Pseudogene accumulation might promote the adaptive microevolution of \textit{Yersinia pestis}

Zongzhong Tong, 1, 2 † Dongsheng Zhou, 1 † Yajun Song, 1 Ling Zhang, 1 Decui Pei, 1 Yanping Han, 1 Xin Pang, 1 Min Li, 3 Baizhong Cui, 3 Jin Wang, 1 Zhaobiao Guo, 1 Zhizhen Qi, 3 Lixia Jin, 3 Junhui Zhai, 1 Zongmin Du, 1 Jun Wang, 2 Xiaoyi Wang, 1 Jun Yu, 2 Jian Wang, 2 Peitang Huang, 1 Huanming Yang 2 and Ruifu Yang 1

1 Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, No. 20 Dongdajie, Fengtai, Beijing 100071, China
2 Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 101300, China
3 Qinghai Institute for Endemic Diseases Prevention and Control, Xining 811602, China

Received 22 May 2004
Accepted 15 November 2004

INTRODUCTION

The causative agent of plague, \textit{Yersinia pestis}, is a Gram-negative bacterium evolved from \textit{Yersinia pseudotuberculosis} 1500–20,000 years ago, shortly before the first known pandemics of human plague (Achtman et al., 1999; Skurnik et al., 2000). There have been three historical human plague pandemics, which have killed about 200 million people (Perry & Fetherston, 1997). Plague is circulating in most foci, although the human plague pandemic has been controlled. The recent outbreak of human plague and appearance of multidrug-resistant strains suggest that plague still poses a significant threat to public health (Galimand et al., 1997; Ratsitorahina et al., 2000).

The environments, host and vector of each plague focus differ markedly, and these are the driving forces that contribute to the final shape of bacterial genomes. To adapt to a different niche, \textit{Y. pestis} has to develop different mechanisms to change its genome, such as gene acquisition or loss and point mutation (Hinchliffe et al., 2003; Radnedge et al., 2002). The accumulation of these changes allowed \textit{Y. pestis} to diversify into different genotypes. The epidemiological, aetiological and virulence features are different for each genotype. It will be very useful to the control and prevention of plague if the genotype of an isolate can be
quickly determined and used to track down the origin of pathogens. The whole-genome sequences of two fully virulent *Y. pestis*, CO92 and KIM, were decoded in 2001 and 2002 (Deng et al., 2002; Parkhill et al., 2001). The complete genome sequence of another strain of *Y. pestis*, 91001, which is avirulent to humans, has been finished recently in our laboratory (Song et al., 2004). These genome data provide a good opportunity to study the polymorphisms of the *Y. pestis* genome. A large number of pseudogenes were identified in the genomes of the three *Y. pestis* strains. Gene inactivation is one of the mechanisms for *Y. pestis* to adapt to new niches (Kukkonen et al., 2004; Oyston et al., 2003; Perry et al., 1998; Simonet et al., 1996; Skurnik et al., 2000).

**METHODS**

**Bacterial strains.** There are 11 plague foci in China and many *Y. pestis* have been isolated from these foci in the past years. Two hundred and sixty isolates of *Y. pestis*, isolated between 1943 and 2001 from 10 natural plague foci in China, were included in this study (Table 1 and Supplementary Table S1, available in JMM Online). Four subfoci of the Marmota himalayana plague focus of the Qinghai-Tibet plateau were investigated separately and referred to as foci C, D, G and K. All of these isolates were collected by the Qinghai Center for Disease Prevention and Control, China. In addition, seven strains of *Y. pseudotuberculosis* (CMCC53518, CMCC53519, CMCC35220, CMCC53521, CMCC53522, ATCC29833 and CMCC53502) were included, the first five belonging to biotypes I to V, respectively. These strains were originally from the China Medical Culture Collection except for ATCC29833, which was purchased from the American Type Culture Collection, and their dates of isolation and origins are unknown. All the strains were grown in Luria–Bertani broth. The extraction of genomic DNA was performed using a conventional phenol/chloroform extraction method (Adair et al., 2000).

**Comparative analysis of pseudogenes and primer design.** The genome sequence data for *Y. pestis* CO92 (AL590842), KIM (AE009952) and 91001 (AE017042) were downloaded from GenBank. The nucleotide sequences of all predicted pseudogenes on the chromosomes of the three *Y. pestis* strains were extracted and compared with their alleles using BLAST and CLUSTAL_X to find detailed variations between different strains. (There are 149, 54 and 141 pseudogenes predicted in CO92, KIM and 91001, respectively). Primers were designed to discriminate between wild-type and mutant in each mutation site using Primer 3.0 (Tables 2 and 3).

**Allele-specific PCR.** The genomic DNAs of 260 *Y. pestis* isolates and seven *Y. pseudotuberculosis* strains were pre-arranged in 96-well plates with 10 ng per well. Each primer pair was pre-tested to ensure the specificity of amplification with CO92 and 91001 genomic DNA as template. The 30 μl PCR reaction mixture contained 50 mmol l⁻¹ KCl, 10 mmol l⁻¹ Tris/HCl (pH 8.0), 2-0 mmol l⁻¹ MgCl₂, 0-001 % gelatin, 0-1 % BSA, 100 μmol l⁻¹ each of dATP, dCTP, dGTP and dTTP, 0-3 μmol l⁻¹ of each primer, 1 unit of Taq DNA polymerase (MBI) and 10 ng of template DNA. The parameters for amplification were as follows: pre-denaturation at 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s and elongation at 72 °C for 1 min, and a final extension at 72 °C for 5 min to ensure the complete extension of the amplicons. Minor changes of cycling parameters were made for different primer pairs. After amplification, 15 μl of each PCR product was subjected to electrophoresis on a 1-2 % agarose gel and visualized by ethidium bromide staining and UV irradiation. There were positive, negative and blank controls in each 96-well plate.

**Data analysis.** For every primer pair, each isolate was scored as '+4' for positive amplification or '−' for negative amplification (Supplementary Table S2, available in JMM Online). Then the amplification results of all the primer pairs that were used to study the same mutation site were integrated to determine the wild-type or mutant status for each isolate. For every mutation site that was analysed, a score of '1' was given for wild-type and '0' for mutant (Table 4). The similarity of mutation profiles was analysed using PHYLIP. The phylogenetic tree was reconstructed with *Y. pseudotuberculosis* as an outgroup (Fig. 3).

**RESULTS AND DISCUSSION**

**Mechanisms of gene inactivation**

There are three kinds of mutation that lead to gene inactivation: nonsense mutation, frameshift and disruption by insertion of insertion sequence (IS) elements. Fig. 1(a) shows the primers designed to discriminate wild-type and nonsense-mutation alleles. Primer pairs LW/R and LM/R should give positive and negative amplification results, respectively, for the wild-type, and the opposite results for the mutant. If both primer pairs give no product it is very likely that the target gene is absent from the genome, for gene acquisition and loss are common in *Y. pestis* (Hinchliffe et al., 2003; Radnedge et al., 2002). Fig. 1(b) shows an example of primers designed to discriminate wild-type and frameshift mutation alleles, in which a change was caused by a 19 bp fragment insertion. The 3° end of primer LM is located on the inserted fragment, while the primer LW flanks the two sides of the insertion site. The primer pairs LW/R and LM/SH should give negative and positive amplification results, respectively, in the mutant allele, but they should give the reverse results in the wild-type one.

The two disrupted parts of a gene that was interrupted by an IS element insertion either flank the two sides of the inserted IS element (mutant 1) or are located remotely due to genome rearrangement (mutant 2) (Fig. 1c). The primer pairs L1/R2 and L2/R1 should give identical amplicons in mutants 1 and 2, but no product in the wild-type. Primer pair L1/R1 should produce an amplicon for mutant 1 that is much longer than that for the wild-type, while giving no product in mutant 2.

**Pseudogene distribution in *Y. pestis* strains CO92, KIM and 91001**

Different annotators have taken different gene prediction rules, so there are many differences in definition and numbers of pseudogenes between the three *Y. pestis* strains. For example, YPO3775 was predicted as a pseudogene in CO92 due to IS100 element insertion and its allele (YP3274) was interrupted by IS100 and predicted as a pseudogene in 91001 too, but the homologous sequences of the two disrupted parts were predicted as two intact coding sequences in KIM (y0455, y0458). To employ more information, we took the longest sequence to be the whole length of a gene if they were...
### Table 1. Isolates of *Y. pestis* used in this study

<table>
<thead>
<tr>
<th>Focus code</th>
<th>Biovar</th>
<th>Main host</th>
<th>Geographic region (province)</th>
<th>No. of isolates*</th>
<th>Focus name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Antiqua</td>
<td>Marmota caudata</td>
<td>Xinjiang</td>
<td>11</td>
<td><em>Marmota caudata</em> plague focus of the Pamirs plateau</td>
</tr>
<tr>
<td>B</td>
<td>Antiqua</td>
<td>Marmota baibacina and Spermophilus undulatus</td>
<td>Xinjiang</td>
<td>30</td>
<td><em>Marmota baibacina-Spermophilus undulatus</em> plague focus of the Tianshan mountains</td>
</tr>
<tr>
<td>C</td>
<td>Antiqua</td>
<td>Marmota himalayana</td>
<td>Qinghai, Gansu and Tibet</td>
<td>41</td>
<td><em>Marmota himalayana</em> plague focus of the Qinghai-Gansu-Tibet grassland</td>
</tr>
<tr>
<td>D</td>
<td>Antiqua</td>
<td>Marmota himalayana</td>
<td>Qinghai and Gansu</td>
<td>16</td>
<td><em>Marmota himalayana</em> plague focus of the Qilian mountain</td>
</tr>
<tr>
<td>E</td>
<td>Antiqua</td>
<td>Apodemus chevrieri and Eothenomys miletus</td>
<td>Yunnan</td>
<td>10</td>
<td><em>Apodemus chevrieri-Eothenomys miletus</em> plague focus of the highland of Northwestern Yunnan province</td>
</tr>
<tr>
<td>F</td>
<td>Orientalis</td>
<td>Rattus flavipectus</td>
<td>Yunnan, Guangdong, Fujian, Guizhou and Guangxi</td>
<td>22</td>
<td><em>Rattus flavipectus</em> plague focus of the Yunnan-Guangdong-Fujian provinces</td>
</tr>
<tr>
<td>G</td>
<td>Antiqua</td>
<td>Marmota himalayana</td>
<td>Tibet</td>
<td>11</td>
<td><em>Marmota himalayana</em> plague focus of the Gangdici mountains</td>
</tr>
<tr>
<td>H</td>
<td>Antiqua</td>
<td>Spermophilus daurus</td>
<td>Heilongjiang, Jilin and Liaoning</td>
<td>30</td>
<td><em>Spermophilus daurus</em> plague focus of the Song-Liao plain</td>
</tr>
<tr>
<td>I</td>
<td>Mediaevalis</td>
<td>Meriones unguiculatus</td>
<td>Inner Mongolia, Ningxia and Hebei</td>
<td>21</td>
<td><em>Meriones unguiculatus</em> plague focus of the Inner Mongolian plateau</td>
</tr>
<tr>
<td>J</td>
<td>Mediaevalis</td>
<td>Spermophilus daurus alaschanicus</td>
<td>Gansu and Ningxia</td>
<td>20</td>
<td><em>Spermophilus daurus alaschanicus</em> plague focus of the Loess plateau in Gansu and Ningxia provinces</td>
</tr>
<tr>
<td>K</td>
<td>Mediaevalis</td>
<td>Marmota himalayana</td>
<td>Xinjiang</td>
<td>9</td>
<td><em>Marmota himalayana</em> plague focus of the Kunlun mountains</td>
</tr>
<tr>
<td>L</td>
<td>Mediaevalis</td>
<td>Microtus brandti</td>
<td>Inner Mongolia</td>
<td>10</td>
<td><em>Microtus brandti</em> plague focus of the Xilin Gol grassland</td>
</tr>
<tr>
<td>M</td>
<td>Mediaevalis</td>
<td>Microtus fuscus</td>
<td>Qinghai and Sichuan</td>
<td>29</td>
<td><em>Microtus fuscus</em> plague focus of the Qinghai-Tibet plateau</td>
</tr>
</tbody>
</table>

*For the year of isolation and isolate numbers see Supplementary Table S1 (available in JMM Online).
different in the three strains. For example, we took y0455 and y0458 as a pseudogene in KIM, and so the total number of pseudogenes present in Fig. 2 is greater than the predicted number of 54.

It was found that pseudogene profiles were different in the three Y. pestis strains (Fig. 2). Ninety-one genes were inactivated in all the three strains, while 22, 10 and 39 pseudogenes were specific to CO92, KIM and 91001, respectively. It is very interesting that 23 pseudogenes are shared by CO92 and KIM although they belong to different biovars, but only two pseudogenes are shared by 91001 and KIM while they belong to the same biovar. The distribution of pseudogenes suggests that 91001 and KIM might have originated differently.

All of these pseudogenes are non-essential for the survival of Y. pestis, but their inactivation may relate to the adaptation of Y. pestis to different niches (Kukkonen et al., 2004; Oyston et al., 1998; Perry et al., 1998; Simonet et al., 1996; Skurnik et al., 2000), and so it can be hypothesized that pseudogene profile should have some correlation with an isolate’s geographic origin and could disclose the adaptive micro-evolution of Y. pestis. Therefore we focused on the gene inactivation mutations and tried to use them as genetic markers for genotyping and analysis of the adaptive micro-evolution of Y. pestis in different natural plague foci.

In total, 24 mutations in 24 genes were screened in this study (Table 3), including 17 IS element insertions, two frameshifts, one in-frame deletion, one fragment deletion, two nonsense mutations and one nonsynonymous mutation, which represent all mechanisms of mutation found in Y. pestis. The pseudogenes involved in this study were classified into six groups according to the strains in which they were present (Fig. 2). The first three groups contain genes only inactivated in one strain, the fourth contains genes inactivated in all three strains, the fifth contains those inactivated in both CO92 and KIM, and the sixth those inactivated in both CO92 and 91001. For two genes, YPO3367 and YPO3720, which are inactivated both in 91001 and in KIM due to frameshifts, allele-specific primers that discriminate the wild-type and mutant cannot be designed because the two frameshifts occurred within a homopolymeric tract of 11G and 7C, respectively.

Table 2. Oligonucleotide primers used

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS100-c-L</td>
<td>actcattccctgtcgg</td>
</tr>
<tr>
<td>IS100-c-R</td>
<td>cactgaataacacaggctg</td>
</tr>
<tr>
<td>IS285-c-L</td>
<td>atctcaatacttttcgg</td>
</tr>
<tr>
<td>IS285-c-U</td>
<td>aacgctgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO0089-C-L</td>
<td>ccagccgcttttctgtg</td>
</tr>
<tr>
<td>YPO0089-C-R</td>
<td>gatccacccgattctcg</td>
</tr>
<tr>
<td>YPO0098L</td>
<td>gattgtggaacactccgg</td>
</tr>
<tr>
<td>YPO0098R</td>
<td>gcaagggcagcagctg</td>
</tr>
<tr>
<td>YPO0170L</td>
<td>cattctgtgcatctg</td>
</tr>
<tr>
<td>YPO0170R</td>
<td>ccagccgcttttctgtg</td>
</tr>
<tr>
<td>YPO1337L</td>
<td>ctctctctctctttgctc</td>
</tr>
<tr>
<td>YPO1337R</td>
<td>ctctctctctctctcagg</td>
</tr>
<tr>
<td>YPO1380L</td>
<td>ctccacaaacctcagg</td>
</tr>
<tr>
<td>YPO1380R</td>
<td>ctgcatccatcatcagg</td>
</tr>
<tr>
<td>YPO1528F</td>
<td>aacatcatatcgggctacaac</td>
</tr>
<tr>
<td>YPO1528R</td>
<td>gttgacgctttagatttacc</td>
</tr>
<tr>
<td>YPO1676LM</td>
<td>cattctgtgcatctg</td>
</tr>
<tr>
<td>YPO1676LW</td>
<td>aagctgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO1676R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO1728L</td>
<td>cgctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO1728R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO1752L</td>
<td>gcctctctctctctctctc</td>
</tr>
<tr>
<td>YPO1752R</td>
<td>gctctctctctctctctc</td>
</tr>
<tr>
<td>YPO1923LM</td>
<td>caggtttttcctttagct</td>
</tr>
<tr>
<td>YPO1923L1W</td>
<td>aagctgcttttctttagct</td>
</tr>
<tr>
<td>YPO1923R</td>
<td>tattctgctactcagg</td>
</tr>
<tr>
<td>YPO1967L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO1967R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2034L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2034R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2142L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2142R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2398-L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2398-R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2493L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2493R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2493RW</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2537L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2537R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2943L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO2943R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO3049F</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO3049RM2</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO3049RW2</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO3250L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
</tbody>
</table>

Table 2. cont.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YPO3250R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO3375L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO3375R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO4008L</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
<tr>
<td>YPO4008R</td>
<td>gctctgtgctttcgacggtaagc</td>
</tr>
</tbody>
</table>
Table 3. Genetic variations analysed in this study and allele-specific PCR-screening results

Gene codes and descriptions of product are according to *Y. pestis* CO92.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variation (91001/CO92/KIM)</th>
<th>Product</th>
<th>Amplification result</th>
</tr>
</thead>
<tbody>
<tr>
<td>YPO0089-YPO0090</td>
<td>Intact/940 bp deletion/intact</td>
<td>GlpX protein glycerol kinase</td>
<td>No isolate had this 940 bp deletion</td>
</tr>
<tr>
<td>YPO0098</td>
<td>Intact/disrupted by IS100/intact</td>
<td>ABC-transport protein, ATP-binding component</td>
<td>The allele was disrupted by IS100 in isolates from focus F</td>
</tr>
<tr>
<td>YPO0170</td>
<td>Intact/disrupted by IS100/intact</td>
<td>Acetylornithine aminotransfer</td>
<td>The allele was intact in all <em>Y. pestis</em> isolates</td>
</tr>
<tr>
<td>YPO1337</td>
<td>Intact/disrupted by IS100/intact</td>
<td>Arginine-binding periplasmic protein 2 precursor</td>
<td>The allele was disrupted by IS100 in isolates from focus F</td>
</tr>
<tr>
<td>YPO1380</td>
<td>Intact/intact/disrupted by IS100</td>
<td>Putative membrane protein</td>
<td>The allele was disrupted by IS100 in isolates from foci G, H, I, J and K</td>
</tr>
<tr>
<td>YPO1562</td>
<td>Intact/disrupted by IS285/disrupted by IS285</td>
<td>Putative intimin</td>
<td>The allele was intact in isolates from foci L, M</td>
</tr>
<tr>
<td>YPO1582</td>
<td>Intact/disrupted by IS285/disrupted by IS285</td>
<td>Lactose permease</td>
<td>The allele was intact in isolates from foci L, M and B</td>
</tr>
<tr>
<td>YPO1676</td>
<td>Intact/nonsense mutation/intact</td>
<td>Methyl-accepting chemotaxis protein</td>
<td>The nonsense mutation was specific to isolates of biovar <em>orientalis</em></td>
</tr>
<tr>
<td>YPO1728</td>
<td>Intact/disrupted by IS100/disrupted by IS100</td>
<td>Putative transcriptional regulator of sugar transport protein</td>
<td>The allele was intact in isolates from foci L, M and B</td>
</tr>
<tr>
<td>YPO1752</td>
<td>Intact/disrupted by IS100/intact</td>
<td>Hypothetical protein</td>
<td>The allele was intact in all <em>Y. pestis</em> isolates</td>
</tr>
<tr>
<td>YPO1923-1</td>
<td>Intact/18 bp in-frame deletion/18 bp in-frame deletion</td>
<td>Putative sensor protein</td>
<td>The allele was intact in isolates from foci L and M</td>
</tr>
<tr>
<td>YPO1923-2</td>
<td>Intact/framenhift/intact</td>
<td>Putative sensor protein</td>
<td>The thyminde deletion was specific to isolates of biovar <em>orientalis</em></td>
</tr>
<tr>
<td>YPO1967</td>
<td>Intact/disrupted by IS100/disrupted by IS100</td>
<td>Porin</td>
<td>The allele was intact in isolates from foci L, M and B</td>
</tr>
<tr>
<td>YPO2034</td>
<td>All were disrupted by IS1541</td>
<td>Putative ABC transporter ATP-binding protein</td>
<td>The allele was disrupted by IS1541 insertion in all <em>Y. pestis</em> isolates</td>
</tr>
<tr>
<td>YPO2142</td>
<td>Intact/intact/nonsense mutation</td>
<td>Na&lt;sup&gt;+&lt;/sup&gt;/H&lt;sup&gt;+&lt;/sup&gt; antiporter</td>
<td>The nonsense mutation was specific to isolates from focus K</td>
</tr>
<tr>
<td>YPO2309</td>
<td>Disrupted by IS100/intact/intact</td>
<td>Two-component regulatory system, sensor kinase protein</td>
<td>The allele was disrupted by IS100 in isolates from foci L and M</td>
</tr>
<tr>
<td>YPO2398</td>
<td>Disrupted by IS100/disrupted by IS100</td>
<td>Putative exported protein</td>
<td>The allele was intact in isolates from foci G, H, I, J and K</td>
</tr>
<tr>
<td>YPO2493</td>
<td>Intact/intact/framenhift</td>
<td>Putative dioxygenase alpha subunit</td>
<td>The frameshift was specific to isolates from foci G, H, I, J and K</td>
</tr>
<tr>
<td>YPO2572</td>
<td>Intact/disrupted by IS285/intact</td>
<td>Bifunctional penicillin-binding protein 1C</td>
<td>The allele was disrupted by IS285 in isolates from foci C, D, E and F</td>
</tr>
<tr>
<td>YPO2943</td>
<td>All were disrupted by IS285</td>
<td>Outer membrane usher protein</td>
<td>The allele was disrupted by IS285 in all <em>Y. pestis</em> isolates</td>
</tr>
<tr>
<td>YPO3049</td>
<td>Nonsynonymous mutation: CGC/CAC/CAC</td>
<td>Putative binding protein-dependent transport system, inner-membrane component</td>
<td>The allele was CGC in isolates from foci L, M and B</td>
</tr>
<tr>
<td>YPO3250</td>
<td>All were disrupted by IS100</td>
<td>Putative amino acid transaminase</td>
<td>The allele was intact in isolates from foci G and H</td>
</tr>
<tr>
<td>YPO3775</td>
<td>All were disrupted by IS100</td>
<td>Deoxyribonuclease TatD</td>
<td>The allele was intact in isolates from focus B</td>
</tr>
<tr>
<td>YPO4008</td>
<td>Intact/disrupted by IS100/disrupted by IS100</td>
<td>Two-component system, sensor kinase</td>
<td>The allele was intact in isolates from foci L, M and B</td>
</tr>
</tbody>
</table>
**Table 4.** Digitized mutation profile of eight genotypes and comparison with typing result based on DFR profiles

<table>
<thead>
<tr>
<th>Focus or strain*</th>
<th>Genotype</th>
<th>Genomovar†</th>
<th>Gene score‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td><em>Y. pseudotuberculosis</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>04</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>01, 02, 03</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>05, 06, 08</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>08</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>07</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>7</td>
<td>09</td>
<td>1</td>
</tr>
<tr>
<td><em>Y. pestis</em> CO92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>J</td>
<td>5</td>
<td>11, 12</td>
<td>1</td>
</tr>
<tr>
<td>K</td>
<td>6</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td><em>Y. pestis</em> KIM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>8</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>8</td>
<td>14</td>
<td>0</td>
</tr>
</tbody>
</table>

*The codes for the foci are as in Table 1.
†Genomovars were typed based on 22 DFR profiles (Zhou et al., 2004b).
‡Genes are as follows: A, YPO2309; B, YPO3049; C, YPO2398; D, YPO3250; E, YPO1562; F, YPO1923-1; G, YPO1582; H, YPO1728; I, YPO1967; J, YPO4008; K, YPO3775; L, YPO2572; M, YPO0098; N, YPO1337; O, YPO1676; P, YPO1923-2; Q, YPO1380; R, YPO2493; S, YPO2142.
Genotyping of homologous isolates of *Y. pestis* based on mutation profile

Five of the 24 mutations analysed showed identical alleles in the 260 *Y. pestis* isolates. Three of these five mutations (YPO1752, YPO0170 and YPO0089-YPO0090) may be specific to *Y. pestis* CO92, while inactivation of the other two genes (YPO2943 and YPO2034) may have accelerated the speciation of *Y. pestis* because both of them are intact in *Y. pseudotuberculosis*. These five mutations were discarded from further genotyping and evolutionary analysis, because they have no discriminatory power among the *Y. pestis* isolates from different geographic areas (Table 3). A 0/1 matrix for the 260 *Y. pestis* isolates and the seven *Y. pseudotuberculosis* strains was constructed (Table 4) by using the remaining mutation profiles. From the results of this PCR-based genotyping method, it was found that isolates from the same focus had identical mutation profiles, and there were several mutations specific to isolates from the same focus.

The 260 *Y. pestis* isolates were grouped into eight genotypes based on their mutation profiles (Table 4). *Y. pestis* CO92, 91001 and KIM showed identical mutation profiles to genotypes 7, 8 and 6, respectively, and so they were assigned into the corresponding genotype. The seven *Y. pseudotuberculosis* strains harboured the wild-type alleles for all the mutations analysed, and they were assigned into a single genotype and used as an outgroup in the following phylogenetic analysis. The phylogenetic tree of the nine genotypes was reconstructed using PHYLIP MIX algorithm (Fig. 3).

Based on whole-genome DNA microarray hybridization, Zhou et al. (2004b) have identified 22 difference regions (DFRs) in the *Y. pestis* genome. The 260 *Y. pestis* isolates were divided into 14 genomovars based on DFR profile typing, and a paradigm of plague transmission, colonization and expansion in China was deduced (Zhou et al., 2004b). The results of the two typing methods that were based on different molecular genetic markers agreed with each other (Table 4).

---

**Fig. 1.** (a) Schematic diagram of primers designed for identifying a nonsense mutation. The gene was inactivated in the mutant due to a nonsense mutation (tac/taa). (b) Schematic diagram of primers designed for identifying a frameshift. The gene was inactivated in the mutant due to a 19 bp fragment insertion (cgtatgctgltttcttgg). (c) Schematic diagram of primers designed for identifying genes disrupted by an insertion sequence. The gene was disrupted by an IS element insertion in mutants 1 and 2, and in mutant 2 the two disrupted parts are located remotely on the chromosome due to a genome rearrangement.

**Fig. 2.** Distribution of pseudogenes in three strains of *Y. pestis*. The number of pseudogenes harboured by each strain or shared with the neighbouring strains and (in parentheses) the number of pseudogenes analysed in this study are shown.

**Fig. 3.** Reconstructed phylogenetic tree based on mutation profile using PHYLIP MIX algorithm. The upper-case letters in parentheses indicate the focus code as listed in Table 1.
However, DFR typing had more discriminatory power than the mutation profile, since some isolates of one genotype were divided into two or more genomovars.

Many genotyping methods that are based on different molecular biological techniques have been applied to typing Y. pestis, such as ribotyping, RFLP and PFGE (Guioule et al., 1994, 1997; Huang et al., 2002). Ribotyping and RFLP analysis provide information about the local genome environments of specific gene sequences on the basis of the probe used. PFGE separates DNA fragments upon digestion of the chromosome with a restriction endonuclease that cleaves infrequently. All of these methods rely on the separation of different length DNA fragments using electrophoresis, which cannot resolve the co-migration of DNA bands and presence of too small or too large bands. Recently, a variable-number tandem repeat (VNTR) technique was used to type Y. pestis (Adair et al., 2000), and a typing method based on divergence of IS100 locations has also been reported (Motin et al., 2002).

The genotyping method described here is based on the mutations that led to gene inactivation, which may relate to Y. pestis phenotype and niche adaptation. These mutations not only include IS element insertions, but also include frameshift and nonsense mutations, which represent all possible mechanisms of pseudogene evolution that were employed by Y. pestis to reshape its genome to adapt to a new niche.

Different lifestyles determine different genotypes

Plague is a typical natural focus-based disease. Long-term survival of Y. pestis and periodic epidemics of animal plague both occur in natural plague foci. Plague primarily affects rodent reservoirs, and disease transmission among rodents is accomplished by the bites of fleas vectors.

The lifestyle of Y. pestis can be viewed as its surrounding environment and the relationships that it establishes with rodent reservoirs and flea vectors. The lifestyles in different foci vary markedly, although the rodent reservoirs or flea vectors may be the same in different foci, which determine the final shape of the Y. pestis genome.

Y. pestis isolates from the same focus showed identical mutation profiles and were always grouped into one genotype (Table 4). We did not find that isolates from one focus were split into different genotypes that were shared by isolates from different foci, which suggests that adaptation of Y. pestis to a particular lifestyle contributes to the reshaping of its genome and leads to the accumulation of mutations within functional genes. However, isolates from adjacent foci (foci G, D and E; foci I and J) or geographically distant foci (foci L and M; foci H and G) did have identical mutation profiles and belonged to the same genotype (Supplementary Fig. S1, available in JMM Online). This may be due to the lifecycle of adjacent foci having some common feature in the rodent reservoir or environment. They might be grouped into different genotypes with more molecular markers. These mutated genes are non-essential for survival of Y. pestis under the selective pressure exerted by its specific lifestyle. The different selective pressures correlated with the different lifestyles account for the discrete segregation of genotypes of Y. pestis.

Gene inactivation accelerates the formation of different biovars

Y. pestis have been historically divided into three biovars, antiqua, mediaevalis and orientalis, according to their ability to ferment glycerol and to reduce nitrate (Achtman et al., 1999). This kind of biovar assignment is based on their metabolic variations, but does not seem to correlate with virulence. In this study, the isolates of the three biovars were clustered into four distinct branches in the phylogenetic tree (Fig. 3), which strongly supports the notion that biovars mediaevalis and orientalis directly arose from biovar antiqua individually (Gonzalez et al., 2002).

Focus F covers a large area in the southeast of China and has been thought of as the third human plague pandemic origin (Supplementary Fig. S1, available in JMM Online). Y. pestis isolates from this focus are of biovar orientalis. All the tested isolates showed identical mutation profiles (genotype 7) and had accumulated some specific mutations. YPO0098, YPO1337 and YPO1676 were found to have become pseudogenes only in biovar orientalis isolates. The inactivation of these genes, leading to the unique mutation profile (genotype) of biovar orientalis strains, might have correlated with the emergence of biovar orientalis from antiqua and promoted its adaptation to the newly expanded niche.

The specific origin of isolates from foci L and M

All of the Y. pestis isolates from foci L and M had an identical mutation profile and were grouped into genotype 8, although these two plague foci are geographically distant (Supplementary Fig. S1, available in JMM Online). The isolates from focus B showed an identical mutation profile and were grouped in genotype 1. A nonsynonymous mutation within YPO3049 was found only in the Y. pestis isolates from foci L, M and B, whereas the interruptions of YPO1582, YPO1728, YPO1967 and YPO4008 by the IS element were found in all Y. pestis isolates except for those from foci L, M and B. The clade formed by Y. pestis isolates from foci L and M (genotype 8) and the clade of the isolates from focus B (genotype 1) diverged next to Y. pseudotuberculosis in the phylogenetic tree (Fig. 3). Given the fact that Y. pestis is a clone recently evolved from Y. pseudotuberculosis, it can be speculated that isolates of genotypes 1 and 8 are the oldest lineages of Y. pestis isolates used in this study. The isolates from focus B are of biovar antiqua, and so this is consistent with the conclusion that biovars mediaevalis and orientalis originated directly from biovar antiqua. However, it is striking that isolates of genotype 8 are the oldest lineages, because the isolates of genotype 8 are of biovar mediaevalis according to the conventional biotyping method.

The isolates from foci L and M (genotype 8) have accumulated mutations that are unique except for some shared with
genotype 1, such as YPO2309, which encodes a two-component regulatory system sensor kinase protein interrupted by IS100 that is intact in the isolates from other foci, and another seven unique mutations identified in our previous study (Zhou et al., 2004a). The isolates from these two foci showed identical biochemical features (they do not ferment arabinose) and almost identical genomic content as suggested by DNA microarray hybridization, besides the above unique mutation profile (Zhou et al., 2004b). Another interesting fact is that human plague has never been reported in these two foci. Isolates from these foci are thought to be of low pathogenicity to humans and *Y. pestis* strain 91001 isolated from focus L has even been verified as avirulent to human beings by volunteer trial through subcutaneous inoculation (Fan et al., 1994, 1995). Based on these unique features, we propose that the isolates from these two foci are a special old lineage of *Y. pestis* and specified differently to others. The main host is *Microtus brandti* in focus L and *Microtus fuscus* in focus M, so these strains were proposed as a new biovar *microtus* based on three biochemical features (*microtus*, negative for reducing nitrate, ferment glycerol and negative for fermenting arabinose; *antiqua*, reduce nitrate, and ferment glycerol and arabinose; *mediaevalis*, negative for reducing nitrate, ferment glycerol and arabinose; *antiqua*, reduce nitrate, negative for fermenting glycerol but positive for fermenting arabinose) (Zhou et al., 2004a).

**The accumulation of pseudogenes promotes speciation of *Y. pestis***

For all analysed mutations, the wild-type alleles were found in all of the seven *Y. pseudotuberculosis* strains used in this study, which is in accordance with the proposal that *Y. pseudotuberculosis* is the ‘young mother’ of *Y. pestis* and they cannot be discriminated at the DNA level except for point mutations (Achtman et al., 1999). There is the accumulation of pseudogenes in *Y. pestis* during the course of speciation, such as the O-antigen gene cluster, *yadA* and *inv*; all of these genes are intact in *Y. pseudotuberculosis* but inactivated in *Y. pestis*, which is the outcome of the switch of *Y. pestis* from an enteric lifestyle to a mammalian blood-borne lifestyle (Kukkonen et al., 2004; Oyston et al., 2003; Perry et al., 1998; Simonet et al., 1996; Skurnik et al., 2000). It was revealed in this study that YPO2034 and YPO2943, which encode a putative ABC transporter ATP-binding protein and an outer membrane usher protein, are disrupted by IS1541 and IS285 insertions, respectively, in all *Y. pestis*, but are intact in the seven *Y. pseudotuberculosis* strains. These two genes may be non-essential or even harmful to *Y. pestis* and so gene inactivation would be an adaptive mutation for *Y. pestis* to evolve from *Y. pseudotuberculosis* and adapt to a new niche.

**ACKNOWLEDGEMENTS**

This work was supported by the National High Technology Research and Development Program of China (863 program) (contract no. 2001AA223061) and National Natural Science Foundation of China (contract no. 30371284). We thank Dr David Bastin (Tianjing Biochip Co., Tianjing, China) for critical reading of the manuscript.

**REFERENCES**


Deng, W., Burland, V., Plunkett, G., III & 18 other authors (2002). Genome sequence of *Yersinia pestis* KIM. J Bacteriol 184, 4601–4611.


