Microbiology (2002), **148**, 1687–1698 Printed in Great Britain

Genome plasticity in *Yersinia pestis*

Lyndsay Radnedge,[1] Peter G. Agron,[1] Patricia L. Worsham[2] and Gary L. Andersen[1]

Author for correspondence: Gary L. Andersen. Tel: +1 925 423 2525. Fax: +1 925 422 2282. e-mail: Andersen2@LLNL.GOV

[1] Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, L-441, 7000 East Avenue, Livermore, CA 94550, USA

[2] United States Army Research Institute of Infectious Diseases, Fort Detrick, MD 21702, USA

Yersinia pestis*, the causative agent of bubonic plague, emerged recently (<20000 years ago) as a clone of *Yersinia pseudotuberculosis*. There is scant evidence of genome diversity in *Y. pestis*, although it is possible to differentiate three biovars (antiqua, mediaevalis or orientalis) based on two biochemical tests. There are a few examples of restriction fragment length polymorphisms (RFLPs) within *Y. pestis*; however, their genetic basis is poorly understood. In this study, six difference regions (DFRs) were identified in *Y. pestis*, by using subtractive hybridization, which ranged from 4·6 to 19 kb in size. Four of the DFRs are flanked by insertion sequences, and their sequences show similarity to bacterial genes encoding proteins for flagellar synthesis, ABC transport, insect toxicity and bacteriophage functions. The presence or absence of these DFRs (termed the DFR profile) was demonstrated in 78 geographically diverse strains of *Y. pestis*. Significant genome plasticity was observed among these strains and suggests the acquisition and deletion of these DNA regions during the recent evolution of *Y. pestis*. *Y. pestis* biovar orientalis possesses DFR profiles that are different from antiqua and mediaevalis biovars, reflecting the recent origins of this biovar. Whereas some DFR profiles are specific for antiqua and mediaevalis, some DFR profiles are shared by both biovars. Furthermore, the progenitor of *Y. pestis*, *Y. pseudotuberculosis* (an enteric pathogen), possesses its own DFR profile. The DFR profiles detailed here demonstrate genome plasticity within *Y. pestis*, and they imply evolutionary relationships among the three biovars of *Y. pestis*, as well as between *Y. pestis* and *Y. pseudotuberculosis*.

Keywords: bacterial genome, comparative genomics, subtractive hybridization, *Yersinia pseudotuberculosis*

INTRODUCTION

Yersinia pestis, the causative agent of bubonic plague, possesses a genome that is highly conserved among different isolates. *Y. pestis* is predominantly a pathogen of rodents, but it is transmitted into the bloodstream of humans via flea bites. *Y. pestis* has been classified into three biovars (antiqua, mediaevalis and orientalis), based on the ability of strains to ferment glycerol and to reduce nitrate. It has been proposed that antiqua strains are derived from those responsible for the plague pandemic of the sixth century and that mediaevalis strains are responsible for the Black Death of the 14th century; the current pandemic has been associated with the most-recently evolved biovar, orientalis (Devignat, 1951).

Several studies have demonstrated a lack of diversity between different strains of *Y. pestis*. Different strains cannot be classified by differences in serotypes or in phagetypes and, at the genetic level, there are only two reported sequence polymorphisms in seven loci examined in 58 strains (Achtman *et al.*, 1999; Adair *et al.*, 2000). Moreover, the genome of *Y. pestis* is highly related to that of *Yersinia pseudotuberculosis* (an enteric pathogen transmitted by the faecal–oral route), based on hybridization studies (Bercovier *et al.*, 1980) and 16S rRNA sequences (Trebesius *et al.*, 1998). These observations led to the proposal that *Y. pestis* is a recent, highly uniform clone of *Y. pseudotuberculosis* that arose

Abbreviations: DFR, difference region; RFLP, restriction fragment length polymorphism; SSH, suppression subtractive hybridization.

The GenBank accession numbers for the sequences reported in this paper can be found in Table 1; the GenBank accession number for DFR4 is AF426171.

Table 1. Characteristics of DFRs on the *Y. pestis* CO92 genome

DFR	Tester*	Driver*	CO92†	Accession nos	Size (bp)	
					Minimum	Maximum
DFR1	O	A/M	+	AF333796–AF333797 AF334808	12000	13669
DFR2	A/M	O	+	AF333798–AF333801	12148	15299
DFR3	A/O	M	+	AF333802–AF333804	10128	12198
DFR4	A	M/O	−	AF333805–AF333807 AF337898–AF337900	15603	15603
DFR5	O	A/M	+	AF333808–AF333810	15790	19062
DFR6	A/O	M	+	AF333811–AF333813	3252	4603

* O, orientalis; M, mediaevalis; A, antiqua. Biovars from which the tester and driver DNA were prepared.

† +, The DFR is mapped to the *Y. pestis* CO92 genome; −, the DFR is absent from the *Y. pestis* CO92 genome.

between 1500 and 20000 years ago (Achtman *et al.*, 1999). Some genome plasticity was demonstrated by restriction fragment length polymorphism (RFLP) analysis of IS100 locations, enabling the construction of a phylogenetic tree in which the different biovars of *Y. pestis* are clustered together (Achtman *et al.*, 1999; McDonough & Hare, 1997). Similarly, PFGE analysis of *Spe*I fragments of *Y. pestis* genomic DNA shows no variability within biovars, but some variation is seen between biovars (Lucier & Brubaker, 1992). Another study of variable number tandem repeats (VNTRs), which are regions with potentially high variability, has shown these regions to be a useful tool for strain discrimination, but, in contrast to RFLP analysis, the VNTR alleles were not biovar specific (Adair *et al.*, 2000). Small genomic differences between different isolates of *Y. pestis* can also be demonstrated by ribotyping (Guiyoule *et al.*, 1994, 1997).

As the nucleotide sequences of more bacterial genomes are uncovered, it is becoming increasingly apparent that they are composed of DNA sequence mosaics, some of which are acquired by lateral (or horizontal) gene transfer (Jain *et al.*, 1999; Lawrence & Ochman, 1998; Ochman *et al.*, 2000; Perna *et al.*, 2001). Large genomic differences that have arisen from lateral gene transfer events commonly originate from mobile genetic elements, such as transposons, insertion elements or bacteriophage integration. The excision and acquisition of large genomic fragments quickly generate new strain variants, with the balance between the two processes maintaining the genome size (Lawrence & Ochman, 1998). There is now evidence for genome plasticity in *Y. pestis* CO92, since three large genomic rearrangements are seen – one translocation and two inversions – which apparently arise during growth of the organism (Parkhill *et al.*, 2001). Furthermore, a large unstable fragment has been characterized – the 102 kb *pgm* locus – within which lies the high-pathogenicity island (Buchrieser *et al.*, 1999). Instability of the *pgm* locus is mediated by a recombination event between two copies of IS100 that flank the locus (Fetherston *et al.*, 1992). Insertion elements are also found to be unusually abundant in the genome of *Y. pestis* CO92 and are found at the boundaries of the three variable regions in this strain (Parkhill *et al.*, 2001).

In this study, suppression subtractive hybridization (SSH) (Diatchenko *et al.*, 1996; Akopyants *et al.*, 1998) was used to identify genomic differences between different strains of *Y. pestis*. Here we report the identification of six difference regions (DFRs) that are variable between different isolates of *Y. pestis*, four of which contain insertion elements. Five of the DFRs are present in the genome of *Y. pestis* CO92 (biovar orientalis) (Parkhill *et al.*, 2001), but are absent in some isolates. One DFR is absent in *Y. pestis* CO92, but is present in other strains, providing evidence that both the acquisition and excision of large genomic segments in *Y. pestis* have contributed to the evolution of its genome.

METHODS

Bacterial strains. The bacterial strains examined in the comparisons using SSH were *Y. pestis* D14 Salazar (orientalis), *Y. pestis* D15 Yokohama (antiqua), *Y. pestis* D27 KIM (mediaevalis), *Y. pestis* D46 KIM10 (mediaevalis), *Y. pestis* Antiqua (antiqua) and *Y. pestis* CO92 (orientalis) (Table 1). The first four strains were generous gifts from Dr Robert Brubaker (Michigan State University) and the last two strains were from the USAMRIID culture collection. DNA from the additional strains tested for difference products is listed in Table 2, and this DNA was generously provided by the following colleagues: strains 1, 2, 11, 14, 30, 31, 39, 45, 46, 56, 59 and 72 were from the CDC collection, courtesy of Dr May Chu; strains 5–10 were from the Dept of Health Services, Berkeley, CA, courtesy of Dr Will Probert; strains 18, 19, 20, 21, 41, 42, 51, 60, 61, 68, 70 and 79 were courtesy of Dr Robert Brubaker. Strains 80–82 were purchased from the American Type Culture Collection (ATCC), Manassas, VA. DNA from the remaining strains was supplied by the USAMRIID collection.

SSH. The method of SSH used here is essentially as described by Akopyants *et al.* (1998). SSH identifies DNA sequences (termed difference products) that are specific to one genome (designated the tester) and which are absent in another genome (designated the driver). Briefly, genomic DNA from the tester and driver are cut with a frequently cutting restriction enzyme

Table 2. The appearance of each DFR in 78 strains of *Y. pestis*, and in four isolates of *Y. pseudotuberculosis*

Strain no. and name		Origin	Biotype*	DFR1	DFR2	DFR3	DFR4	DFR5	DFR6	Profile†
Y. pestis										
1	15-19	FSU‡	Ort	+	+	+	−	+	+	A
2	1634	Vietnam	Ort	+	+	+	−	+	+	A
3	195 P	India	Ort	+	+	+	−	+	+	A
4	770 CO3311	Indonesia	Ort	+	+	+	−	+	+	A
5	90A 0414	USA	Ort	+	+	+	−	+	+	A
6	90A 5268	USA	Ort	+	+	+	−	+	+	A
7	91A 2647	USA	Ort	+	+	+	−	+	+	A
8	92A 2909	USA	Ort	+	+	+	−	+	+	A
9	93A 3795	USA	Ort	+	+	+	−	+	+	A
10	96A 8520	USA	Ort	+	+	+	−	+	+	A
11	A1122	USA	Ort	+	+	+	−	+	+	A
12	A12 (2)§	USA	Ort	+	+	+	−	+	+	A
13	Antigua R	FSU	Ort	+	+	+	−	+	+	A
14	AZ94-0666	USA	Ort	+	+	+	−	+	+	A
15	Cambodia	Cambodia	Ort	+	+	+	−	+	+	A
16	CO92	USA	Ort	+	+	+	−	+	+	A
17	CO92 derivatives (16)§	USA	Ort	+	+	+	−	+	+	A
18	D5 Java	Indonesia	Ort	+	+	+	−	+	+	A
19	D73 M23	Unknown	Ort	+	+	+	−	+	+	A
20	D77 MP6	Unknown	Ort	+	+	+	−	+	+	A
21	D88 Java	Indonesia	Ort	+	+	+	−	+	+	A
22	Dodson	USA	Ort	+	+	+	−	+	+	A
23	Exu#3	Brazil	Ort	+	+	+	−	+	+	A
24	I18	India	Ort	+	+	+	−	+	+	A
25	Jaffa	Israel	Ort	+	+	+	−	+	+	A
26	Java 9	Indonesia	Ort	+	+	+	−	+	+	A
27	K119-66	Vietnam	Ort	+	+	+	−	+	+	A
28	LaPaz	Bolivia	Ort	+	+	+	−	+	+	A
29	P1178VN	Vietnam	Ort	+	+	+	−	+	+	A
30	PB6 India	India	Ort	+	+	+	−	+	+	A
31	PEXU2	Brazil	Ort	+	+	+	−	+	+	A
32	R216-66	Vietnam	Ort	+	+	+	−	+	+	A
33	RatB PBM23	Burma	Ort	+	+	+	−	+	+	A
34	RFPBM-19	Burma	Ort	+	+	+	−	+	+	A
35	SC 1 Alexander	USA	Ort	+	+	+	−	+	+	A
36	South Park	USA	Ort	+	+	+	−	+	+	A
37	YP1-RR5	India	Ort	+	+	+	−	+	+	A
38	Yreka	USA	Ort	+	+	+	−	+	+	A
39	ZE94-2122	Zimbabwe	Ort	+	+	+	−	+	+	A
40	590/Exu#13	Brazil	Ort	+	+	−	−	+	+	B
41	D20/Dodson	USA	Ort	+	+	−	−	+	+	B
42	D59/EV76H	Madagascar	Ort	+	+	−	−	+	+	B
43	Eyc 1962/EV76 LVS	Madagascar	Ort	+	+	−	−	+	+	B
44	H3	Madagascar	Ort	+	+	−	−	+	+	B
45	Namibia F361/66	Namibia	Ort	+	+	−	−	+	+	B
46	NM99-0030	USA	Ort	+	+	−	−	+	+	B
47	Russian Plague Vaccine	Unknown	Ort	+	+	−	−	+	+	B
48	SEV	FSU	Ort	+	+	−	−	+	+	B
49	Stavropol	FSU	Ort	+	+	−	−	+	+	B
50	586 Exu#9	Brazil	Ort	+	−	+	−	+	+	C
51	D14/Salazar	Bolivia	Ort	+	−	+	−	+	+	C

Table 2 (*cont.*)

Strain no. and name	Origin	Biotype*	DFR1	DFR2	DFR3	DFR4	DFR5	DFR6	Profile†
52 Pestoides A	FSU	Med	+	+	+	+	−	+	D
53 Pestoides B	FSU	Med	+	+	+	+	−	+	D
54 Pestoides C	FSU	Med	+	+	−	+	−	+	E
55 Pestoides D	FSU	Med	+	+	−	+	−	+	E
56 Harbin 35	Manchuria	Med	−	+	+	+	−	+	F
57 Nicholisk 41	Manchuria	Med	−	+	−	+	−	+	G
58 366	Yemen	Med	−	+	+	−	−	−	H
59 Amal	Saudi Arabia	Med	−	+	−	−	−	−	I
60 D1 Iran	Iran	Med	−	+	−	−	−	−	I
61 KIM (5)§	Kurdistan	Med	−	+	−	−	−	−	I
62 PKH-10	Kurdistan	Med	−	+	−	−	−	−	I
63 PKR108	Kurdistan	Med	−	+	−	−	−	−	I
64 PKR158	Kurdistan	Med	−	+	−	−	−	−	I
65 Nicholisk 51	Manchuria	Ant	+	+	+	−	+	+	A
66 Antiqua	Congo	Ant	+	+	+	+	−	+	D
67 A16	Congo	Ant	+	+	−	+	−	+	E
68 D15 Yokohama	Japan	Ant	−	+	+	+	−	+	F
69 Yeo 154 Yokohama	Japan	Ant	−	+	+	+	−	+	F
70 D94 Kuma	Manchuria	Ant	−	+	+	+	−	+	F
71 Nairobi	Kenya	Ant	−	+	+	+	−	+	F
72 Nepal 516	Nepal	Ant	−	+	+	+	−	+	F
73 Angola	Angola	Ant	+	+	−	−	−	+	J
74 Pestoides E	FSU	Ant	+	+	+	−	−	+	M
75 Pestoides G	FSU	Ant	+	+	+	−	−	+	M
76 Pestoides F	FSU	Ant	−	+	+	−	−	+	K
77 316	Unknown		+	+	+	−	+	+	A
78 Pestoides J	FSU		−	+	−	−	−	−	I
Y. pseudotuberculosis									
79 PB1	USA	1	+	−	+	+	−	+	L
80 ATCC 6904	Unknown	2	+	−	+	+	−	+	L
81 ATCC 29833	Unknown	1	+	−	+	+	−	+	L
82 ATCC 29910	Unknown	2	+	−	+	+	−	+	L
		Key:	+	+	+	−	+	+	A
			+	+	−	−	+	+	B
			+	−	+	−	+	+	C
			+	+	+	+	−	+	D
			+	+	−	+	−	+	E
			−	+	+	+	−	+	F
			−	+	−	+	−	+	G
			−	+	+	−	−	−	H
			−	+	−	−	−	−	I
			+	+	−	−	−	+	J
			−	+	+	−	−	+	K
			+	−	+	+	−	+	L
			+	+	+	−	−	+	M

* Ort, orientalis; Med, mediaevalis; Ant, antiqua. These abbreviations refer to the *Y. pestis* biovar. The numbers 1 and 2 refer to the *Y. pseudotuberculosis* serotype.

† The DFR profiles are listed in the key and are designated by the appropriate letter (A–M). Each + or − within the profile indicates the presence or absence of the specific DFR in each of the 83 strains tested, as determined by PCR. Each reaction was repeated in triplicate, using three different primer pairs from each DFR.

‡ FSU, former Soviet Union.

§ Numbers in parentheses, given after the strain name, indicate the number of different isolates of that strain that were tested.

(*Rsa*I). After ligation of the tester DNA to oligonucleotide adaptors and hybridization with restricted driver DNA, SSH uses PCR amplification to enrich for unique segments of restricted tester DNA and simultaneously limits non-target amplification by suppression PCR. Difference products were cloned into the pGEMT-Easy vector (Promega) and transformed into 50 µl of competent DH5α cells (Life Technologies). DNA from clones containing putative tester-specific difference products was purified using magnetic beads (Skowronski *et al*., 2000) and sequenced on an ABI 3700 automated sequencer (Applied Biosystems). The resulting data were analysed using the ABI Sequencing Analysis software (version 3.2), and assembled and edited using Phred, Phrap (Ewing & Green, 1998; Ewing *et al*., 1998) and Consed (version 7.0; Gordon *et al*., 1998).

Analysis of tester-specific sequences. Oligonucleotide primers were designed using the putative tester-specific sequences and were supplied by Genosys. The primers were designed either manually or by using Consed (Gordon *et al*., 1998); they had a melting temperature of >60 °C. The primers were initially screened against genomic DNAs prepared from both the tester and the driver. To determine whether a primer pair was tester-specific, 75 pg of the tester and the driver DNA was used as template in PCR reactions using the following parameters: 94 °C (15 s), 65 °C (15 s) and 72 °C (30 s) for 27 cycles. The products were visualized on a 1·5 % agarose gel run in 0·5× TBE. If a PCR product was present when tester DNA was used as template and absent when the driver DNA was used a template then the sequence was designated tester-specific. The confirmed tester-specific oligonucleotides were then used to amplify PCR products from genomic DNA prepared from the strain collection listed in Table 2. A positive control, to test the integrity of the genomic DNA template, was performed using primers (23S Forward, 5′-ctaccttaggaccgttatagttac-3′; 23S Reverse, 5′-gaaggaactaggcaaaatggt-3′) specific for a region of the 23S gene conserved within the *Enterobacteriaceae*. The nucleotide sequences of all of the primers were searched against the *Y. pestis* CO92 sequence, to ensure that each sequence occurred only once within the genome.

BLAST searches using tester-specific DNA sequences were performed via the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST/). ORFs were identified and compared to non-redundant protein sequence databases using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Sequences that were present in the tester were searched against the Sanger Centre *Y. pestis* CO92 database (http://www.sanger.ac.uk/Projects/Y_pestis/blast_server.shtml) (Altschul *et al*., 1997; Parkhill *et al*., 2001) and the preliminary sequence of *Y. pseudotuberculosis* IP32953 (held at http://bbrp.llnl.gov/bbrp/bin/y.pseudotuberculosis_blast).

The sequence context of the difference products that did not map to the *Y. pestis* CO92 genome was determined by the identification of clones in a library of *Y. pestis* D15 Yokohama DNA fragments with an average size of 5 kb. The library was prepared by generating a partial digest of genomic DNA with *Sau*3A I, gel-purifying the fragments that fell within the 4–6 kb size range and then cloning them into the *Bam*HI site of vector pUC9. The library was probed with oligonucleotides designed from the difference product identified as tester-specific.

Nomenclature. The term 'difference product' refers to the tester-specific restriction fragment derived from SSH. The term 'difference region' (DFR) is used to describe the region of the genome to which several closely located difference products map. The term 'DFR profile' refers to the pattern of alleles for each of the six DFRs identified.

RESULTS

Mapping of difference products to DFRs

Subtractions with various strains of *Y. pestis* identified a collection of clones whose nucleotide sequences could be aligned with that of *Y. pestis* CO92. Sixteen such clones were found to map to five DFRs (DFR1, DFR2, DFR3, DFR5 and DFR6) that were distributed throughout the *Y. pestis* CO92 genome (Fig. 1A). The sixth DFR (DFR4) was absent from *Y. pestis* CO92, and was mapped immediately downstream of DFR3 (Fig. 1B; see below). Difference products within the *pgm* locus of *Y. pestis* were identified by BLAST analyses and were discarded (data not shown).

The boundaries of the DFRs that mapped to *Y. pestis* CO92 were determined by designing oligonucleotides based on the *Y. pestis* CO92 sequence upstream and downstream of the locations of the difference products. These oligonucleotides were used to amplify PCR products from templates prepared from the tester, the driver and *Y. pestis* CO92. This approach assumed that the tester sequence was identical to that of *Y. pestis* CO92. Indeed, the tester DNA was always reported as positive in these amplifications. No products were seen when the nucleotide sequence of the driver DNA differed from that of *Y. pestis* CO92. It is highly unlikely that false-negative results would be seen, given the clonal nature of the *Y. pestis* genome (Achtman *et al*., 1999; Adair *et al*., 2000). Amplification products were seen when the nucleotide sequences of the driver and *Y. pestis* CO92 resumed collinearity. The borders can thus be crudely, but rapidly, mapped to within approximately 1 kb, based on their location between the last negative result and the first positive PCR result, when DNA from the driver is used as the template. Although this approach does not provide a precise location for sequence divergence, it can rapidly establish the context of multiple loci within a DFR. The GenBank accession numbers for each difference product are listed in Table 1.

One clone was identified from a subtraction where DNA from *Y. pestis* D15 Yokohama (antiqua) was used as the tester and *Y. pestis* KIMD27 (mediaevalis) was used as the driver which did not map to the *Y. pestis* CO92 genome (difference product no. 2 of DFR4, Fig. 1B). A library of partial *Sau*3A I digests of *Y. pestis* D15 Yokohama DNA was probed with oligonucleotides specific to this difference product. One clone was identified that contained a 5645 bp insert, and this was termed DFR4. The insert sequence was determined, and it showed a 5′ 1180 bp region that overlapped the 3′ region of DFR3, indicating that DFR4 is deleted from *Y. pestis* CO92 immediately downstream of the rightmost copy of IS1541 in DFR3. Overlapping clones from the *Y. pestis* D15 Yokohama library were not identified using

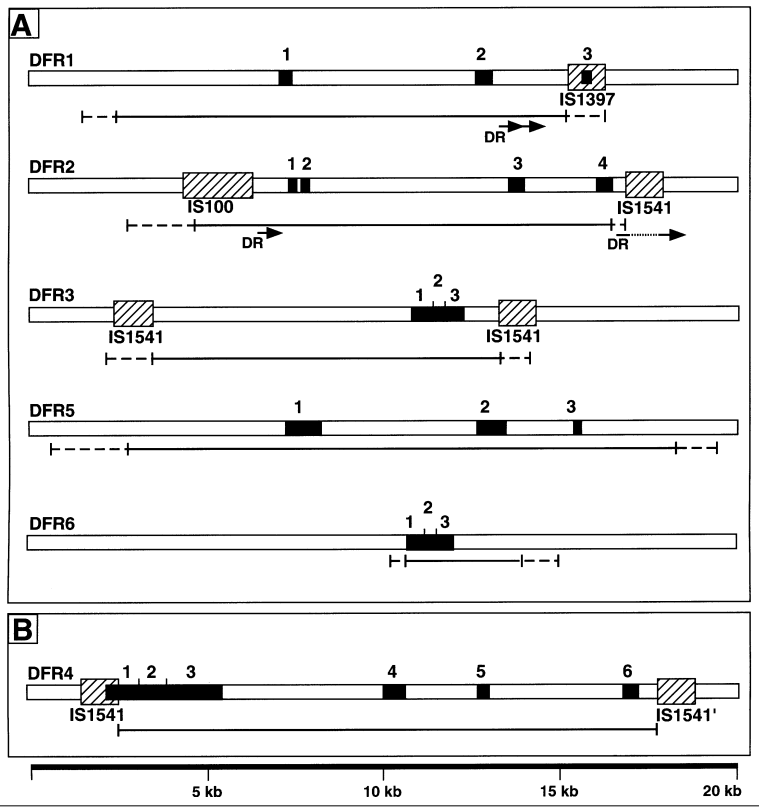

Fig. 1. Distribution of the difference products within each DFR. (A) An illustration of the five DFRs that are present in the *Y. pestis* CO92 genome. The numbered solid boxes show the location of each difference product (clone); short vertical lines delineate the boundaries of contiguous difference products. The locations of insertion elements are shown as expanded hatched boxes. Arrows show the approximate locations of direct repeats (DR). The rightmost DR of DFR2 is interrupted by IS1541 (dotted line). The boundary of each DFR is shown as its maximum size (dashed line) and minimum size (solid line), as determined by PCR (see Methods for explanation). (B) A graphical representation of DFR4, which does not map to *Y. pestis* CO92. The expanded hatched boxes show the locations of IS1541, the rightmost of which, IS1541′, is truncated.

the DFR4 difference product no. 2 specific primer pair, but two additional difference products (nos 1 and 3) were located within the same library clone. Furthermore, three additional difference products (DFR4 nos 4, 5 and 6) were identified that had identical properties, namely they were present in *Y. pestis* D15 Yokohama and were absent from *Y. pestis* KIM D27. These three difference products were also found to be similar to three non-overlapping contigs in the *Y. pseudotuberculosis* database. Primers were designed at the ends of these contigs, to map their relative position in *Y. pestis* D15 Yokohama. Successfully amplified products were cloned, sequenced and assembled into a single contig to complete the sequence of DFR4 (accession no. AF426171). The genome of *Y. pestis* D15 Yokohama resumes collinearity with *Y. pestis* CO92 15·6 kb downstream of IS1541 in DFR4.

Diversity of DFRs in different strains of *Y. pestis*

The presence or absence of each DFR was determined in 78 strains representing all three biovars of *Y. pestis* and many worldwide origins, to examine the possibility of biovar specificity. Four strains of *Y. pseudotuberculosis*, the progenitor of *Y. pestis*, representing two serogroups were also included in the study. PCR using oligonucleotides specific to the multiple difference products (clones) that map within each DFR demonstrated the presence or absence of biovar specificity. In all cases, the PCR results were identical from each difference product within each DFR, with each plus or minus in Table 2 representing identical data for at least nine PCRs.

With the exception of DFR5, there seems to be little correlation between the presence and absence of each individual DFR with the biovars of *Y. pestis*. DFR5 is present in all orientalis biovars, whereas it is absent in all mediaevalis biovars (Table 2). DFR5 is found in only one strain of the antiqua biovar: *Y. pestis* Nicholisk 51 (strain 65, Table 2). Interestingly, this strain represents a variant of the antiqua biovar with restored ability to ferment glycerol (Motin *et al.*, 2002). The PCR data from the *Y. pestis* collection were correlated as profiles designated A–M (Table 2). Profiles A, B and C were found almost exclusively in orientalis strains, and all contained DFR5. The only exception was Nicholisk 51 (antiqua biovar), which fell into profile A. Profiles G, H and I were seen for the mediaevalis biovars; profile I was composed of Central Asian isolates, including five KIM strains (strain 61; Table 2). Profiles J, K and M were seen for strains within the antiqua biovar, whereas profiles D, E and F were found in both the antiqua and mediaevalis biovars. Two strains whose biovar could not be determined, since they could not utilize glycerol or reduce nitrate, are listed in Table 2 (strains 77 and 78). Strain 77 had the same profile (A) as the orientalis biovars, whereas strain 78 had profile I. All four strains of *Y. pseudotuberculosis* fell into the same profile (L), indicating that the presence or absence of the six DFRs is independent of its serovar. Profile L was not found for any strain of *Y. pestis* tested. DFR2 and DFR5 were both absent in *Y. pseudotuberculosis*, and may represent regions acquired since *Y. pestis* emerged as a derivative of *Y. pseudotuberculosis*. DFR2 was seen in all but two strains that originated in South America (biovar orien-

talis, profile C, strains 50 and 51; Table 2). DFR2 was isolated from subtractions using driver DNA from *Y. pestis* D14 Salazar (strain 51; Table 2).

BLAST analysis of the DFRs

The results of BLAST analyses for each DFR are shown in Table 3, and show good concordance with the genome of *Y. pestis* CO92 (Parkhill *et al.*, 2001). DFR1 contains 15 ORFs, eight of which show similarity to genes involved in lateral flagellar synthesis (McCarter & Wright, 1993). Two ORFs in DFR1 encode proteins that are 39 and 59 % identical to ORFA and ORFB of IS1397, respectively (Bachellier *et al.*, 1997). Seven ORFs were identified within DFR2 which show identity to bacteriophage sequences, including a putative replicon, primase and integrase. There are two directly repeated regions of 287 bp, the rightmost of which is interrupted by a copy of IS1541, and which appears to be located at the right boundary of DFR2 (Fig. 1A). A copy of IS100 is seen at the approximate location of the left boundary of DFR2, all of which are features commonly found in pathogenicity islands (Hacker & Kaper, 2000). DFR3 contains 10 ORFs, some of which show identity to enzymes involved in purine salvage pathways, which are encoded by *xapRAB*. There are two directly repeated copies of IS1541 located at the boundaries of DFR3; the rightmost IS1541 interrupts the *xapB* ORF (YPO1172; Parkhill *et al.*, 2001). Downstream there are two ORFs with no similarity to the non-redundant database. These are followed by three ORFs with similarity to *yohI*, penicillin-binding protein 7 and D-lactate dehydrogenase of *Escherichia coli*. Interestingly, ORF analysis of DFR4 reveals sequence similarity to the C-terminal portion of XapB immediately downstream of IS1541. Thus, it would appear that IS1541 has been inserted into the *xapB* ORF in *Y. pestis* D15 Yokohama and lies immediately downstream of DFR3. DFR4 has been deleted from *Y. pestis* CO92 and is absent from all orientalis biovars tested. The sequence downstream of this copy of IS1541 in DFR4 encodes 17 additional ORFs, the first of which shows similarity to the C terminus of XapB. The rest of DFR4 encodes homologues of *hydH*/*hydG* from *Salmonella typhimurium*, and hypothetical ORFs from *Pseudomonas aeruginosa* and *E. coli*. The ORF of *Y. pestis hydH* contains an insertion when compared to *hydH* of *S. typhimurium*, and the *hydG* ORF is incomplete. The right boundary of DFR4 is delimited by another (truncated) copy of the insertion sequence element IS1541. DFR4 is present in some antiqua and mediaevalis strains, and is present in all four strains of *Y. pseudotuberculosis* tested (Table 2). DFR5 has 14 ORFs and is located near to the replication terminus of *Y. pestis* (accession no. AAF68950). The smallest DFR, DFR6, encodes three ORFs, one of which shows 73 % amino acid sequence identity to SepC from the insect pathogen *Serratia entomophila*. The *sepABC* operon is a member of a family of toxin complex (*tc*) genes that encode insecticidal proteins (Waterfield *et al.*, 2001). The G+C content of the entire *Y. pestis* CO92 genome is 47·6 mol %, whereas the G+C contents of the DFRs range from 43·1 to 52·5 mol %.

DISCUSSION

It is becoming increasingly apparent that the acquisition of genes via lateral transfer is the basis of bacterial genome diversity (Ochman *et al.*, 2000; Perna *et al.*, 2001). Lateral gene transfer enables strains to rapidly acquire large DNA segments that encode functions facilitating their survival. The very close genetic similarity of *Y. pestis* and *Y. pseudotuberculosis* has long been established (Bercovier *et al.*, 1980). The few sequence polymorphisms that have been reported for *Y. pestis* represent small changes in nucleotide sequence, and these would not be detected by SSH. It is likely that lateral gene transfer has contributed to the evolution of *Y. pestis*, perhaps even since it emerged as a clone derived from *Y. pseudotuberculosis*. Indeed, RFLP experiments indicate genome plasticity due to the mobility of larger regions of DNA (Achtman *et al.*, 1999; McDonough & Hare, 1997; Lucier & Brubaker, 1992), as does the instability of the 102 kb *pgm* locus (Fetherston *et al.*, 1992). The genome sequence of *Y. pestis* CO92 demonstrates the genetic flexibility of this strain (Parkhill *et al.*, 2001). The three genomic anomalies reported in *Y. pestis* CO92 were not detected by SSH since they represent rearrangements; SSH depends on the presence of a sequence in one genome (tester) and its absence in another (driver).

Data presented here indicate the presence of six large genomic DFRs ranging in size from 4·6 to 19 kb (Fig. 1), which demonstrates for the first time the genetic basis of genomic plasticity among the three biovars of *Y. pestis*. Interestingly, the DFR profile (L) of *Y. pseudotuberculosis* is not seen in *Y. pestis*. One of the *Y. pestis* DFRs, DFR2, is found in all but two of the 78 *Y. pestis* strains tested, but it is absent from all four isolates of *Y. pseudotuberculosis* tested (Table 2). Therefore, DFR2 appears to be a region that was acquired early in the evolution of *Y. pestis*. It possesses many hallmarks of lateral gene transfer: the 15 kb region encodes ORFs with similarities to many bacteriophage genes (including replication, primase and integration functions), it possesses an insertion element (IS1397) and it is delineated by two direct repeats. However, there is no evidence that DFR2 is inserted into a tRNA gene and its 44·5 mol % G+C content is not significantly different from that of the rest of the *Y. pestis* CO92 genome (47·6 %), as is common with pathogenicity islands (Hacker & Kaper, 2000). Since *Y. pestis* emerged less than 20000 years ago (Achtman *et al.*, 1999), it is more likely that new sequences were acquired from genomes with a similar G+C content, rather than amelioration whereby an incoming sequence is adjusted to the base composition of the resident genome (Lawrence & Ochman, 1997).

In contrast, the presence or absence of some of the DFRs (DFR1, DFR3 and DFR6) shows no correlation with the biovar of the isolates. These DFRs are also present in *Y. pseudotuberculosis* and have been subsequently deleted

Table 3. ORFs found in the DFRs

The ORFs were found using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). When the ORF could be equated to the gene identity of the *Y. pestis* CO92 genome (Parkhill *et al*., 2001) it is shown, along with the DNA strand on which it is encoded. BLASTP analysis against the non-redundant protein database indicates the putative functions of each ORF, followed by its size, the co-ordinates of the protein homology, the accession number of the homologue and the E-value (BLAST score; statistical significance threshold).

DFR	G+C content (mol%)	ORF no.	CO92 gene identity	Strand	BLASTP (ORF Finder)	ORF size (aa)	Homology		Accession no.	E-value
							Start of	End of		
DFR1	45·2	1	YPO0738	+	FliC, flagellin; *P. aeruginosa*	399	1	398	AAC28556	8·00E-40
		2	YPO0739	+	Fla, flagellin; *Clostridium tyrobutyricum*	401	39	374	CAB44444	6·00E-28
		3	YPO0740	+	LafB, flagellin; *Vibrio parahaemolyticus*	425	1	418	AAB07351	5·00E-55
		4	YPO0743	+	LafE, flagellin; *V. parahaemolyticus*	412	32	377	AAB07354	7·00E-17
		5	YPO0744	+	LafV, flagellin; *V. parahaemolyticus*	156	52	156	AAB07355	5·00E-05
		6	YPO0745	+	LafS, flagellin; *V. parahaemolyticus*	231	16	227	AAB07356	9·00E-37
		7	YPO0746	+	LafT, flagellin; *V. parahaemolyticus*	288	1	280	AAB07357	3·00E-88
		8	YPO0747	+	LafU, chemotaxis protein; *Escherichia coli*	303	49	303	BAA77900	8·00E-60
		9	YPO0749	+	Hypothetical protein; *Salmonella typhimurium* LT2	198	93	192	AAF33525	4·00E-05
		10	YPO0750	+	Novel	150	–	–	–	–
		11	YPO0755	+	OrfA, IS1397; *Escherichia coli*	178	8	176	CAA63546	5·00E-29
		12	YPO0756	+	OrfB, IS1397; *Escherichia coli*	219	5	218	CAA63547	6·00E-61
		13	Absent	–	Hypothetical protein Rv3096; *Mycobacterium tuberculosis*	391	7	345	CAB08388	2·00E-53
		14	YPO0758	+	RbsR, *rbs* repressor; *Haemophilus influenzae* Rd	339	8	333	AAC22164	2·00E-37
		15	YPO0759	+	Transporter, NadC family; *Vibrio cholerae*	456	5	455	AAC22267	1·00E-100
DFR2	44·5	1	YPO1085	–	OrfB, IS100 transposase; *Y. pestis*	259	1	259	AAC82752	1·00E-144
		2	YPO1086	–	OrfA, IS100 transposase; *Y. pestis*	340	1	340	AAC82751	0
		3	YPO1089	–	RepA protein, plasmid TF-FC2; *Escherichia coli*	364	44	209	A36134	2·00E-07
		4	YPO1090	–	DNA primase; phage phi-R73	329	25	327	C41830	2·00E-46
		5	YPO1091	–	Novel	195	–	–	–	–
		6	YPO1092	–	Orf3, hypothetical protein; *Francisella tularensis*	298	13	279	AAC12937	1·00E-06
		7	YPO1096	–	Integrin analogue gene, Uso1p; *Saccharomyces cerevisiae*	556	184	554	NP_010225	9·00E-07
		8	YPO1097	–	Novel	362	–	–	–	–
		9	YPO1098	–	SlpA, integrase; *Escherichia coli*	410	12	400	P32053	1·00E-131
		10	YPO1099	–	IS1541 transposase; *Y. pestis*	152	1	152	AAC82673	5·00E-89
DFR3	47·9	1	YPO1164	+	IS1541 transposase; *Y. pestis*	152	1	152	AAC82673	6·00E-88
		2	YPO1165	–	BetA, choline dehydrogenase; *P. aeruginosa*	567	2	558	AAG08757	0
		3	YPO1166	–	BetB, betaine aldehyde dehydrogenase; *P. aeruginosa*	490	1	490	AAG08758	0
		4	YPO1167	–	BetI, transcriptional regulator; *P. aeruginosa*	193	1	187	AAG08759	1·00E-65
		5	YPO1168	+	High-affinity choline transport protein; *Erwinia amylovora*	673	1	672	AAG31040	0
		6	YPO1169	+	XapR; *Escherichia coli*	297	9	295	S11407	3·00E-50
		7	YPO1170	–	YyaM, conserved hypothetical protein; *Bacillus subtilis*	310	23	303	S66005	9·00E-19
		8	YPO1171	+	XapA, xanthosine phosphorylase; *Escherichia coli*	287	14	285	P45563	1·00E-106
		9	YPO1172	+	XapB, xanthosine permease; *Escherichia coli*	282	17	231	P45562	3·00E-75
		10	YPO1173	+	IS1541 transposase; *Y. pestis*	152	1	152	AAC82673	6·00E-88

Table 3 (*cont.*)

DFR	G+C content (mol%)	ORF no.	CO92 gene identity	Strand	BLASTP (ORF Finder)	ORF size (aa)	Homology		Accession no.	E-value
							Start of	End of		
DFR4	43·1	1	Absent	+	IS1541 transposase; *Y. pestis*	152	1	152	AAC82673	6·00E-88
		2	Absent	+	XapB, xanthosine permease; *Escherichia coli*	201	1	201	P45562	8·00E-80
		3	Absent	−	Hypothetical ORF; *Salmonella typhimurium* LT-2	157	1	150	AAF33522	4·00E-17
		4	Absent	+	HydH, sensor protein; *Salmonella typhimurium*	607	26	591	AAF33505	4·00E-71
		5	Absent	−	Novel	109	–	–	–	–
		6	Absent	+	HydG; *Escherichia coli*	152	13	137	P14375	5·00E-43
		7	Absent	+	Morphinone reductase PA2932; *P. aeruginosa*	369	4	362	E83279	5·00E-96
		8	Absent	−	Probable transcription regulator PA3630; *P. aeruginosa*	303	9	293	C83191	6·00E-49
		9	Absent	+	Novel	172	–	–	–	–
		10	Absent	+	Novel	109	–	–	–	–
		11	Absent	−	YbhR, hypothetical protein (*moaE–rhlE* intergenic region); *Escherichia coli*	368	1	367	P75774	1·00E-178
		12	Absent	−	Novel	153	–	–	–	–
		13	Absent	−	YbhS, hypothetical protein (*moaE–rhlE* intergenic region); *Escherichia coli*	419	49	419	P75775	1·00E-163
		14	Absent	−	YbhF, hypothetical ABC transporter, ATP-binding protein; *Escherichia coli*	580	7	580	P75776	0
		15	Absent	−	Hypothetical protein b0795 precursor; *Escherichia coli*	328	1	325	C64816	1·00E-121
		16	Absent	−	YbiH, probable transcription regulator; *Escherichia coli*	236	16	234	D64816	3·00E-56
		17	Absent	+	RhlE, putative ATP-dependent RNA helicase; *Escherichia coli*	451	1	445	P25888	0
		18	Absent	+	IS1541 transposase; *Y. pestis* (partial)	91	1	91	AAC82673	9·00E-48
DFR5	43·6	1	YPO2265	−	Tus DNA replication terminus site-binding protein; *Y. pestis*	256	15	256	AAF68950	1·00E-137
		2	YPO2266	−	Sulfonamide resistance; *Escherichia coli*	300	1	299	BAA15330	1·00E-120
		3	YPO2267	+	Als, operon regulatory protein; *Escherichia coli*	166	2	163	BAA15329	7·00E-44
		4	YPO2268	+	Mlc protein; *Escherichia coli*	403	2	403	BAA15328	1·00E-172
		5	YPO2269	+	Dethiobiotin synthase; *Escherichia coli*	222	2	220	BAA15327	5·00E-75
		6	YPO2270	−	Putative chloride channel/membrane protein; *Escherichia coli*	430	10	430	B64915	1·00E-144
		7	YPO2274	+	Gene II product; bacteriophage I2-2	350	1	344	P15419	1·00E-118
		8	YPO2277	+	Novel	570	–	–	–	–
		9	YPO2279	+	pI protein; filamentous phage phiLf	341	8	298	JE0085	2·00E-18
		10	YPO2280	+	S-protein secretion D; *Aeromonas hydrophila*	414	144	387	AAA79322	2·00E-19
		11	YPO2283	−	Transcriptional regulator, LacI family; *V. cholerae*	335	1	328	AAF95481	6·00E-73
		12	YPO2284	−	Novel	156	–	–	–	–
		13	YPO2286	+	ABC transporter, periplasmic substrate-binding protein, putative; *Deinococcus radiodurans*	343	70	275	AAF10849	2·00E-12
		14	YPO2287	−	Novel	242	–	–	–	–
DFR6	52·5	1	YPO2379	+	NemA, *N*-ethylmaleimide reducing enzyme; *Escherichia coli*	365	1	365	JC5605	1·00E-161
		2	YPO2380	−	SepC; *Serratia entomophila*	984	1	984	AAG09644	0
		3	YPO2382	+	RNase T, degrades tRNA; *Escherichia coli*	226	12	224	P30014	6·00E-98

in some strains of *Y. pestis*. Eight of the 15 ORFs in DFR1 have similarity to genes involved in lateral flagellar synthesis responsible for surface swarming in *Vibrio parahaemolyticus* (McCarter & Wright, 1993). Since *Y. pestis* is non-motile, the presence of DFR1 does not apparently confer motility on those strains in which it is present. ORFs identified in DFR3 encode putative proteins involved in the purine salvage pathway (Seeger *et al.*, 1995). A single ORF was found in DFR6 with significant similarity to SepC, an insect toxin encoded on a plasmid in *Serratia entomophila* (Hurst *et al.*, 2000). Given the insect vector-borne aetiology of bubonic plague, this is an interesting observation to bear in mind as the mode of action of the SepABC insecticidal proteins is being uncovered.

Another DFR, DFR5, is present in only the orientalis biovar of *Y. pestis*, which is considered to have emerged most recently (Buchrieser *et al.*, 1999). DFR5 is absent from all mediaevalis and all but one of the antiqua biovars (Nicholisk 51), as well as from *Y. pseudotuberculosis* (Table 2). Therefore, DFR5 appears to be more recently acquired than DFR2 (which is seen in all three biovars). Two of the 14 ORFs in DFR5 are similar to bacteriophage genes, again implicating a lateral gene transfer event. Interestingly, DFR5 is located adjacent to the replication terminus of the genome, a site where increased recombination has been observed (Perkins *et al.*, 1993), and which may explain genome rearrangement involving DFR5 in the absence of insertion sequences.

DFR4 is absent from all orientalis biovars tested, indicating that the evolution of the biovars results from both the acquisition and deletion of genetic material. DFR4 is also deleted from Nicholisk 51, an antiqua strain, which has recently been shown to have an IS100-based fingerprint more typical of the orientalis biovar (Motin *et al.*, 2002). This observation led to the hypothesis that this antiqua strain has restored ability to ferment glycerol. The data presented here support this hypothesis, as Nicholisk 51 possesses the same six DFR alleles (profile A) as the majority of orientalis strains.

Currently, biovar designation is based on only two phenotypes: the ability to utilize glycerol (antiqua and mediaevalis biovars) and the ability to reduce nitrate (antiqua and orientalis biovars). Thus, the only discernable difference between Nicholisk 51 (antiqua) and the profile A orientalis biovars is the ability of the former to utilize glycerol. It is possible that the loss of glycerol utilization seen in orientalis strains is due to a deletion of 93 bp in the glycerol-3-phosphate dehydrogenase gene *glpD* (Motin *et al.*, 2002). Similarly, the single mediaevalis strain with profile F, Harbin 35 (strain 56, Table 2), only differs from antiqua strains with the same profile by its inability to reduce nitrate, which may also be a simple loss of gene function. Two strains in this collection, 316 (strain 77, Table 2) and Pestoides J (strain 78, Table 2), could not be characterized with respect to a biovar, since they are negative for both glycerol utilization and nitrate reduction. Since 316 has profile A it seems appropriate to align it with the orientalis strains, whereas Pestoides J seems to conform best with the Central Asian mediaevalis strains of profile I. Clearly, the comprehensive and accurate determination of the evolutionary development of the three biovars of *Y. pestis* requires analysis of more than two genetic markers.

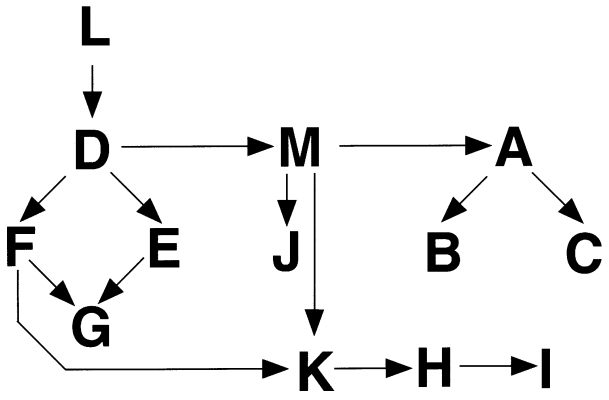

Fig. 2. Model of the evolutionary relationships of strains of *Y. pestis* based on different DFR profiles. Profile L is seen in the four strains of *Y. pseudotuberculosis* tested, whereas the remaining profiles represent those found in *Y. pestis.*

It is possible to use the DFR profiles presented here to propose a model of the evolutionary relationships between natural isolates. The number of profiles possible for six DFRs is 64 (2^6), of which 12 are represented in this strain collection (though the possibility exists that other strains possess profiles not found here). The model originates with *Y. pseudotuberculosis* as the progenitor, which possesses DFR profile L. Assuming that only one DFR allele is changed at a time, it is possible to suggest the simple evolutionary model presented in Fig. 2 using the 12 profiles found in this study. The only *Y. pestis* derivative with a single DFR change from profile L is profile D, which results from the acquisition of DFR2. This would imply that the earliest *Y. pestis* strains would possess profile D (Antiqua, Pestoides A and Pestoides B). The profiles that appear early in the evolution of *Y. pestis* (profiles D, E and F) are seen in the older antiqua and mediaevalis biovars. A lineage leading to the Central Asian mediaevalis strains, including KIM10 (profile I), can be proposed via profiles K and H (Fig. 2). Profile K can arise from profile D via F or M. Since profiles K and M are seen only in antiqua biovars, this is the more likely evolutionary route. The orientalis lineage progresses from profile D to profile A via profile M by the loss of DFR4 followed by the acquisition of DFR5. Twelve of the 51 orientalis strains have profiles that contain only one DFR allele different from profile A. Specifically, the loss of DFR3 leads to profile B, and the loss of DFR2 leads to profile C. *Y. pestis* strains with profile B are geographically diverse, whereas the two strains with profile C are both South American isolates. Significantly, these observations conform to the results of PFGE analyses of strains EV76 (Profile B) and Salazar

(Profile C), which showed different restriction patterns to the other orientalis strains studied (Lucier & Brubaker, 1992). There is more homogeneity within the DFR profiles of the orientalis biovar, which would be expected since the orientalis biovar has evolved over a shorter time period (since 1894) than the antiqua or mediaevalis biovars (1500–20000 years). Even though only three profiles are seen within the orientalis biovar, this is evidence that significant changes in the *Y. pestis* genome have occurred in as short a time frame as 100 years.

Insertion elements have played important roles in the pathogenicity of many bacteria (Hacker & Kaper, 2000), and are highly represented in the *Y. pestis* CO92 genome (Parkhill *et al.*, 2001). They have been previously implicated in the strain variability of *Y. pestis*, namely by RFLP analysis of the IS100 location (Achtman *et al.*, 1999; McDonough & Hare, 1997), and their presence contributes to the instability of the *pgm* locus (Buchrieser *et al.*, 1998; Fetherston & Perry, 1994). In *E. coli*, the majority of insertion sequences are associated with lateral gene transfer, and they are frequently located at junctions of native and transferred DNA (Lawrence & Ochman, 1998). In this study insertion elements are seen close to the borders of DFR1, DFR2, DFR3 and DFR4. DFR2 is flanked by two different insertion elements, IS100 and IS1541, and appears to be very stable as it is present in all but two of the strains tested. There is more variability in the presence or absence of DFR3 and DFR1, which are also flanked by insertion elements. One possible mechanism for the loss of DFR3 is recombination between the two directly repeated copies of IS1541, in a manner analogous to the instability of the *pgm* locus. The six DFRs of *Y. pestis* were stable during laboratory manipulations such as serial subculture or passage through laboratory animals. The original *Y. pestis* CO92 isolate was compared to 16 derivatives (*pgm*-negative derivatives, plasmid-less derivatives, strains in which plasmid pMT1 appears to have integrated into the chromosome and strains that had been subjected to long-term serial subculture on various media at room temperature and 37 °C). All *Y. pestis* CO92 derivatives gave the same result: profile A (no. 17; Table 2). Further evidence for stability under laboratory conditions is seen in the Yreka derivatives A1122 and A12 (strains 38, 11 and 12; Table 2), which all possess profile A. Presumably the selective pressure driving genome evolution in the natural environment of *Y. pestis* is far greater than that found under laboratory conditions.

In summary, here we report the first successful identification and characterization of large regions of DNA responsible for genomic differences between different strains of *Y. pestis*. This approach is particularly powerful when a reference genome is available, and in this study the genome sequences of *Y. pestis* CO92 and *Y. pseudotuberculosis* IP32953 were exploited. When no reference genome is available, a library of tester fragments can be interrogated which, although more laborious, will still provide useful data. Six DFRs were identified among *Y. pestis* strains, which result from both the acquisition and deletion of large regions of DNA, and these carry many hallmarks of lateral gene transfer and provide clues to the evolutionary relationships between different biovars and geographical isolates of *Y. pestis*. Future work will determine the exact boundaries of these DFRs and their benefits in an evolutionary context. Specifically, their role in the survival or pathogenicity of *Y. pestis* will be ascertained.

ACKNOWLEDGEMENTS

This work was performed under the auspices of the US Dept of Energy by the University of California, Lawrence Livermore National Laboratory, under contract no. W-7405-Eng-48, and was funded by the Department of Energy, NN-20, Chemical and Biological Non-Proliferation Program. We appreciate the generosity of May Chu (CDC, Fort Collins), Robert Brubaker (Michigan State University) and Will Probert (Dept of Health Services, Berkeley, CA) who supplied genomic DNA for some strains included in this study. We appreciate the technical assistance of Sylvia Gamez-Chin, Aubree Hubbell, Madison Macht and Jessica Wollard. We greatly appreciate the sharing of unpublished data by Vladimir Motin, and gratefully acknowledge his constructive suggestions for this manuscript.

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 USA* **96**, 14043–14048.

Adair, D. M., Worsham, P. L., Hill, K. K., Klevytska, A. M., Jackson, P. J., Friedlander, A. M. & Keim, P. (2000). Diversity in a variable-number tandem repeat from *Yersinia pestis*. *J Clin Microbiol* **38**, 1516–1519.

Akopyants, N. S., Fradkov, A., Diatchenko, L., Hill, J. E., Siebert, P. D., Lukyanov, S. A., Sverdlov, E. D. & Berg, D. E. (1998). PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc Natl Acad Sci USA* **95**, 13108–13113.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Bachellier, S., Clément, J. M., Hofnung, M. & Gilson, E. (1997). Bacterial interspersed mosaic elements (BIMEs) are a major source of sequence polymorphism in *Escherichia coli* intergenic regions including specific associations with a new insertion sequence. *Genetics* **145**, 551–562.

Bercovier, H., Mollaret, H. H., Alonso, J. M., Brault, J., Fanning, G. R., Steigerwalt, A. G. & Brenner, D. J. (1980). Intra- and interspecies relatedness of *Yersinia pestis* by DNA hybridization and its relationship to *Y. pseudotuberculosis*. *Curr Microbiol* **4**, 225–229.

Buchrieser, C., Prentice, M. & Carniel, E. (1998). The 102-kilobase unstable region of *Yersinia pestis* comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. *J Bacteriol* **180**, 2321–2329.

Buchrieser, C., Rusniok, C., Frangeul, L., Couve, E., Billault, A., Kunst, F., Carniel, E. & Glaser, P. (1999). The 102-kilobase *pgm* locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect Immun* **67**, 4851–4861.

Devignat, R. (1951). Variétés de l'espèce *Pasteurella pestis*. Nouvelle hypothèse. *Bull W H O* **4**, 247–253.

Diatchenko, L., Lau, Y. F., Campbell, A. P. & 8 other authors (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A* **93**, 6025–6030.

Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175–185.

Fetherston, J. D. & Perry, R. D. (1994). The pigmentation locus of *Yersinia pestis* KIM6$^+$ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. *Mol Microbiol* **13**, 697–708.

Fetherston, J. D., Schuetze, P. & Perry, R. D. (1992). Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. *Mol Microbiol* **6**, 2693–2704.

Gordon, D., Abajian, C. & Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res* **8**, 195–202.

Guiyoule, A., Grimont, F., Iteman, I., Grimont, P. A., Lefèvre, M. & Carniel, E. (1994). Plague pandemics investigated by ribotyping of *Yersinia pestis* strains. *J Clin Microbiol* **32**, 634–641.

Guiyoule, A., Rasoamanana, B., Buchrieser, C., Michel, P., Chanteau, S. & Carniel, E. (1997). Recent emergence of new variants of *Yersinia pestis* in Madagascar. *J Clin Microbiol* **35**, 2826–2833.

Hacker, J. & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641–679.

Hurst, M. R., Glare, T. R., Jackson, T. A. & Ronson, C. W. (2000). Plasmid-located pathogenicity determinants of *Serratia entomophila*, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of *Photorhabdus luminescens*. *J Bacteriol* **182**, 5127–5138.

Jain, R., Rivera, M. C. & Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* **96**, 3801–3806.

Lawrence, J. G. & Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**, 383–397.

Lawrence, J. G. & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**, 9413–9417.

Lucier, T. S. & Brubaker, R. R. (1992). Determination of genome size, macrorestriction pattern polymorphism, and nonpigmentation-specific deletion in *Yersinia pestis* by pulsed-field gel electrophoresis. *J Bacteriol* **174**, 2078–2086.

McCarter, L. L. & Wright, M. E. (1993). Identification of genes encoding components of the swarmer cell flagellar motor and propeller and a sigma factor controlling differentiation of *Vibrio parahaemolyticus*. *J Bacteriol* **175**, 3361–3371.

McDonough, K. A. & Hare, J. M. (1997). Homology with a repeated *Yersinia pestis* DNA sequence IS100 correlates with pesticin sensitivity in *Yersinia pseudotuberculosis*. *J Bacteriol* **179**, 2081–2085.

Motin, V. L., Georgescu, A. M, Elliott, J. M. & 8 other authors (2002). Genetic variability of *Yersinia pestis* isolates as predicted by PCR-based IS100 genotyping and analysis of structural genes encoding glycerol-3-phosphate dehydrogenase (*glpD*). *J Bacteriol* **184**, 1019–1027.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.

Parkhill, J., Wren, B. W., Thomson, N. R. & 32 other authors (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527.

Perkins, J. D., Heath, J. D., Sharma, B. R. & Weinstock, G. M. (1993). *Xba*I and *Bln*I genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. *J Mol Biol* **232**, 419–445.

Perna, N. T., Plunkett, G., III, Burland, V. & 25 other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157 : H7. *Nature* **409**, 529–533.

Seeger, C., Poulsen, C. & Dandanell, G. (1995). Identification and characterization of genes (*xapA*, *xapB*, and *xapR*) involved in xanthosine catabolism in *Escherichia coli*. *J Bacteriol* **177**, 5506–5516.

Skowronski, E. W., Armstrong, N., Andersen, G., Macht, M. & McCready, P. M. (2000). Magnetic microplate-format plasmid isolation protocol for high-yield, sequencing-grade DNA. *Biotechniques* **29**, 786–790.

Trebesius, K., Harmsen, D., Rakin, A., Schmelz, J. & Heesemann, J. (1998). Development of rRNA-targeted PCR and *in situ* hybridization with fluorescently labelled oligonucleotides for detection of *Yersinia* species. *J Clin Microbiol* **36**, 2557–2564.

Waterfield, N. R., Bowen, D. J., Fetherston, J. D., Perry, R. D. & ffrench-Constant, R. H. (2001). The *tc* genes of *Photorhabdus*: a growing family. *Trends Microbiol* **9**, 185–191.

Received 4 September 2001; revised 13 January 2002; accepted 15 February 2002.