Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Molecular typing of *Yersinia pestis* using MLVA-19

Douglas E. Bader & Chad W. Stratilo
DRDC Suffield

Defence R&D Canada
Technical Memorandum
DRDC Suffield TM 2013-150
December 2013

Canada

Molecular typing of *Yersinia pestis* using MLVA-19

Douglas E. Bader & Chad W. Stratilo
DRDC Suffield

Defence Research and Development Canada – Suffield

Technical Memorandum
DRDC Suffield TM 2013-150
December 2013

Principal Author

Original signed by Doug Bader

Doug Bader

Defence Scientist

Approved by

Original signed by David Pedersen

Dr David Pedersen

Head Bio Threat Defence Section

Approved for release by

Original signed by Dr. Robin Clewley

Dr. Robin Clewley

Chairman, Document Review Panel

Abstract

Yersinia pestis, a potential biothreat agent, shows highly conserved genomic sequences between strains of diverse origin. Variable number tandem repeats (VNTRs) have been used in numerous bacteria species (e.g., *Bacillus anthracis*) for strain-specific discrimination. Using several VNTR loci in combination, a robust PCR-based typing system called MLVA or multiple-locus variable-number tandem repeat analysis allows bacteria that are the same species to be differentiated from each other even if they are genetically very homogeneous. MLVA typing of *Y. pestis* using 42 VNTR loci for strain discrimination was reduced to a subset of 19 polymorphic markers (MLVA-19). The most polymorphic markers were selected because they are considered more effective for identifying genetic similarity on small geographic scales. DNA from nine *Y. pestis* strains analyzed using MLVA-19 generated six distinct genotypes demonstrating the YP-MLVA-19 typing scheme was able to discriminate *Y. pestis* isolates from a very small sample set. Decreasing the number of MLVA loci did not negatively impact the resolving power of this typing method. The methods and cluster database that were developed in this project can be used to type and characterize genetic relatedness of additional *Y. pestis* strains in the future.

Résumé

Yersinia pestis, un agent de guerre biologique potentiel, affiche des séquences génomiques largement conservées entre les souches d'origine diverse. Le nombre variable de répétitions en tandem (VNTR) a été utilisé dans de nombreuses espèces de bactéries (p. ex., Bacillus anthracis) pour la discrimination liée à une souche spécifique. Lorsqu'on utilise plusieurs locus VNTR de façon combinée, un système robuste de typage fondé sur la PCR nommé MLVA, ou analyse du nombre variable de répétitions en tandem multilocus, permet aux bactéries de la même espèce d'être différenciées les unes des autres même si elles sont génétiquement très homogènes. Le typage MLVA de Y. pestis à l'aide de 42 locus VNTR pour la discrimination des souches a été réduit à un sous-ensemble de 19 marqueurs polymorphes (MLVA-19). Les marqueurs les plus polymorphes ont été choisis parce qu'ils sont considérés plus efficaces pour l'identification de la similarité génétique à petites échelles géographiques. L'ADN de neuf souches Y. pestis analysées à l'aide du typage MLVA-19 a produit six génotypes distincts, ce qui démontre que le schéma de typage YP-MLVA-19 a pu discriminer les isolats Y. pestis à partir d'un ensemble d'échantillonnage très petit. La réduction du nombre de locus MLVA n'a pas eu d'impact négatif sur le pouvoir de résolution de cette méthode de typage. Les méthodes et la base de données en grappe qui ont été conçues dans le cadre de ce projet pourront être utilisées plus tard pour effectuer le typage et la caractérisation des liens génétiques d'autres souches de Y. pestis.

Executive Summary

Molecular Typing of *Yersinia pestis* using MLVA-19

Douglas E. Bader and Chad W. Stratilo; DRDC Suffield TM 2013-150; Defence R&D Canada – Suffield; December 2013.

Introduction or background: There is a need to characterize biothreat agents for biodefence and biosecurity purposes. Genetic typing techniques can be used to characterize biothreats and have the potential to determine the possible foci or origin of the agent. *Yersinia pestis,* a potential biothreat agent, shows highly conserved genomic sequences between strains of diverse origin. Variable number tandem repeats (VNTRs) have been used in numerous bacteria species (e.g., *Bacillus anthracis*) for strain-specific discrimination. Using several VNTR loci in combination, a robust PCR-based marker typing system called MLVA or multiple-locus variable-number tandem repeat analysis allows bacteria that are the same species to be differentiated from each other even if they are genetically very homogeneous. MLVA typing of *Y. pestis* has been successfully carried out by others using 42 VNTR loci for strain discrimination. A CRTI project (CRTI 02-0069RD) was undertaken to establish molecular typing capabilities of *Y. pestis* strains in Canadian federal laboratories. DNA from nine *Y. pestis* strains was analyzed using MLVA at 19 loci (MLVA-19).

Results: Nine isolates generated six distinct genotypes demonstrating the YP-MLVA-19 typing scheme was able to discriminate *Y. pestis* isolates from a very small sample set.

Significance: This project has allowed DRDC Suffield to establish the capability to perform high resolution subtyping of *Y. pestis*. Strains were typed using single-plex PCR reactions that were pooled or in multiplex reactions greatly reducing reagent use while increasing throughput. Isolates originally screened against 42 possible VNTR loci were reduced to a subset of 19 polymorphic markers to allow a more manageable set of markers for characterization. Decreasing the number of MLVA loci did not negatively impact the resolving power of this typing method. The methods and cluster database that were developed in this project can be used to type and characterize genetic relatedness of additional *Y. pestis* strains in the future.

Future plans: MLVA databases will be expanded as additional strains are acquired in order to increase the resolving power of this technique.

Sommaire

Typage moléculaire de Yersinia pestis à l'aide de MLVA-19

Douglas E. Bader et Chad W. Stratilo; RDDC Suffield TM 2013-150; Recherche et développement pour la défense Canada - Suffield; décembre 2013.

Introduction ou contexte : Il est nécessaire d'effectuer la caractérisation des agents de guerre biologique aux fins de défense biochimique et de biosécurité. Les techniques de typage génétique peuvent être utilisées pour caractériser les menaces biologiques et elles peuvent servir à déterminer le foyer ou l'origine possible de l'agent. Yersinia pestis, un agent de guerre biologique potentiel, affiche des séquences génomiques largement conservées entre les souches d'origine diverse. Le nombre variable de répétitions en tandem (VNTR) a été utilisé dans de nombreuses espèces de bactéries (p. ex., Bacillus anthracis) pour la discrimination liée à une souche spécifique. Lorsqu'on utilise plusieurs locus VNTR de façon combinée, un système robuste de typage fondé sur la PCR nommé MLVA, ou analyse du nombre variable de répétitions en tandem multilocus, permet aux bactéries de la même espèce d'être différenciées les unes des autres même si elles sont génétiquement très homogènes. Certains ont réussi à effectuer le typage MLVA de Y. pestis à l'aide de 42 locus VNTR pour la discrimination de souches. Un projet de l'IRTC (CRTI 02-0069RD) a été lancé pour établir les capacités de typage moléculaire des souches de Y. pestis dans des laboratoires fédéraux du Canada. L'ADN de neuf souches de Y. pestis a été analysée à l'aide du typage MLVA à 19 locus (MLVA-19).

Résultats : Neuf isolats ont produit six génotypes distincts, ce qui démontre que le schéma de typage YP-MLVA-19 a pu discriminer les isolats Y. pestis à partir d'un ensemble d'échantillonnage très petit.

Importance : Ce projet a permis à RDDC Suffield d'établir la capacité d'exécuter le sous-typage de haute résolution de Y. pestis. Le typage des souches a été effectué à l'aide de réactions PCR en monoplex qui ont été regroupées ou à l'aide de réactions en multiplex, ce qui a grandement réduit l'utilisation de réactifs tout en augmentant le débit d'alimentation. Les isolats qui ont d'abord été testés à l'aide de 42 locus VNTR potentiels ont été réduits à un sous-ensemble de 19 marqueurs polymorphes pour obtenir un ensemble de marqueurs mieux gérable aux fins de caractérisation. La réduction du nombre de locus du typage MLVA n'a pas eu d'impact négatif sur la puissance de résolution de cette méthode de typage. Les méthodes et la base de données en grappe qui ont été conçues dans le cadre de ce projet pourront être utilisées plus tard pour effectuer le typage et la caractérisation des liens génétiques d'autres souches de Y. pestis.

Plans futurs : Les bases de données MLVA seront augmentées à mesure que d'autres souches seront acquises, et ce, afin d'accroître la puissance de résolution de cette technique.

Table of Contents

List of Tables

Acknowledgements

The authors wish to thank Kelvin Leung for his technical assistance. This project would not have been possible without the support and collaboration of M. Mulvey, L. Brydon, and D. Johnstone at Public Health Agency of Canada National Microbiology Laboratory (PHAC-NML). This project was supported by CRTI 02-0069RD.

Introduction

Yersinia pestis, a human bacterial pathogen, is the causative agent of bubonic, pneumonic and septicaemic plague [1]. It is primarily a zoonotic disease and is endemic throughout the world [2]. *Yersinia pestis* is a Gram-negative bacterium that harbors three virulence plasmids—pFra that encodes the anti-phagocytic capsular protein fraction 1 (F1) and the murine toxin that enables bacteria to survive in the flea gut; pYV that encodes V antigen and *Yersinia* outer proteins (Yops), which disrupt phagocytosis and reduce inflammation; and pPla that makes a plasminogen activator that allows bacteria to spread in tissues by dissolving fibrin clots [3]. *Y. pestis* can be divided into three biovariants or biovars (*antiqua*, *medievalis*, and *orientalis*) based on their ability to reduce nitrate and utilize glycerol [4].

Antiqua (nitrate+ glycerol+), *medievalis* (nitrate- glycerol +) and *orientalis* (nitrate+ glycerol-) are thought to have caused the three major plague pandemics. *Antiqua* is associated with the Justinian plague which started in Africa and spread to the Mediterranean (AD 541 to 767); *medievalis* is associated with the Black Death which may have originated in central Asia and spread from the Caspian Sea to Europe (1346 to early 19th century); and *orientalis* is associated with modern plague (since 1894) which began in southwest China and spread globally via shipping routes [2, 4]. DNA sequences from ancient human remains dispute the assertion that different biovars were responsible for the three different pandemics and suggests that *orientalis*-like *Y. pestis* may have been involved in all three, a suggestion that remains highly controversial [2].

A fourth biovar, *microtus* has been proposed based on the unique pathogenic, biochemical and molecular features of *microtus* strains [5]. *Microtus* strains are lethal to *Microtus* (rodents of the genus *Microtus* belonging to the order *Rodentia* and family *Muridae*), as well as mice and other rodents, but are avirulent to larger mammals such as guinea pigs, rabbits, rhesus monkeys, sheep, and humans. There is currently no evidence that human plague can arise from *microtus* strains [5].

Y. pestis is a clone that evolved from *Y. pseudotuberculosis* between 1500-20,000 years ago [4]. In addition to containing virulence plasmids, *Y. pestis* has 21 genomic islands, eighteen of which are found in *Y. pseudotuberculosis*. Three genomic islands are found only in *Y. pestis* strains [4].

Yersinia pestis shows highly conserved genomic sequences between strains of diverse origin. Numerous typing methods have been described including repetitive element PCR [6], pulsed-field gel electrophoresis (PFGE) [7], insertion sequence-element probe restriction fragment length polymorphism analysis [8] and whole genome microarray [9] that show concordant results. Variable number tandem repeats (VNTRs) have been used in numerous bacteria species (e.g., *Bacillus anthracis)* for strain-specific discrimination [10]. Using several VNTR loci in combination, a robust PCR-based marker typing system called MLVA or multiple-locus variable-number tandem repeat analysis allows bacteria that are the same species to be differentiated from each other even if they are genetically very homogeneous.

MLVA typing of *Y. pestis* has been successfully carried out using 42 VNTR loci for strain discrimination [11]. These loci are amplified by PCR using fluorescently labeled primers. The PCR products can be analyzed by fragment size analysis or they can be sequenced directly.

This report describes MLVA typing of *Yersina pestis* from the DRDC Suffield collection and DNA samples provided by NML using 19 VNTR loci (MLVA-19). Strains were typed using single-plex PCR reactions that were pooled, or in multiplex reactions which greatly reducing reagent use while increasing throughput. The amplicons were able to be distinguished from one another using different fluorescent dyes or expected amplicon sizes, prior to capillary electrophoresis.

Isolates originally screened against 42 possible VNTR loci were reduced to a subset of 19 polymorphic markers to allow a more manageable set of markers for characterization. In *Y. pestis*, those markers with the highest number of tandem repeats showing the highest degree of polymorphism across the isolates were tested. The most polymorphic markers were selected because they are considered more effective for forensic analysis and for identifying genetic similarity on small geographic scales. Nine isolates from the collection generated six distinct genotypes demonstrating the YP MLVA-19 typing scheme was able to discriminate *Y. pestis* isolates from a very small sample set.

Materials and Methods

The strains typed in this study are part of the DRDC Suffield strain collection and Public Health Agency of Canada – National Microbiology Laboratory (PHAC-NML) strain collection (Table 1). DRDC Suffield strain DNA was isolated in the BSL3 bacterial suite using MasterPure™ DNA & RNA Purification Kit (Epicentre Biotechnologies, Madison, WI, USA.), Phase lock gel® (Eppendorf, Westbury, NY, USA) or GNOME® DNA Isolation Kit (QBiogene, Irvine, CA, USA.). Concentrations of approximately 60 ng/µL were obtained from each strain tested. Sterile DNase-, RNase-free water (NFW) was used as a negative control. PHAC-NML strain DNA was provided to DRDC Suffield by PHAC-NML.

PCR amplification of the 19 variable loci was carried out as follows: 2 mM $MgCl_2$, 1× PCR buffer, 2 mM dNTPs, 0.5 U Taq polymerase (Life Technologies, Inc., Rockville, Md.), 1.0 µL of template DNA, 0.5 µM forward primers, 0.5 µM reverse primer and filtered sterile water to a volume of 25 µL. PCRs were either carried out with individual primer pairs or as multiplex PCRs (8 primer pairs per reaction). The primers that were used, as well as the 5' phosphoramidite fluorescent dyes used with each primer, are listed in Table 2. The nature of each VNTR is listed in Table 3.

The PCR reaction mixtures were incubated at 94 °C for 5 min and then cycled at 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s for 35 cycles with a final incubation of 72 °C for 5 min using an ABI 9700 thermocycler. PCR reagents were obtained from Life Technologies, Inc. or Applied Biosystems Inc. PCR products (2 µL) were pooled as described in Table 3 unless performed as part of a multiplex in which pooling was not necessary. The reactions were diluted with NFW to a final volume of 80 µL. One µL of the diluted pooled PCR products was combined with 8 µL of HiDi Formamide and 1 µL of the combined size standard from BioVentures Inc., Murfreesboro, TN. (0.5 µL Rhodamine-X MapMarker (70 to 400 bp) and 0.5 µL CST ROX (420-800 bp).

The PCR amplification products were sized using the Applied Biosystems 3100 Genetic Analyzer with a 36 cm capillary using filter set D for fragment analysis. Allele names were established by direct sequencing of PCR amplicons by PHAC-NML and DRDC Suffield. The bin sets for *Y. pestis* alleles are found in Table 4.

Phylogenetic analysis was carried out using Unweighted Pair Group Method with Arithmetic Mean (UPGMA).

Table 1. Y. pestis Strains Tested

Strain Name	Source
1. PP65BC	PHAC-NML
2. PP2868	PHAC-NML
3. PX14-3	PHAC-NML
4. PP1967	PHAC-NML
5. PP2321 (1966)	PHAC-NML
6. PP2321 (1970)	PHAC-NML
7. GB	DRDC Suffield
8. CO92	DRDC Suffield
9. C12	DRDC Suffield

Table 2. Primers Used for Amplification of VNTR Loci

Locus Name	Primer name and 5' modification	Primer Sequence 5'-3'
Yp-M06	Yp-M06-f1-6-FAM Yp-M06-r1	GATAGATCTCCGAAGGCAGATCGCAATAGGTC GGGCGATAGGATAGCTTGATGCGTTGTTTTAC
Yp-M09	Yp-M09-f1-6-FAM Yp-M09-r1	GACCTCGATCTGCTTAGAACCTTTGTAGCTGTTGC GTTGCATTTGTTGGCTAACTGCTGACTGAGTTC
Yp-M12	Yp-M12-f1-6-FAM Yp-M12-r1	GAAGCGGCAACAATTTACCGTTATTTATGCT TTTATTCGCCTCCCCTTCGAACTTGAA
Yp-M18	Yp-M18-f1-HEX Yp-M18-r1	GGGGTGTTAATTGTGAGGCGTGTTGTC CCCTACCCGCCACTCTCTTGGTAGC
Yp-M19	Yp-M19-f1-NED Yp-M19-r1	TTTCGTTCAATTCAGTCACCGCTTTCTC GTCCTGCCCGTTTTCCTTGCCTTCTC
Yp-M21	Yp-M21-f1-6-FAM Yp-M21-r1	GATTTATGAATGGCTACAACGTCGTCGCA GTAGTGATACAGGCAAATCCAAGAGCGCA
Yp-M22	Yp-M22-f1-NED Yp-M22-r1	GCGTGATACCAAAGGCTGGCTCACC GGCACTTTGGGTACGGAACGTCATCAC
Yp-M23	Yp-M23-f1-HEX Yp-M23-r1	GTTAAAACTTAATTAACCAACTTAAGAGTCGCCATATC GTTATCAGATTTCGCTTGAAGTAGGTTTAACGATGAC
Yp-M25	Yp-M25-f1-6-FAM Yp-M25-r1	GTTTAGCTGTAAATAGATTTAGAAGCCTCGTCTTTTGAC GATATAAATGAGTTGATTCAGGTGTTCATATTTAACGAAAC
Yp-M27	Yp-M27-f1-HEX Yp-M27-r1	GTCTAACTGGCGCGGCATTCTTGC GGGTGTTCTTATGTCATCCGCCAACAAAC
Yp-M28	Yp-M28-f1-6FAM Yp-M28-r1	GTTTGGCGGTTGGGCGTACCTTGGTA AGCGCCCGTAGACGCTTTCGAAATAGC
Yp-M29	Yp-M29-f1-HEX Yp-M29-r1	GAGCGGCGGGTTCTCATGCTGAT GTTTAAGCAGTAGATCTAAAGCGTTATGAATATTGGTGTTA
Yp-M31	Yp-M31-f1-HEX Yp-M31-r1	GGTTTGCAGGTTTTTGTTGTGGATTATGGACTTAGAT GGCGGGATGGCGTATCGGTTGC
Yp-M33	Yp-M33-f1-HEX Yp-M33-r1	AGCAACCTGTGCCGCCTCGATATAAG GAGACGGGCGAAATTGAAGCACAGTTAT
Yp-M34	Yp-M34-f1-6FAM Yp-M34-r1	GAATCGCGGGTTGACGCTGTTGAGC GCTGAACAGCCCCATAAAACCGGAGC
Yp-M37	Yp-M37-f1-6FAM Yp-M37-r1	GCCACAGGAAGAGGACATTTCAGAGAAAAC GTTGCTAAAACGATACCGCTACGATCAGC
Yp-M58	Yp-M58-f1-6FAM Yp-M58-r1	GCGATAACCCACATTATCACAATAACCAACAC GCTGATGGAACCGGTATGCTGAATTTGC
Yp-M59	Yp-M59-f1-6FAM Yp-M59-r1	GCTTAGCCGCCAGAAAAGGTGAGTTGGC GATAATGGCGGTAGCCGGAATCTGATAATCATC
Yp-M79	Yp-M79-f1-NED Yp-M79-r1	GCCCTTATCTACTGGGCCAAGCTAACGC GCCATGGCGGGATGTAATGGCAC

Table 3. VNTR and PCR Product Characterization

VNTR Marker	Repeat Motif	Repeat Size	Pool	Range (bp)	Repeat copy #	Allele no.
Yp-M06	AT	2	3	146-148	5-6	2
Yp-M09	TGC	3	1	109-112	6-7	4
Yp-M12	CAAA	4	2	212-232	8-12	7
Yp-M18	TCATCC	6	1	181-127	4-13	9
Yp-M19	CTTTTA or CCTTTA	6	2	311-551	29-62	9
Yp-M21	ACACCCA	7	1	172-179	3-4	3
Yp-M22	TTCATTGTTCATTG or CTCATTGCTCATTG or CTCATTGTTCATTC	14	1	214-361	4-15	9
Yp-M23	AGCAAAT	7	2	162-232	5-19	9
Yp-M25	CAACCTA or CCACCTA or ACCACCT or GTCACCT	7	2	251-386	7-23	10
Yp-M27	TTATTGAC	8	3	173-325	5-18	11
Yp-M28	GTTAATTG	8	1	188-220	5-8	7
Yp-M29	CTTTCTAT	8	1	191-263	3-12	9
Yp-M31	TGATTAGT	8	2	310-390	6-16	7
Yp-M33	TGTGGCTTG or TGCGGTTTG or TGTGGTTGG or TGGCTTGTG	9	1	339-375	6-20	9
Yp-M34	GTTGAAGAA	9	3	207-351	4-20	11
Yp-M37	TGGTGATAGT or TGGTGATAGT or CGGTGATAGT	10	1	309-339	2-7	5
Yp-M58	TTGTTTATCTATCTATG or TTGTTTAGCTATCTATG	17	1	344-412	3-7	6
Yp-M59	CCCATATCAGTAATTAG or TCCATATCAGTAACTAG	17	1	279-296	6-7	4
Yp-M79	TGCTCAAC	8	3	220-324	6-19	10

Table 4. Y. pestis Bin Set Panels.

Marker Name	Bin Parameters	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 7	Allele 8	Allele 9	Allele 10
Yp-M06	Sequence (bp)	146	148								
	Location	141.38	143								
	Offset(left/right)	1/1	0.5/0.94								
Yp-M09	Sequence (bp)	109									
	Location	105.13									
	Offset(left/right)	0.5/0.5									
Yp-M12	Sequence (bp)	212	216	220	224	228	232				
	Location	207.43	211	215	218	222.19	226.28				
	Offset(left/right)	1/1	0.5/1.05	0.5/0.5	0.5/1.74	0.5/1.55	0.5/0.5				
Yp-M18	Sequence (bp)	127	139	145	151	169	181				
	Location	125.71	137	143.37	149.81	168.01	179				
	Offset(left/right)	0.5/0.5	0.5/1.4	0.5/0.5	1/1	1/1	0.5/1.34				
Yp-M19	Sequence (bp)	210	311	317	347	353	365	371	389	497	
	Location	211	311	317.42	348	354.02	366.59	372.82	390.86	503	
	Offset(left/right)	0.5/0.5	0.5/0.64	1/1	0.5/0.86	1/1	1/1	1.43/1	1/1	0.5/0.5	
Yp-M21	Sequence (bp)	172	179								
	Location	168	175.38								
	Offset(left/right)	0.5/0.82	0.5/0.5								
Yp-M22	Sequence (bp)	214	228	235	242	256	249	361	270	276	284
	Location	212.5	227	234	241.12	255	247.95	361	268.89	276.5	283.53
	Offset(left/right)	1/1	0.5/0.5	0.5/0.5	1/1	0.5/0.5	1/1	0.5/1.42	1/1	0.5/0.5	1/1
Yp-M23	Sequence (bp)	162	169	176	182	197	204	211	232	218	267
	Location	157	164	171.59	179.16	193	200			214.82	263.07
	Offset(left/right)	0.5/1.13	0.5/1.32	1/1	0.5/0.5	0.5/0.5	0.5/1.1			0.5/0.5	1/1
Yp-M25	Sequence (bp)	251	258	265	349	356	363	377	386		
	Location	249.2	255.83	263	348	354.61					
	Offset(left/right)	1/1	1/1	0.5/0.73	2.05/0.5	1/1					
Yp-M27	Sequence (bp)	173	181	205	213	221	237	261	277	285	325
	Location	170	178.98	203.0	212	219	236.28	260.54	276	285.03	326.75
	Offset(left/right)	0.5/0.5	1/1	1/1.51	0.5/0.5	0.5/1.28	1/1	1/1	0.5/1.3	0.5/0.5	1/1
Yp-M28	Sequence (bp)	188	196	204	212	220					
	Location	187.93	195	203.83	213.1	220.59					
	Offset(left/right)	1/1	0.5/0.85	0.5/0.5	1.52/0.5	1/1					
Yp-M29	Sequence (bp)	191	199	207	223	231	239	243	247	255	263
	Location	188.29	196	204	220.86	229	236	240.92	245.21	253.14	261.08
	Offset(left/right)	1/1	0.5/0.81	0.5/0.05	0.5/0.5	0.5/0.5	0.5/1.61	0.5/0.5	1/1	1/1	0.5/0.5
Yp-M31	Sequence (bp)	310	315	326	334	358	390				
	Location	307.74	315.73	323.5	332.59	357.42	390.32				
	Offset(left/right)	1/1	1.11/0.81	0.5/1.23	1/1	1/1	1/1				
Yp-M33	Sequence (bp)	339	348	357	375						
	Location	343.52	352	362.83	380.61						
	Offset(left/right)	1/1	0.5/1.69	1.72/0.5	1/1						
Yp-M34	Sequence (bp)	207	225	234	252	261	270	279	315	324	351
	Location	204.85	222.87	231	249	258.65	267.84	276.6	312	321.55	348
	Offset(left/right)	1/1	1/1	0.5/1.0	0.5/1.04	1/1	0.5/1.12	0.84/1.12	0.5/0.89	1/1	0.5/1.47
Yp-M37	Sequence (bp)	309	319	329	339						
	Location	309	319.52	329	340.36						
	Offset(left/right)	0.5/0.91	0.5/0.74	0.5/1.13	1/1						
Yp-M58	Sequence (bp)	344	361	378	395	412					
	Location	343.98	360.65	377	396	412.61					
	Offset(left/right)	0.95/0.5	1/1	0.5/1.62	0.81/0.14						
Yp-M59	Sequence (bp)	279	296								
	Location	276.53	293.63								
	Offset(left/right)	0.99/0.5	1.08/0.5								
Yp-M79	Sequence (bp)	220	236	244	252	260	284	324			
	Location	219.43	235	244	251.46	260	285.5	323.65			
	Offset(left/right)	1/1	0.5/1.3	0.7/0.5	1/1	1.25/0.5	1/1	1/1			

This page left intentionally blank

Results and Discussion

The nature and extent of variability of the 19 VNTR loci are described in Table 3. Direct sequencing of the 19 loci allowed for sizing of the alleles present at those loci. Allele names were based on amplicon sizes established by direct sequencing (Table 4). The size of the amplicons during fragment analysis are also listed (Table 4). The differences between amplicon length established by fluorescent primer fragment length analysis and direct sequencing can be determined by comparison between the allele name established by direct sequencing, and the location of the allele established by fragment length analysis (Table 4). The difference between length established by sequencing and length determined by fragment analysis is <+/- 4 bp although there may be no difference at a particular locus or allele. This discrepancy may be attributable to the type of phosphoramidite fluorescent dye used in 5' labelling for fragment analysis. Thus the name of the allele for a given locus is based on the size of the VNTR amplicon as determined by sequence analysis as opposed to fragment size analysis.

The results of the MLVA-19 analysis are presented in Table 5. Four of the 19 loci (M06, M09, M21, and M33) were unable to distinguish the nine strains while five of the 19 loci (M22, M23, M27, M34, M79) generated the highest number of alleles among the nine strains (4 alleles each).

There were six discrete MLVA-19 genotypes for the nine strains tested. Three pairs of strains had identical MLVA-19 genotypes (PP65BC/PP1967, PP2321-1966/1970, and CO92/C12). Three strains had unique MLVA-19 genotypes (PP2868, PX14-3 and GB) with the latter two being more closely related. All three of the *Yersinia pestis* strains in the DRDC Suffield collection belong to the biovar *orientalis* [12, 13]. *Y. pestis* CO92 was originally isolated from the sputum of a human patient who was apparently exposed to a plague-infected cat in Colorado and who later died of primary pneumonic plague [14]. *Y. pestis* C12 is an $F1^-$ isogenic strain derived by site-directed mutagenesis of the F1 structural gene from the CO92 strain [12]. Based on their history, it is not surprising that CO92 and C12 had the same MLVA-19 genotype. *Y. pestis* GB was isolated from a fatal human case [15] but it had a different MLVA-19 genotype from CO92 and C12, having differences at 10 of 19 loci.

The MLVA typing method presented in this report is a robust and well accepted approach to characterizing *Y. pestis* isolates. The use of fragment analysis allows for quick and reproducible results. Analysis of strains between laboratories world wide is convenient and easily converted into simple numerical data unlike amplified fragment length polymorphisms (AFLP) or pulsed-field gel electrophoresis. Decreasing the number of MLVA loci used from 42 to 19 increased the throughput but did not negatively impact the resolving power of this typing method.

Discrepancies between the allele calls used in this paper and those found in other publications are due to designations adopted based on fragment sizing analysis as opposed to sequence data. Although the choice of allele name also takes into account the repeat size, it may be beneficial in the future to move away from allele names that are based on amplicon or sequence size but rather adopt a scheme that uses a letter or number designation based on the repeat structure and length.

Y. pestis is a category A biological threat agent and could be used against the CAF or as a bioterrorist threat against Canadians. This project has allowed DRDC Suffield to establish the capability to perform high resolution subtyping of *Y. pestis*. The methods and cluster database

that were developed in this project can be used to type and characterize genetic relatedness of additional *Y. pestis* strains in the future.

Table 5. Results of MLVA-19 Analysis

Y. pestis Strains	M06	M09	M12	M18	M19	M21	M22	M23	M25	M27	M28	M29	M31	M33	M34	M37	M58	M59	M79
PP65BC	148	109	224	139	347	179	235	204	349	221	212	215	326	357	315	319	378	296	244
PP1967	148	109	224	139	347	179	235	204	349	221	212	215	326	357	315	319	378	296	244
PP2321 (1966)	148	109	224	139	347	179	235	204	349	221	212	231	326	357	315	319	378	296	244
PP2321 (1970)	148	109	224	139	347	179	235	204	349	221	212	231	326	357	315	319	378	296	244
CO92	148	109	224	139	353	179	235	169	335	245	212	215	326	357	225	329	395	296	252
C12	148	109	224	139	353	179	235	169	335	245	212	215	326	357	225	329	395	296	252
PP2868	148	109	224	139	347	179	249	182	370	277	204	207	315	357	279	319	395	296	220
PX14-3	148	109	232	145	497	179	276	218	370	285	212	215	326	357	270	319	395	279	236
GB	148	109	236	145	497	179	269	218	335	285	212	215	334	357	270	319	395	279	236
Alleles	**1**	**1**	**3**	**2**	**3**	**1**	**4**	**4**	**3**	**4**	**2**	**3**	**3**	**1**	**4**	**2**	**2**	**2**	**4**

Allele sizes (measured in base pairs) were determined by sequence analysis and used as the allele name for each locus (M06 through M79). The predominant allele among the nine test strains is highlighted in yellow for each locus. Alleles that differ from the predominant allele are highlighted in a different color for visual differentiation. The total number of alleles for a given locus among the nine strains tested is provided at the bottom of the table (black, boldface type). Strains having the same MLVA-19 type have matching font colors.

This page left intentionally blank

References

[1] Chauvaux, S., Rosso, M-L., Frangeul, L., Lacroix, C., Labarre, L., Schiavo, A., Marceau, M., Dillies, M-A., Foulon, J., Coppee, J-Y., Medigue, C., Simonet., M., and Carniel, E. (2007), Transcriptome Analysis of *Yersinia pestis* in Human Plasma: An Approach for Discovering Bacterial Genes Involved in Septicaemic Plague, *Microbiology*, 153, 3112-3123.

[2] Chain, P.S.G., Hu, P., Malfatti, S.A., Radnedge, L., Larimer, F., Vergez, L.M., Worsham, P., Chu, M.C., Andersen, G.L. (2006), Complete Genome Sequence of *Yersinia pestis* Strains Antiqua and Nepal516: Evidence of Gene Reduction in an Emerging Pathogen, *J. Bacteriology*, 188, 4453-4463.

[3] Carniel E. (2003). Evolution of Pathogenic *Yersinia*, Some Lights in the Dark. *Adv. Exp. Med. Biol*., 529: 3–12.

[4] Zhou, D., Han, Y., Song, Y., Tong, Z., Wang, J., Guo, Z., Pei, D., Pang, X., Zhai, J., Cui, B., Qi, Z., Jin, L., Dai, R., Du, Z., Bao, J., Zhang, X., Yu, J., Wang, J., Huang, P., and Yang, R. (2004), DNA Microarray Analysis of Genome Dynamics in *Yersinia pestis*: Insights into Bacterial Genome Microevolution and Niche Adaptation, *J. Bacteriology*, 186, 5138-5146.

[5] Zhou, D., Tong, Z., Song, Y., Han, Y., Pei, D., Pang, X., Zhai, J., Li., M., Cui, B., Qi, Z., Jin, L., Dai, R., Du, Z., Wang, J., Guo, Z., Wang, J., Huang, P., and Yang, R. (2004), Genetics of Metabolic Variations between *Yersinia pestis* Biovars and the Proposal of a New Biovar, *microtus*, *J. Bacteriology*, 186, 5147-5152.

[6] Kim, W., Song, M.O., Song, W., Kim, K.J., Chung, S.I., Choi, C.S., and Park, Y.H. (2003). Comparison of 16S rDNA analysis and rep-PCR genomic fingerprinting for molecular identification of *Yersinia pseudotuberculosis*, *Antonie Van Leeuwenhoek,* 83(2), 125-133.

[7] Huang, X-Z., Chu, M.C., Engelthaler, D.M., Lindler, L.E., (2002). Genotyping of a Homogeneous Group of *Yersinia pestis* Strains Isolated in the United States, *J. Clin. Micro.*, 40 (4), 1164–1173.

[8] Torrea, G., Chenal-Francisque, V., Leclercq, A., and Carniel, E. (2006). Efficient Tracing of Global Isolates of *Yersinia pestis* by Restriction Fragment Length Polymorphism Analysis Using Three Insertion Sequences as Probes, *J. Clin. Micro*., 44 (6), 2084–2092.

[9] Zhou, D., Han, Y., Dai, E., Song, Y., Pei, D., Zhai, J., Du, Z., Wang, J., Guo, Z., Yang. R. (2004). Defining the Genome Content of Live Plague Vaccines by Use of Whole-Genome DNA Microarray, *Vaccine*, 22 (25-26), 3367-3374.

[10] Kim, W., Hong,Y.P., Yoo, J.H., Lee, W.B., Choi, C.S., Chung, S.I. (2002). Genetic Relationships of *Bacillus anthracis* and Closely Related Species Based on Variable-Number Tandem Repeat Analysis and BOX-PCR Genomic Fingerprinting, *FEMS Microbiol Lett.* 207(1), 21-27.

[11] Klevytska, A.M., Price, L.B., Schupp, J.M., Worsham, P.L., Wong, J., and Keim, P. (2001). Identification and Characterization of Variable-Number Tandem Repeats in the *Yersinia pestis* Genome, *J. Clin. Micro*., 39 (9), 3179–3185.

[12] Anderson Jr., G.W., Leary, S.E.C., Williamson, E.D., Titball, R.W., Welkos, S.L., Worsham, P.L. and Friedlander, A.M. (1996), Recombinant V Antigen Protects Mice Against Pneumonic and Bubonic Plague Caused by F1-Capsule-Positive and -Negative Strains of *Yersinia pestis*, *Infect. Immun*., 64, 4580-4585.

[13] Oyston, P.C.F., Dorrell, N., Williams, K., Li, S-L., Green, M., Titball, R.W., and Wren, B.W. (2000), The Response Regulator PhoP Is Important for Survival under Conditions of Macrophage-Induced Stress and Virulence in *Yersinia pestis*, *Infect. Immun*., 68, 3419-3425.

[14] Doll, J.M., Zeitz, P.S., Ettestad, P., Bucholtz, A.L., Davis, T. and Gage, K. (1994), Cat-Transmitted Fatal Pneumonic Plague in a Person Who Travelled from Colorado to Arizona, *Am. J. Trop. Med. Hyg.*, 51, 109-114.

[15] Russell, P. Eley, S.M., Hibbs, S.E., Manchee, R.J., Stagg, A.J. and Titball, R.W. (1995), A Comparison of Plague Vaccine, USP and EV76 Vaccine Induced Protection Against *Yersinia pestis* in a Murine Model, *Vaccine*, 13, 1551-1556.

List of Symbols/Abbreviations/Acronyms/Initialisms

AFLP	Amplified fragment length polymorphism
bp	Base pairs
BSL3	BioSafety Level 3
CAF	Canadian Armed Forces
CRTI	CBRN Research and Technology Initiative
DNA	Deoxyribonucleic acid
dNTPs	Dideoxynucletotide triphosphates
DRDC	Defence Research & Development Canada
MLVA	Multi-locus variable-number tandem repeat analysis
NFW	Nuclease-free water
PCR	Polymerase Chain Reaction
PFGE	Pulsed-field Gel Electrophoresis
PHAC-NML	Public Health Agency of Canada - National Microbiology Laboratory
R&D	Research & Development
RNA	Ribonucleic acid
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
VNTR	Variable Number Tandem Repeat

This page intentionally left blank

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Designated)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Defence Research and Development Canada – Suffield
P.O. Box 4000, Station Main
Medicine Hat, Alberta T1A 8K6

2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.)

UNCLASSIFIED

2b. CONTROLLED GOODS

(NON-CONTROLLED GOODS)
DMC A
REVIEW: GCEC December 2013

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Molecular typing of *Yersinia pestis* using MLVA-19

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Bader, D.E.; Stratilo, C.W.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
December 2013	26	15

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence Research and Development Canada – Suffield
P.O. Box 4000, Station Main
Medicine Hat, Alberta T1A 8K6

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC Suffield TM 2013-150	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Yersinia pestis, a potential biothreat agent, shows highly conserved genomic sequences between strains of diverse origin. Variable number tandem repeats (VNTRs) have been used in numerous bacteria species (e.g., *Bacillus anthracis*) for strain-specific discrimination. Using several VNTR loci in combination, a robust PCR-based typing system called MLVA or multiple-locus variable-number tandem repeat analysis allows bacteria that are the same species to be differentiated from each other even if they are genetically very homogeneous. MLVA typing of *Y. pestis* using 42 VNTR loci for strain discrimination was reduced to a subset of 19 polymorphic markers (MLVA-19). The most polymorphic markers were selected because they are considered more effective for identifying genetic similarity on small geographic scales. DNA from nine *Y. pestis* strains analyzed using MLVA-19 generated six distinct genotypes demonstrating the YP-MLVA-19 typing scheme was able to discriminate *Y. pestis* isolates from a very small sample set. Decreasing the number of MLVA loci did not negatively impact the resolving power of this typing method. The methods and cluster database that were developed in this project can be used to type and characterize genetic relatedness of additional *Y. pestis* strains in the future

Yersinia pestis, un agent de guerre biologique potentiel, affiche des séquences génomiques largement conservées entre les souches d'origine diverse. Le nombre variable de répétitions en tandem (VNTR) a été utilisé dans de nombreuses espèces de bactéries (p. ex., Bacillus anthracis) pour la discrimination liée à une souche spécifique. Lorsqu'on utilise plusieurs locus VNTR de façon combinée, un système robuste de typage fondé sur la PCR nommé MLVA, ou analyse du nombre variable de répétitions en tandem multilocus, permet aux bactéries de la même espèce d'être différenciées les unes des autres même si elles sont génétiquement très homogènes. Le typage MLVA de Y. pestis à l'aide de 42 locus VNTR pour la discrimination des souches a été réduit à un sous-ensemble de 19 marqueurs polymorphes (MLVA-19). Les marqueurs les plus polymorphes ont été choisis parce qu'ils sont considérés plus efficaces pour l'identification de la similarité génétique à petites échelles géographiques. L'ADN de neuf souches Y. pestis analysées à l'aide du typage MLVA-19 a produit six génotypes distincts, ce qui démontre que le schéma de typage YP-MLVA-19 a pu discriminer les isolats Y. pestis à partir d'un ensemble d'échantillonnage très petit. La réduction du nombre de locus MLVA n'a pas eu d'impact négatif sur le pouvoir de résolution de cette méthode de typage. Les méthodes et la base de données en grappe qui ont été conçues dans le cadre de ce projet pourront être utilisées plus tard pour effectuer le typage et la caractérisation des liens génétiques d'autres souches de Y. pestis.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS

Yersinia pestis; molecular typing; VNTR; variable number of tandem repeats;
MLVA, multi-locus VNTR analysis