ISSN 1022-7954, Russian Journal of Genetics, 2015, Vol. 51, No. 7, pp. 642–646. © Pleiades Publishing, Inc., 2015.
Original Russian Text © E.G. Oglodin, A.V. Cherkasov, G.A. Eroshenko, G.N. Odinokov, N.Yu. Shavina, L.A. Novichkova, V.V. Kutyrev, 2015, published in Genetika, 2015, Vol. 51, No. 7, pp. 754–758.

GENETICS OF MICROORGANISMS

Analysis of the Nucleotide Sequence of a Cryptic Plasmid from *Yersinia pestis* Strains in the Central-Caucasian High-Mountain Plague Focus

E. G. Oglodin, A. V. Cherkasov, G. A. Eroshenko, G. N. Odinokov, N. Yu. Shavina, L. A. Novichkova, and V. V. Kutyrev

Microbe, Russian Antiplague Research Institute, Saratov, 410005 Russia
e-mail: rusrapi@microbe.ru

Received July 14, 2014; in final form, October 20, 2014

Abstract—An analysis of a 5.4-kb cryptic plasmid detected in the course of whole-genome sequencing of the *Yersinia pestis* medieval biovar strain isolated in the Russian Central-Caucasian high-mountain plague focus was performed. The identification of the nucleotide sequence of this cryptic plasmid and its structural and functional analysis revealed that it contained eight open reading frames, among which the following genes were identified: the *rep* gene of a replication protein, the *virB6* gene of a type IV secretion system inner membrane protein, the *virB5* gene of the type IV secretion system minor pilin, and a number of genes probably associated with secretion and transport. A general analysis of the pCKF plasmid DNA showed that the adenine content was 28.34%, the cytosine content was 20.5%, the guanine content was 17.87%, and that of thymine was 33.28%, while the total G+C content appeared to be 38.38%. The G+C content of the chromosome of the *Y. pestis* strain C-627 is 47.6%, which indicates that the pCKF plasmid was obtained from a microorganism phylogenetically distant from the *Yersinia* bacteria and any other bacteria from the Enterobacteriaceae family. A comparison of the amino acid sequences of hypothetical proteins encoded by pCKF plasmid with analogous proteins encoded by other bacteria was carried out. The possible contribution of the pCKF plasmid to the maintenance of the most ancient known phylogenetic line of *Y. pestis* medieval biovar, 2.MED0, was discussed.

DOI: 10.1134/S1022795415060125

INTRODUCTION

The plague-etiologic agent *Yersinia pestis* causes a highly infectious disease and circulates in nature between rodents and their fleas. *Yersinia pestis* strains usually contain three different plasmids, pCad, pFra, and pPst, each playing an important role at a certain stage of plague pathogenesis [1]. Two of these plasmids, namely pFra (92.6 kb) and pPst (9.6 kpb), are species specific. pFra plasmid, also called pMT1, encodes two virulence factors, the capsule protein F1, which allows the pathogen to escape phagocytosis by the host immune system cells, and the murine toxin Ymt, which is necessary for *Y. pestis* transmission by fleas. The pPst plasmid (also referred to as pPla and pPCP1) encodes plasminogen activator Pla, which is essential for bubonic and pneumonic plague pathogenesis. The pCad plasmid (also called pYV, pCD) with a size of 70.2 kb is common for the three species of pathogenic Yersinia: *Y. pestis, Y. pseudotuberculosis*, and *Y. enterocolitica*. This plasmid contains a cluster of genes involved in the control of calcium-dependent growth (the dependence of the ability to grow at low temperatures from the presence of Ca^{2+} ions in the culturing medium), which are responsible for the injection of the bacterium virulence factors via the type III secretion system of the host cell.

Apart from these three plasmids, which are typical for plague-causative agents, *Y. pestis* strains may contain additional plasmids of various sizes. Some of them originate from recombination between the typical plasmids, while others are the original plasmids present in certain *Y. pestis* populations [2–4]. A plasmid with a size of about 19.5 kb, which is a dimer composed by two 9.5-kb pPst plasmids, was described in *Y. pestis* strains obtained in the western United States [3]. Cryptic 6 kb pYC plasmid was identified in *Y. pestis* strains isolated in Yunnan province, China [4]. This plasmid comprises 12 open reading frames, among which ORF1 was shown to be the gene encoding the replication initiation protein and ORF11 and ORF12 were shown to be the genes encoding DinJ1 and DinJ2 proteins, which are likely to contribute to the maintenance of DNA stability. The plasmid region upstream from the ORF1 encompasses a number of copies of a direct repeat, with characteristics similar to those of the DNA replication origin [4].

In the *Y. pestis* strains obtained in the Tuva mountain plague focus in Russia, there also occurs a plasmid with a size of about 33 kb, the functions of which are

unknown [5]. A cryptic pCRY plasmid, about 22 kb in size, was detected in the genome sequence of *Y. pestis* strain 91001, which belongs to microtus strains from the two plague foci in China [6]. The G+C content of this plasmid constitutes 49.1%, which makes it different from the other three plasmids present in the plague-causative agent. pCRY contains an approximately 30 genes, among which are the *repA* replication protein gene and the *parA* gene, the product of which is necessary for newly synthesized plasmid segregation into the daughter cells, as well as the *vir* cluster of type VI secretion system genes (T4SS). T4SS is an important virulence factor that is required in bacterial pathogens for the secretion of effector proteins into the host cells and for the transfer of plasmids bearing genes controlling antibiotic resistance. T4SS in *Agrobacterium tumifaciens* usually comprises 12 Vir proteins, namely, Vir1–Vir11 and VirD4 proteins, though in other bacteria it may include additional genes or, on the contrary, may contain a limited set of these proteins' homologs [7, 8]. The type IV secretion system apparatus differs from that of the type III secretion system by its constituting elements, as well as by the architecture. It appears that T4SS participates in the secretion of a broad range of compounds, including large DNA-protein complexes [8]. Song et al. [6] have for the first time demonstrated the presence of T4SS genes in *Y. pestis* strains. However, since the *vir* cluster of the pCRY plasmid lacked *virB7* and *virB8* genes, the function of this incomplete T4SS gene cluster remained unclear. In PCR analysis of pCRY plasmid occurrence in 257 *Y. pestis* strains from China with *repA* gene-specific primers, only 11 revealed a positive signal, thus indicating that this plasmid is uncharacteristic of *Y. pestis* [6].

In the Central-Caucasian high-mountain plague focus in Russia, *Y. pestis* strains of medieval biovar can be isolated; they are different from other strains of this biovar, frequently occurring in the plague foci in Russia, the countries of the Commonwealth of Independent States, and the bordering countries. These strains are characterized by decreased virulence, extraordinary proline dependence, and the presence of an additional plasmid with a size of 5.4 kb [9, 10]. We have recently performed whole-genome sequencing of the proline-dependent *Y. pestis* strain C-627, which was obtained in the Central-Caucasian high-mountain plague focus. This showed that these strains are the most ancient of all of the known strains of medieval biovar [11]. Their ancient origin is an argument in favor of the previously generally held and presently challenged opinion that a second plague pandemic—the Black Death, which ravaged Europe in 13–14 centuries and was caused by *Y. pestis* strains of medieval biovar—has broken out in the Caucasus and the Caspian Sea region.

The purpose of this work was to identify the complete nucleotide sequence of the 5.4-kb plasmid of *Y. pestis* strains from the Central-Caucasian plague focus and to carry out its structural and functional analysis.

MATERIALS AND METHODS

Y. pestis Strain, Culturing Conditions, DNA Isolation and Sequencing

Y. pestis strain C-627 was obtained from *Citellophilus tesquorum* fleas feeding on the Caucasian mountain ground squirrel *Citellus musticus* on the territory of the Kuban-Malkin region of the Central-Caucasian natural plague focus in 1986. The strain was provided by the State Collection of Pathogenic Bacteria of the Russian Research Antiplague Institute "Microbe", where it is stored in a lyophilized state. The strain was cultured in broth and agar LB medium (pH 7.2) for 24–48 h at 28°C. DNA was isolated using AxyPrep Bacterial Genomic Miniprep Kit. A 200-bp fragment library was obtained with an Ion Xpress Library Kit according to the manufacturer's protocol. The nucleotide sequence of the strain genome was identified with an Ion PGM (Life Technologies, the United States) system and an ABI 3500x1 Genetic Analyzer (Life Technologies, the United States) automatic genetic analyzer. In order to obtain the full sequence of the pCKF plasmid circle, the primers were designed in correspondence with the terminal regions of the plasmid contigs. The terminal sequences of contigs were identified with them and were used to finalize the plasmid sequence reconstruction.

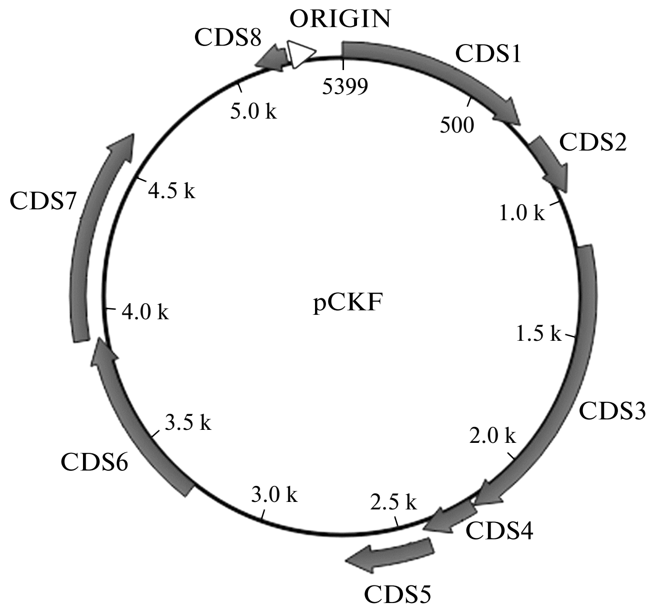
Computer-Assisted Analysis and Software

The primary sequencing data were processed using the Ion Torrent Suite software 3.4.2. Single reads were then assembled *de novo* into contigs with Newbler gsAssembler 2.6 software. Only those contigs containing no less than 100 reads were analyzed. The depth of coverage for each nucleotide was about 2000 fold. Finalization of the plasmid sequence was carried out with UGENE 1.12.2 software.

The analysis of the obtained nucleotide sequence of the plasmid was carried out on the RAST server, while the analysis of the amino acid sequences of the protein-coding regions was performed with Psi-BLAST based on data contained in PATRIC, NCBI GenBank, RAST, Uniprot, Pfam, and EMBL.

RESULTS AND DISCUSSION

The identification of the full-genome sequence of *Y. pestis* strain C-627 obtained in the Central-Caucasian high-mountain plague focus previously allowed us to determine that it belongs to the ancient branch of medieval biovar, which is designated by us as 2.MED0. This has not been previously described, in either Russian or foreign works, and is not present on the world genetic diversity tree of *Y. pestis* [11]. The C-627 strain genome comprises (along with the chromosomal DNA and pFra, pPst, and pCad plasmids) a cryptic



Map of pCKF plasmid from *Y. pestis* strain C-627 obtained in the Central-Caucasian high-mountain plague focus. Arrows indicate the open reading frames detected with the RAST server and the NCBI GenBank glimmer3 algorithm. White triangle designates the presumable plasmid replication start site.

plasmid with a size of 5.4 kb. Since this plasmid had no specific name, we designated it pCKF after the name of the Central-Caucasian high-mountain plague focus where strains containing this plasmid can be isolated. The analysis of the data obtained in the course of the whole-genome sequencing of *Y. pestis* strain C-627 showed that the depth of pCKF plasmid sequence coverage in the C-627 genome was more than 100 fold, which is about two times higher than the corresponding parameter value for another small plasmid, the 9.5-kb pPst plasmid. With UGENE software, the sequences of the terminal regions of the plasmid contigs identified by the capillary sequencing analyzer ABI 3500x1 Genetic Analyzer were assembled together with the nucleotide sequences of pCKF fragments into a circular replicon with a size of 5.4 kb, the map of which is given in the figure.

The pCKF plasmid DNA analysis showed its composition: 28.34% adenine, 20.5% cytosine, 17.87% guanine, and 33.28% thymine. The total G+C content appeared to be 38.38%. At the same time, the G+C content of the *Y. pestis* strain C-627 chromosome DNA is 47.6%, which implies that pCKF plasmid was most probably transferred from some microorganism phylogenetically distant from the *Yersinia* bacteria, as well as from other bacteria of the Enterobacteriaceae family.

In order to identify the number of protein-coding sequences in the pCKF plasmid of the *Y. pestis* strain C-627 and to unveil their possible functions, an analysis of the complete nucleotide sequence of this plasmid was carried out with the open reading frames finder application of the RAST server and the NCBI GenBank glimmer3 algorithm (table). As a result of the analysis, it was shown that the plasmid contains eight open reading frames. Among the genes identified are the *rep* gene encoding a replication protein, the *virB6* gene encoding the inner membrane channel protein of the type IV secretion system, and the *virB5* gene of the minor pilin of the type IV secretion system. These genes are considered identified, because the rate of their similarity with other nucleotide sequences was not less than 70%.

The coding sequence CDS1 (nucleotides 1–693) of pCKF plasmid showed an 89% similarity with the ORF1 gene encoding the replication protein in the 6-kb pYC cryptic plasmid contained in *Y. pestis* strains from Yunnan province, China [4]. Such a significant homology rate between the pCKF plasmid CDS1 and the pYC plasmid ORF1 gene allows us to conclude that this coding sequence is a replication protein gene

Nucleotide sequences contained in pCKF plasmid

CDS	Coding sequences localization (nucleotides starting from the replication start site)	Gene	Encoded protein product	Homologous sequences, % of homology
CDS1*	1–693	*rep*	Replication protein	orf1 of the *Y. pestis* pYC plasmid, 89%
CDS2	765–980		Hypothetical protein	
CDS3	1175–2218	*virB6*	VirB6, type IV secretion system protein	*virB6 Proteus mirabilis*, more than 70%
CDS4	2220–2414		Hypothetical protein	
CDS5	2404–2691		"	
CDS6	3271–3906		"	
CDS7	3903–4619	*virB5*	VirB5, type IV secretion system minor pilin	*virB5 Campylobacter upsaliensis*, more than 70%
CDS8	5204–5088		Hypothetical protein	

* CDS coding sequence.

as well. This protein participates in the process of the initiation of DNA replication by DNA polymerase and in the catalysis of DNA synthesis from deoxyribonucleotide triphosphates.

CDS3 (nucleotides 1175–2218) has more than 70% homology with the *Proteus mirabilis virB6* gene encoding VirB6 protein of the inner membrane of T4SS. VirB6 is supposed to participate in the formation of a channel in the inner membrane and in protein secretion into the surrounding medium or into the host cell. When the predicted amino acid sequence of pCKF plasmid CDS3 was taken into the analysis, it demonstrated 45% homology with the *Y. enterocolitica* CC4T TriE protein, which is encoded by the p29930 plasmid and carries out functions associated with protein secretion similar to those of VirB6.

CDS7 (nucleotides 3903–4619) shows more than 70% homology with the *virB5* gene of *Campylobacter upsaliensis*. VirB5 is a minor pilin and participates in periplasmic channel formation in T4SS. The amino acid sequence of the CDS7 protein product shows 55% homology with *Y. enterocolitica* T4SS TriD protein, which is also encoded by p29930 plasmid. It appears that the *virB5* and *virB6* genes identified in pCRF plasmid contribute to protein secretion into the surrounding medium or into the host cells.

When utilizing the RAST server application, which allows the identification of open reading frames and NCBI GenBank glimmer3 algorithm (both of which indentify only those genes showing no less than 70% homology with the genes in the databases), we were not able to determine the functions of the five other transcription units in pCKF plasmid. We identified the putative functions of some of these genes by comparing the predicted amino acid sequences of their protein products with the amino acid sequences present in the PATRIC, NCBI GenBank, Uniprot, Pfam, and EMBL databases. Although the percent of similarity with the identified homologous proteins appeared to be less than the threshold value, it still allowed us to make suppositions of the possible functions of the proteins encoded by some pCKF genes.

CDS2 (nucleotides 765–980) encodes a protein that has 41% homology with the *Escherichia coli* hypothetical protein with unknown functions.

CDS4 (nucleotides 2220–2414) protein shows 42% homology with the *Streptococcus parasanguinis* hypothetical protein, the functions of which are unknown.

CDS5 (nucleotides 2404–2691) protein has 39% homology with un uncharacterized protein product of the BN482_02023 gene of the *Clostridium* sp. strain CAG:127.

CDS6 (nucleotides 3271–3906) protein product shows 30% homology with the *Y. enterocolitica* T4SS TriL protein encoded by the p29930 plasmid.

CDS8 (nucleotides 5204–5088) protein has 43% homology with the *Bacillus anthracis* ABC transporter C-terminal permease protein. The proteins of this family occur only in prokaryotes and take part in the transport of a broad range of substrates, varying from small ions to macromolecules.

We have also revealed the presence of a unique region with high A+T contents (65.9%) in the nucleotide position 5219–5308 of the pCKF plasmid of the *Y. pestis* strain C-627, which was made up by two 44 bp repeats. The repeats were composed by two shorter 22 bp repeats containing three unique nucleotide substitutions each. The entire AT-rich region is located upstream from the *rep* gene encoding the replication protein and may play the role of replication origin of this plasmid (see figure).

To summarize, the obtained data indicate that the genes located in pCKF plasmid may be associated with processes of transport and secretion, in particular with T4SS. Similar functions were described for genes encoded by pCRY plasmid from the *Y. pestis* microtus strain 91001 [6]. The latter encodes 10 out of 12 *vir* cluster genes of T4SS but lacks *virB7* and *virB8* genes. In the pCKF plasmid, only two *vir* cluster genes, namely *virB5* and *virB6*, were identified, with their functions remaining unclear. Since the possibility of the presence of T4SS genes in the genomes of the strains bearing pCKF plasmid could not be excluded, we have examined the full-genome sequence of the *Y. pestis* strain C-627 for the presence of corresponding genes. As a result of the analysis, we detected the presence of the *virB11* gene, which encodes adenosine triphosphatase (ATPase) participating in the secretion of various substrates, in the chromosome. The significance of these single T4SS genes, which are localized in the strain C-627 genome and in the pCKF plasmid, for the vital activity and pathogenic properties of the strains bearing this plasmid is still unclear. However, the stable maintenance of the pCKF plasmid in the genomes of *Y. pestis* strains of medieval biovar from the Central-Caucasian high-mountain plague foci may indicate that the factors encoded by it provide their persistence under the specific landscape and geographical conditions of this region, while other ancient branches of medieval biovar have not survived. The fact that at least four genes, *virB6* (CDS3), *virB5* (CDS7), CDS6, and CDS8, in pCKF plasmid are associated with secretion and transport makes it possible to suppose that this plasmid may contribute to the pathogenicity of the strains which bear it.

To summarize, we have for the first time determined the full nucleotide sequence of the 5.4 kb cryptic plasmid from *Y. pestis* strains of medieval biovar and performed its structural and functional analysis. It appears that this plasmid encodes proteins participating in the secretion and transport into the host cells or the extracellular medium. The international information resources lack any data on this plasmid, the genes it contains, and their protein products. The pCKF plasmid nucleotide sequence was deposited in the

NCBI GenBank database with the accession no. KM112087.

REFERENCES

1. Kutyrev, V.V., Eroshenko, G.A., Popov, N.V., et al., Molecular mechanisms of interactions of plague causative agents with invertebrates, *Mol. Genet., Microbiol. Virusol.*, 2009, vol. 24, no. 4, pp. 169–176.
2. Filippov, A.A., Solodovnikov, N.S., Kookleva, L.M., and Protsenko, O.A., Plasmid content in *Yersinia pestis* strains of different origin, *FEMS Microbiol. Lett.*, 1990, vol. 55, pp. 45–48.
3. Chu, M.C., Dong, X.Q., Zhou, X., and Garon, C.F., A cryptic 19-kilobase plasmid associated with U.S. isolates of *Yersinia pestis*: a dimer of the 9.5-kilobase plasmid, *Am. J. Trop. Med. Hyg.*, 1998, vol. 59, pp. 679–686.
4. Dong, X.Q., Lindler, L.E., and Chu, M.C., Complete DNA sequence and analysis of an emerging cryptic plasmid isolated from *Yersinia pestis*, *Plasmid*, 2000, vol. 43, pp. 144–148.
5. Balakhonov, S.V., Afanas'ev, M.V., Shestopalov, M.Yu., et al., The first case of *Yersinia pestis* subsp. *pestis* isolation from the territory of Altai Mountain natural plague focus: 1. Microbiological characteristics, molecular-genetic and mass-spectrometric identification of the isolate, *Probl. Osobo Opasnykh Infekts.*, 2013, no. 1, pp. 60–65.
6. Song, Y., Tong, Z., Wang, J., et al., Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans, *DNA Res.*, 2004, vol. 11, pp. 179–197.
7. Fronzes, R., Schafer, E., Wang, L., et al., Structure of a type IV secretion system core complex, *Science*, 2009, vol. 323, pp. 266–268.
8. Christie, P.J., Atmakuri, K., Krishnamoorthy, V., et al., Biogenesis, architecture, and function of bacterial type IV secretion systems, *Annu. Rev. Microbiol.*, 2005, vol. 59, pp. 451–485.
9. Protsenko, S.L., Serdyukova, T.V., and Rozanova, G.N., Detection of a new plasmid in proline-dependent plague agent strains found in Central-Caucasian natural focus and heterogeneity in plasmid composition of the circulating in the focus strains, in *Osobo opasnye infektsii na Kavkaze*, (Particularly Dangerous Infections in the Caucasus), 1987, pp. 265–267.
10. Zharinova, N.V., Bryukhanova, G.D., Eremenko, E.I., et al., Properties of *Yersinia pestis* strains, circulating in the Central Caucasus mountainous natural foci of plague, in *Aktual'nye problemy epidemiologii i profilaktiki infektsionnykh boleznei (Modern Problems of Epidemiology and Prophylaxis of Infectious Diseases)*, 2004, vol. 2, pp. 14–16.
11. Odinokov, G.N., Eroshenko, G.A., Krasnov, Ya.M., et al., Analysis of the genome wide sequence of *Yersinia pestis* strains based on the consecutive 680-SNP algorithm, *Probl. Osobo Opasnykh Infekts.*, 2013, no. 3, pp. 49–54.

Translated by E. Martynova