

## Open-Source Genomic Analysis of Shiga-Toxin-Producing *E. coli* O104:H4

Loman, Nicholas; Pallen, Mark; Rohde, Holger; Qin, Junjie ; Cui, Yujun; Li, Dongfang; Hentschke, Moritz ; Chen, Wentong; Pu, Fei ; Peng, Yangqing ; Li, Junhua ; Xi, Feng ; Li, Shenghui ; Li, Yin Li; Zhang, Zhaoxi ; Yang, Xianwei ; Zhao, Meiru ; Wang, Peng ; Guan, Yuanlin ; Cen, Zhong

DOI:

[10.1056/NEJMoa1107643](https://doi.org/10.1056/NEJMoa1107643)

### Document Version

Peer reviewed version

### Citation for published version (Harvard):

Loman, N, Pallen, M, Rohde, H, Qin, J, Cui, Y, Li, D, Hentschke, M, Chen, W, Pu, F, Peng, Y, Li, J, Xi, F, Li, S, Li, YL, Zhang, Z, Yang, X, Zhao, M, Wang, P, Guan, Y, Cen, Z, Zhao, X, Christner, M, Kobbe, R, Loos, S, Oh, J, Yang, L, Danchin, A, Gao, GF, Song, Y, Li, Y, Yang, H, Wang, J, Xu, J, Wang, J, Aepfelbacher, M & Yang, R 2011, 'Open-Source Genomic Analysis of Shiga-Toxin-Producing *E. coli* O104:H4', *New England Journal of Medicine*, vol. 365, no. 8, pp. 718-724. <https://doi.org/10.1056/NEJMoa1107643>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

BRIEF REPORT

## Open-Source Genomic Analysis of Shiga-Toxin–Producing *E. coli* O104:H4

Holger Rohde, M.D., Junjie Qin, Ph.D., Yujun Cui, Ph.D., Dongfang Li, M.E., Nicholas J. Loman, M.B., B.S., Moritz Hentschke, M.D., Wentong Chen, B.S., Fei Pu, B.S., Yangqing Peng, B.S., Junhua Li, B.E., Feng Xi, B.E., Shenghui Li, B.S., Yin Li, B.S., Zhaoxi Zhang, B.S., Xianwei Yang, B.S., Meiru Zhao, M.S., Peng Wang, B.M., Yuanlin Guan, B.E., Zhong Cen, M.E., Xiangna Zhao, B.S., Martin Christner, M.D., Robin Kobbe, M.D., Sebastian Loos, M.D., Jun Oh, M.D., Liang Yang, Ph.D., Antoine Danchin, Ph.D., George F. Gao, Ph.D., Yajun Song, Ph.D., Yingrui Li, B.S., Huanming Yang, Ph.D., Jian Wang, Ph.D., Jianguo Xu, M.D., Ph.D., Mark J. Pallen, M.D., Ph.D., Jun Wang, Ph.D., Martin Aepfelbacher, M.D., Ruifu Yang, M.D., Ph.D., and the *E. coli* O104:H4 Genome Analysis Crowd-Sourcing Consortium\*

### SUMMARY

An outbreak caused by Shiga-toxin–producing *Escherichia coli* O104:H4 occurred in Germany in May and June of 2011, with more than 3000 persons infected. Here, we report a cluster of cases associated with a single family and describe an open-source genomic analysis of an isolate from one member of the family. This analysis involved the use of rapid, bench-top DNA sequencing technology, open-source data release, and prompt crowd-sourced analyses. In less than a week, these studies revealed that the outbreak strain belonged to an enteroaggregative *E. coli* lineage that had acquired genes for Shiga toxin 2 and for antibiotic resistance.

**E**SCHERICHIA COLI IS A WIDESPREAD COMMENSAL OF THE MAMMALIAN GUT and a versatile pathogen.<sup>1,2</sup> Enterovirulent strains of *E. coli* are classified into a number of overlapping pathotypes, which include Shiga-toxin–producing, enterohemorrhagic, and enteroaggregative varieties.<sup>2</sup> Enteroaggregative *E. coli* strains have been associated with sporadic and epidemic diarrhea and, in the laboratory, show a distinctive pattern of adherence to Hep-2 cells (termed aggregative, or “stacked brick”).<sup>3</sup> In Shiga-toxin–producing *E. coli*, the toxin is encoded on a prophage and inhibits protein synthesis within susceptible eukaryotic cells. Strains of enterohemorrhagic *E. coli* produce Shiga toxin and a specific protein secretion system (called a type III secretion system) that is encoded by the locus of enterocyte effacement (LEE) and that is responsible for attachment to the intestine.<sup>2</sup> Shiga-toxin–producing and enterohemorrhagic *E. coli* strains are commonly associated with the hemolytic–uremic syndrome, a combination of renal impairment, thrombocytopenia, and hemolytic anemia that is often accompanied by neurologic and myocardial damage.

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Pallen at the Centre for Systems Biology, University of Birmingham, B15 2TT, United Kingdom, or at [m.pallen@bham.ac.uk](mailto:m.pallen@bham.ac.uk); or to Dr. Yang at BGI-Shenzhen, Shenzhen 518083, China, or at the Beijing Institute of Microbiology and Epidemiology, 20 Dongda St., Beijing 100071, China, or at [yangruifu@genomics.org.cn](mailto:yangruifu@genomics.org.cn).

The following two groups of authors contributed equally to this article: Drs. Rohde, Qin, Cui, D. Li, and Loman; and Drs. Pallen, J. Wang, Aepfelbacher, and R. Yang.

\*Members of the *E. coli* O104:H4 Genome Analysis Crowd-Sourcing Consortium are listed in the Supplementary Appendix, available at NEJM.org.

This article (10.1056/NEJMoa1107643) was published on July 27, 2011, at NEJM.org.

N Engl J Med 2011.

Copyright © 2011 Massachusetts Medical Society.

More than 3000 cases of infection with an unusual strain of Shiga-toxin-producing *E. coli* O104:H4 were reported to the Robert Koch Institute in Berlin during a nationwide outbreak in Germany in May and June of 2011.<sup>4</sup> This outbreak resulted in more than 40 deaths, and associated cases were reported in more than a dozen countries in Europe and North America (mostly in travelers returning from Germany). Household transmission was described in the Netherlands, and life-threatening colonic ischemia was reported as a complication in addition to the hemolytic-uremic syndrome and bloody diarrhea.<sup>5,6</sup> Epidemiologic and microbiologic evidence indicated that the O104:H4 strain was distributed throughout Germany on bean sprouts.<sup>7</sup>

The outbreak was characterized by several unusual features: a high incidence in adults (especially women), a greatly increased incidence of the hemolytic-uremic syndrome (in approximately 25% of patients, as compared with 1 to 15% in previous outbreaks of Shiga-toxin-producing *E. coli*), a predominance of female patients among cases of the hemolytic-uremic syndrome, and a rare serotype of Shiga-toxin-producing *E. coli* that had been linked to only two sporadic cases of the hemolytic-uremic syndrome (one in Germany and the other in South Korea).<sup>4,8,9</sup> Recognition of infection during the outbreak was hampered by a laboratory approach that targeted phenotypes associated with the most common lineage of enterohemorrhagic *E. coli* (the non-sorbitol-fermenting O157:H7 serotype) rather than one aimed at finding all strains of Shiga-toxin-producing *E. coli*.<sup>10</sup> Here, we report a local cluster of cases associated with a family from northern Germany and describe an open-source genomic analysis of an isolate from the family cluster.

#### CASE REPORTS

On May 17, 2011, a 16-year-old girl was admitted to the pediatric emergency ward at the University Medical Center Hamburg-Eppendorf with bloody diarrhea and abdominal pain. Her laboratory values were normal. Later on the same day, her 12-year-old brother was admitted with a 2-day history of malaise and headache and a 1-day history of vomiting and nonbloody diarrhea. The boy presented with acute renal failure (serum creatinine level, 4.1 mg per deciliter [362  $\mu$ mol per liter]; and potassium level, 6 mmol per liter), thrombocytopenia (22,000

platelets per cubic millimeter), and hemolytic anemia (hemoglobin, 11.6 g per deciliter; bilirubin, 2.8 mg per deciliter [49  $\mu$ mol per liter]; and lactate dehydrogenase, 2297 U per liter). His hemoglobin level fell to 8.4 g per deciliter within 48 hours after admission, thereby fulfilling the case definition of the hemolytic-uremic syndrome.

The children, their parents, and a teenage friend had eaten a meal together a week earlier. The meal included a freshly prepared salad containing bean sprouts. The children's mother had no symptoms, and no Shiga-toxin-producing *E. coli* was isolated from her stool. However, the hemolytic-uremic syndrome developed in the father, and his stool sample was culture-positive for Shiga-toxin-producing *E. coli*. The teenage friend had diarrhea but was not admitted to the medical center.

Stool samples from the siblings were plated on Sorbitol-MacConkey agar and incubated in a liquid enrichment culture. The next day, supernatants from the liquid cultures tested positive for Shiga toxin on enzyme-linked immunosorbent assay. Uniformly sorbitol-positive colonies were identified as *E. coli* on MALDI-TOF (matrix-assisted laser desorption ionization-time of flight) mass spectrometry. Several single colonies were positive for the *stx2* gene and negative for the *stx1* and *eae* genes on polymerase-chain-reaction (PCR) assay. None of the isolates agglutinated with polyvalent serum samples directed against the serotypes that are most frequently associated with Shiga-toxin-producing *E. coli*. Subsequent analyses showed that the strain belonged to the rare serotype O104:H4 harboring an extended-spectrum beta-lactamase (ESBL) gene of the CTX-M-15 class.

Although our 16-year-old patient had a mild course of disease without the hemolytic-uremic syndrome and was discharged from the hospital on the same day, the clinical picture for her brother was much less benign. The boy's renal function, hemoglobin level, and thrombocytopenia improved after 9 days of peritoneal dialysis, but severe neurologic symptoms, including somnolence, visual impairment, speech disturbances, hemiplegia, and incontinence, developed. He underwent four cycles of plasmapheresis and therapy with the anti-C5-antibody eculizumab. After this treatment, his clinical condition improved, and he was discharged after 24 days with serum creatinine levels just above the normal range. However, he was left with neurologic sequelae and required rehabilitation.

## METHODS AND RESULTS

**OPEN-SOURCE GENOMICS**

To investigate the evolutionary origins and pathogenic potential of the outbreak strain, we set in motion an open-source genomics program of research that incorporated new high-throughput sequencing approaches, public data release, and rapid outsourcing of analyses to bioinformaticians worldwide (crowd-sourcing) (Fig. 1). Initially, we sequenced the genome of the isolate from the 16-year-old girl (TY2482), using the Ion Torrent Personal Genome Machine (PGM), and obtained an initial draft of the genome 3 days after receipt of the DNA sample. Three DNA libraries were prepared and seven sequencing runs performed, following the protocols of the manufacturer (Life Technologies), to generate 79 Mb of sequence data, with an average read length of 101 bp. (For details regarding the sequencing procedures, see the Supplementary Appendix, available with the full text of this article at NEJM.org.)

We released these data into the public domain under a Creative Commons 0 license, which elicited a burst of crowd-sourced, curiosity-driven analyses carried out by bioinformaticians on four continents.<sup>11</sup> Twenty-four hours after the release of the genome, it had been assembled; 2 days after its dissemination, it had been assigned to an existing sequence type. Five days after the release of the sequence data, we had designed and released strain-specific diagnostic primer sequences, and within a week, two dozen reports had been filed on an open-source wiki (a Web site that facilitates collaborative effort) dedicated to analysis of the strain. These analyses provided timely information on the strain's virulence and resistance genes, in addition to its phylogenetic lineage.

We also performed sequencing on the Illumina HiSeq platform in accordance with the manufacturer's instructions. An initial single-end run was used to correct errors in the Ion Torrent sequence, principally in homopolymeric tracts. We later performed paired-end and mate-pair sequencing on this platform, exploiting libraries with insert sizes of 470 bp, 2 kb, and 6 kb, and generated enough data (1 Gb, 576 Mb, and 576 Mb from each library, respectively) to create a high-quality draft genome sequence within 2 weeks after receipt of the DNA samples. (Additional details are provided in the Supplementary Appendix.) The reads were deposited in GenBank's Short Read Archive with acces-

sion numbers SRA037315 for Ion Torrent reads and SRA039136 for Illumina platform reads.

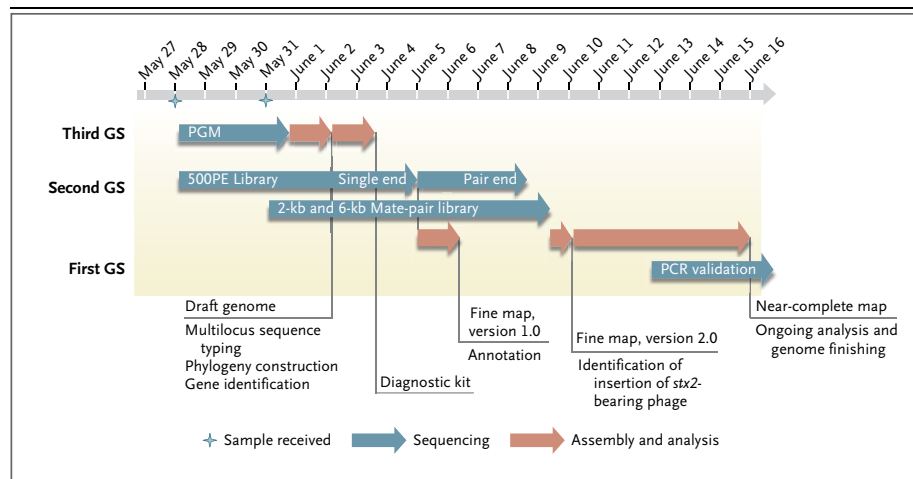
**PHYLOGENETIC ANALYSIS**

The assembled Ion Torrent data provided gene sequences that could be analyzed with an existing multilocus-sequence-typing scheme for *E. coli* that relied on sequence comparisons for seven conserved housekeeping genes (*adh*, *fumC*, *gyrB*, *mdh*, *purA*, *recA*, and *icd*).<sup>12</sup> This analysis revealed a close relationship to a strain, 01-09591, which was isolated in Germany in 2001 and which fell into sequence type ST678. The TY2482 sequences differed from the profile of the 2001 strain by a single base pair in the *adh* gene and a single-base difference in a homopolymeric sequence in the *recA* gene. (We subsequently discovered that the latter difference was a sequencing error generated by the PGM.) The 2001 strain, which produced Shiga toxin and was associated with the hemolytic-uremic syndrome, fell into the O104:H4 serotype but did not have the genes associated with type III secretion in typical enterohemorrhagic *E. coli*.<sup>13,14</sup> Additional scrutiny of the multilocus-sequence-typing database revealed that strains with the broad O104 serotype were scattered across several sequence types, whereas strains with the narrower O104:H4 serotype appeared to be limited to ST678.<sup>10</sup>

Comparisons of the TY2482 genome with all previously sequenced complete genomes of *E. coli* isolates revealed a very close relationship to *E. coli* strain 55989, with an average nucleotide identity of 99.8% (see the Supplementary Appendix). This strain was isolated in the Central African Republic from a stool sample obtained from an adult with human immunodeficiency virus infection who had persistent watery diarrhea.<sup>15</sup> It has been classified as an enteroaggregative *E. coli*, but unlike TY2482, it does not have Shiga toxin genes.<sup>15</sup> However, it is worth noting that Mossoro et al.,<sup>15</sup> who first described *E. coli* strain 55989, also described strains of enteroaggregative *E. coli* with Shiga toxin genes in the same human population.<sup>15</sup>

**COMPARISON OF THE CHROMOSOMES OF TY2482 AND 55989**

Isolates from the German outbreak were initially described as enterohemorrhagic *E. coli*. However, the close relationship between TY2482 and 55989 led us to consider the likelihood that TY2482 is an enteroaggregative *E. coli*. Our analysis of the gene content of TY2482 showed that it, like 01-09591,



**Figure 1. Timeline of the Open-Source Genomics Program.**

After receiving the first batch of DNA samples on May 28, 2011, sequencing runs with the use of the Ion Torrent Personal Genome Machine (PGM) and Illumina (small-insert library) were initiated simultaneously. On May 31, the second batch of DNA was received and used for Illumina large-insert sequencing. An assembly of the Ion Torrent reads was released on June 2, which enabled subsequent analyses (multilocus sequence typing, phylogenetic analysis, and genome comparisons). Errors in the Ion Torrent data were corrected with the use of later Illumina data, and a high-quality draft genome sequence was created. GS denotes generation of sequencing technology. The symbols at May 28 and May 31 in the timeline indicate the arrival of DNA samples.

lacked the LEE and genes encoding effectors associated with type III secretion.<sup>16</sup> Instead, we found that the TY2482 genome encodes virulence factors that are typical of enteroaggregative *E. coli*. Other investigators working on the outbreak strain have also observed genes typically found in enteroaggregative strains on PCR assay and have noted a behavioral phenotype that is characteristic of this pathotype on cell-adherence assay.<sup>17</sup>

To identify strain-specific genes, we performed a detailed comparison of the chromosomes of TY2482 and enteroaggregative *E. coli* strain 55989. First, we aligned the TY2482 assembly against the 55989 chromosome (for details, see the Supplementary Appendix). We then adopted the gene predictions and annotation from the 55989 genome for these conserved sequences. Next, we identified several isolate-specific regions of difference (i.e., regions present in the TY2482 chromosome and absent from the 55989 genome or vice versa) that were more than 5 kb (Table 1 and Fig. 2, and the Supplementary Appendix). TY2482-specific regions of difference included prophage remnants or apparently intact prophages, such as the *stx2* prophage, which, like its close relatives in the genomes of O157:H7 strains EDL933 and

Sakai, is inserted into the *wrbA* locus. The *stx2* genes differ by only one single-nucleotide polymorphism from the *stx2* allele seen in O157 enterohemorrhagic *E. coli* strain EDL933.

**TY2482 PLASMIDS**

From our de novo assembly (i.e., assembly without the use of a reference genome), we concluded that the TY2482 genome contains two large conjugative plasmids, pESBL TY2482 and pAA TY2482, and a small plasmid, pG2011 TY2482 (Fig. 2). From scrutiny of copy numbers of sequence reads, it was clear that the two large plasmids were replicating at an approximate ratio of 1:1 with the chromosome, whereas the small plasmid was maintained at a copy number at least nine times that of the other replicons. No phenotype could be ascribed to the small plasmid.

The largest plasmid, pESBL TY2482, was an IncI plasmid similar to pEC\_Bactec, which was found in an *E. coli* strain isolated from the joint of a horse with arthritis.<sup>18</sup> The pESBL TY2482 plasmid encodes a CTX-M-15 ESBL, as well as a beta-lactamase from the TEM class. The second large plasmid, pAA TY2482, resembled a plasmid from strain 55989 but carried a gene cluster encoding a

**Table 1. Genetic Elements in Strain TY2482 of Shiga-Toxin–Producing *Escherichia coli* O104:H4.**

| Genetic Element             | Notable Features or Functions   | Size or 55989 Coordinates*  |
|-----------------------------|---|-----------------------------|
| <b>Plasmid</b>              |   |                             |
| pESBL TY2482                | IncI1 plasmid, homologous to pEC_Bactec carrying <i>bla</i> CTX-M-15  | 88 kb                       |
| pAA TY2482                  | Plasmid encoding aggregative adherence fimbriae I   | 76 kb                       |
| pG2011 TY2482               | Plasmid with no obvious phenotype   | 1.5 kb                      |
| <b>Region of difference</b> |   |                             |
| I-ROD1                      | Degenerate prophage   | 296227 (tRNA- <i>Thr</i> )  |
| I-ROD2                      | <i>Stx2</i> -encoding prophage  | 1176265 ( <i>wrbA</i> )     |
| I-ROD3                      | Microcin gene cluster; tellurite resistance gene cluster  | 1207704 (tRNA- <i>Ser</i> ) |
| I-ROD4                      | Prophage  | 1811905 ( <i>ymfG</i> )     |
| I-ROD5                      | Prophage  | 2102453 ( <i>yecE</i> )     |
| I-ROD6                      | Molybdate metabolism regulator; <i>yehL</i>   | 2426442 (IS1)               |
| I-ROD7                      | Multidrug-resistant gene cluster ( <i>dfA7</i> , <i>sull</i> , <i>sullI</i> , <i>strA</i> , <i>strB</i> , <i>tetA</i> ); mercury resistance | 4211244 (tRNA- <i>Sec</i> ) |
| D-ROD1                      | Prophage  | 1094587–1140306             |
| D-ROD2                      | Prophage  | 1413924–1446834             |
| D-ROD3                      | Prophage  | 1754689–1800354             |
| D-ROD4                      | Prophage  | 2688656–2701228             |
| D-ROD5                      | Type VI secretion genes   | 3401720–3427357             |
| D-ROD6                      | Prophage  | 4944269–5004333             |

\* Coordinates from the genome of *E. coli* strain 55989 are given for predicted boundaries of regions of difference, with the gene carrying the insertion site shown in parentheses for a region of difference involving an insertion into 55989 (I-ROD). D-ROD denotes a region of difference involving a deletion.

rare type of aggregative adherence fimbria (AAF/I) instead of the more common type (AAF/III) encoded by genes in the 55989 plasmid. We exploited this AAF/I cluster as a target for strain-specific PCR primers as part of a suite of primers to identify the outbreak isolate.

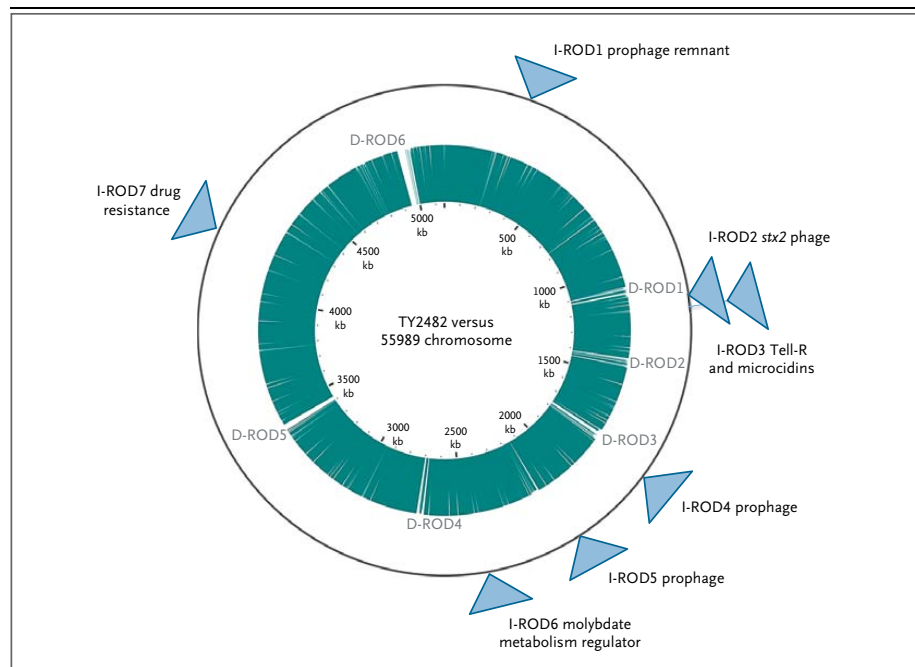
#### DISCUSSION

Our genomic analyses suggest that the German outbreak strain evolved from a progenitor that belonged to the enteroaggregative pathotype and resembled strain 55989. The emergence of the outbreak strain depended on the acquisition of a *stx2* prophage and of a plasmid encoding a CTX-M-15 ESBL. Sometime during this process, the strain also appears to have lost one gene cluster, encoding AAF/III fimbriae, and gained another, encoding the rarer AAF/I fimbriae.

Although this outbreak strain has surprised the general public and public health officials, related potential progenitor strains have been reported from three continents. The appearance of

an O104:H4 strain associated with the hemolytic–uremic syndrome in Korea in 2005 is unexplained, and its link to the German outbreak is unclear.<sup>9</sup> Also, the O104:H4 strain 01-09591 that was isolated in Germany in 2001 urgently requires further investigation. Both strains should undergo genome sequencing and comparison with TY2482. The link to strain 55989, which was isolated in the Central African Republic in the late 1990s, is also intriguing. Genome sequencing of additional Central African isolates from the study that yielded 55989 is likely to illuminate the evolution of this lineage and of enterovirulent *E. coli* in general (see the article by Rasko et al. elsewhere in this issue of the *Journal*<sup>19</sup>).

Although the genome sequence alone cannot provide a full explanation for the high degree of virulence of this strain, it prompts a reassessment of our assumptions and provides a framework for future hypothesis-driven research. Both commensal and pathogenic varieties of *E. coli* have to survive in the gut. However, mere survival, even if twinned with the production of Shiga toxin, is probably



**Figure 2. Comparison of the TY2482 and 55989 Genomes.**

The outer circle depicts the *Escherichia coli* 55989 chromosome. The inner circle represents the TY2482 assembly mapped against the 55989 chromosome. Regions of difference (ROD) that are present in 55989 but not TY2482 (D-RODs) are shown as gaps in the inner circle. The positions of RODs that were found in TY2482 but not 55989 (I-RODs) are shown as wedges on the outer circle at positions corresponding to the predicted insertion sites. Tell-R denotes tellurite resistance.

not enough to cause the hemolytic–uremic syndrome or bloody diarrhea. For that, the bacteria would probably need to adhere to the gut mucosa. In the past, much research has been concentrated on the adhesion systems of typical enterohemorrhagic *E. coli*, particularly the LEE-encoded type III secretion system.<sup>16,20</sup> This German outbreak strain shows us that Shiga-toxin–producing *E. coli* can exploit alternative adhesion mechanisms, very likely including aggregative adherence fimbriae, to the same end. This strain also shows that pathotypes of *E. coli* can overlap and that they evolve rather than stand as fixed archetypes.

It remains unclear why this strain has proved to be so virulent. As noted, a novel suite of adhesins might provide an explanation. Alternatively, perhaps this strain exploits more efficient mechanisms for toxin release. It is worth remembering that strains of enteroaggregative *E. coli* have caused large sprout-associated outbreaks before, including

one outbreak<sup>21</sup> that affected more than 2000 persons in Japan in 1993. Thus, there is clearly an urgent need to understand how the German outbreak strain and other strains of enteroaggregative *E. coli* adhere to and colonize seeds and seedlings.

Our rapid open-source analysis of an outbreak-associated bacterial pathogen was characterized by a propitious confluence of high-throughput genomics, crowd-sourced analyses, and a liberal approach to data release. Although phenotypic or molecular analyses that exploit known virulence, resistance, or epidemiologic targets are useful in diagnostic and public health microbiology, genome sequencing offers the advantages of open-endedness (revealing the “unknown unknowns”), universal applicability, and the ultimate in resolution. Our study shows how benchtop sequencing platforms can generate data with sufficient speed to have an important effect on clinical and epidemiologic problems.

#### BRIEF REPORT

Supported by grants from the State Key Development Program for Basic Research of China (2009CB522600), the National Key Program for Infectious Diseases of China (2008ZX10004-009), Shenzhen Biological Industry Development Special Foundation—Basic Research Key Projects (JC201005250088A), Key Laboratory Project Supported by Shenzhen City (ZD200806180054A), the European Union Microme Program (FP7-KBBE-2007-3-2-08-222886), the Alexander von Humboldt Foundation (to Dr. L. Yang), the Medi-

cal Faculty of the University Medical Center Hamburg—Eppendorf, and the British Biotechnology and Biological Sciences Research Council (BB/E011179/1).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank David Vallenet, Claudine Médigue, Xiaoning Wang, and Jennifer Gardy for their helpful discussions.

#### APPENDIX

The authors' affiliations are as follows: the Institute of Medical Microbiology, Virology and Hygiene (H.R., M.H., M.C., L.Y., M.A.) and the Department of Pediatrics (R.K., S. Loos, J.O.), University Medical Center Hamburg—Eppendorf, Hamburg, Germany; BGI-Shenzhen, Shenzhen, China (J.Q., Y.C., D.L., W.C., F.P., Y.P., J.L., F.X., S. Li, Yin Li, Z.Z., Z.C., Yingrui Li, H.Y., Jian Wang, Jun Wang, R.Y.), and the School of Bioscience and Biotechnology, South China University of Technology, Guangzhou, China (J.L.); the State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology (Y.C., X.Z., Y.S., R.Y.), CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Science (G.F.G.), and the State Key Laboratory for Infectious Disease Prevention and Control and National Institute for Communicable Diseases Control and Prevention, Chinese Center for Disease Control and Prevention (G.F.G., J.X.) — all in Beijing; the Centre for Systems Biology, University of Birmingham, Birmingham, United Kingdom (N.J.L. M.J.P.); and AMAbiotics, Evry, France (A.D.).

#### REFERENCES

1. Hobman JL, Penn CW, Pallen MJ. Laboratory strains of *Escherichia coli*: model citizens or deceitful delinquents growing old disgracefully? *Mol Microbiol* 2007; 64:881-5.
2. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2004;2:123-40.
3. Weintraub A. Enterohaggregative *Escherichia coli*: epidemiology, virulence and detection. *J Med Microbiol* 2007;56:4-8.
4. Frank C, Werber D, Cramer JP, et al. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany — preliminary report. *N Engl J Med* 2011. DOI: 10.1056/NEJMoa1106483.
5. Kuijper E, Soonawala D, Vermont C, van Dissel J. Household transmission of haemolytic uraemic syndrome associated with *Escherichia coli* O104:H4 in the Netherlands. *Euro Surveill* 2011;16:pii:19897.
6. Cordesmeier S, Peitz U, Godde N, Kasper H, Hoffmann M, Allemeyer E. Colonic ischaemia as a severe Shiga toxin/verotoxin producing *Escherichia coli* O104:H4 complication in a patient without haemolytic uraemic syndrome, Germany, June 2011. *Euro Surveill* 2011;16:pii:19895.
7. Robert Koch Institute. Information update on EHEC/HUS outbreak, 2011. ([http://www.rki.de/nm\\_217400/EN/Home/PM082011.html](http://www.rki.de/nm_217400/EN/Home/PM082011.html)).
8. Scheut F, Moller Nielsen E, Frimodt-Moller J, et al. Characteristics of the enterohaggregative Shiga toxin/verotoxin-producing *Escherichia coli* O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011. *Euro Surveill* 2011;16:pii:19889.
9. Bae WK, Lee YK, Cho MS, et al. A case of hemolytic uraemic syndrome caused by *Escherichia coli* O104:H4. *Yonsei Med J* 2006;47:437-9.
10. Chattaway MA, Dallman T, Okeke IN, Wain J. Enterohaggregative *E. coli* O104 from an outbreak of HUS in Germany 2011, could it happen again? *J Infect Dev Ctries* 2011;5:425-36.
11. GitHub. *E. coli* O104:H4 genome analysis crowd sourcing, 2011. (<https://github.com/ehec-outbreak-crowdsourced/BGI-data-analysis/wiki>).
12. University College Cork. *Escherichia coli* MLST Database, 2011. (<http://mlst.ucc.ie/mlst/dbs/Ecoli>).
13. Mellmann A, Bielaszewska M, Kock R, et al. Analysis of collection of hemolytic uraemic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg Infect Dis* 2008;14:1287-90.
14. Creuzburg K, Middendorf B, Mellmann A, et al. Evolutionary analysis and distribution of type III effector genes in pathogenic *Escherichia coli* from human, animal and food sources. *Environ Microbiol* 2011;13:439-52.
15. Mossoro C, Glaziou P, Yassibanda S, et al. Chronic diarrhea, hemorrhagic colitis, and hemolytic-uraemic syndrome associated with HEP-2 adherent *Escherichia coli* in adults infected with human immunodeficiency virus in Bangui, Central African Republic. *J Clin Microbiol* 2002;40:3086-8.
16. Tobe T, Beatson SA, Taniguchi H, et al. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc Natl Acad Sci U S A* 2006;103:14941-6.
17. Bielaszewska M, Mellmann A, Zhang W, et al. Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis* 2011 June 22 (Epub ahead of print).
18. Smet A, Van Nieuwerburgh FV, Vandekerckhove TTM, et al. Complete nucleotide sequence of CTX-M-15-plasmids from clinical *Escherichia coli* isolates: insertional events of transposons and insertion sequences. *PLoS ONE* 2010;5(6):e11202.
19. Rasko DA, Webster DR, Sahl JW, et al. Origins of the *E. coli* strain causing an outbreak of hemolytic-uraemic syndrome in Germany. *N Engl J Med* 2011. DOI: 10.1056/NEJMoa1106920.
20. Schmidt MA. LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* 2010;12:1544-52.
21. Itoh Y, Nagano I, Kunishima M, Ezaki T. Laboratory investigation of enterohaggregative *Escherichia coli* O untypeable:H10 associated with a massive outbreak of gastrointestinal illness. *J Clin Microbiol* 1997;35:2546-50.

Copyright © 2011 Massachusetts Medical Society.



## **Supplementary Appendix: Open-source genomics of a Shiga-toxin-producing**

### ***Escherichia coli* O104:H4**

#### **Crowd-sourcing consortium**

The following members of the *E. coli* O104:H4 Genome Analysis Crowd-sourcing consortium made contributions that influenced the analyses reported here: Kathryn E. Holt, David J. Studholme, Michael Feldgarden and Marina Manrique.

A full account of crowd-sourcing efforts can be accessed here: <https://github.com/ehec-outbreak-crowdsourced/BGI-data-analysis/wiki/>

#### **Methods and Results**

##### *Ion Torrent library construction and sequencing*

Genomic DNA was extracted and purified using a conventional SDS lysis and phenol-chloroform method. 5µg of DNA (OD260/OD280 = 1.85) was dissolved in TE buffer to a total volume of 100 µl and fragmented by sonication (Covaris S2, Massachusetts, USA) to a size distribution of 50-300 bp. Library preparation and template preparation of live Ion Sphere™ Particles was performed according to the manufacturer's protocol (Ion Torrent, USA). During the library preparation, nick-translation was followed by 5 cycles of PCR amplification. Finally the sequencing was performed on the PGM Sequencer. Seven 314 chips were run to generate 79.1 Mb of sequence, with average length of 101 bp.

##### *Illumina library construction and sequencing*

Whole-genome sequencing was performed using Illumina HiSeq 2000 (Illumina Inc. U.S.A) by generating paired-end libraries with an average insert size of 470 bp, 2 kbp and 6 kbp

following the manufacturer's instruction. The read lengths were 90bp, 50bp, 50bp and 1Gb, 576Mb and 576Mb high quality data were generated from each library respectively.

*Creation of a draft genome assembly using Ion Torrent PGM data (2nd June 2011)*

An assembly was performed using MIRA 3.2.1.17\_dev using command-line parameters --*job=denovo,genome,accurate,iontor -GE:not=1*. The Ion Torrent PGM assembly from 5 chips of Ion Torrent 314 data produced an assembly of 3,057 contigs, total bases: 5,491,032 with an N50 value of 3,675.

*Creation of a hybrid assembly using Ion Torrent PGM data and Illumina single-end data (6th June 2011)*

Ion Torrent and Illumina read data were quality filtered before assembly including removal of adapter contamination. The Ion Torrent PGM assembly from 7 chips of Ion Torrent 314 data were assembled with Newbler 2.0.00.22. Illumina single-end data (taken from the in-progress paired-end run) were assembled using SOAPdenovo 1.06<sup>1</sup> (with *k*-mer of 51 and parameters "-d 1, -R". Assemblies were combined using AMOS minimus2 1.59 with parameters REFCOUNT=0, OVERLAP=50, MINID=94, MAXTRIM=10<sup>2</sup>. The resulting assembly consisted of 451 contigs greater than 200bp with an N50 of 53266bp. The largest contig was 204342bp.

*Creation of a draft genome scaffold assembly using Illumina paired-end and mate-pair reads*

A draft *de novo* assembly was produced using SOAPdenovo version 1.05. Contigs were first assembled using the 470bp paired-end library initially using a *k*-mer value of 45 for de Bruijn graph construction. These were subsequently scaffolded in a hierarchical fashion using 2kb followed by 6kb mate-pair libraries by way of the rank parameter in the SOAPdenovo configuration file. Other parameters supplied to SOAPdenovo included -F to attempt to fill

gaps in scaffolds. Where possible, in order to fill remaining scaffold gaps, local information available from the abundant mate-pair data was utilised by the GapCloser utility which was run over the assembly output with a  $k$ -mer size of 23. Both scaffolds and un-scaffolded contigs were used in further analysis, with the exception of contigs smaller than 200bp which were excluded.

*De novo* assembly produced 24 scaffolds plus 75 un-scaffolded contigs. The largest scaffold was 757969bp, the smallest was 552bp. Scaffold N50 was 403980bp. After gap filling the scaffolds contained 143 distinct stretches of gaps (represented as ambiguous 'N' bases) comprising 94491bp of sequence.

#### *Insert sizes*

The estimated insert size with standard deviations predicted by SOAPdenovo are demonstrated in Table S1.

**Table S1. The estimated insert size determined by the *de novo* assembly process.**

| <b>Library</b> | <b>Estimated insert size</b> | <b>Standard deviation</b> |
|----------------|------------------------------|---------------------------|
| 470bp          | 468                          | 31                        |
| 2kb            | 2548                         | 246                       |
| 6kb            | 6193                         | 566                       |

#### *Determination of closest reference by average nucleotide identity (ANI)*

Average nucleotide identity with all complete *E. coli* genomes available in GenBank was calculated using the ANIb algorithm which uses BLAST as the underlying alignment method<sup>3-4</sup>. Scrutiny of results (Table S2) revealed that *E. coli* 55989 showed the highest nucleotide identity with an ANI of 99.8% between the TY2482 draft chromosome and *E. coli*

55989. The ANIb algorithm shreds sequences into 1kb segments. BLAST alignments needed to be longer than 700bp and have >70% nucleotide identity to count towards ANIb calculation. ANIb parameters to BLAST were "-F F -e 0.001 -v 1 -b 1 -X 150 -q -1".

**Table S2. Average nucleotide identities for TY2482 compared against all complete *E.***

***coli* genomes**

| <b>TY2482 vs</b>                          | <b>ANIb</b> |
|---|-------------|
| Escherichia coli 55989                    | 99.84       |
| Escherichia coli IAI1                     | 99.2        |
| Escherichia coli W                        | 99.14       |
| Escherichia coli E24377A                  | 99.09       |
| Escherichia coli SE11                     | 99.09       |
| Escherichia coli O103:H2 str. 12009       | 98.95       |
| Escherichia coli O26:H11 str. 11368       | 98.98       |
| Escherichia coli O111:H- str. 11128       | 98.85       |
| Escherichia coli HS                       | 98.67       |
| Escherichia coli ATCC 8739                | 98.55       |
| Escherichia coli str. K-12 substr. W3110  | 98.54       |
| Escherichia coli str. K-12 substr. MG1655 | 98.54       |
| Escherichia coli DH1                      | 98.54       |
| Escherichia coli BL21-Gold(DE3)plyss AG   | 98.53       |
| Escherichia coli BL21(DE3)                | 98.53       |
| Escherichia coli BL21(DE3)                | 98.53       |
| Escherichia coli B str. REL606            | 98.53       |
| Escherichia coli BW2952                   | 98.49       |
| Escherichia coli str. K-12 substr. DH10B  | 98.5        |
| Escherichia coli H10407                   | 98.5        |
| Escherichia coli ETEC H10407              | 98.5        |
| Escherichia coli O55:H7 str. CB9615       | 97.92       |
| Escherichia coli O157:H7 str. TW14359     | 97.86       |
| Escherichia coli O157:H7 str. Sakai       | 97.87       |
| Escherichia coli O157:H7 str. EC4115      | 97.86       |
| Escherichia coli O157:H7 str. EDL933      | 97.82       |
| Escherichia coli 042                      | 97.45       |
| Escherichia coli UMN026                   | 97.39       |
| Escherichia coli IAI39                    | 97.3        |
| Escherichia coli SMS-3-5                  | 97.21       |
| Escherichia coli SE15                     | 97.06       |
| Escherichia coli CFT073                   | 97.02       |
| Escherichia coli S88                      | 96.97       |

|  |       |
|--|-------|
| Escherichia coli O83:H1 str. NRG 857C  | 96.99 |
| Escherichia coli O127:H6 str. E2348/69 | 96.95 |
| Escherichia coli UM146                 | 96.94 |
| Escherichia coli 536                   | 96.95 |
| Escherichia coli UTI89                 | 96.93 |
| Escherichia coli APEC O1               | 96.98 |
| Escherichia coli ED1a                  | 96.82 |

*Annotation of putative regions of difference between TY2482 and 59989*

The TY2482 scaffold assembly was aligned against *E. coli 55989* using progressiveMauve<sup>5</sup> (part of Mauve 2.3.1) using default settings. For ease of viewing, scaffolds were moved and where necessary reverse complemented to fit the order of the *E. coli 55989* chromosome using the Mauve contig mover, again run with default parameters. Unaligned regions of the TY2482  $\geq$  5kb were examined as putative regions of difference. Gene prediction within these regions was performed using Glimmer 3.02<sup>6</sup> using the g3-iterated.sh workflow with default options. Genes with a raw score of  $\geq$  1.0 were extracted for further analysis. Due to Glimmer mis-predictions when run on the plasmid sequences, plasmids pESBL and pAA instead used Heuristic GeneMark.hmm<sup>7</sup> PROKARYOTIC (version 2.8a) for gene calling. This was run with default settings and model file "heuristic\_no\_rbs.mat" ([http://opal.biology.gatech.edu/GeneMark/heuristic\\_hmm2.cgi](http://opal.biology.gatech.edu/GeneMark/heuristic_hmm2.cgi)). Putative protein products  $\geq$  50 aa in length were searched against the Genbank non-redundant protein database using PHMMER using HMMER (<http://hmmer.janelia.org/>). Genome visualisation plots were generated using CGview<sup>8</sup>.

Manual inspection of scaffolds revealed each plasmid was contained within a single scaffold. Manual curation of pG2011 demonstrated an ~1.5kb plasmid with >99% nucleotide identity to *E. coli* strain H30 plasmid pO26-S1. This plasmid sequence was present as a 2-copy tandem repeat in the assembly, likely an artefact of the mate-pair assembly process (as insert

sizes longer than the plasmid were used) and has been manually edited to form a single copy.

The location of the plasmids in the assembly are as follows: pESBL-TY2482 = scaffold19, pAA-TY2482 = scaffold16, pG2011 = scaffold21.

### **Accession numbers**

The sequencing reads have been deposited into NCBI's Short Read Archive with accession numbers SRR227300, SRR227337, SRR227338, SRR227339, SRR227340, SRR231653, SRR231654 (Ion Torrent) and SRX079806 (Illumina mate-pair), SRX079805 (Illumina mate-pair) , SRX079804 (Illumina paired-end).

The scaffolded assembly and annotation has been deposited to Genbank, accession number AFVR00000000 (draft Illumina scaffold assembly), AFVS00000000 (Ion Torrent assembly) and AFOG01000000 (hybrid Ion Torrent and Illumina single-end assembly).

### **References**

1. Li R, Zhu H, Ruan J, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 2010;20:265-72.
2. Sommer DD, Delcher AL, Salzberg SL, Pop M. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics*. 2007;8:64.
3. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;57:81-91.
4. Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009;106:19126-31.
5. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010;5:e11147.

6. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 2007;23:673-679
7. Besemer J, Borodovsky M. Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* 1999;27:3911-20.
8. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. *Bioinformatics* 2005;21:537-9.

**Table S3. Annotated genes on the RODs and plasmids of TY2482**

| <b>ROD_ID</b> | <b>CDS_ID</b>  | <b>Best hit (NR)</b>                               | <b>Curated annotation</b> | <b>Best hit (species)</b>            |
|---------------|----------------|--|---------------------------|--------------------------------------|
| I-ROD1        | irod1_orf00001 | conserved domain protein                           |                           | Escherichia coli MS 84-1             |
| I-ROD1        | irod1_orf00002 | hypothetical protein ERKG_00886                    |                           | Escherichia coli H252                |
| I-ROD1        | irod1_orf00003 | transposon Tn21 resolvase                          |                           | Escherichia coli B7A                 |
| I-ROD1        | irod1_orf00005 | hypothetical protein ECoL_00180                    |                           | Escherichia coli EC4100B             |
| I-ROD1        | irod1_orf00007 | conserved domain protein                           |                           | Escherichia coli MS 187-1            |
| I-ROD1        | irod1_orf00008 | conserved domain protein                           |                           | Escherichia coli MS 187-1            |
| I-ROD1        | irod1_orf00009 | hypothetical protein HMPREF9550_01817              |                           | Escherichia coli MS 187-1            |
| I-ROD1        | irod1_orf00011 | putative acyl-carrier-protein S-malonyltransferase |                           | Escherichia coli B7A                 |
| I-ROD1        | irod1_orf00012 | hypothetical protein EcB7A_3346                    |                           | Escherichia coli B7A                 |
| I-ROD1        | irod1_orf00013 | hypothetical protein HMPREF9550_01813              |                           | Escherichia coli MS 187-1            |
| I-ROD1        | irod1_orf00015 | hypothetical protein ECoL_00172                    |                           | Escherichia coli EC4100B             |
| I-ROD1        | irod1_orf00016 | hypothetical protein HMPREF9542_01440              |                           | Escherichia coli MS 117-3            |
| I-ROD1        | irod1_orf00017 | phage integrase                                    |                           | Escherichia coli H252                |
| I-ROD2        | irod2_orf00001 | integrase  |                           | Escherichia coli O26:H11 str. 11368  |
| I-ROD2        | irod2_orf00002 | hypothetical bacteriophage protein                 |                           | Escherichia coli O157:H7 str. EC4501 |
| I-ROD2        | irod2_orf00004 | conserved hypothetical protein                     |                           | Escherichia coli O157:H7 str. EC4501 |
| I-ROD2        | irod2_orf00006 | putative antirepressor                             |                           | Escherichia coli O103:H2 str. 12009  |
| I-ROD2        | irod2_orf00007 | hypothetical protein ECO26_1566                    |                           | Escherichia coli O26:H11 str. 11368  |
| I-ROD2        | irod2_orf00008 | hypothetical protein ECO26_1570                    |                           | Escherichia coli O26:H11 str. 11368  |
| I-ROD2        | irod2_orf00009 | gp43   |                           | Escherichia coli B171                |
| I-ROD2        | irod2_orf00010 | conserved hypothetical protein                     |                           | Escherichia coli O157:H7 str. EC508  |
| I-ROD2        | irod2_orf00011 | conserved hypothetical protein                     |                           | Escherichia coli O157:H7 str. EC508  |
| I-ROD2        | irod2_orf00012 | hypothetical protein ECO103_2866                   |                           | Escherichia coli O103:H2 str. 12009  |
| I-ROD2        | irod2_orf00013 | hypothetical protein ECO103_2865                   |                           | Escherichia coli O103:H2 str. 12009  |
| I-ROD2        | irod2_orf00014 | putative exonuclease                               |                           | Escherichia coli O103:H2 str. 12009  |
| I-ROD2        | irod2_orf00015 | RecT protein                                       |                           | Escherichia coli O157:H7 str. EC508  |
| I-ROD2        | irod2_orf00016 | conserved hypothetical protein                     |                           | Escherichia coli O157:H7 str. EC4501 |



|        |                |                                      |                                       |
|--------|----------------|--------------------------------------|---------------------------------------|
| I-ROD2 | irod2_orf00017 | hypothetical protein ECH7EC4501_4934 | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00019 | conserved domain protein             | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00020 | conserved hypothetical protein       | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00021 | phage regulatory protein, Rha family | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00022 | type II restriction enzyme BsuBI     | Escherichia coli O157:H7 str. TW14588 |
| I-ROD2 | irod2_orf00023 | modification methylase BsuBI         | Escherichia coli O157:H7 str. TW14588 |
| I-ROD2 | irod2_orf00025 | repressor protein CI                 | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00026 | hypothetical protein SDY_1924        | Shigella dysenteriae Sd197            |
| I-ROD2 | irod2_orf00027 | helicase domain protein              | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00028 | hypothetical protein ECOK1180_4044   | Escherichia coli 1180                 |
| I-ROD2 | irod2_orf00029 | hypothetical protein ECO103_2848     | Escherichia coli O103:H2 str. 12009   |
| I-ROD2 | irod2_orf00030 | protein ninG                         | Escherichia coli 1180                 |
| I-ROD2 | irod2_orf00033 | DNA modification methylase           | Stx2-converting phage 86              |
| I-ROD2 | irod2_orf00034 | Shiga toxin 2 subunit A              | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00036 | Shiga toxin 2 subunit B              | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00037 | hypothetical protein                 | Shigella phage 7888                   |
| I-ROD2 | irod2_orf00038 | hypothetical protein ECs2970         | Escherichia coli O157:H7 str. Sakai   |
| I-ROD2 | irod2_orf00039 | conserved domain protein             | Escherichia coli O157:H7 str. EC4113  |
| I-ROD2 | irod2_orf00040 | lysozyme                             | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00041 | anti-repressor protein Ant           | Enterobacteria phage VT2phi_272       |
| I-ROD2 | irod2_orf00044 | endopeptidase (Protein gp15)         | Escherichia coli S88                  |
| I-ROD2 | irod2_orf00045 | Rha protein                          | Escherichia coli O157:H7 str. TW14588 |
| I-ROD2 | irod2_orf00048 | putative terminase small subunit     | Stx2-converting phage 86              |
| I-ROD2 | irod2_orf00049 | hypothetical protein ECOK1180_4067   | Escherichia coli 1180                 |
| I-ROD2 | irod2_orf00050 | large subunit terminase              | Escherichia coli O157:H7 str. EC4113  |
| I-ROD2 | irod2_orf00051 | putative phage portal protein        | Stx2-converting phage 86              |
| I-ROD2 | irod2_orf00052 | hypothetical protein 933Wp53         | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00053 | hypothetical protein 933Wp54         | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00054 | hypothetical protein 933Wp55         | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00055 | hypothetical protein 933Wp56         | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00056 | hypothetical protein ECs1226         | Escherichia coli O157:H7 str. Sakai   |
| I-ROD2 | irod2_orf00057 | hypothetical protein ECO103_2826     | Escherichia coli O103:H2 str. 12009   |

|        |                |  |                                       |
|--------|----------------|--|---------------------------------------|
| I-ROD2 | irod2_orf00059 | tail fiber protein                         | Escherichia coli O157:H7 str. EC4113  |
| I-ROD2 | irod2_orf00061 | hypothetical protein Stx2-86_gp25          | Stx2-converting phage 86              |
| I-ROD2 | irod2_orf00063 | conserved hypothetical protein             | Escherichia coli O157:H7 str. EC4196  |
| I-ROD2 | irod2_orf00065 | outer membrane protein Lom precursor       | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00066 | conserved hypothetical protein             | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00067 | hypothetical protein 933Wp68               | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00068 | conserved hypothetical protein             | Escherichia coli O157:H7 str. EC4501  |
| I-ROD2 | irod2_orf00069 | hypothetical protein 933Wp70               | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00071 | hypothetical protein 933Wp71               | Enterobacteria phage 933W             |
| I-ROD2 | irod2_orf00072 | hypothetical protein Stx2-86_gp35          | Stx2-converting phage 86              |
| I-ROD2 | irod2_orf00073 | hypothetical protein                       | Enterobacteria phage Min27            |
| I-ROD3 | irod3_orf00001 | predicted integrase                        | Escherichia sp. 4_1_40B               |
| I-ROD3 | irod3_orf00002 | unknown                                    | Shigella flexneri 2a                  |
| I-ROD3 | irod3_orf00003 | prophage CP4-57 regulatory protein alpA    | Escherichia coli 3431                 |
| I-ROD3 | irod3_orf00004 | unknown                                    | Shigella flexneri 2a                  |
| I-ROD3 | irod3_orf00005 | type III restriction enzyme, res subunit   | Pelobacter propionicus DSM 2379       |
| I-ROD3 | irod3_orf00006 | hypothetical protein E4_08923              | Escherichia sp. 4_1_40B               |
| I-ROD3 | irod3_orf00007 | Transposase                                | Shigella dysenteriae CDC 74-1112      |
| I-ROD3 | irod3_orf00008 | IS66 family element, orf2                  | Shigella boydii CDC 3083-94           |
| I-ROD3 | irod3_orf00011 | hypothetical protein                       | Escherichia coli                      |
| I-ROD3 | irod3_orf00012 | hypothetical protein HMPREF9552_03072      | Escherichia coli MS 198-1             |
| I-ROD3 | irod3_orf00013 | hypothetical protein E4_08823              | Escherichia sp. 4_1_40B               |
| I-ROD3 | irod3_orf00014 | hypothetical protein Z1185                 | Escherichia coli O157:H7 EDL933       |
| I-ROD3 | irod3_orf00015 | conserved hypothetical protein             | Escherichia coli O157:H7 str. TW14588 |
| I-ROD3 | irod3_orf00016 | hypothetical protein Z1188                 | Escherichia coli O157:H7 EDL933       |
| I-ROD3 | irod3_orf00017 | putative glucosyltransferase               | Escherichia coli O157:H7 EDL933       |
| I-ROD3 | irod3_orf00018 | putative ferric enterochelin esterase MchK | Escherichia coli                      |
| I-ROD3 | irod3_orf00019 | MchS2 protein                              | Escherichia coli                      |
| I-ROD3 | irod3_orf00020 | hypothetical protein p1ECUMN_0112          | Escherichia coli UMN026               |
| I-ROD3 | irod3_orf00022 | MchC protein                               | Escherichia coli CFT073               |
| I-ROD3 | irod3_orf00023 | microcin H47 secretion protein             | Escherichia coli 042                  |
| I-ROD3 | irod3_orf00024 | MtfB                                       | Escherichia coli                      |

|        |                |  |       |  |
|--------|----------------|--|-------|--|
| I-ROD3 | irod3_orf00025 | conserved hypothetical protein   |       | Escherichia coli O157:H7 str. EC4196             |
| I-ROD3 | irod3_orf00026 | hypothetical protein ECDG_03856  |       | Escherichia coli B185                            |
| I-ROD3 | irod3_orf00027 | hypothetical protein ROD_49891   |       | Citrobacter rodentium ICC168                     |
| I-ROD3 | irod3_orf00028 | hypothetical protein ROD_49911   |       | Citrobacter rodentium ICC168                     |
| I-ROD3 | irod3_orf00029 | ImpA-related N- superfamily  |       | Escherichia coli M605                            |
| I-ROD3 | irod3_orf00030 | hypothetical protein AHA_1063  |       | Aeromonas hydrophila subsp. hydrophila ATCC 7966 |
| I-ROD3 | irod3_orf00031 | immunoglobulin-binding regulator A   |       | Escherichia coli M605                            |
| I-ROD3 | irod3_orf00032 | insertion element IS1 7 protein insA   |       | Shigella dysenteriae 1617                        |
| I-ROD3 | irod3_orf00034 | hypothetical protein ECNA114_2538  |       | Escherichia coli NA114                           |
| I-ROD3 | irod3_orf00035 | putative transposase   |       | Shigella flexneri K-671                          |
| I-ROD3 | irod3_orf00036 | putative ATP synthase F0, A subunit  |       | Escherichia coli MS 116-1                        |
| I-ROD3 | irod3_orf00037 | aspartate racemase   |       | Shigella flexneri K-272                          |
| I-ROD3 | irod3_orf00038 | hypothetical protein HMPREF9541_00362  |       | Escherichia coli MS 116-1                        |
| I-ROD3 | irod3_orf00039 | putative transcriptional regulator   |       | Shigella flexneri 2a                             |
| I-ROD3 | irod3_orf00042 | conserved domain protein   |       | Escherichia coli MS 116-1                        |
| I-ROD3 | irod3_orf00043 | predicted protein  |       | Nematostella vectensis                           |
| I-ROD3 | irod3_orf00044 | protein kinase   |       | Yersinia pseudotuberculosis IP 31758             |
| I-ROD3 | irod3_orf00045 | hypothetical protein ESA_01782   |       | Cronobacter sakazakii ATCC BAA-894               |
| I-ROD3 | irod3_orf00046 | putative tellurium resistance protein  | TerY3 | Serratia marcescens                              |
| I-ROD3 | irod3_orf00047 | putative tellurium resistance protein  | TerY2 | Serratia marcescens                              |
| I-ROD3 | irod3_orf00049 | tellurium resistance protein   | TerX  | Serratia marcescens                              |
| I-ROD3 | irod3_orf00050 | putative tellurium resistance protein TerY                                     | TerY1 | Enterobacter cloacae subsp. cloacae ATCC 13047   |
| I-ROD3 | irod3_orf00051 | terW   | TerW  | Citrobacter sp. 30_2                             |
| I-ROD3 | irod3_orf00052 | hypothetical protein SMR0069   |       | Serratia marcescens                              |
| I-ROD3 | irod3_orf00053 | hypothetical protein Z1166   |       | Escherichia coli O157:H7 EDL933                  |
| I-ROD3 | irod3_orf00054 | putative ATP-binding protein   |       | Escherichia coli APEC O1                         |
| I-ROD3 | irod3_orf00055 | hypothetical protein APECO1_O1R65  |       | Escherichia coli APEC O1                         |
| I-ROD3 | irod3_orf00056 | hypothetical protein APECO1_O1R66  |       | Escherichia coli APEC O1                         |
| I-ROD3 | irod3_orf00057 | hypothetical protein APECO1_O1R67  |       | Escherichia coli APEC O1                         |
| I-ROD3 | irod3_orf00058 | putative phage inhibition, colicin resistance and tellurite resistance protein | TerZ  | Escherichia coli O157:H7 EDL933                  |

|        |                |  |      |  |
|--------|----------------|--|------|--|
| I-ROD3 | irod3_orf00059 | putative phage inhibition, colicin resistance and tellurite resistance protein | TerA | Escherichia coli O157:H7 EDL933        |
| I-ROD3 | irod3_orf00060 | putative phage inhibition, colicin resistance and tellurite resistance protein | TerB | Escherichia coli O157:H7 EDL933        |
| I-ROD3 | irod3_orf00061 | putative phage inhibition, colicin resistance and tellurite resistance protein | TerC | Escherichia coli O157:H7 EDL933        |
| I-ROD3 | irod3_orf00063 | putative phage inhibition, colicin resistance and tellurite resistance protein | TerD | Escherichia coli O157:H7 EDL933        |
| I-ROD3 | irod3_orf00064 | putative phage inhibition, colicin resistance and tellurite resistance protein | TerE | Escherichia coli O157:H7 EDL933        |
| I-ROD3 | irod3_orf00065 | putative tellurium resistance protein TerF                                     | TerF | Escherichia coli O103:H2 str. 12009    |
| I-ROD3 | irod3_orf00067 | putative GTP-binding protein   |      | Escherichia coli 042                   |
| I-ROD3 | irod3_orf00068 | antigen 43 precursor   |      | Escherichia coli                       |
| I-ROD3 | irod3_orf00071 | putative autotransporter   |      | Shigella sp. D9                        |
| I-ROD3 | irod3_orf00072 | hypothetical protein EscherichiacoliO157_22726                                 |      | Escherichia coli O157:H7 str. FRIK2000 |
| I-ROD3 | irod3_orf00073 | hypothetical protein ECS88_2092  |      | Escherichia coli S88                   |
| I-ROD3 | irod3_orf00074 | hypothetical protein ECS88_2092  |      | Escherichia coli S88                   |
| I-ROD3 | irod3_orf00075 | conserved hypothetical protein   |      | Escherichia coli H591                  |
| I-ROD3 | irod3_orf00077 | hypothetical protein SD1617_3951   |      | Shigella dysenteriae 1617              |
| I-ROD3 | irod3_orf00078 | hypothetical protein ECS88_2094  |      | Escherichia coli S88                   |
| I-ROD3 | irod3_orf00079 | hypothetical protein APECO1_1098   |      | Escherichia coli APEC O1               |
| I-ROD3 | irod3_orf00080 | hypothetical protein ECO103_3758   |      | Escherichia coli O103:H2 str. 12009    |
| I-ROD3 | irod3_orf00081 | hypothetical protein ECNA114_2131  |      | Escherichia coli NA114                 |
| I-ROD3 | irod3_orf00083 | toxin of the YeeV-YeeU toxin-antitoxin system                                  |      | Escherichia sp. 4_1_40B                |
| I-ROD3 | irod3_orf00084 | conserved hypothetical protein   |      | Escherichia coli ETEC H10407           |
| I-ROD3 | irod3_orf00086 | hypothetical protein UTI89_C4999   |      | Escherichia coli UTI89                 |
| I-ROD3 | irod3_orf00087 | hypothetical protein Z1226   |      | Escherichia coli O157:H7 EDL933        |
| I-ROD4 | irod4_orf00001 | AntB   |      | Escherichia coli                       |
| I-ROD4 | irod4_orf00003 | conserved hypothetical protein   |      | Escherichia coli FVEC1302              |
| I-ROD4 | irod4_orf00004 | valyl-tRNA synthetase  |      | Escherichia coli E110019               |
| I-ROD4 | irod4_orf00006 | hypothetical protein Stx2-86_gp35  |      | Stx2-converting phage 86               |
| I-ROD4 | irod4_orf00008 | hypothetical protein SDY_1670  |      | Shigella dysenteriae Sd197             |
| I-ROD4 | irod4_orf00010 | hypothetical protein ECED1_1152  |      | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00011 | hypothetical protein ECED1_1151  |      | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00012 | hypothetical protein 933Wp68   |      | Enterobacteria phage 933W              |
| I-ROD4 | irod4_orf00013 | hypothetical protein Stx2-86_gp30  |      | Stx2-converting phage 86               |
| I-ROD4 | irod4_orf00014 | putative outer membrane precursor Lom  |      | Escherichia coli O103:H2 str. 12009    |

|        |                |  |  |
|--------|----------------|--|--|
| I-ROD4 | irod4_orf00016 | hypothetical protein Stx2Ip034                   | Stx2 converting phage I                |
| I-ROD4 | irod4_orf00018 | hypothetical protein Stx2-86_gp25                | Stx2-converting phage 86               |
| I-ROD4 | irod4_orf00020 | putative long tail fiber protein                 | Stx2-converting phage 86               |
| I-ROD4 | irod4_orf00021 | hypothetical protein ECED1_1137                  | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00022 | hypothetical protein ECED1_1136                  | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00023 | hypothetical protein ECED1_1135                  | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00024 | hypothetical protein Stx2-86_gp17                | Stx2-converting phage 86               |
| I-ROD4 | irod4_orf00025 | hypothetical protein ECED1_1133                  | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00026 | hypothetical protein ECED1_1132                  | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00027 | putative phage portal protein                    | Stx2-converting phage 86               |
| I-ROD4 | irod4_orf00028 | hypothetical protein ECOK1180_4067               | Escherichia coli 1180                  |
| I-ROD4 | irod4_orf00029 | large subunit terminase                          | Escherichia coli O157:H7 str. EC4113   |
| I-ROD4 | irod4_orf00030 | putative terminase small subunit                 | Stx2-converting phage 86               |
| I-ROD4 | irod4_orf00033 | bacteriophage lysis protein                      | Shigella dysenteriae 1012              |
| I-ROD4 | irod4_orf00036 | putative endolysin                               | Shigella dysenteriae Sd197             |
| I-ROD4 | irod4_orf00037 | protein S  | Enterobacteria phage 933W              |
| I-ROD4 | irod4_orf00038 | conserved hypothetical protein                   | Shigella dysenteriae 1617              |
| I-ROD4 | irod4_orf00039 | hypothetical protein SGF_04061                   | Shigella flexneri CDC 796-83           |
| I-ROD4 | irod4_orf00040 | YjhS   | Shigella boydii CDC 3083-94            |
| I-ROD4 | irod4_orf00041 | putative NinH protein                            | Phage BP-4795                          |
| I-ROD4 | irod4_orf00042 | crossover junction endodeoxyribonuclease         | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00044 | hypothetical protein E2348C_2522                 | Escherichia coli O127:H6 str. E2348/69 |
| I-ROD4 | irod4_orf00045 | putative ninB protein                            | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00046 | putative antirepressor protein Ant from prophage | Escherichia coli ED1a                  |
| I-ROD4 | irod4_orf00047 | hypothetical protein ECO26_2262                  | Escherichia coli O26:H11 str. 11368    |
| I-ROD4 | irod4_orf00048 | death-on-curing family protein                   | Escherichia coli STEC_7v               |
| I-ROD4 | irod4_orf00049 | hypothetical protein ECSTEC7V_1837               | Escherichia coli STEC_7v               |
| I-ROD4 | irod4_orf00050 | hypothetical protein ECO111_1061                 | Escherichia coli O111:H- str. 11128    |
| I-ROD4 | irod4_orf00051 | hypothetical protein G2583_1712                  | Escherichia coli O55:H7 str. CB9615    |
| I-ROD4 | irod4_orf00052 | hypothetical protein EcE24377A_1426              | Escherichia coli E24377A               |
| I-ROD4 | irod4_orf00053 | hypothetical protein ECO103_1369                 | Escherichia coli O103:H2 str. 12009    |
| I-ROD4 | irod4_orf00055 | putative replication protein                     | Escherichia coli ED1a                  |

|        |                |  |                                     |
|--------|----------------|--|-------------------------------------|
| I-ROD4 | irod4_orf00056 | hypothetical protein ECED1_1103                                | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00057 | hypothetical protein ECED1_1102                                | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00058 | regulatory protein CII from prophage                           | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00059 | prophage repressor CI  | Enterobacteria phage HK97           |
| I-ROD4 | irod4_orf00060 | hypothetical protein ECED1_1098                                | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00061 | hypothetical protein ECED1_1097                                | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00063 | monocarboxylate transporter                                    | Culex quinquefasciatus              |
| I-ROD4 | irod4_orf00064 | hypothetical protein ECED1_1095                                | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00065 | hypothetical protein ECED1_1094                                | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00067 | FtsZ inhibitor protein   | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00068 | hypothetical protein ECED1_1091                                | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00069 | Exodeoxyribonuclease VIII (putative partial) from phage origin | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00070 | putative host-nuclease inhibitor protein Gam                   | Shigella dysenteriae Sd197          |
| I-ROD4 | irod4_orf00071 | Recombination protein bet from phage origin                    | Escherichia coli ED1a               |
| I-ROD4 | irod4_orf00072 | putative exonuclease encoded by prophage CP-933K               | Escherichia coli O157:H7 EDL933     |
| I-ROD4 | irod4_orf00074 | prophage DLP12 integrase                                       | Escherichia coli 101-1              |
| I-ROD5 | irod5_orf00001 | hypothetical protein SSON_1273                                 | Shigella sonnei Ss046               |
| I-ROD5 | irod5_orf00002 | hypothetical protein EC55989_1079                              | Escherichia coli 55989              |
| I-ROD5 | irod5_orf00005 | triple helix repeat-containing collagen                        | Clostridium beijerinckii NCIMB 8052 |
| I-ROD5 | irod5_orf00006 | hypothetical protein SD15574_2985                              | Shigella dysenteriae 155-74         |
| I-ROD5 | irod5_orf00008 | hypothetical protein ECE128010_5420                            | Escherichia coli E128010            |
| I-ROD5 | irod5_orf00011 | Putative tail component of prophage                            | Escherichia coli NA114              |
| I-ROD5 | irod5_orf00012 | hypothetical protein ECLG_05105                                | Escherichia coli TA271              |
| I-ROD5 | irod5_orf00015 | Superoxide dismutase (Cu-Zn)                                   | Escherichia coli O55:H7 str. CB9615 |
| I-ROD5 | irod5_orf00019 | minor tail protein   | Escherichia coli UTI89              |
| I-ROD5 | irod5_orf00021 | minor tail protein   | Escherichia coli UTI89              |
| I-ROD5 | irod5_orf00022 | putative tail fiber component H of prophage CP-933U            | Escherichia coli O157:H7 EDL933     |
| I-ROD5 | irod5_orf00024 | Phage minor tail protein                                       | Escherichia coli EC4100B            |
| I-ROD5 | irod5_orf00025 | Phage minor tail protein                                       | Escherichia coli EC4100B            |
| I-ROD5 | irod5_orf00026 | phage major tail protein                                       | Escherichia coli 042                |
| I-ROD5 | irod5_orf00027 | hypothetical protein DAPPUDRAFT_279812                         | Daphnia pulex                       |
| I-ROD5 | irod5_orf00029 | polysaccharide Transporter, PST family                         | Enterococcus faecium E1679          |

|        |                |   |  |
|--------|----------------|---|--|
| I-ROD5 | irod5_orf00032 | hypothetical protein                                | Arthrospira platensis NIES-39  |
| I-ROD5 | irod5_orf00034 | terminase large subunit domain protein              | Escherichia coli RN587/1   |
| I-ROD5 | irod5_orf00035 | conserved hypothetical protein                      | Escherichia albertii TW07627   |
| I-ROD5 | irod5_orf00036 | phage major capsid protein E                        | Escherichia coli H489  |
| I-ROD5 | irod5_orf00039 | Hypothetical protein CBG02325                       | Caenorhabditis briggsae  |
| I-ROD5 | irod5_orf00041 | conserved domain protein                            | Escherichia coli MS 153-1  |
| I-ROD5 | irod5_orf00042 | hypothetical protein c1457                          | Escherichia coli CFT073  |
| I-ROD5 | irod5_orf00043 | Phage minor tail protein                            | Escherichia coli EC4100B   |
| I-ROD5 | irod5_orf00045 | hypothetical protein MK0973                         | Methanopyrus kandleri AV19   |
| I-ROD5 | irod5_orf00047 | hypothetical protein SCA50_1305                     | Salmonella enterica subsp. enterica serovar Choleraesuis str. SCSA50 |
| I-ROD5 | irod5_orf00049 | hypothetical protein ECOK1_1278                     | Escherichia coli IHE3034   |
| I-ROD5 | irod5_orf00052 | phage DNA packaging protein Nu1                     | Escherichia coli MS 21-1   |
| I-ROD5 | irod5_orf00053 | putative phage protein                              | Escherichia coli 042   |
| I-ROD5 | irod5_orf00055 | hypothetical protein ECS88_0566                     | Escherichia coli S88   |
| I-ROD5 | irod5_orf00056 | endopeptidase                                       | Escherichia coli 2362-75   |
| I-ROD5 | irod5_orf00058 | hypothetical protein SBO_1923                       | Shigella boydii Sb227  |
| I-ROD5 | irod5_orf00059 | putative membrane-associated lysozyme; Qin prophage | Escherichia coli 55989   |
| I-ROD5 | irod5_orf00061 | hypothetical protein Stx2-86_gp06                   | Stx2-converting phage 86   |
| I-ROD5 | irod5_orf00062 | hypothetical protein Stx2-86_gp05                   | Stx2-converting phage 86   |
| I-ROD5 | irod5_orf00063 | lysis protein S                                     | Stx2-converting phage 86   |
| I-ROD5 | irod5_orf00066 | hypothetical protein DAPPUDRAFT_52038               | Daphnia pulex  |
| I-ROD5 | irod5_orf00068 | heterokaryon incompatibility protein                | Glomerella graminicola M1.001  |
| I-ROD5 | irod5_orf00070 | DNA methylase family protein                        | Shigella flexneri J1713  |
| I-ROD5 | irod5_orf00071 | hypothetical protein HMPREF9542_00842               | Escherichia coli MS 117-3  |
| I-ROD5 | irod5_orf00072 | hypothetical protein EcF11_4284                     | Escherichia coli F11   |
| I-ROD5 | irod5_orf00075 | hypothetical protein ECRN5871_4170                  | Escherichia coli RN587/1   |
| I-ROD5 | irod5_orf00076 | hypothetical protein E4_10746                       | Escherichia sp. 4_1_40B  |
| I-ROD5 | irod5_orf00077 | endodeoxyribonuclease RusA family protein           | Escherichia coli STEC_7v   |
| I-ROD5 | irod5_orf00078 | LexA repressor                                      | Escherichia coli S88   |
| I-ROD5 | irod5_orf00079 | DNA adenine methylase                               | Escherichia coli UTI89   |
| I-ROD5 | irod5_orf00080 | hypothetical protein ECS88_0547                     | Escherichia coli S88   |

|        |                |  |   |
|--------|----------------|--|---|
| I-ROD5 | irod5_orf00081 | hypothetical protein PcdtI_gp46            | Phage cdtI  |
| I-ROD5 | irod5_orf00082 | putative antirepressor                     | Escherichia coli EC4100B                                |
| I-ROD5 | irod5_orf00083 | nucleic acid-binding protein; e14 prophage | Escherichia coli S88                                    |
| I-ROD5 | irod5_orf00084 | hypothetical protein ECD227_2469           | Escherichia fergusonii ECD227                           |
| I-ROD5 | irod5_orf00085 | regulatory protein cI                      | Escherichia coli EC4100B                                |
| I-ROD5 | irod5_orf00086 | hypothetical protein ECoL_03975            | Escherichia coli EC4100B                                |
| I-ROD5 | irod5_orf00087 | hypothetical protein ECoL_03976            | Escherichia coli EC4100B                                |
| I-ROD5 | irod5_orf00089 | Hypothetical protein yfdR                  | Escherichia coli EC4100B                                |
| I-ROD5 | irod5_orf00090 | hypothetical protein ShiD9_12075           | Shigella sp. D9   |
| I-ROD5 | irod5_orf00091 | conserved hypothetical protein             | Escherichia coli E22                                    |
| I-ROD5 | irod5_orf00093 | conserved hypothetical protein             | Escherichia coli E22                                    |
| I-ROD5 | irod5_orf00095 | Phage EaA protein                          | Escherichia coli EC4100B                                |
| I-ROD5 | irod5_orf00096 | Phage EaA protein                          | Escherichia coli EC4100B                                |
| I-ROD5 | irod5_orf00097 | Integrase                                  | Escherichia coli EC4100B                                |
| I-ROD6 | irod6_orf00001 | molybdate metabolism regulator             | Escherichia coli 536                                    |
| I-ROD6 | irod6_orf00003 | hypothetical protein ECP_2154              | Escherichia coli 536                                    |
| I-ROD6 | irod6_orf00005 | yehL protein                               | Escherichia coli B088                                   |
| I-ROD6 | irod6_orf00006 | hypothetical protein ECP_2157              | Escherichia coli 536                                    |
| I-ROD6 | irod6_orf00007 | hypothetical protein ECIAI1_2197           | Escherichia coli IAI1                                   |
| I-ROD6 | irod6_orf00008 | hypothetical protein ECP_2159              | Escherichia coli 536                                    |
| I-ROD7 | irod7_orf00001 | integrase                                  | Escherichia coli  |
| I-ROD7 | irod7_orf00002 | Evolved beta-D-galactosidase, beta subunit | Shigella dysenteriae CDC 74-1112                        |
| I-ROD7 | irod7_orf00003 | transposase TnpA                           | Corynebacterium glutamicum                              |
| I-ROD7 | irod7_orf00004 | resolvase for Tn21                         | Plasmid R100  |
| I-ROD7 | irod7_orf00006 | Urf2 protein                               | Escherichia fergusonii ECD227                           |
| I-ROD7 | irod7_orf00007 | integrase                                  | Plasmid R100  |
| I-ROD7 | irod7_orf00008 | dihydrofolate reductase type A7            | Salmonella enterica subsp. enterica serovar Weltevreden |
| I-ROD7 | irod7_orf00010 | putative transposase                       | Klebsiella pneumoniae subsp. pneumoniae MGH 78578       |
| I-ROD7 | irod7_orf00014 | 3-hydroxyisobutyrate dehydrogenase         | Mycobacterium tuberculosis 210                          |
| I-ROD7 | irod7_orf00016 | protein RepC                               | Salmonella enterica subsp. enterica serovar Enteritidis |



|        |                |  |   |
|--------|----------------|--|---|
| I-ROD7 | irod7_orf00017 | dihydropteroate synthase   | Salmonella enterica subsp. enterica serovar Typhi str. CT18 |
| I-ROD7 | irod7_orf00018 | aminoglycoside/hydroxyurea antibiotic resistance kinase          | Escherichia coli MS 200-1                                   |
| I-ROD7 | irod7_orf00019 | beta-lactamase   | Escherichia coli 3431                                       |
| I-ROD7 | irod7_orf00021 | hypothetical protein R100p008                                    | Plasmid R100  |
| I-ROD7 | irod7_orf00022 | putative mercury resistance protein                              | Plasmid R100  |
| I-ROD7 | irod7_orf00023 | transcriptional regulator MerD                                   | Plasmid R100  |
| I-ROD7 | irod7_orf00026 | RecName: Full=Mercuric reductase; AltName: Full=Hg(II) reductase |   |
| I-ROD7 | irod7_orf00027 | putative mercury transport protein MerC                          | Aeromonas salmonicida subsp. salmonicida A449               |
| I-ROD7 | irod7_orf00029 | Tn501 orf, hypotheical   | Shigella flexneri 5a  |
| I-ROD7 | irod7_orf00033 | InsL   | Escherichia coli 53638                                      |
| I-ROD7 | irod7_orf00034 | hypothetical protein pFL129_4                                    | Escherichia coli  |
| I-ROD7 | irod7_orf00036 | TetA   | Salmonella enterica subsp. enterica serovar Choleraesuis    |
| I-ROD7 | irod7_orf00037 | integral membrane protein DUF6                                   | Escherichia coli MS 78-1                                    |
| I-ROD7 | irod7_orf00038 | hypothetical protein HMPREF9544_05491                            | Escherichia coli MS 153-1                                   |
| I-ROD7 | irod7_orf00039 | conserved hypothetical protein                                   | Escherichia coli ETEC H10407                                |
| I-ROD7 | irod7_orf00040 | conserved hypothetical protein                                   | Escherichia coli SE15                                       |
| I-ROD7 | irod7_orf00042 | hypothetical protein HMPREF9553_03865                            | Escherichia coli MS 200-1                                   |
| I-ROD7 | irod7_orf00044 | putative regulatory protein                                      | Escherichia coli 536  |
| I-ROD7 | irod7_orf00045 | conserved hypothetical protein                                   | Escherichia coli SE15                                       |
| I-ROD7 | irod7_orf00046 | transposase  | Escherichia coli SE15                                       |
| I-ROD7 | irod7_orf00048 | hypothetical protein   | Escherichia coli SE15                                       |
| I-ROD7 | irod7_orf00049 | hypothetical protein ECUMN_4880                                  | Escherichia coli UMN026                                     |
| I-ROD7 | irod7_orf00051 | putative autotransporter   | Escherichia coli 536  |
| I-ROD7 | irod7_orf00052 | antigen 43 domain protein  | Escherichia coli LT-68                                      |
| I-ROD7 | irod7_orf00053 | hypothetical protein EcE24377A_4893                              | Escherichia coli E24377A                                    |
| I-ROD7 | irod7_orf00054 | hypothetical protein ECNA114_2131                                | Escherichia coli NA114                                      |
| I-ROD7 | irod7_orf00056 | conserved domain protein   | Escherichia coli MS 187-1                                   |
| I-ROD7 | irod7_orf00058 | conserved hypothetical protein                                   | Escherichia coli SE15                                       |
| I-ROD7 | irod7_orf00059 | putative radC-like protein yeeS                                  | Escherichia coli CFT073                                     |
| I-ROD7 | irod7_orf00060 | hypothetical protein c0272                                       | Escherichia coli CFT073                                     |

|        |                    |                                       |  |
|--------|--------------------|---------------------------------------|--|
| I-ROD7 | irod7_orf00061     | unknown                               | Escherichia coli   |
| I-ROD7 | irod7_orf00063     | DNA repair protein                    | Escherichia coli MS 78-1   |
| I-ROD7 | irod7_orf00064     | hypothetical protein c4574            | Escherichia coli CFT073  |
| I-ROD7 | irod7_orf00065     | conserved hypothetical protein        | Shigella dysenteriae 1617  |
| I-ROD7 | irod7_orf00067     | hypothetical protein APECO1_3486      | Escherichia coli APEC O1   |
| I-ROD7 | irod7_orf00068     | hypothetical protein SF3000           | Shigella flexneri 2a str. 301  |
| I-ROD7 | irod7_orf00069     | hypothetical protein ECO103_3594      | Escherichia coli O103:H2 str. 12009                                  |
| I-ROD7 | irod7_orf00070     | hypothetical protein ECED1_4984       | Escherichia coli ED1a  |
| pESBL  | scaffold19_orf0002 | YciB                                  | Escherichia coli   |
| pESBL  | scaffold19_orf0003 | hypothetical protein pECBactecp21     | Escherichia coli   |
| pESBL  | scaffold19_orf0004 | hypothetical protein SC121            | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |
| pESBL  | scaffold19_orf0005 | single-stranded DNA-binding protein   | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |
| pESBL  | scaffold19_orf0006 | hypothetical protein pO157p50         | Escherichia coli O157:H7 str. Sakai                                  |
| pESBL  | scaffold19_orf0007 | plasmid SOS inhibition protein B      | Escherichia coli   |
| pESBL  | scaffold19_orf0008 | plasmid SOS inhibition protein A      | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |
| pESBL  | scaffold19_orf0009 | hypothetical protein SC115            | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |
| pESBL  | scaffold19_orf0010 | antirestriction protein               | Escherichia coli MS 107-1  |
| pESBL  | scaffold19_orf0011 | hypothetical protein ECSE_P1-0063     | Escherichia coli SE11  |
| pESBL  | scaffold19_orf0012 | hypothetical protein HMPREF9542_03988 | Escherichia coli MS 117-3  |
| pESBL  | scaffold19_orf0013 | hypothetical protein SeHA_A0062       | Salmonella enterica subsp. enterica serovar Heidelberg str. SL476    |
| pESBL  | scaffold19_orf0014 | hypothetical protein EcE22_3665       | Escherichia coli E22   |
| pESBL  | scaffold19_orf0015 | CcgAII protein                        | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |
| pESBL  | scaffold19_orf0016 | putative transposase                  | Escherichia coli E22   |
| pESBL  | scaffold19_orf0018 | hypothetical protein SC107            | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |
| pESBL  | scaffold19_orf0019 | hypothetical protein R64_p076         | Salmonella enterica subsp. enterica serovar Typhimurium              |
| pESBL  | scaffold19_orf0020 | hypothetical protein SC105            | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |

|       |                    |                                       |  |
|-------|--------------------|---------------------------------------|--|
| pESBL | scaffold19_orf0021 | hypothetical protein pECBactecp14     | Escherichia coli   |
| pESBL | scaffold19_orf0022 | hypothetical protein LH0067           | Escherichia coli   |
| pESBL | scaffold19_orf0023 | relaxosome component                  | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0024 | NikB                                  | Escherichia coli O157:H7 str. Sakai                                  |
| pESBL | scaffold19_orf0025 | hypothetical protein EcE24377A_D0057  | Escherichia coli E24377A   |
| pESBL | scaffold19_orf0026 | hypothetical protein pECBactecp09     | Escherichia coli   |
| pESBL | scaffold19_orf0027 | hypothetical protein pECBactecp08     | Escherichia coli   |
| pESBL | scaffold19_orf0028 | putative protein FinQ                 | Escherichia coli MS 84-1   |
| pESBL | scaffold19_orf0029 | counter protein for PndA              | Escherichia coli   |
| pESBL | scaffold19_orf0030 | hypothetical protein SC084            | Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 |
| pESBL | scaffold19_orf0031 | conserved hypothetical protein        | Escherichia coli MS 107-1  |
| pESBL | scaffold19_orf0032 | hypothetical protein ECSE_P1-0081     | Escherichia coli SE11  |
| pESBL | scaffold19_orf0033 | putative regulator protein            | Escherichia coli SE11  |
| pESBL | scaffold19_orf0034 | exclusion-determining family protein  | Escherichia coli MS 84-1   |
| pESBL | scaffold19_orf0035 | TraY                                  | Escherichia coli O157:H7 str. EC4486                                 |
| pESBL | scaffold19_orf0036 | F pilin acetylation protein           | Escherichia coli   |
| pESBL | scaffold19_orf0037 | F pilus assembly                      | Escherichia coli   |
| pESBL | scaffold19_orf0038 | F pilus assembly                      | Escherichia coli   |
| pESBL | scaffold19_orf0039 | TraU                                  | Escherichia coli O157:H7 str. EC4401                                 |
| pESBL | scaffold19_orf0040 | hypothetical protein HMPREF9542_01329 | Escherichia coli MS 117-3  |
| pESBL | scaffold19_orf0041 | TraR protein                          | Escherichia coli   |
| pESBL | scaffold19_orf0042 | hypothetical protein Collb-P9_p070    | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0043 | hypothetical protein Collb-P9_p071    | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0044 | hypothetical protein Collb-P9_p072    | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0045 | hypothetical protein Collb-P9_p073    | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0046 | hypothetical protein Collb-P9_p074    | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0047 | thick pilus signal peptide            | Escherichia coli W   |
| pESBL | scaffold19_orf0048 | DNA primase                           | Escherichia coli O157:H7 str. EC4401                                 |
| pESBL | scaffold19_orf0049 | EDTA-resistant nuclease               | Escherichia coli   |
| pESBL | scaffold19_orf0051 | ATP-binding protein                   | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0052 | lipoprotein                           | Salmonella enterica subsp. enterica serovar                          |

|       |                    |                                       |  |  |
|-------|--------------------|---------------------------------------|--|--|
|       |                    |                                       |  | Typhimurium  |
| pESBL | scaffold19_orf0053 | hypothetical protein Collb-P9_p082    |  | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0054 | hypothetical protein Collb-P9_p083    |  | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0055 | F pilus assembly                      |  | Escherichia coli   |
| pESBL | scaffold19_orf0056 | TraE protein                          |  | Escherichia coli   |
| pESBL | scaffold19_orf0057 | shufflon-specific DNA recombinase     |  | Escherichia coli AA86  |
| pESBL | scaffold19_orf0058 | hypothetical protein HMPREF9536_01879 |  | Escherichia coli MS 84-1   |
| pESBL | scaffold19_orf0059 | conserved hypothetical protein        |  | Escherichia coli MS 107-1  |
| pESBL | scaffold19_orf0060 | hypothetical protein R64_p118         |  | Salmonella enterica subsp. enterica serovar Typhimurium            |
| pESBL | scaffold19_orf0061 | shufflon protein C'                   |  | Escherichia coli O157:H7 str. EC4486                               |
| pESBL | scaffold19_orf0062 | conserved hypothetical protein        |  | Escherichia coli MS 107-1  |
| pESBL | scaffold19_orf0063 | shufflon protein A                    |  | Salmonella enterica subsp. enterica serovar Kentucky str. CVM29188 |
| pESBL | scaffold19_orf0064 | peptidase A24A prepilin type IV       |  | Escherichia coli W   |
| pESBL | scaffold19_orf0065 | type IV prepilin cluster              |  | Escherichia coli   |
| pESBL | scaffold19_orf0066 | type IV prepilin cluster; prepilin    |  | Escherichia coli   |
| pESBL | scaffold19_orf0067 | integral membrane protein             |  | Escherichia coli E22   |
| pESBL | scaffold19_orf0068 | ATP-binding protein PilQ              |  | Escherichia coli SE11  |
| pESBL | scaffold19_orf0069 | IncII conjugal transfer protein PilP  |  | Escherichia coli   |
| pESBL | scaffold19_orf0070 | IncII conjugal transfer protein PilO  |  | Escherichia coli   |
| pESBL | scaffold19_orf0071 | lipoprotein PilN                      |  | Escherichia coli SE11  |
| pESBL | scaffold19_orf0072 | hypothetical protein Collb-P9_p101    |  | Plasmid Collb-P9   |
| pESBL | scaffold19_orf0073 | IncII conjugal transfer protein PilL  |  | Escherichia coli   |
| pESBL | scaffold19_orf0074 | predicted protein                     |  | Nematostella vectensis   |
| pESBL | scaffold19_orf0075 | IncII conjugal transfer protein TraC  |  | Escherichia coli   |
| pESBL | scaffold19_orf0076 | transcription termination factor NusG |  | Escherichia coli MS 84-1   |
| pESBL | scaffold19_orf0077 | TraA protein                          |  | Escherichia coli SE11  |
| pESBL | scaffold19_orf0078 | replication initiation protein        |  | Salmonella enterica subsp. enterica serovar Kentucky str. CVM29188 |
| pESBL | scaffold19_orf0079 | hypothetical protein ND12IncII_3      |  | Escherichia coli   |
| pESBL | scaffold19_orf0080 | hypothetical protein pECBactecp34     |  | Escherichia coli   |
| pESBL | scaffold19_orf0081 | YagA                                  |  | Escherichia coli O157:H7 str. EC4486                               |

|       |                    |  |             |  |
|-------|--------------------|--|-------------|--|
| pESBL | scaffold19_orf0082 | transposase  |             | Salmonella enterica subsp. enterica serovar Infantis   |
| pESBL | scaffold19_orf0083 | conserved hypothetical protein                         |             | Escherichia coli MS 21-1                               |
| pESBL | scaffold19_orf0084 | hypothetical protein                                   |             | Escherichia coli                                       |
| pESBL | scaffold19_orf0085 | hypothetical protein pC15-1a_016                       | blaCTX-M-15 | Escherichia coli                                       |
| pESBL | scaffold19_orf0086 | ISEcp1 transposase                                     |             | Escherichia coli                                       |
| pESBL | scaffold19_orf0087 | transposase for transposon Tn3                         |             | Escherichia coli                                       |
| pESBL | scaffold19_orf0088 | hypothetical protein pC15-1a_019                       |             | Escherichia coli                                       |
| pESBL | scaffold19_orf0089 | TEM-1 beta-lactamase                                   | blaTEM-1    | Salmonella enterica subsp. enterica serovar Montevideo |
| pESBL | scaffold19_orf0090 | conserved domain protein                               |             | Escherichia coli MS 21-1                               |
| pESBL | scaffold19_orf0091 | cobyrinic acid a,c-diamide synthase                    |             | Escherichia coli                                       |
| pESBL | scaffold19_orf0093 | protein impB domain protein                            |             | Escherichia coli 1357                                  |
| pESBL | scaffold19_orf0094 | hypothetical protein ColIb-P9_p029                     |             | Plasmid ColIb-P9                                       |
| pESBL | scaffold19_orf0095 | DinI-like family protein                               |             | Escherichia coli MS 21-1                               |
| pESBL | scaffold19_orf0096 | hypothetical protein p026VIR_p092                      |             | Escherichia coli                                       |
| pESBL | scaffold19_orf0097 | hypothetical protein ECO103_p71                        |             | Escherichia coli O103:H2 str. 12009                    |
| pESBL | scaffold19_orf0098 | conserved hypothetical protein                         |             | Escherichia coli H299                                  |
| pESBL | scaffold19_orf0099 | hypothetical protein ND12Inc11_24                      |             | Escherichia coli                                       |
| pESBL | scaffold19_orf0100 | conserved hypothetical protein                         |             | Escherichia coli W                                     |
| pAA   | scaffold16_orf0001 | putative secreted protein                              |             | Streptomyces hygroscopicus ATCC 53653                  |
| pAA   | scaffold16_orf0002 | hypothetical protein c3579                             |             | Escherichia coli CFT073                                |
| pAA   | scaffold16_orf0003 | unknown protein encoded in ISEc8                       |             | Escherichia coli O157:H7 EDL933                        |
| pAA   | scaffold16_orf0004 | hypothetical protein SbBS512_A0019                     |             | Shigella boydii CDC 3083-94                            |
| pAA   | scaffold16_orf0005 | AggA457 protein  | AggA        | Escherichia coli                                       |
| pAA   | scaffold16_orf0006 | RecName: Full=Protein AggB; Flags: Precursor           | AggB        |  |
| pAA   | scaffold16_orf0007 | HdaC, HUS-associated diffuse adherence                 | AggC        | Escherichia coli                                       |
| pAA   | scaffold16_orf0008 | RecName: Full=Chaperone protein AggD; Flags: Precursor | AggD        |  |
| pAA   | scaffold16_orf0010 | putative resolvase                                     |             | Escherichia coli                                       |
| pAA   | scaffold16_orf0011 | 3-hydroxyisobutyrate dehydrogenase                     |             | Mycobacterium tuberculosis 210                         |
| pAA   | scaffold16_orf0012 | hypothetical protein ColIb-P9_p027                     |             | Plasmid ColIb-P9                                       |
| pAA   | scaffold16_orf0013 | StbA protein   |             | Escherichia coli MS 84-1                               |

|     |                    |  |  |
|-----|--------------------|--|--|
| pAA | scaffold16_orf0015 | putative 60 kDa chaperonin                                   | Escherichia coli E24377A   |
| pAA | scaffold16_orf0016 | hypothetical protein Collb-P9_p024                           | Plasmid Collb-P9   |
| pAA | scaffold16_orf0017 | resolvase  | Salmonella enterica subsp. enterica serovar Kentucky str. CVM29188 |
| pAA | scaffold16_orf0018 | plasmid maintenance protein CcdB                             | Escherichia coli   |
| pAA | scaffold16_orf0019 | plasmid maintenance protein CcdA                             | Escherichia coli   |
| pAA | scaffold16_orf0021 | hypothetical protein E4_23171                                | Escherichia sp. 4_1_40B  |
| pAA | scaffold16_orf0022 | hypothetical protein p1ECUMN_0160                            | Escherichia coli UMN026  |
| pAA | scaffold16_orf0024 | orf906   | Escherichia coli   |
| pAA | scaffold16_orf0026 | phage integrase  | Escherichia coli M863  |
| pAA | scaffold16_orf0027 | COG1506: Dipeptidyl aminopeptidases/acylaminoacyl-peptidases | Magnetospirillum magnetotacticum MS-1                              |
| pAA | scaffold16_orf0028 | hypothetical protein pECL46p020                              | Escherichia coli   |
| pAA | scaffold16_orf0029 | hypothetical protein pEC55989_0007                           | Escherichia coli 55989   |
| pAA | scaffold16_orf0030 | hypothetical protein IPF_103                                 | Escherichia coli 1520  |
| pAA | scaffold16_orf0031 | incFII family plasmid replication initiator RepA             | Escherichia coli MS 78-1   |
| pAA | scaffold16_orf0032 | replication initiation protein                               | Escherichia coli E128010   |
| pAA | scaffold16_orf0033 | replication protein  | Escherichia sp. 4_1_40B  |
| pAA | scaffold16_orf0034 | conjugal transfer pilus acetylation protein TraX             | Shigella flexneri 2a str. 301                                      |
| pAA | scaffold16_orf0035 | hypothetical protein pYT1_p113                               | Salmonella enterica subsp. enterica serovar Typhimurium            |
| pAA | scaffold16_orf0036 | DNA helicase TraI  | Escherichia coli MS 57-2   |
| pAA | scaffold16_orf0037 | conserved hypothetical protein                               | Salmonella enterica subsp. enterica serovar Kentucky               |
| pAA | scaffold16_orf0038 | hypothetical protein c3659                                   | Escherichia coli CFT073  |
| pAA | scaffold16_orf0039 | hypothetical protein c3661                                   | Escherichia coli CFT073  |
| pAA | scaffold16_orf0040 | hypothetical protein pB171_031                               | Escherichia coli   |
| pAA | scaffold16_orf0041 | conserved hypothetical protein                               | Escherichia coli H299  |
| pAA | scaffold16_orf0042 | conjugal transfer fertility inhibition protein FinO          | Escherichia coli   |
| pAA | scaffold16_orf0043 | conjugal transfer pilus acetylation protein TraX             | Salmonella enterica subsp. enterica serovar Kentucky str. CVM29188 |
| pAA | scaffold16_orf0044 | hypothetical protein pYT1_p113                               | Salmonella enterica subsp. enterica serovar Typhimurium            |
| pAA | scaffold16_orf0045 | conjugal transfer nickase/helicase TraI                      | Escherichia coli   |

|     |                    |   |  |
|-----|--------------------|---|--|
| pAA | scaffold16_orf0046 | conjugal transfer nickase/helicase TraI                                       | Salmonella enterica subsp. enterica serovar Kentucky str. CVM29188 |
| pAA | scaffold16_orf0047 | hypothetical protein R100p115.2br   | Plasmid R100   |
| pAA | scaffold16_orf0048 | Protein traJ  | Escherichia coli 55989   |
| pAA | scaffold16_orf0049 | TraM  | Escherichia coli   |
| pAA | scaffold16_orf0050 | putative lytic transglycosylase   | Escherichia coli ETEC H10407                                       |
| pAA | scaffold16_orf0051 | conserved hypothetical protein  | Escherichia coli MS 185-1  |
| pAA | scaffold16_orf0052 | putative recombinase  | Escherichia coli   |
| pAA | scaffold16_orf0053 | SepA  | Escherichia coli 536   |
| pAA | scaffold16_orf0054 | putative transposase  | Escherichia coli   |
| pAA | scaffold16_orf0057 | conserved hypothetical protein  | Escherichia coli MS 153-1  |
| pAA | scaffold16_orf0058 | AatD  | Escherichia sp. 4_1_40B  |
| pAA | scaffold16_orf0059 | AatC ATB binding protein of ABC transporter                                   | Escherichia coli 55989   |
| pAA | scaffold16_orf0060 | AatB  | Escherichia coli 55989   |
| pAA | scaffold16_orf0061 | AatA outermembrane protein  | Escherichia coli 55989   |
| pAA | scaffold16_orf0062 | AatP permease   | Escherichia sp. 4_1_40B  |
| pAA | scaffold16_orf0063 | serine protease eatA  | Shigella dysenteriae 1617  |
| pAA | scaffold16_orf0064 | protease IgA1   | Escherichia coli   |
| pAA | scaffold16_orf0065 | hypothetical protein E4_23001   | Escherichia sp. 4_1_40B  |
| pAA | scaffold16_orf0066 | Serine protease sat precursor (Secreted autotransporter toxin sat) (fragment) | Escherichia coli 55989   |
| pAA | scaffold16_orf0067 | ISPsy2, transposase   | Escherichia coli MS 185-1  |
| pAA | scaffold16_orf0069 | 14 kDa aggregative adherence fimbriae I protein (Fragment) (modular protein)  | Escherichia coli 55989   |
| pAA | scaffold16_orf0070 | putative transposase domain protein   | Escherichia coli 3431  |
| pAA | scaffold16_orf0071 | Serine protease sepA precursor (fragment)                                     | Escherichia sp. 4_1_40B  |
| pAA | scaffold16_orf0072 | IS186 transposase   | Escherichia coli UMNK88  |
| pAA | scaffold16_orf0073 | CvaB, IS186 transposase   | Escherichia coli BW2952  |
| pAA | scaffold16_orf0074 | hypothetical protein  | Escherichia coli   |
| pAA | scaffold16_orf0075 | hypothetical protein Mtub2_09757  | Mycobacterium tuberculosis 210                                     |
| pAA | scaffold16_orf0076 | putative IS639 ORF1   | Escherichia coli ETEC 1392/75                                      |
| pAA | scaffold16_orf0077 | putative transcriptional activator aggR (AAF-III) regulatory protein)         | Escherichia coli 55989   |
| pAA | scaffold16_orf0078 | transposase ORF A, IS1  | Escherichia coli 55989   |

|     |                    |   |  |                                |
|-----|--------------------|---|--|--------------------------------|
| pAA | scaffold16_orf0079 | transposase   |  | Escherichia coli M863          |
| pAA | scaffold16_orf0080 | hypothetical protein Mtub2_09757  |  | Mycobacterium tuberculosis 210 |
| pAA | scaffold16_orf0081 | hypothetical protein E4_23056   |  | Escherichia sp. 4_1_40B        |
| pAA | scaffold16_orf0083 | putative transposase (fragment)   |  | Escherichia coli 55989         |
| pAA | scaffold16_orf0084 | putative Isopentenyl-diphosphate delta-isomerase (IPP isomerase)<br>(Isopentenyl pyrophosphate isomerase) (IPP:DMAPP isomerase) |  | Escherichia coli 55989         |
| pAA | scaffold16_orf0085 | hypothetical protein pEC55989_0080  |  | Escherichia coli 55989         |
| pAA | scaffold16_orf0086 | conserved hypothetical protein  |  | Escherichia coli MS 119-7      |
| pAA | scaffold16_orf0087 | transposase   |  | Escherichia coli M863          |
| pAA | scaffold16_orf0088 | putative transposase insK for insertion sequence element IS150  |  | Shigella sonnei 53G            |
| pAA | scaffold16_orf0089 | putative protein encoded within IS  |  | Shigella sonnei Ss046          |