Draft Genome Sequences of *Yersinia pestis* Strains from the 1994 Plague Epidemic of Surat and 2002 Shimla Outbreak in India

Kiran N. Mahale · Pradyumna S. Paranjape · Nachiket P. Marathe ·
Dhiraj P. Dhotre · Somak Chowdhury · Sudarshan A. Shetty · Avinash Sharma ·
Kaushal Sharma · Urmil Tuteja · Harsh V. Batra · Yogesh S. Shouche

Received: 27 March 2014 / Accepted: 19 May 2014 / Published online: 31 May 2014
© Association of Microbiologists of India 2014

Abstract We report the first draft genome sequences of the strains of plague-causing bacteria, *Yersinia pestis*, from India. These include two strains from the Surat epidemic (1994), one strain from the Shimla outbreak (2002) and one strain from the plague surveillance activity in the Deccan plateau region (1998). Genome size for all four strains is ~4.49 million bp with 139–147 contigs. Average sequencing depth for all four genomes was 21x.

Keywords IonTorrent · NGS, next generation sequencing · Plague · Shimla · Surat

Introduction

Plague is a highly infectious disease caused by a Gram-negative coccobacillus (family Enterobacteriaceae), *Yersinia pestis*. Plague has earned disrepute in history because of its causative role in the three pandemics [1, 2]: the

Justinian plague (sixth century, originated in central Asia and spread through Asia, Africa and Europe, infecting nearly 100 million people); the Black Death (fourteenth century pandemic with a death toll of 50 million; Europe was deprived of 25 million of its population, the other 25 million deaths were in Asia and Africa); and the Modern Pandemic (begun in China's Yunnan Province in the mid-nineteenth century expanding its geographic range to previously unaffected regions). India was one of the severely hit countries by the Modern Pandemic of plague causing 12.5 million casualties during 1896–1950 [3].

The 1994 Gujarat plague epidemic was one of the most significant demonstrations that plague was not totally wiped out from India. The official panic responses from several countries, cancellation of 400 flights to India by fifteen airlines and the getaway of 600,000 inhabitants of the city of Surat during the 1994 epidemic showed that plague remains one of the dreaded infectious diseases even in the post-antibiotic era [4, 5].

The availability of genome sequences of Indian strains of this highly infectious pathogen from the past epidemics and breakouts and the periodic surveillances will help in the differentiation for epidemiology and phylogeography studies. This knowledge will also aid in designing simpler and efficient diagnosis of Indian strains of *Y. pestis* [3, 6]. In this endeavour, we report the first draft genome sequences of four *Y. pestis* strains isolated from India (Table 1).

Materials and Methods

Bacterial Strains

Two of the four strains are from the 1994 Surat epidemic; *Y. pestis* 9 and 113 have been isolated from a pneumonic patient

K. N. Mahale (✉) · P. S. Paranjape · N. P. Marathe ·
D. P. Dhotre · S. Chowdhury · S. A. Shetty · A. Sharma ·
K. Sharma · Y. S. Shouche (✉)
Microbial Culture Collection, National Centre for Cell Science,
University of Pune Campus, GaneshKhind,
Pune 411007, Maharashtra, India
e-mail: kkmahale@gmail.com

Y. S. Shouche
e-mail: yogesh@nccs.res.in

U. Tuteja
Microbiology Division, Defence Research and Development
Establishment, Gwalior, Madhya Pradesh, India

H. V. Batra
Microbiology Division, Defence Food Research Laboratory,
Mysore, Karnataka, India

Table 1 Summary of the sequencing results and NCBI PGAAP annotations of the four Indian strains of *Y. pestis*

	<i>Y. pestis</i> 9	<i>Y. pestis</i> 113	<i>Y. pestis</i> S3	<i>Y. pestis</i> 24H
Area of isolation	Surat, Gujarat	Surat, Gujarat	Shimla, Himachal Pradesh	Hosur, Deccan Plateau
Year	1994	1994	2002	1998
Host	<i>Homo sapiens</i>	<i>Rattus rattus</i>	<i>Homo sapiens</i>	<i>Tatera indica</i>
Total IonTorrent reads ($\times 10^6$)	1.64	1.56	1.11	1.13
Mean read length	92	91	107	90
Depth	23x	27x	18x	16x
Chromosomal contigs	140	139	147	140
Total chromosomal nucleotides	4,491,896	4,494,902	4,492,823	4,492,822
Chromosome NCBI acc. no.	AXDF01000001-140	AXDG01000001-139	AXDH01000001-147	AXDI01000001-140
Genes	4,284	4,247	4,507	4,245
CDS	3,861	3,979	4,160	3,916
rRNAs	19	19	19	19
tRNAs	68	68	68	68
ncRNA	22	30	34	26
Pseudogenes	314	151	226	216
Frameshifted genes	311	204	223	207

and a rodent, *Rattus rattus* respectively. The third isolate, *Y. pestis* 24H, is from a rodent, *Tatera indica*, collected during a surveillance activity in 1998 in Hosur (Deccan plateau). The fourth strain, *Y. pestis* S3, is from a 2002 plague outbreak in Shimla, Himachal Pradesh, isolated from a human [7–9]. DNA extraction was done using phenol–chloroform method described by Sambrook et al. [10].

Genome Assembly and Analysis

Whole genome sequencing of all the four *Y. pestis* strains was performed by next generation sequencing using an IonTorrent PGM system, according to the manufacturer's protocol for 200 bp chemistry. The run generated 1.11–1.64 million reads for the four strains with a mean read length of ~ 100 bp. The reads were assembled by MIRA version 3.4.1.1 by reference-based assembly.

The reference chromosome used as a reference for the assembly was that of *Y. pestis* CO92, a 4.6 Mb chromosome (NCBI accession no. NC_003143.1). The prime repeat elements, the four types of insertion sequences, IS100 (1954 bases), IS285 (1315 bases), IS1541 (or IS200G, 712 bases) and IS1661 (1441 bases) were masked in the reference genomes prior to assembly.

Results and Discussion

The number of chromosomal contigs obtained for the four strains of *Y. pestis* were in the range 139–147. Gene

prediction and annotation for all the four genomes were done using NCBI Prokaryotic genome automatic annotation pipeline (PGAAP). A summary of the sequencing results and PGAAP annotations has been provided in Table 1.

The whole genome shotgun projects for the four strains have been deposited at DDBJ/EMBL/GenBank under the accessions AXDF000000000 (*Y. pestis* 9), AXDG000000000 (*Y. pestis* 113), AXDH000000000 (*Y. pestis* S3) and AXDI000000000 (*Y. pestis* 24H). The respective versions described in this paper are AXDF01000000, AXDG01000000, AXDH01000000 and AXDI01000000.

The genomes were also subjected to annotation by RAST (Rapid Annotation using Subsystem Technology; <http://rast.nmpdr.org/>) [11]. The subsystem annotations of the four strains were compared with that of *Y. pestis* CO92 (NCBI accession no. NC_003143.1) and presented in Table 2.

Preliminary analysis using RAST annotation (Table 2) showed the presence of a higher number of virulence, disease and defense genes in the Indian strains of *Y. pestis* than the reference genome *Y. pestis* CO92 (91 genes). Highest number of virulence, disease and defence genes were found in Surat strain 9 (107 genes) followed by Hosur strain 24 H (105 genes). Similarly higher numbers of stress response genes were detected in the Indian strains (strain S3, 162 genes; strain 9, 159 genes) than the reference genome CO92 (145 genes). Further sequencing of more *Y. pestis* strains from India, the in-depth annotations of their genomes and their comparison with publicly available *Y. pestis* genomes from other geographic locations will shed more light on the evolution and phylogeography of the Indian strains.

Table 2 Comparative counts of the subsystem features of the four Indian strains (9, 113, S3, 24H) with that of *Y. pestis* CO92 in the RAST annotation system

Strain of <i>Y. pestis</i>	CO92	9	113	S3	24H
RAST ID	6666666.51834	6666666.48993	6666666.48994	6666666.48996	6666666.48997
Subsystem features					
Cofactors, vitamins, prosthetic groups, pigments	241	265	244	250	241
Cell Wall and Capsule	188	212	187	193	207
Virulence, disease and defense	91	107	97	97	105
Potassium metabolism	26	29	26	29	28
Photosynthesis	0	0	0	0	0
Miscellaneous	58	62	59	64	63
Phages, prophages, transposable elements, plasmids	17	22	18	17	17
Membrane Transport	220	232	232	242	225
Iron acquisition and metabolism	105	128	112	110	115
RNA metabolism	202	227	209	225	215
Nucleosides and nucleotides	111	132	121	124	126
Protein Metabolism	248	288	270	284	284
Cell division and cell cycle	35	42	36	38	35
Motility and chemotaxis	151	161	154	160	153
Regulation and cell signaling	115	131	122	131	126
Secondary metabolism	5	6	6	5	6
DNA metabolism	136	149	141	142	139
Regulons	6	6	6	6	6
Fatty acids, lipids, and isoprenoids	105	118	113	128	111
Nitrogen metabolism	26	30	29	29	28
Dormancy and sporulation	2	2	2	2	2
Respiration	126	141	132	137	136
Stress response	145	159	153	162	155
Metabolism of aromatic compounds	22	23	23	25	31
Amino acids and derivatives	405	455	414	423	427
Sulfur metabolism	59	68	60	67	61
Phosphorus metabolism	56	62	60	59	59
Carbohydrates	486	570	513	556	525

Acknowledgments KNM is the recipient of a post-doctoral fellowship (DBT-RA program) from the Department of Biotechnology, Government of India. The work was funded by the Department of Biotechnology (DBT), Government of India; Microbial Culture Collection Project.

References

- Gage KL, Kosoy MY (2005) Natural history of plague: perspectives from more than a century of research. *Annu Rev Entomol* 50:505–528
- Eisen RJ, Gage KL (2012) Transmission of flea-borne zoonotic agents. *Annu Rev Entomol* 57:61–82
- Dennis DT (1994) Plague in India. *BMJ* 309:893–894
- Ramalingaswami V (1996) The plague outbreaks of India, 1994—a prologue. *Curr Sci* 71:781–782
- Perry RD, Fetherston JD (1997) *Yersinia pestis*—etiologic agent of plague. *Clin Microbiol Rev* 10:35–66
- Arya SC (1994) Plague in India. *BMJ* 309:1369
- Batra HV, Tuteja U, Agarwal GS (1996) Isolation and identification of *Yersinia pestis* responsible for the recent plague outbreaks in India. *Curr Sci* 71:787–791
- Panda SK, Nanda SK, Ghosh A, Sharma C, Shivaji S, Seshu Kumar G, Kannan K, Batra HV, Tuteja U, Ganguly NK, Chakrabarty A, Sharat Chandra H (1996) The 1994 plague epidemic of India: molecular diagnosis and characterization of *Yersinia pestis* isolates from Surat and Beed. *Curr Sci* 71:794–799
- Kingston JJ, Tuteja U, Kapil M, Murali HS, Batra HV (2009) Genotyping of Indian *Yersinia pestis* strains by MLVA and repetitive DNA sequence based PCRs. *Antonie Van Leeuwenhoek* 96:303–312
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75