Draft Genome Sequences of Three *Escherichia coli* Strains Investigated for the Effects of Lysogeny on Niche Diversification

Jennifer Y. H. Lai,ᵃ Hao Zhang,ᵃ Miranda H. Y. Chiang,ᵃ Min Yu,ᵃ Rui Zhang,ᵇ Stanley C. K. Lauᵃ

Division of Life Science, the Hong Kong University of Science and Technology, Hong Kongᵃ; State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, Chinaᵇ

J.Y.H.L. and H.Z. contributed equally to this work.

During the course of investigating the effects of lysogeny on niche diversification of *Escherichia coli*, we used the temperate phages induced from one *E. coli* strain to infect another and created an isogenic lysogen of the latter. The draft genome sequences of the three *E. coli* strains are reported herein.

Address correspondence to Rui Zhang, ruizhang@xmu.edu.cn, or Stanley C. K. Lau, scklau@ust.hk.

Prophages have contributed 26% of strain-specific genes to the pangenome currently identified for commensal and pathogenic *Escherichia coli* (1). Through the addition of new genes and the regulation of host genes, prophages have profound effects on the metabolism, physiology, and hence ecological niche of *E. coli* (2–4). In our investigation of lysogeny as a mechanism mediating the survival of *E. coli* outside animal hosts, we isolated *E. coli* E1140 from marine sediment and E455 from pig feces. Prophages associated with E1140 were induced to release phage particles using mitomycin C (5, 6). The induced phages were taken to infect E455. An isogenic lysogen, namely, E455L was thus created.

The draft genome sequences of the three strains were obtained using 300-bp insert pair-end libraries on the Illumina HiSeq2000 system and 10 kb-insert single molecule, real-time (SMRT) technology on the PacBio RS system. The reads were *de novo* assembled manually into contigs using CLC Genomics Workbench version 6.5.1 (CLC bio, USA) and the Celera Assembler version 8 (7). We obtained 11 contigs of 1,804 to 2,924,481 bp in size ($N_{50} = 2,924,481$ bp) with a total size of 5,111,617 bp for E1140, 6 contigs of 5,215 to 3,090,951 bp in size ($N_{50} = 3,090,951$ bp) with a total size of 4,587,712 bp for E455, and 15 contigs of 5,720 to 2,472,449 bp in size ($N_{50} = 2,472,449$ bp) with a total size of 4,592,228 bp for E455L.

Using the Glimmer2 (8), the numbers of protein coding sequences predicted were 4,989 for E1140, 4,353 for E455, and 4,390 for E455L. The prophages in the draft genome sequences were identified using the PHAge Search Tool (PHAST) (9). E455 was found to carry two prophages. E455L carried not only the two prophages of E455, but also an extra P2 prophage that was 30,983 bp in size and comprised of 37 phage genes and 6 genes of unknown functions. E1140 carried seven prophages, including a P2 that was identical to the one found in E455L. These results suggest that E455 had received a P2 prophage of E1140 during lysogenic infection, resulting in an isogenic lysogen E455L.

Nucleotide sequence accession numbers. These whole-genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession numbers JEND00000000, JENE00000000, and JDFU00000000 for *E. coli* strains E455, E455L, and E1140, respectively. The versions described in this paper are JEND02000000, JENE03000000, and JDFU02000000 for E455, E455L, and E1140, respectively.

ACKNOWLEDGMENTS

REFERENCES

1. **Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Médigue C, Rocha EP, Denamur E.** 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet. 5:e1000344. http://dx.doi.org/10.1371/journal.pgen.1000344.
2. **Chen Y, Golding I, Sawai S, Guo L, Cox EC.** 2005. Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. PLoS Biol. 3:e229. http://dx.doi.org/10.1371/journal.pbio.0030229.
3. **Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK.** 2010. Cryptic prophages help bacteria cope with adverse environments. Nat. Commun. 1:147. http://dx.doi.org/10.1038/ncomms1146.
4. **Wick LM, Qi W, Lacher DW, Whittam TS.** 2005. Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. J. Bacteriol. 187:1783–1791. http://dx.doi.org/10.1128/JB.187.5.1783-1791.2005.
5. **Keller KL, Overbeck-Carrick TL, Beck DJ.** 2001. Survival and induction of SOS in *Escherichia coli* treated with cisplatin, UV-irradiation, or mitomycin C are dependent on the function of the RecBC and RecFOR pathways of homologous recombination. Mutat. Res. 486:21–29. http://dx.doi.org/10.1016/S0921-8777(01)00077-5.
6. **Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S.** 2005. Switches

in bacteriophage lambda development. Annu. Rev. Genet. **39:**409–429. http://dx.doi.org/10.1146/annurev.genet.39.073003.113656.

7. **Denisov G, Walenz B, Halpern AL, Miller J, Axelrod N, Levy S, Sutton G.** 2008. Consensus generation and variant detection by Celera Assembler. Bioinformatics **24:**1035–1040. http://dx.doi.org/10.1093/bioinformatics/btn074.

8. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:** 4636–4641. http://dx.doi.org/10.1093/nar/27.23.4636.

9. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res. **39**(Suppl 2):W347–W352. http://dx.doi.org/10.1093/nar/gkr485.