International Journal of Medical Microbiology 304 (2014) 452–463

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

The *Yersinia pseudotuberculosis* complex: Characterization and delineation of a new species, *Yersinia wautersii*

Cyril Savin[a,*], Liliane Martin[a], Christiane Bouchier[b], Sofia Filali[a], Jérôme Chenau[a,d], Zhemin Zhou[c], François Becher[d], Hiroshi Fukushima[e], Nicholas R. Thomson[f], Holger C. Scholz[g], Elisabeth Carniel[a]

[a] Yersinia Research Unit and National Reference Laboratory, Institut Pasteur, Paris, France
[b] Genomic Platform, Institut Pasteur, Paris, France
[c] Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK
[d] Service de Pharmacologie et d'Immunoanalyse, iBiTec-S, CEA, Gif-sur-Yvette, France
[e] Shimane Prefectural Institute of Public Health and Environmental Science, Matsue, Japan
[f] Pathogen Genomics, Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK
[g] Bundeswehr Institute of Microbiology, Munich, Germany

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form 24 January 2014
Accepted 2 February 2014

Keywords:
Y. pseudotuberculosis
Y. wautersii
Korean Group
Y. similis
Species

ABSTRACT

The genus *Yersinia* contains three species pathogenic for humans, one of which is the enteropathogen *Yersinia pseudotuberculosis*. A recent analysis by Multi Locus Sequence Typing (MLST) of the '*Y. pseudotuberculosis* complex' revealed that this complex comprises three distinct populations: the *Y. pestis*/*Y. pseudotuberculosis* group, the recently described species *Yersinia similis*, and a third not yet characterized population designated 'Korean Group', because most strains were isolated in Korea. The aim of this study was to perform an in depth phenotypic and genetic characterization of the three populations composing the *Y. pseudotuberculosis* complex (excluding *Y. pestis*, which belonged to the *Y. pseudotuberculosis* cluster in the MLST analysis). Using a set of strains representative of each group, we found that the three populations had close metabolic properties, but were nonetheless distinguishable based on D-raffinose and D-melibiose fermentation, and on pyrazinamidase activity. Moreover, high-resolution electrospray mass spectrometry highlighted protein peaks characteristic of each population. Their 16S rRNA gene sequences shared high identity (≥99.5%), but specific nucleotide signatures for each group were identified. Multi-Locus Sequence Analysis also identified three genetically closely related but distinct populations. Finally, an Average Nucleotide Identity (ANI) analysis performed after sequencing the genomes of a subset of strains of each group also showed that intragroup identity (average for each group ≥99%) was higher than intergroup diversity (94.6–97.4%). Therefore, all phenotypic and genotypic traits studied concurred with the initial MLST data indicating that the *Y. pseudotuberculosis* complex comprises a third and clearly distinct population of strains forming a novel *Yersinia* species that we propose to designate *Yersinia wautersii* sp. nov. The isolation of some strains from humans, the detection of virulence genes (on the pYV and pVM82 plasmids, or encoding the superantigen *ypmA*) in some isolates, and the absence of pyrazinamidase activity (a hallmark of pathogenicity in the genus *Yersinia*) argue for the pathogenic potential of *Y. wautersii*.

© 2014 Elsevier GmbH. All rights reserved.

Introduction

The genus *Yersinia* belongs to the *Enterobacteriaceae* family and comprises 17 species. Three of them are pathogenic for humans: *Yersinia pestis*, the agent of plague, and the two enteropathogens *Y. pseudotuberculosis* and *Y. enterocolitica* (Carniel et al., 2006).

The two enteropathogenic species are transmitted by the fecal–oral route, usually through the consumption of contaminated food, and less often by direct contact with an infected animal (Carniel et al., 2006; Fukushima et al., 1989, 1988). Human infections are characterized by diarrhea, fever and abdominal pain that could mimic appendicitis. Systemic infections such as septicemia are rare in children, but are frequently observed in elderly patients with underlying disorders such as diabetes, iron overload

* Corresponding author at: Unité de Recherche Yersinia, Institut Pasteur, 25-28 rue du Dr Roux, 75724 Paris cedex 15, France.
Tel.: +33 1 40 61 37 67; fax: +33 1 45 68 89 54.
E-mail address: cyril.savin@pasteur.fr (C. Savin).

http://dx.doi.org/10.1016/j.ijmm.2014.02.002
1438-4221/© 2014 Elsevier GmbH. All rights reserved.

Table 1
Characteristics of the 40 *Yersinia* strains studied.

Species	Strain number	Geno-serotype	Year	Country	Source	Sequence type[a]
Y. pseudotuberculosis	IP32953	O:1b	1990	France	Human	42
	79136	O:1b	1993	Korea	Human	88
	WP-931108	O:15	1993	Korea	Water	6
	OK5608	O:10	1990	Japan	Raccoon dog	49
	Ryster	O:4a	1968	Japan	Monkey	62
	IP30642	O:1a	1963	Tunisia	Mouse	9
	IP31553	O:6	1969	Japan	Guinea pig	27
	IP32589	O:2a	1981	New Zealand	Human	14
	IP32591	O:12/13	1981	France	Human	41
	IP32670	O:1b	1983	England	Pig	43
	IP32887	O:3	1988	Argentina	Cattle	19
	IP33098	O:2b	2000	France	Hare	16
	IP33178	O:1a	2001	China	Rat	42
	IP33185	O:3	2001	Russia	Human	3
	IP33250	O:3	2001	Russia	Human	32
	IP33278	O:5a	2001	Russia	Human	2
Y. similis	$Y228^T$	O:6	UN	Germany	Rabbit	92[b]
	R2091-2	O:7	1992	Japan	Mole	75
	R2096-1	O:9	1992	Japan	Mole	79
	R850	O:5b	1988	Japan	Mole	75
	OK6609	O:5b	1990	Japan	Marten	74
	R116	O:6	1986	Japan	Mole	76
	R253	O:6	1987	Japan	Mouse	71
	MW1018-2	O:6	1992	Japan	Water	77
	MW48	O:9	1990	Japan	Water	80
	R626R	O:9	1988	Japan	Mole	83
	R2031	O:11	1991	Japan	Mouse	71
	R1505	O:11	1989	Japan	Mole	75
	MW109-2	O:11	1990	Japan	Water	71
	MW332-1	O:10	1991	Japan	Water	81
	Kuratani-2	O:1c	1988	Japan	Water	75
	RH805Ly	O:12/13	1996	Japan	Mole	71
Korean Group	IP30151	O:4a	1960	Sweden	Otter	23
	#51	O:4a	1993	Japan	Human	23
	WP-930601	O:11	1993	Korea	Water	20
	WP-931109	O:15	1993	Korea	Water	22
	WP-931205	O:11	1993	Korea	Water	20
	12-219N1	O:15	1993	Korea	Human	21
	WP-931201	O:15	1993	Korea	Water	21
	Y428	O:4a	2002	Germany	Badger	96[b]

[a] According to Laukkanen-Ninios et al. (2011).
[b] Data from genome sequence. UN: unknown.

or cirrhosis (Ljungberg et al., 1995). It has been estimated that the two enteropathogens, *Y. enterocolitica* and *Y. pseudotuberculosis*, diverged from a common ancestor 41–186 million years ago (Achtman et al., 1999). The third human pathogen, *Y. pestis*, is genetically almost identical to *Y. pseudotuberculosis* (Chain et al., 2004), from which it emerged probably less than 5000 years ago (Cui et al., 2013; Morelli et al., 2010).

Our recent analysis by Multi Locus Sequence Typing (MLST) of the *Y. pseudotuberculosis* population revealed that, in addition to the *Y. pestis*/*Y. pseudotuberculosis* group, this '*Y. pseudotuberculosis* complex' comprises two other distinct populations (Laukkanen-Ninios et al., 2011). One of these corresponds to the recently described species *Yersinia similis* (Sprague et al., 2008), which was formerly thought to be a non-pathogenic subgroup of *Y. pseudotuberculosis* (Fukushima et al., 2001). Another MLST study of *Y. pseudotuberculosis* also identified a distinct population that was called 'cluster B' and that actually was the *Y. similis* group (Ch'ng et al., 2011). The second distinct population received the designation 'Korean Group', because most of the strains composing it were isolated in Korea (Laukkanen-Ninios et al., 2011). This newly identified group has not been studied yet.

The aim of this work was to perform an in depth phenotypic and genetic characterization of the three distinct populations composing the *Y. pseudotuberculosis* complex: *Y. pseudotuberculosis sensu stricto*, *Y. similis* and the Korean Group, to get more insights into the characteristics of these populations. This analysis did not include *Y. pestis*, as its sequence type placed it in the *Y. pseudotuberculosis* cluster in the MLST analysis.

Materials and methods

Bacterial strains

The characteristics of the 40 *Yersinia* strains used in this study are given in Table 1. IP strains were from the collection of the *Yersinia* Research Unit and National Reference Laboratory (Institut Pasteur, Paris, France). *Y. similis* strain $Y228^T$ (CIP109846) was obtained from the Collection of the Institut Pasteur. Korean Group strain Y428 was kindly provided by the Thüringer Landesamt für Verbraucherschutz (TLV), Bad Langensalza, Germany. Other strains originated from the collection of the Shimane Prefectural Institute of Public Health and Environmental Science, Matsue, Japan. All strains were grown overnight in 10 ml Luria Bertani (LB) broth under agitation at 28 °C before performing the assays.

Metabolic and other phenotypic characteristics

Metabolic tests performed on all strains included the API20E and API50CH strips (BioMérieux, Marcy L'étoile, France) incubated at 30 °C, pyrazinamidase (Kandolo and Wauters, 1985) and

tween-esterase activities (Lee, 1977). Additional metabolic characterization of a few strains was carried out with the Phenotypic Microarray Biolog® (AES Chemunex, Ivry-sur-Seine, France), using the PM1 and PM2A (carbon sources), PM3B (nitrogen sources) and PM4A (phosphorus and sulphur sources) 96-well plates. Sodium pyruvate (2 M) was added to PM3B and PM4A. Incubation was performed at 28 °C. Optical density (OD) was recorded at 590 nm and 750 nm after 24 h and 48 h. A corrected OD value (OD_{590}–OD_{750}) of 0.1 was considered as positive. To validate the Biolog® results, auxanograms were performed on a larger strain sample. Briefly, a minimal medium containing 4.69 g NaH_2PO_4, 11.15 g Na_2HPO_4, 2.65 g $(NH_4)_2SO_4$, 0.075 g $MgSO_4$, 3 g KCl, 0.2 ml of 0.1% $FeCl_3$, 0.1 g casaminoacids, 0.02 g yeast extract in 1 l of H_2O was prepared. pH was adjusted to 7.2 before filtration through a 0.22 μm pore filter. The minimal medium was then supplemented with 0.1% (w/v) of the filtered-sterilized carbon source to be tested. A *Yersinia* colony was picked and inoculated into 5 ml of the medium that was incubated at 28 °C under agitation. OD_{600} was recorded at 24 h and 48 h. An OD value of 0.2 was considered as positive. The geno-serotype of the strains was determined as described in (Bogdanovich et al., 2003). Motility was assessed in Mannitol motility semi solid medium (Bio-Rad, Marnes-la-Coquette, France) incubated at 25 °C for 48 h.

Antimicrobial susceptibility

A 2 μl loop of bacteria grown on LB agar plates at 28 °C for 24 h was suspended in 4.5 ml of saline, diluted 1:100 in saline and thoroughly vortexed. Two Mueller Hinton 2 plates were inoculated with 3–4 ml of this bacterial suspension and excess liquid was removed. Plates were dried for 5 min at room temperature before being loaded with antibiotic discs (BioRad, Marnes-la-Coquette, France). Plate 1 was loaded with tetracycline 30 μg, ticarcillin 75 μg, amoxicillin 25 μg, cefoxitin 30 μg, and cefalotin 30 μg discs, and Plate 2 with ciprofloxacin 5 μg, nalidixic acid 30 μg, ceftriaxone 30 μg, amoxicillin/clavulanate 20/10 μg, trimethoprim 5 μg and sulphonamides 200 μg disks. Inhibition diameters were measured after 24 h at 28 °C. Antimicrobial susceptibility was established according to the criteria of the "Société Française de Microbiologie".

Protein profiling by mass spectrometry

Sample preparation was carried out using the trifluoroacetic acid (TFA) protocol (Lasch et al., 2008). Briefly, *Yersinia* strains were grown in LB broth at 28 °C for 24 h and bacterial concentrations were evaluated at OD_{600}. TFA (80% final concentration) was added to 10^9 bacteria and vigorously shaken for 10 min. The cellular extract was centrifuged at 14,000 × g, 4 °C for 15 min, and the supernatant was filtered using Millipore's Ultrafree MC 0.22 μm filter tubes (Millipore, Billerica, MA, USA). Protein extracts were diluted 1:10 in 0.1% formic acid solution and injected in triplicate in reverse-phase liquid chromatography coupled to an LTQ-Orbitrap Discovery mass spectrometer (Thermo Scientific, San Jose, CA, USA). The source conditions were as follows: capillary temperature, 275 °C; sheath gas flow, 30 arbitrary units; auxiliary gas flow, 3 arbitrary units; capillary voltage, 49 V; ESI spray voltage, 4.5 kV. The MS survey scan was performed in the positive ion mode from m/z 500 to 2000 using a resolution set at 30,000 at m/z 400 and FTMS data were collected in profile mode. The Thermo Scientific XCalibur 2.0.7 software was used to acquire data and to produce a mass spectrum with all deconvoluted monoisotopic masses in the range of 5000–20,000 Da. The peak list was exported to Microsoft Excel®. Each sample was analyzed in triplicate, and only masses found in at least two replicates were considered as significant and used to generate the final peak list fingerprint of each strain.

Molecular typing methods

Comparison of 16S rRNA genes was done after PCR amplification of the genes (Janvier and Grimont, 1995) and sequencing (GATC Biotech, Mulhouse, France). The 16S rRNA sequences of all the strains involved in the study have been deposited in Genbank under accession numbers HG326151–HG326187. Alignment of the sequences was performed with Bionumerics 6.6 (Applied Maths, Sint Marten Letem, Belgium) followed by the reconstruction of a phylogenetic tree. Multi-locus sequence analysis (MLSA) was performed as described (Kotetishvili et al., 2005), on four housekeeping genes (*glnA*, *gyrB*, *recA* and *hsp60*). The sequences were concatenated and analyzed using Bionumerics 6.6. A phylogenetic tree was reconstructed using the neighbor joining method.

Genome sequencing and Average Nucleotide Identity (ANI) analysis

A whole genome strategy based on Illumina paired-end sequencing was performed for five *Y. similis* and five Korean Group strains. Illumina library preparation and sequencing followed standard protocols developed by the supplier. Insert were sized using AMPure XP Beads (Agencourt) (±500 bp) and enriched with 10 cycles of PCR before library quantification and validation. Hybridization of the library to the flow cell and bridge amplification was performed to generate clusters, and paired-end reads of 100 cycles were collected on a HiSeq 2000 (Illumina, San Diego, CA). After sequencing was complete, image analysis, base calling, and error estimation were performed using Illumina Analysis Pipeline version 1.7. A de novo assembly was performed with the quality-filtered reads using CLC Assembly CELL v4.0.1 (CLC BIO, Aarhus, Denmark).

For the Korean Group strain Y428 the whole genome sequence was generated using the Illumina sequencing platform (GAII instruments) using standard 200–300 bp fragments and 76 cycles of sequencing. The Illumina data were assembled using the de novo Velvet assembler and the VELVET v0.7.03 algorithm (Zerbino and Birney, 2008), creating multi-contig draft genome.

Average Nucleotide Identity (ANI) analyses were performed using the jspecies package (Richter and Rossello-Mora, 2009). To calculate the ANI between two genomes, the query genome was spliced in 1020 nucleotide fragments, each of them was blasted against the subject genome, and ANI was defined as the mean identity of all BLASTN matches showing more than 30% sequence identity over regions aligned on at least 70% of their length (Goris et al., 2007).

Genome sequences are available in Genbank (accession numbers starting with CBLA) and BioProject (number starting with PRJEB) for strains Y228 (CBLA010000001-974 and PRJEB3982), R2091-2 (CBLB010000001-228 and PRJEB3984), R2096-1 (CBLC010000001-273 and PRJEB3985), R116 (CBLD010000001-182 and PRJEB3986), R2031 (CBLE010000001-231 and PRJEB3987), #51 (CBLF010000001-2423 and PRJEB3988), WP-930601 (CBLG010000001-340 and PRJEB3989), WP-931109 (CBLH010000001-3034 and PRJEB3990), WP-931205 (CBLI010000001-886 and PRJEB3991), and 12-219N1 (CBLJ010000001-3382 and PRJEB3992). The Y428 sequence has been deposited in the European Nucleotide Archive, under study accession number ERR024901.

Amplification of virulence factors

The genomic DNA of each strain was extracted with the PureLink genomic DNA mini kit (Invitrogen, Carlbad, CA, USA), and used for PCR amplification of genes coding for superantigens (*ypmA/C* and *ypmB*) (Fukushima et al., 2001), the high-pathogenicity island

(*ybtE* and *irp2*) (Fukushima et al., 2001; Lesic and Carniel, 2005), the Yersinia Adhesion Pathogenicity Island (*api5*, *pilS*, *api24* to *api26*, *api64* and *api74*) (Collyn et al., 2004), the pVM82 plasmid (*mucAB*) (Eppinger et al., 2007), the pYV plasmid (*yopK*) (Derbise et al., 2012), and the *ail* (attachment invasion locus) gene (491 bp). The PCR primers used are listed in Table S1.

Results

To perform an in depth phenotypic and genetic comparison of the three populations forming the *Y. pseudotuberculosis* complex, we selected a subset of strains (Table 1) among those that were recently subjected to MLST (Laukkanen-Ninios et al., 2011). The 16 *Y. pseudotuberculosis* strains selected represented a variety of geographical origins, years of isolation, and sources. Because the availability of *Y. similis* isolates was limited, the 16 isolates selected were much less diverse. For the Korean Group, we used all seven isolates previously characterized by MLST plus an additional strain (Y428) isolated in Germany, which clustered with the Korean Group, but had a unique ST type (ST96) (Table 1). As the Korean Group strain IP30151 could not be grown again, no new phenotypic analyses could be performed. Nonetheless, the remaining DNA from this strain allowed us to perform a few genetic tests.

Phenotypic comparison of the three populations

We first subjected the three populations of strains to 60 metabolic tests available in the API20E and API50CH strips, plus tween esterase and pyrazinamidase activities. An overall high metabolic homogeneity was observed among the three groups, despite a few characters that were strain-dependent within each group (Fig. S1).

Nonetheless, the 16 *Y. similis* strains studied could be differentiated from both *Y. pseudotuberculosis* and the Korean Group by two metabolic tests. All *Y. similis* strains analyzed had a positive pyrazinamidase activity, while both *Y. pseudotuberculosis* and the Korean Group were pyrazinamidase-negative (Fig. S1). This contrasts with the results of Sprague et al. (Sprague et al., 2008) that defined the *Y. similis species* as being pyrazinamidase-negative. The second *Y. similis*-specific trait was an inability to ferment D-melibiose, as previously observed (Laukkanen-Ninios et al., 2011; Sprague et al., 2008). To further determine whether the inability to ferment D-melibiose was a reliable test to distinguish *Y. similis* from *Y. pseudotuberculosis*, the database of the *Yersinia* National Reference Laboratory (Institut Pasteur) that contains the metabolic characteristics of 3049 *Y. pseudotuberculosis* strains isolated since 1930 was investigated and 33 isolates unable to ferment D-melibiose were identified. They were isolated from animals and humans and were all of serotype O:3 except one strain of serotype O:1. All 33 strains were pyrazinamidase-negative. An MLST analysis performed in a previous study for 11 of them (Laukkanen-Ninios et al., 2011) and in this study for the 22 other strains confirmed that they all belong to the species *Y. pseudotuberculosis* despite their lack of D-melibiose fermentation. Sprague et al. also reported that *Y. similis* strains were negative for the β-galactosidase activity (Sprague et al., 2008). However, we found in this study that 14 out of the 16 *Y. similis* strains studied were β-galactosidase-positive, including the Y228^T type strain (Fig. S1). Furthermore, we found that all *Y. pseudotuberculosis* and Korean Group strains analyzed also exhibited a β-galactosidase activity (Fig. S1), indicating that this test cannot be used for the differentiation of the 3 populations.

The eight isolates from the Korean Group had also some metabolic properties that differentiated them from the two other groups. While none of the *Y. pseudotuberculosis* and *Y. similis* strains tested fermented D-tagatose, most Korean Group strains (5/7) were able to do so (Fig. S1). Furthermore, all Korean Group strains fermented D-raffinose, but none of the *Y. pseudotuberculosis* and *Y. similis* strains tested did so. A search in the *Y. pseudotuberculosis* database identified only one D-raffinose-positive strain, but since this strain was no longer viable, it was not possible to determine whether it belonged to the species *Y. pseudotuberculosis* or *Y. similis*.

To further identify metabolic properties that could differentiate the Korean Group from *Y. pseudotuberculosis* and *Y. similis*, one strain each of *Y. pseudotuberculosis* (IP32953), *Y. similis* (Y228^T) and the Korean Group (WP-931109) was subjected to in depth metabolic characterization, using phenotypic microarrays (Biolog system) assessing utilization of various carbon, nitrogen, phosphorus and sulphur sources. Several additional metabolic tests distinguished *Y. similis* strain Y228^T and Korean Group WP-931109 from each other and from *Y. pseudotuberculosis*, however, when auxanograms were performed on a larger number of isolates, these differing metabolic properties appeared to be strain-specific and not group-specific (data not shown).

The genus *Yersinia* is characterized by a temperature-dependent motility: bacteria are known to be motile at temperatures between 25 °C and 30 °C (except *Y. pestis*), but they are all non-motile at 37 °C (Bejot et al., 1975). Using the mannitol-motility test, we observed that most of the *Y. pseudotuberculosis* strains (10/16) were not motile at 28 °C. This lack of motility is actually not uncommon in the *Y. pseudotuberculosis* strain collection of the French Reference Laboratory (personal communication). Similarly, only 10/16 *Y. similis* strains were motile. The Korean Group was characterized by a complete absence of motility of the 7 strains studied at 28 °C (Fig. S1).

Antimicrobial susceptibility

The susceptibility of the 39 strains to 11 antibiotics commonly used to treat Gram-negative infections (tetracycline, ticarcillin, amoxicillin, amoxicillin/clavulanate, cefoxitin, cefalotin, ciprofloxacin, nalidixic acid, ceftriaxone, trimethoprim and sulphonamides) was determined. All 16 *Y. pseudotuberculosis*, 16 *Y. similis* and 7 Korean Group strains analyzed were susceptible to the 11 antibiotics tested, indicating the absence of intrinsic resistance to antibiotics in these three populations.

Comparison of the three populations by mass spectrometry

Protein profiling by mass spectrometry (MALDI-TOF MS) has been previously applied to species differentiation within the genus *Yersinia* (Ayyadurai et al., 2010; Lasch et al., 2010). However, when a subset of strains from each group (5 *Y. pseudotuberculosis*, 3 *Y. similis* and 4 Korean Group) was subjected to MALDI-TOF MS, no spectral fingerprints typical of each population were obtained (data not shown). To improve peak discrimination we then used high-resolution electrospray mass spectrometry with LTQ-Orbitrap on the 39 strains from the three populations. Between 50 and 70 significant peaks were observed per mass spectrum fingerprint. Protein profiling confirmed the high similarity between the three populations, as illustrated on Fig. 1 which shows mass spectra representative of each population obtained after peak deconvolution. For instance, all strains shared peaks at 7282, 9262, 9655, and 15,176 Da. However, some peaks were group-specific: peak α (9625 Da) for the Korean group (7/7 strains) and peak β (13,278 Da) for *Y. similis* (16/16 strains) (Fig. 1). Peak γ (9639 Da) was unique to *Y. pseudotuberculosis*, but was observed in only 14 of the 16 strains studied. Protein profiling by high-resolution electrospray mass spectrometry thus allowed an unambiguous differentiation of the three populations forming the *Y. pseudotuberculosis* complex.

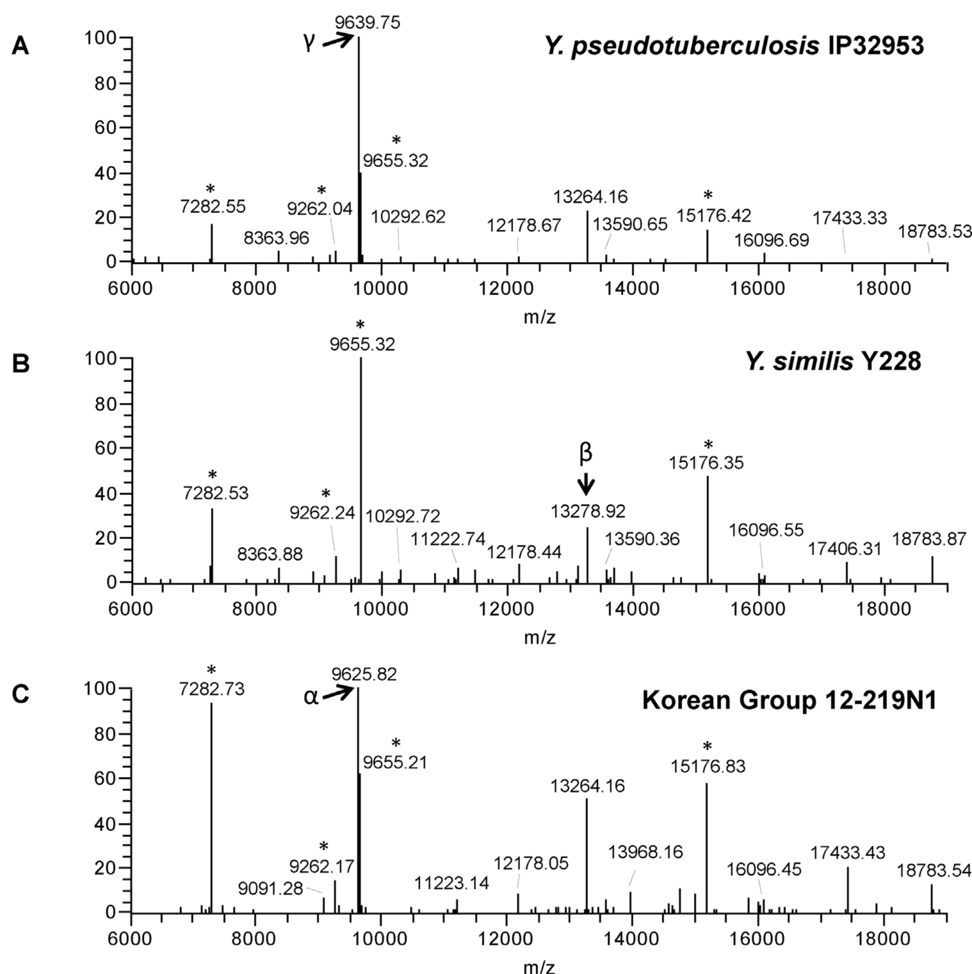

Fig. 1. Deconvoluted electrospray mass spectra representative of each group of strains in the *Y. pseudotuberculosis* complex. Spectra deconvolution was performed with Xtract software. Discriminating peaks (α, β and γ) are indicated on the mass spectrometry spectra. Asterisks indicate peaks shared by the three groups.

Comparison of the 16S rRNA gene sequences of three populations

To compare the three populations at the genetic level, the sequences of the 16S rRNA genes were determined for the 40 strains. As expected, the 16S rRNA genes were highly conserved (average nt identity >99.5%) among all strains of the *Y. pseudotuberculosis* complex (Table 2). However, 16S rRNA gene sequences specifying each population were identified. As previously reported (Sprague et al., 2008), differences in the 16S rRNA sequences of *Y. similis* and *Y. pseudotuberculosis* were observed. For instance a “CTT” stretch at position 1003–1005 in the sequence of *Y. similis* corresponded to “TAA” in *Y. pseudotuberculosis* and to either “TAA” or “TGA” in the Korean Group (Fig. S2), indicating that 1003-CTT-1005 typifies *Y. similis*. This “CTT” stretch may partly destabilize the 16S rRNA secondary structure of this species, as it is located in a region that forms a stem in *Y. pseudotuberculosis* and Korean Group strains, and is not compensated by complementary mutations. Of note, within the Korean Group, the “TAA” stretch characterized strains isolated in Korea, whereas the isolates of other geographical origins had the “TGA” stretch (Fig. S2). Furthermore, the Korean Group had a specific signature at position 189, which was a “C” in these strains instead of an “A” in *Y. similis* and *Y. pseudotuberculosis* (Fig. S2). Since this nucleotide substitution is within a hairpin loop, it should not affect the stability of the 16S rRNA secondary structure.

Table 2
Nucleotide identity of the 16S rRNA genes of the 40 strains studied.

	Y. pseudotuberculosis	*Y. similis*	Korean Group
Y. pseudotuberculosis	99.94 (±0.09)		
Y. similis	99.63 (±0.06)	99.99 (±0.04)	
Korean Group	99.75 (±0.21)	99.56 (±0.14)	99.74 (±0.25)

Indicated are the average percentages of nucleotide identity between or within each group. Numbers into brackets correspond to standard deviations.

Multi Locus Sequence Analysis (MLSA) of the three populations

The 40 selected strains were then subjected to Multi Locus Sequence Analysis (MLSA) (Kotetishvili et al., 2005). The phylogenetic tree obtained from the concatenated sequences of the four housekeeping genes (*glnA*, *gyrB*, *recA* and *hsp60*) displayed three main branches, each corresponding to one of the three populations (Fig. 2). The *Y. similis* group was the most distant (bootstrap value at the node of 100), while the *Y. pseudotuberculosis* and Korean Group branches were closer to each other (bootstrap at the node of 89). Within the Korean Group branch, the three strains of non-Korean origin clustered together (Fig. 2).

Comparison of the three populations at the genomic level

To refine further the genetic comparison of these three populations, the genomes of five *Y. similis* (Y228^{T}, R2091-2, R2096-1, R116 and R2031) and six Korean Group strains (#51, WP-930601,

Table 3
Average Nucleotide Identity among 15 strains of the three populations.

Species		*Y. pseudotuberculosis*				*Y. similis*					Korean Group					
	Strains	IP31758	IP32953	PB1	YPIII	R2091-2	R2096-1	R116	R2031	$Y228^T$	WP-930601	WP-931109	WP-931205	12-219N1	#51	Y428
Y. pst	IP31758	---														
	IP32953	99.03	---													
	PB1	98.95	99.62	---												
	YPIII	99.05	99.15	99.18	---											
Y. sim	R2091-2	94.71	94.53	94.55	94.45	---										
	R2096-1	94.70	94.52	94.46	94.42	99.53	---									
	R116	94.75	94.56	94.58	94.51	99.44	99.53	---								
	R2031	94.89	94.61	94.63	94.61	99.27	99.36	99.26	---							
	$Y228^T$	94.60	94.54	94.53	94.45	99.40	99.39	99.35	99.27	---						
KG	WP-930601	97.60	97.53	97.55	97.48	94.73	94.72	94.74	94.84	94.69	---					
	WP-931109	97.56	97.28	97.54	97.45	94.61	94.67	94.62	94.80	94.72	99.53	---				
	WP-931205	97.50	97.48	97.36	97.54	94.69	94.72	94.79	94.79	94.68	99.33	99.07	---			
	12-219N1	97.44	97.46	97.39	97.37	94.56	94.68	94.74	94.78	94.76	99.53	99.25	99.14	---		
	#51	97.36	97.36	97.31	97.16	94.74	94.88	94.87	94.89	94.77	98.96	99.00	98.89	98.76	---	
	Y428	97.41	97.47	97.47	97.37	95.20	95.09	95.16	95.28	95.10	98.65	98.62	98.69	98.68	98.96	---

Y. pst: *Y. pseudotuberculosis*; Y. sim: *Y. similis*; KG: Korean Group. Results are expressed as percentages. Gray shadings indicate intra-group ANI.

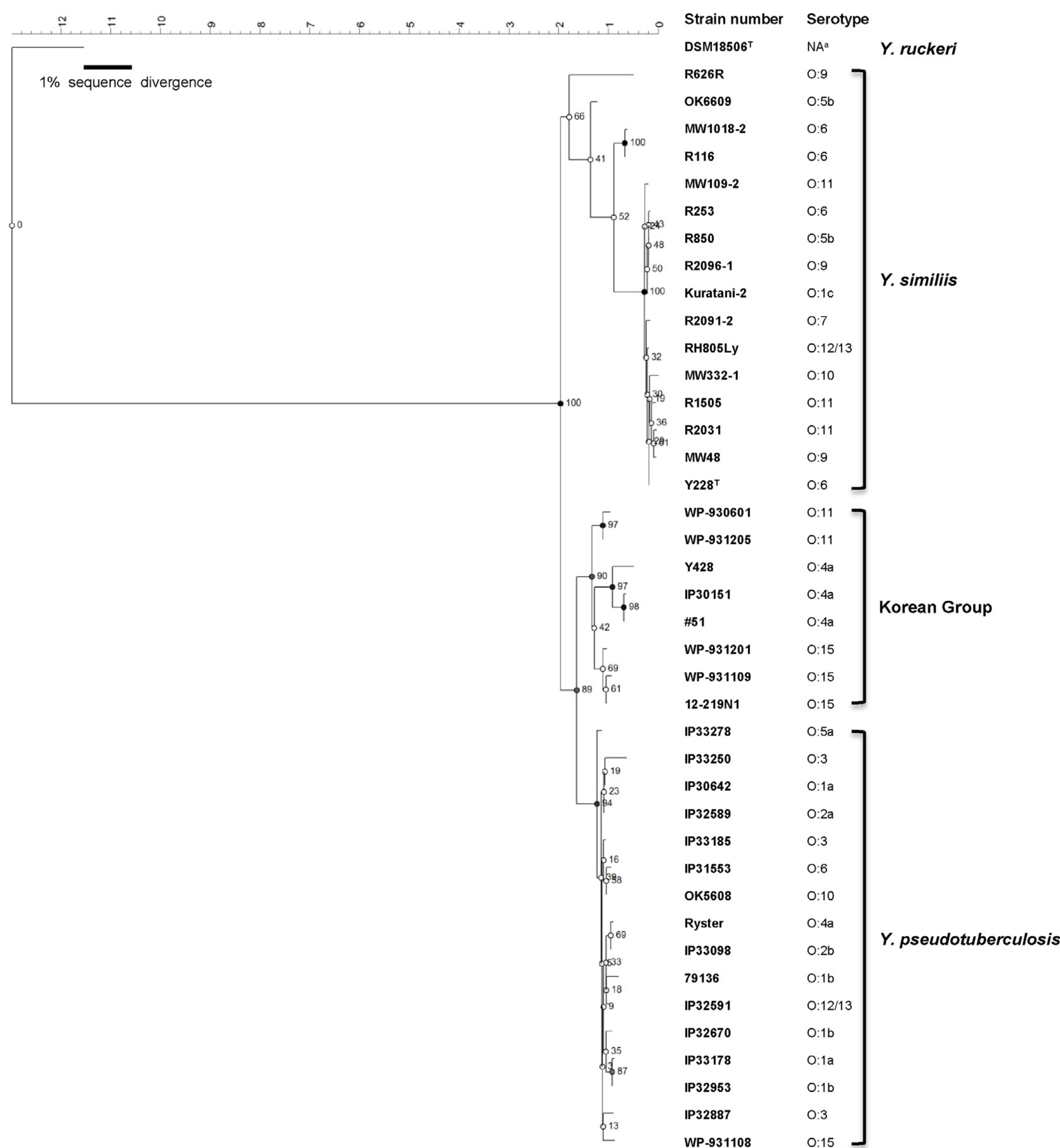

Fig. 2. Neighbor-joining tree obtained by multilocus sequence analysis of 40 strains from the three populations forming the *Y. pseudotuberculosis* complex. Concatenated sequences of four genes (≈2000 bp) were compared using the method based on Jukes-Cantor distance matrix. Bootstrap values obtained after 1000 replicates are given at the nodes. The tree was rooted using *Yersinia ruckeri* (DSM8506^T). Bar represent 1% sequence divergence. NA: not applicable.

WP-931109, WP-931205, 12-219N1 and Y428) were sequenced and compared to four *Y. pseudotuberculosis* genome sequences available in the databases (IP32953 (NC_006155.1), YPIII (NC_010465.1), PB1 (NC_010634.1) and IP31758 (NC_009708.1)). Pair wise Average Nucleotide Identity (ANI) was calculated for the 15 strains. Within each population, the ANI value was always >98.6% (Table 3), with an average identity >99%, demonstrating the genomic homogeneity of these populations. The Korean Group was the most heterogeneous, with an intra-group identity of 99%. Inter-group ANIs were always lower than intra-group ANIs (Table 3), with an average of 94.8% for *Y. similis*/Korean Group, 94.6% for *Y. similis*/*Y. pseudotuberculosis*, and 97.4% for *Y. pseudotuberculosis*/Korean Group.

Presence of virulence factors in the three populations

The pYV virulence plasmid is a requisite for the virulence of human pathogenic *Yersinia* species (Gemski et al., 1980). The presence of this plasmid was analyzed by PCR in the 40 strains belonging to the three populations. The pYV was detected in 9/16 *Y. pseudotuberculosis* strains (Table 4). Its absence in the 7 remaining strain was not unexpected as it is well established that this replicon is easily lost upon subculture in vitro (Fredriksson-Ahomaa and Korkeala, 2003). None of the 16 *Y. similis* strains harbored the pYV, in agreement with a prior report indicating that this species is non-pathogenic (Sprague et al., 2008), Among the 8 Korean Group strains analyzed, one (WP-931109) gave a positive PCR reaction. However, the amplified *yopK* fragment was weak, and when the strain was cultured again, no more PCR fragment was detected, suggesting that the plasmid was present in a small proportion of the bacterial population and was rapidly lost upon subculture.

The chromosomal *ail* (attachment invasion locus) gene is a virulence marker found predominantly in pathogenic *Yersinia*, and commonly used as a PCR target for the detection of pathogenic strains (Ho et al., 2012; Lambertz et al., 2008). PCR amplification of this gene was positive for all *Y. pseudotuberculosis* and Korean Group strains tested, and negative for all *Y. similis* strains (Table 4) (Sprague et al., 2008).

Table 4
Presence of virulence genes in the three populations.

Species	Strain	Serotype	pYV	*ail*	pVM82	HPI		YAPI					Superantigen	
			yopK		*mucAB*	*irp2*	*ybtE*	*api5*	*pilS*	*api24-26*	*api64*	*api74*	*ypmA/C*	*ypmB*
Y. pst	IP32953	O:1b	+	+	–	+	+	–	–	–	–	–	–	–
	79136	O:1b	–	+	–	+	+	+	+	+	+	+	+	–
	WP-931108	O:15	+	+	–	–	–	+	+	+	–	–	+	–
	OK5608	O:10	+	+	–	–	–	+	+	+	–	–	+	–
	Ryster	O:4a	+	+	–	+	+	+	+	+	–	–	–	–
	IP30642	O:1a	–	+	–	+	+	–	–	–	–	–	–	–
	IP31553	O:6	–	+	+	–	–	+	+	+	–	–	+	–
	IP32589	O:2a	–	+	–	–	–	–	–	–	–	–	–	–
	IP32591	O:12/13	+	+	–	+	+	–	–	–	–	–	–	–
	IP32670	O:1b	+	+	–	+	+	+	+	+	+	+	–	–
	IP32887	O:3	–	+	–	–	–	+	+	+	–	–	+	–
	IP33098	O:2b	+	+	–	–	–	–	–	–	–	–	–	–
	IP33178	O:1a	–	+	–	+	+	–	–	–	–	–	–	–
	IP33185	O:3	–	+	–	–	–	–	–	–	–	–	+	–
	IP33250	O:3	+	+	–	–	–	+	+	+	–	–	+	–
	IP33278	O:5a	+	+	–	–	–	+	+	+	–	–	+	–
Y. sim	Y228[T]	O:6	–	–	–	–	–	+	+	–	–	–	–	+
	R2091-2	O:7	–	–	–	–	–	–	+	–	–	–	–	+
	R2096-1	O:9	–	–	–	–	–	+	+	+	–	–	–	+
	R850	O:5b	–	–	–	–	–	+	+	+	–	–	–	+
	OK6609	O:5b	–	–	–	–	–	–	+	+	–	–	–	+
	R116	O:6	–	–	–	–	–	–	–	–	–	–	–	+
	R253	O:6	–	–	–	–	–	–	+	+	–	–	–	+
	MW1018-2	O:6	–	–	–	–	–	+	+	+	–	–	–	+
	MW48	O:9	–	–	–	–	–	–	+	+	–	–	–	+
	R626R	O:9	–	–	–	–	–	–	+	+	–	–	–	+
	R2031	O:11	–	–	–	–	–	+	+	–	–	–	–	+
	R1505	O:11	–	–	–	–	–	–	+	+	–	–	–	+
	MW109-2	O:11	–	–	–	–	–	+	+	–	–	–	–	+
	MW332-1	O:10	–	–	–	–	–	–	+	+	–	–	–	+
	Kuratani-2	O:1c	–	–	–	–	–	+	+	+	–	–	–	–(+)[a]
	RH805Ly	O:12/13	–	–	–	–	–	–	+	+	–	–	–	+
KG	WP-930601	O:11	–	+	–	–	–	+	+	+	–	–	–	–
	WP-931205	O:11	–	+	–	–	–	–	–	–	–	–	–	–
	12-219N1	O:15	–	+	–	–	–	+	+	+	–	–	–	–
	WP-931201	O:15	–	+	–	–	–	–	–	–	–	–	–	–
	WP-931109	O:15	±	+	–	–	–	–	–	–	–	–	±	–
	IP30151	O:4a	–	ND	+[b]	–	–	ND	ND	ND	ND	ND	+	–
	#51	O:4a	–	+	–	–	–	+	+	+	–	–	–	–
	Y428	O:4a	–	+	–	–	–	–	–	–	–	–	–	–

Y. pst, *Y. pseudotuberculosis*; *Y. sim*, *Y. similis*; KG, Korean Group. +, PCR fragment at the expected size; –, no PCR fragment; ±, faint PCR fragment, inconsistently detected.
[a] Gene not detected by PCR, but present in a truncated form.
[b] According to Eppinger et al. (2007). ND: not done due to a lack of DNA.

Another plasmid (pVM82) has been reported to play a role in the virulence of certain strains of *Y. pseudotuberculosis* causing an infection known as Far East Scarlet-Like Fever (FESLF) (Cornelis et al., 1998; Gintsburg et al., 1988) and characterized by specific clinical manifestations (rash, skin desquamation, exanthema, hyperhemic tongue, diarrhea, fever and toxic shock syndrome) (Carniel et al., 2006; Dubrovina et al., 1999; Eppinger et al., 2007; Gintsburg et al., 1988). pVM82 (amplifications of *mucAB*) was detected in one *Y. pseudotuberculosis* strain (IP31553) isolated in Japan, and in one Korean Group strain (IP30151), but in none of the 16 *Y. similis* strains analyzed (Table 4).

Two chromosomal mobile elements that carry virulence determinants, the High Pathogenicity Island (HPI) (Carniel, 1999), and the *Yersinia* Adhesion Pathogenicity Island (YAPI) (Collyn et al., 2004) are also found in a subset of *Y. pseudotuberculosis* strains. The presence of these two genomic islands was investigated in the 40 strains. The HPI was present in 7/16 *Y. pseudotuberculosis* isolates, and was absent from all *Y. similis* and Korean Group strains. A complete YAPI (i.e. possessing the 5 fragments amplified by PCR) was detected in only 2 *Y. pseudotuberculosis* strains (79136 and IP32670). The right-end part of YAPI, encompassing *api64* to *api74* was absent from all 37 other strains belonging to the three populations, while portions of the left-end part (*api5* to *api26*) were frequently detected in *Y. pseudotuberculosis*, *Y. similis* and Korean Group strains (Table 4).

Finally, another virulence determinant whose presence is strain-dependent in *Y. pseudotuberculosis* is the superantigen YPM (Fukushima et al., 2001). Three superantigen genes have been described in *Y. pseudotuberculosis* (Fukushima et al., 2001; Uchiyama et al., 1993): *ypmA* and *ypmC* differ by a single nucleotide (Fukushima et al., 2001). The virulence-associated *ypmA/C* locus (Carnoy et al., 2000), was detected in 9/16 *Y. pseudotuberculosis* and 2/8 Korean Group strains, but in none of the 16 *Y. similis* strains (Table 4), as previously reported (Fukushima et al., 2001; Sprague et al., 2008). In contrast, the *ypmB* locus, which encodes another superantigen not associated with virulence (Fukushima et al., 2001), was absent from all *Y. pseudotuberculosis* and Korean Group strains and was present in all but one (Kuratani-2) *Y. similis* isolates (Table 4). Amplification and sequencing of the chromosomal region encompassing the *ypm* locus in strain Kuratani-2 revealed that a portion of the *ypmB* gene is missing in this strain (data not shown), thus preventing the amplification of the gene with the primers used. The presence of a *ypmB* locus thus specifies the species *Y. similis* within the *Y. pseudotuberculosis* complex.

Altogether, our results confirm the presence of virulence genes in the pathogenic *Y. pseudotuberculosis* species, and in

agreement with prior reports, that these genes are absent from the non-pathogenic *Y. similis* species (Sprague et al., 2008). Five determinants (pYV, *ail*, YAPI, pVM82 and *ypmA*) that encode virulence factors in *Y. pseudotuberculosis* were also detected in some isolates of the Korean Group, arguing for a pathogenic potential of this population of strains.

Discussion

The aim of this study was to perform an in depth characterization of the three populations recently shown to form distinct MLST branches in the *Y. pseudotuberculosis* complex (Laukkanen-Ninios et al., 2011). Two of these populations, *Y. pseudotuberculosis* and *Y. similis*, were already identified as distinct species, but limited information was available for the recently identified *Y. similis* species. The third population, referred to as Korean Group (because most strains composing it originated from Korea), was not yet characterized. As expected from their clustering into a single complex, we found that these three populations displayed a high level of phenotypic and genetic homogeneity. However, whatever the method used, each population displayed unique phenotypic and genetic markers that distinguished them from the other two.

The recently described species *Y. similis* was reported to be pyrazinamidase-negative (Sprague et al., 2008). However, we found here that all 16 *Y. similis* strains studied exhibited a pyrazinamidase activity. This discrepancy could not be attributed to individual strain variations as we also tested the *Y. similis* type strain Y228^T, previously used by Sprague et al. to define the species, and found that it was also pyrazinamidase-positive. Since the intensity of the pyrazinamidase reaction is variable, it is possible that a positive but faint activity in *Y. similis* isolates was not detected in the original study. The second *Y. similis*-specific trait identified by Sprague et al. in four strains was an inability to ferment D-melibiose (Sprague et al., 2008). Our previous MLST analysis of 23 *Y. similis* strains and this study, which used a subset of these strains plus the type strain, confirm that this species does not ferment this sugar. However, lack of D-melibiose fermentation is not a *Y. similis*-specific trait, as some *Y. pseudotuberculosis* strains are also D-melibiose-negative (Laukkanen-Ninios et al., 2011). Indeed, Fukushima et al. found that ≈11% of the *Y. pseudotuberculosis* strains analyzed (after exclusion of the genetic group G4, now known to be *Y. similis*) did not ferment D-melibiose (Fukushima et al., 2001). We also identified 1% D-melibiose-negative *Y. pseudotuberculosis* strains in the collection of the French Reference Laboratory for *Yersinia*. However, 267 of the 268 D-melibiose-negative *Y. pseudotuberculosis* strains (235 from the Fukushima's study, and 33 from the French Reference Laboratory), were of serotype O:3, while none of the *Y. similis* strains analyzed had this serotype (Fukushima et al., 2001; Sprague et al., 2008). Therefore, an inability to ferment D-melibiose associated with a serotype other than O:3 strongly orients toward the species *Y. similis*. This species was also defined as unable to hydrolyze o-nitrophenyl-β-D-galactopyranoside (ONPG reaction indicative of a β-galactosidase activity), while the majority of the *Y. pseudotuberculosis* tested (75% of the 36 strains) were reported to have this activity (Sprague et al., 2008). Here we found that all 16 *Y. pseudotuberculosis*, all 8 Korean Group, and 14/16 *Y. similis* strains tested were able to hydrolyze ONPG, and therefore that this test could not be used to identify *Y. similis*. Therefore, the species *Y. similis* can be differentiated phenotypically from *Y. pseudotuberculosis* and the Korean Group by its positive pyrazinamidase activity and its inability to ferment D-melibiose associated with a non-O:3 serotype.

MALDI-TOF, which is a technique now often used for species identification (Sauer et al., 2008), was not resolutive enough to discriminate the three populations within the *Y. pseudotuberculosis* complex. However, protein fingerprint obtained by high-resolution electrospray mass spectrometry allowed to unambiguously differentiate *Y. similis* from *Y. pseudotuberculosis* and the Korean Group. Although this method is not commonly used for routine identification, it further demonstrates that *Y. similis* has additional specific phenotypic markers.

From a genetic point of view, *Y. similis* is known to have 16S rRNA sequences and MLST profiles that distinguish it from *Y. pseudotuberculosis* (Ch'ng et al., 2011; Laukkanen-Ninios et al., 2011; Sprague et al., 2008). Our analysis of the 16S rRNA sequence performed on additional *Y. similis* strains, and our MLSA further demonstrate that this species forms a distinct genetic cluster in the *Y. pseudotuberculosis* complex, and that simple genotypic methods can also be used to identify *Y. similis* strains. At the whole genome level, the five *Y. similis* strains sequenced displayed a high genetic homogeneity (average ANI value ≥99.4%), as expected for strains belonging to the same species. The ANI values between *Y. similis* and *Y. pseudotuberculosis* (average of 94.6%) and between *Y. similis* and the Korean Group (average of 94.8%) were lower, but still high, confirming that this species forms a distinct group genetically close to *Y. pseudotuberculosis* and the Korean Group.

As the Korean Group was only recently defined, very little information was available concerning this population (Laukkanen-Ninios et al., 2011). These strains were classified as *Y. pseudotuberculosis* based on their biochemical profile. Our extensive analysis of their metabolic properties confirmed that the Korean Group has numerous biochemical reactions similar to those of *Y. pseudotuberculosis*. However, some differences specific for this population were observed. D-raffinose fermentation was systematically positive in the Korean Group and negative in both *Y. pseudotuberculosis* and *Y. similis*. D-raffinose fermentation is due to an α-galactosidase (Schmid and Schmitt, 1976) enzyme that catalyzes the transformation of D-raffinose into D-galactose and sucrose (Schmid and Schmitt, 1976). An analysis of the genome sequences of the three populations generated in this study showed that *Y. similis* strains lack this gene, thus explaining their absence of D-raffinose fermentation. However, the α-galactosidase-encoding gene was present in the *Y. pseudotuberculosis* genomes, did not display any frameshift, and shared ≥99% nucleotide identity with the Korean Group gene. One possible explanation for the absence of α-galactosidase activity in *Y. pseudotuberculosis* might be a difference in gene regulation. Most Korean Group strains fermented D-tagatose (5/7), while none of the *Y. pseudotuberculosis* and *Y. similis* strains tested were able to do so. The use of D-tagatose as a carbon source is mediated by two transporters and a kinase in *Klebsiella oxytoca* (Van der Heiden et al., 2013). However, none of these *K. oxytoca* genes were detected in the Korean Group genomes, suggesting that these strains use another pathway to ferment D-tagatose. Korean Group strains were systematically non-motile whereas 6/16 *Y. pseudotuberculosis* strains and 10/16 *Y. similis* were motile. Motility is mediated by one or several flagella whose various components are encoded by 22 *fli* genes (Carniel and Hinnebusch, 2004). Comparison of these 22 genes in *Y. pseudotuberculosis* and the Korean Group identified three divergent genes: (i) the predicted products of *fliD* had the same size, but they shared only 80% amino acid (aa) identity; (ii) the *fliC* product was larger in the Korean Group (569 aa) than in *Y. pseudotuberculosis* (369 aa), due to a large insertion in the coding sequence that did not change the reading frame; and (iii) the *fliK* of the Korean Group had an 18 bp insertion (HHELIS protein motif), a 6 bp insertion (KT motif), and a 3 bp deletion (S) that did not modify the reading frame. These differences in *fliC*, *fliD*, and *fliK* between *Y. pseudotuberculosis* and the Korean Group were found in all genome sequences, indicating that they were population-specific. They may be responsible for the lack of motility of the Korean Group, but additional experiments would be necessary to demonstrate it. The Korean Group could also be differentiated phenotypically from

Y. pseudotuberculosis and *Y. similis* by its specific high-resolution electrospray mass spectrometry protein profile. Finally, in addition to these unique phenotypic properties, the Korean Group displayed a specific 16S rRNA signature that can also be used to differentiate it from the two other populations.

Our previous MLST analysis identified the Korean Group as a distinct cluster within the *Y. pseudotuberculosis* complex (Laukkanen-Ninios et al., 2011): the seven Korean Group isolates had nine unique alleles in *adk*, *glnA*, *trpE*, *thrA* and *aroA*. Furthermore, the pairwise distance between the concatenated sequences of individual STs ranged from 0.08% to 0.5% within the Korean Group, while this distance was systematically higher in pairwise comparisons with *Y. pseudotuberculosis* (0.8–1.8%) and with *Y. similis* (4.2–4.5%) isolates. These results also indicated that the Korean Group is genetically closer to *Y. pseudotuberculosis* than to *Y. similis*. This is confirmed by our MLSA study: the resulting neighbor joining tree formed three distinct branches, with the *Y. pseudotuberculosis* and Korean Group branches being closer to each other than to the *Y. similis* branch. This contrasts with the previous phylogenetic tree based on the 1436 bp fragment of 16S RNA genes of various *Yersinia* species in which the only Korean Group strain studied clustered with *Y. similis* (Laukkanen-Ninios et al., 2011). Our whole genome analysis confirms that the Korean Group is genetically closer to *Y. pseudotuberculosis* than to *Y. similis*. Indeed, the ANI between *Y. pseudotuberculosis* and the Korean Group (average of 97.4%) was higher than the ANI between *Y. similis* and the Korean Group (average of 94.8%). Despite its closer genetic relatedness to *Y. pseudotuberculosis*, the Korean Group formed a homogeneous cluster (average intra-group ANI of 99%), distinct from the *Y. pseudotuberculosis* cluster.

These results therefore confirm that the three populations first identified by MLST correspond to three genetically and phenotypically distinct groups. The definition of species was previously based on percentages of DNA–DNA hybridization, but this technique was not very easy to apply and showed limits that led to the use of more convenient approaches such as 16S rRNA gene sequence comparisons (DeLong and Pace, 2001). However, this technique is not suitable to define species within the genus *Yersinia*, as all species have 16S rRNA sequence identities superior to the 97% cut-off value proposed for species delineation (Hurst et al., 2011; Merhej et al., 2008; Murros-Kontiainen et al., 2011). An ANI value ≥95% has also been proposed to define bacterial species (Goris et al., 2007). Following this criterion, the Korean Group would not represent a species distinct from *Y. pseudotuberculosis*. However, using this criterion *Burkholderia pseudomallei* and *Burkholderia mallei* should also be a single species, as their genomes share 98.7% ANI (Holden et al., 2004; Nierman et al., 2004), thus demonstrating that the use of strict cut-off values do not reflect the more complex concept of species. Our observation that intra-group ANIs for the Korean Group were always superior to inter-group ANIs argued for a genetically close but distinct group. Furthermore, the MLST study of the *Y. pseudotuberculosis* complex demonstrated high specific fixation index values (FST = 0.93, $p < 0.01$) and high pairwise genetic distances between populations (0.79–0.97) (Laukkanen-Ninios et al., 2011). The MLSA approach has been increasingly used to establish the phylogenetic position of new species (Hanage et al., 2006). The neighbor-joining tree generated from our MLSA analysis of the *Y. pseudotuberculosis* complex formed three defined and independent branches corresponding to each population. It was previously suggested that the Korean Group might be in the process of speciation (Laukkanen-Ninios et al., 2011). The unequivocal genetic distinction of the Korean Group from *Y. pseudotuberculosis* and *Y. similis*, plus the existence of specific phenotypic markers, thus strengthen the premise th at the Korean Group is a new *Yersinia* species for which we propose the name *Yersinia wautersii* sp. nov. (Table 5). Strain 12-219N1 was chosen as the type strain because it was isolated in Korea, from a human case, and its genome sequence is available and deposited in public databases.

Table 5
Description of *Yersinia wautersii* sp. nov.

Name	*Yersinia wautersii* (wau.ter.si'i, N.L. gen. masc. n. *wautersii*. Named in honor of Professor George Wauters, Belgian yersiniologist to whom we owe in depth characterization of the genus *Yersinia*, including the biotyping scheme of *Y. enterocolitica*.
Morphology	Cells are Gram-negative and coccoid rods. Small Colonies appear on TSA agar after 24 h of growth at 28 °C or 37 °C.
Positive metabolic properties	β-Galactosidase, urease, glycerol, L-arabinose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, D-mannitol, N-acetylglucosamine, arbutin, esculin, D-maltose, D-trehalose, D-raffinose and D-arabitol.
Negative metabolic properties	Arginin decarboxylase, lysin decarboxylase, ornithin decarboxylase, tryptophan deaminase, H_2S production, indole production, Voges-Proskauer, gelatinase, pyrazinamidase, erythritol, D-arabinose, L-xylose, methyl-β-D-xylopyranoside, L-sorbose, dulcitol, inositol, D-sorbitol, methyl-β-D-mannopyranoside, methyl-β-D-glucopyranoside and amygdalin.
Species description	Based on seven strains.
Type strain	12-219N1^T (=CIP110607^T = DSM27350^T) Serotype O:15. Isolated from human stools in Korea in 1993.
G + C content	46.6 mol%.
16S rRNA sequence	Accession number HG326166 in Genbank.

From a medical perspective, our study shows that the three closely related species composing the *Y. pseudotuberculosis* complex can be readily distinguished with simple phenotypic (D-melibiose, D-raffinose, and pyrazinamidase), and genotypic (MLSA and 16S rRNA sequencing) methods (Table S2), easily applicable in clinical pathology laboratories. In contrast, although MALDI-TOF MS is becoming a frequently used method for bacterial identification (Sauer et al., 2008) and has been applied to species differentiation within the genus *Yersinia* (Ayyadurai et al., 2010; Lasch et al., 2010), this approach is not applicable to the distinction of the three species within the *Y. pseudotuberculosis* complex. This is not unexpected as the strains composing this complex are phylogenetically much closer to each other than to other *Yersinia* species. From a therapeutic point of view, our results also indicate that while *Y. enterocolitica* is resistant to beta-lactams, and first and second-generation cephalosporins (Pham et al., 1995, 1991, 2000), all species within the *Y. pseudotuberculosis* complex are naturally susceptible to these molecules and to other antibiotics commonly used to treat Gram-negative infections.

While *Y. pseudotuberculosis* is known to be pathogenic for humans (Carniel et al., 2006; Mair, 1965; Tsubokura et al., 1989), *Y. similis* has only been isolated from environmental and animal reservoirs and is considered to be a non-pathogenic species (Fukushima et al., 2001; Sprague et al., 2008). In agreement with this, we found that none of the virulence regions that characterize the different pathotypes of *Y. pseudotuberculosis* were present in the *Y. similis* strains studied. In contrast, all *Y. similis* isolates harbored the non-virulence associated *ypmB* locus, which was absent from *Y. pseudotuberculosis* and *Y. wautersii*, and could thus be used as an additional genetic marker for *Y. similis* identification. It has previously been shown that there is an almost perfect association between pyrazinamidase activity and pathogenicity in the genus *Yersinia*: non-pathogenic species are all pyrazinamidase-positive, and pathogenic *Yersinia* are systematically pyrazinamidase-negative (Kandolo and Wauters, 1985). A positive pyrazinamidase activity, and the absence of virulence

markers and clinical cases thus confirm that *Y. similis* is a non-pathogenic species for humans.

In contrast to *Y. similis*, some *Y. wautersii* strains were isolated from human stools, arguing for their pathogenic potential. However, since no data regarding the clinical symptoms of these *Y. wautersii*-infected individuals are available, this observation is not sufficient to reach conclusions about the pathogenicity of this species. Further arguing for a pathogenic potential was the presence of virulence genes in some *Y. wautersii* isolates. The pYV virulence plasmid, which is a hallmark of human pathogenic *Yersinia* strains (Cornelis et al., 1998; Gemski et al., 1980), was detected in only one *Y. wautersii* isolate but, this plasmid is known to be easily lost upon subculture in vitro (Fredriksson-Ahomaa and Korkeala, 2003). This is further exemplified by the fact that the pYV-positive *Y. wautersii* strain rapidly lost its plasmid during this study, and that 7 of the 16 *Y. pseudotuberculosis* strains studied were also pYV-negative, although several of them were isolated from humans with clinical symptoms, including one (IP32589), which was isolated from a human lymph node. Therefore, the observation that one *Y. wautersii* strain harbored the pYV, even if transiently, is an argument for the pathogenicity of this species. Because of the absence/loss of the pYV, we were unable to test its virulence in the mouse experimental model of infection. Isolation of *Y. wautersii* strains from deep organs in humans or carrying the pYV would definitively demonstrate the pathogenicity of this species. Its pathogenic potential is nonetheless further supported by the detection of the virulence-associated *ail* gene in all *Y. wautersii* strains tested and of the *ypmA/C* gene in two strains, one of which (IP30151) also harbored the *mucAB* locus located on the pVM82 plasmid. The presence of both the super antigen-encoding gene and the pVM82 plasmid would classify this strain in the pathotype causing Far-East Scarlet-Like Fever (Eppinger et al., 2007). Finally, the absence of pyrazinamidase activity, a hallmark of pathogenicity in other *Yersinia* species, further argues for the potential pathogenicity of *Y. wautersii*. Therefore, our observations are compatible with this newly described species being the fourth *Yersinia* species pathogenic for humans.

Although most *Y. wautersii* strains came from Korea, the isolation of three strains in other countries (Japan, Germany and Sweden) indicates the geographic spread of this species. It is possible that other *Y. wautersii* have already been isolated elsewhere, but have been misidentified as *Y. pseudotuberculosis*. This study provides the tools for their proper identification.

Acknowledgements

This work was partly funded by the Institut de Veille Sanitaire (InVS, Saint-Maurice, France) and by the Scientific Foundation of Ireland (05/FE1/B882 to Z.Z.). We thank Odile Bouvet (Inserm U722, Paris, France) for her advices regarding auxanograms, Laurence Ma (Genomic Platform, Institut Pasteur, France) for performing the sequencing of *Y. similis* and Korean Group genomes, Mark Achtman for his valuable comments on the manuscript, the Thüringer Landesamt für Verbraucherschutz (TLV, Bad Langensalza. Germany) for providing the Y428 strain, and Chantal Bizet (Collection de l'Institut Pasteur, Paris, France) for providing CIP *Yersinia* strains. The sequencing of Korean Group strain Y428 was supported by Wellcome Trust grant 098051.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijmm.2014.02.002.

References

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

Ayyadurai, S., Flaudrops, C., Raoult, D., Drancourt, M., 2010. Rapid identification and typing of *Yersinia pestis* and other *Yersinia* species by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. BMC Microbiol. 10, 285.

Bejot, J., Alonso, J.M., Mollaret, H.H., 1975. Le diagnostic bactériologique des Yersinioses humaines (Infections à *Yersinia pseudotuberculosis* et *enterocolitica*). Méd. Maladies Infect. 5, 233–236.

Bogdanovich, T., Carniel, E., Fukushima, H., 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.

Carniel, E., 1999. The *Yersinia* high-pathogenicity island. Int. Microbiol. 2, 161–167.

Carniel, E., Autenrieth, I., Cornelis, G., Fukushima, H., Guinet, F., Isberg, R., Pham, J., Prentice, M., Simonet, M., Skurnik, M., Wauters, G., 2006. *Y. enterocolitica* and *Y. pseudotuberculosis*. In: Dworkin (Ed.), The Prokaryotes. Springer, New York, pp. 270–398.

Carniel, E., Hinnebusch, B.J., 2004. *Yersinia*: Molecular and Cellular Biology. Horizon Biosciences.

Carnoy, C., Mullet, C., Muller-Alouf, H., Leteurtre, E., Simonet, M., 2000. Superantigen YPMa exacerbates the virulence of *Yersinia pseudotuberculosis* in mice. Infect. Immun. 68, 2553–2559.

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

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., Brubaker, R.R., Fowler, J., Hinnebusch, J., Marceau, M., Medigue, C., Simonet, M., Chenal-Francisque, V., Souza, B., Dacheux, D., Elliott, J.M., Derbise, A., Hauser, L.J., Garcia, E., 2004. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. U. S. A. 101, 13826–13831.

Collyn, F., Billault, A., Mullet, C., Simonet, M., Marceau, M., 2004. YAPI, a new *Yersinia pseudotuberculosis* pathogenicity island. Infect. Immun. 72, 4784–4790.

Cornelis, G.R., Boland, A., Boyd, A.P., Geuijen, C., Iriarte, M., Neyt, C., Sory, M.P., Stainier, I., 1998. The virulence plasmid of *Yersinia*, an antihost genome. Microbiol. Mol. Biol. Rev. 62, 1315–1352.

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Xu, L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z., Du, Z., Wu, H., Yang, X., Cao, H., Wang, H., Wang, J., Yao, S., Rakin, A., Falush, D., Balloux, F., Achtman, M., Song, Y., Yang, R., 2013. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. Proc. Natl. Acad. Sci. U. S. A. 110, 577–582.

DeLong, E.F., Pace, N.R., 2001. Environmental diversity of bacteria and archaea. Syst. Biol. 50, 470–478.

Derbise, A., Cerda Marin, A., Ave, P., Blisnick, T., Huerre, M., Carniel, E., Demeure, C.E., 2012. An encapsulated *Yersinia pseudotuberculosis* is a highly efficient vaccine against pneumonic plague. PLoS Negl. Trop. Dis. 6, e1528.

Dubrovina, V.I., Golubinskii, E.P., Borsuk, G.I., Balakhonov, S.V., Konovalova Zh, A., 1999. Characteristics of phagocytosis of *Yersinia pseudotuberculosis* with various sets of plasmids. Med. Parazitol. (Mosk), 50–53.

Eppinger, M., Rosovitz, M.J., Fricke, W.F., Rasko, D.A., Kokorina, G., Fayolle, C., Lindler, L.E., Carniel, E., Ravel, J., 2007. The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. PLoS Genet. 3, e142.

Fredriksson-Ahomaa, M., Korkeala, H., 2003. Low occurrence of pathogenic *Yersinia enterocolitica* in clinical, food, and environmental samples: a methodological problem. Clin. Microbiol. Rev. 16, 220–229.

Fukushima, H., Gomyoda, M., Ishikura, S., Nishio, T., Moriki, S., Endo, J., Kaneko, S., Tsubokura, M., 1989. Cat-contaminated environmental substances lead to *Yersinia pseudotuberculosis* infection in children. J. Clin. Microbiol. 27, 2706–2709.

Fukushima, H., Gomyoda, M., Shiozawa, K., Kaneko, S., Tsubokura, M., 1988. *Yersinia pseudotuberculosis* infection contracted through water contaminated by a wild animal. J. Clin. Microbiol. 26, 584–585.

Fukushima, H., Matsuda, Y., Seki, R., Tsubokura, M., Takeda, N., Shubin, F.N., Paik, I.K., Zheng, X.B., 2001. Geographical heterogeneity between Far Eastern and Western countries in prevalence of the virulence plasmid, the superantigen *Yersinia pseudotuberculosis*-derived mitogen, and the high-pathogenicity island among *Yersinia pseudotuberculosis* strains. J. Clin. Microbiol. 39, 3541–3547.

Gemski, P., Lazere, J.R., Casey, T., Wohlhieter, J.A., 1980. Presence of a virulence-associated plasmid in *Yersinia pseudotuberculosis*. Infect. Immun. 28, 1044–1047.

Gintsburg, A.L., Shubin, F.N., Shovadaeva, G.A., Kulichenko, A.N., Ianishevskii, N.V., 1988. A new pathogenic trait encoded by *Yersinia pseudotuberculosis* pVM82 plasmid. Genetika 24, 1562–1571.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91.

Hanage, W.P., Fraser, C., Spratt, B.G., 2006. Sequences, sequence clusters and bacterial species. Philos. Trans. R. Soc. Lond. B Biol. Sci. 361, 1917–1927.

Ho, D.K., Riva, R., Kirjavainen, V., Jarva, H., Ginstrom, E., Blom, A.M., Skurnik, M., Meri, S., 2012. Functional recruitment of the human complement inhibitor C4BP to *Yersinia pseudotuberculosis* outer membrane protein Ail. J. Immunol. 188, 4450–4459.

Holden, M.T., Titball, R.W., Peacock, S.J., Cerdeno-Tarraga, A.M., Atkins, T., Crossman, L.C., Pitt, T., Churcher, C., Mungall, K., Bentley, S.D., Sebaihia, M., Thomson, N.R., Bason, N., Beacham, I.R., Brooks, K., Brown, K.A., Brown, N.F., Challis, G.L., Cherevach, I., Chillingworth, T., Cronin, A., Crossett, B., Davis, P., DeShazer, D., Feltwell, T., Fraser, A., Hance, Z., Hauser, H., Holroyd, S., Jagels, K., Keith, K.E., Maddison, M., Moule, S., Price, C., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Simmonds, M., Songsivilai, S., Stevens, K., Tumapa, S., Vesaratchavest, M., Whitehead, S., Yeats, C., Barrell, B.G., Oyston, P.C., Parkhill, J., 2004. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. Proc. Natl. Acad. Sci. U. S. A. 101, 14240–14245.

Hurst, M.R., Becher, S.A., Young, S.D., Nelson, T.L., Glare, T.R., 2011. *Yersinia entomophaga* sp. nov., isolated from the New Zealand grass grub *Costelytra zealandica*. Int. J. Syst. Evol. Microbiol. 61, 844–849.

Janvier, M., Grimont, P.A., 1995. The genus *Methylophaga*, a new line of descent within phylogenetic branch gamma of Proteobacteria. Res. Microbiol. 146, 543–550.

Kandolo, K., Wauters, G., 1985. Pyrazinamidase activity in *Yersinia enterocolitica* and related organisms. J. Clin. Microbiol. 21, 980–982.

Kotetishvili, M., Kreger, A., Wauters, G., Morris Jr., J.G., Sulakvelidze, A., Stine, O.C., 2005. Multilocus sequence typing for studying genetic relationships among *Yersinia* species. J. Clin. Microbiol. 43, 2674–2684.

Lambertz, S.T., Nilsson, C., Hallanvuo, S., 2008. TaqMan-based real-time PCR method for detection of *Yersinia pseudotuberculosis* in food. Appl. Environ. Microbiol. 74, 6465–6469.

Lasch, P., Drevinek, M., Nattermann, H., Grunow, R., Stammler, M., Dieckmann, R., Schwecke, T., Naumann, D., 2010. Characterization of *Yersinia* using MALDI-TOF mass spectrometry and chemometrics. Anal. Chem. 82, 8464–8475.

Lasch, P., Nattermann, H., Erhard, M., Stammler, M., Grunow, R., Bannert, N., Appel, B., Naumann, D., 2008. MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. Anal. Chem. 80, 2026–2034.

Laukkanen-Ninios, R., Didelot, X., Jolley, K.A., Morelli, G., Sangal, V., Kristo, P., Brehony, C., Imori, P.F., Fukushima, H., Siitonen, A., Tseneva, G., Voskressenskaya, E., Falcao, J.P., Korkeala, H., Maiden, M.C., Mazzoni, C., Carniel, E., Skurnik, M., Achtman, M., 2011. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. Environ. Microbiol. 13, 3114–3127.

Lee, W.H., 1977. Two plating media modified with Tween 80 for isolating *Yersinia enterocolitica*. Appl. Environ. Microbiol. 33, 215–216.

Lesic, B., Carniel, E., 2005. Horizontal transfer of the high-pathogenicity island of *Yersinia pseudotuberculosis*. J. Bacteriol. 187, 3352–3358.

Ljungberg, P., Valtonen, M., Harjola, V.P., Kaukoranta-Tolvanen, S.S., Vaara, M., 1995. Report of four cases of *Yersinia pseudotuberculosis* septicemia and a literature review. Eur. J. Clin. Microbiol. Infect. Dis. 14, 804–810.

Mair, N.S., 1965. Sources and serological classification of 177 strains of *Pasteurella pseudotuberculosis* isolated in Great Britain. J. Pathol. Bacteriol. 90, 275–278.

Merhej, V., Adekambi, T., Pagnier, I., Raoult, D., Drancourt, M., 2008. *Yersinia massiliensis* sp. nov., isolated from fresh water. Int. J. Syst. Evol. Microbiol. 58, 779–784.

Morelli, G., Song, Y., Mazzoni, C.J., Eppinger, M., Roumagnac, P., Wagner, D.M., Feldkamp, M., Kusecek, B., Vogler, A.J., Li, Y., Cui, Y., Thomson, N.R., Jombart, T., Leblois, R., Lichtner, P., Rahalison, L., Petersen, J.M., Balloux, F., Keim, P., Wirth, T., Ravel, J., Yang, R., Carniel, E., Achtman, M., 2010. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. 42, 1140–1143.

Murros-Kontiainen, A., Fredriksson-Ahomaa, M., Korkeala, H., Johansson, P., Rahkila, R., Bjorkroth, J., 2011. *Yersinia nurmii* sp. nov. Int. J. Syst. Evol. Microbiol. 61, 2368–2372.

Nierman, W.C., DeShazer, D., Kim, H.S., Tettelin, H., Nelson, K.E., Feldblyum, T., Ulrich, R.L., Ronning, C.M., Brinkac, L.M., Daugherty, S.C., Davidsen, T.D., Deboy, R.T., Dimitrov, G., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Khouri, H., Kolonay, J.F., Madupu, R., Mohammoud, Y., Nelson, W.C., Radune, D., Romero, C.M., Sarria, S., Selengut, J., Shamblin, C., Sullivan, S.A., White, O., Yu, Y., Zafar, N., Zhou, L., Fraser, C.M., 2004. Structural flexibility in the *Burkholderia mallei* genome. Proc. Natl. Acad. Sci. U. S. A. 101, 14246–14251.

Pham, J.N., Bell, S.M., Hardy, M.J., Martin, L., Guiyoule, A., Carniel, E., 1995. Susceptibility to beta-lactam agents of *Yersinia enterocolitica* biotype 4, serotype O3 isolated in various parts of the world. J. Med. Microbiol. 43, 9–13.

Pham, J.N., Bell, S.M., Lanzarone, J.Y., 1991. A study of the beta-lactamases of 100 clinical isolates of *Yersinia enterocolitica*. J. Antimicrob. Chemother. 28, 19–24.

Pham, J.N., Bell, S.M., Martin, L., Carniel, E., 2000. The beta-lactamases and beta-lactam antibiotic susceptibility of *Yersinia enterocolitica*. J. Antimicrob. Chemother. 46, 951–957.

Richter, M., Rossello-Mora, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. 106, 19126–19131.

Sauer, S., Freiwald, A., Maier, T., Kube, M., Reinhardt, R., Kostrzewa, M., Geider, K., 2008. Classification and identification of bacteria by mass spectrometry and computational analysis. PLoS ONE 3, e2843.

Schmid, K., Schmitt, R., 1976. Raffinose metabolism in *Escherichia coli* K12. Purification and properties of a new alpha-galactosidase specified by a transmissible plasmid. Eur. J. Biochem. 67, 95–104.

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

Tsubokura, M., Otsuki, K., Sato, K., Tanaka, M., Hongo, T., Fukushima, H., Maruyama, T., Inoue, M., 1989. Special features of distribution of *Yersinia pseudotuberculosis* in Japan. J. Clin. Microbiol. 27, 790–791.

Uchiyama, T., Miyoshi-Akiyama, T., Kato, H., Fujimaki, W., Imanishi, K., Yan, X.J., 1993. Superantigenic properties of a novel mitogenic substance produced by *Yersinia pseudotuberculosis* isolated from patients manifesting acute and systemic symptoms. J. Immunol. 151, 4407–4413.

Van der Heiden, E., Delmarcelle, M., Lebrun, S., Freichels, R., Brans, A., Vastenavond, C.M., Galleni, M., Joris, B., 2013. A pathway closely related to the (D)-tagatose pathway of gram-negative enterobacteria identified in the gram-positive bacterium *Bacillus licheniformis*. Appl. Environ. Microbiol. 79, 3511–3515.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.