JOURNAL OF BACTERIOLOGY, Aug. 2011, p. 4272–4273 Vol. 193, No. 16
0021-9193/11/$12.00 doi:10.1128/JB.05413-11
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

Genome of Multidrug-Resistant Uropathogenic *Escherichia coli* Strain NA114 from India▽

Tiruvayipati Suma Avasthi,[1]† Narender Kumar,[1]† Ramani Baddam,[1]† Arif Hussain,[1]*
Nishant Nandanwar,[1] Savita Jadhav,[2] and Niyaz Ahmed[1,3,4]*

Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Hyderabad, India[1]; Department of Microbiology, Dr. D. Y. Patil Medical College, Dr. D. Y. Patil University, Pimpri, Pune, India[2]; Institute of Life Sciences, University of Hyderabad Campus, Hyderabad, India[3]; and Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia[4]

Received 31 May 2011/Accepted 6 June 2011

Uropathogenic *Escherichia coli* (UPEC) causes serious infections in people at risk and has a significant environmental prevalence due to contamination by human and animal excreta. In developing countries, UPEC assumes importance in certain dwellings because of poor community/personal hygiene and exposure to contaminated water or soil. We report the complete genome sequence of *E. coli* strain NA114 from India, a UPEC strain with a multidrug resistance phenotype and the capacity to produce extended-spectrum beta-lactamase. The genome sequence and comparative genomics emanating from it will be significant in understanding the genetic makeup of diverse UPEC strains and in boosting the development of new diagnostics/vaccines.

Pathogenic *Escherichia coli* constitutes a significant threat to public health, and the emergence of extended-spectrum beta-lactamase (ESBL)-producing *E. coli* with high virulence potential is alarming (14, 16, 20, 22, 23). Comparative genomics holds significant promise in understanding the genome organization of such bacteria and thereby identifying coordinates highly relevant in the development of intervention strategies (1). Our group has recently studied uropathogenic *E. coli* (UPEC) from the western Indian city of Pune (11), whereupon strain NA114 emerged as an ideal representative of the entire Pune collection. The three major characteristics of strain NA114 that make it epidemiologically and clinically significant are its affiliation with serogroup O25, its placement in phylogenetic group B2, and its sequence type, ST131 (19). The latter denotes a pandemic clone frequently associated with community-acquired antimicrobial-resistant infections (23). Motivated by these facts, we performed complete in-depth sequencing, annotation, and analysis of the genome of UPEC strain NA114, which was originally obtained from the urine of a 70-year-old male patient with prostatitis from Pune. Antibiotic sensitivity tests revealed that it was a multidrug-resistant strain refractory to several common antibiotics and was an ESBL producer (11).

(This work constitutes part of the unpublished doctoral work of Arif Hussain.)

The genome sequence was determined by Illumina Genome Analyzer (GA2x, pipeline ver. l.6) and consisted of sequence traces equivalent to 8 gigabytes of data, encompassing 54-bp paired-end reads with an insert size of 300 bp, and the genome coverage achieved was 500×. The sequence was assembled using Velvet (26), and the contigs were ordered with respect to the best-aligned positions compared to the reference genome of *E. coli* SE15 (25) using Mauve (5, 6). The genome alignment tools BLAT (15) and MUMmer (17) were also used to validate the aligned contigs. The genome was annotated with the help of the RAST server (2), and putative CDSs were identified by comparing outputs from Glimmer (7), Genemark (4), and EasyGene (18). Artemis (24) was used to glean the following details of the genome.

The size of the NA114 chromosome was 4,935,666 bp with a G+C content of 51.16% and a coding percentage of 88.4% with 4,875 protein coding sequences with an average length of 901 bp. The genome revealed 67 tRNA and 3 rRNA genes. We also found several virulence genes, including *iha*, *sat*, *fimH*, *kpsM*, *iutA*, and *malX*, which correspond to the genes of *E. coli* CFT073 (9). In addition, genes corresponding to another UPEC strain, UTI89, such as *fyuA* and *usp* etc., were located. PCR-based analysis showed that this strain carried multiple virulence genes infrequently described in a clone of this type, including *sfa*, *aer*, *cnf*, and an intact polyketide synthase (*pks*) island (12). *E. coli* NA114 also contains other virulence factors, such as *pap*, *fim*, and genes for iron uptake systems such as the hemin uptake system and the yersiniabactin siderophore (*ybt*). In addition to a 4.935-Mb chromosomal genome, strain NA114 also harbored a single plasmid of 3.5 kb which has yet to be analyzed with regard to its replicon type and resistance gene profiles, if it has any.

These observations and the comparative genomic studies emanating therefrom could be extremely useful both in improving our fundamental understanding of multidrug resistance mechanisms encoded by UPEC and in the design of effective drugs to control and manage the alarming health hazards caused by ESBL-producing bacteria in both the developing and developed parts of the world.

* Corresponding author. Mailing address: Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Professor CR Rao Road, Gachibowli, Hyderabad 500 046, India. Phone: 91 40 23134585. Fax: 91 40 66794585. E-mail for Niyaz Ahmed: niyazSL@uohyd.ernet.in. E-mail for Arif Hussain: pbl_uoh@gmail.com.
† These authors contributed equally.
▽ Published ahead of print on 17 June 2011.

Downloaded from http://jb.asm.org/ on October 19, 2019 by guest

Nucleotide sequence accession number. The genome sequence of *E. coli* NA114 has been deposited in GenBank under accession no. CP002797.

This genome program was supported by the University of Hyderabad through an interim grant to N.A. as a part of the Indo-German International Research Training Group—Internationales Graduiertenkolleg (GRK1673)—Functional Molecular Infection Epidemiology, an initiative of the German Research Foundation (DFG) and the University of Hyderabad/University Grants Commission India, of which N.A. is a Speaker. N.A. is an Adjunct Professor of Molecular Biosciences at the University of Malaya, Kuala Lumpur, Malaysia, and an Adjunct Professor of Chemical Biology at the Institute of Life Sciences, Hyderabad, India.

We are thankful to Seyed E. Hasnain, Joerg Hacker, Lothar H. Wieler, and Christa Ewers for helpful advice and suggestions. Akash Ranjan is gratefully acknowledged for timely support and discussions. We are grateful to M/s Genotypic Technology Pvt. Ltd., Bengaluru, India, for their help with Illumina sequencing. We are also thankful to the authorities at Dr. D. Y. Patil University, Pune, India, for their cooperation and facilitation of this study. We acknowledge project management/laboratory facilitation support rendered by Gutti Navamallika.

REFERENCES

1. **Ahmed, N., U. Dobrindt, J. Hacker, and S. E. Hasnain.** 2008. Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. Nat. Rev. Microbiol. **6:**387–394.
2. **Aziz, R. K., et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9:**75.
3. Reference deleted.
4. **Besemer, J., and M. Borodovsky.** 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. **33:**W451–W454.
5. **Darling, A. C., B. Mau, F. R. Blattner, and N. T. Perna.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14:**1394–1403.
6. **Darling, A. E., B. Mau, and N. T. Perna.** 2010. Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147.
7. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
8. Reference deleted.
9. **Guyer, D. M., J. S. Kao, and H. L. Mobley.** 1998. Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. Infect. Immun. **66:**4411–4417.
10. Reference deleted.
11. **Jadhav, S., et al.** 2011. Virulence characteristics and genetic affinities of multiple drug resistant uropathogenic *Escherichia coli* from a semi urban locality in India. PLoS One **6:**e18063.
12. **Johnson, J. R., B. Johnston, M. A. Kuskowski, J. P. Nougayrede, and E. Oswald.** 2008. Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island. J. Clin. Microbiol. **46:**3906–3911.
13. Reference deleted.
14. **Kaper, J. B., J. P. Nataro, and H. L. T. Mobley.** 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. **2:**123.
15. **Kent, W.** 2002. BLAT—the BLAST-like alignment tool. Genome Res. **12:**656–664.
16. **Kumarasamy, K. K., M. A. Toleman, T. R. Walsh, J. Bagaria, F. Butt, et al.** 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect. Dis. **10:**597–602.
17. **Kurtz, S., et al.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5:**R12.
18. **Larsen, T. S., and A. Krogh.** 2003. EasyGene—a prokaryotic gene finder that ranks ORFs by statistical significance. BMC Bioinformatics **4:**21.
19. **Lau, S. H., et al.** 2008. UK epidemic *Escherichia coli* strains A-E, with CTX-M-15 b-lactamase, all belong to the international O25:H4-ST131 clone. J. Antimicrob. Chemother. **62:**1241–1244.
20. **Livermore, D. M.** 2009. Beta-lactamases—the threat renews. Curr. Protein Pept. Sci. **10:**397–400.
21. Reference deleted.
22. **Miriagou, V., et al.** 2010. Acquired carbapenemases in Gram-negative bacterial pathogens: detection and surveillance issues. Clin. Microbiol. Infect. **16:**112–122.
23. **Rogers, B. A., H. E. Sidjabat, and D. L. Paterson.** 2011. *Escherichia coli* O25b-ST131: a pandemic, multiresistant, community-associated strain. J. Antimicrob. Chemother. **66:**1–14.
24. **Rutherford, K., et al.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945.
25. **Toh, H., et al.** 2010. Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. J. Bacteriol. **192:**1165–1166.
26. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821–829.

Downloaded from http://jb.asm.org/ on October 19, 2019 by guest