DNA Research **11**, 179–197 (2004)

Complete Genome Sequence of *Yersinia pestis* Strain 91001, an Isolate Avirulent to Humans

Yajun Song,[1,†] Zongzhong Tong,[1,†] Jin Wang,[1] Li Wang,[3] Zhaobiao Guo,[1] Yanpin Han,[1] Jianguo Zhang,[2] Decui Pei,[1] Jingsong Zhou,[1] Haiou Qin,[2] Xin Pang,[1] Yujun Han,[2] Junhui Zhai,[1] Min Li,[4] Baizhong Cui,[4] Zhizhen Qi,[4] Lixia Jin,[4] Ruixia Dai,[4] Feng Chen,[2] Shengting Li,[2] Chen Ye,[2] Zongmin Du,[1] Wei Lin,[2] Jin Wang,[2] Jun Yu,[2] Huanming Yang,[2] Jian Wang,[2] Peitang Huang[3] and Ruifu Yang[1,*]

Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing 100071, P. R. China,[1] Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 100029, P. R. China,[2] Institute of Bioengineering, Academy of Military Medical Sciences, Beijing 100071, P. R. China,[3] Qinghai Institute for Endemic Diseases Prevention and Control, Xining 811602, P. R. China

(Received 19 January 2004; revised 3 April 2004)

Abstract

Genomics provides an unprecedented opportunity to probe in minute detail into the genomes of the world's most deadly pathogenic bacteria-*Yersinia pestis*. Here we report the complete genome sequence of *Y. pestis* strain 91001, a human-avirulent strain isolated from the rodent Brandt's vole-*Microtus brandti*. The genome of strain 91001 consists of one chromosome and four plasmids (pPCP1, pCD1, pMT1 and pCRY). The 9609-bp pPCP1 plasmid of strain 91001 is almost identical to the counterparts from reference strains (CO92 and KIM). There are 98 genes in the 70,159-bp range of plasmid pCD1. The 106,642-bp plasmid pMT1 has slightly different architecture compared with the reference ones. pCRY is a novel plasmid discovered in this work. It is 21,742 bp long and harbors a cryptic type IV secretory system. The chromosome of 91001 is 4,595,065 bp in length. Among the 4037 predicted genes, 141 are possible pseudogenes. Due to the rearrangements mediated by insertion elements, the structure of the 91001 chromosome shows dramatic differences compared with CO92 and KIM. Based on the analysis of plasmids and chromosome architectures, pseudogene distribution, nitrate reduction negative mechanism and gene comparison, we conclude that strain 91001 and other strains isolated from *M. brandti* might have evolved from ancestral *Y. pestis* in a different lineage. The large genome fragment deletions in the 91001 chromosome and some pseudogenes may contribute to its unique nonpathogenicity to humans and host-specificity.

Key words: *Yersinia pestis*; genome; evolution; pathogenicity

1. Introduction

Yersinia pestis, the causative agent of bubonic and pneumonic plague, is thought to be one of the most dangerous and deadly pathogenic bacteria in the world. There have been three known major plague pandemics in human history, which have claimed hundreds of thousands of lives. *Y. pestis* has been classified into three biovars according to their ability to reduce nitrate and utilize glycerol: *Antiqua* (positive for both), *Mediaevalis* (negative for nitrate reduction and positive for glycerol utilization), and *Orientalis* (positive for nitrate reduction and negative for glycerol utilization). These three biovars are thought to be responsible for the three major plague pandemics: the Justinian plague, the Black Death and the modern plague, respectively.[1] The third plague pandemic was believed to have originated from China in the 19th century.

Two independent groups have decoded the whole genome sequences of two fully virulent *Y. pestis* strains: CO92 (*Orientalis* strain) and KIM (*Mediaevalis* strain) respectively.[2,3] These important works provide copious data for comparative genomic research. *Y. pestis* strain 91001 was isolated from *Microtus brandti* in Inner Mongolia, China. It has a LD_{50} of 23.2 cells for mice by subcutaneous challenge, whereas 10^9 live cells of 91001 failed to cause any infectious symptoms in rabbits. The most striking characteristic of 91001 is that $1.5\times10^{$ cells challenging through the subcutaneous route caused neither

Communicated by Hideo Shinagawa

* To whom correspondence should be addressed. Tel. +86-10-66948595, Fax. +86-10-83820748, E-mail: yangrf@nic.bmi.ac.cn

† These authors contributed equally to this work.

bubonic plague nor pneumonic plague in a volunteer trial.[4] To better understand the secrets of this highly lethal pathogen, we carried out genome sequencing of *Y. pestis* strain 91001 for comparison by the "whole genome shotgun" method.

2. Materials and Methods

2.1. Bacterial strain

Y. pestis strain 91001 has the following major phenotypes: $F1^+$ (able to produce fraction 1 antigen or the capsule), VW^+ (presence of V antigen), Pst^+ (able to produce pesticin) and Pgm^+ (pigmentation on Congo red media). This strain falls into biovar *Mediaevalis* according to traditional criteria (negative for nitrate reduction and positive for glycerol utilization). This strain also has unique carbohydrate utilization features: arabinose negative, rhamnose positive and melibiose positive.

2.2. Genome sequencing, assembling and finishing

The genome sequence of 91001 was determined by a whole genome shotgun method.[5,6] Briefly, chromosomal DNA and plasmids were extracted following standard protocols. The DNA was fragmented by sonication, and fragments ranging from 1.5 to 3 kb were extracted from an agarose gel after size-fractionation, and then randomly cloned into a pUC18 vector after end repairing. At the same time, the chromosomal DNA was partially digested with *Alu* I to construct a library with DNA inserts ranging from 1.5 to 3 kb. A total of 55,000 clones were sequenced from both ends using dye-terminator chemistry on a MegaBace auto sequencing machine (Beckman, USA) and an ABI 377 sequencer (ABI, USA). Base calling was performed with the software Phred.[7,8] There were 84,220 qualified reads (>200 bp at Phred Value Q20) collected, which gave rise to chromosome coverage of 8.6-fold. And 3,905 qualified reads, of which 2200 reads were picked from genome library reads after running BLAST with plasmids sequence from *Yersinia pestis* strain CO92,[2,3] gave rise to plasmid coverage of 14.8-fold. The genome sequence was then assembled by using program RePS.[9] All the assembled contigs were checked for accuracy with the software package Consed.[11] For finishing, all the contigs were analyzed with RePS to construct "genome scaffolds," and also mapped onto the genomes of strain CO92 and KIM.[2,3] Primers from every contig were designed by Consed to close the gaps by PCR.[11] At the finishing stage, an additional 5324 chromosome and 362 plasmid reads were added to the final assembly of chromosome and plasmids, respectively. Finally, the overall sequence quality of the genome was further improved by using the following criteria: (1) three independent, high-quality reads as minimal coverage, (2) sequence coverage accountable from both strands and (3) Phred quality value >Q40 for each given base. Based on the final consensus quality scores generated by Phrap, we estimated an overall error rate of 0.90 in 10,000 bases for the final gap-free genome assembly.[7,8] The complete sequence assembly was verified by digesting genomic DNA with *I -Ceu* I restriction enzymes followed by pulsed-field gel electrophoresis (PFGE) analysis, and also by PCR amplification.[12]

2.3. Annotation and comparative genomic analysis

The final genome sequence was confirmed and annotated as described previously.[6] Briefly, three different sets of potential CDS (coding sequences) were established with GLIMMER 3.0, ORPHEUS and CRITICA at their default settings respectively.[13–15] All the predicted CDS and putative intergenic sequences were subjected to further manual inspections. Exhaustive BLAST searches with an incremental stringency against the NCBI nonredundant protein database were performed to determine homology of the predicted coding sequences. Translational start codons were identified based on protein homology, proximity to ribosome-binding site, relative positions to predicted signal peptide, and also putative promoter sequences. Then the three sets of CDS (longer than 150 bp) were integrated and combined. When frameshifts and point mutations were discovered from two adjacent CDS, they were classified as inactive genes or pseudogenes after careful inspections of the raw sequence data. To find putative orthologs in other completed genome sequences, CDS from the genomes were searched against the NCBI nonredundant protein database, and also assigned according to the COG database search results. Protein motifs and domains of all CDS were documented based on intensive searches against InterPro databases.[17] Transfer RNAs, RNase P genes and other stable RNAs were predicted with the tRNAscan-SE software.[18] Transmembrane domains, putative membrane proteins, and ABC transporters were identified with the TMHMM software.[19] VNTR (variable number tandem repeat) elements in the genome were identified using Tandem Repeat Finder.[20] Finally, Artemis was used to integrate and visualize all the annotation features.[21] Comparative genomic analysis was performed using the BLAST algorithm[9] and the Artemis Comparison Tool (ACT) (http://www.sanger.ac.uk/Software/ACT/). Major genomic rearrangements identified by *in silico* analysis were further confirmed by PCR amplification.

3. Results and Discussion

3.1. General features of 91001 chromosome

The genome of *Y. pestis* strain 91001 is composed of one chromosome and four plasmids (accession number: AE017042 for chromosome, AE017043 for pCD1, AE017044 for pCRY, AE017045 for pMT1, and pCP1

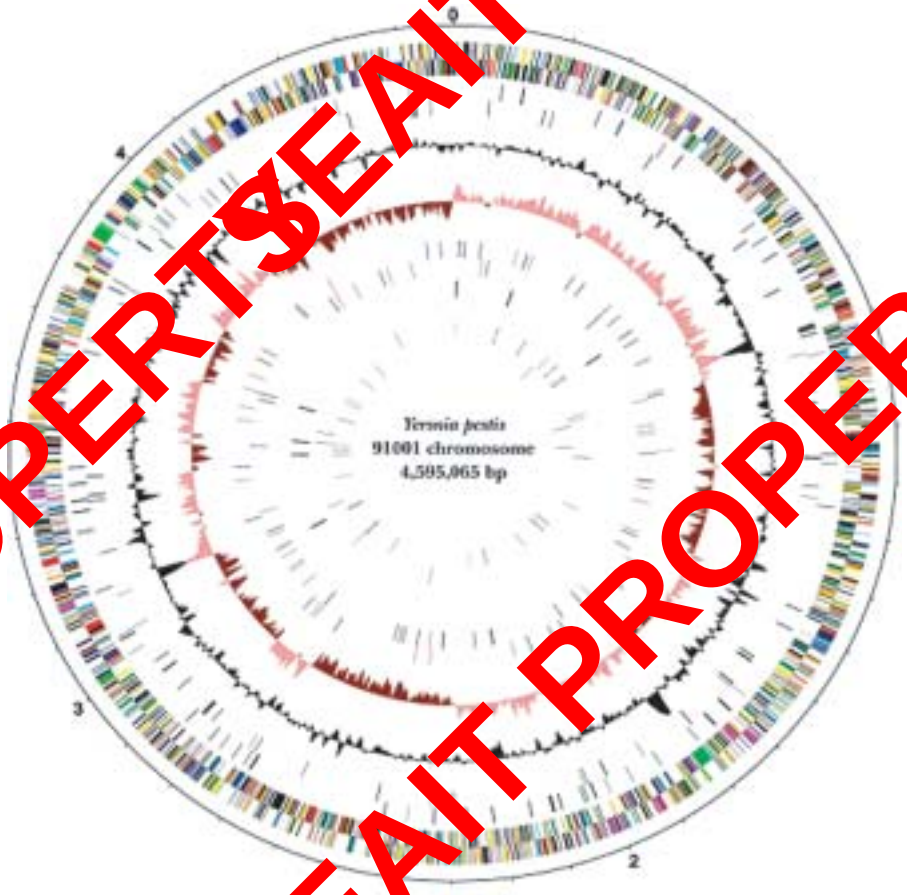

Figure 1. Circular representation of the *Y. pestis* 9100[illegible] genome. Circles display (from the outside): (1) Physical map scaled in [illegible] from base 1, the start of the putative [illegible] origin. (2) Coding sequences transcribed in the clockwise direction. [illegible]oding sequences transcribed in the counterclockwise direction. Genes displayed in 2 and 3 are colour-coded according to different functional categories: translation/ribosome structure/biogenesis, pink; transcription, olive drab; DNA replication/recombination/repair, forest green; cell division/chromosome partitioning, light blue; posttranslational modification/protein turnover/chaperones, purple; cell envelope biogenesis/outer membrane, red; cell motility/secretion, plum; inorganic ion transport/metabolism, dark sea green; signal transduction mechanisms, medium purple; energy production/conversion, dark olive green; carbohydrate transport/metabolism, gold; amino acid transport/metabolism, yellow; nucleotide transport/metabolism, orange; coenzyme metabolism, tan; lipid metabolism, salmon; secondary metabolites biosynthesis/transport/catabolism, light green; general function prediction only, dark blue; conserved hypothetical, medium blue; hypothetical, black; unclassified, light blue; pseudogenes, gray. (4) Pseudogenes in the clockwise direction. (5) Pseudogenes in the counterclockwise direction. (6) G + C percent content (in a 10-kb window and 1-kb incremental shift); values >47.6% (average) are in plotted outwards and values <47.6% in inwards. (7) GC skew (G − C/G + C, in a 10-kb window and 1-kb incremental shift); values greater than zero are in magenta and less than in brown. (8) IS elements (IS) in the clockwise direction. (9) IS in the counterclockwise direction. (10) rRNA genes in the clockwise direction. (11) rRNA genes in the counter clockwise direction. (12) tRNA genes in the clockwise direction. (13) tRNA genes in the counter clockwise direction. IS elements in 8 and 9 are colored as following: IS 100, green; IS 285, black; IS 1541, red; IS 1661, blue.

or AE017046). The chromosome of strain 91001 is 4,595,065 bp in length, shown in circular illustration in Fig. 1. There are obvious anomalies in GC skew of the chromosome of 91001 (see Fig. 1), which has been observed in CO92 but not in KIM.[2,3] Anomalies of GC bias in the CO92 chromosome are caused by rearrangements (inversion or translocation) of large genome DNA blocks mediated by IS (insertion sequence) elements, which had been confirmed by elaborate PCR amplifications. Parkhill et al. even identified genome rearrangements during the growth of a single culture of CO92, which suggested that the *Y. pestis* genome possesses frequent *in vitro* intra-genomic recombination and abnormal fluidity.[2] We also confirmed the anomalies of chromosome GC bias in 91001 by PCR amplifications; however we failed to identify any genome rearrangements within the same batch culture of this strain. Figure 2 portrays the dramatic structural variations between three finished chromosomes of *Y. pestis*, from which we deduced that genomes of *Y. pestis* are highly dynamic because of their abundant IS elements. Genome rearrangements including translocation, inversion and inverted translocation. Due

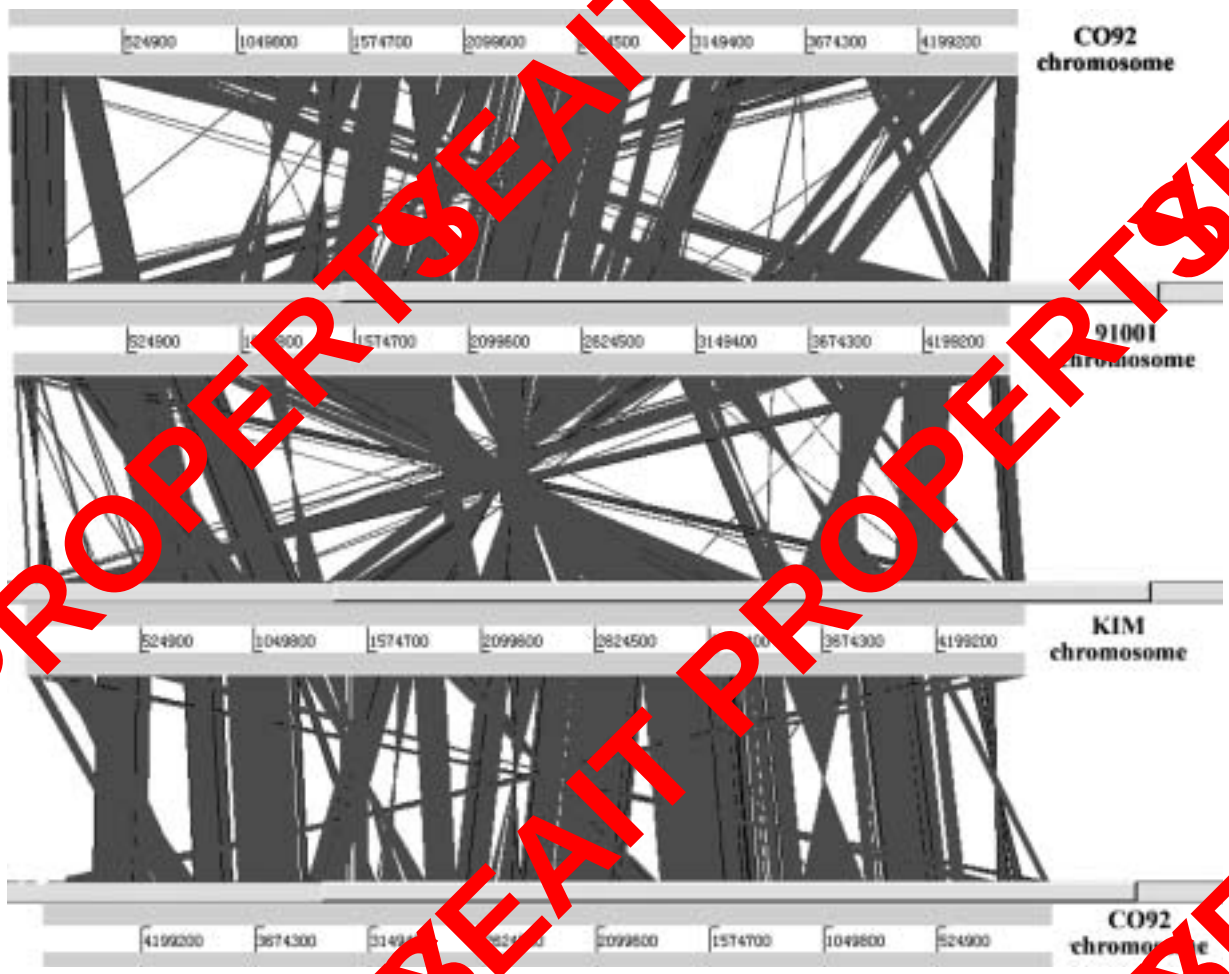

Figure 2. Linear genomic comparison of the circular chromosomes of CO92, KIM and 91001. This picture was generated by running the Artemis Comparison Tool (ACT); the horizontal bars and numbers represent the forward and backward strands to scale; similar genome fragments (99% identity at DNA level between two genomes) were connected with black lines and blocks; regions connected by two triangles indicated inversion rearrangements.

to the frequent genomic rearrangement events, it is very difficult to track down the "genome backbone architecture" of *Y. pestis*.

4. Comparative Genomic Analysis of Chromosome

4.1. Genome structure

A brief comparison of three finished *Y. pestis* chromosomes (91001, CO92 and KIM) is summarized in Table 1. G+C% contents are almost identical in the three chromosomes (about 47.6%). CO92 has six copies of rRNA operons, while 91001 and KIM both have seven copies. The number of tRNA genes also differs slightly in 91001 (72), CO92 (70) and KIM (73). Chromosome of 91001 is somewhat shorter than that of CO92 and KIM, which can be partly explained by the differences in copy numbers of IS elements. For IS 100, copy number variation to strain is: CO92 (44) > KIM (35) > 91001 (30). This is also the case for IS 1541: CO92 (72) > KIM (49) > 91001 (43). While in *Y. pseudotuberculosis*, the copy numbers of IS 100 and IS 1541 are 0-6 and 7-13, respectively, far less than in *Y. pestis*.[22,23] It is believed that *Y. pestis* is a clone that evolved from *Y. pseudotuberculosis* 1500–20,000 years ago shortly before the first known pandemics of human plague.[24] Biovar *Antiqua* seems to be closest to the ancestral *Y. pestis*, while *Mediaevalis* and *Orientalis* were derived from *Antiqua* separately.[24] Among the three strains, CO92 (biovar *Orientalis*) is the latest one to diverge;[25] therefore, we assume that the accumulation of IS elements is an important process in the course of within-species microevolution of *Y. pestis*. Hence, strain 91001 seems to be an "older" strain in this sense.

Another important factor counting for the chromosome length variation is that, there are large indels (insertions and deletions) among the chromosomes of these three strains. The 91001 chromosome has a 61.8-kb unique fragment, which had been discovered in previ-

Table 1. General features of chromosome of *Yersinia pestis* 91001, CO92 and KIM.

Accession Number	AE017042	AL590842	AE009952
Source	91001	CO92	KIM
Length (bp)	4,59[illegible]	4,653,728	4,600,755
G+C content	47.[illegible]%	47.64%	47.64%
Coding sequences*	403[illegible]	4,012	4,198
of which pseudogenes	141	149	54**
Coding density	81.6%	83.8%	86%
Average gene length (bp)	966	998	940
rRNA operon	7×(16S-23S-5S) + 5S	6×(16S-23S-5S)	7×(16S-23S-5S)
Transfer RNAs	72	70	73
Other stable RNAs	6	6	—
IS100	30 intact	44 intact	35 intact
IS1541	43 intact	62 intact	49 intact
	2 disrupted by IS 100	2 disrupted by IS 100	3 disrupted by IS 100
	3partial	2 partial	6 partial
IS1661	7 intact	7 intact	8 intact
	2 partial	2 partial	2 partial
IS285	[illegible] intact	21 intact	19 intact

*: Differences of CDS number in different chromosome are mainly due to different genome annotation standards, which also influence the coding density and average gene length.
**: Although the corresponding sequences of many pseudogenes in CO92 are identical in KIM, they are not annotated as pseudogenes by the authors, and we took a criterion similar to CO92 sequencing group.
–: not annotated.

ous suppression subtractive hybridization assays[26] and was termed as DFR4 (different region 4). All the four tested *Y. pseudotuberculosis* harbor this fragment.[26] Our further DFR typing result shows that, among the 257 tested *Y. pestis* strains isolated in China, only those isolated from *Microtus brandti* and *M. fuscus* possess this fragment (unpublished data). As DFR4 is shared by *Y. pseudotuberculosis* and *Y. pestis* strains isolated from *Microtus* but is not present in other *Y. pestis* strains, we deduce that strains isolated from *Microtus* (including 91001) might be those most closely related to ancestral *Y. pestis* strains. We also identified a 36-kb fragment shared by CO92 and KIM, which is absent in strain 91001. This fragment seems to be a prophage. The predicted genes of CO92- and KIM-specific genome fragments are listed in Table 2. As CO92 and KIM are both virulent to humans and 91001 is only lethal to mice, fragments specific for CO92 and KIM might contribute to the pathogenicity to humans of *Y. pestis*. Prophages are thought to be a major drive for the evolution of bacterial genomes through lateral gene transfer.[27] Moreover, it has been widely accepted that the bacteriophage genome encode certain virulence factors, including the well-characterized bacterial toxins and proteins that alter antigenicity, several new classes such as superantigens, effectors translocated by a type III secretion system, and proteins required for intracellular survival and host cell attachment.[28] Hayashi et al. revealed that, compared with its nonpathogenic counterpart K-12, about half of the O157 Sakai-specific sequences are of bacteriophage origin, which strongly suggests that bacteriophages play a predominant role in the pathogenicity and evolution of O157:H7 strains.[5] Therefore, it is reasonable to assume that this 33-kb prophage-like fragment might provide a

Table 2. Gene list of CO92 and KIM specific fragments.

genes in CO92	Length (bp)	predicted products	presence in KIM*
YPO2095	210	hypothetical phage protein	+
YPO2096	276	hypothetical phage protein	+
YPO2097	321	putative phage protein	+
YPO2098	513	putative phage lysozyme	+
YPO2099	459	putative prophage endopeptidase	+
YPO2100	798	phage regulatory protein	+
YPO2101	636	hypothetical phage protein	+
YPO2102	4[illegible]	hypothetical phage protein	+
YPO2103	1[illegible]	putative phage terminase (pseudogene)	+
YPO2104	[illegible]209	transposase for the IS285 insertion element	+
YPO2106	1583	putative phage protein (pseudogene)	
YPO2[illegible]	1113	hypothetical phage protein	+
YPO[illegible]109	774	hypothetical phage protein	+
YPO2[illegible]	1206	putative phage protein	+
YPO[illegible]11[illegible]	531	hypothetical phage protein	+
YPO[illegible]112	255	conserved hypothetical phage protein	+
YPO[illegible]2113	351	hypothetical phage protein	+
YPO2114	585	hypothetical phage protein	+
YPO2115	408	hypothetical phage protein	+
YPO2116	921	putative phage protein	+
YPO2117	312	hypothetical phage protein	+
YPO2118	222	hypothetical phage protein	+
YPO2119	3504	putative phage tail protein	+
YPO2120	342	putative phage protein	+
YPO2122	753	putative phage protein	+
YPO2123	711	putative phage minor tail protein	+
YPO2124	633	hypothetical phage protein	+
YPO2125	540	putative phage regulatory protein	+
YPO2126	1095	putative phage protein	+
YPO2127	552	putative phage-related membrane protein	+
YPO2128	351	putative phage-related lipoprotein	+
YPO2129	621	putative phage tail assembly protein	+
YPO2130	31[illegible]	hypothetical phage protein	+
YPO2131	3204	putative phage host specificity protein	+
YPO2132	9[illegible]	hypothetical phage protein	+
YPO2132	[illegible]09	hypothetical phage protein	+
YPO2133	[illegible]18	hypothetical phage protein	+
YPO2134	420	putative phage tail fiber assembly protein	
YPO21[illegible]	156	hypothetical phage protein	+
YPO[illegible]487	339	putative membrane protein	+
YPO2[illegible]	252	hypothetical protein	+
YPO[illegible]48[illegible]	504	conserved hypothetical protein	+

*: Genome of KIM was annotated following different criteria from CO92, so the gene numbers are omitted. Symbol "+" indicates the presence of corresponding genes.

virulence enhancement mechanism to CO92, KIM and other fully virulent strains thus helping broaden their host range. The finding that 91001 lacks this fragment only shows defective pathogenicity to human beings. We are on the way of constructing mutants to verify this hypothesis.

4.2. CDS and pseudogenes

Table 3 summarizes the assigned functions of predicted CDS in the chromosome of strain 91001. Just like other finished bacterial genomes, a large portion (about 16%) of CDS in 91001 was annotated as "hypothetical" or "conserved hypothetical." We carried out extensive gene level comparisons in the three finished genomes, and found that despite the dramatic genome structure differences, genes in the three chromosomes share great similarity. More than 90% of the amino acid sequences of predicted genes in the three chromosomes are totally identical (Table 4). Only a few genes are absent in different genomes, mainly localized on the large genomic fragments mentioned above, while the remaining genes are

Table 3. COG assigned functional categories of predicted CDS in the chromosome of strain 91001.

Functional categories	Number	Percentage
Information storage and processing	**663**	**16.42%**
Translation, ribosomal structure and biogenesis	161	3.99%
Transcription	205	5.08%
DNA replication, recombination and repair	296	7.33%
Cellular processes	**825**	**20.44%**
Cell division and chromosome partitioning	33	0.82%
Posttranslational modification, protein turnover, chaperones	108	2.68%
Cell envelope biogenesis, outer membrane	192	4.76%
Cell motility and secretion	204	5.05%
Inorganic ion transport and metabolism	188	4.66%
Signal transduction mechanisms	100	2.48%
Metabolism	**1146**	**28.39%**
Energy production and conversion	174	4.31%
Carbohydrate transport and metabolism	302	7.48%
Amino acid transport and metabolism	310	7.68%
Nucleotide transport and metabolism	77	1.91%
Coenzyme metabolism	115	2.85%
Lipid metabolism	71	1.76%
Secondary metabolites biosynthesis, transport and catabolism	97	2.40%
Poorly characterized	**1263**	**31.29%**
Unclassified function	306	7.58%
General function prediction only	309	7.65%
Conserved hypothetical protein	243	6.02%
Hypothetical protein	405	10.03%
Pseudogenes	**141**	**3.49%**
Total	**4037**	**100%**

Table 4. Gene level comparison of chromosomes of three *Y. pestis* strains

Pairwise comparison of predicted genes (including pseudogenes)

		91001			CO92			KIM		
strain	total	identical	similar*	absent$	identical	similar*	absent$	identical	similar*	absent$
91001	4037	4037	—	—	3653	363	21	3630	363	44
CO92	4011	3597	358	56	4011	—	—	3762	209	40
KIM	4144	3725	381	38	3919	23	2	4144	—	—

Pairwise comparison of predicted genes (excluding pseudogenes)

		91001			CO92			KIM		
strain	total	identical	similar*	absent$	identical	similar*	absent$	identical	similar*	absent$
91001	3896	3896	—	—	3566	311	19	3545	312	39
CO92	3874	3508	313	53	3874	—	—	3657	179	38
KIM	4090	3703	350	37	3880	207	3	4090	—	—

*: tBlastN comparing results indicating more than 90% similarity with more than 90% coverage;
$: absent genes are mainly included in the large genome fragment deletions.

highly similar in the three genomes. The high conservation of coding sequences in *Y. pestis* is consistent with the relatively shorter evolutionary routes of this pathogen.[24] Table 4 also shows that, although 91001 and KIM are both *Mediaevalis* strains, gene-level similarity between them is less than those between them and CO92. For example, when we compare the 4090 genes (excluding pseudogenes) predicted in KIM with their counterparts in strain 91001, 3703 are identical, 350 are highly similar and 37 are absent in strain 91001, whereas 3880 are identical, only 207 are similar and only 3 are absent in CO92. This is also the case when comparing predicted CDS of CO92 to KIM and 91001 respectively. That is to say, KIM resembles CO92 more than 91001 considering CDS similarity. This strongly implies that, although 91001 and KIM are both *Mediaevalis* strains, they may be located on different evolutionary lineages.

In the course of evolution of *Y. pestis*, one essential process is the inactivation of many genes related to enteropathogenic lifestyle, such as the O- antigen clusters *wbA* and *inv*.[2] We identified 141 pseudogenes in the chromosome of 91001, and they are deactivated by IS element insertion, nonsense mutation or frameshift. Interestingly, CO92 and KIM, belonging to different biovars, share 114 pseudogenes, while possessing 23 and 13 unique pseudogenes, respectively. 91001 and KIM, which belong to the same biovar, share only 94 pseudogenes, and have 53 and 33 unique genes, respectively. The distribution of pseudogenes again suggests that 91001 and KIM might locate in different evolutionary branches. Table 5 shows 91001-specific pseudogenes which are intact, and therefore presumably active, in both CO92 and KIM. CO92 and KIM are fully virulent strains and strain 91001 is avirulent to humans, some of these 91001-specific pseudogenes are probably related to the pathogenicity and host range of *Y. pestis*. Some of these pseudogenes, which encode regulatory proteins, membrane-related proteins, etc., are of special significance for further investigation.

4.3. pgm *locus*

The *pgm* locus is an established virulence-related gene cluster in *Y. pestis*, which determines the pigmentation phenotype when growing on Congo red media. During successive *in vitro* passages, the pigmentation phenotype in most *Y. pestis* strains can spontaneously lost with a frequency of approximately 10^{-5} per generation.[29] While the pigmentation phenotype of 91001 is very stable, ten passages *in vitro* did not produce a *pgm*$^-$ colony.[4] The 102-kb *pgm* segment consists of two parts: *hms* locus (hemin storage locus) and HPI (high pathogenicity island). The *pgm* locus of strain 91001 is a bit different from the reference strains. One striking feature is that, in CO92 and KIM, there are two IS 100 elements with the same orientation flanking the *pgm* locus, while in 91001 there is only one IS 100 element adjacent to *pgm* locus. According to the IS mediated intra-genome recombination mechanism, it needs two IS elements with the same orientation to trigger the embedded genome fragment to be deleted.[30] This finding presents a rational explanation for the stability of the *pgm* locus in strain 91001 and also other strains isolated from *Microtus*. Compared with CO92 and KIM, the *pgm* locus of 91001 lacks an additional IS 285 element in the HPI region. Further large-scale screening of the 257 strains revealed that this IS 285 element is unique in the *pgm* locus of strains from *M. Brandti* (unpublished data). The regions flanking the *pgm* locus in 91001 also show variation from the reference strains. In CO92 and KIM, the IS 100 element adjacent to *hms* disrupted a gene encoding a protein resembling Porin of *Escherichia coli*, while in strain 91001, there is no IS 100 element adjacent to the *hms* region and this gene remains intact as it does in *Y. pseudotuberculosis*. An additional IS 100 element 2500 bp away from *hms* disrupted the *hutC* gene in strain 91001 and mediated a translocation which caused the two parts of the *hutC* gene to be separated by 200 kb. We confirmed this rearrangement by PCR in strain 91001 and other *Microtus* strains. The *hutC* gene encodes a GntR family regulator, and the biological consequence of this inactivation still need to be clarified.

4.4. NapA gene

The nitrate reduction phenotype is one of the two key characteristics used to assign *Y. pestis* strains into different biovars. *Mediaevalis* strains possess a nitrate reduction-negative phenotype. Genome sequence analysis of KIM revealed that the nitrate reduction-negative phenotype is due to a nonsense mutation in nt_{613} of the *napA* gene.[3] However, strain 91001 does not have this mutation in the *napA* gene. Instead, a 1021_G-1021_A mutation in this gene leads to a 341_{Ala}-341_{Thr} substitution in the NapA protein, and the change in polarity of the amino acid might alter the activity of NapA. We used site-specific primers to screen the distribution of these two kinds of mutations of the *napA* gene in 257 *Y. pestis* strains isolated in China. The results show that 43 strains isolated from *Microtus* all bear the 1021_G-1021_A mutation, and the other 54 *Mediaevalis* strains possess a 613_G-613_T nonsense mutation, while the remaining *Orientalis* and *Antiqua* strains have neither of the two mutations (unpublished data). Our data strongly suggest a probable novel inactivation mechanism of the *napA* gene, and also support our hypothesis that strains from *Microtus* are slightly different from other *Mediaevalis* strains.

4.5. Carbohydrate metabolism

Y. pestis 91001 and other strains isolated from *Microtus* are all unable to utilize arabinose. Extensive genome analysis reveals that a regulatory gene *araC* and arabinose operon *araABCDFGH* in 91001 is flanked by

Table 5. Pseudogenes specific for 91001, which are all intact in both CO92 and KIM.

91001	CO92	KIM	predicted products	mutation in 91001
YP1715	**YPO1973**	**y2339**	**GntR-family transcriptional regulatory protein**	**disrupted by IS 100 and rearrangement**
YP1823	**YPO1973**	**y2339**	**GntR-family transcriptional regulatory protein**	**disrupted by IS 100 and rearrangement**
YP3522	YPO0918	y3305	LysE type translocator	disrupted by IS 100 and rearrangement
YP3615	YPO0918	y3305	LysE type translocator	disrupted by IS 100 and rearrangement
YP1639	**YPO1753**	**y2556**	**ferrichrome receptor protein**	**disrupted by IS 100 and rearrangement**
YP0084	YPO0082	y0055	putative transferase	disrupted by IS 100
YP0442	YPO0286	y0577	putative coproporphyrinogen III oxidase	disrupted by IS 100
YP2094	**YPO2309**	**y1404**	**two-component regulatory system, sensor kinase**	**disrupted by IS 100**
YP2435	YPO2728	y1537	putative membrane protein	disrupted by IS 100
YP2529	**YPO2826**	**y1304**	**RpiR-family transcriptional regulatory protein**	**disrupted by IS 100**
YP3376	YPO3615	y0036	amino acid permease	disrupted by IS 100
YP1700	YPO1956	y2354	hypothetical protein	disrupted by IS 285
YP1849	YPO1687	y1849	putative alanine racemase	disrupted by IS 285
YP2456	**YPO2654**	**y1228**	**putative membrane protein**	**disrupted by IS 285**
YP3352	**YPO0804**	**y3192**	**putative regulatory membrane protein**	**disrupted by IS 285**
YP2955	YPO0641a	y3540	hypothetical protein	disrupted by IS 285
YP1833	YPO1985	y2326	putative glycosyl transferase	partial deletion
YP2063	YPO2266	y2108	Permeases of the major facilitator superfamily	partial deletion
YP2669	YPO3047	y1434	sulfatase related protein	partial deletion
YP2669a	YPO3046	y1433	putative sulfatase modifier protein	partial deletion
YP0168	YPO0166	y3950	putative glycosyl hydrolase	nonsense mutation
YP1089	YPO2624	y1199	putative N-acetylglucosamine metabolism protein	nonsense mutation
YP3566	YPO0869	y3253	hypothetical protein	nonsense mutation
YP0185	YPO0186	y3967	putative sugar transferase	frame shifit mutation
YP0614	YPO3469	y0715	maltose/maltodextrin transport ATP-binding protein	frame shifit mutation
YP0820	YPO3110	y1074	putative O-antigen flippase	frame shifit mutation
YP0827	YPO3099	y1081	mannose-phosphate guanylyltransferase	frame shifit mutation
YP1372	YPO1483	y2687	conserved hypothetical protein	frame shifit mutation
YP1888	**YPO2045**	**y2267**	**putative hemolysin**	**frame shifit mutation**
YP2054	YPO2258	y2100	arabinose operon regulatory protein	frame shifit mutation
YP2345	YPO2534	y1853	conserved hypothetical protein	frame shifit mutation
YP2433	**YPO2731**	**y1656**	**putative membrane protein**	**frame shifit mutation**
YP2578	YPO2925	y1531	hypothetical protein	frame shifit mutation
YP2671	YPO3047	y1434	binding protein-dependent transport system	frame shifit mutation
YP2914	YPO0194	y3585	conserved hypothetical protein	frame shifit mutation
YP3011	**YPO0698**	**y3480**	**outer membrane usher protein**	**frame shifit mutation**
YP3044	YPO0733	y3445	putative flagellar hook-associated protein	frame shifit mutation
YP3048	YPO0737	y3441	putative flagellin related protein	frame shifit mutation
YP3349	YPO0962	y3349	hypothetical protein	frame shifit mutation
YP3923	YPO3624	y0245	putative aliphatic sulfonates binding protein	frame shifit mutation

Because CO92 and KIM are fully virulent strains and 91001 is avirulent to human, some of these 91001-specific pseudogenes are probably related to pathogenicity and host range of *Y. pestis*. Pseudogenes in boldface are of special interest for further study.

a 112-bp deletion from the 7th nt and a frameshift at codon 226. The *araC* gene encodes a regulator, which initiates the transcription of other genes in the operon in the presence of arabinose and also suppresses the transcription without arabinose.[32] We confirmed the 112-bp deletion in all the strains isolated from *Microtus* in China (unpublished data), which implies the stable genetic basis for arabinose-negative phenotype in strain 91001 and other *Microtus* strains.

Strain 91001 can utilize melibiose, but most other *Y. pestis* strains, including CO92 and KIM, fail to utilize this carbohydrate. Following a detailed comparison, we propose that YP1470 is a key gene related to melibiose metabolism. YP1470 encodes a 435 a.a (amino acid pro-

Table 6. Two-component systems identified in the chromosome of strain 91001.

Histidine protein kinase sensor		Response regulator protein		note
gene ID	gene name	gene ID	gene name	
YP0024	*cpxB*	YP0023	*ntrC*	
YP0073	*cpxA*	YP0074	*cpxR1*	
YP0136	*envZ*	YP0137	*ompR1*	
YP0409*	*baeS1*	YP0408	*citB1*	
YP1530*	*barA*	YP1528	*uvrY*	
YP0575	*basS*	YP0576	*basR*	
YP0728	*phoR*	YP0727	*phoB*	
YP0920	*rcsC*	YP0919	*rcsB*	
YP0984*	*baeS2*	YP0983	*atoC1*	
YP1666*	*baeS3*	YP1667	*fimZ*	
YP1763	*phoQ*	YP1764	*phoP*	
YP1796*	*cheA*	YP1809	*cheB*	Paired
		YP1810	*cheY*	members
YP1847	*baeS4*	YP1848	*copR*	
YP2094	*rstB*	YP2093	*rstA*	
YP2492	*kdpD1*	YP2491	*kdpE*	
YP2541	*baeS5*	YP2543	*atoC2*	
YP2622	*baeS6*	YP2623	*cpxR2*	
YP2632	*baeS7*	YP2633	*ompR2*	
YP2718	*baeS8*	YP2719	*ompR3*	
YP3328*	*baeS9*	YP3327	*ompR4*	
YP3370	*uhpB*	YP3371	*uhpA2*	
YP3591	*creC*	YP3592	*creB*	
YP3809*	*arcB*	YP3725	*arcA*	
YP1490	*qseS1*	YP0397	*lytT*	
YP1704	*pyrB*	YP1291	*psaE*	
YP2330	*atoS*	YP1464	*uhpA1*	
YP2518	*creC*	YP1972	*hnr*	Orphan
YP0918	*yehU*	YP2131	*tyrR*	members
		YP2140	*pspF*	
		YP2664	*narP*	
		YP3024	*atoC3*	
		YP3047	--	

*: Genes probably encodes hybrid sensory kinases with both hybrid sensory kinase domain and response regulator receiver domain. YP1796 has two pairing regulators (YP1809 and YP1810).

tein with 12 transmembrane helices, whose structure is quite similar to the melibiose carrier MelB protein in *E. coli*, which is responsible for carrying extracellular melibiose molecules into the bacteria cells.[33] The counterparts in CO92 and KIM are both disrupted by an IS 285 element at nt_{77}. PCR screening in 257 strains determined that all the *Microtus* strains have an intact YP1470 gene, and 31 strains isolated from Tianshan, Xinjiang also have intact YP1470; while the counterpart CDS in others are all disrupted by IS 285 (unpublished data). The role of YP1470 in the melibiose metabolism of *Y. pestis* needs to be verified by mutant analysis.

4.6. Two-component systems

Bacteria have evolved sophisticated sensory mechanisms and intracellular signal pathways in order to respond to a large number of extracellular signals in their continuously changing surroundings. Two-component systems are a basic stimulus-response coupling mechanism used by bacteria to sense and respond to changing environmental conditions. This sophisticated signaling system has been widely found in prokaryotes and eukaryotes; its prototypical system comprises a histidine protein kinase sensor (HK) containing a conserved kinase core that senses the environmental stimulus, and a response regulator protein (RR) containing a regulatory domain.[34] A total of 61 CDS were identified as putative members of the two-component signal transducers in strain 91001, as shown in Table 6. The number of two-component members in 91001 is close to that of *E. coli* (62) and a little smaller than that of *Synechocystis* sp. strain PCC 6803 (80).[35,36]

As shown in Table 6, 47 members of a two-component system were identified as 23 cognate pairs of putative cognate sensor/regulator, of which *cheA* (YP1796) has two paring regulators, *cheB* (YP1809) and *cheY* (YP1810). In most cases, the cognate sensor/regulator pairs are located next to each other on the chromosome and are most likely in the same transcriptional orientation except for *arcB*/*arcA*, *rcsC*/*rcsB* and *barA*/*uvrY*. The order of sensor and response regulator of cognate pairs on the chromosome appears to be random, which is quite similar to the case in *E. coli* with approximately half of sensor genes located upstream of the response regulator gene and half downstream.[35] Some cognate pairs need further verification, and we take these cognate pairs simply because they are adjacent on the chromosome.

Seven genes were identified as encoding possible hybrid sensory kinases. Hybrid sensor proteins have more complex architectures and functions, and they contain both a sensor histidine kinase domain and a response regulator receiver domain. The additional complexity of the phosphorelay system may provide for multiple regulatory checkpoints as well as a means of communication between individual signaling pathways.[37] BarA is a hybrid sensor protein and its analogue in *E. coli* had always been taken as an orphan without a functional partner. However, recently it was demonstrated that BarA and UvrY constitute a two-component system associated with the control of energy metabolism, although they are apart from each other in the chromosome.[38,39]

Response regulator PsaE (YP1291) is an isolated element without a known functional partner, whose function is to positively regulate the downstream gene *psaA* encoding the virulence protein pH 6 antigen.[40] Another established virulence-related pair of genes is *phoQ*/*phoP*. The isogenic *phoP* mutant of *Y. pestis* showed a reduced ability to survive in macrophages and under conditions of low pH and oxidative stress *in vitro*. The mean lethal dose of the *phoP* mutant in mice increased 75-fold in comparison with that of the wild-type strain.[41] The PhoPQ regulatory system controls the lipo-oligosaccharide (LOS) modification, which may be also required for survival of *Y. pestis* within the mammalian and/or flea host.[42]

The *baeS2*/*atoC1* (YP0983/YP0984) is absent from KIM and CO92. These two genes are located in the DFR4 according to Radnedge's study, and they are absent from some strains of *Mediaevalis*, *Antiqua* and all of *Orientalis*.[26] Two histidine protein kinase sensors, *baeS3* (YP1666) and *baeS8* (YP2718) have apparently become pseudogenes in CO92 due to frameshift mutations, and there also is a frameshift in *baeS8* in strain KIM. It has been found that the frameshift mutation of *baeS3* is only present in *Y. pestis* strains of biovar *Orientalis* and not in those of *Antiqua* and *Mediaevalis*. YP1666 is located in the 102-kb *pgm* locus and together with its cognate response regulator *fimZ* encode a putative two-component system similar to the BvgAS regulatory system of *B. pertussis*.[31] The two-component system BvgAS positively controls transcription of the virulence genes of *B. pertussis* and *B. bronchiseptica*, which include several genes for toxins and adhesins. On the other hand, the BvgAS system negatively controls the expression of a poorly characterized set of genes, the so-called virulence repressed genes.[43,44] There is still little evidence to explain the role of this BvgAS-like system in *Y. pestis*.

The gene *atoS1* (YP1490) is a hybrid histidine protein kinase containing both sensor kinase domain and response regulator domain. This gene was disrupted by IS100 in CO92 and there is a frameshift within a homopolymeric tract of 7G in the gene of strain KIM. The *uhpB* (YP3370), which is disrupted by IS 100 in CO92 and KIM, constitutes a two-component system with its cognate response regulator UhpA2. In *E. coli*, UhpAB form a signal transmitter cassette with UphC, controlling the expression of the phosphate transporter UbpT.[45]

Two-component systems serve as a basic stimulus-response coupling mechanism to allow organisms to sense and respond to changes in diverse conditions. For pathogenic bacteria, two-component systems are essential for sensing the changing environments while infecting hosts by helping them avoid the host's immune response. However, losing some two-component systems may increase the bacterial virulence, which suggests that some two-component systems negatively regulate bacterial virulence gene.[46,47] Therefore, whether deletion or inactivation of the two-component systems accounts for the virulence in strains CO92 and KIM needs further investigation.

4.7. Quorum sensing system

A further layer of microbial sensing and response mechanisms has been recently uncovered in the form of cell-to-cell communication via the use of small signaling molecules, which has been termed a "quorum sensing system." N-Acyl-homoserine lactones (AHSL) are usually employed as signals to control cell density during the growth of Gram-negative bacteria.[48] It is now known that many of the species belonging to the genus *Yersinia* express quorum-sensing systems. Throup et al. first identified YenI/YenR as a quorum sensing system in *Y. enterocolitica*.[49] Genes encoding LuxRI homologues (YpsR/I and YtbR/I) have also been identified in *Y. pseudotuberculosis*. Mutations in *ypsI* or *ypsR* indicate that this quorum-sensing regulon is involved in temperature-dependent control of motility and cellular aggregation of *Y. pseudotuberculosis*.[50] In the chromosome of strain 91001, we also identified two quorum sensing systems, *ypeI*/*ypeR* (YP2275/YP2276) and *yspI*/*yspR* (YP3454/YP3455). These two regulons are quite similar to their counterparts in *Y. pseudotuberculosis*, and they are all intact in strains

Table 7. Overview of comparison of 91001 with published *Y. pestis* plasmids sequences.

Plasmid	pPCP1			pCD1				pMT1			
Accession Number	AE017046	AL109969	AF053945	AE017043	AL117189	AF053946	AF074612	AE017045	AL117211	AF053947	AF074613
Source	91001	CO92	KIM	91001	CO92	KIM5-D45	KIM5	91001	CO92	KIM5-D46	KIM10+
Length (bp)	9609	9612	9612	70159	70305	70504	70559	106642	96210	100984	100990
G+C content	45.26%	45.27%	45.28%	44.85%	44.84%	44.81%	44.81%	50.31%	50.23%	50.[illegible]%	50.[illegible]%
CDS*	10	9	5	98	97	70	76	133	103	78	115
pseudogenes	0	0	0	13	8	6	2	6	3	0	0
Coding density	61.1%	57.2%	44.1%	87.9%	81.4%	70.0%	64.6%	93.4%	88.8%	68.4%	89.5%
Average gene length(bp)	587	621	848	620	643	771	600	760	835	886	786
IS 100 copy	1	1	1	1	1	1	1	3	2	2	2
IS 285 copy	0	0	0	1 partial	1 partial	2 partial	2 partial	1	1	1	1
IS 1541 copy	0	0	0	0	0	0	0	1	1	1	1

*: Differences of CDS number in counterpart plasmids are mainly caused by different genome annotation standards, which also influence the coding density and average gene length.

CO92 and KIM.

4.8. in silico *comparison of plasmid pPCP1*

Typical *Y. pestis* strains contain three plasmids, pPCP1, pCD1 and pMT1, which have all been reported to play significant roles in different stages of *Y. pestis* pathogenesis.[51–53] In this study, we performed a detailed comparison between pPCP1, pCD1 and pMT1 from strain 91001 and their previously published counterparts, shown in overview in Table 7. Plasmid pPCP1 is a virulence-related plasmid, which encodes the putative *Y. pestis*-specific adhesin/invasion, plasminogen activator (Pla); Pla has been proven essential for effectively invading human epithelial and endothelial cells, which plays a vital role in establishing subcutaneous infection.[54] Plasmid pPCP1 sequences of the three strains (91001, CO92 and KIM) are nearly identical; however, due to the differences in annotation criteria of different sequencing centers, the coding density and average gene length of these three pPCP1 entries vary dramatically. There are six single nucleotide polymorphisms (SNPs) in the three plasmids, and three of the SNPs are deletions or mutations in mononucleotide repeat regions. The mutations in mononucleotide repeats caused by a deficiency in post-synthesis mismatch repair mechanism had been thought of as a kind of adaptive mutation in the bacterial genome.[55] Interestingly, only one of the six point mutations is located in the coding area, which results in a 279_{Thr}–279_{Ile} mutation in the important virulence factor Pla protein. As this mutation involves the substitution of a hydrophilic hydroxyl-amino acid to a nonpolar amino acid, it is worthwhile to perform further study to clarify the possible relationship between this point mutation and the ability of *Y. pestis* to infect humans.

4.9. in silico *comparison of plasmid pCD1*

Plasmid pCD1 is a common virulent plasmid shared by the three pathogenic *Yersinia* species, and it is termed pYV and pIB in the enteropathogenic bacteria *Y. enterocolitica* and *Y. pseudotuberculosis*, respectively. This plasmid harbors a gene cluster named LCRS (low calcium response stimulons) which can secrete virulent factors through a type III secretory system into host cells when coming into contact with them.[56] Plasmid pCD1 of 91001 is slightly shorter than those of reference strains (Table 7). Compared with the two pCD1 plasmids from strain KIM, there is a 212-bp (partial IS 285) deletion between *yopM* and *yopD* in strain 91001, which is also the case in strain CO92. Another major deletion in 91001 pCD1 is located in the gene *yopM*, which is 126 bp shorter than those of strains CO92 and KIM. Because of the IS 100- mediated rearrangements, the structure of the four pCD1 entries varies a little among the strains.

The LCRS elements are nearly identical in the four pCD1 plasmids. The most significant variation of LCRS components is that YopM, an important cytotoxin effector of type III system, is 42 amino acids shorter than the reference counterparts. YopM is an acidic protein able to bind to thrombin, causing the virulence of *yopM* mutant strains to decrease 1000-fold compared to wild-type strains.[57] Typical YopM molecules of *Y. pestis* are 409 a.a long with 15 duplicated leucine rich regions (LRRs);[57] due to the 42 a.a deletion, the YopM molecule of 91001 only possesses 13 LRRs with 367 a.a in length. The number of amino acid residues and LRR repeats are the same with YopM of *Y. enterocolitica* (accession number

NP_052388), and the similarity of amino acids between them is 95%. By PCR screening, we discovered that all of the ten strains isolated from *Microtus brandti* have this deletion in the *yopM* gene (unpublished data). Bla et al. also discovered heterogeneity in the YopM proteins of the *Y. enterocolitica* and *Y. pseudotuberculosis*, and they further concluded that the heterogeneity in the YopM protein might not alter the virulence of *Y. pestis* strains.[58] A previous study revealed that mutants with these 22 a.a and 20 a.a LRR deletions in YopM produced no decrease in thrombin-binding activities compared with wild-type strains.[59]

There are also other mutations in certain LCRS elements in 91001. LcrV is the only protective antigen in LCRS, which acts as a bifunctional molecule of regulator and antihost factor.[60] The *lcrV* gene of 91001 is identical to that of *Y. pestis* strain Pestiodes F (accession number, AF067309).[61] These two *lcrV* genes are 16 bp shorter than those from other *Y. pestis* strains, and this deletion is caused by two direct repeats (ATGCAAG) at the 3′ terminus of *lcrV* gene. Pestoides F strain was isolated from vole and although it does not harbor plasmid pPCP1, it is fully virulent by the aerosol challenge.[62] Strain 91001 is also lethal to mice, thus the deletion in *lcrV* gene does not appear to decrease the lethality of these strains in mice. There is still no evidence whether this loss affects the host range of strain 91001. Another case is YopN, a secretory protein acting as calcium sensor.[63] There is a substitution in YopN of 91001 (52_{Phe}–52_{Ile}). Another mutation in LCRS of 91001 occurred in *yopJ*, which encoded a cytotoxin effector inducing *in vitro* apoptosis. There is a Gly-Glu mutation in *yopJ* in 91001, and this will lead to a Gly-Glu substitution in the corresponding position of the YopJ protein in strain 91001.

4.10. In silico *comparison of plasmid pMT1*

Plasmid pMT1 is a *Y. pestis*-specific plasmid which encodes two major virulence-related factors: F1 capsular protein, which can help *Y. pestis* escape from phagocytosis of the host immune system, and *Yersinia* murine toxin (Ymt), which is essential for transmission of *Y. pestis* by flea vectors.[64] As shown in Table 7, all four of the pMT1 plasmids contain one copy of IS 1541 and one copy of IS 285, while that of strain 91001 has three copies of IS 100 elements.

The full length of pMT1 in strain 91001 is 106,642 bp, about 6–10 kb larger than the other three pMT1 plasmids.[51,52,65] The 5.7-kb fragment of 91001, which is absent in pMT1 of *Mediaevalis* KIM strain, is 99% similar to the corresponding region of plasmid pHCM2 of *Salmonella enterica* serovar Typhi,[53] which suggests the origin of this fragment. Plasmid pMT1 from *Orientalis* CO92 has an additional IS-mediated 6.3-kb fragment deletion, and this fragment is also highly similar to plasmid pHCM2. Plasmid pMT1 of strain 91001 also lacks two segments (around 340 bp and 700 bp) common to KIM and CO92. The 340-bp region is highly homologous to part of plasmid pHCM2, and this deletion in strain 91001 leads to a 112 a.a deletion in the coded membrane protein. The 700-bp region shows no similarity to any sequence in the NCBI database.

The major different fragments in the four pMT1 entries are most closely related to plasmid pHCM2, which implies the evolution of pMT1. The ancestral pMT1 plasmid might have evolved from a pHCM2-like plasmid, and obtained some virulence-related genes (*ymt* and *calf* operon) by lateral transfer during the evolution process. Plasmid pMT1 from strain 91001 has retained more pHCM2-like sequences, but it has also lost some pHCM2-like sequences and the sequences common to pMT1 of strain CO92 and KIM (such as YPMT1.73). As *Orientalis* strains are newly occurred, it seems that plasmid pMT1 has undergone successive reductive evolution to simplify the genome structure. Interestingly, based on the reductive evolutionary hypothesis, although 91001 and KIM are both *Mediaevalis* strains, pMT1 of 91001 seems to resemble the ancestral pMT1 plasmid more than that of KIM, and it might have evolved in a different lineage from KIM.

Previous data revealed rearrangements mediated by IS elements in different pMT1 sequences.[53] Figure 3 portrays detailed rearrangement events in four pMT1 plasmids. Ignoring the fragment deletions, architectures of pMT1 from strains 91001 and KIM (accession number AF074611) are quite similar. However, another pMT1 entry from strain KIM derivate KIM5-D[27] (accession number AF053947) has undergone a 24-kb fragment inversion, which is flanked by two opposite IS 200 elements. Plasmid pMT1 of strain 91001 and CO92 share a common IS element insertion, which disrupted the gene coding for the alpha subunit of DNA polymerase III; while this gene is intact in both KIM strains. All of these observations suggest that rearrangement of pMT1 occurs at high frequency.

These two established virulent factors (F1 antigen and Ymt) in the four pMT1 entries have no differences, suggesting that they sustained tougher selective pressure in the life cycle of *Y. pestis* and remained identical in evolution.

4.11. Cryptic plasmid

As well as the above three known plasmids, some *Y. pestis* strains harbor more diverse plasmid profiles. Filippov et al. studied 242 *Y. pestis* strains isolated from various natural plague foci of former U.S.S.R. and other countries, and shown that twenty strains (8%) of them harbored additional cryptic plasmids, mostly about 20 MDa in size.[66] A cryptic plasmid about 45.5 kb, dimer of a 9.5-kb plasmid pPCP1, was found in *Y. pestis*

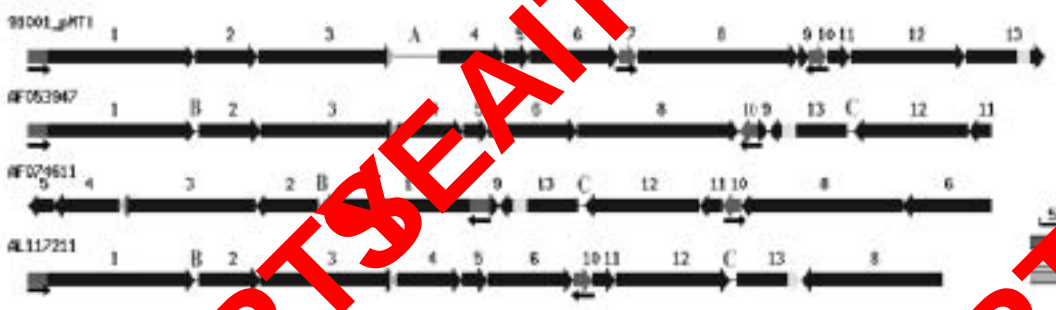

Figure 3. Comparison of four published pMT1 entries of *Y. pestis* to show rearrangements among them. Each pMT1 starts from the first base according to NCBI deposited sequences, and they are portrayed in scale. The numbered solid blocks with arrows represent the same fragments in different entries, and the arrows indicate the orientation. IS elements are also shown in different grades of gray. The narrow arrows represent the orientation of IS 100. Line signed as "A" is the 91001-specific fragment, While "B" and "C" are the fragments common to strains CO92 and KIM, which are absent in strain 91001.

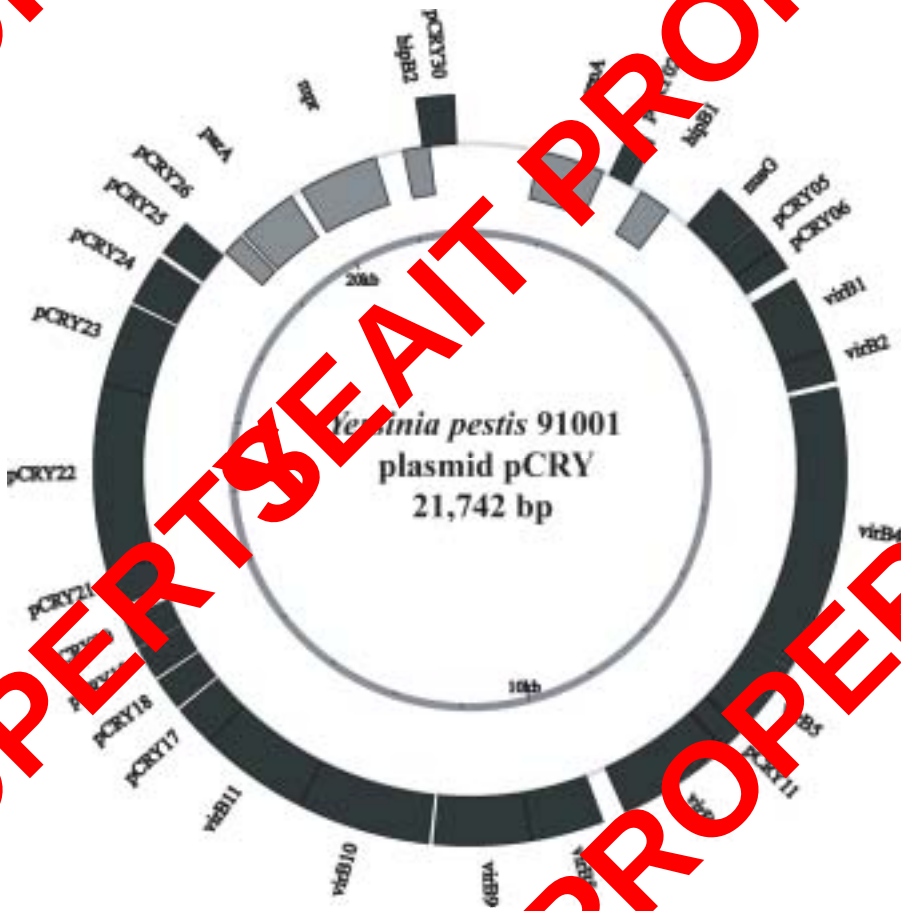

Figure 4. Circular illustration of plasmid pCRY of strain 91001. The inner circle with scale indicates the length of pCRY. Dark blocks represent the genes on the plus strand and the gray ones represent those on the minus strand. Note the coding bias on the plus strand.

strains isolated from the western United States.[67] A 6-kb cryptic plasmid has been recovered from *Y. pestis* isolated from regions of Yunnan province in China.[68]

The 21,742-bp plasmid pCRY is a novel plasmid identified in this study, and we termed it pCRY for its cryptic function. G+C% of plasmid pCRY is 49.1%, which is slightly different from the other three plasmids. We arbitrarily assigned the first base of string "TCGTTCCACT" as the "origin" of this plasmid. A total of 30 genes were predicted in pCRY as shown in Fig. 4 and Table 8. It is very interesting that pCRY shows abnormal coding bias on the plus strand, and 24 of the predicted genes located on the plus strand.

BLASTN homology search results revealed that a re-

Table 8. Gene list of plasmid pCRY in strain 91001.

gene ID	gene name	length	predicted coding products	coding strand
pCRY01	repA	714 bp	putative RepA protein	minus
pCRY02		234 bp	hypothetical protein	plus
pCRY03	hipB1	465 bp	putative transcriptional regulators	minus
pCRY04		474 bp	transcription antiterminator	plus
pCRY05		270 bp	hypothetical protein	plus
pCRY06		219 bp	putative ATP/GTP-binding protein remnant	plus
pCRY07	virB1	711 bp	Type IV secretory pathway, VirB1 components	plus
pCRY08	virB2	306 bp	Type IV secretory pathway, VirB2 component, putative mating pair formation protein TraC	plus
pCRY09	virB4	2685 bp	Type IV secretory pathway, VirB4 components	plus
pCRY10	virB5	705 bp	Type IV secretion system, component VirB5	plus
pCRY11		228 bp	hypothetical protein	plus
pCRY12	virB6	1074 bp	Type IV secretory pathway, VirB6 components	plus
pCRY13	virB8	684 bp	Type IV secretion system, component VirB8	plus
pCRY14	virB9	909 bp	Type IV secretory pathway, VirB9 components	plus
pCRY15	virB10	1251 bp	Type IV secretory pathway, VirB10 components	plus
pCRY16	virB11	1026 bp	Type IV secretory pathway, VirB11 components, and related ATPases involved in archaeal flagella biosynthesis	plus
pCRY17		399 bp	hypothetical protein	plus
pCRY18		306 bp	hypothetical protein	plus
pCRY19		306 bp	putative dopa decarboxylase protein remnant	plus
pCRY20		294 bp	hypothetical protein	plus
pCRY21		345 bp	hypothetical protein	plus
pCRY22		1452 bp	putative mobilization mobB protein	plus
pCRY23		765 bp	putative mobilization protein mobC	plus
pCRY24		474 bp	micrococcal nuclease (thermonuclease) homologs	plus
pCRY25		342 bp	putative membrane prtotein	plus
pCRY26		234 bp	hypothetical protein	minus
pCRY27	parA	648 bp	ATPases involved in chromosome partitioning	minus
pCRY28	mpr	861 bp	zinc metalloproteinase Mpr protein	minus
pCRY29	hipB2	282 bp	predicted transcriptional regulators	minus
pCRY30		360 bp	putaive membrane protein	plus

gion of pCRY (nucleotide number 165–265) was quite similar to that of several plasmids with greater than 89% identity. These plasmids are mostly harbored by Enterobactericeae members, such as plasmid p307 in *E. coli*, plasmid pGSH500 in *Klebsiella pneumoniae*, plasmid pYVe439-80 in *Y. enterocolitica* and plasmid pCP301 in *Shigella flexneri*. The similarity of these regions in different bacteria implies that they might act as *cis*-acting elements in these plasmids as there is no gene predicted in these regions. All the above plasmids belong to theta-replicon A plasmids. This kind of plasmid can encode RepA protein independently, and there are varying numbers of DnaA Box elements around the *repA* gene, to which the RepA protein binds. An A+T rich region can act as replication origin.[69,70] We annotated a *repA* gene in pCRY based on BLASTP, COG and InterPro analysis. The "TTATCACA" sequence downstream the *repA* gene is identical to the six bases of 3′ end of DnaA Box R1 (TTATCCACA) in *E. coli*. It is probably the binding site of the RepA protein.[71] Downstream of the *repA* gene, there is an A+T rich region (nucleotide number 370–690 in pCRY, A+T 59%), and A+T% of region spanning nucleotides 400–500 is even higher (65%). Figure 5 illustrates the G+C% plot of the first 1000 bp of pCRY, which clearly shows the A+T% higher region. The A+T rich region might contain the replication origin of pCRY.[72] We also identified a *parA* gene in the pCRY plasmid. The *parA-parB* genes were found to be responsible for the partition of replicated plasmids into daughter cells. Although we failed to identify a *parB* gene in plasmid

Figure 5. G+C% plot of the first 1,000 bp of plasmid pCRY. The bottom line with scale indicates the overall 1,000 bp. Note the abnormally A+T rich region upstream of *repA* gene, which might contains the replication origin of plasmid pCRY.

pCRY. There is an unknown gene right downstream *parA*. As plasmid pCRY encodes its own replication and partition systems, it might be able to maintain itself in different bacteria as an independent genetic element, and it might have been incorporated into *Y. pestis* strain 91001 by occasional lateral transfer from unidentified bacteria. We designed primers targeting the *repA* gene and screened 257 strains of *Y. pestis* isolated in China by PCR amplification. Only 11 strains showed positive amplification (unpublished data). Therefore pCRY might be an atypical plasmid in *Y. pestis*, and it might contribute little to the common life cycle of *Y. pestis*.

We also identified a type IV secretory system coding a gene cluster in pCRY, which includes 20 genes. Although quite a few pathogens have the type IV system,[73] this is the first report of this system in *Y. pestis*. A type IV system is an essential virulence factor in *Bartonella* for establishing intraerythrocytic infection.[73] However, the type IV system on plasmid pCRY lacks two important genes, *virB3* and *virB7*. We do not know the function of the type IV system in pCRY.

4.12. Concluding remarks

The genome sequence of 91001, a strain with unique pathogenicity and carbohydrate metabolism, sheds light on the mysteries of *Y. pestis*. Strain 91001 and others like it isolated from *Microtus* are supposed to be avirulent to humans, while they are highly lethal to mice. By comparing the genome sequence of this strain with those of the fully virulent *Y. pestis* strains (CO92 and KIM), we have been able to find clues how *Y. pestis* might have evolved from a single host pathogen to a multihost pathogen. Following extensive analysis of plasmid structure, pseudogene distribution, gene level comparison, the *pgm* locus characteristics, nitrate reduction-negative mechanism, genes related to arabinose and melibiose metabolism, and chromosome architectures, we can safely draw a conclusion that 91001 evolved from ancestral *Y. pestis* through a different lineage. The whole genome microarray-based comparative genomic research carried out by us has also proved that *Microtus* strains should be reclassified into a novel biovar of *Y. pestis*, biovar *Microtus* (unpublished data). The ancestral *Y. pestis* strain was probably virulent only to rodents, then some strains occasionally obtained gene(s) by horizontal gene transfer, and were able to cross species barriers and broaden their host range.[74] Thus the 33-kb prophage-like fragment, absent in 91001, is a candidate that might determine the ability to infect humans of fully virulent strains. However, mutations in the established virulence-related genes, such as *yopM* and *ail* can not be ignored either, as well as some interesting strain-specific pseudogenes. Our paper identifies some candidate DNA regions and factors determining the appalling lethality of *Y. pestis* to humans, which will be of help in developing an efficient vaccine against plague.

Acknowledgments: We thank the sequencing team of the Beijing Genomics Institute for their contribution to genome library construction and DNA sequencing. We are also grateful to Dr. David Bastin (Tianjin Biochip Co., Tianjin, China) and Mr. Qi Guo (Beijing Genomics Institute, Chinese Academy of Sciences, China) for careful reading of the manuscript and valuable suggestions. We wish to express our respect and appreciation to Chinese researchers for their excellent works on the ecology and epidemiology of the plague in China.

References

1. Perry, R. D. and Fetherston, J. D. 1997, *Yersinia pestis*—etiologic agent of plague, *Clin. Microbiol. Rev.*, **10**, 35–66.

2. Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B. et al. 2001, Genome sequence of *Yersinia pestis*, the causative agent of plague, *Nature*, **413**, 523–527.
3. Deng, W., Burland, V., Plunkett, G. 3rd, Boutin, A., Mayhew, G. F., Liss, P. et al. 2002, Genome sequence of *Yersinia pestis* KIM, *J. Bacteriol.*, **184**, 4601–4611.
4. Fan, Z., Luo, Y., Wang, S., Jin, L., Zhu, [illegible] L., J. et al. 1995, *Microtus brandti* plague in the [illegible] Grassland was inoffensive to human (in Chinese), *Chin. J. Control Endemic Dis.*, **10**, 56–57.
5. Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K. et al. 2001, Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12, *DNA Res.*, **8**, 11–22.
6. Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S. et al. 2002, A complete sequence of the *T. tengcongensis* genome, *Genome Res.*, **12**, 689–700.
7. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998, Base-calling of automated sequencer traces using phred. I. Accuracy assessment, *Genome Res.*, **8**, 175–185.
8. Ewing, B. and Green, P. 1998, Base-calling of automated sequencer traces using phred. II. Error probabilities, *Genome Res.*, **8**, 186–194.
9. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389–3402.
10. Wang, J., Wong, G. K., Ni, P., Han, Y., Huang, J., Zhang, J. et al. 2002, RePS: a sequence assembler that masks exact repeats identified from the shotgun data, *Genome Res.*, **12**, 824–831.
11. Gordon, D., Abajian, C., and Green, P. 1998, Consed: a graphical tool for sequence finishing, *Genome Res.*, **8**, 195–202.
12. Liu, S. L., Hessel, A., and Sanderson, K. E. 1993, The *XbaI-BlnI-CeuI* genomic cleavage map of *Salmonella enteritidis* shows an inversion relative to *Salmonella typhimurium* LT2, *Mol. Microbiol.*, **10**, 655–664.
13. Badger, J. H. and Olsen, G. J. 1999, CRITICA: coding region identification tool invoking comparative analysis, *Mol. Biol. Evol.*, **16**, 512–524.
14. Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. 1999, Improved microbial gene identification with GLIMMER, *Nucleic Acids Res.*, **27**, 4636–4641.
15. Frishman, D., Mironov, A., Mewes, H. W., and Gelfand, M. S., Combining diverse evidence for gene recognition in completely sequenced bacterial genomes, *Nucleic Acids Res.*, **26**, 2941–2947.
16. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S. et al. 2001, The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res.*, **29**, 22–28.
17. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D. et al. 2002, InterPro, an integrated documentation resource for protein families, domains and functional sites, *Brief Bioinform.*, **3**, 225–235.
18. Lowe, T. M. and Eddy, S. R. 1997, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.*, **25**, 955–964.
19. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. 2001, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol.*, **305**, 567–580.
20. Benson, G. 1999, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.*, **27**, 573–580.
21. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. et al. 2000, Artemis: sequence visualization and annotation, *Bioinformatics*, **16**, 944–945.
22. Odaert, M., Berche, P., and Simonet, M. 1996, Molecular typing of *Yersinia pseudotuberculosis* by using an IS200-like element, *J. Clin. Microbiol.*, **34**, 2231–2235.
23. McDonough, K. A. and Brubaker, R. R. 1997, Homology with a repeated *Yersinia pestis* DNA sequence IS100 correlates with pesticin sensitivity in Yersinia pseudotuberculosis, *J. Bacteriol.*, **179**, 2081–2085.
24. Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A. and Carniel, E. 1999, *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*, *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14043–14048.
25. Doll, J. M., Zeitz, P. S., Ettestad, P., Bucholtz, A. L., Davis, T., and Gage, K. 1994, Cat-transmitted fatal pneumonic plague in a person who travelled from Colorado to Arizona, *Am. J. Trop. Med. Hyg.*, **51**, 109–114.
26. Radnedge, L., Agron, P. G., Worsham, P. L., and Andersen, G. L. 2002, Genome plasticity in *Yersinia pestis*, *Microbiology.*, **148**, 1687–1698.
27. Gentry-Weeks, C., Coburn, P. S., and Gilmore, M. S. 2002, Phages and other mobile virulence elements in gram-positive pathogens, *Curr. Top Microbiol Immunol.*, **264**, 79–94.
28. Boyd, E. F. and Brussow, H. 2002, Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved, *Trends Microbiol.*, **10**, 521–529.
29. Brubaker, R. R. 1969, Mutation rate to nonpigmentation in *Pasteurella pestis*, *J. Bacteriol.*, **98**, 1404–1406.
30. Gray, Y. H. 2000, It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements, *Trends Genet.*, **16**, 461–468.
31. Buchrieser, C., Rusniok, C., Frangeul, L., Couve, E., Billault, A., Kunst, F. et al. 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.
32. Schleif, R. 2000, Regulation of the L-arabinose operon of *Escherichia coli*, *Trends Genet.*, **16**, 559–565.
33. Matsuzaki, S., Weissborn, A. C., Tamai, E., Tsuchiya, T., and Wilson, T. H. 1999, Melibiose carrier of *Escherichia coli*: use of cysteine mutagenesis to identify the amino acids on the hydrophilic face of transmembrane helix 2, *Biochim. Biophys. Acta.*, **1420**, 63–72.
34. Stock, A. M., Robinson, V. L., and Goudreau, P. N. 2000, Two-component signal transduction, *Annu. Rev. Biochem.*, **69**, 183–215.
35. Mizuno, T. 1997, Compilation of all genes encoding

two-component phosphotransfer signal transducers in the genome of *Escherichia coli*, *DNA Res.*, **4**, 161–168.
36. Mizuno, T., Kaneko, T., and Tabata, S. 1996, Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803, *DNA Res.*, **3**, 407–414.
37. Hoch, J. A. 2000, Two-component and phosphorelay signal transduction, *Curr. Opin. Microbiol.*, **3**, 165–170.
38. Pernestig, A. K., Georgellis, D., Romeo, T., Suzuki, K., Tomenius, H., Normark, S. et al. 2003, The *Escherichia coli* BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources, *J. Bacteriol.*, **185**, 843–853.
39. Pernestig, A. K., Melefors, O., and Georgellis, D. 2001, Identification of UvrY as the cognate response regulator for the BarA sensor kinase in *Escherichia coli*, *J. Biol. Chem.*, **276**, 225–231.
40. Price, S. B., Freeman, M. D., and Yeh, K. S. 1995, Transcriptional analysis of the *Yersinia pestis* pH 6 antigen gene, *J. Bacteriol.*, **177**, 5997–6000.
41. Lyst, P. C., Dorrell, N., Williams, K., Li, S. R., Green, M., Titball, R. W. et al. 2000, The response regulator PhoP is important for survival under conditions of macrophage-induced stress and virulence in *Yersinia pestis*, *Infect Immun.*, **68**, 3419–3425.
42. Hitchen, P. G., Prior, J. L., Oyston, P. C., Panico, M., Wren, B. W., Titball, R. W. et al. 2002, Structural characterization of lipo-oligosaccharide (LOS) from *Yersinia pestis*: regulation of LOS structure by the PhoP/Q system, *Mol. Microbiol.*, **44**, 1637–1650.
43. Bock, A. and Gross, R. 2001, The BvgAS two-component system of *Bordetella* spp.: a versatile modulator of virulence gene expression, *Int. J. Med. Microbiol.*, **291**, 119–130.
44. Yuk, M. H., Harvill, E. T., and Miller, J. F. 1998, The BvgAS virulence control system regulates type III secretion in *Bordetella bronchiseptica*, *Mol. Microbiol.*, **28**, 945–959.
45. Verhamme, D. T., Postma, B., W., Crielaard, W., and Hellingwerf, K. J. 2002, Cooperativity in signal transfer through the Uhp system of *Escherichia coli*, *J. Bacteriol.*, **184**, 4205–4210.
46. Parish, T., Smith, D. A., Kendall, S., Casali, N., Bancroft, G. J., and Stoker, N. G. 2003, Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*, *Infect Immun.*, **71**, 1134–1140.
47. Perez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B., and Martin, C., 2001, An essential role for phoP in *Mycobacterium tuberculosis* virulence, *Mol. Microbiol.*, **41**, 179–187.
48. Whitehead, N. A., Barnard, A. M., Slater, H., Simpson, N. J., and Salmond, G. P. 2001, Quorum-sensing in Gram-negative bacteria, *FEMS Microbiol. Rev.*, **25**, 365–404.
49. Throup, J. P., Camara, M., Briggs, G. S., Winson, M. K., Chhabra, S. R., Bycroft, B. W. et al. 1995, Characterisation of the yenI/yenR locus from *Yersinia enterocolitica* mediating the synthesis of two N-acyl homoserine lactone signal molecules, *Mol. Microbiol.*, **17**, 345–356.
50. Atkinson, S., Throup, J. P., Stewart, G. S., and Williams, P., 1999, A hierarchical quorum-sensing system in *Yersinia pseudotuberculosis* is involved in the regulation of motility and clumping, *Mol. Microbiol.*, **33**, 1267–1277.
51. Lindler, L. E., Plano, G. V., Burland, V., Mayhew, G. F., and Blattner, F. R. 1998, Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen, *Infect Immun.*, **66**, 5731–5742.
52. Hu, P., Elliott, J., McCready, P., Skowronski, E., Garnes, J., Kobayashi, A. et al. 1998, Structural organization of virulence-associated plasmids of *Yersinia pestis*, *J. Bacteriol.*, **180**, 5192–5202.
53. Prentice, M. B., James, K. D., Parkhill, J., Baker, S. G., Stevens, K., Simmonds, M. et al. 2001, *Yersinia pestis* pFra shows biovar-specific differences and recent common ancestry with a *Salmonella enterica* serovar Typhi plasmid, *J. Bacteriol.*, **183**, 2586–2594.
54. Lahteenmaki, K., Kukkonen, M., and Korhonen, T. K. 2001, The Pla surface protease/adhesin of *Yersinia pestis* mediates bacterial invasion into human endothelial cells, *FEBS Lett.*, **504**, 69–72.
55. Riley, M. S., Longerich, S., Gee, P., and Harris, R. S. 1994, Adaptive mutation by deletions in small mononucleotide repeats, *Science.*, **265**, 405–407.
56. Cornelis, G. R., Boland, A., Boyd, A. P., Geuijen, C., Iriarte, M., Neyt, C. et al. 1998, The virulence plasmid of *Yersinia*, an antihost genome, *Microbiol. Mol. Biol. Rev.*, **62**, 1315–1352.
57. Evdokimov, A. G., Anderson, D. E., Routzahn, K. M., and Waugh, D. S. 2001, Unusual molecular architecture of the *Yersinia pestis* cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit, *J. Mol. Biol.*, **312**, 807–821.
58. Boland, A., Havaux, S., and Cornelis, G. R. 1998, Heterogeneity of the *Yersinia* YopM protein, *Microb. Pathog.*, **25**, 343–348.
59. Hines, J., Skrzypek, E., Kajava, A. V., and Straley, S. C. 2001, Structure-function analysis of *Yersinia pestis* YopM's interaction with alpha-thrombin to rule on its significance in systemic plague and to model YopM's mechanism of binding host proteins, *Microb. Pathog.*, **30**, 193–209.
60. Fields, K. A. and Straley, S. C. 1999, LcrV of *Yersinia pestis* enters infected eukaryotic cells by a virulence plasmid independent mechanism, *Infect Immun.*, **67**, 4801–4813.
61. Adamson, D. M., Worsham, P. L., Hill, K. K., Klevytska, A. M., Jackson, P. J., Friedlander, A. M. et al. 2000, Diversity in a variable-number tandem repeat from *Yersinia pestis*, *J. Clin. Microbiol.*, **38**, 1516–1519.
62. Worsham, P. L. and Roy, C. 2003, Pestoides F, a *Yersinia pestis* strain lacking plasminogen activator, is virulent by the aerosol route, *Adv. Exp. Med. Biol.*, **529**, 129–131.
63. Forsberg, A., Viitanen, A. M., Skurnik, M., and Wolf-Watz, H. 1991, The surface-located YopN protein is involved in calcium signal transduction in *Yersinia pseudotuberculosis*, *Mol. Microbiol.*, **5**, 977–986.
64. Hinnebusch, B. J., Rudolph, A. E., Cherepanov, P., Dixon, J. E., Schwan, T. G., and Forsberg, A. 2002, Role

of Yersinia murine toxin in survival of *Yersinia pestis* in the midgut of the flea vector, *Science.*, **296**, 733–735.
65. Youngren, B., Radnedge, L., Hu, P., Garcia, E., and Austin, S. 2000, A plasmid partition system of the P1-P7par family from the pMT1 virulence plasmid of *Yersinia pestis*, *J. Bacteriol.*, **182**, 3924–3928.
66. Filippov, A. A., Solodovnikov, N. S., Lobanova, E. M., and Protsenko, O. A. 1990, Plasmid content in *Yersinia pestis* strains of different origin, *FEMS Microbiol. Lett.*, **55**, 45–48.
67. Chu, M. C., Dong, X. Q., Zhou, A., and Garon, C. F. 1998, A cryptic 19-kilobase plasmid associated with U.S. isolates of *Yersinia pestis* is a derivative of the 9.5-kilobase plasmid, *Am. J. Trop. Med. Hyg.*, **59**, 679–686.
68. Dong, X. Q., Lindler, L. E., and Chu, M. C. 2000, Complete DNA sequence and analysis of an emerging cryptic plasmid isolated from *Yersinia pestis*, *Plasmid.*, **43**, 144–148.
69. Bruand, C., Le Chatelier, E., Ehrlich, S. D., and Janniere, L. 1993, A fourth class of theta-replicating plasmids: the pAM beta 1 family from gram-positive bacteria, *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 11668–11672.
70. Novick, R. P., Iordanescu, S., Projan, S. J., Kornblum, J., and Edelman, I. 1989, pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator, *Cell.*, **59**, 395–404.
71. Pacek, M., Konopa, G., and Konieczny, I. 2000, DnaA box sequences as the site for helicase delivery during plasmid RK2 replication initiation in *Escherichia coli*, *J. Biol. Chem.*, **276**, 23639–23644.
72. Speck, C. and Messer, W. 2001, Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA, *EMBO J.*, **20**, 1469–1476.
73. Mattick, J. S. 2002, Type IV pili and twitching motility, *Annu. Rev. Microbiol.*, **56**, 289–314.
74. Woolhouse, M. E., Taylor, L. H., and Haydon, D. T. 2001, Population biology of multihost pathogens, *Science.*, **292**, 1109–1112.