Author/s:
Williamson, DA;Baines, SL;Carter, GP;Da Silva, AG;Ren, X;Sherwood, J;Dufour, M;Schultz, MB;French, NP;Seemann, T;Stinear, TP;Howden, BP

Title:
Genomic insights into a sustained national outbreak of yersinia pseudotuberculosis

Date:
2016-01-01

Citation:
Williamson, D. A., Baines, S. L., Carter, G. P., Da Silva, A. G., Ren, X., Sherwood, J., Dufour, M., Schultz, M. B., French, N. P., Seemann, T., Stinear, T. P. & Howden, B. P. (2016). Genomic insights into a sustained national outbreak of yersinia pseudotuberculosis. Genome Biology and Evolution, 8 (12), pp.3806-3814. https://doi.org/10.1093/gbe/evw285.

Persistent Link:
https://hdl.handle.net/11343/257977

Genomic Insights into a Sustained National Outbreak of *Yersinia pseudotuberculosis*

Deborah A. Williamson[1,2,†], Sarah L. Baines[1,†], Glen P. Carter[1], Anders Gonçalves da Silva[1,2], Xiaoyun Ren[3], Jill Sherwood[3], Muriel Dufour[3], Mark B. Schultz[1,2], Nigel P. French[4], Torsten Seemann[1,5], Timothy P. Stinear[1], and Benjamin P. Howden[1,2]

[1]Doherty Applied Microbial Genomics, Department of Microbiology & Immunology, The University of Melbourne at The Doherty Institute for Infection and Immunity, Melbourne, Australia

[2]Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology & Immunology, The University of Melbourne at The Doherty Institute for Infection and Immunity, Melbourne, Australia

[3]Institute of Environmental Science and Research, Wellington, New Zealand

[4]Infectious Disease Research Centre, Massey University, Palmerston North, New Zealand

[5]Victorian Life Sciences Computation Initiative, The University of Melbourne, Melbourne, Australia

†These authors contributed equally to this work.

*Corresponding author: E-mail: deborah.williamson@unimelb.edu.au.

Accepted: November 28, 2016

Abstract

In 2014, a sustained outbreak of yersiniosis due to *Yersinia pseudotuberculosis* occurred across all major cities in New Zealand (NZ), with a total of 220 laboratory-confirmed cases, representing one of the largest ever reported outbreaks of *Y. pseudotuberculosis*. Here, we performed whole genome sequencing of outbreak-associated isolates to produce the largest population analysis to date of *Y. pseudotuberculosis*, giving us unprecedented capacity to understand the emergence and evolution of the outbreak clone. Multivariate analysis incorporating our genomic and clinical epidemiological data strongly suggested a single point-source contamination of the food chain, with subsequent nationwide distribution of contaminated produce. We additionally uncovered significant diversity in key determinants of virulence, which we speculate may help explain the high morbidity linked to this outbreak.

Key words: yersiniosis, zoonosis, foodborne disease, genomics, epidemiology.

Introduction

In 2014, a sustained outbreak of yersiniosis due to *Yersinia pseudotuberculosis* occurred in New Zealand (NZ) involving all major cities. With a total of 220 laboratory-confirmed cases, this outbreak currently represents one of the largest globally reported outbreaks of human yersiniosis due to *Y. pseudotuberculosis*. Here, by performing a comprehensive genomic analysis of *Y. pseudotuberculosis*, we report the emergence of a novel clone belonging to multi-locus sequence type (ST) 42, and the cause of the 2014 outbreak. Our genomic and clinical epidemiological analyses strongly support a single point-source contamination of the food chain, with subsequent nationwide distribution of contaminated produce. Moreover, within the context of a globally and taxonomically diverse dataset, we demonstrate the evolution of pathogenic *Y. pseudotuberculosis* clones via the acquisition of key virulence factors, and provide additional insight into host-specific adaptation of *Y. pseudotuberculosis*.

Background

Yersinia pseudotuberculosis is a zoonotic pathogen, belonging to the Enterobacteriaceae. It is capable of infecting both human and animal hosts, and has a broad range of domestic and wild animal reservoirs (Fukushima and Gomyoda 1991; Childs-Sanford et al. 2009; Magistrali et al. 2014; Chakraborty et al. 2015). Transmission is by the fecal-oral route, and human infection can occur by the ingestion of contaminated produce or water, or alternatively by direct contact with an infected animal or human (Toma 1986; Fukushima et al. 1989; Jalava et al. 2004; Vincent et al. 2008;

Rimhanen-Finne et al. 2009). Although *Y. pseudotuberculosis* is a less common cause of human gastrointestinal yersiniosis than *Y. enterocolitica*, both are important foodborne pathogens with similar clinical features of infection. These include fever, abdominal pain and diarrhea. In addition, extra-intestinal symptoms such as reactive arthritis and erythema nodosum can occur (Hannu et al. 2003; Jalava et al. 2006), and dissemination of infection to the bloodstream and deep tissues has been described (Kaasch et al. 2012). Classically, identification and typing of *Y. pseudotuberculosis* has been performed by O-antigen serotyping, with 15 major serotypes identified (Bogdanovich et al. 2003). However, because the vast majority of strains isolated from human cases belong to serotypes O:1 and O:3, the utility of serotyping in investigating potential outbreaks of *Y. pseudotuberculosis* is very limited (Bogdanovich et al. 2003; Halkilahti et al. 2013).

A recent detailed phylogenomic analysis of *Yersinia*, including 31 isolates of *Y. pseudotuberculosis*, provided information on the population structure and virulence gene repertoire of the genus (Reuter et al. 2014). Similar to other human pathogenic *Yersinia* species, the pathogenicity of *Y. pseudotuberculosis* is influenced by the presence of several virulence factors; in particular, a 70kb virulence plasmid (pYV) that harbors genes encoding a type III secretion system and associated effector proteins (the *Yersinia* outer proteins (Yops)), the chromosomal attachment and invasion locus (*ail*) (Ch'ng et al. 2011; McNally et al. 2016), and the invasins (*inv*). Additional virulence determinants variably present within the species include a superantigen (the *Y. pseudotuberculosis*-derived mitogen (YPM)), a high-pathogenicity island (HPI) encoding the iron uptake system yersiniabactin (Collyn et al. 2005), and the *Yersinia* adhesion pathogenicity island which harbors a pilin gene cluster (the *pil* locus) (Collyn et al. 2002). Although most infections with *Y. pseudotuberculosis* are thought to be sporadic, previous outbreaks of foodborne infection due to *Y. pseudotuberculosis* have been reported, predominantly in temperate climates (Fukushima et al. 1985; Toma 1986; Jalava et al. 2004; Kangas et al. 2008; Vincent et al. 2008; Chakraborty et al. 2015). In this study, we combine genomic and clinical epidemiological data to shed light on the origins and genetic basis for virulence of an epidemic clone of *Y. pseudotuberculosis* responsible for, to date, one of the largest laboratory-confirmed outbreaks of *Y. pseudotuberculosis*.

Methods

Setting and Case Definitions

New Zealand is an island nation in the Southwest Pacific, with a population of approximately 4.7 million. Under the New Zealand Health Act of 1956, medical practitioners are legally required to report suspected or proven cases of notifiable diseases, including human yersiniosis. During the 2014 outbreak, confirmed cases of *Y. pseudotuberculosis* infections were epidemiologically defined as patients who had a clinically compatible illness between August 14 and November 1, 2014 (e.g., diarrhoea, vomiting, fever, and abdominal pain) and laboratory confirmation of disease, either by direct isolation of *Y. pseudotuberculosis* from a clinical specimen ($n = 190$), or by serological evidence of recent exposure to *Y. pseudotuberculosis* ($n = 30$).

Bacterial Isolates and Whole Genome Sequencing

A total of 126 isolates of *Y. pseudotuberculosis* recovered in NZ underwent whole genome sequencing. This collection included 93 isolates recovered from patients during the 2014 outbreak, 28 historic isolates recovered between 2004 and July 2014, and five isolates recovered in August 2015. This collection was supplemented with the publicly available sequence reads and partial or complete genomes of 96 isolates; 84 representing the *Y. pseudotuberculosis* complex, and 12 representing other species of *Yersinia* (supplementary dataset S1, Supplementary Material online). Genomic DNA was extracted from isolates using the JANUS automated workstation (PerkinElmer). DNA libraries were created using the Nextera XT DNA sample preparation kit (Illumina, San Diego, CA), and sequenced on the NextSeq (Illumina, San Diego, CA) using 2 × 150 bp chemistry. In addition, one outbreak isolate (YP4713) was sequenced on the RS-II (Pacific Biosciences) using P6-C4 chemistry. Sequence data generated in this study has been deposited in the European Nucleotide Archive (ENA) under study accession PRJEB14046.

Genomic Analyses

The methodology for all genomic analyses performed in this study are described in detail in supplementary text S1, Supplementary Material online, and only briefly outlined here. Read-mapping and variant calling was performed using Snippy v2.5 (http://github.com/tseemann/snippy), and a maximum likelihood (ML) phylogeny was constructed from core genome single nucleotide polymorphisms (SNPs) using PhyML v3.0 (Guindon et al. 2010) run under a HKY85 model of nucleotide substitution. Recombination detection was performed using ClonalFrameML v1.7 (Didelot and Wilson 2015), and Bayesian phylogenetic analysis using BEAST 2 v2.3.2 (Bouckaert et al. 2014). *De novo* assembly was performed using SPAdes v3.7.1 (Bankevich et al. 2012), and a pan-genome constructed with Roary v3.6.2 (Page et al. 2015), using BLASTp with a minimum percentage identity of 90% for clustering orthologs. Discriminant Analysis of Principle Components (DAPC), as described by Jombart et al. (2010) was implemented in R v3.2.3 using the package *adegenet* (Jombart and Ahmed 2011).

Results and Discussion

Outbreak Overview

A total of 220 laboratory-confirmed cases of *Y. pseudotuberculosis* were reported across both the North and South islands of NZ. A typical epidemic curve was observed (supplementary fig. S1, Supplementary Material online), with the peak number of cases occurring in mid-September 2014. Of the 220 confirmed cases, 72 (33%) individuals were hospitalized with presumed complications of *Y. pseudotuberculosis* infection, representing considerable morbidity.

Although no *Y. pseudotuberculosis* isolates were recovered from food produce, the results of a previously conducted "infield" case–control study implicated several potential foodstuffs, most notably, fresh carrots and lettuce (Institute of Environmental Science and Research 2014). A prior large outbreak of *Y. pseudotuberculosis* in schoolchildren in Finland was strongly associated with raw carrot consumption (Kangas et al. 2008), and these authors hypothesized that prolonged cold storage of contaminated carrots may have allowed the survival and growth of *Y. pseudotuberculosis*, which similar to *Y. enterocolitica*, is psychrotrophic and can multiply in refrigerated food (Palonen et al. 2010). It is plausible that the relatively sustained nature of the NZ outbreak was partially due to storage of contaminated produce in domestic refrigerators, with associated temporal differences in individual consumption. Moreover, the incubation period of *Y. pseudotuberculosis* varies widely from 3 to 21 days, which may also have contributed to the outbreak duration.

Complete Genome Sequence of *Y. pseudotuberculosis* Outbreak Isolate YP4713

To provide genomic insight into the putative virulence of the outbreak clone, and create a suitable reference genome for comparative genomic analyses, a representative isolate from the outbreak was fully sequenced and assembled. Reference *Y. pseudotuberculosis* YP4713 was recovered at the start of the outbreak from the feces of a patient with the earliest reported date of symptom onset. The genome of YP4713 comprised a 4,724,276 bp circular chromosome (47.6% GC content) and a single 68,815 bp circular plasmid (44.7% GC content), which displayed 97.3% nucleotide sequence identity to the pYV virulence plasmid previously described in *Y. pseudotuberculosis* IP32953 (GenBank accession CP009711). A total of 4,076 predicted protein coding-regions (CDS) were identified in the chromosome, and 82 CDS in the plasmid. *In silico* multi-locus sequence typing (MLST) confirmed that this isolate belonged to ST42.

A DNA:DNA comparison of the chromosomes of all complete *Y. pseudotuberculosis* genomes in GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and representatives of other members of the *Yersinia* genus to YP4713 is illustrated in figure 1. There was strong conservation of the genome between members of the *Y. pseudotuberculosis* complex (*Y. pseudotuberculosis*, *Y. pestis*, and *Y. similis*) compared with other *Yersinia* species, with > 90% nucleotide sequence identity observed across the majority of the YP4713 genome. Variable regions were predominantly associated with presumptive phages and the HPI locus. Only one mobile element appeared to be unique to YP4713 in this comparison; a presumptive phage located at nucleotide position 3,572,758 to 3,623,567 (ENA chromosome accession LT596221). Apart from genes encoding phage machinery, this novel phage region encompassed a locus involved in a methionine salvage pathway (*mtnADCB*). However, the *mtnADCB* locus was also detected in all analysed *Y. pseudotuberculosis* isolates (supplementary fig. S2, Supplementary Material online); only the phage backbone was unique to NZ isolates of *Y. pseudotuberculosis*.

Phylogenomic context of the NZ *Y. pseudotuberculosis* Outbreak

A phylogenetic model of the global population structure of the *Y. pseudotuberculosis* complex was re-constructed with 209 isolates (178 *Y. pseudotuberculosis*, 26 *Y. pestis*, and five *Y. similis*). A total of 366,289 variant nucleotide positions were detected, including SNPs, insertions and deletions. Of these, 175,185 were SNPs that resided within the core genome (representing 62.5% of YP4713). These core genome SNPs were used to construct the ML tree illustrated in figure 2*A*.

Similar to a previous study (Reuter et al. 2014), our phylogenetic model of the *Y. pseudotuberculosis* complex demonstrated that *Y. pseudotuberculosis* and *Y. similis* were the most divergent species within the complex while *Y. pestis* was most closely related to, and is a presumed descendant of ancestral *Y. pseudotuberculosis* (fig. 2*A*). At a species level, distinct clusters of *Y. pseudotuberculosis* were evident within the ML tree and typically grouped together in ST-specific clades. Of the 93 outbreak-associated isolates, 87 isolates belonged to ST42. The six additional isolates that were epidemiologically classified as outbreak-associated, but were not ST42, belonged to three STs: ST9 ($n = 2$), ST14 ($n = 2$), and ST43 ($n = 2$).

Within the ST42 clade, there were three main sub-clades: (i) the ST42 NZ outbreak clade ($n = 82$), with a maximum core genome SNP distance of only two SNPs to reference YP4713, (ii) a circulating non-outbreak ST42 clade from NZ comprising isolates recovered from all time periods ($n = 12$), with SNP distances of 55–66 SNPs, and (iii) ST42 strains previously described from international studies ($n = 7$), with SNP distances of 111–2,145 SNPs (fig. 2*B*).

The lack of genetic diversity observed within the core genome of the ST42 NZ outbreak clade is consistent with the epidemiological hypothesis of a point-source contamination of the food chain. Furthermore, one of the five NZ isolates identified in August 2015 (NZYP8105) also clustered with the

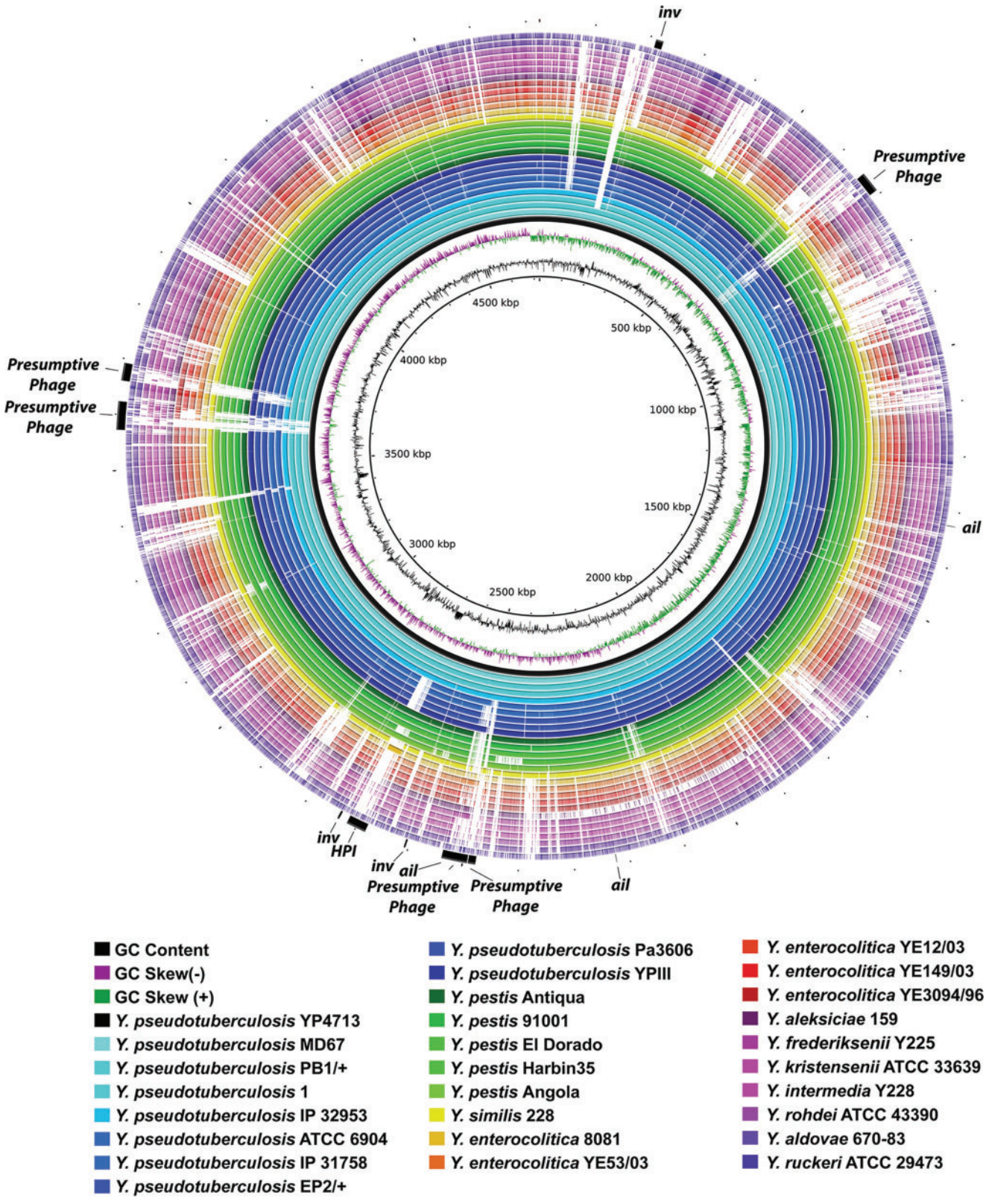


Fig. 1.—DNA:DNA comparison of representative complete genomes of the *Yersinia* genus, relative to YP4713, illustrating a high level of sequence conservation amongst the *Yersinia pseudotuberculosis* complex compared with other *Yersinia* species.

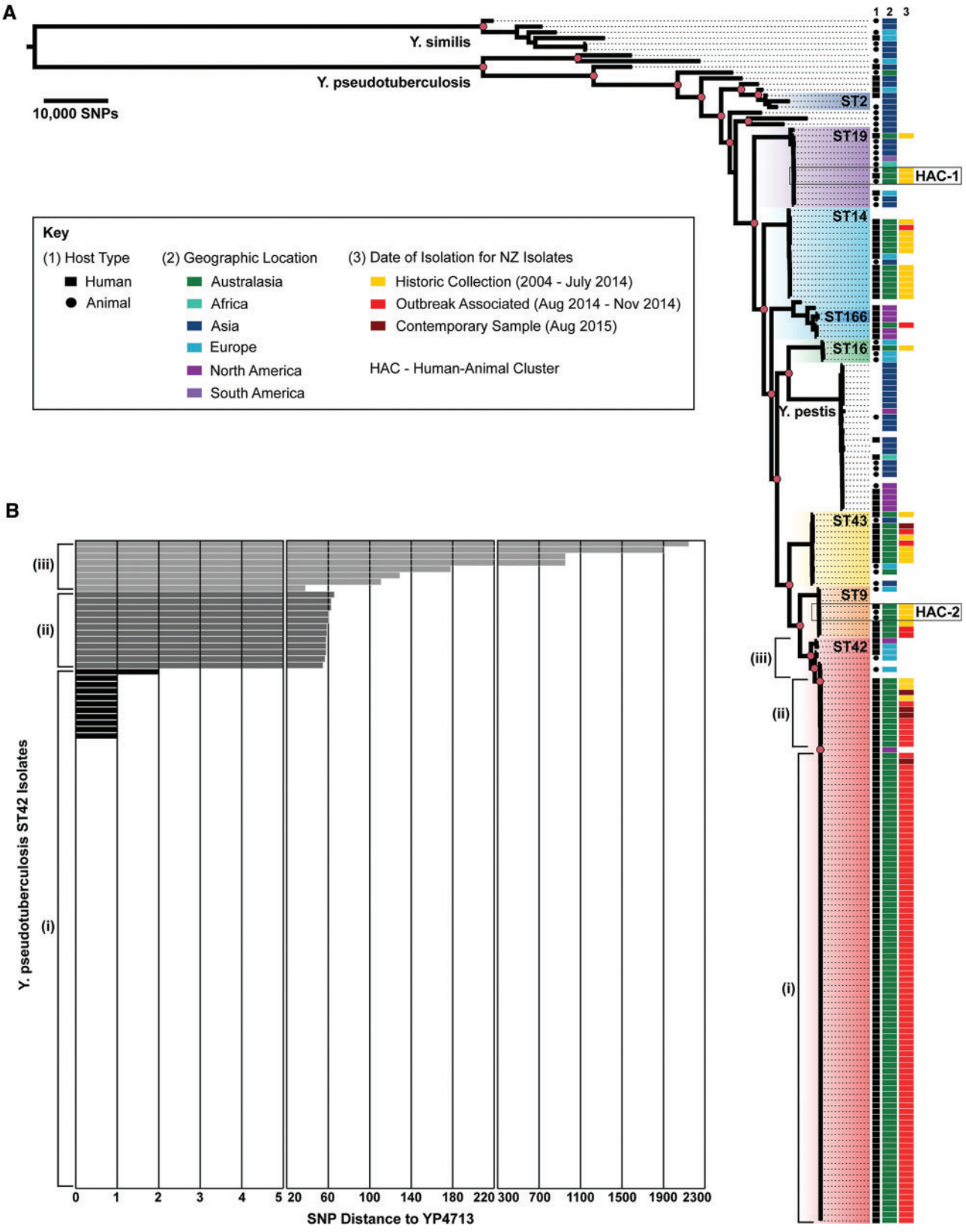


Fig. 2.—(*A*) Population structure of the *Yersinia pseudotuberculosis* complex. The maximum likelihood tree, constructed from 175,185 core genome SNPs, illustrates the population structure of a representative global population of the *Y. pseudotuberculosis* complex. Demographic information for each

Table 1
Evolutionary Rates of ST-Specific Clades of *Yersinia pseudotuberculosis*

Clade (n)	Nucleotide Substitution Rate (s^{-s-y}) Median (95% HPD Range)	Emergence Date in NZ Median (95% HPD Range)
ST19 (10)	3.87×10^{-7} (3.03×10^{-7}–4.64×10^{-7})	1860 (1832–1883)
ST14 (12)	8.67×10^{-7} (4.70×10^{-7}–1.31×10^{-6})	1909 (1846–1951)
ST43 (9)	5.63×10^{-7} (3.93×10^{-7}–7.75×10^{-7})	1886 (1838–1919)
ST9 (6)	2.01×10^{-6} (1.44×10^{-6}–2.63×10^{-6})	2002 (2000–2003)
ST42 (95)	3.57×10^{-7} (2.25×10^{-7}–4.96×10^{-7})	1978 (1964–1991)
ST42 Outbreak clade	–	2013 (2012–2014)

HPD = highest posterior density.

ST42 NZ outbreak clade, indicating that the outbreak clone was still circulating within NZ at least one year after the first outbreak case was reported (fig. 2*A*).

The 28 historic NZ isolates were located throughout the ML tree, and belonged to a variety of STs: ST9 ($n = 4$), ST14 ($n = 11$), ST16 ($n = 1$), ST19 ($n = 4$), ST42 ($n = 3$), and ST43 ($n = 5$). Interestingly, within these historic strains, there were two groups of highly related human and animal isolates (within ST19 and ST43), consistent with a possible zoonotic transmission pathway of *Y. pseudotuberculosis* [annotated as human–animal cluster (HAC) 1 and 2, fig. 2*A*). A maximum of 25 and 26 core genome SNPs were identified between the human and animal isolates in each respective HAC.

At present, there are no routinely used molecular typing methods for *Y. pseudotuberculosis.* PFGE has been used in previous outbreaks of *Y. pseudotuberculosis*, but there are recognized inter-laboratory reproducibility issues with PFGE, and the discriminatory power afforded by PFGE is unclear (Jalava et al. 2004; Halkilahti et al. 2013). Moreover, our genomic data suggests that several cases were misclassified as "outbreak-associated" when only clinical epidemiological data were considered, highlighting the utility of WGS for investigation of disease outbreaks caused by this pathogen. The prospective use of WGS-based surveillance in this outbreak would have enabled rapid classification of cases, and potentially prevented unnecessary investigation by local public health units.

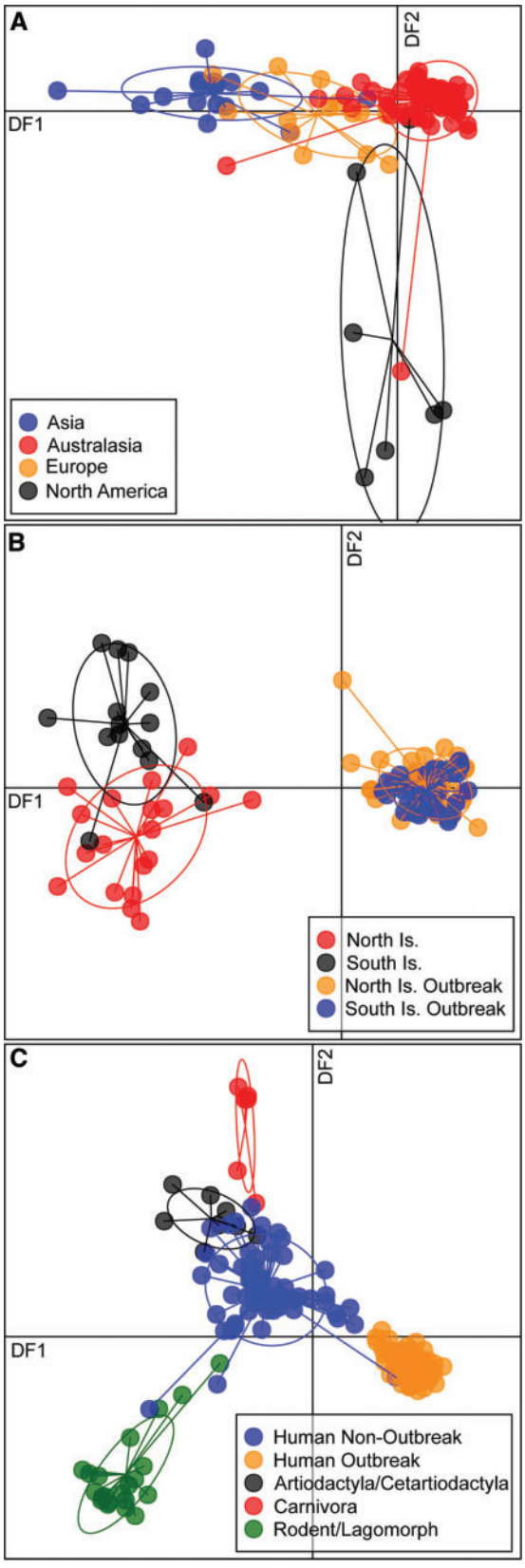
Temporal origins of *Y. pseudotuberculosis* Outbreak Clone

Y. pseudotuberculosis has previously been shown to have a high frequency of recombination compared with single nucleotide mutations (Laukkanen-Ninios et al. 2011). Therefore, to further understand the evolution of the *Y. pseudotuberculosis* complex, we undertook an assessment of recombination within this population. We detected extensive recombination in the predicted ancestors for each of the major ST clades but very few recent events within extant genomes (supplementary fig. S3, Supplementary Material online). No recombination events were detected in isolates of the ST42 NZ outbreak clade or its most recent ancestral node indicating that this form of evolution has not played a major role in the emergence of this new clone (supplementary fig. S3, Supplementary Material online).

A Bayesian framework was employed to assess heterochronicity within the data and predict an emergence date of the ST42 NZ outbreak clone and other major clades in NZ within the *Y. pseuodotuberculosis* complex (supplementary methods, Supplementary Material online). To account for the extensive ancestral recombination detected, clade-specific genome alignments of isolates with temporal information were tested independently. The median nucleotide substitution rates of the different *Y. pseudotuberculosis* clades were similar (table 1), ranging from 2.01×10^{-6} substitution per site per year (s^{-s-y}) in ST9, to 3.57×10^{-7} s^{-s-y} in ST42. The *Y. pestis* clade had a significantly slower rate of 8.18×10^{-8} s^{-s-y} (95% highest posterior density (HPD) range of 1.13×10^{-7} s^{-s-y} to 5.16×10^{-8} s^{-s-y}); however, evolutionary rates in *Y. pestis* are known to vary widely (Cui et al. 2013; Wagner et al. 2014).

From these data, we estimate an emergence date for each *Y. pseudotuberculosis* clade in NZ (table 1). Similar to the evolutionary pattern portrayed in the ML tree, ST19 and ST14 were among the earliest *Y. pseudotuberculosis* clades to emerge in NZ, with the time of their most recent common ancestor (TMRCA) estimated at 1909 and 1860, respectively. Notably, the ST43 clade is also estimated to have emerged around this time (TMRCA = 1886), while the two clades that have descended from ST43 (ST9 and ST42) are more recent, with TMRCAs of 2002 and 1980, respectively. The TMRCA for the ST42 NZ outbreak clone was estimated at 2013 (95% HPD of 2012–2014), suggesting the very recent emergence of this

Fig. 2.—Continued
isolate is provided alongside the phylogeny and indicated in the key. Red circles in the tree indicate clade-nodes with ≥ 70% bootstrap support (support is not shown within clades). (*B*) Histogram illustrating the core genome single nucleotide polymorphism (SNP) distance of sequence type (ST) 42 isolates relative to YP4713.



clone in NZ; a finding consistent with the timing of the outbreak and further supporting a single-source as the cause.

Evolution towards virulence in *Y. pseudotuberculosis*

The 2014 outbreak had a high level of morbidity, with approximately one-third of cases hospitalized with presumed complications of *Y. pseudotuberculosis* infection. Therefore, to explore the pathogenic potential of the ST42 NZ outbreak clade, the sequences of all 209 isolates were searched for genes and loci known to be associated with virulence. Consistent with previous observations (Ch'ng et al. 2011), the presence or absence of key virulence factors within the *Y. pseudotuberculosis* complex correlated broadly with ST groups (supplementary fig. S2, Supplementary Material online). The most variable element was the virulence plasmid pYV. However, this plasmid can be lost through *in vitro* passage which may account for its heterogeneity amongst the various clades (Eppinger et al. 2007). All isolates were found to carry three or more different attachment and invasion (*ail*) and invasin (*inv*) genes. Conversely, the presence of the two pathogenicity islands (HPI and YAPI) split the global phylogeny into two broad groups. Many of the clades and singletons located close to the root of the tree, including *Y. similis*, contained only YAPI, whereas clades located further from the root harbored HPI. These more distal clades, including the ST42 outbreak clade, represent serotype O:1, one of the two serotypes most commonly associated with human disease. This finding suggest that these distal clades (containing *ail, inv*, pYV and HPI) may possibly represent an evolutionary progression of more pathogenic lineages within the *Y. pseudotuberculosis* complex.

Geographic- and host-specific evolution of *Y. pseudotuberculosis*

To further explore the nationwide spread of the outbreak and assess possible sources of this epidemic clone, we performed a Discriminant Analysis of Principal Components (DAPC) focusing on the accessory genome content of the isolates, and complementing the core genome SNP phylogenomic analysis. DAPC is a multivariate technique that attempts to reconstruct hypothesized population subdivisions (typically formed from demographic or phenotypic information) using genomic data.

Fig. 3.—Discriminant analysis of principal components exploring geographic- and host-specific genetic signatures within the accessory genome of *Yersinia pseudotuberculosis* isolates. Each graph represents the attempted genomic reconstruction of demographic groups; (*A*) source continent, (*B*) source region for NZ isolates, and (*C*) host source. All graphs display the two most discriminant functions (DF), and data points are colored based on their assigned demographic group.

When isolates were grouped based on the continent from which they were recovered, broad clusters could be resolved in accessory genome content (fig. 3*A*), and when using a 90% membership threshold (a metric representing the strength of the genomic association to a given group) 90.4% of isolates were correctly assigned to their geographic group (88.9% for Asia, 96.1% for Australasia, 50.0% for Europe, and 71.4% for North America), suggesting there is genomic evidence for the regional evolution of *Y. pseudotuberculosis*.

When DAPC was performed only on *Y. pseudotuberculosis* isolates from NZ, geographic clustering was also observed (fig. 3*B*), with 69.6% of non-outbreak isolates correctly re-assigned to either the North or South Island. Conversely, only 4.9% of the outbreak isolates could be correctly re-assigned to their originating region using a 90% membership threshold. The absence of a genomic signal to support geographic clustering of outbreak isolates is consistent with the clinical epidemiological evidence of a point-source outbreak with nationwide commercial distribution of contaminated produce.

DAPC was also used to identify a potential host reservoir for the outbreak clone, grouping the isolates based on the host from which they were recovered. Distinct clusters could be resolved, (fig. 3C), and 93.9% of isolates were correctly assigned to their host group (84.1% for humans (non-outbreak), 95.2% for rodents and lagomorphs (rabbits), and 100% for all other groups), demonstrating potential genomic evidence for within-host evolution of *Y. pseudotuberculosis*. Among the outbreak isolates, none shared more than 1% membership with other host groups, suggesting that the outbreak clone originated from a source not sampled in this collection. Interestingly, human isolates involved in the human-animal clusters identified in the global phylogeny were assigned back to the respective animal group, highlighting the utility of DAPC in studying potential zoonotic transmission pathways. However, additional broader sampling of animal and environmental sources (ideally contiguous in space and time to human cases) is required to allow more robust inferences of host source.

Conclusions

In this study, we undertook detailed genomic analysis of an extensive national outbreak of *Y. pseudotuberculosis*. Importantly, the high resolution afforded by WGS provided valuable insights into the emergence and spread of the outbreak clone, and allowed us to explore putative temporal, geographic and host associations. Taken together, the combination of our genomic models and the clinical epidemiological evidence strongly supported the hypothesis of a single point-source outbreak, with an as-yet unidentified host reservoir. Although we did not detect any unique genomic determinants of virulence in the outbreak strain, our population-based comparative genomic analysis of a large number of *Y. pseudotuberculosis* isolates provided valuable and novel insights into the distribution of known virulence factors across the complex. We also show how DAPC modeling, combining genomic and epidemiological information, is a potentially powerful approach to explore potential zoonotic transmission pathways and to help in source attribution during outbreak investigations.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank staff in the Enteric Reference Laboratory and Health Intelligence Team at the Institute of Environmental Science and Research, New Zealand. We thank Public Health Units, laboratories and clinicians throughout New Zealand. We also thank the New Zealand Ministry of Health. Sequencing work was partially funded by the New Zealand Ministry of Health. Doherty Applied Microbial Genomics is funded by the Department of Microbiology and Immunology at The University of Melbourne. The Microbiological Diagnostic Unit Public Health Laboratory is funded by the Department of Health and Human Services, Victoria. BPH is funded by a Fellowship from the National health and Medical Research Council (NHMRC), Australia (GNT1105905).

Literature Cited

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bogdanovich TM, et al. 2003. Genetic (sero) typing of *Yersinia pseudotuberculosis*. Adv Exp Med Biol. 529:337–340.

Bouckaert R, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 10(4):e1003537.

Chakraborty A, et al. 2015. The descriptive epidemiology of yersiniosis: a multistate study, 2005-2011. Public Health Rep. 130(3):269–277.

Childs-Sanford SE, et al. 2009. *Yersinia pseudotuberculosis* in a closed colony of Egyptian fruit bats (*Rousettus aegyptiacus*). J Zoo Wildl Med. 40(1):8–14.

Ch'ng SL, et al. 2011. Population structure and evolution of pathogenicity of *Yersinia pseudotuberculosis*. Appl Environ Microbiol. 77(3):768–775.

Collyn F, et al. 2002. *Yersinia pseudotuberculosis* harbors a type IV pilus gene cluster that contributes to pathogenicity. Infect Immun. 70(11):6196–6205.

Collyn F, et al. 2005. Linkage of the horizontally acquired *ypm* and *pil* genes in *Yersinia pseudotuberculosis*. Infect Immun. 73(4):2556–2558.

Cui Y, et al. 2013. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. Proc Natl Acad Sci U S A. 110(2):577–582.

Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol. 11(2):e1004041.

Eppinger M, et al. 2007. The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. PLoS Genet. 3(8):e142.

Fukushima H, et al. 1985. Epidemiological study of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* infections in Shimane Prefecture, Japan. Zentralbl Bakteriol Mikrobiol Hyg B. 180(5–6):515–527.
Fukushima H, et al. 1989. Cat-contaminated environmental substances lead to *Yersinia pseudotuberculosis* infection in children. J Clin Microbiol. 27(12):2706–2709.
Fukushima H, Gomyoda M. 1991. Intestinal carriage of *Yersinia pseudotuberculosis* by wild birds and mammals in Japan. Appl Environ Microbiol. 57(4):1152–1155.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.
Halkilahti J, Haukka K, Siitonen A. 2013. Genotyping of outbreak-associated and sporadic *Yersinia pseudotuberculosis* strains by novel multilocus variable-number tandem repeat analysis (MLVA). J Microbiol Methods 95(2):245–250.
Hannu T, et al. 2003. Reactive arthritis after an outbreak of *Yersinia pseudotuberculosis* serotype O:3 infection. Ann Rheum Dis. 62(9):866–869.
Institute of Environmental Science and Research. 2014. Yersinia pseudotuberculosis outbreak: Results of a case-control study. Available at: http://archive.mpi.govt.nz/portals/2/Documents/food/yersinia/yersinia-pseudotuberculosis-outbreak-results-case-control-study-08102014.pdf.
Jalava K, et al. 2004. Multiple outbreaks of *Yersinia pseudotuberculosis* infections in Finland. J Clin Microbiol. 42(6):2789–2791.
Jalava K, et al. 2006. An outbreak of gastrointestinal illness and erythema nodosum from grated carrots contaminated with *Yersinia pseudotuberculosis*. J Infect Dis. 194(9):1209–1216.
Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21):3070–3071.
Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11:94.
Kaasch AJ, et al. 2012. *Yersinia pseudotuberculosis* bloodstream infection and septic arthritis: case report and review of the literature. Infection 40(2):185–190.
Kangas S, et al. 2008. *Yersinia pseudotuberculosis* O:1 traced to raw carrots, Finland. Emerg Infect Dis. 14(12):1959–1961.
Laukkanen-Ninios R, et al. 2011. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. Environ Microbiol. 13(12):3114–3127.
Magistrali CF, et al. 2014. Atypical *Yersinia pseudotuberculosis* serotype O:3 isolated from hunted wild boars in Italy. Vet Microbiol. 171(1–2):227–231.
McNally A, et al. 2016. 'Add, stir and reduce': *Yersinia* spp. as model bacteria for pathogen evolution. Nat Rev Microbiol. 14(3):177–190.
Page AJ, et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31(22):3691–3693.
Palonen E, Lindstrom M, Korkeala H. 2010. Adaptation of enteropathogenic *Yersinia* to low growth temperature. Crit Rev Microbiol. 36(1):54–67.
Reuter S, et al. 2014. Parallel independent evolution of pathogenicity within the genus *Yersinia*. Proc Natl Acad Sci U S A. 111(18):6768–6773.
Rimhanen-Finne R, et al. 2009. *Yersinia pseudotuberculosis* causing a large outbreak associated with carrots in Finland, 2006. Epidemiol Infect. 137: 342.
Toma S. 1986. Human and nonhuman infections caused by *Yersinia pseudotuberculosis* in Canada from 1962 to 1985. J Clin Microbiol. 24(3):465–466.
Vincent P, et al. 2008. Sudden onset of pseudotuberculosis in humans, France, 2004-05. Emerg Infect Dis. 14(7):1119–1122.
Wagner DM, et al. 2014. *Yersinia pestis* and the plague of Justinian 541-543 AD: a genomic analysis. Lancet Infect Dis. 14(4):319–326.

Associate editor: Ruth Hershberg