

Supplementary Online Content

Loman NJ, Constantinidou C, Christner M, et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of shiga-toxigenic *Escherichia coli* O104:H4. *JAMA*. doi:10.1001/jama.2013.3231

eSupplement

eTable 1. Clinical diagnosis and information recovered from STEC-positive samples using conventional microbiology

eTable 2. Information recovered from STEC-positive samples using diagnostic metagenomics

eTable 3. Information recovered from other pathogens using diagnostic metagenomics

eTable 4. Outbreak-specific genes in the draft genome of the outbreak strain obtained by metagenomics

eFigure 1. Genome depth of coverage plots for selected samples

eFigure 2. Genome coverage plots focusing on the region of the E coli O104:H4 STEC 280 genome around the Shiga-toxin encoding phage

eReferences

This supplementary material has been provided by the authors to give readers additional information about their work.

eSupplement

Bioinformatics methods

Sequences derived from the nuclear and mitochondrial genomes of the study subjects were removed from all datasets through alignment to the human reference database (hg19) using Bowtie 2.0.0-beta5 and retaining only reads that did not map to the human genome for further analysis¹. Human-filtered sequence reads, which have subsequently undergone a second, BLASTX-based filtering step to include only reads that have a best-hit to a bacterial taxon will be made available in the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/>) under accession number ERP001956.

A combined catalogue of environmental sequence tags was produced through *de novo* assembly. Due to memory constraints of this process, the combined reads were randomly sampled without replacement to produce a subset containing 10% of the original reads. Sub-sampling was performed using the 'sample' module of seqtk (<https://github.com/lh3/seqtk>). Reads were subsequently assembled using Ray Meta with a k-mer value of 31 without use of paired-end information². Contigs (EGTs) 100bp in length or greater were used in further analysis.

Sequence reads from each individual sample were aligned against the combined reference assembly using the bwasw module of BWA 0.6.2³. Additionally, sequence reads collected from stool samples of 124 healthy individuals in the MetaHIT project were aligned against the same reference assembly⁴. Contig statistics (length, GC coverage) and alignment statistics (depth of coverage) were determined using the assemblage suite of scripts (<https://github.com/sujaikumar/assemblage>). Approximate taxonomic assignment of contigs was performed through MEGABLAST best-hit against the non-redundant nucleotide database at the NCBI using BLAST+. Any contigs that showed best hits to non-bacterial sequences were removed.

In order to identify EGTs from outbreak-associated strains, n samples were chosen randomly from both the German collection and the MetaHIT collection ($n=1,2,5,10,20,40,45$). EGTs present in at least 20/45 German samples (empirically determined) and not present in 45 randomly selected MetaHIT samples were visualised using scatterplots demonstrating GC coverage, alignment depth of coverage and taxonomic assignment. Scatterplots were produced using custom versions of the blobology scripts in assemblage.

Contig clustering

The set of subtracted contigs were clustered according to a model-based classification algorithm using Mclust in the R statistical package with default parameters. EGTs in each cluster were used as seeds to recruit other EGTs from the combined assembly, using paired-end information and per-sample depth of coverage to provide linkage information. Pair-wise comparison of the log-scaled per-sample coverage of each EGT was performed using the Bray-Curtis dissimilarity index. EGTs with a dissimilarity index of ≤ 0.35 by comparison with the initial cluster seeds were recruited. Paired-end information was used to recruit further EGTs; if one pair in a read was the

reported hit for a contig within the cluster, the contig hitting its mate was recruited into the contig cluster. This process was iterated several times until no more contigs were recruited into the set.

EGT clustering is implemented through Python scripts, made available in a public Github repository (<http://github.com/nickloman>) in addition to the full bioinformatics pipeline, which allows users to repeat and extend the analysis in this paper, including the generation of figures and tables.

To aid clarity of presentation of figure 3, EGTs belonging to the outbreak genome were re-ordered using the Mauve Contig Mover⁵, using the *Escherichia coli* O104:H4 2011C-3493 chromosome and plasmid sequences as reference⁶.

Plots were produced in R, using ggplot2. Coverage plots were produced using the Bioconductor R library package with the ShortRead library⁷.

Contigs without a taxonomic assignment were further searched against the NCBI non-redundant protein database using BLASTX.

For the five STEC-negative samples that yielded other pathogens on culture, phylogenetic profiling of reads was performed using Metaphlan 1.7.6⁸, again using Bowtie as the alignment engine. Metaphlan performs taxonomic assignment of short sequencing reads by the use of a lineage-specific marker database. Reads mapping to entries in the marker database are considered to be unambiguous markers of a particular taxon, which in some cases are discriminatory at species level. Calculations of taxonomic frequency take into account the predicted genome length.

Recovery of typing and toxin data for *Clostridium difficile*, *Salmonella enterica*, *Campylobacter jejuni* and *C. concisus* was performed through BWASW alignment to manually-curated reference databases of genomes and genes, with manual inspection of alignments (details available in the Github repository). MLST sequence information was retrieved through BLAST searches of sequences on the BIGSdb MLST sequence server at the University of Oxford⁹.

Supplementary Results

Time taken for analysis

	Task	Time taken	Notes
Sequencing	MiSeq sequencing (2x151)	~27 hours	~5 minutes per cycle
	HiSeq 2500 sequencing (2x151)	~40 hours	~7 minutes per cycle
Bioinformatics	De novo assembly of 10% dataset using 2x8-core server with 256Gb RAM (Amazon EC2 High Memory Cluster Eight Extra Large)	3 hours, 13 minutes, 53 seconds	Time could be approximately halved by doubling cores

Numbers of OTUs

The number of contigs annotated to a particular taxonomic order, or not classified through BLASTN of the assembly.

	Not annotated	Bacteroidales	Clostridiales	Enterobacteriales	Lactobacillales	Selenomonadales
Panel A	81,035	28,777	21,639	14,023	1,430	437
Panel B	13,494	13,777	13,660	7,843	233	81
Panel C	150	0	4	294	2	0

eTable 1 - Clinical Diagnosis and Information recovered from STEC-positive samples using conventional microbiology

Sample	Days after diarrhea onset	HUS	Onset of HUS (days)	CFU/ml	Stx2 ELISA	Stx2 PCR
2535	3	+	1	20000	+	+
2638	5	+	-1	10000000	+	+
2661	7	+	-2	0*	-	+
2668	8	-	na	80	+	+
2669	6	+	2	600000	+	+
2723	3	+	4	20000000	+	+
2741	7	-	na	100000	-	+
2752	7	+	0	2000000	+	+
2758	1	-	na	60	-	+
2764	8	+	-1	4000000	+	+
2772	1	+	4	2000000	+	+
2828	6	+	-1	2000000	+	+
2840	5	-	na	20000000	+	+
2848	4	+	-1	200	-	+
2849	5	-	na	10000	-	+
2878	2	-	na	10000000	+	+
2880	1	-	na	3000000	-	+
2896	2	-	na	10000000	+	+
2971	1	-	na	4000000	+	+
3014	1	-	na	2000000	+	+
3093	2	-	na	100	-	+
3132	6	-	na	0*	-	+
3134	10	-	na	200	+	+
3135	3	-	na	4000000	+	+
3185	10	+	-7	2000000	+	+
3303	3	-	na	20000000	+	+
3411	8	-	na	200000	+	+
3549	14	-	na	40	-	+
3587	10	-	na	140	-	+
3646	6	-	na	0*	-	-
3751	19	+	-12	1000000	+	+
3852	1	-	na	400	+	+
3958	12	-	na	60	+	+
4112	14	-	na	2000	+	+
4141	5	-	na	400000	+	+
4168	8	-	na	400000	+	+
4198	6	-	na	1000	+	+
4328	20	-	Na	2000	+	+
4508	26	+	-23	4000000	-	+
5066	3	-	na	2000	+	+

* Initially culture positive, but could not be subsequently grown

eTable 2 - Information recovered from STEC-positive samples using diagnostic metagenomics

Sample	Run ID	Platform	Total number of reads (million)	Total throughput (gigabases)	% of non-human reads E. coli	H-antigen type (fliC)	Stx type	Stx copy number relative to chromosome	MLST assignment	E. coli O104:H4 chromosomal coverage (mean)	pAA coverage	pESBL coverage	C.difficile frequency
2535	N18	MiSeq	17.7	2.65	33.04	H4, H48, H49, H19	2	4.97	Mixture	167.54	314.71	120.72	nd
2638	N12	MiSeq	18.74	2.81	67.51	H4	2	1.55	ST678	20.51	32.5	24.91	nd
2723	N4	MiSeq	11.81	1.77	1.86	H4	2	5.7	Mixture (2/7 ST678 alleles detected)	4.64	9.4	5.47	nd
2764	N5	MiSeq	10.3	1.55	1.66	H4	2	1.23	ST678	4.03	1.29	5.13	nd
4328	N7	MiSeq	13.82	2.07	0	nd	nd	na	nd	0.03	0.03	0	nd
Mean average			(14.4)	(2.17)									
2535	H	HiSeq	55.01	8.25	47.79	H49, H48, H32, H4	2	4.69	Mixture	619.21	1209.12	487.79	0.0014
2638	H	HiSeq	30.82	4.62	64.41	H4	2	1.61	ST678	28.7	49.44	42.42	nd
2661	H	HiSeq	24.95	3.74	1.83	H4, H18	2	2.69	Mixture	8.64	13.46	8.42	nd
2668	H	HiSeq	38	5.7	1.02	nd	nd	na	nd	0.1	0.13	0.05	nd
2669	H	HiSeq	43.76	6.56	0.76	H4	2	11.79	ST678	7.53	18.02	10.66	0.0025
2723	H	HiSeq	9.55	1.43	1.92	H4	2	3.24	Mixture	3.44	8.74	5.16	nd
2741	H	HiSeq	7.54	1.13	14.72	H4	nd	na	Novel ST (Adk-6, fumc-249, gyrb-12, icd-1, mdh-20, pura-12, reca-7)	16.53	1.34	5.09	nd
2752	H	HiSeq	9.39	1.41	0.76	nd	2*	na	Mixture	1.51	0.91	1.24	nd
2758	H	HiSeq	9.63	1.44	2.35	H4	2	2.43	ST678 - partial	4.24	8.42	6.93	nd
2772	H	HiSeq	9.42	1.41	1.64	nd	2*	na	nd	0.12	0.2	0.2	nd
2828	H	HiSeq	9.04	1.36	4.09	H6	2*	na	Mixture	5.09	2.05	1.1	nd
2840	H	HiSeq	10.96	1.64	22.22	H4	2	1.12	ST678	39.12	75.64	54.25	nd

2848	H	HiSeq	10.92	1.64	2.4	H7	2*	0.41	Novel ST (adk-76, icd-36, pura-14, gyrB-novel, mdh-novel, recA - novel)	2.69	0.02	0.04	nd
2849	H	HiSeq	10.58	1.59	5.21	H4	2	0.92	ST678	10.93	15.16	13.91	nd
2878	H	HiSeq	12.57	1.89	0.76	H4	2*	na	Partial (gyrB-266)	1.91	4.09	2.81	0.0009
2880	H	HiSeq	11.69	1.75	9.58	nd	2*	na	Partial (adk-6)	1.84	2.86	3.23	nd
2896	H	HiSeq	11.86	1.78	30.57	H4	2	13.37	ST678	21.89	38.1	40.69	nd
2971	H	HiSeq	16.81	2.52	5.33	H4	2	1.34	ST678	10.51	25.53	20.61	nd
3014	H	HiSeq	18.97	2.84	55.99	H4	2	7.86	ST678	18.75	38.29	44.74	nd
3093	H	HiSeq	30.2	4.53	0	nd	nd	na	nd	0.02	0.05	0.03	nd
3132	H	HiSeq	27.01	4.05	0	nd	2*	na	nd	0.45	1.1	0.8	nd
3134	H	HiSeq	15.42	2.31	2.28	H4, H33	2*	na	Mixture	6.74	3.32	4.42	0.0036
3135	H	HiSeq	20.1	3.01	1.68	H4	2	0.88	ST678 Partial (adk-6, pura-7, reca-7)	4.4	8.65	6.48	0.0006
3185	H	HiSeq	14.23	2.13	3.26	H4	2	0.5	Mixture (adk-6)	7.97	12.62	9.59	nd
3303	H	HiSeq	11.58	1.74	21.04	H4	2	1.34	ST678 - partial (adk-6, fumc-6, gyrb-5, icd-136)	6.13	11.41	11.23	nd
3411	H	HiSeq	13.38	2.01	0	nd	nd	na	nd	0.02	0	0	nd
3549	H	HiSeq	14.56	2.18	0.37	H7	nd	na	Partial (adk-27, mdh-26, pura-19, reca-22)	3.21	7.77	3.1	nd
3587	H	HiSeq	16.84	2.53	0.07	nd	nd	na	nd	0.13	0.06	0.08	0.0014
3646	H	HiSeq	15.35	2.3	0.46	nd	nd	na	nd	1.1	0.73	0.47	nd
3751	H	HiSeq	14.13	2.12	4.59	H12	2*	na	Mixture	9.92	3.36	4.89	nd
3852	H	HiSeq	15.2	2.28	0	nd	2*	na	nd	0.03	0.03	0.06	nd
3958	H	HiSeq	18.98	2.85	0.11	nd	nd	na	nd	0.31	0.11	0.11	0.004
4112	H	HiSeq	21.64	3.25	0.21	nd	nd	na	nd	0.47	0.16	0.23	0.0041
4141	H	HiSeq	14.53	2.18	0.22	nd	2*	na	nd	0.52	0.48	0.37	nd
4168	H	HiSeq	19.36	2.9	0.15	nd	2*	na	nd	0.38	1.23	0.53	nd
4198	H	HiSeq	13.32	2	10.57	-	nd	na	ST73 clonal complex (adk-	16.3	2.98	6.69	0.0069

									36, fumc-24, gyrb-9, icd-13, mdh-95, pura- 11, reca-25				
4328	H	HiSeq	8.56	1.28	0.01	nd	nd	na	nd	0.01	0.02	0.01	nd
4508	H	HiSeq	8.75	1.31	1.3	-	nd	na	nd	1.37	0.2	0.64	0.0037
5066	H	HiSeq	10.53	1.58	23.91	-	2*	na	Mixture (mdh-169, reca-16)	2.19	0.77	0.46	nd
Mean coverage			(17.31)	(2.60)									

* Indicates *stxAB* were found only as partial length fragments

eTable 3 - Information recovered from other pathogens using diagnostic metagenomics

Sample	Clinical diagnosis	RunID	Platform	Total number of reads (million)	Total throughput (gigabases)	Pathogen(s) detected	MLST alleles	Additional information
1122	<i>C. difficile</i>	O9	MiSeq	8.00	1.21	<i>C. difficile</i> (0.1%)	nd	<i>toxAB</i> positive
		H	HiSeq	93.09	14.06	<i>C. difficile</i> (0.13%)	Novel ST: (<i>adk-1</i> , <i>recA-5</i> , <i>sodA-18</i> , <i>glyA-23</i> , <i>atpA-11</i>)	<i>toxAB</i> positive
1196	<i>S. enterica</i>	N21	MiSeq	17.48	2.64	nd	nd	
		H	HiSeq	73.61	11.12	<i>S. enterica enterica</i> serogroup B	nd	Reads match to serovars <i>Typhimurium</i> and <i>Heidelberg</i> , suggesting serogroup B strain
1253	<i>C. difficile</i>	N6	MiSeq	12.15	1.84	<i>Campylobacter concisus</i> (0.21%)	nd	<i>cdtAB</i> positive
		H	HiSeq	82.38	12.44	<i>Campylobacter concisus</i> (0.24%), <i>Clostridium difficile</i> (0.002%),	Novel ST (<i>Cco_{PGM}-30</i> , <i>Cco_{atpA}-28</i> , <i>Cco_{aspA} - novel</i> , <i>Cco_{glnA}-novel</i> , <i>Cco_{gitA}-novel</i> , <i>Cco_{glyA}-novel</i> , <i>Cco_{iIVD}-novel</i>)	<i>cdtAB</i> positive
4096	<i>S. enterica</i>	U4	MiSeq	12.66	1.91	nd	-	-
		N2	MiSeq	18.91	2.86	nd	-	-
		H	HiSeq	82.84	12.51	nd	-	Operator error during library construction, results discarded
4961	<i>C. jejuni</i>	N8	MiSeq	12.46	1.88	<i>C. jejuni</i> (0.65%)	nd	Campylobacter toxins: <i>cdtA</i> and <i>cdtB</i> detected
		H	HiSeq	110.05	16.6	<i>C. jejuni</i> (1.2%)	Partial sequences	Campylobacter toxins: <i>cdtABC</i> detected

eTable 4 – Outbreak-specific genes in the draft genome of the outbreak strain obtained by metagenomics. EGTs were determined by the initial clustering approach (>20 German samples, minus those present in MetaHIT samples). Annotations were determined by BLASTX searches against the NCBI's non-redundant protein database. A full list of annotations for the draft genome outbreak strain are available in the Github repository.

EGT ID	Hit length (residues)	Percentage identity	Genbank ID	Description
contig-103000027	73	100%	gb EJY97814.1	hypothetical protein B185_23682 [Escherichia coli J96]
contig-103000057	134	99%	ref YP_006784600.1	phage-like protein [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779525.1 phage-like protein [Escherichia coli O104:H4 str. 2011C-3493] gb EGR60070.1 hypothetical protein HUSEC41_27409 [Escherichia coli O104:H4 str. 01-09591] gb EHF17837.1 ...
contig-103000057	113	97%	gb EGW94600.1	hypothetical protein EC30301_0279 [Escherichia coli 3030-1]
contig-103000057	74	86%	ref YP_006784601.1	hypothetical protein O3O_11415 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779526.1 hypothetical protein O3K_14200 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR60071.1 hypothetical protein HUSEC41_27414 [Escherichia coli O104:H4 str. ...
contig-103000057	61	77%	ref YP_006784599.1	hypothetical protein O3O_11425 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779524.1 hypothetical protein O3K_14190 [Escherichia coli O104:H4 str. 2011C-3493] gb AFS57514.1 hypothetical protein O3M_14170 [Escherichia coli O104:H4 str. 2009 ...
contig-1038000058	126	100%	gb EHY21132.1	rhsC protein in rhs element [Escherichia coli DEC15E]
contig-1067000018	39	100%	gb EFZ71369.1	addiction module antidote protein, HigA family [Escherichia coli OK1357]
contig-1067000018	22	91%	ref ZP_06657472.1	predicted protein [Escherichia coli B185] gb EFF05456.1 predicted protein [Escherichia coli B185]
contig-1075000027	47	66%	gb EII48584.1	hypothetical protein EC23916_4858 [Escherichia coli 2.3916]
contig-108000037	477	57%	ref YP_006779881.1	putative tail fiber protein [Escherichia coli O104:H4 str. 2011C-3493] gb EHF44810.1 hypothetical protein EUIG_01397 [Escherichia coli O104:H4 str. 11-4522] gb EHF61127.1 hypothetical protein EUJG_00018 [Escherichia coli O104:H4 str. 11-4623] gb EH ...
contig-109000050	193	91%	ref YP_006784265.1	lipopolysaccharide core heptose(II)-phosphate phosphatase [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779194.1 lipopolysaccharide core heptose(II)-phosphate phosphatase [Escherichia coli O104:H4 str. 2011C-3493] gb EHF25933.1 hypothetical ...
contig-109000050	30	77%	gb EGX11033.1	hypothetical protein ECSTECH18_2347 [Escherichia coli STEC_H.1.8]
contig-109000050	24	67%	gb EGX11033.1	hypothetical protein ECSTECH18_2347 [Escherichia coli STEC_H.1.8]
contig-1111000058	109	100%	gb EHY18377.1	outer membrane autotransporter barrel domain protein [Escherichia coli DEC15E]

contig-112000045	529	100% ref YP_006785038.1	hypothetical protein O3O_09175 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779899.1 hypothetical protein O3K_16125 [Escherichia coli O104:H4 str. 2011C-3493] gb EGT68084.1 hypothetical protein C22711_2113 [Escherichia coli O104:H4 str. C2 ...
contig-1136000025	89	100% gb EIQ36831.1	phosphohydrolase [Shigella sonnei 3226-85]
contig-1143000028	125	100% gb EHY18377.1	outer membrane autotransporter barrel domain protein [Escherichia coli DEC15E]
contig-1168000057	69	91% ref NP_613034.1	hypothetical protein Stx2lp156 [Stx2 converting phage I] dbj BAB88003.1 hypothetical protein [Stx2 converting phage I]
contig-1168000057	8	100% gb EIH00761.1	lysis protein S [Escherichia coli 5.0588]
contig-1179000044	119	92% ref ZP_09829171.1	transposase [Pantoea stewartii subsp. stewartii DC283] gb EHT99936.1 transposase [Pantoea stewartii subsp. stewartii DC283]
contig-1179000044	26	96% ref ZP_09829171.1	transposase [Pantoea stewartii subsp. stewartii DC283] gb EHT99936.1 transposase [Pantoea stewartii subsp. stewartii DC283]
contig-1182000049	28	61% gb EHX82896.1	ATPase [Escherichia coli DEC14C]
contig-1182000049	15	100% gb EGJ83006.1	hypothetical protein SF274771_4232 [Shigella flexneri 2747-71] gb EGK18126.1 hypothetical protein SFK218_4600 [Shigella flexneri K-218] gb EIQ22233.1 hypothetical protein SFK404_4533 [Shigella flexneri K-404] gb EJL11259.1 hypothetical protein SF6 ...
contig-1191000017	45	100% ref YP_006784282.1	hypothetical protein O3O_13030 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779211.1 hypothetical protein O3K_12605 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR70924.1 hypothetical protein HUSEC_28327 [Escherichia coli O104:H4 str. LB ...
contig-1195000023	62	100% gb EFZ76734.1	S-formylglutathione hydrolase [Escherichia coli RN587/1] gb EII84510.1 S-formylglutathione hydrolase [Escherichia coli 3003] gb EKI32023.1 S-formylglutathione hydrolase [Escherichia coli ARS4.2123]
contig-1198000013	31	100% gb EKB27377.1	transposase insH for insertion sequence element IS5H [Aeromonas hydrophila SSU]
contig-1198000013	28	100% ref ZP_07243412.1	transposase, IS4 family [Escherichia coli MS 146-1] gb EFK93064.1 transposase, IS4 family [Escherichia coli MS 146-1]
contig-1202000048	96	100% gb EKJ56425.1	hypothetical protein EC01288_3201 [Escherichia coli 0.1288]
contig-1210000013	59	98% gb EKI41632.1	antigen 43 [Escherichia coli 3006]
contig-1225000017	41	100% gb EIH44518.1	hypothetical protein EC970259_2571 [Escherichia coli 99.0741]
contig-1230000054	40	75% gb EHU29356.1	antigen 43 domain protein [Escherichia coli DEC1E]
contig-1230000054	15	100% gb EIN65235.1	hypothetical protein ECPA10_6158, partial [Escherichia coli PA10]

contig-1235000032	125	100% gb EHY18377.1	outer membrane autotransporter barrel domain protein [Escherichia coli DEC15E]
contig-1238000020	30	100% gb EIZ93082.1	hypothetical protein SEEN539_10471 [Salmonella enterica subsp. enterica serovar Newport str. CVM 21539]
contig-1238000020	17	100% gb ABI99686.1	conserved hypothetical protein [Escherichia coli APEC O1]
contig-1246000017	113	99% gb EKI55117.1	hypothetical protein ECN1_0497 [Escherichia coli N1]
contig-1246000017	69	90% ref ZP_08376922.1	hypothetical protein ECPG_03356 [Escherichia coli H591] gb EGI47550.1 hypothetical protein ECPG_03356 [Escherichia coli H591]
contig-1246000017	10	100% gb EHV66413.1	hypothetical protein ECDEC6C_0542 [Escherichia coli DEC6C]
contig-1274000059	113	100% gb EIL10415.1	T3SS effector EspX, partial [Escherichia coli O103:H2 str. CVM9450]
contig-1283000044	51	96% ref ZP_09829171.1	transposase [Pantoea stewartii subsp. stewartii DC283] gb EHT99936.1 transposase [Pantoea stewartii subsp. stewartii DC283]
contig-1283000044	23	96% ref YP_002933937.1	hypothetical protein NT01EI_2533 [Edwardsiella ictaluri 93-146] gb ACR69702.1 conserved hypothetical protein [Edwardsiella ictaluri 93-146]
contig-1288000044	63	100% gb EHF17589.1	antigen 43 [Escherichia coli O104:H4 str. C227-11]
contig-1304000061	253	91% gb EIE54906.1	putative ATP-dependent Clp proteinase Aec27 ATP-binding chain, with chaperone activity [Escherichia coli AI27] gb EIL81997.1 putative ATP-dependent Clp proteinase Aec27 ATP-binding chain, with chaperone activity [Escherichia coli CUMT8]
contig-1304000061	150	91% ref YP_002401351.1	hypothetical protein EC55989_0222 [Escherichia coli 55989] ref YP_006785884.1 hypothetical protein O3O_04910 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006780758.1 hypothetical protein O3K_20470 [Escherichia coli O104:H4 str. 2011C-3493] em ...
contig-1314000003	60	100% gb EIL72989.1	biotin biosynthesis protein BioC [Escherichia coli CUMT8]
contig-1315000009	50	100% gb EKI35282.1	H+ symporter family protein [Escherichia coli 07798]
contig-1329000010	111	89% gb EHV00667.1	hypothetical protein ECDEC4B_1103 [Escherichia coli DEC4B]
contig-1332000009	52	100% ref ZP_10811016.1	Cu(I)/Ag(I) efflux system membrane protein CusA, partial [Enterobacter sp. SST3] gb EJO44633.1 Cu(I)/Ag(I) efflux system membrane protein CusA, partial [Enterobacter sp. SST3]
contig-1344000013	76	100% gb EKJ62283.1	protein RhsE [Escherichia coli 0.1288]
contig-1344000046	64	100% gb EHF57078.1	hypothetical protein EUKG_02060, partial [Escherichia coli O104:H4 str. 11-4632 C1]
contig-1348000051	37	68% gb EGR62318.1	hypothetical protein HUSEC41_15823 [Escherichia coli O104:H4 str. 01-09591] gb EGR73143.1 hypothetical protein HUSEC_16178 [Escherichia coli O104:H4 str. LB226692]

contig-1350000040	31	81% ref NP_612920.1	hypothetical protein Stx2lp041 [Stx2 converting phage I] dbj BAB87889.1 hypothetical protein [Stx2 converting phage I] dbj BAC77857.1 hypothetical protein [Stx1 converting phage] dbj BAC78023.1 hypothetical protein [Stx2 converting phage II]
contig-1350000040	19	95% ref NP_612920.1	hypothetical protein Stx2lp041 [Stx2 converting phage I] dbj BAB87889.1 hypothetical protein [Stx2 converting phage I] dbj BAC77857.1 hypothetical protein [Stx1 converting phage] dbj BAC78023.1 hypothetical protein [Stx2 converting phage II]
contig-1350000040	16	100% ref YP_006785011.1	hypothetical protein O3O_09310 [Escherichia coli O104:H4 str. 2009EL-2071] gb AFS57861.1 hypothetical protein O3M_15965 [Escherichia coli O104:H4 str. 2009EL-2050] gb AFS85629.1 hypothetical protein O3O_09310 [Escherichia coli O104:H4 str. 2009EL-2 ...
contig-1353000043	56	100% gb EKI53726.1	anthranilate synthase component I [Escherichia coli N1]
contig-1361000012	105	90% ref YP_006784913.1	hypothetical protein O3O_09825 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779772.1 hypothetical protein O3K_15465 [Escherichia coli O104:H4 str. 2011C-3493] gb EHF17591.1 hypothetical protein EUAG_01495 [Escherichia coli O104:H4 str. C22 ...
contig-1364000053	68	100% gb EKJ83043.1	nucleoside transporter [Escherichia coli AD30]
contig-1378000016	112	89% ref YP_006785609.1	hypothetical protein O3O_06305 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006780469.1 hypothetical protein O3K_18990 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR75713.1 hypothetical protein HUSEC_02703 [Escherichia coli O104:H4 str. LB ...
contig-1383000032	51	100% gb EIO64761.1	ygbA-ard, partial [Escherichia coli TW09098]
contig-1384000034	70	87% ref YP_006785027.1	portal protein [Escherichia coli O104:H4 str. 2009EL-2071] gb AFS57877.1 portal protein [Escherichia coli O104:H4 str. 2009EL-2050] gb AFS85613.1 portal protein [Escherichia coli O104:H4 str. 2009EL-2071]
contig-1404000005	36	92% ref ZP_09808339.1	putative acetyltransferase [Escherichia hermannii NBRC 105704] dbj GAB53504.1 putative acetyltransferase [Escherichia hermannii NBRC 105704]
contig-1404000005	33	82% ref YP_980147.1	TnpA family transposase [Verminephrobacter eiseniae EF01-2] gb ABM60736.1 Transposase and inactivated derivatives TnpA family-like protein [Verminephrobacter eiseniae EF01-2]
contig-1426000033	64	78% gb EKH07481.1	DNA-directed RNA polymerase, beta subunit [Escherichia coli PA7]
contig-1442000048	59	100% gb EFZ74044.1	prepilin peptidase-dependent protein B [Escherichia coli RN587/1] gb EII88209.1 prepilin-type cleavage/methylation N-terminal domain protein [Escherichia coli 3003] gb EKI25174.1 prepilin peptidase dependent protein B [Escherichia coli ARS4.2123]
contig-1448000041	50	100% gb EIQ16280.1	putative molybdate metabolism regulator domain protein [Shigella flexneri K-1770]

contig-1474000039	50	98% gb EKB62633.1	glycerol dehydrogenase [Klebsiella pneumoniae subsp. pneumoniae WGLW1] gb EKB81635.1 glycerol dehydrogenase [Klebsiella pneumoniae subsp. pneumoniae WGLW3]
contig-1481000036	166	100% ref YP_002404021.1	hypothetical protein EC55989_3029 [Escherichia coli 55989] emb CAU98911.1 conserved hypothetical protein [Escherichia coli 55989]
contig-1487000017	54	100% gb EGT67669.1	hypothetical protein C22711_1697 [Escherichia coli O104:H4 str. C227-11]
contig-1491000004	55	100% gb EII65555.1	crotonobetainyl-CoA:carnitine CoA-transferase [Escherichia coli 2.4168]
contig-150000046	422	100% gb EKH68165.1	hypothetical protein ECPA23_3299 [Escherichia coli PA23]
contig-150000054	257	100% gb EHV74711.1	hypothetical protein ECDEC6E_1941 [Escherichia coli DEC6E]
contig-1519000033	113	88% gb EIL57335.1	hypothetical protein EC54115_06283 [Escherichia coli 541-15]
contig-1519000033	50	98% gb EHX95881.1	CRISPR-associated protein Cas1 [Escherichia coli DEC15A] gb EHY12848.1 CRISPR-associated protein Cas1 [Escherichia coli DEC15D] gb EHY18195.1 CRISPR-associated protein Cas1 [Escherichia coli DEC15E]
contig-1539000039	143	100% gb EIQ40443.1	phosphohydrolase [Shigella sonnei 3233-85]
contig-1549000025	62	92% ref YP_001449276.1	hypothetical protein PBV4795_ORF39 [Phage BP-4795] dbj BAC77957.1 hypothetical protein [Stx1 converting phage] emb CAJ21175.1 hypothetical protein [Enterobacteria phage BP-4795]
contig-1554000000	48	100% gb EIL56690.1	IS911 orfA, partial [Escherichia coli 541-15]
contig-156000009	218	78% gb EII59438.1	hypothetical protein EC33884_4237 [Escherichia coli 3.3884]
contig-156000009	118	100% ref YP_003233679.1	hypothetical protein ECO111_1180 [Escherichia coli O111:H- str. 11128] ref YP_006785010.1 hypothetical protein O3O_09315 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779871.1 hypothetical protein O3K_15985 [Escherichia coli O104:H4 str. 20 ...
contig-156000009	63	78% ref NP_612920.1	hypothetical protein Stx2lp041 [Stx2 converting phage I] dbj BAB87889.1 hypothetical protein [Stx2 converting phage I] dbj BAC77857.1 hypothetical protein [Stx1 converting phage] dbj BAC78023.1 hypothetical protein [Stx2 converting phage II]
contig-1577000000	160	100% ref YP_006785007.1	hypothetical protein O3O_09330 [Escherichia coli O104:H4 str. 2009EL-2071] gb AFS57857.1 hypothetical protein O3M_15945 [Escherichia coli O104:H4 str. 2009EL-2050] gb AFS85633.1 hypothetical protein O3O_09330 [Escherichia coli O104:H4 str. 2009EL-2 ...
contig-1608000043	122	100% gb EIQ53929.1	putative capsid protein of prophage [Shigella dysenteriae 225-75]
contig-1612000033	55	100% gb EKI53766.1	FAD dependent oxidoreductase [Escherichia coli N1]
contig-1618000061	103	100% gb EIQ55329.1	type VI secretion protein lcmF [Shigella sonnei 4822-66]
contig-1634000034	73	85% gb EIL81459.1	phosphate transporter [Escherichia coli CUMT8]
contig-1655000007	63	98% gb EKI40985.1	hypothetical protein EC07798_2381 [Escherichia coli 07798]

contig-1659000043	61	100% ref ZP_10407849.1	conserved protein nucleoside triphosphate hydrolase domain protein [Citrobacter sp. A1] gb EJF22483.1 conserved protein nucleoside triphosphate hydrolase domain protein [Citrobacter sp. A1]
contig-1679000029	64	100% gb EHV72051.1	hypothetical protein ECDEC6D_1854 [Escherichia coli DEC6D] gb EHV74692.1 hypothetical protein ECDEC6E_1922 [Escherichia coli DEC6E]
contig-168000050	154	99% gb EFZ50538.1	putative transposase insK for insertion sequence element IS150 [Shigella sonnei 53G] gb EIQ36917.1 integrase core domain protein [Shigella sonnei 3226-85] gb EIQ40180.1 integrase core domain protein [Shigella sonnei 3233-85] gb EJL12536.1 putative ...
contig-168000050	143	97% gb EGB59949.1	transposase [Escherichia coli M863]
contig-1682000039	57	100% gb EGX00298.1	phosphonate ABC transporter, ATP-binding protein [Escherichia coli STEC_MHI813]
contig-1683000002	49	100% gb EHN95301.1	hypothetical protein ESPG_00506 [Escherichia coli H397]
contig-1683000046	51	100% gb EIF84899.1	lead, cadmium, zinc and mercury-transporting ATPase [Escherichia coli M919]
contig-1693000060	32	81% ref ZP_06988343.1	predicted protein [Escherichia coli FVEC1302] gb EFI22294.1 predicted protein [Escherichia coli FVEC1302]
contig-1693000060	18	78% ref ZP_06988343.1	predicted protein [Escherichia coli FVEC1302] gb EFI22294.1 predicted protein [Escherichia coli FVEC1302]
contig-1697000031	60	100% ref YP_006786895.1	protease IgA1 [Escherichia coli O104:H4 str. 2009EL-2071] gb AFS88882.1 protease IgA1 [Escherichia coli O104:H4 str. 2009EL-2071]
contig-1697000031	20	100% ref YP_006116737.1	putative transposase [Escherichia coli ETEC H10407] emb CBJ02694.1 putative transposase [Escherichia coli ETEC H10407]
contig-1707000036	48	100% gb EHF64922.1	hypothetical protein EUMG_01383 [Escherichia coli O104:H4 str. 11-4632 C3]
contig-1712000018	110	98% gb EKH07179.1	hypothetical protein ECFRIK920_1921 [Escherichia coli FRIK920]
contig-1713000013	39	97% ref ZP_07153084.1	conserved hypothetical protein [Escherichia coli MS 21-1] gb EFK20183.1 conserved hypothetical protein [Escherichia coli MS 21-1]
contig-1714000052	54	100% ref YP_006786875.1	conjugal transfer nickase/helicase Tral [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006792604.1 conjugal transfer nickase/helicase Tral [Escherichia coli O104:H4 str. 2011C-3493] gb AFS59879.1 conjugal transfer nickase/helicase Tral [Escheri ...
contig-1715000057	51	100% gb EKJ83396.1	hypothetical protein ECAD30_13020 [Escherichia coli AD30]
contig-1721000030	47	100% gb EKI36106.1	N-acetylmannosamine kinase [Escherichia coli 07798]
contig-1724000000	72	100% gb EKH68165.1	hypothetical protein ECPA23_3299 [Escherichia coli PA23]
contig-1730000052	177	83% gb EGR60404.1	hypothetical protein HUSEC41_25692 [Escherichia coli O104:H4 str. 01-09591]

contig-173000052	25	36% ref YP_002397159.1	hypothetical protein ECED1_1138 [Escherichia coli ED1a] emb CAR07339.1 conserved hypothetical protein [Escherichia coli ED1a]
contig-1740000052	35	100% gb EID67969.1	polynucleotide adenyltransferase/metal dependent phosphohydrolase [Escherichia coli W26]
contig-1742000042	53	100% gb EKH53027.1	hypothetical protein ECFRIK2001_3424 [Escherichia coli FRIK2001]
contig-1742000042	25	100% ref YP_006784787.1	hypothetical protein O3O_10460 [Escherichia coli O104:H4 str. 2009EL-2071] gb AFS85853.1 hypothetical protein O3O_10460 [Escherichia coli O104:H4 str. 2009EL-2071]
contig-1744000016	51	100% gb EHF48026.1	antitoxin YeeU [Escherichia coli O104:H4 str. 11-4404]
contig-1747000040	75	88% gb EKI45172.1	tail assembly protein I [Escherichia coli 07798]
contig-1756000056	110	100% gb EGI99003.1	vitamin B12 transport periplasmic protein btuE [Shigella dysenteriae 155-74]
contig-1760000013	58	100% gb EKI53765.1	gamma-glutamyl-gamma-aminobutyraldehyde dehydrogenase [Escherichia coli N1]
contig-1761000023	42	100% gb EKJ83357.1	Tat pathway signal sequence protein [Escherichia coli AD30]
contig-1764000056	78	100% gb EKJ58000.1	hypothetical protein EC01288_2403 [Escherichia coli 0.1288]
contig-1766000017	23	91% gb EHY02023.1	insertion sequence 2 OrfB domain protein [Escherichia coli DEC15A] gb EHY13417.1 insertion sequence 2 OrfB domain protein [Escherichia coli DEC15D]
contig-1778000007	80	86% gb EII93782.1	outer membrane autotransporter barrel domain protein [Escherichia coli TW07793]
contig-1792000046	52	100% emb CCJ78361.1	ATP-dependent Clp protease ATP-binding subunit ClpX [Cronobacter muytjensii 530]
contig-1798000025	51	49% gb EKI55121.1	putative outer membrane efflux protein OprM [Escherichia coli N1]
contig-1802000014	36	100% gb EHF49448.1	hypothetical protein EUIG_04407, partial [Escherichia coli O104:H4 str. 11-4522]
contig-1802000014	41	88% ref ZP_03060247.1	hypothetical protein EcB171_2428 [Escherichia coli B171] ref ZP_03060248.1 hypothetical protein EcB171_2440 [Escherichia coli B171] gb EDX30483.1 hypothetical protein EcB171_2428 [Escherichia coli B171] gb EDX30484.1 hypothetical protein EcB171_24 ...
contig-1806000001	53	100% gb EKH78678.1	molybdate ABC transporter, periplasmic molybdate-binding protein [Escherichia coli PA23]
contig-1806000056	51	100% gb EJZ02424.1	type II secretion protein GspC, partial [Escherichia coli J96]
contig-1807000002	35	100% gb EJK97017.1	endodeoxyribonuclease RusA family protein [Escherichia coli STEC_O31]
contig-1807000002	23	100% gb EIL65371.1	hypothetical protein EC75_11808, partial [Escherichia coli 75]
contig-1817000007	46	96% ref ZP_07179195.1	conserved hypothetical protein [Escherichia coli MS 45-1] gb EFJ90655.1 conserved hypothetical protein [Escherichia coli MS 45-1]
contig-1817000031	52	67% gb EHW98216.1	repA3 [Escherichia coli DEC10F]
contig-1824000041	106	100% gb EGT69535.1	hypothetical protein C22711_3565 [Escherichia coli O104:H4 str. C227-11] gb EIH24460.1 putative permease [Escherichia coli 1.2264] gb EIH32198.1 putative permease [Escherichia coli 96.0497] gb EJK94371.1 permease family protein [Escherichia coli S ...

contig-1833000035	49	100% ref ZP_10980532.1	shikimate transporter [Escherichia sp. 1_1_43] gb EJZ48003.1 shikimate transporter [Escherichia sp. 1_1_43]
contig-1835000008	48	100% gb EKI52485.1	hypothetical protein ECN1_2015 [Escherichia coli N1]
contig-1845000015	63	100% gb EIG93190.1	general secretion pathway protein M [Escherichia coli 97.0246] gb EKJ83942.1 general secretion pathway protein M [Escherichia coli AD30]
contig-1845000015	25	100% gb EKJ83941.1	general secretion pathway protein L [Escherichia coli AD30]
contig-186000016	301	95% gb EHV74689.1	hypothetical protein ECDEC6E_1919 [Escherichia coli DEC6E]
contig-1868000022	58	100% gb EIL57343.1	gamma-aminobutyrate transporter [Escherichia coli 541-15]
contig-1874000025	40	100% ref YP_006170596.1	Putative membrane fusion protein silB [Escherichia coli P12b] ref ZP_10411248.1 copper/silver efflux system membrane fusion protein CusB [Enterobacter sp. Ag1] gb EHT04230.1 putative membrane fusion protein silB [Klebsiella oxytoca 10-5246] gb AFG4 ...
contig-19000057	411	100% ref YP_002403328.1	NnaC [Escherichia coli 55989] ref YP_006783623.1 N-acylneuraminate cytidyltransferase [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006778550.1 N-acylneuraminate cytidyltransferase [Escherichia coli O104:H4 str. 2011C-3493] emb CAU98167.1 ...
contig-19000057	387	100% ref YP_002403327.1	NnaA [Escherichia coli 55989] ref YP_006783624.1 NnaA [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006778551.1 NnaA [Escherichia coli O104:H4 str. 2011C-3493] emb CAU98166.1 NnaA [Escherichia coli 55989] gb EGT66551.1 nnaA [Escherichia coli ...
contig-19000057	346	100% ref YP_002403329.1	NnaB [Escherichia coli 55989] ref YP_006783622.1 NnaB [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006778549.1 NnaB [Escherichia coli O104:H4 str. 2011C-3493] emb CAU98168.1 NnaB [Escherichia coli 55989] gb EGR63368.1 NnaB [Escherichia coli ...
contig-19000057	203	41% gb EKA93510.1	N-acetylneuraminate synthase [Fusobacterium periodonticum D10]
contig-19000057	316	100% ref YP_002403326.1	hypothetical protein EC55989_2292 [Escherichia coli 55989] ref YP_006783625.1 hypothetical protein O3O_16375 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006778552.1 hypothetical protein O3K_09245 [Escherichia coli O104:H4 str. 2011C-3493] em ...
contig-19000057	183	100% gb EHF55759.1	hypothetical protein EUKG_02691 [Escherichia coli O104:H4 str. 11-4632 C1] gb EHF57443.1 hypothetical protein EUJG_01346 [Escherichia coli O104:H4 str. 11-4623] gb EHF59645.1 hypothetical protein EULG_02708 [Escherichia coli O104:H4 str. 11-4632 C2 ...
contig-1904000045	45	100% gb EII85390.1	ABC transporter, permease protein [Escherichia coli 3003]

contig-1913000017	71	100% ref ZP_08352161.1	type VI secretion protein, VC_A0110 family [Escherichia coli M718] gb EGI22929.1 type VI secretion protein, VC_A0110 family [Escherichia coli M718]
contig-1927000019	47	96% gb EGK29645.1	hypothetical protein SFK218_5541 [Shigella flexneri K-218] gb EGK30349.1 hypothetical protein SFK218_5515 [Shigella flexneri K-218]
contig-1927000019	37	100% gb EHY18057.1	serine protease sepA autotransporter domain protein [Escherichia coli DEC15E]
contig-1929000014	47	96% gb EGK29645.1	hypothetical protein SFK218_5541 [Shigella flexneri K-218] gb EGK30349.1 hypothetical protein SFK218_5515 [Shigella flexneri K-218]
contig-1929000014	37	100% gb EHF62878.1	serine protease sepA autotransporter, partial [Escherichia coli O104:H4 str. 11-4632 C3]
contig-1938000048	41	100% gb EKJ06615.1	protein RhsE, partial [Escherichia coli EC1856] gb EKJ64386.1 protein RhsE, partial [Escherichia coli FRK523]
contig-1964000010	28	89% ref YP_006779281.1	hypothetical protein O3K_12955 [Escherichia coli O104:H4 str. 2011C-3493] gb AFS74480.1 hypothetical protein O3K_12955 [Escherichia coli O104:H4 str. 2011C-3493]
contig-1966000014	54	100% gb EIG91901.1	protein-(glutamine-N5) methyltransferase, release factor-specific [Escherichia coli 97.0246] gb EIH26741.1 protein-(glutamine-N5) methyltransferase, release factor-specific [Escherichia coli 1.2264] gb EII24982.1 protein-(glutamine-N5) methyltransf ...
contig-1969000030	52	100% gb EKJ82697.1	flagellar biosynthesis protein FlhA [Escherichia coli AD30]
contig-1969000041	56	96% ref YP_006118257.1	phage host specificity protein [Escherichia coli ETEC H10407] emb CBJ04229.1 phage host specificity protein [Escherichia coli ETEC H10407]
contig-1974000018	46	100% gb EHN96646.1	inner membrane protein ycfT [Escherichia coli E101]
contig-1983000057	85	71% gb EKI54485.1	putative membrane protein YbiO [Escherichia coli N1]
contig-1989000016	64	100% ref YP_006781880.1	hypothetical protein O3O_25250 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006776808.1 hypothetical protein O3K_00360 [Escherichia coli O104:H4 str. 2011C-3493] gb EGT69163.1 hypothetical protein C22711_3193 [Escherichia coli O104:H4 str. C2 ...
contig-1992000029	135	91% gb EJE78650.1	putative fimbrial outer membrane usher protein StfC [Escherichia coli O26:H11 str. CVM10021]
contig-1994000018	45	100% gb EKJ84528.1	DNA topoisomerase [Escherichia coli AD30]
contig-1995000054	38	100% ref YP_006170275.1	Tryptophan-specific transport protein [Escherichia coli P12b] gb AFG42105.1 Tryptophan-specific transport protein [Escherichia coli P12b]
contig-20000040	539	95% ref YP_002405757.1	hypothetical protein EC55989_4900 [Escherichia coli 55989] ref ZP_04534429.1 conserved hypothetical protein [Escherichia sp. 3_2_53FAA] ref YP_003352149.1 hypothetical protein ECSF_4159 [Escherichia coli SE15] ref YP_006786238.1 hypothetical prote ...

contig-20000040	444	100% ref YP_002405758.1	hypothetical protein EC55989_4901 [Escherichia coli 55989] emb CAV01946.1 conserved hypothetical protein [Escherichia coli 55989]
contig-20000040	290	100% ref YP_002405760.1	hypothetical protein EC55989_4903 [Escherichia coli 55989] ref YP_006786235.1 hypothetical protein O3O_03125 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006781111.1 hypothetical protein O3K_22270 [Escherichia coli O104:H4 str. 2011C-3493] em ...
contig-20000040	263	100% gb EHV41832.1	phage integrase family protein [Escherichia coli DEC5E]
contig-20000052	431	97% ref YP_002403324.1	Wzx [Escherichia coli 55989] emb CAU98163.1 Wzx [Escherichia coli 55989]
contig-20000052	288	97% gb EGR63361.1	WbwB [Escherichia coli O104:H4 str. 01-09591] gb EGR74203.1 WbwB [Escherichia coli O104:H4 str. LB226692]
contig-20000052	272	100% ref YP_002403323.1	glycosyl transferase family protein [Escherichia coli 55989] emb CAU98162.1 glycosyl transferase, putative [Escherichia coli 55989]
contig-20000052	92	100% gb EGR63362.1	WbwB [Escherichia coli O104:H4 str. 01-09591]
contig-2019000046	46	100% gb EKJ79875.1	fibronectin type III domain protein [Escherichia coli AD30]
contig-2025000059	45	100% gb EKI51514.1	ATP-dependent RNA helicase srmB [Escherichia coli N1]
contig-2036000060	42	100% gb EIL60426.1	UTP--glucose-1-phosphate uridylyltransferase subunit GalF [Escherichia coli 541-1]
contig-2041000050	38	100% ref ZP_08364030.1	putative membrane protein [Escherichia coli TA143] gb EGI30896.1 putative membrane protein [Escherichia coli TA143]
contig-2045000003	39	100% gb EKJ56868.1	CRISPR-associated helicase Cas3 [Escherichia coli O.1288]
contig-2051000051	62	92% gb EHN81884.1	hypothetical protein ESRG_04199 [Escherichia coli TA124]
contig-2057000041	36	100% gb EHY01711.1	ATPase [Escherichia coli DEC15C]
contig-2060000046	56	100% gb EIL03426.1	putative inner membrane protein [Escherichia coli O103:H25 str. CVM9340]
contig-2061000035	70	100% ref YP_006169146.1	Imidazoleglycerol-phosphate synthase [Escherichia coli P12b] gb AFG40976.1 Imidazoleglycerol-phosphate synthase [Escherichia coli P12b]
contig-2065000026	56	100% gb EIQ65699.1	bacterial regulatory s, gntR family protein [Shigella dysenteriae 225-75]
contig-2068000014	76	100% gb EGT66460.1	hypothetical protein C22711_0487 [Escherichia coli O104:H4 str. C227-11]
contig-2071000054	84	100% gb EJK95538.1	yehO [Escherichia coli STEC_O31]
contig-2077000030	73	100% gb EIQ63871.1	K+-transporting ATPase, A subunit [Shigella dysenteriae 225-75]
contig-2090000053	46	100% ref YP_540521.1	hypothetical protein UTI89_C1512 [Escherichia coli UTI89] gb ABE06990.1 hypothetical protein UTI89_C1512 [Escherichia coli UTI89]
contig-2091000039	40	100% gb EIL68009.1	hypothetical protein EC5411_04394 [Escherichia coli 541-1] gb EKJ56435.1 putative NAD-binding domain 4 [Escherichia coli O.1288]

contig-2094000047	46	100% gb EKJ83389.1	membrane-bound PQQ-dependent dehydrogenase, glucose/quininate/shikimate family [Escherichia coli AD30]
contig-2098000000	49	96% gb EHO03153.1	quinone oxidoreductase 2 [Escherichia coli E101]
contig-2106000051	32	97% ref ZP_09829171.1	transposase [Pantoea stewartii subsp. stewartii DC283] gb EHT99936.1 transposase [Pantoea stewartii subsp. stewartii DC283]
contig-2106000051	26	96% ref ZP_09829171.1	transposase [Pantoea stewartii subsp. stewartii DC283] gb EHT99936.1 transposase [Pantoea stewartii subsp. stewartii DC283]
contig-2120000002	45	100% gb EKJ41949.1	hypothetical protein ECEC1870_3204 [Escherichia coli EC1870]
contig-2123000031	26	100% ref ZP_08373629.1	putative transient receptor potential locus [Escherichia coli TA280] gb EGI41433.1 putative transient receptor potential locus [Escherichia coli TA280]
contig-2128000003	38	100% gb EJZ02255.1	LysR family transcriptional regulator, partial [Escherichia coli J96]
contig-2129000051	96	100% gb EKJ62282.1	rhsG core with extension domain protein [Escherichia coli O.1288]
contig-2138000006	31	100% gb EGB55812.1	hypothetical protein ERGG_03300 [Escherichia coli H489]
contig-2145000005	40	100% gb EHO04180.1	hypothetical protein ESNG_04668 [Escherichia coli B093]
contig-215000032	94	93% ref YP_006784293.1	hypothetical protein O3O_12975 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779222.1 hypothetical protein O3K_12660 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR71359.1 hypothetical protein HUSEC_26012 [Escherichia coli O104:H4 str. LB ...
contig-215000032	86	100% ref YP_006784295.1	phage anti-repressor protein AntB [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779224.1 phage anti-repressor protein AntB [Escherichia coli O104:H4 str. 2011C-3493] gb EGR71361.1 phage anti-repressor protein AntB [Escherichia coli O104:H4 ...
contig-215000032	73	100% ref YP_006784294.1	putative C4-type zinc finger protein [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779223.1 putative C4-type zinc finger protein [Escherichia coli O104:H4 str. 2011C-3493] gb EGR71360.1 putative C4-type zinc finger protein [Escherichia coli ...
contig-2171000002	35	100% gb EKI34165.1	putative glutathionylspermidine synthase [Escherichia coli 07798]
contig-2176000051	40	100% gb EII12460.1	sulfate adenyltransferase, large subunit [Escherichia coli 5.0959]
contig-2189000028	35	100% gb EKA95142.1	3-octaprenyl-4-hydroxybenzoate carboxy-lyase [Proteus mirabilis WGLW4]
contig-2190000061	72	100% gb EHF78759.1	protein rhsD [Escherichia coli O104:H4 str. 11-4632 C5]
contig-2193000021	43	100% gb EKI42401.1	molybdopterin synthase sulfurlyase MoeB [Escherichia coli 3006]
contig-2205000008	42	100% gb EKJ79933.1	plasmid-derived single-stranded DNA-binding protein [Escherichia coli AD30]

contig-2206000058	44	98% gb EHW41149.1	TMAO reductase system periplasmic protein TorT [Escherichia coli DEC9A] gb EHW43094.1 TMAO reductase system periplasmic protein TorT [Escherichia coli DEC9B] gb EHW49691.1 TMAO reductase system periplasmic protein TorT [Escherichia coli DEC9C] gb E ...
contig-2210000015	50	100% gb EKJ60896.1	putative GTP-binding protein [Escherichia coli O.1288]
contig-2216000045	44	100% gb EIO86796.1	non-specific ribonucleoside hydrolase rihC [Escherichia coli TW10119]
contig-2220000036	59	58% gb ADR67068.1	putative transcriptional regulator [Klebsiella pneumoniae subsp. pneumoniae]
contig-2222000016	88	86% ref ZP_09829171.1	transposase [Pantoea stewartii subsp. stewartii DC283] gb EHT99936.1 transposase [Pantoea stewartii subsp. stewartii DC283]
contig-2222000016	78	88% gb EGK29645.1	hypothetical protein SFK218_5541 [Shigella flexneri K-218] gb EGK30349.1 hypothetical protein SFK218_5515 [Shigella flexneri K-218]
contig-2222000040	77	100% gb EII59033.1	ShET2 enterotoxin, N-terminal domain protein [Escherichia coli 3.3884]
contig-2223000036	43	100% ref NP_289883.1	leader peptidase [Escherichia coli O157:H7 str. EDL933] ref NP_312215.1 HopD [Escherichia coli O157:H7 str. Sakai] ref ZP_02778098.1 peptidase, A24 (type IV prepilin peptidase) family [Escherichia coli O157:H7 str. EC4401] ref ZP_02794196.1 peptid ...
contig-2223000043	39	100% gb EHW84621.1	hypothetical protein ECDEC11A_4857 [Escherichia coli DEC11A] gb EHW98295.1 hypothetical protein ECDEC11B_4918 [Escherichia coli DEC11B]
contig-2236000032	52	100% gb EIQ62149.1	crossover junction endodeoxyribonuclease RuvC [Shigella dysenteriae 225-75]
contig-2244000050	35	100% gb EKI45750.1	leucine--tRNA ligase [Escherichia coli 07798]
contig-2246000047	40	65% ref YP_002413044.1	cobalamin synthase [Escherichia coli UMN026] ref ZP_06649457.1 cobS [Escherichia coli FVEC1412] ref ZP_06653920.1 cobalamin 5'-phosphate synthase [Escherichia coli B354] ref ZP_06990752.1 cobalamin synthase [Escherichia coli FVEC1302] sp B7NC26.1 ...
contig-2249000018	114	100% gb EII00399.1	prophage CP4-57 integrase [Escherichia coli 96.154]
contig-2250000024	39	100% gb EKJ83676.1	class II glutamine amidotransferase [Escherichia coli AD30]
contig-2251000056	51	100% gb EJF03470.1	hypothetical protein ECO9952_04361, partial [Escherichia coli O26:H11 str. CVM9952]
contig-2262000013	34	100% gb EKJ82619.1	cytosine deaminase [Escherichia coli AD30]
contig-2268000049	39	97% gb EJZ61809.1	sugar phosphatase [Shigella flexneri 1485-80]
contig-2275000047	34	100% gb EHP66824.1	hypothetical protein HMPREF0986_01153 [Escherichia coli 4_1_47FAA]
contig-2276000040	50	100% gb EII85709.1	IS66 family element, transposase [Escherichia coli 3003]
contig-2289000018	35	100% gb EKI37021.1	type III secretion apparatus lipoprotein, YscJ/HrcJ family [Escherichia coli 3006]
contig-2295000031	34	100% gb EIH13995.1	pertactin [Escherichia coli 97.0259]
contig-2296000039	36	100% gb EJE95541.1	cytoplasmic protein [Escherichia coli O111 str. CVM9455]

contig-2301000020	27	96% ref ZP_03036142.1	hypothetical protein EcF11_4182 [Escherichia coli F11] ref ZP_07778923.1 hypothetical protein EC236275_0447 [Escherichia coli 2362-75] ref YP_006147608.1 hypothetical protein i02_0390 [Escherichia coli str. 'clone D i2'] ref YP_006152527.1 hypothe ...
contig-2301000020	16	88% ref ZP_03036142.1	hypothetical protein EcF11_4182 [Escherichia coli F11] ref ZP_07778923.1 hypothetical protein EC236275_0447 [Escherichia coli 2362-75] ref YP_006147608.1 hypothetical protein i02_0390 [Escherichia coli str. 'clone D i2'] ref YP_006152527.1 hypothe ...
contig-2301000051	34	100% gb EKI50607.1	putative aspartate/ornithine carbamoyltransferase [Escherichia coli N1]
contig-2314000035	57	91% ref ZP_07133507.1	conserved hypothetical protein [Escherichia coli MS 115-1] gb EFJ99217.1 conserved hypothetical protein [Escherichia coli MS 115-1]
contig-2320000007	39	100% ref ZP_07784541.1	acetyltransferase family protein [Escherichia coli 1827-70] ref YP_006132084.1 hypothetical protein UMNK88_268 [Escherichia coli UMNK88] ref ZP_10978825.1 hypothetical protein ESCG_02290 [Escherichia sp. 1_1_43] gb EFQ02147.1 acetyltransferase fam ...
contig-2320000059	34	100% gb EKH51047.1	putative deacetylase [Escherichia coli FRIK2001]
contig-2323000028	24	100% gb EKI44145.1	hydrogenase expression/formation protein [Escherichia coli 07798]
contig-2323000028	20	100% gb EIQ29461.1	hydrogenase-1 operon protein hyaE [Shigella boydii 965-58]
contig-2328000012	39	100% gb EJO25644.1	hypothetical protein MU9_1864 [Morganella morganii subsp. morganii KT]
contig-2328000039	36	100% gb EKJ82962.1	RND transporter, hydrophobe/amphiphile efflux-1 family protein [Escherichia coli AD30]
contig-2330000045	320	100% ref ZP_02796956.2	ribosome inactivating protein [Escherichia coli O157:H7 str. EC4486] ref ZP_02815427.2 shiga toxin subunit A [Escherichia coli O157:H7 str. EC869] gb EDU77859.1 ribosome inactivating protein [Escherichia coli O157:H7 str. EC4486] gb EDU88534.1 shi ...
contig-2330000045	89	100% ref NP_049501.1	Shiga toxin 2 subunit B [Enterobacteria phage 933W] ref NP_050540.1 Shiga toxin 2 subunit B [Enterobacteria phage VT2-Sakai] ref NP_286977.1 shiga-like toxin II B subunit encoded by bacteriophage BP-933W [Escherichia coli O157:H7 str. EDL933] ref N ...
contig-2330000045	103	85% gb EKI50736.1	hypothetical protein ECEC1735_3442, partial [Escherichia coli EC1735] gb EKI65069.1 hypothetical protein ECEC1737_3361, partial [Escherichia coli EC1737]
contig-2334000050	40	100% gb EKI39064.1	inner membrane amino-acid ABC transporter permease protein yecS [Escherichia coli 3006]
contig-2338000056	44	100% gb EKI49729.1	putative mutase [Escherichia coli N1]
contig-2341000043	33	100% gb EIN79705.1	hypothetical protein ECPA10_1571 [Escherichia coli PA10] gb EIP61648.1 hypothetical protein ECEC1738_1580 [Escherichia coli EC1738]
contig-2341000059	34	100% gb EIQ74759.1	6-phosphogluconate dehydrogenase, C-terminal domain protein [Shigella flexneri 1235-66]
contig-2342000002	36	100% gb EKJ81118.1	ATP-dependent DNA helicase RecG [Escherichia coli AD30]

contig-2344000023	38	100% ref ZP_07134673.1	FeS assembly ATPase SufC [Escherichia coli MS 115-1] ref ZP_08354087.1 FeS assembly ATPase SufC [Escherichia coli M718] gb EFJ98053.1 FeS assembly ATPase SufC [Escherichia coli MS 115-1] gb EGI21252.1 FeS assembly ATPase SufC [Escherichia coli M71 ...
contig-2344000038	77	100% gb EIP66492.1	DDG protein [Escherichia coli EC1734] gb EKI46805.1 DDG protein [Escherichia coli EC1735] gb EKI59475.1 DDG protein [Escherichia coli EC1737] gb EKI60992.1 DDG protein [Escherichia coli EC1736]
contig-2344000038	64	97% gb EHG00200.1	aminotransferase [Escherichia coli cloneA_i1]
contig-2350000061	39	100% gb EKI44103.1	3-oxoacyl-[acyl-carrier-protein] synthase 3 [Escherichia coli 07798]
contig-2353000035	49	100% gb EII67573.1	SWIM zinc finger domain protein [Escherichia coli 2.4168]
contig-2357000018	55	100% gb EKI12534.1	lysozyme, partial [Escherichia coli 5412]
contig-2357000023	34	68% gb EKJ83396.1	hypothetical protein ECAD30_13020 [Escherichia coli AD30]
contig-2364000017	36	100% gb EKJ82646.1	choline dehydrogenase [Escherichia coli AD30]
contig-2365000003	38	100% ref YP_006784581.1	tail fiber component K of prophage [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779506.1 tail fiber component K of prophage [Escherichia coli O104:H4 str. 2011C-3493] gb AFS57495.1 tail fiber component K of prophage [Escherichia coli O104: ...
contig-2375000051	53	100% gb EKI35740.1	gamma-glutamyltransferase [Escherichia coli 07798]
contig-2375000054	45	76% gb EKI54449.1	23S rRNA (uracil-5-)-methyltransferase RumB [Escherichia coli N1]
contig-2376000005	35	100% gb EIL70054.1	imidazole glycerol phosphate synthase subunit HisF [Escherichia coli 75]
contig-2383000010	40	100% gb EKI55989.1	carbamoyl-phosphate synthase large chain [Escherichia coli N1]
contig-2387000031	24	96% ref ZP_06660436.1	yjiX protein [Escherichia coli B185] gb EFF03530.1 yjiX protein [Escherichia coli B185] gb EII94684.1 PF04328 family protein [Escherichia coli TW07793]
contig-2387000031	16	100% gb EIQ65626.1	inner membrane protein YjiY [Shigella dysenteriae 225-75]
contig-2393000017	37	100% gb EKI50029.1	glutamate synthase [NADPH] large chain [Escherichia coli N1]
contig-2395000059	33	100% gb EIG82285.1	sensor kinase protein RcsC [Escherichia coli 1.2741]
contig-2397000001	45	71% ref YP_002402175.1	hydrogenase 1 maturation protease [Escherichia coli 55989] ref YP_006785108.1 hydrogenase 1 maturation protease [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779968.1 hydrogenase 1 maturation protease [Escherichia coli O104:H4 str. 2011C-34 ...
contig-2397000004	79	92% gb EIG83332.1	integrase core domain protein [Escherichia coli 1.2741]
contig-2397000004	29	100% gb EIG83332.1	integrase core domain protein [Escherichia coli 1.2741]
contig-2398000030	45	100% gb EKJ83445.1	thiamine ABC transporter, inner membrane subunit [Escherichia coli AD30]
contig-2400000035	38	100% gb EKI33078.1	cation/acetate symporter ActP [Escherichia coli 3006]

contig-2401000052	35	91% gb EGW69218.1	hypothetical protein ECSTECC16502_2793 [Escherichia coli STEC_C165-02] gb EIH13910.1 hypothetical protein EC990741_2332 [Escherichia coli 97.0259] gb EIQ07571.1 hypothetical protein SF285071_2660 [Shigella flexneri 2850-71]
contig-2403000057	61	100% gb EIL03175.1	hypothetical protein ECO9340_22200 [Escherichia coli O103:H25 str. CVM9340]
contig-2404000041	67	100% gb EIP14840.1	protein involved in detoxification of methylglyoxal [Escherichia coli TW14301]
contig-2414000008	45	100% gb EJL19045.1	potG [Shigella sonnei str. Moseley]
contig-2415000051	35	100% gb EKI46212.1	inner membrane protein CreD [Escherichia coli 07798]
contig-2423000002	24	100% gb EKI50328.1	glutamate-ammonia-ligase adenylyltransferase [Escherichia coli N1]
contig-2423000042	41	100% gb EKJ56227.1	obg family GTPase CgtA [Escherichia coli 0.1288]
contig-2427000018	44	100% gb EIL03851.1	alcohol dehydrogenase [Escherichia coli O103:H25 str. CVM9340]
contig-2437000016	53	100% gb EKH07481.1	DNA-directed RNA polymerase, beta subunit [Escherichia coli PA7]
contig-2441000033	48	100% gb EKJ61928.1	rhs domain-containing protein, partial [Escherichia coli 0.1288]
contig-2443000014	54	100% gb EKJ80457.1	hypothetical protein ECAD30_43620 [Escherichia coli AD30]
contig-2443000017	55	100% gb AFS57512.1	Terminase large subunit (Gp2) [Escherichia coli O104:H4 str. 2009EL-2050]
contig-2443000045	65	100% gb EKH63341.1	putative L-ascorbate-6-phosphate lactonase ulaG [Escherichia coli FRIK2001]
contig-2445000022	40	100% ref YP_006144489.1	hypothetical protein CE10_2431 [Escherichia coli O7:K1 str. CE10] gb AEQ13205.1 conserved protein [Escherichia coli O7:K1 str. CE10]
contig-2452000035	27	100% ref ZP_07143859.1	conserved domain protein [Escherichia coli MS 187-1] gb EFK27149.1 conserved domain protein [Escherichia coli MS 187-1]
contig-2460000015	44	100% gb EIQ30700.1	transaldolase, putative [Shigella boydii 4444-74]
contig-2464000018	54	100% gb EKI45471.1	hypothetical protein EC07798_0467 [Escherichia coli 07798]
contig-2465000041	43	100% gb EKJ80184.1	alpha/beta hydrolase fold family protein [Escherichia coli AD30]
contig-2468000002	40	100% gb EKI48355.1	transcriptional regulator, AraC family [Escherichia coli N1]
contig-2479000016	31	61% gb EKJ80347.1	lysozyme [Escherichia coli AD30]
contig-2479000032	45	100% gb EIQ53180.1	hypothetical protein SF123566_7261, partial [Shigella flexneri 1235-66]
contig-248000019	363	100% gb AFS57491.1	putative phage tail fiber protein [Escherichia coli O104:H4 str. 2009EL-2050]
contig-2495000041	61	100% gb EJE71254.1	phosphonate metabolism protein PhnM [Escherichia coli O111:H8 str. CVM9634]
contig-2499000055	44	100% gb EJY97867.1	L-asparaginase [Escherichia coli J96]
contig-25000057	444	99% ref NP_049461.1	integrase [Enterobacteria phage 933W] ref NP_050500.1 integrase [Enterobacteria phage VT2- Sakai] ref NP_286942.1 integrase for bacteriophage BP-933W [Escherichia coli O157:H7 str. EDL933] ref NP_612960.1 hypothetical protein Stx2lp082 [Stx2 conver ...
contig-25000057	69	51% ref ZP_03357401.1	integrase [Salmonella enterica subsp. enterica serovar Typhi str. E02-1180]

contig-25000057	208	100% ref YP_006779933.1	adenine methylase [Escherichia coli O104:H4 str. 2011C-3493] gb EGR60035.1 adenine methylase [Escherichia coli O104:H4 str. 01-09591] gb EGR70902.1 adenine methylase [Escherichia coli O104:H4 str. LB226692] gb EGT68114.1 hypothetical protein C2271 ...
contig-25000057	212	93% ref YP_006785007.1	hypothetical protein O3O_09330 [Escherichia coli O104:H4 str. 2009EL-2071] gb AFS57857.1 hypothetical protein O3M_15945 [Escherichia coli O104:H4 str. 2009EL-2050] gb AFS85633.1 hypothetical protein O3O_09330 [Escherichia coli O104:H4 str. 2009EL-2 ...
contig-2519000015	53	100% gb EKH07481.1	DNA-directed RNA polymerase, beta subunit [Escherichia coli PA7]
contig-300000054	140	100% ref YP_006784295.1	phage anti-repressor protein AntB [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779224.1 phage anti-repressor protein AntB [Escherichia coli O104:H4 str. 2011C-3493] gb EGR71361.1 phage anti-repressor protein AntB [Escherichia coli O104:H4 ...
contig-307000038	208	100% gb EHF34548.1	hypothetical protein EUEG_02031, partial [Escherichia coli O104:H4 str. 09-7901] gb EHF79072.1 hypothetical protein EUMG_04238, partial [Escherichia coli O104:H4 str. 11-4632 C3]
contig-309000007	187	93% ref ZP_06656337.1	ampG [Escherichia coli B185] gb EFF06719.1 ampG [Escherichia coli B185]
contig-319000009	170	65% gb EHF41436.1	hypothetical protein EUFG_02075 [Escherichia coli O104:H4 str. 11-3677]
contig-337000048	244	100% ref NP_049526.1	outer membrane protein Lom precursor [Enterobacteria phage 933W] ref NP_050564.1 similar to Lom protein of bacteriophage lambda [Enterobacteria phage VT2-Sakai] ref NP_286999.1 outer membrane protein Lom of bacteriophage BP-933W [Escherichia coli O ...
contig-337000048	90	100% gb EIO16281.1	hypothetical protein ECPA33_3401 [Escherichia coli PA33]
contig-349000007	308	97% ref ZP_03034725.1	hypothetical protein EcF11_4284 [Escherichia coli F11] ref ZP_03047733.1 hypothetical protein EcE22_3634 [Escherichia coli E22] ref ZP_07180024.1 hypothetical protein HMPREF9553_04180 [Escherichia coli MS 200-1] ref YP_006778775.1 hypothetical pro ...
contig-349000007	74	93% gb EIL64763.1	hypothetical protein EC75_12630, partial [Escherichia coli 75]
contig-349000007	45	96% gb EHW49393.1	hypothetical protein ECDEC9C_1319 [Escherichia coli DEC9C] gb EHW94418.1 hypothetical protein ECDEC11A_1543 [Escherichia coli DEC11A] gb EHX04853.1 hypothetical protein ECDEC11B_1292 [Escherichia coli DEC11B]
contig-349000007	29	100% gb EGT69841.1	hypothetical protein C22711_3873 [Escherichia coli O104:H4 str. C227-11]
contig-371000046	41	100% gb EIL44421.1	Prophage integrase [Escherichia coli KD1]
contig-373000052	181	99% gb EIG83186.1	integrase core domain protein [Escherichia coli 1.2741]

contig-42000012	188	93% ref ZP_10056473.1	hypothetical protein ESGB_03541 [Escherichia sp. 4_1_40B] ref YP_006098852.1 hypothetical protein EC042_4581 [Escherichia coli O42] ref YP_006782680.1 hypothetical protein O3O_21195 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006777616.1 hy ...
contig-42000012	51	98% ref YP_006098851.1	hypothetical protein EC042_4580 [Escherichia coli O42] emb CBG37403.1 conserved hypothetical protein [Escherichia coli O42]
contig-42000012	60	100% gb EGR60054.1	Protein impB (fragment) [Escherichia coli O104:H4 str. 01-09591]
contig-42000012	49	100% ref ZP_10056471.1	hypothetical protein ESGB_05012 [Escherichia sp. 4_1_40B] gb EGR60053.1 hypothetical protein HUSEC41_27543 [Escherichia coli O104:H4 str. 01-09591] gb EGR71391.1 hypothetical protein HUSEC_25867 [Escherichia coli O104:H4 str. LB226692] gb EHF20714. ...
contig-42000012	14	93% gb EHX60415.1	hypothetical protein ECDEC13D_3775 [Escherichia coli DEC13D]
contig-42000012	32	100% ref YP_006098850.1	hypothetical protein EC042_4579B, partial [Escherichia coli O42] emb CBG37402.1 conserved hypothetical protein [Escherichia coli O42]
contig-42000012	126	37% gb EKI41416.1	hypothetical protein EC3006_1159 [Escherichia coli 3006]
contig-42000012	28	96% ref ZP_07105485.1	hypothetical protein HMPREF9346_05260 [Escherichia coli MS 119-7] gb EFK43190.1 hypothetical protein HMPREF9346_05260 [Escherichia coli MS 119-7]
contig-42000012	28	64% gb EKJ82887.1	hypothetical protein ECAD30_21150 [Escherichia coli AD30]
contig-42000012	35	51% ref YP_003739814.1	ImpA domain-containing protein [Erwinia billingiae Eb661] emb CAX57954.1 ImpA domain protein [Erwinia billingiae Eb661]
contig-42000012	19	53% ref YP_629028.1	hypothetical protein MXAN_0762 [Myxococcus xanthus DK 1622] gb ABF90984.1 conserved hypothetical protein [Myxococcus xanthus DK 1622]
contig-426000033	513	100% gb EHF62316.1	serine protease sepA autotransporter [Escherichia coli O104:H4 str. 11-4632 C3]
contig-427000040	67	100% ref YP_002404314.1	hypothetical protein EC55989_3350 [Escherichia coli 55989] emb CAU99407.1 hypothetical protein EC55989_3350 [Escherichia coli 55989] gb EGR61604.1 hypothetical protein HUSEC41_26405 [Escherichia coli O104:H4 str. 01-09591] gb EGR73001.1 hypotheric ...
contig-427000040	40	80% gb EKI25667.1	hypothetical protein ECTW00353_3338 [Escherichia coli TW00353]
contig-427000040	32	84% ref ZP_06988343.1	predicted protein [Escherichia coli FVEC1302] gb EFI22294.1 predicted protein [Escherichia coli FVEC1302]
contig-44000002	161	92% ref YP_006784248.1	bacteriophage protein [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779177.1 bacteriophage protein [Escherichia coli O104:H4 str. 2011C-3493] gb EGR60064.1 bacteriophage protein [Escherichia coli O104:H4 str. 01-09591] gb EGR71374.1 bacter ...

contig-44000002	105	100% ref YP_006784249.1	DNA-binding protein [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779178.1 DNA-binding protein [Escherichia coli O104:H4 str. 2011C-3493] gb EGR71375.1 putative DNA-binding protein [Escherichia coli O104:H4 str. LB226692] gb EHF25947.1 hyp ...
contig-44000002	100	100% ref YP_006784251.1	putative antirepressor protein Ant from prophage [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779180.1 putative antirepressor protein Ant from prophage [Escherichia coli O104:H4 str. 2011C-3493] gb EGR71377.1 putative antirepressor protein ...
contig-44000002	84	100% ref YP_006784250.1	hypothetical protein O3O_13190 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779179.1 hypothetical protein O3K_12445 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR71376.1 hypothetical protein HUSEC_25942 [Escherichia coli O104:H4 str. LB ...
contig-44000002	32	53% ref YP_403979.1	antirepressor protein [Shigella dysenteriae Sd197] gb ABB62488.1 putative antirepressor protein [Shigella dysenteriae Sd197]
contig-44000002	59	100% ref YP_006784247.1	bacteriophage protein [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779176.1 bacteriophage protein [Escherichia coli O104:H4 str. 2011C-3493] gb EGE64987.1 hypothetical protein ECSTEC7V_1837 [Escherichia coli STEC_7v] gb EGR60065.1 bacteri ...
contig-44000002	24	67% ref NP_929390.1	hypothetical protein plu2130 [Photorhabdus luminescens subsp. laumondii TTO1] emb CAE14423.1 unnamed protein product [Photorhabdus luminescens subsp. laumondii TTO1]
contig-445000011	329	69% ref YP_852419.1	hypothetical protein APECO1_406 [Escherichia coli APEC O1] gb ABJ00705.1 conserved hypothetical protein [Escherichia coli APEC O1]
contig-448000017	183	100% ref YP_002405616.1	Inner membrane protein yagU [Escherichia coli 55989] ref YP_006786312.1 Inner membrane protein yagU [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006781185.1 Inner membrane protein yagU [Escherichia coli O104:H4 str. 2011C-3493] emb CAV01695.1 ...
contig-464000031	72	81% ref NP_612923.1	hypothetical protein Stx2lp044 [Stx2 converting phage I] dbj BAB87892.1 hypothetical protein [Stx2 converting phage I] dbj BAC77860.1 hypothetical protein [Stx1 converting phage] dbj BAC78026.1 hypothetical protein [Stx2 converting phage II]
contig-464000031	77	53% ref NP_612926.1	hypothetical protein Stx2lp047 [Stx2 converting phage I] dbj BAB87895.1 hypothetical protein [Stx2 converting phage I] dbj BAC77863.1 hypothetical protein [Stx1 converting phage] dbj BAC78029.1 hypothetical protein [Stx2 converting phage II]

contig-47000004	358	100% ref YP_002405769.1	Filamentation induced by cAMP protein Fic [Escherichia coli 55989] ref ZP_04534441.1 filamentation induced by cAMP protein Fic [Escherichia sp. 3_2_53FAA] ref YP_003352160.1 hypothetical protein ECSF_4170 [Escherichia coli SE15] ref YP_006786230.1 ...
contig-47000004	209	100% ref YP_002405770.1	hypothetical protein EC55989_4913 [Escherichia coli 55989] ref YP_003352161.1 hypothetical protein ECSF_4171 [Escherichia coli SE15] ref YP_006786229.1 hypothetical protein O3O_03155 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006781105.1 h ...
contig-50000005	422	96% ref YP_001648957.1	hypothetical protein [Enterobacteria phage Min27] gb ABY49919.1 hypothetical protein [Enterobacteria phage Min27]
contig-50000005	13	100% ref YP_001648957.1	hypothetical protein [Enterobacteria phage Min27] gb ABY49919.1 hypothetical protein [Enterobacteria phage Min27]
contig-50000005	355	100% gb EHF25244.1	hypothetical protein EUBG_01391 [Escherichia coli O104:H4 str. C236-11]
contig-515000027	153	100% ref YP_006784774.1	hypothetical protein O3O_10525 [Escherichia coli O104:H4 str. 2009EL-2071] gb AFS85866.1 hypothetical protein O3O_10525 [Escherichia coli O104:H4 str. 2009EL-2071]
contig-518000050	136	100% ref YP_006784909.1	hypothetical protein O3O_09845 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779768.1 hypothetical protein O3K_15445 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR70780.1 hypothetical protein HUSEC_29199 [Escherichia coli O104:H4 str. LB ...
contig-518000050	46	100% gb EIN76862.1	hypothetical protein ECPA14_3000, partial [Escherichia coli PA14]
contig-523000042	321	100% gb EHF64872.1	hypothetical protein EUMG_01741 [Escherichia coli O104:H4 str. 11-4632 C3]
contig-557000031	146	100% emb CCG06222.1	hypothetical protein [Escherichia phage P13374]
contig-56000014	217	100% ref NP_049517.1	hypothetical protein 933Wp57 [Enterobacteria phage 933W] ref NP_050555.1 hypothetical protein VT2-Sap58 [Enterobacteria phage VT2-Sakai] ref NP_612894.1 hypothetical protein Stx2lp015 [Stx2 converting phage I] gb AAD25462.1 AF125520_57 hypothetical ...
contig-56000014	206	100% ref YP_006779882.1	hypothetical protein O3K_16040 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR70819.1 hypothetical protein HUSEC_28934 [Escherichia coli O104:H4 str. LB226692] gb EGT68068.1 hypothetical protein C22711_2097 [Escherichia coli O104:H4 str. C227-11] ...
contig-56000014	112	100% gb EKJ51226.1	putative tail fiber protein, partial [Escherichia coli FRIK523]
contig-56000014	90	97% ref NP_612892.1	hypothetical protein Stx2lp013 [Stx2 converting phage I] dbj BAB87861.1 hypothetical protein [Stx2 converting phage I] dbj BAC77829.1 hypothetical protein [Stx1 converting phage] dbj BAC77995.1 hypothetical protein [Stx2 converting phage II] gb AD ...
contig-613000045	122	87% gb EJE84850.1	major tail fiber protein, partial [Escherichia coli O111 str. CVM9455]
contig-624000011	155	80% gb EHY21132.1	rhsC protein in rhs element [Escherichia coli DEC15E]

contig-677000031	72	58% gb EGX14756.1	transposase family protein [Escherichia coli STEC_H.1.8]
contig-677000031	27	89% gb EGX14756.1	transposase family protein [Escherichia coli STEC_H.1.8]
contig-68000018	205	100% ref NP_049525.1	hypothetical protein 933Wp65 [Enterobacteria phage 933W] ref NP_050563.1 hypothetical protein VT2-Sap66 [Enterobacteria phage VT2-Sakai] ref NP_286998.1 hypothetical protein Z1488 [Escherichia coli O157:H7 str. EDL933] ref NP_309262.1 hypothetical ...
contig-68000018	192	100% gb EGW70921.1	hypothetical protein EC253486_3259 [Escherichia coli 2534-86] gb EHU75491.1 putative outer membrane protein Lom [Escherichia coli DEC3C]
contig-68000018	195	100% ref YP_001648957.1	hypothetical protein [Enterobacteria phage Min27] gb ABY49919.1 hypothetical protein [Enterobacteria phage Min27]
contig-68000018	13	100% ref YP_001648957.1	hypothetical protein [Enterobacteria phage Min27] gb ABY49919.1 hypothetical protein [Enterobacteria phage Min27]
contig-68000018	75	99% ref NP_049524.1	hypothetical protein 933Wp64 [Enterobacteria phage 933W] ref NP_050562.1 hypothetical protein VT2-Sap65 [Enterobacteria phage VT2-Sakai] ref NP_612914.1 hypothetical protein Stx2lp035 [Stx2 converting phage I] gb AAD25469.1 AF125520_64 hypothetical ...
contig-711000016	60	93% ref YP_006097599.1	hypothetical protein EC042_3308 [Escherichia coli 042] emb CBG36136.1 hypothetical protein EC042_3308 [Escherichia coli 042]
contig-711000016	45	98% ref ZP_07098992.1	hypothetical protein HMPREF9345_03879 [Escherichia coli MS 107-1] ref ZP_07139707.1 hypothetical protein HMPREF9548_01872 [Escherichia coli MS 182-1] ref ZP_07688012.1 hypothetical protein HMPREF9348_00745 [Escherichia coli MS 145-7] gb EFK03358.1 ...
contig-716000011	100	89% gb EHV62735.1	hypothetical protein ECDEC6C_1800 [Escherichia coli DEC6C]
contig-72000024	164	87% gb EHF25242.1	hypothetical protein EUBG_01389 [Escherichia coli O104:H4 str. C236-11] gb EHF25923.1 hypothetical protein EUBG_02070 [Escherichia coli O104:H4 str. C236-11] gb EHF30902.1 hypothetical protein EUAG_04527 [Escherichia coli O104:H4 str. C227-11] gb E ...
contig-72000024	128	100% gb EHF57098.1	hypothetical protein EUKG_02053, partial [Escherichia coli O104:H4 str. 11-4632 C1]
contig-72000024	89	100% ref NP_612910.1	hypothetical protein Stx2lp031 [Stx2 converting phage I] dbj BAB87879.1 hypothetical protein [Stx2 converting phage I] dbj BAC77847.1 hypothetical protein [Stx1 converting phage] dbj BAC78013.1 hypothetical protein [Stx2 converting phage II]
contig-72000024	75	100% ref YP_006784276.1	hypothetical protein O3O_13060 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779205.1 hypothetical protein O3K_12575 [Escherichia coli O104:H4 str. 2011C-3493] gb EGR60406.1 hypothetical protein HUSEC41_25702 [Escherichia coli O104:H4 str. ...

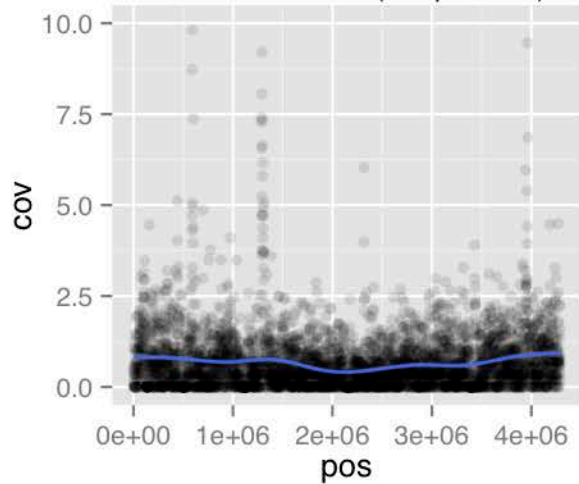
contig-72000024	54	87% ref YP_794071.1	hypothetical protein Stx2-86_gp24 [Stx2-converting phage 86] dbj BAF34056.1 hypothetical protein [Stx2-converting phage 86]
contig-755000031	101	100% gb EII87680.1	aromatic ring-opening dioxygenase, catalytic subunit LigB [Escherichia coli 3003] gb EKI24512.1 hypothetical protein ECARS42123_3543 [Escherichia coli ARS4.2123]
contig-769000049	128	100% gb EHF41264.1	hypothetical protein EUIG_02066, partial [Escherichia coli O104:H4 str. 11-4522] gb EHF64475.1 hypothetical protein EUMG_02060, partial [Escherichia coli O104:H4 str. 11-4632 C3]
contig-770000046	269	96% ref YP_006792611.1	protease IgA1 [Escherichia coli O104:H4 str. 2011C-3493] gb EHF39393.1 serine protease sepA autotransporter [Escherichia coli O104:H4 str. 11-4522] gb AFS59871.1 protease IgA1 [Escherichia coli O104:H4 str. 2009EL-2050] gb AFS77072.1 protease IgA1 ...
contig-776000012	41	100% gb EGX21436.1	hemolysin expression-modulating protein [Escherichia coli STEC_S1191]
contig-776000012	32	91% gb EHV87409.1	hypothetical protein ECDEC7D_4501 [Escherichia coli DEC7D]
contig-777000029	117	100% gb EII00399.1	prophage CP4-57 integrase [Escherichia coli 96.154]
contig-782000002	68	100% gb EIL09704.1	hypothetical protein ECO9340_18308 [Escherichia coli O103:H25 str. CVM9340]
contig-782000002	25	100% ref ZP_03062642.1	conserved hypothetical protein [Escherichia coli B171] ref ZP_07780755.1 putative membrane protein [Escherichia coli 2362-75] ref YP_006138346.1 hypothetical protein ECNA114_1248 [Escherichia coli NA114] gb EDX28124.1 conserved hypothetical protei ...
contig-789000028	43	95% ref ZP_04535728.1	LOW QUALITY PROTEIN: conserved hypothetical protein [Escherichia sp. 3_2_53FAA] gb EEH86263.1 LOW QUALITY PROTEIN: conserved hypothetical protein [Escherichia sp. 3_2_53FAA]
contig-789000028	20	95% gb EIN41723.1	putative DNA methylase, partial [Escherichia coli FRIK1990]
contig-793000045	102	100% gb EHV72047.1	hypothetical protein ECDEC6D_1850 [Escherichia coli DEC6D] gb EHV74688.1 hypothetical protein ECDEC6E_1918 [Escherichia coli DEC6E]
contig-812000037	88	100% gb EHF30076.1	hypothetical protein EUAG_04529, partial [Escherichia coli O104:H4 str. C227-11]
contig-818000020	97	100% gb EKH68165.1	hypothetical protein ECPA23_3299 [Escherichia coli PA23]
contig-849000032	43	98% gb EHU73964.1	hypothetical protein ECDEC3D_3077, partial [Escherichia coli DEC3D]
contig-849000032	68	72% gb EHU73964.1	hypothetical protein ECDEC3D_3077, partial [Escherichia coli DEC3D]
contig-878000009	105	88% gb EHV74241.1	hypothetical protein ECDEC7A_4273 [Escherichia coli DEC7A] gb EHV83770.1 hypothetical protein ECDEC7C_4328 [Escherichia coli DEC7C] gb EHV87404.1 hypothetical protein ECDEC7D_4496 [Escherichia coli DEC7D] gb EHV97102.1 hypothetical protein ECDEC7E ...
contig-888000032	90	100% gb EGR70797.1	hypothetical protein HUSEC_29144 [Escherichia coli O104:H4 str. LB226692]
contig-894000042	97	100% gb EGT68557.1	hypothetical protein C22711_2587 [Escherichia coli O104:H4 str. C227-11]

contig-90000008	98	88% ref YP_002405765.1	transcriptional regulator [Escherichia coli 55989] ref ZP_04534439.1 conserved hypothetical protein [Escherichia sp. 3_2_53FAA] ref YP_003352157.1 transcriptional regulator [Escherichia coli SE15] ref YP_006786231.1 transcriptional regulator [Esch ...
contig-90000008	51	92% ref YP_002405766.1	hypothetical protein EC55989_4909 [Escherichia coli 55989] emb CAV01960.1 hypothetical protein EC55989_4909 [Escherichia coli 55989] gb EFU47777.1 conserved domain protein [Escherichia coli MS 110-3] gb EGR60207.1 hypothetical protein HUSEC41_2659 ...
contig-90000008	49	80% ref YP_002405764.1	hypothetical protein EC55989_4907 [Escherichia coli 55989] emb CAV01957.1 hypothetical protein EC55989_4907 [Escherichia coli 55989] gb EGR60205.1 hypothetical protein HUSEC41_26585 [Escherichia coli O104:H4 str. 01-09591] gb EGR71602.1 hypothetic ...
contig-90000008	14	71% ref ZP_06571313.1	hypothetical protein CIT292_09223 [Citrobacter youngae ATCC 29220] gb EFE07339.1 hypothetical protein CIT292_09223 [Citrobacter youngae ATCC 29220]
contig-90000008	41	61% ref YP_002149438.1	hypothetical protein SeAg_B4812 [Salmonella enterica subsp. enterica serovar Agona str. SL483] gb ACH50603.1 hypothetical protein SeAg_B4812 [Salmonella enterica subsp. enterica serovar Agona str. SL483]
contig-94000001	560	86% ref YP_006778803.1	tail length tape measure protein [Escherichia coli O104:H4 str. 2011C-3493] gb EHF30593.1 lambda family phage tail tape measure protein [Escherichia coli O104:H4 str. C227-11] gb EHF80337.1 lambda family phage tail tape measure protein [Escherichia ...
contig-942000021	97	100% gb EHF31141.1	hypothetical protein EUDG_00692 [Escherichia coli O104:H4 str. 04-8351]
contig-948000026	85	100% gb EKJ61932.1	RHS Repeat containing protein, partial [Escherichia coli O.1288]
contig-962000055	37	100% ref YP_006784898.1	hypothetical protein O3O_09900 [Escherichia coli O104:H4 str. 2009EL-2071] ref YP_006779759.1 hypothetical protein O3K_15400 [Escherichia coli O104:H4 str. 2011C-3493] gb EGT67944.1 hypothetical protein C22711_1973 [Escherichia coli O104:H4 str. C2 ...
contig-962000055	41	88% ref ZP_03060247.1	hypothetical protein EcB171_2428 [Escherichia coli B171] ref ZP_03060248.1 hypothetical protein EcB171_2440 [Escherichia coli B171] gb EDX30483.1 hypothetical protein EcB171_2428 [Escherichia coli B171] gb EDX30484.1 hypothetical protein EcB171_24 ...
contig-989000036	39	87% gb EHW67200.1	hypothetical protein ECDEC10C_5975 [Escherichia coli DEC10C]
contig-989000036	35	91% gb EIQ65768.1	hypothetical protein ECEPECA12_1283 [Escherichia coli EPECa12]
contig-991000042	104	100% ref ZP_08380073.1	putative regulator [Escherichia coli H591] ref ZP_08394300.1 conserved hypothetical protein [Shigella sp. D9] gb EGI43900.1 putative regulator [Escherichia coli H591] gb EGJ07585.1 conserved hypothetical protein [Shigella sp. D9]

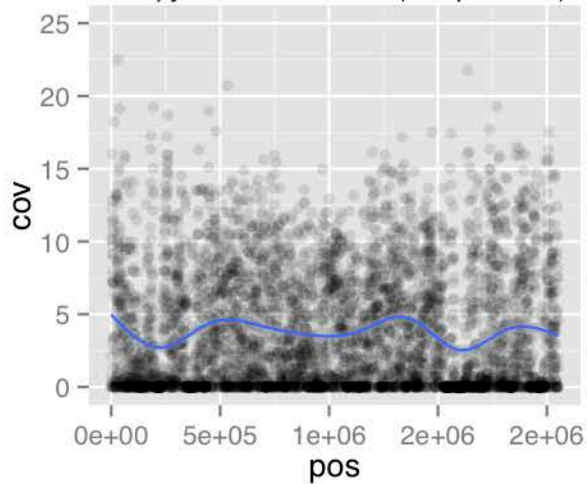
eFigure 1

Genome depth of coverage plots for selected samples. The reference genome indicated by the title was divided into 5000 equal size bins and the maximal coverage (the number of reads mapping to this bin) is plotted. Top-left: Genome coverage for sample 1122 against the *C. difficile* 630 reference genome. Top-middle: Genome coverage for sample 1253 against the *C. concisus* 13826 reference genome. Top-right: Genome coverage for sample 4961 against the *C. jejuni* NCTC11168 genome. Bottom-left: Genome coverage for sample 2535 against the *E. coli* O104:H4 STEC 280 genome, revealing increased Shiga-toxin encoding phage copy number and the presence of a mixed population of *E. coli* O104:H4 is predicted to be present at approximately ~300-fold coverage, with the *E. coli* core genome present at approximately double that number. Bottom-middle: Genome coverage for sample 2669 against the *E. coli* O104:H4 STEC 280 genome, demonstrating 10-fold increase in phage copy number compared to the *E. coli* chromosome. Bottom-right: Genome coverage of sample 3134 against the *E. coli* O104:H4 STEC 280 genome demonstrating increased copy number of a second prophage (I-ROD4) and normal copy number of the Shiga-toxin encoding phage.

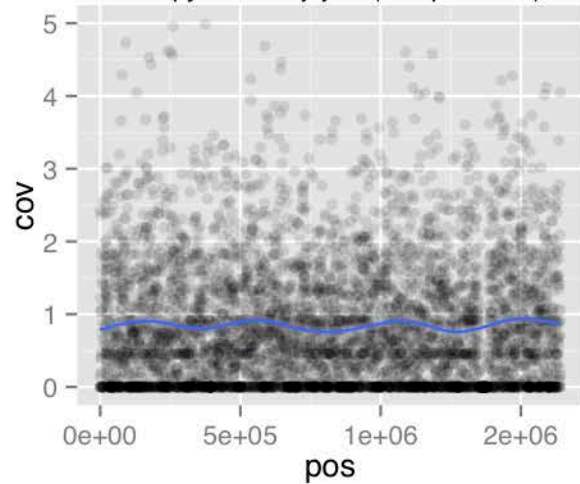
Clostridium difficile (sample 1122)



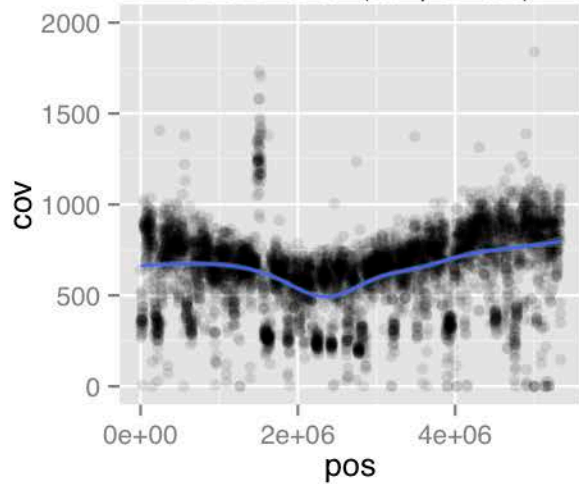
Campylobacter concisus (sample 1253)



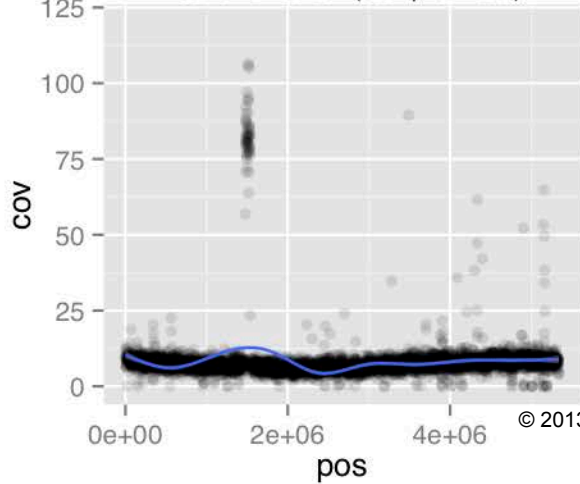
Campylobacter jejuni (sample 4961)



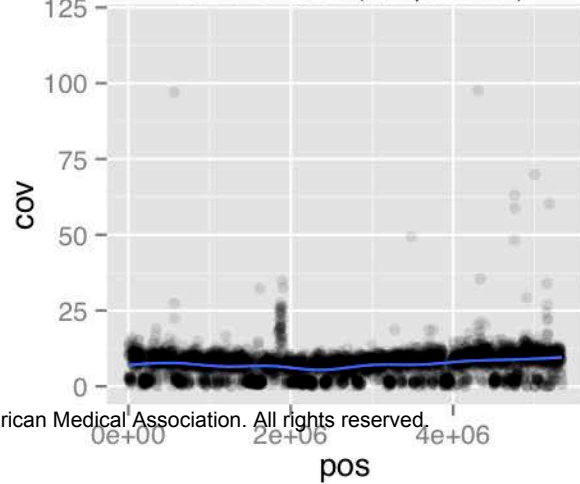
Escherichia coli (sample 2535)



Escherichia coli (sample 2669)

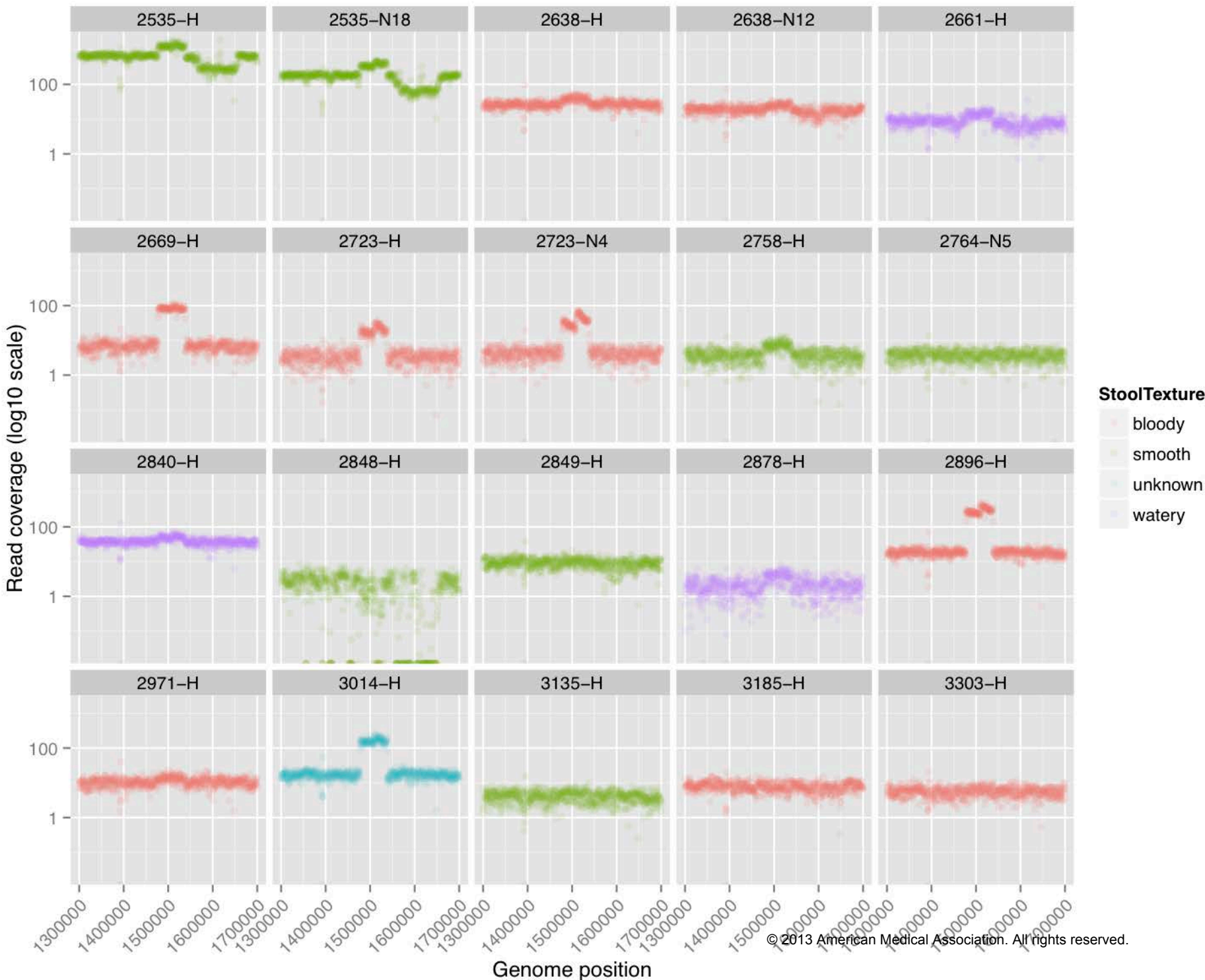


Escherichia coli (sample 3134)



eFigure 2

Genome coverage plots focusing on the region of the *E. coli* O104:H4 STEC 280 genome around the Shiga-toxin encoding phage. The Y-axis is scaled log₁₀. Twenty-one runs are shown for the 17 STEC-positive samples where reads cover at least 90% of the Stx2 subunit A sequence. Additional technical replicates between MiSeq and HiSeq are shown for samples 2535, 2638 and 2723.



eReferences

- ¹ Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. doi:10.1038/nmeth.1923
- ² Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., & Corbeil, J. (2012).
² Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., & Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biology*, 13(12), R122. doi:10.1186/gb-2012-13-12-r122
- ³ Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589–595. doi:10.1093/bioinformatics/btp698
- ⁴ Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59–65. doi:10.1038/nature08821
- ⁵ Rissman, A. I. A., Mau, B. B., Biehl, B. S. B., Darling, A. E. A., Glasner, J. D. J., & Perna, N. T. N. (2009). Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics*, 25(16), 2071–2073. doi:10.1093/bioinformatics/btp356
- ⁶ Ahmed, S. A., Awosika, J., Baldwin, C., Bishop-Lilly, K. A., Biswas, B., Broomall, S., et al. (2012). Genomic Comparison of *Escherichia coli* O104:H4 Isolates from 2009 and 2011 Reveals Plasmid, and Prophage Heterogeneity, Including Shiga Toxin Encoding Phage stx2. *PLoS ONE*, 7(11), e48228–e48228. doi:10.1371/journal.pone.0048228
- ⁷ Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H., & Gentleman, R. (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, 25(19), 2607–2608. doi:10.1093/bioinformatics/btp450
- ⁸ Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, 9(8), 811–814. doi:10.1038/nmeth.2066
- ⁹ Jolley, K. A., & Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, 11(1), 595–595. doi:10.1186/1471-2105-11-595