PCR based on previously published *Y. pestis* CRISPR primer sequences (Le Flèche *et al*., 2001; Pourcel *et al*., 2005). New primers were designed for CRISPR YP2 as the earlier published *Y. pestis* primers failed with many *Y. pseudotuberculosis* strains. The CRISPR loci-specific primers are presented in Table S2. PCR reactions were run in a final volume of 50 µl containing 50 pmol of each primer, 5 µl of 10× Dynazyme II buffer, 200 µM of each dNTP, 1 U of Dynazyme II and 150 ng of template DNA. PCR consisted of the following steps: 94°C for 3 min, 32 cycles of denaturation at 94°C for 40 s, annealing at 53°C for 40 s and extension at 72°C for 3 min, and final extension step at 72°C for 10 min. PCR-amplified fragments were visualized after agarose gel electrophoresis (1.2% agarose) by ethidium bromide staining. The DNA fragments were sent to Institute for Molecular Medicine Finland core facility for sequencing after exonuclease I (Neo Lab) and Shrimp Alkaline Phosphatase (Promega) treatment. The fragments were sequenced using the Applied Biosystems Dye Terminator Kit (BigDye v.3.1) and ABI 3730xl DNA Analyser. The CRISPR-loci specific primers were used as sequencing primers from the fragment ends and internal primers were designed for sequencing long PCR fragments.

Raw sequence read data were analysed and assembled to contigs using either Sequencer 5.1 (Gene Codes Corporation) or the Staden Package (Staden, 1996). Before submitting the contig sequence data to CRISPRFinder tool at CRISPRs Web Server (http://crispr.u-psud.fr/) (Grissa *et al*., 2007a,b), the data were combined with available CRISPR sequence data from *Y. pestis* and *Y. pseudotuberculosis* strains (Pourcel *et al*., 2005; Cui *et al*., 2008; Riehm *et al*.,

2012) (Table S1). The CRISPRFinder tool returned the recognized spacers with unique randomly selected identification numbers. The GenBank non-redundant (nr) nucleotide sequence database was searched for individual spacer sequences using the BLASTN tool (Altschul *et al*., 1990).

Accession numbers

All the sequences reported in this article including the earlier published but not submitted sequence data (Pourcel *et al*., 2005; Cui *et al*., 2008; Riehm *et al*., 2012) were deposited to nucleotide sequence databases. The accession numbers for the YP1, YP2 and YP3 loci are listed in Table S1.

*Construction of pYptb32953::*cat *and pTM100-CRISPR*

Primers specific for pYptb32953 (Accession No. BX936400.1), the 27 kb cryptic plasmid of *Y. pseudotuberculosis* IP32953 (Table S2) were used to amplify a 797 bp fragment of pYptb32953 from a plasmid miniprep template. The PCR fragment was purified and digested with EcoRI followed by ligation with EcoRI-digested and Shrimp Alkaline Phosphatase (SAP)-treated suicide vector pSW23T (Demarre *et al*., 2005). The ligation mixture was electroporated into *E. coli* strain ω7249 that is Kan^R (Babic *et al*., 2008). Transformants carrying the correct insert were identified by PCR and the isolated plasmid named as pSW23T-pIP was further confirmed by restriction digestions. The suicide construct was introduced to *Y. pseudotuberculosis* IP32953 by conjugation from *E. coli* ω7249/pSW23T-pIP, and Clm^R transconjugants were selected with LA-Clm plates where the donor was unable to grow due to its requirement for Dap. One of the Clm^R transconjugants was named as IP32953/pYptb32953::*cat* and used as a donor to introduce the tagged plasmid into *E. coli* strains PM191NaR, a Nal^R spontaneous derivative of PM191 (Meacock and Cohen, 1980) to obtain *E. coli* PM191NaR/pYptb32953::*cat*, and to strain ω7249 to obtain *E. coli* ω7249/pYptb32953::*cat*. pTM100-CRISPR was constructed by cloning a PCR-amplified 909 bp DNA fragment of pYptb32953 (nucleotides 13 002–13 910; for PCR primers, see Table S2) into EcoRV site of pTM100 (Michiels and Cornelis, 1991). pTM100-CRISPR was electroporated into *E. coli* strain ω7249. pTM100-waaF (Noszczynska *et al*., 2015) was used as spacer-free control plasmid in mobilization experiments.

Conjugation frequency assays

The *E. coli* ω7249/pYptb32953::*cat*, ω7249/pTM100-CRISPR and ω7249/pTM100-waaF strains were used as donor strains to determine conjugation/mobilization frequencies into a set of *Y. pseudotuberculosis* strains. The donor bacteria were grown in LB-Kan-Clm-Dap at 37°C for 16 h, the culture was diluted 1:10 in fresh medium and incubated for an additional 3 h, washed and resuspended into PBS to OD_{600} of ~1.0. The recipient bacteria were grown in LB at 22°C for 16 h, the culture was diluted 1:10 in fresh medium and incubated for an additional 3 h, washed and resuspended into PBS to OD_{600} of ~1.0. For each recipient strain, three parallel matings were prepared. Equal amounts of the donor and recipient suspensions were mixed, and 100 µl aliquots were pipetted in the middle of three parallel LA plates supplemented with Dap but without antibiotics. The plates were incubated at 37°C for 16 h. The bacteria on the plates' surface were resuspended into 1 ml of PBS. Two hundred microlitre aliquots were recovered from each and diluted with PBS to OD_{600} of 0.2. The concentrations of donor, recipient and transconjugant bacteria in these mating mixtures were determined by pipetting 5 µl drops of 10^0–10^{-8} diluted mixtures on LA-Kan-Clm-Dap plates (for donor counts), CIN plates (for total recipient counts) and CIN-Clm plates (for transconjugant counts). The donor plates were incubated for 24 h at 37°C and the recipient and transconjugant plates at 22°C for 48 h. The colonies in the last dilutions showing growth were counted. Conjugation frequencies were expressed as ratios between the transconjugant and recipient concentrations.

Acknowledgements

This work is part of the European Biodefence project EDA B-1325-ESM4-GP. Scientific contribution to this work has been accomplished by representatives from France and Finland. Work by GV andOG was supported by the French ministry of defense. Anita Liljegren and Juha Laitinen are thanked for expert technical help. The authors declare no conflict of interest.

References

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carniel, E. (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* **96:** 14043–14048.

Aleksic, S., Bockemühl, J., and Wuthe, H.H. (1995) Epidemiology of *Y. pseudotuberculosis* in Germany, 1983–1993. *Contrib Microbiol Immunol*. **13:** 55–58.

Almendros, C., Mojica, F.J., Diez-Villasenor, C., Guzman, N.M., and Garcia-Martinez, J. (2014) CRISPR-Cas functional module exchange in *Escherichia coli*. *MBio* **5:** e00767–13.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Babic, A., Guerout, A.M., and Mazel, D. (2008) Construction of an improved RP4 (RK2)-based conjugative system. *Res Microbiol* **159:** 545–549.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., *et al*. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:** 1709–1712.

Bertani, G. (2004) Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol* **186:** 595–600.

Blouin, Y., Cazajous, G., Dehan, C., Soler, C., Vong, R., Hassan, M.O., *et al*. (2014) Progenitor '*Mycobacterium canettii*' clone responsible for lymph node tuberculosis epidemic, Djibouti. *Emerg Infect Dis* **20:** 21–28.

Bogdanovich, T., Carniel, E., Fukushima, H., and Skurnik, M. (2003) Use of O-antigen gene cluster-specific PCRs for the identification and O-genotyping of *Yersinia pseudotuberculosis* and *Yersinia pestis*. *J Clin Microbiol* **41:** 5103–5112.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151:** 2551–2561.

Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O'Toole, G.A. (2012) The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J Bacteriol* **194:** 5728–5738.

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., *et al.* (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* **101:** 13826–13831.

Chain, P.S., Hu, P., Malfatti, S.A., Radnedge, L., Larimer, F., Vergez, L.M., *et al.* (2006) Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J Bacteriol* **188:** 4453–4463.

Chauvaux, S., Dillies, M.A., Marceau, M., Rosso, M.L., Rousseau, S., Moszer, I., *et al.* (2011) In silico comparison of Yersinia pestis and *Yersinia pseudotuberculosis* transcriptomes reveals a higher expression level of crucial virulence determinants in the plague bacillus. *Int J Med Microbiol* **301:** 105–116.

Ch'ng, S.L., Octavia, S., Xia, Q., Duong, A., Tanaka, M.M., Fukushima, H., and Lan, R. (2011) Population structure and evolution of pathogenicity of *Yersinia pseudotuberculosis*. *Appl Environ Microbiol* **77:** 768–775.

Cui, Y., Li, Y., Gorge, O., Platonov, M.E., Yan, Y., Guo, Z., *et al.* (2008) Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE* **3:** e2652.

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., *et al.* (2013) Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci USA* **110:** 577–582.

Delannoy, S., Beutin, L., Burgos, Y., and Fach, P. (2012a) Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains by use of the CRISPR locus as a target for a diagnostic real-time PCR. *J Clin Microbiol* **50:** 3485–3492.

Delannoy, S., Beutin, L., and Fach, P. (2012b) Use of clustered regularly interspaced short palindromic repeat sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR. *J Clin Microbiol* **50:** 4035–4040.

Demarre, G., Guerout, A.M., Matsumoto-Mashimo, C., Rowe-Magnus, D.A., Marliere, P., and Mazel, D. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol* **156:** 245–255.

Diez-Villasenor, C., Almendros, C., Garcia-Martinez, J., and Mojica, F.J. (2010) Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* **156:** 1351–1361.

Fabre, L., Zhang, J., Guigon, G., Le Hello, S., Guibert, V., Accou-Demartin, M., *et al.* (2012) CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PLoS ONE* **7:** e36995.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., *et al.* (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468:** 67–71.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007a) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8:** 172.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007b) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35:** W52–W57.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1:** e60.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., *et al.* (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139:** 945–956.

Harbeck, M., Seifert, L., Hansch, S., Wagner, D.M., Birdsell, D., Parise, K.L., *et al.* (2013) *Yersinia pestis* DNA from skeletal remains from the 6(th) century AD reveals insights into Justinianic Plague. *PLoS Pathog* **9:** e1003349.

Held, N.L., Herrera, A., and Whitaker, R.J. (2013) Reassortment of CRISPR repeat-spacer loci in *Sulfolobus islandicus*. *Environ Microbiol* **15:** 3065–3076.

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., *et al.* (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* **190:** 1401–1412.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., *et al.* (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* **35:** 907–914.

Karginov, F.V., and Hannon, G.J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* **37:** 7–19.

Laukkanen-Ninios, R., Didelot, X., Jolley, K.A., Morelli, G., Sangal, V., Kristo, P., *et al.* (2011) Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. *Environ Microbiol* **13:** 3114–3127.

Le Flèche, P., Hauck, Y., Onteniente, L., Prieur, A., Denoeud, F., Ramisse, V., *et al.* (2001) A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol* **1:** 2.

Levin, B.R., Moineau, S., Bushman, M., and Barrangou, R. (2013) The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLoS Genet* **9:** e1003312.

Liu, F., Barrangou, R., Gerner-Smidt, P., Ribot, E.M., Knabel, S.J., and Dudley, E.G. (2011) Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl Environ Microbiol* **77:** 1946–1956.

McGhee, G.C., and Sundin, G.W. (2012) *Erwinia amylovora* CRISPR elements provide new tools for evaluating strain diversity and for microbial source tracking. *PLoS ONE* **7:** e41706.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9:** 467–477.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013) The basic building blocks and evolution of CRISPR-cas systems. *Biochem Soc Trans* **41:** 1392–1400.

Meacock, P.A., and Cohen, S.N. (1980) Partitioning of bacterial plasmids during cell division a cis-acting locus that accomplishes stable plasmid inheritance. *Cell* **20:** 529–642.

Michiels, T., and Cornelis, G.R. (1991) Secretion of hybrid proteins by the *Yersinia* Yop export system. *J Bacteriol* **173:** 1677–1685.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155:** 733–740.

Morelli, G., Song, Y., Mazzoni, C.J., Eppinger, M., Roumagnac, P., Wagner, D.M., *et al.* (2010) *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* **42:** 1140–1143.

Naktin, J., and Beavis, K.G. (1999) Yersinia enterocolitica *and* Yersinia pseudotuberculosis. *Clin Lab Med* **19:** 523–536, vi.

Noszczynska, M., Kasperkiewicz, K., Duda, K.A., Podhorodecka, J., Rabsztyn, K., Gwizdala, K., *et al.* (2015) Serological characterization of the enterobacterial common antigen substitution of the lipopolysaccharide of *Yersinia enterocolitica* O:3. *Microbiology* **161:** 219–227.

Pourcel, C., and Drevet, C. (2013) Occurence, diversity of CRISPR-Cas systems and genotyping implications. In *CRISPR-Cas Systems*. Barrangou, R., and van der Oost, J. (eds). Berlin, Heidelberg: Springer-Verlag, pp. 33–59.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151:** 653–663.

Rezzonico, F., Smits, T.H., and Duffy, B. (2011) Diversity, evolution, and functionality of clustered regularly interspaced short palindromic repeat (CRISPR) regions in the fire blight pathogen *Erwinia amylovora*. *Appl Environ Microbiol* **77:** 3819–3829.

Riehm, J.M., Vergnaud, G., Kiefer, D., Damdindorj, T., Dashdavaa, O., Khurelsukh, T., *et al.* (2012) *Yersinia pestis* lineages in Mongolia. *PLoS ONE* **7:** e30624.

Savin, C., Martin, L., Bouchier, C., Filali, S., Chenau, J., Zhou, Z., *et al.* (2014) The *Yersinia pseudotuberculosis* complex: characterization and delineation of a new species, *Yersinia wautersii. Int J Med Microbiol* **304:** 452–463.

Schouls, L.M., Reulen, S., Duim, B., Wagenaar, J.A., Willems, R.J., Dingle, K.E., *et al.* (2003) Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. *J Clin Microbiol* **41:** 15–26.

Shariat, N., and Dudley, E.G. (2014) CRISPRs: molecular signatures used for pathogen subtyping. *Appl Environ Microbiol* **80:** 430–439.

Shariat, N., Kirchner, M.K., Sandt, C.H., Trees, E., Barrangou, R., and Dudley, E.G. (2013a) Subtyping of *Salmonella enterica* serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results. *J Clin Microbiol* **51:** 2328–2336.

Shariat, N., Sandt, C.H., DiMarzio, M.J., Barrangou, R., and Dudley, E.G. (2013b) CRISPR-MVLST subtyping of *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Heidelberg and application in identifying outbreak isolates. *BMC Microbiol* **13:** 254.

Skurnik, M., Peippo, A., and Ervelä, E. (2000) Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. *Mol Microbiol* **37:** 316–330.

Song, Y., Tong, Z., Wang, J., Wang, L., Guo, Z., Han, Y., *et al.* (2004) Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res* **11:** 179–197.

Sontheimer, E.J., and Marraffini, L.A. (2010) Microbiology: slicer for DNA. *Nature* **468:** 45–46.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* **82:** 237–266.

Souza, R.A., Pitondo-Silva, A., Falcao, D.P., and Falcao, J.P. (2010) Evaluation of four molecular typing methodologies as tools for determining taxonomy relations and for identifying species among *Yersinia* isolates. *J Microbiol Meth* **82:** 141–150.

Sprague, L.D., Scholz, H.C., Amann, S., Busse, H.J., and Neubauer, H. (2008) Yersinia similis *sp. nov. Int J Sys Evol Microbiol* **58:** 952–958.

Staden, R. (1996) The Staden sequence analysis package. *Mol Biotechnol* **5:** 233–241.

Tauxe, R.V. (2004) Salad and pseudoappendicitis: *Yersinia pseudotuberculosis* as a foodborne pathogen. *J Infect Dis* **189:** 761–763.

Thomson, N.R., Howard, S., Wren, B.W., Holden, M.T., Crossman, L., Challis, G.L., *et al.* (2006) The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genet* **2:** e206.

Touchon, M., Charpentier, S., Clermont, O., Rocha, E.P., Denamur, E., and Branger, C. (2011) CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J Bacteriol* **193:** 2460–2467.

Voskressenskaya, E., Leclercq, A., Tseneva, G., and Carniel, E. (2005) Evaluation of ribotyping as a tool for molecular typing of *Yersinia pseudotuberculosis* strains of worldwide origin. *J Clin Microbiol* **43:** 6155–6160.

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., *et al.* (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci USA* **108:** 10092–10097.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482:** 331–338.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Comparison of YP1 spacer-based grouping of *Y. pseudotuberculosis* complex strains (see Table S6) to MLST-based minimal spanning tree in Fig. 1 of Laukkanen-Ninios and colleagues (2011). The groups are indicated by their root spacers and from them lines are drawn to the sequence types the strains in the group represent.
Fig. S2. Comparison of YP3 spacer-based grouping of *Y. pseudotuberculosis* complex strains (see Table S6) to MLST-based minimal spanning tree in Fig. 1 of Laukkanen-Ninios and colleagues (2011). The groups are indicated by their root spacers and from them lines are drawn to the sequence types the strains in the group represent.
Table S1. *Yersinia pseudotuberculosis* genetic complex strains included in this study. The nucleotide database accession numbers for the YP1, YP2 and YP3 CRISPR loci sequences for which the sequence was available are indicated.
Table S2. Primers used in the study.
Table S3. DR variants occurring at least twice in *Y. pseudotuberculosis* complex. Variant nucleotides are indicated by grey shading.
Table S4. *Y. pseudotuberculosis* complex CRISPR spacers. The identification keys and the sequences of 1902 spacers are given. Grey-shaded spacers are unique.
Table S5. Most common CRISPR spacers among the *Y. pseudotuberculosis* genetic complex (excluding *Y. pestis*) strains.
Table S6. Spacer alignments of the YP1, YP2 and YP3 CRISPR loci of *Y. pseudotuberculosis*, *Y. similis* and *Y. wautersii* strains. The strains were sorted based on (A) YP1 alignments and on (B) YP3 alignments. Strains were clustered based on the most ancient (root) spacers. The most ancient spacers in the columns are at left side, and the most recent spacers, i.e. the leader proximal ones, at right. Gaps have been introduced between spacers to maximize alignments. The consecutive numbered spacers are shown by bold underlined numbers with the first and last spacer separated by a dash. Multilocus sequence type (ST) according to MLST Databases at (http://mlst.warwick.ac.uk/mlst/dbs/Ypseudotuberculosis, see Laukkanen-Ninios *et al.*, *Env Microbiol.* 13:3114–3127, 2011). Ns, no sequence information available. The 'nnn' in the middle of some spacer arrays indicates that we did not get the middle part of the CRISPR locus sequenced (see for example strain Pa3606, A). To facilitate comparison of spacers, the numbers are highlighted with different colours.
Table S7. Organization and sequences of the YP2 loci of *Y. pseudotuberculosis* and *Y. similis.*
Table S8. Spacers of *Y. pseudotuberculosis* complex with matching sequences (> 88% identity in BLASTN search) in plasmids (for the list of *Y. pseudotuberculosis* IP32953 cryptic plasmid-specific spacers, see the main text Table 3). The distribution of spacers 82, 283, 556, 585, 1001, 1154 and 1206, present in a number of different plasmids, is indicated in separate columns.
Table S9. Spacers of *Y. pseudotuberculosis* complex with matching sequences (> 88% identity in BLASTN search) in phage sequences.
Table S10. Spacers of *Y. pseudotuberculosis* complex with matching sequences (> 88% identity in BLASTN search) in bacterial genomes (*Y. pestis* and *Y. enterocolitica* excluded).
Table S11. Spacers of the YP1, YP2 and YP3 CRISPR loci of *Y. pseudotuberculosis* strains sharing spacers with *Y. pestis*. Gaps have been introduced between the *Y. pestis* spacers to maximize alignments. The spacers shared between *Y. pseudotuberculosis* and *Y. pestis* strains are highlighted with different colours. The consecutive numbered spacers are shown by bold underlined numbers with the first and last spacer separated by a dash. Ns, no sequence information available.
Table S12. Spacers of *Y. pseudotuberculosis*, *Y. similis* and *Y. wautersii* with hits in *Y. pestis* genomes.
Table S13. Conversion table for spacers from Cui and colleagues (2008) Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE* **3**, e2652.

National Defence University
PO Box 7, 00861 HELSINKI

Puh. +358 299 800

www.mpkk.fi

ISBN 978-951-25-2902-5 (pbk.)
ISBN 978-951-25-2903-2 (pdf)
ISSN 2342-9992 (Print)
ISSN 2343-0001 (Web)

Puolustusvoimat
The Finnish Defence Forces