nature microbiology

SUPPLEMENTARY INFORMATION

ARTICLE NUMBER: 15014 | DOI: 10.1038/NMICROBIOL.2015.14

Genomic diversity of EPEC associated with clinical presentations of differing severity

RESULTS

Distribution of EPEC virulence-associated genes

The T6SS have been previously demonstrated to contribute to the pathogenic mechanism of other enteric bacteria such as *V. cholerae* [1]. The T6SS genes have been detected *in silico* in the genomes of diverse *E. coli* including EPEC [2,3]; however, their role in EPEC virulence is currently undefined. The occurrence of T6SS genes among diverse AEEC genomes, and the presence of two distinct T6SS gene clusters in the genome of the EPEC2 prototype isolate B171 and other genomes of the EPEC2 phylogenomic lineage has recently been described [3,4]. Interestingly, the genomes of the EPEC7 lineage, which is within *E. coli* phylogroup B1 with EPEC2, contain only a single T6SS gene cluster. Additional genomes sequenced in the current study that were not classified into defined lineages but contained genes with similarity to T6SS genes were present only in isolates of *E. coli* phylogroup B1 (Table S1). In contrast, the genomes of the EPEC1, EPEC4, EPEC8, and EPEC9 lineages, which are all present in *E. coli* phylogroup B2, did not contain any T6SS homologs. Overall, the presence or absence of the secretion system genes and known virulence genes follows the whole genome phylogenomic distribution and provides additional clues to the variable virulence gene repertoire likely involved in the pathogenic mechanisms in this group of organisms (Table S1).

The four newly described EPEC lineages in this study exhibited lineage-specific distributions of known virulence-associated genes including the T3SS effectors (Table S1), which was previously described using a more historical collection of EPEC isolates [4]. Meanwhile the T3SS effectors identified in the two EPEC genomes that were present in *E. coli* phylogroup E, and were both *bfp*- (Fig. 1), exhibited greater similarity to effectors of the O157:H7 EHEC isolate Sakai than to the effectors of the prototype EPEC1 and EPEC2 isolates (Fig. S2, Table S1).

EPEC-specific genes associated with different clinical outcomes

Hierarchical cluster analysis of the 246 gene clusters that were more frequently associated with symptomatic or asymptomatic EPEC isolates in the analysis with the decreased threshold and filtering for genes that are present or absent in the commensal isolate placed nearly all of the genomes into two large groups, with the exception of five genomes that were not in any of the EPEC phylogenomic clades (Fig. S5). The larger group designated, group I in Figure S5 contained 48 EPEC isolates, of which 40 (83%) were from symptomatic outcomes (LI or NSI) and only eight (17%) from asymptomatic outcomes. Meanwhile, group II in Figure S5 was comprised of 17 EPEC isolates, with seven (41%) from symptomatic outcomes and 10 (59%) from asymptomatic outcomes (Fig. S5). Within group I, sub-groups of EPEC isolates from the same phylogenomic lineages could be identified; however, group II contained a mixture of EPEC isolates from each of the three *E. coli* phylogroups (A, B1, B2) and multiple EPEC phylogenomic lineages. In contrast, hierarchical cluster analysis of the 141 gene clusters that were significantly associated with symptomatic or asymptomatic genomes when analyzing only the 61 tEPEC isolates at this decreased threshold separated the genomes into two large and similarly sized groups (Fig. S6). In group I there were 34 EPEC isolate genomes consisting of 20 (59%) from symptomatic outcomes, and 14 (41%) from asymptomatic outcomes (Fig. S6). Meanwhile, group II contained 27 EPEC isolate genomes consisting of 24 (89%) from symptomatic outcomes, and three (11%) from asymptomatic outcomes (Fig. S6). Within each of these groups, the genomes formed sub-groups by phylogenomic lineage (Fig. S6). The genomes of group I primarily belonged to *E. coli* phylogroup B2, while the genomes of group II belonged to *E. coli* phylogroups B1 and A (Fig. S6). Overall, these gene-based comparisons demonstrate there is considerable phylogroup- and lineage-specificity of EPEC genomic content, which raises the question of how much variation exists in the survival and virulence mechanisms of these EPEC isolates.

DOI: 10.1038/NMICROBIOL.2015.14

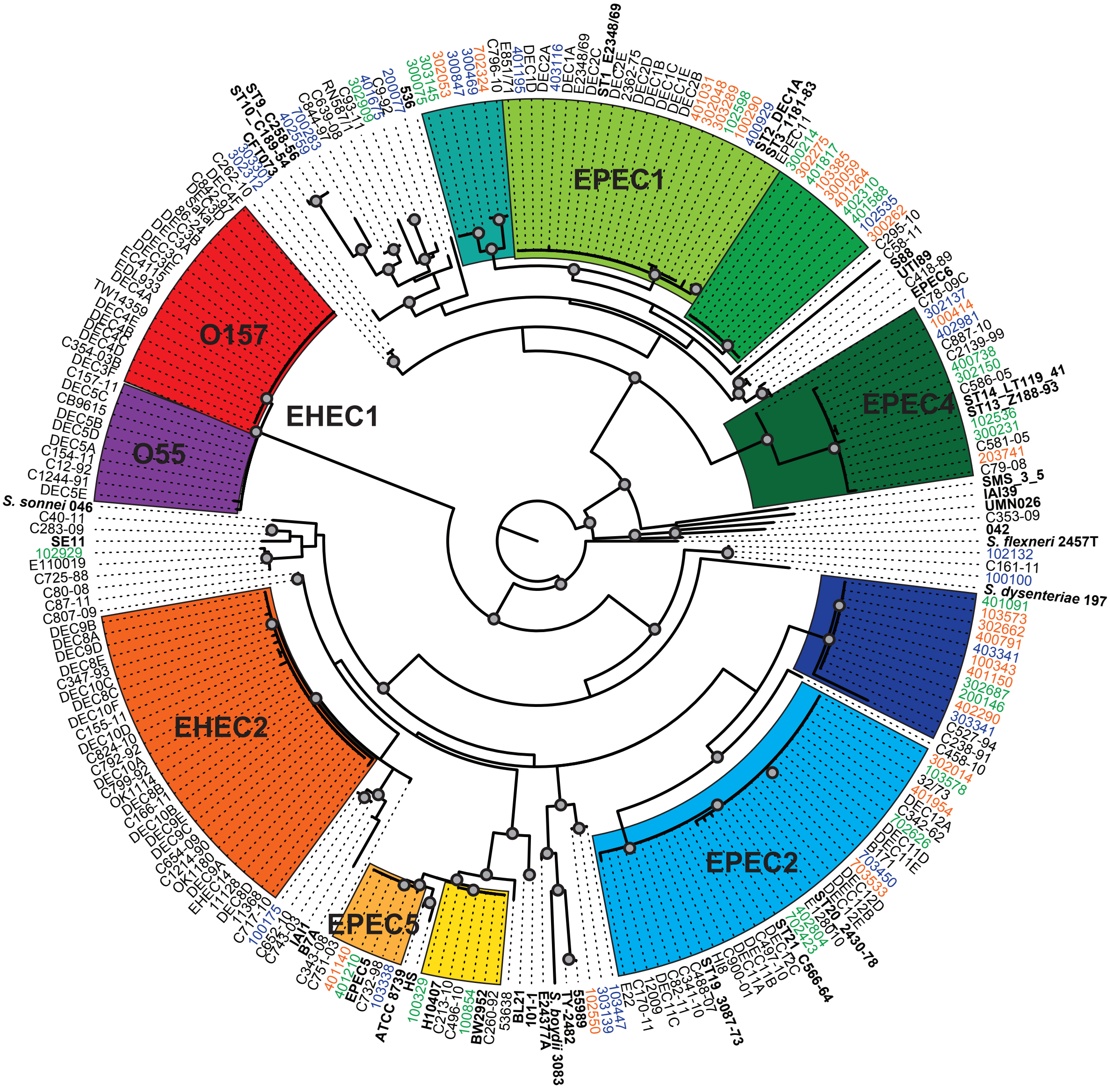

Fig. S1

Figure S1. Housekeeping gene-based phylogenetic analysis. A phylogenetic analysis of the 70 EPEC isolates characterized in this study with a collection of previously sequenced EPEC [4] and other *E. coli* isolates constructed from a concatenation of the MLST loci of the http://shigatox.net/new/tools/ecmlst.html MLST scheme [5]. A maximum-likelihood phylogeny was constructed using RAxML v7.2.8 [6] and visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). EPEC isolates from lethal outcomes (LI) are indicated in orange, while those from non-lethal symptomatic (NSI) outcomes are indicated in green, and isolates from asymptomatic (AI) outcomes are indicated in blue. The EPEC1-4 reference sequences types (STs) included are those identified by Lacher et al. [5], while the EPEC5 and EPEC6 reference sequences were from strains described by Tennant et al. [7].

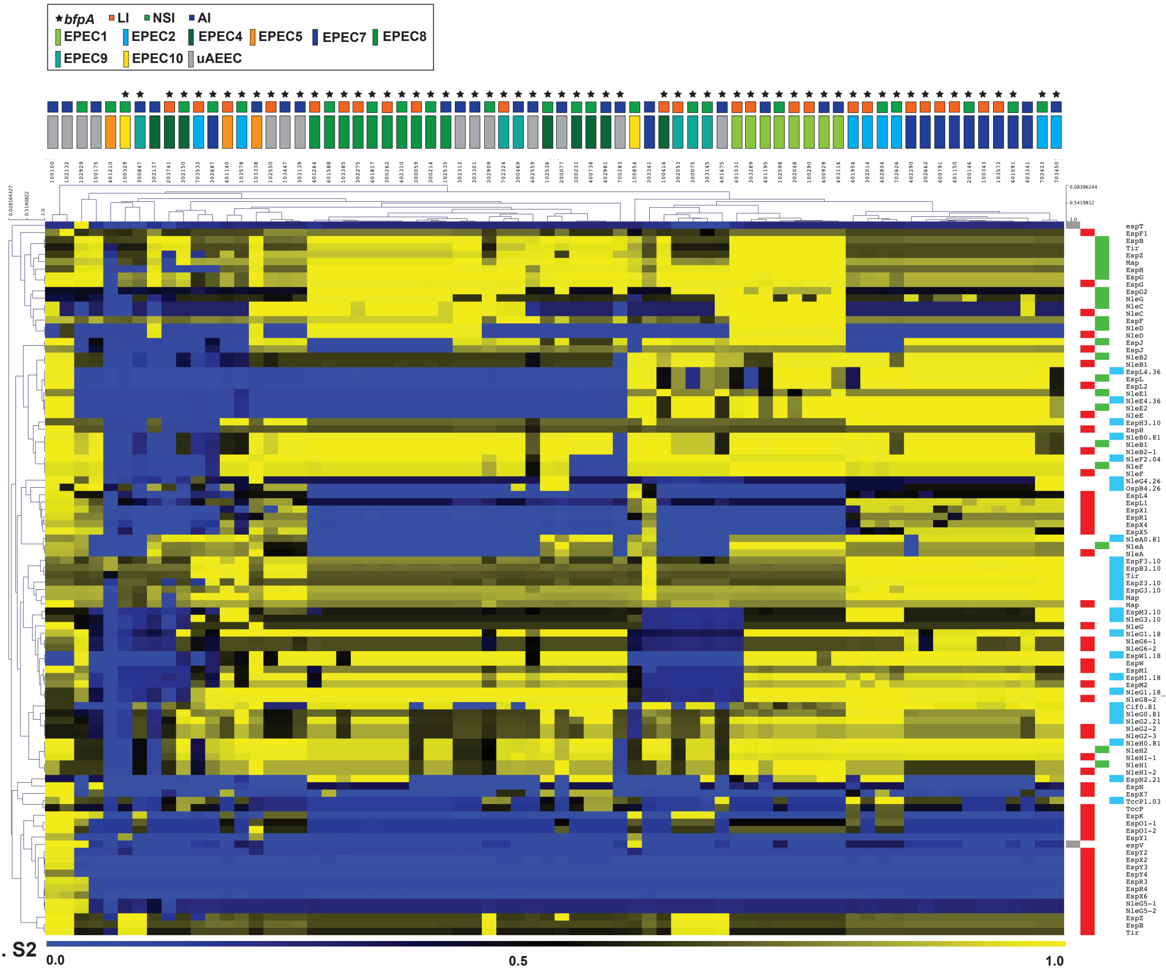

Fig. S2

Figure S2. ***In silico*** **identification of the previously characterized EPEC and EHEC T3SS effectors.** The TBLASTN BSR values identified for each T3SS effector identified in each EPEC genome analyzed were compared by hierarchical clustering [8] using Pearson correlation with average linkage, performed using MeV [9]. Each column represents one of the 70 EPEC genomes analyzed with the phylogenomic lineages designated by colored rectangles and the clinical outcome each EPEC isolate was associated with is indicated by a orange (lethal, LI), green (non-lethal symptomatic, NSI), or blue (asymptomatic, AI) square. Each row represents one of the previously characterized T3SS effectors of the EPEC strain E2348/69 (green label on the right side) [10], EPEC strain B171 (blue label on the right side) [11,12], and EHEC O157:H7 strain Sakai (red label on the right side) [13]. The presence of *bfpA* in each of the genomes is indicated by a star symbol above the clinical presentation of each isolate.

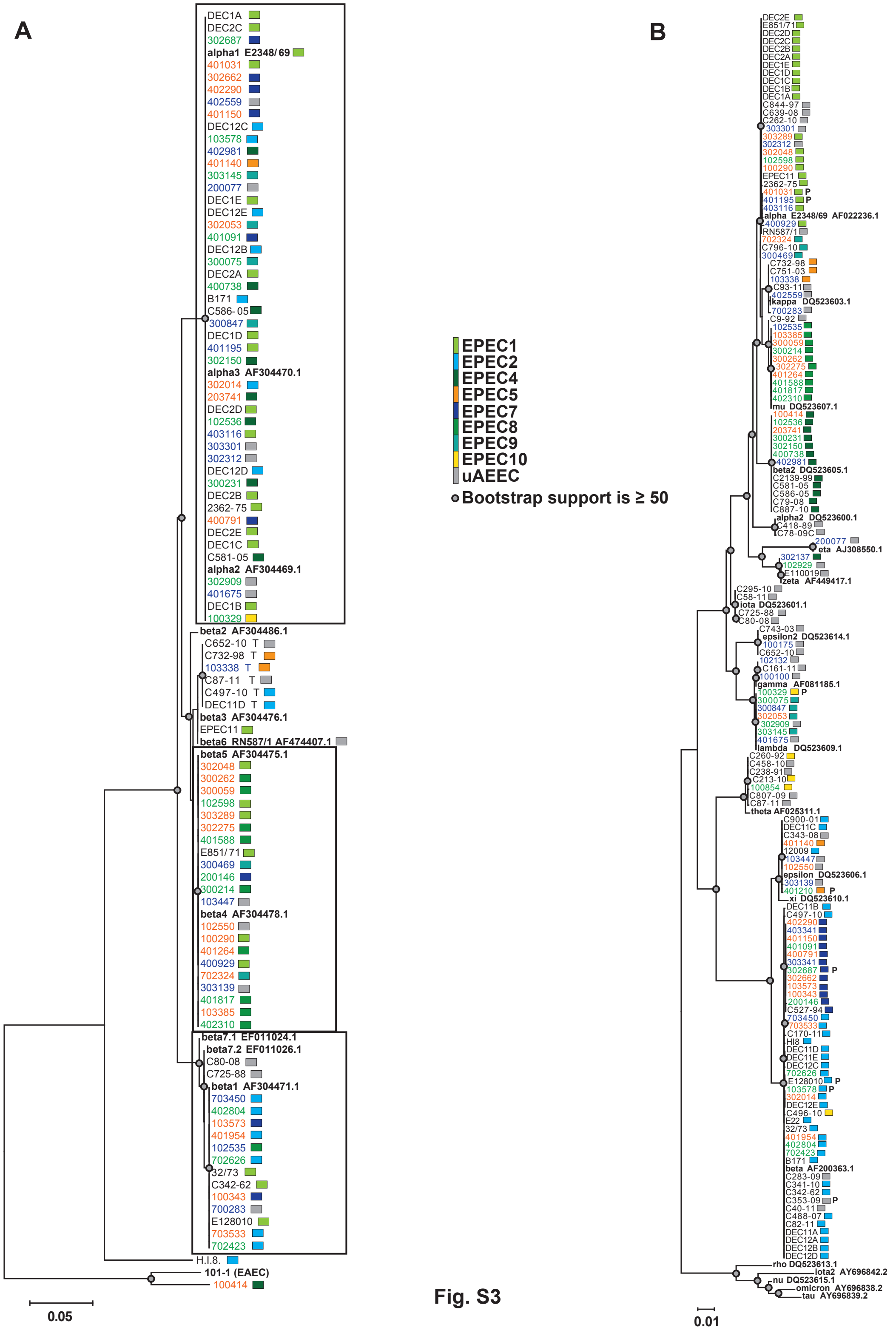

Fig. S3

Figure S3. Phylogenetic analysis of A) *bfpA,* and B) *eae* nucleotide sequences from the EPEC genomes analyzed in this study and those from a recent study [4].

Sequences of *eae* alleles [14] and *bfpA* alleles [5,15] that were previously described are included as references and their respective GenBank accession numbers are indicated. The *eae* and *bfpA* sequences that could be identified by *in silico* analysis of each genome were aligned using the ClustalW [16] algorithm in MEGA5 [17]. A maximum-likelihood phylogeny was constructed using the Kimura 2-parameter model with 1,000 bootstrap replications in MEGA5 [17]. Clinical outcomes of the isolates are designated as follows: lethal infections (LI) are indicated in orange, while those from non-lethal symptomatic infections (NSI) are indicated in green, and isolates from asymptomatic infections (AI) are indicated in blue. The colored box next to each sequence indicates the phylogenomic lineage of the EPEC isolate that each *eae* or *bfpA* allele was obtained from, and the color designations are indicated in the Figure. A "T" indicates a genome that contained a truncated *bfpA* sequence. A "P" indicates a genome contained a partial *perA* sequence.

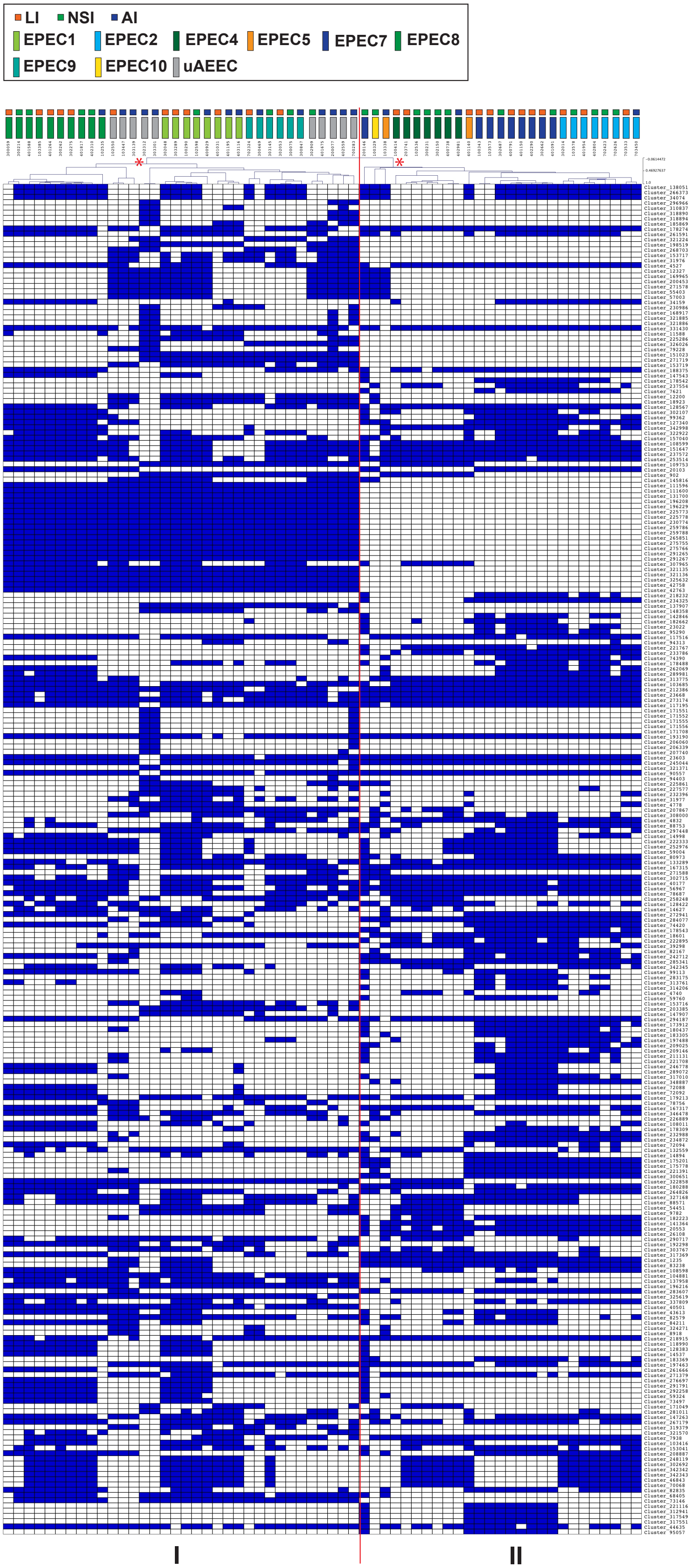
LI
NSI
AI
EPEC1
EPEC2
EPEC4
EPEC5
EPEC7
EPEC8
EPEC9
EPEC10
uAEEC
*
*
I
II

Fig. S4

Figure S4. Hierarchical cluster analysis of the 258 LS-BSR gene clusters that were significantly (chi-squared or Fisher's exact test, $p<0.05$) more associated with genomes of symptomatic (LI and NSI) compared to asymptomatic (AI) cases for the 61 typical EPEC genomes analyzed. The LS-BSR gene clusters, generated using a clustering threshold of 90% nucleotide identity, that were significantly (chi-squared or Fisher's exact test p-value <0.05) associated with genomes of symptomatic compared to asymptomatic cases, were compared by hierarchical clustering [8]. Hierarchical clustering with Pearson correlation and average linkage was performed using MeV [9]. Each column represents a genome, while each row is an LS-BSR gene cluster. The gene clusters that were present with an LS-BSR value ≥ 0.9 are indicated in blue, while the gene clusters that were absent (LS-BSR value <0.9) are white. The red boxes indicate three groups of genomes that are designated I and II. The red asterisks identify nodes that separate the genomes into the three groups. The color-coded rectangles at the top of the plot denote the phylogenomic lineage, and the color-coded squares indicate the clinical outcome of each isolate. The color-coding of each symbol is in the box at the top of the figure.

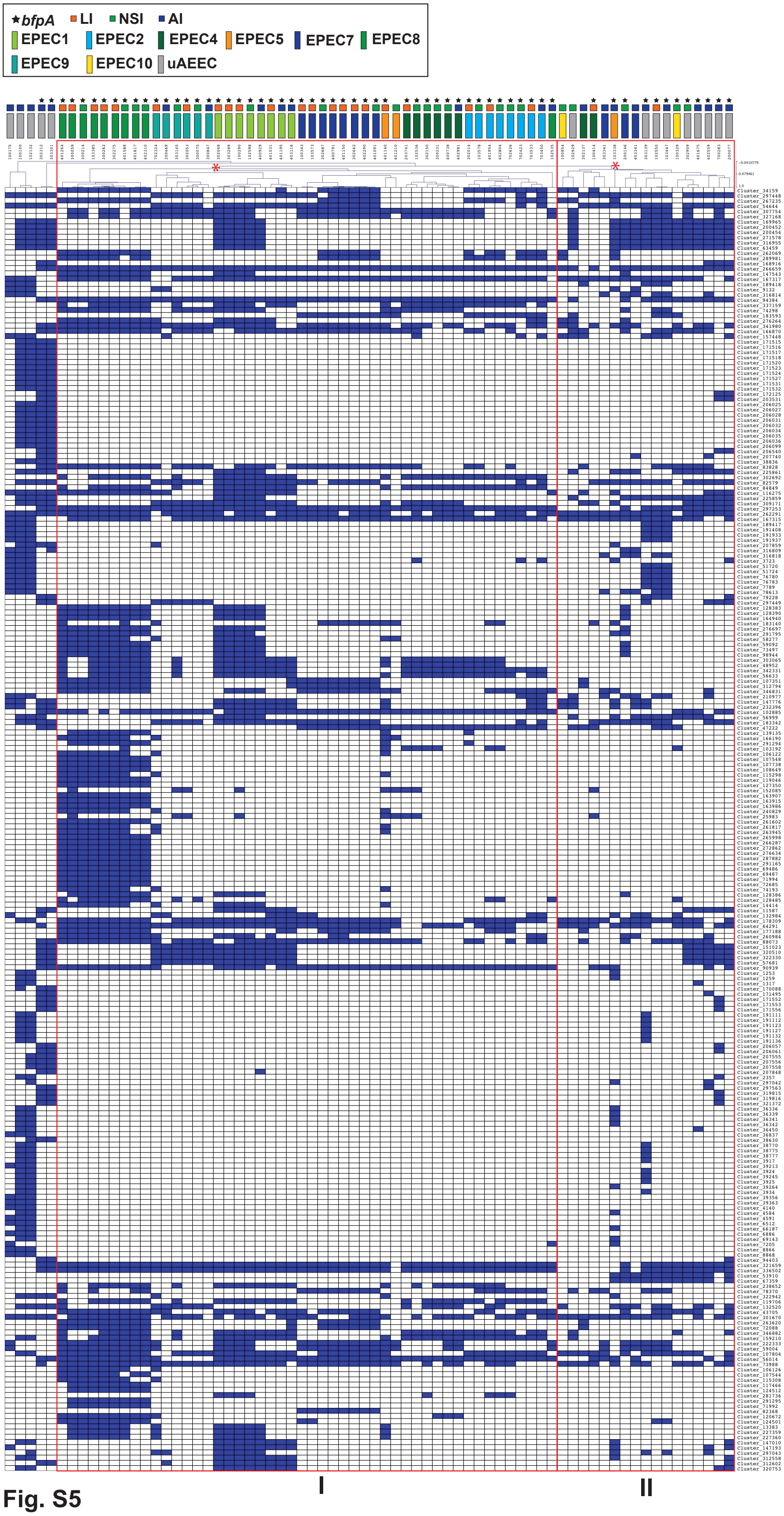
★ bfpA
LI
NSI
AI
EPEC1
EPEC2
EPEC4
EPEC5
EPEC7
EPEC8
EPEC9
EPEC10
uAEEC
I
II

Fig. S5

Figure S5. Identification of EPEC-specific factors when compared with commensal isolates. The plot is a hierarchical cluster analysis of the 246 LS-BSR gene clusters that were significantly (chi-square test or Fisher's exact test, $p<0.05$) more prevalent in genomes of symptomatic (LI and NSI) compared to asymptomatic (AI) cases for all 70 EPEC genomes analyzed when using a more inclusive clustering threshold of 80% nucleotide identity. The genes were significantly (chi-square test or Fisher's exact test p-value <0.05) associated with genomes of symptomatic compared to asymptomatic cases. These gene clusters were also determined to be absent (LS-BSR value <0.8) from three *E. coli* commensal isolates (HS, NC_009800.1; K-12, NC_000913.3; and SE11, NC_011415.1), indicating they are likely unique to EPEC or other related pathogenic *E. coli*. Hierarchical clustering [8] with Pearson correlation and average linkage was performed using MeV [9]. Each column represents a genome, while each row is an LS-BSR gene cluster. The gene clusters that were present with an LS-BSR value ≥ 0.8 are indicated in blue, while the gene clusters that were absent (LS-BSR value <0.8) are white. The red boxes indicate two groups of genomes that are designated I and II, and the red asterisks identify the nodes that separate the genomes into the two main groups. The color-coded rectangles at the top of the plot denote the phylogenomic lineage, and the color-coded squares indicate the clinical outcome of each isolate. The color-coding of each symbol is in the box at the top of the figure. A star symbol denotes the presence of *bfpA* in each genome.

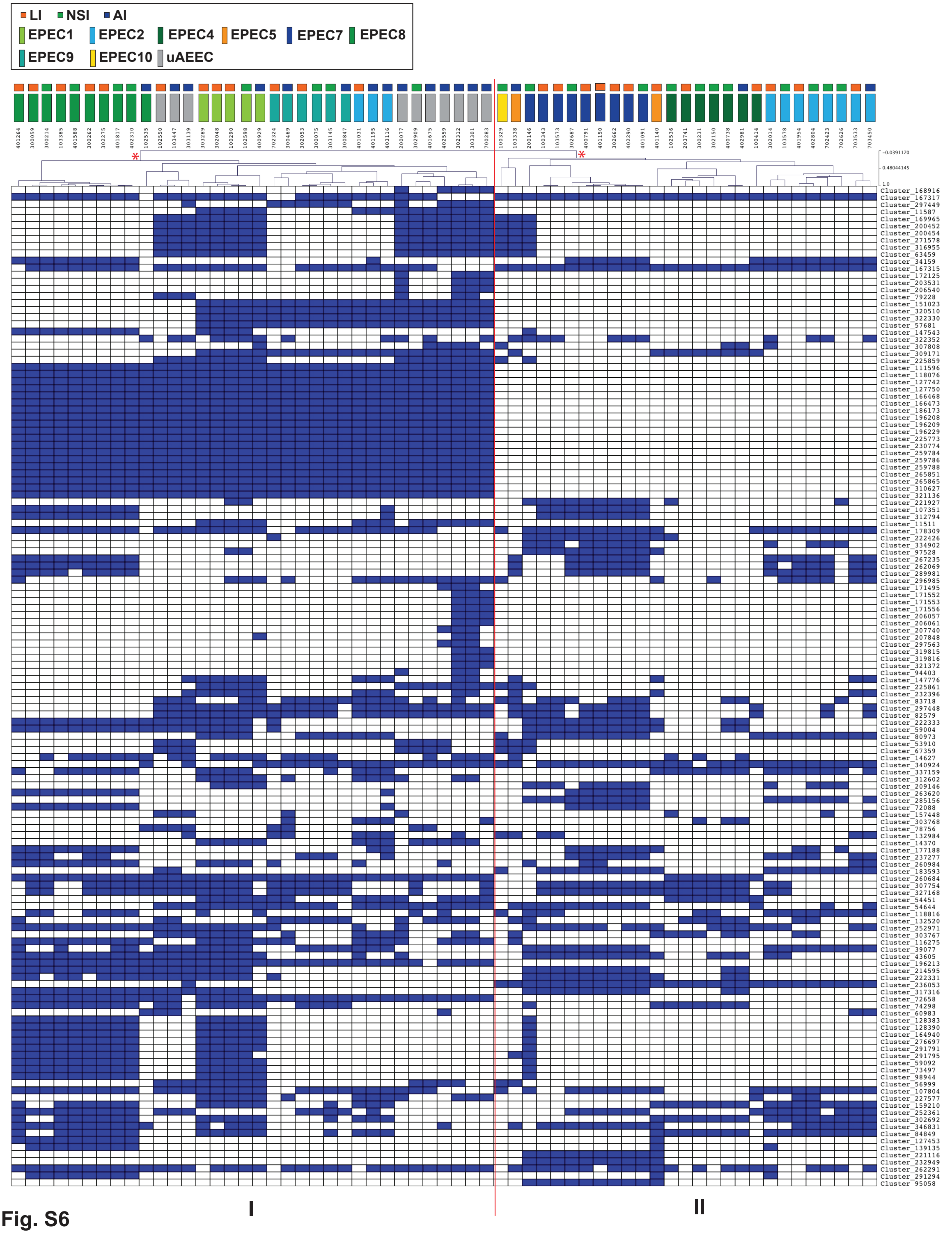

Fig. S6

Figure S6. Identification of tEPEC-specific factors when compared with commensal isolates. The plot is a hierarchical cluster analysis of the 141 tEPEC-specific LS-BSR gene clusters that were significantly (chi-squared or Fisher's exact test,

$p<0.05$) more prevalent in genomes of symptomatic (LI and NSI) compared to asymptomatic (AI) cases for the 61 typical EPEC genomes analyzed. Hierarchical clustering was used to compare the LS-BSR gene clusters, generated using a more inclusive clustering threshold of 80% nucleotide identity, that were significantly (chi-squared or Fisher's exact test p-value <0.05) associated with genomes of symptomatic compared to asymptomatic cases. These gene clusters were also determined to be absent (LS-BSR value <0.8) from three *E. coli* commensal isolates (HS, NC_009800.1; K-12, NC_000913.3; and SE11, NC_011415.1), indicating they are likely unique to EPEC or other related pathogenic *E. coli*. Hierarchical clustering [8] with Pearson correlation and average linkage was performed using MeV [9]. Each column represents a genome, while each row is an LS-BSR gene cluster. The gene clusters that were present with an LS-BSR value ≥0.8 are indicated in blue, while the gene clusters that were absent (LS-BSR value <0.8) are white. The red boxes indicate two groups of genomes that are designated I and II. The red asterisks identify nodes that separate the genomes into the two main groups. The color-coded rectangles at the top of the plot denote the phylogenomic lineage, and the color-coded squares indicate the clinical outcome of each isolate. The color-coding of each symbol is in the box at the top of the figure.

References

1. Pukatzki, S. *et al.* Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci. USA* **103**, 1528-33 (2006).
2. Shrivastava, S. & Mande, S.S. Identification and functional characterization of gene components of Type VI Secretion system in bacterial genomes. *PLoS One* **3**, e2955 (2008).
3. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10**, 104 (2009).
4. Hazen, T.H. *et al.* Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. *Proc Natl Acad Sci USA* **110**, 12810-5 (2013).
5. Lacher, D.W., Steinsland, H., Blank, T.E., Donnenberg, M.S. & Whittam, T.S. Molecular evolution of typical enteropathogenic *Escherichia coli*: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. *J. Bacteriol.* **189**, 342-50 (2007).
6. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-90 (2006).
7. Tennant, S.M. *et al.* Characterisation of atypical enteropathogenic *E. coli* strains of clinical origin. *BMC Microbiol.* **9**, 117 (2009).
8. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863-8 (1998).
9. Saeed, A.I. *et al.* TM4 microarray software suite. *Methods Enzymol.* **411**, 134-93 (2006).
10. Iguchi, A. *et al.* Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J. Bacteriol.* **191**, 347-54 (2009).
11. Ogura, Y. *et al.* Systematic identification and sequence analysis of the genomic islands of the enteropathogenic *Escherichia coli* strain B171-8 by the combined use of whole-genome PCR scanning and fosmid mapping. *J Bacteriol* **190**, 6948-60 (2008).
12. Rasko, D.A. *et al.* The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **190**, 6881-93 (2008).
13. Tobe, T. *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 USA* **103**, 14941-6 (2006).
14. Lacher, D.W., Steinsland, H. & Whittam, T.S. Allelic subtyping of the intimin locus (*eae*) of pathogenic *Escherichia coli* by fluorescent RFLP. *FEMS Microbiol. Lett.* **261**, 80-7 (2006).

15. Blank, T.E., Zhong, H., Bell, A.L., Whittam, T.S. & Donnenberg, M.S. Molecular variation among type IV pilin (*bfpA*) genes from diverse enteropathogenic *Escherichia coli* strains. *Infect. Immun.* **68**, 7028-38 (2000).

16. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-80 (1994).

17. Kumar, S., Nei, M., Dudley, J. & Tamura, K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinf.* **9**, 299-306 (2008).

List of Supplemental Tables and Data Sets

Table S1. Prevalence of virulence-associated genes among the EPEC genomes sequenced in this study by phylogenomic lineage and symptomatic group.

Table S2. Genes identified by LS-BSR that were significantly correlated with LI, NSI, or AI clinical outcomes for all 70 EPEC genomes analyzed using a stringent gene clustering threshold of ≥90%.

Table S3. Genes identified by LS-BSR that were significantly correlated with LI, NSI, or AI clinical outcomes for the 61 typical EPEC genomes analyzed using a stringent gene clustering threshold of ≥90%.

Table S4. Genes identified by LS-BSR that were significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for all 70 EPEC genomes analyzed using a stringent gene clustering threshold of ≥90%.

Table S5. Genes identified by LS-BSR that were significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for the 61 typical EPEC genomes analyzed using a stringent clustering threshold of ≥90%.

Table S6. Genes identified by LS-BSR that were significantly correlated with one clinical outcome over another for all 70 EPEC genomes analyzed using an inclusive clustering threshold of ≥80%.

Table S7. Genes identified by LS-BSR that were significantly correlated with one clinical outcome over another for the 61 typical EPEC genomes analyzed using an inclusive clustering threshold of ≥80%.

Table S8. Genes identified by LS-BSR that were significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for all 70 EPEC genomes analyzed using an inclusive clustering threshold of ≥80%.

Table S9. Genes identified by LS-BSR that were significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for the 61 typical EPEC genomes analyzed using an inclusive clustering threshold of ≥80%.

Table S10. The number of LS-BSR gene clusters with a predicted protein that contained a protein domain of surface-associated or secreted proteins.

Data Set S1. Nucleotide sequences of the 12,964 LS-BSR gene clusters identified for the 70 EPEC genomes analyzed using a stringent gene clustering threshold of ≥90%.

Data Set S2. Nucleotide sequences of the 12,196 LS-BSR gene clusters identified for the 70 EPEC genomes and three commensal *E. coli* genomes analyzed using an inclusive gene clustering threshold of ≥80%.

Table S1. Prevalence of virulence-associated genes among the EPEC genomes sequenced in this study by phylogenomic lineage and symptomatic group

			EAF Plasmid		T3SS Effectors				T6SS		
Phylogenomic Lineage	Clinical Outcome	Specimen ID	BFP	*perA*	EPEC1 (E2348/69, n=21)	EPEC2 (B171, n=26)	O157:H7 EHEC (Sakai, n=49)	T2SS (E2348/69)	T6SSA (B171)	T6SSM (B171)	OI-122
EPEC1	LI	100290	+	+	19 (90)	14 (54)	16 (33)	+	-	-	+
EPEC1	LI	302048	+	+	19 (90)	13 (50)	15 (31)	+	-	-	+
EPEC1	LI	303289	+	-	18 (86)	13 (50)	14 (29)	+	-	-	+
EPEC1	LI	401031	+	-	19 (90)	8 (31)	12 (24)	+	-	-	+
EPEC1	NSI	102598	+	+	19 (90)	14 (54)	16 (33)	+	-	-	+
EPEC1	AI	401195	+	-	17 (81)	11 (52)	13 (27)	+	-	-	2 & 3
EPEC1	AI	403116	+	+	20 (95)	13 (50)	15 (31)	+	-	-	2 & 3
EPEC1	AI	400929	+	+	20 (95)	13 (50)	15 (31)	+	-	-	2 & 3
							No. Genomes Positive	8	0	0	5
							Prevalence (%)	100	0	0	62.5
EPEC2	LI	302014	+	+	9 (43)	23 (88)	17 (35)	-	+	+	+
EPEC2	LI	401954	+	+	8 (38)	22 (85)	15 (31)	-	+	+	+
EPEC2	LI	703533	+	-	1 (5)	9 (35)	4 (8)	-	+	+	1
EPEC2	NSI	103578	+	-	4 (19)	16 (62)	7 (14)	-	+	+	1
EPEC2	NSI	402804	+	+	9 (43)	24 (92)	18 (37)	-	+	+	+
EPEC2	NSI	702423	+	+	9 (43)	23 (88)	16 (33)	-	+	+	+
EPEC2	NSI	702626	+	+	6 (29)	22 (85)	15 (31)	-	+	+	+
EPEC2	AI	703450	+	+	8 (38)	23 (88)	15 (31)	-	+	+	+
							No. Genomes Positive	0	8	8	6
							Prevalence (%)	0	100	100	75
EPEC4	LI	100414	+	-	13 (62)	8 (31)	8 (16)	+	-	-	-
EPEC4	LI	203741	+	-	3 (14)	1 (4)	1 (2)	+	-	-	-
EPEC4	NSI	102536	+	+	10 (48)	16 (62)	11 (22)	+	-	-	-
EPEC4	NSI	302150	+	+	7 (33)	6 (23)	6 (12)	+	-	-	-
EPEC4	NSI	300231	+	+	9 (43)	12 (46)	8 (16)	+	-	-	-
EPEC4	NSI	400738	+	+	9 (43)	13 (50)	11 (22)	+	-	-	-
EPEC4	AI	302137	-	-	9 (43)	5 (19)	5 (10)	+	-	-	-
EPEC4	AI	402981	+	+	9 (43)	13 (50)	11 (22)	+	-	-	-
							No. Genomes Positive	8	0	0	0
							Prevalence (%)	100	0	0	0
EPEC5	LI	401140	+	-	7 (33)	17 (65)	10 (20)	+	-	-	-
EPEC5	NSI	401210	-	-	0 (0)	1 (4)	0 (0)	+	-	-	-
EPEC5	AI	103338	T	+	15 (71)	14 (54)	19 (39)	-	+	-	-
							No. Genomes Positive	2	1	0	0
							Prevalence (%)	66.7	33.3	0	0
EPEC7	LI	100343	+	-	10 (48)	20 (77)	19 (39)	+	+	-	2 & 3
EPEC7	LI	103573	+	+	10 (48)	20 (77)	19 (39)	+	+	-	2 & 3
EPEC7	LI	302662	+	+	10 (48)	19 (73)	19 (39)	+	+	-	2 & 3
EPEC7	LI	400791	+	-	10 (48)	20 (77)	19 (39)	+	+	-	2 & 3
EPEC7	LI	402290	+	+	10 (48)	19 (73)	20 (41)	+	+	-	2 & 3
EPEC7	LI	401150	+	+	10 (48)	20 (77)	20 (41)	+	+	-	2 & 3
EPEC7	NSI	200146	+	+	11 (52)	20 (77)	21 (43)	+	+	-	2 & 3
EPEC7	NSI	302687	+	-	3 (14)	11 (52)	6 (12)	+	+	-	-
EPEC7	NSI	401091	+	+	10 (48)	20 (77)	19 (39)	+	+	-	2 & 3
EPEC7	AI	303341	-	-	10 (48)	15 (58)	14 (29)	+	+	-	2 & 3
EPEC7	AI	403341	-	-	12 (57)	19 (73)	20 (41)	+	+	-	2 & 3
							No. Genomes Positive	11	11	0	0
							Prevalence (%)	100	100	0	0
EPEC8	LI	103385	+	+	15 (71)	11 (52)	10 (20)	+	-	-	-
EPEC8	LI	300059	+	+	3 (14)	9 (35)	9 (18)	+	-	-	-
EPEC8	LI	300262	+	+	15 (71)	13 (50)	13 (27)	+	-	-	-
EPEC8	LI	302275	+	+	15 (71)	13 (50)	13 (27)	+	-	-	-
EPEC8	LI	401264	+	-	15 (71)	12 (46)	12 (24)	+	-	-	-
EPEC8	NSI	300214	+	+	15 (71)	11 (52)	11 (22)	+	-	-	-
EPEC8	NSI	402310	+	+	15 (71)	13 (50)	13 (27)	+	-	-	-
EPEC8	NSI	401817	+	+	15 (71)	13 (50)	13 (27)	+	-	-	-
EPEC8	NSI	401588	+	+	15 (71)	13 (50)	14 (29)	+	-	-	-
EPEC8	AI	102535	+	+	15 (71)	11 (52)	11 (22)	+	-	-	-
							No. Genomes Positive	10	0	0	0
							Prevalence (%)	100	0	0	0
EPEC9	LI	302053	+	+	9 (43)	5 (19)	10 (20)	+	-	-	+
EPEC9	LI	702324	+	+	14 (67)	10 (38)	9 (18)	+	-	-	-

EPEC9	NSI	303145	+	+	10 (48)	9 (35)	14 (29)	+	-	-	+
EPEC9	NSI	300075	+	+	11 (52)	7 (27)	12 (24)	+	-	-	+
EPEC9	AI	300847	+	-	0 (0)	1 (4)	4 (8)	+	-	-	1
EPEC9	AI	300469	+	+	14 (67)	10 (38)	9 (18)	+	-	-	-
						No. Genomes Positive		6	0	0	3
						Prevalence (%)		100	0	0	50
EPEC10	NSI	100329	+	-	0 (0)	0 (0)	4 (8)	-	+	-	-
EPEC10	NSI	100854	-	-	12 (57)	8 (31)	11 (22)	-	-	-	2
						No. Genomes Positive		0	1	0	0
						Prevalence (%)		0	50	0	0
uAEEC	LI	102550	+	-	9 (43)	13 (50)	13 (27)	+	+	+	-
uAEEC	NSI	102929	-	-	9 (43)	10 (38)	19 (39)	+	+	-	-
uAEEC	NSI	302909	+	-	9 (43)	6 (23)	11 (22)	+	-	-	-
uAEEC	AI	100175	-	-	9 (43)	6 (23)	7 (14)	+	+	-	-
uAEEC	AI	102132	-	-	11 (52)	11 (52)	36 (73)	-	-	-	+
uAEEC	AI	103447	+	+	9 (43)	14 (54)	15 (31)	+	+	+	-
uAEEC	AI	100100	-	-	13 (62)	11 (52)	37 (76)	-	-	-	+
uAEEC	AI	200077	+	+	13 (62)	13 (50)	10 (20)	+	-	-	-
uAEEC	AI	303301	+	+	13 (62)	8 (31)	9 (18)	+	-	-	-
uAEEC	AI	303139	+	+	9 (43)	14 (54)	14 (29)	+	+	+	-
uAEEC	AI	302312	+	+	13 (62)	8 (31)	9 (18)	+	-	-	-
uAEEC	AI	402559	+	-	9 (43)	5 (19)	5 (10)	+	-	-	-
uAEEC	AI	401675	+	+	10 (48)	4 (15)	9 (18)	+	-	-	2
uAEEC	AI	700283	+	+	9 (43)	8 (31)	4 (8)	+	-	-	-
						No. Genomes Positive		12	5	3	2
						Prevalence (%)		85.71	35.71	21.43	14.29

[a] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

[b] The EPEC phylogenomic lineages are designated EPEC1-EPEC6, or NC indicates not classified to a lineage.

[c] A "+" indicates a gene or set of genes is present with a TBLASTN BSR value ≥0.4. For the BFP operon, a "+" indicates that one or more genes of the BFP operon are present, "T" indicates an isolate that has a truncated bfp operon, and "-" indicates that no BFP genes were detected.

Table S2. Genes identified by LS-BSR that are significantly correlated with LI, NSI, or AI clinical outcomes for all 70 EPEC genomes analyzed using a stringent gene clustering threshold

Lethal (LI) vs. Non-lethal Symptomatic (NSI)

	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI	NSI	LI	NSI	% LI	% NSI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_105636	19	11	5	12	79	48	LI	3.73	0.0534	0.0355		peptidase M60-like family protein
Cluster_113721	7	15	17	8	29	65	NSI	4.77	0.0290	0.0199		phage tail tape measure protein, lambda family
Cluster_121796	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		caudovirales tail fibre assembly family protein
Cluster_123564	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479	*fyuA*	pesticin receptor
Cluster_123571	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		ABC transporter family protein
Cluster_123776	0	5	24	18	0	22	NSI	3.78	0.0520	0.0219		conserved hypothetical protein
Cluster_1322	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		conserved hypothetical protein
Cluster_136944	17	22	7	1	71	96	NSI	3.52	0.0608	0.0479		phage terminase large subunit family protein
Cluster_137178	4	11	20	12	17	48	NSI	3.91	0.0479	0.0305		hypothetical protein
Cluster_140609	2	8	22	15	8	35	NSI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_14423	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		conserved hypothetical protein
Cluster_163350	2	8	22	15	8	35	NSI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_163539	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		hypothetical protein
Cluster_168638	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		putative tniA transposase domain protein
Cluster_175605	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_183262	23	17	1	6	96	74	LI	2.89	0.0891	0.0479		hypothetical protein
Cluster_184733	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*entS*	enterobactin exporter EntS
Cluster_188256	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_196142	19	11	5	12	79	48	LI	3.73	0.0534	0.0355		conserved hypothetical protein
Cluster_196644	17	23	7	0	71	100	NSI	5.75	0.0165	0.0094	*hscA*	fe-S protein assembly chaperone HscA
Cluster_197373	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*emrB*	multidrug resistance protein B
Cluster_201731	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		helix-turn-helix, Psq domain protein
Cluster_201738	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		conserved hypothetical protein
Cluster_208031	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219		substrate binding domain of ABC-type glycine betaine transport system family protein
Cluster_209143	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		transposase C of IS166 homeodomain protein
Cluster_209349	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*tnaA*	tryptophanase
Cluster_209403	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*selB*	selenocysteine-specific translation elongation factor
Cluster_211547	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*dnaE*	DNA polymerase III, alpha subunit
Cluster_214112	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		phage tail-collar fibre family protein
Cluster_21586	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*citG*	triphosphoribosyl-dephospho-CoA synthase CitG
Cluster_21640	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219	*nagE*	PTS system, N-acetylglucosamine-specific IIBC component
Cluster_216712	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		putative membrane protein
Cluster_217185	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479	*ybtU*	yersiniabactin biosynthetic protein YbtU
Cluster_217186	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		methyltransferase family protein
Cluster_217188	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		helix-turn-helix domain protein
Cluster_217190	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		ABC transporter family protein
Cluster_217191	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		major Facilitator Superfamily protein
Cluster_217459	0	6	24	17	0	26	NSI	5.03	0.0250	0.0094		HTH domain protein
Cluster_218768	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*dacB*	D-alanyl-D-alanine carboxypeptidase/D-alanyl-D-alanine-endopeptidase
Cluster_220923	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*nuoN*	NADH-quinone oxidoreductase subunit N
Cluster_22096	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496		AAA-like domain protein
Cluster_221350	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*ftsY*	signal recognition particle-docking protein FtsY
Cluster_23468	18	10	6	13	75	43	LI	3.63	0.0569	0.0392		phage capsid scaffolding (GPO) serine peptidase family protein
Cluster_236134	16	8	8	15	67	35	LI	3.59	0.0582	0.0422	*dhfrI*	dihydrofolate reductase type 1
Cluster_236296	23	17	1	6	96	74	LI	2.89	0.0891	0.0479		ATPase subunit of terminase family protein
Cluster_237804	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496		nitrate reductase, alpha subunit
Cluster_240829	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		hypothetical protein
Cluster_24314	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		degT/DnrJ/EryC1/StrS aminotransferase family protein
Cluster_24348	5	15	19	8	21	65	NSI	7.74	0.0054	0.0032		phage tail tape measure protein, lambda family
Cluster_245108	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219		electron transport complex, RnfABCDGE type, B subunit
Cluster_245260	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*cheA*	chemotaxis protein CheA
Cluster_247403	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 2 family protein
Cluster_247837	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		putative domain protein
Cluster_254547	17	9	7	14	71	39	LI	3.58	0.0585	0.0415		bacteriophage lysis family protein
Cluster_257914	24	17	0	6	100	74	LI	5.03	0.0250	0.0094	*cadC*	transcriptional activator CadC
Cluster_258452	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*malP*	maltodextrin phosphorylase
Cluster_264521	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_265225	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219		PRD domain protein
Cluster_265334	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*mdfA*	multidrug transporter MdfA
Cluster_269540	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*can*	carbonic anhydrase 2
Cluster_274418	4	11	20	12	17	48	NSI	3.91	0.0479	0.0305		mu-like prophage FluMu gp41 family protein
Cluster_280171	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219	*tatD*	tat-linked quality control protein TatD
Cluster_282502	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219	*yphB*	aldose 1-epimerase
Cluster_283931	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		pyridine nucleotide-disulfide oxidoreductase family protein
Cluster_284866	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		bacterial regulatory, tetR family protein
Cluster_286226	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*cybH*	ni/Fe-hydrogenase, b-type cytochrome subunit
Cluster_286883	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		acetyltransferase domain protein
Cluster_290242	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219	*yphF*	ABC transporter periplasmic-binding protein YphF
Cluster_290560	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496		AAA-like domain protein
Cluster_301785	4	11	20	12	17	48	NSI	3.91	0.0479	0.0305		phage tail tape measure protein, TP901 family, core region
Cluster_306818	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_307418	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_307421	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		alpha/beta hydrolase family protein
Cluster_307429	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		salicylate synthase
Cluster_307430	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		phage integrase family protein
Cluster_313761	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		HNH endonuclease family protein
Cluster_315743	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*ycdY*	chaperone protein YcdY
Cluster_315801	2	10	22	13	8	43	NSI	5.89	0.0152	0.0078		phage tail protein I
Cluster_317131	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		conserved hypothetical protein
Cluster_32291	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219	*dcd*	deoxycytidine triphosphate deaminase
Cluster_324641	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219	*yphB*	aldose 1-epimerase
Cluster_328365	21	13	3	10	88	57	LI	4.19	0.0406	0.0243		hypothetical protein
Cluster_335585	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496		ybgS-like family protein
Cluster_342879	5	12	19	11	21	52	NSI	3.73	0.0534	0.0355		conserved hypothetical protein
Cluster_346133	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_346478	21	14	3	9	88	61	LI	3.09	0.0787	0.0490		conserved hypothetical protein
Cluster_346896	2	8	22	15	8	35	NSI	3.45	0.0631	0.0363		phage tail tape measure protein, lambda family
Cluster_36312	2	8	22	15	8	35	NSI	3.45	0.0631	0.0363		ethanolamine utilization - propanediol utilization family protein
Cluster_4538	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		putative membrane protein
Cluster_46213	3	10	21	13	13	43	NSI	4.19	0.0406	0.0243		putative transmembrane protein
Cluster_48362	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		amino acid adenylation domain protein
Cluster_48588	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		phosphoadenosine phosphosulfate reductase family protein
Cluster_51806	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*ptsP*	phosphoenolpyruvate-protein phosphotransferase

Cluster_5286	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*atpF*	ATP synthase F0, B subunit
Cluster_59364	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		endodeoxyribonuclease RusA family protein
Cluster_59365	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		phage antitermination Q family protein
Cluster_60356	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		helix-turn-helix domain protein
Cluster_64413	0	5	24	18	0	22	NSI	3.78	0.0520	0.0219		phosphoadenosine phosphosulfate reductase family protein
Cluster_8046	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		hypothetical protein
Cluster_80721	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		putative lipopolysaccharide heptosyltransferase III
Cluster_81331	18	23	6	0	75	100	NSI	4.54	0.0332	0.0219		lppC lipofamily protein
Cluster_82579	14	6	10	17	58	26	LI	3.76	0.0524	0.0392	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_83084	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*purD*	phosphoribosylamine--glycine ligase
Cluster_83927	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase
Cluster_83929	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*rfbB*	dTDP-glucose 4,6-dehydratase
Cluster_84374	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496		transglycosylase SLT domain protein
Cluster_87728	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*phrB*	deoxyribodipyrimidine photo-lyase
Cluster_92542	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496	*espS*	putative t3SS effector protein EspS
Cluster_93726	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*tolA*	protein TolA
Cluster_97263	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*cadA*	cadmium-translocating P-type ATPase
Cluster_97497	19	23	5	0	79	100	NSI	3.39	0.0654	0.0496	*tolA*	protein TolA

Lethal (LI) vs. Asymptomatic (AI)

	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI	AI	LI	AI	% LI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_102911	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		phage integrase family protein
Cluster_103347	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein
Cluster_103412	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		repressor protein C2
Cluster_103416	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		conserved hypothetical protein
Cluster_103685	24	16	0	7	100	70	LI	6.35	0.0117	0.0039		hypothetical protein
Cluster_10467	17	9	7	14	71	39	LI	3.58	0.0585	0.0415		phage Terminase family protein
Cluster_105805	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		DNA methylase family protein
Cluster_107544	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		tape measure domain protein
Cluster_107631	18	10	6	13	75	43	LI	3.63	0.0569	0.0392		phage tail tape measure protein, TP901 family, core region
Cluster_108011	13	3	11	20	54	13	LI	7.11	0.0077	0.0050	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_108598	14	5	10	18	58	22	LI	5.10	0.0239	0.0171		DNA-packaging protein FI
Cluster_108599	22	14	2	9	92	61	LI	4.61	0.0317	0.0173		tail attachment protein
Cluster_108649	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_108651	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_108754	17	9	7	14	71	39	LI	3.58	0.0585	0.0415		phage terminase, small subunit, P27 family
Cluster_109753	9	16	15	7	38	70	AI	3.65	0.0561	0.0415		outer membrane transport family protein
Cluster_11079	11	20	13	3	46	87	AI	7.11	0.0077	0.0050		transglycosylase SLT domain protein
Cluster_113457	14	20	10	3	58	87	AI	3.48	0.0619	0.0490	*fimF*	protein FimF
Cluster_113721	7	15	17	8	29	65	AI	4.77	0.0290	0.0199		phage tail tape measure protein, lambda family
Cluster_114691	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_115306	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		putative membrane protein
Cluster_117195	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_118792	23	17	1	6	96	74	LI	2.89	0.0891	0.0479		methyltransferase domain protein
Cluster_118881	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		transposase, Mutator family protein
Cluster_118990	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein
Cluster_119046	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		putative tail fiber protein
Cluster_12327	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		putative prophage protein
Cluster_125122	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		putative predicted protein
Cluster_127340	16	8	8	15	67	35	LI	3.59	0.0582	0.0422		DDE_Tnp_1-associated family protein
Cluster_127445	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_127817	23	17	1	6	96	74	LI	2.89	0.0891	0.0479	*repA*	incFII family plasmid replication initiator RepA
Cluster_128383	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein
Cluster_128567	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		phage portal protein, lambda family
Cluster_131841	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*fryC*	fructose-like permease IIC component 1
Cluster_133256	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_133289	23	17	1	6	96	74	LI	2.89	0.0891	0.0479		bacterial Ig-like domain family protein
Cluster_1339	5	12	19	11	21	52	AI	3.73	0.0534	0.0355		bacteriophage replication gene A family protein
Cluster_135012	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		gene 25-like lysozyme family protein
Cluster_136953	21	13	3	10	88	57	LI	4.19	0.0406	0.0243		haemolysin expression modulating family protein
Cluster_13773	20	12	4	11	83	52	LI	3.91	0.0479	0.0305	*folP*	dihydropteroate synthase
Cluster_138051	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		lysis S family protein
Cluster_14417	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		putative predicted protein
Cluster_14423	10	2	14	21	42	9	LI	5.09	0.0240	0.0173		conserved hypothetical protein
Cluster_14537	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		viral (Super1) RNA helicase family protein
Cluster_145816	4	12	20	11	17	52	AI	5.11	0.0238	0.0145		putative glyco3, capsid size determination protein Sid
Cluster_147543	8	0	16	23	33	0	LI	7.03	0.0080	0.0039		putative yfdA
Cluster_147672	22	15	2	8	92	65	LI	3.45	0.0631	0.0363		terminase small subunit
Cluster_14998	17	5	7	18	71	22	LI	9.48	0.0021	0.0012		bacteriophage replication gene A family protein
Cluster_15006	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		phage major capsid protein, P2 family
Cluster_150583	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		transposase family protein
Cluster_151647	22	14	2	9	92	61	LI	4.61	0.0317	0.0173		phage major capsid E family protein
Cluster_152085	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		conserved hypothetical protein
Cluster_153041	16	8	8	15	67	35	LI	3.59	0.0582	0.0422		conserved hypothetical protein
Cluster_153994	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		conserved hypothetical protein
Cluster_155688	0	4	24	19	0	17	AI	2.60	0.1067	0.0496	*yjeO*	inner membrane protein YjeO
Cluster_155747	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*phnH*	phosphonate C-P lyase system protein PhnH
Cluster_157040	18	9	6	14	75	39	LI	4.80	0.0284	0.0189		methyltransferase domain protein
Cluster_16353	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_163907	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyltransferase Family 4 family protein
Cluster_163983	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_164950	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		UTRA domain protein
Cluster_166190	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		phage integrase family protein
Cluster_166870	19	10	5	13	79	43	LI	4.91	0.0267	0.0171		integron integrase family protein
Cluster_168917	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		nucleoside 2-deoxyribosyltransferase family protein
Cluster_169965	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		ogr/Delta-like zinc finger family protein
Cluster_171515	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		putative lipoprotein
Cluster_171516	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		methyltransferase domain protein
Cluster_171517	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		beta-ketoacyl synthase, N-terminal domain protein
Cluster_171518	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		acyltransferase family protein
Cluster_171521	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		putative membrane protein
Cluster_171522	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		AMP-binding enzyme family protein
Cluster_171524	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		glycosyl transferase 2 family protein
Cluster_171525	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		thioesterase-like superfamily protein
Cluster_171526	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		outer membrane lipocarrier LolA family protein
Cluster_171527	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		MMPL family protein

Cluster_171529	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		beta-ketoacyl synthase, N-terminal domain protein
Cluster_172243	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		phage integrase family protein
Cluster_175599	5	0	19	23	21	0	LI	3.39	0.0654	0.0496	*rfbD*	dTDP-4-dehydrorhamnose reductase
Cluster_175605	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_175612	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		transposase DDE domain protein
Cluster_175618	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		polysaccharide biosynthesis/export family protein
Cluster_177488	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*fliC*	flagellin
Cluster_178274	23	17	1	6	96	74	LI	2.89	0.0891	0.0479		conserved hypothetical protein
Cluster_178309	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		conserved hypothetical protein
Cluster_181608	20	12	4	11	83	52	LI	3.91	0.0479	0.0305		transposase family protein
Cluster_183172	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		DDE domain protein
Cluster_183173	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		transposase family protein
Cluster_183184	16	8	8	15	67	35	LI	3.59	0.0582	0.0422		phage minor tail protein G
Cluster_183211	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		DNA methylase family protein
Cluster_183988	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		putative transposase
Cluster_185869	1	7	23	16	4	30	AI	4.03	0.0447	0.0226		transposase family protein
Cluster_188245	18	8	6	15	75	35	LI	6.14	0.0132	0.0084		conserved hypothetical protein
Cluster_188256	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_188375	18	10	6	13	75	43	LI	3.63	0.0569	0.0392		phage major capsid E family protein
Cluster_193190	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		nlpC/P60 family protein
Cluster_195799	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		putative icd-like protein
Cluster_196644	17	23	7	0	71	100	AI	5.75	0.0165	0.0094	*hscA*	fe-S protein assembly chaperone HscA
Cluster_200453	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		putative derepression protein
Cluster_20103	15	6	9	17	63	26	LI	4.91	0.0267	0.0189		outer membrane transport family protein
Cluster_201168	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*eutK*	ethanolamine utilization protein EutK
Cluster_201403	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		conserved hypothetical protein
Cluster_201731	10	1	14	22	42	4	LI	7.16	0.0074	0.0044		helix-turn-helix, Psq domain protein
Cluster_201738	10	1	14	22	42	4	LI	7.16	0.0074	0.0044		conserved hypothetical protein
Cluster_202509	14	5	10	18	58	22	LI	5.10	0.0239	0.0171		transposase C of IS166 homeodomain protein
Cluster_206025	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		phosphopantetheine attachment site family protein
Cluster_206026	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		phosphopantetheine attachment site family protein
Cluster_206029	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_206034	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_206036	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		fabA-like domain protein
Cluster_206037	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		KR domain protein
Cluster_206038	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		beta-ketoacyl synthase, N-terminal domain protein
Cluster_206099	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_207740	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		caudovirales tail fibre assembly family protein
Cluster_208660	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*dnaX*	DNA polymerase III, subunit gamma and tau
Cluster_209349	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*tnaA*	tryptophanase
Cluster_209403	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*selB*	selenocysteine-specific translation elongation factor
Cluster_210145	0	5	24	18	0	22	AI	3.78	0.0520	0.0219		conserved hypothetical protein
Cluster_210210	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaC*	4-hydroxyphenylacetate 3-monooxygenase, reductase component
Cluster_210211	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaB*	4-hydroxyphenylacetate 3-monooxygenase, oxygenase component
Cluster_210215	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaH*	2-oxo-hepta-3-ene-1,7-dioic acid hydratase
Cluster_210976	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		putative prophage protein
Cluster_210977	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_210978	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_211001	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		phage virion morphogenesis protein
Cluster_211547	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*dnaE*	DNA polymerase III, alpha subunit
Cluster_211780	11	18	13	5	46	78	AI	3.94	0.0470	0.0355		conserved hypothetical protein
Cluster_21640	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*nagE*	PTS system, N-acetylglucosamine-specific IIBC component
Cluster_217459	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		HTH domain protein
Cluster_21749	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		conserved hypothetical protein
Cluster_217662	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_21870	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*ssb*	single-stranded DNA-binding protein
Cluster_218768	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*dacB*	D-alanyl-D-alanine carboxypeptidase/D-alanyl-D-alanine-endopeptidase
Cluster_220312	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		baseplate J-like family protein
Cluster_220402	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*alr*	alanine racemase
Cluster_22096	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		AAA-like domain protein
Cluster_221391	11	2	13	21	46	9	LI	6.35	0.0118	0.0078		conserved hypothetical protein
Cluster_221813	19	9	5	14	79	39	LI	6.24	0.0125	0.0077		conserved hypothetical protein
Cluster_221865	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		bacterial type II and III secretion system family protein
Cluster_222331	15	7	9	16	63	30	LI	3.65	0.0561	0.0415		phage portal protein, HK97 family
Cluster_222333	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		phage major capsid protein, HK97 family
Cluster_222761	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122
Cluster_222808	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		polysaccharide biosynthesis/export family protein
Cluster_225861	3	11	21	12	13	48	AI	5.42	0.0199	0.0114	*psu*	polarity suppression protein
Cluster_226889	3	10	21	13	13	43	AI	4.19	0.0406	0.0243		putative P4-specific DNA primase
Cluster_227923	19	11	5	12	79	48	LI	3.73	0.0534	0.0355		conserved hypothetical protein
Cluster_236094	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		insA N-terminal domain protein
Cluster_236134	16	6	8	17	67	26	LI	6.22	0.0126	0.0084	*dhfrI*	dihydrofolate reductase type 1
Cluster_236296	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		ATPase subunit of terminase family protein
Cluster_237554	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		conserved hypothetical protein
Cluster_237572	22	14	2	9	92	61	LI	4.61	0.0317	0.0173		head decoration protein
Cluster_237804	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		nitrate reductase, alpha subunit
Cluster_23840	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		glycosyl transferase family 2 domain protein
Cluster_239025	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		phage terminase, small subunit, P27 family
Cluster_239026	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		phage Terminase family protein
Cluster_239028	15	7	9	16	63	30	LI	3.65	0.0561	0.0415		phage portal protein, HK97 family
Cluster_239410	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		bacteriophage replication gene A family protein
Cluster_240829	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		hypothetical protein
Cluster_24188	21	13	3	10	88	57	LI	4.19	0.0406	0.0243		conserved hypothetical protein
Cluster_24314	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		degT/DnrJ/EryC1/StrS aminotransferase family protein
Cluster_24348	5	14	19	9	21	61	AI	6.24	0.0125	0.0077		phage tail tape measure protein, lambda family
Cluster_244107	24	18	0	5	100	78	LI	3.78	0.0520	0.0219		putative purine permease YgfU
Cluster_245260	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*cheA*	chemotaxis protein CheA
Cluster_246425	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		conserved hypothetical protein
Cluster_246778	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		conserved hypothetical protein
Cluster_247403	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 2 family protein
Cluster_248097	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein
Cluster_248119	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		dnaB-like helicase N terminal domain protein
Cluster_252971	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		HNH endonuclease family protein
Cluster_252976	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		phage prohead protease, HK97 family
Cluster_257914	24	18	0	5	100	78	LI	3.78	0.0520	0.0219	*cadC*	transcriptional activator CadC
Cluster_258248	6	13	18	10	25	57	AI	3.63	0.0569	0.0392		conserved hypothetical protein
Cluster_259563	18	23	6	0	75	100	AI	4.54	0.0332	0.0219		MASE1 family protein
Cluster_261591	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		methyltransferase domain protein
Cluster_261646	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		phage/conjugal plasmid C-4 type zinc finger , TraR family protein

Cluster_261666	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		AAA domain protein
Cluster_261814	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		conserved hypothetical protein
Cluster_261875	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		hypothetical protein
Cluster_262069	12	2	12	21	50	9	LI	7.71	0.0055	0.0034		istB-like ATP binding family protein
Cluster_262139	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_263937	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		conserved hypothetical protein
Cluster_263945	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		conserved hypothetical protein
Cluster_263946	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		phage major capsid E family protein
Cluster_263948	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		conserved hypothetical protein
Cluster_264521	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_264826	19	9	5	14	79	39	LI	6.24	0.0125	0.0077		yagB/YeeU/YfjZ family protein
Cluster_265225	18	23	6	0	75	100	AI	4.54	0.0332	0.0219		PRD domain protein
Cluster_265334	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*mdfA*	multidrug transporter MdfA
Cluster_266287	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		hypothetical protein
Cluster_266373	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		lysis S family protein
Cluster_266615	16	6	8	17	67	26	LI	6.22	0.0126	0.0084		putative yubO
Cluster_266630	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein
Cluster_267010	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		putative pcar
Cluster_267179	20	11	4	12	83	48	LI	5.11	0.0238	0.0145		conserved hypothetical protein
Cluster_268703	3	9	21	14	13	39	AI	3.09	0.0787	0.0490	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_269540	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*can*	carbonic anhydrase 2
Cluster_271379	15	22	9	1	63	96	AI	5.85	0.0155	0.0102		glycosyl hydrolase family 65, N-terminal domain protein
Cluster_271578	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		conserved hypothetical protein
Cluster_271588	1	7	23	16	4	30	AI	4.03	0.0447	0.0226	*intA*	prophage CP4-57 integrase
Cluster_272862	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		helix-turn-helix domain protein
Cluster_272941	19	6	5	17	79	26	LI	11.24	0.0008	0.0004		phage tail E family protein
Cluster_274418	4	11	20	12	17	48	AI	3.91	0.0479	0.0305		mu-like prophage FluMu gp41 family protein
Cluster_275430	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*adiC*	arginine/agmatine antiporter
Cluster_275815	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*gmm*	GDP-mannose mannosyl hydrolase
Cluster_276264	11	2	13	21	46	9	LI	6.35	0.0118	0.0078		replication regulatory RepB family protein
Cluster_276292	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		stress-induced bacterial acidophilic repeat motif family protein
Cluster_276697	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein
Cluster_276834	21	14	3	9	88	61	LI	3.09	0.0787	0.0490		clp protease family protein
Cluster_277010	20	12	4	11	83	52	LI	3.91	0.0479	0.0305		transposase DDE domain protein
Cluster_280171	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*tatD*	tat-linked quality control protein TatD
Cluster_281686	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		conserved hypothetical protein
Cluster_281878	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		helix-turn-helix domain protein
Cluster_286129	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		P2 phage tail completion R family protein
Cluster_286226	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*cybH*	ni/Fe-hydrogenase, b-type cytochrome subunit
Cluster_286879	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		conserved hypothetical protein
Cluster_286883	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		acetyltransferase domain protein
Cluster_287268	11	3	13	20	46	13	LI	4.57	0.0325	0.0243		conserved hypothetical protein
Cluster_289072	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		transposase DDE domain protein
Cluster_289981	12	2	12	21	50	9	LI	7.71	0.0055	0.0034		integrase core domain protein
Cluster_290560	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		AAA-like domain protein
Cluster_291165	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		phage integrase family protein
Cluster_291297	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 2 family protein
Cluster_291813	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein
Cluster_292258	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein
Cluster_295885	18	23	6	0	75	100	AI	4.54	0.0332	0.0219		acetyltransferase domain protein
Cluster_296966	0	5	24	18	0	22	AI	3.78	0.0520	0.0219		conserved hypothetical protein
Cluster_297203	3	10	21	13	13	43	AI	4.19	0.0406	0.0243		conserved hypothetical protein
Cluster_301785	4	11	20	12	17	48	AI	3.91	0.0479	0.0305		phage tail tape measure protein, TP901 family, core region
Cluster_302042	21	14	3	9	88	61	LI	3.09	0.0787	0.0490		type II secretion system (T2SS), F family protein
Cluster_302107	14	6	10	17	58	26	LI	3.76	0.0524	0.0392		conserved hypothetical protein
Cluster_302692	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		hypothetical protein
Cluster_307965	12	3	12	20	50	13	LI	5.78	0.0162	0.0114		traP family protein
Cluster_309396	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein
Cluster_310837	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*kilR*	killing protein KilR
Cluster_313189	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		putative predicted protein
Cluster_313761	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		HNH endonuclease family protein
Cluster_315512	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaA*	4-hydroxyphenylacetate catabolism regulatory protein HpaA
Cluster_315513	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaX*	4-hydroxyphenylacetate permease
Cluster_315514	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaI*	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase
Cluster_315518	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaE*	5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase
Cluster_315519	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpcE*	homoprotocatechuate catabolism bifunctional isomerase/decarboxylase
Cluster_315520	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaR*	homoprotocatechuate degradation operon regulator, HpaR
Cluster_315743	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*ycdY*	chaperone protein YcdY
Cluster_316170	0	6	24	17	0	26	AI	5.03	0.0250	0.0094		putative predicted protein
Cluster_316814	1	7	23	16	4	30	AI	4.03	0.0447	0.0226		AAA domain protein
Cluster_316818	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		putative predicted protein
Cluster_317131	1	7	23	16	4	30	AI	4.03	0.0447	0.0226		conserved hypothetical protein
Cluster_318890	0	5	24	18	0	22	AI	3.78	0.0520	0.0219		nucleoside transporter family protein
Cluster_318894	0	5	24	18	0	22	AI	3.78	0.0520	0.0219		lysR substrate binding domain protein
Cluster_321224	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_321885	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_321886	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		hNH endonuclease family protein
Cluster_32291	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*dcd*	deoxycytidine triphosphate deaminase
Cluster_322922	15	4	9	19	63	17	LI	8.14	0.0043	0.0027		lysis S family protein
Cluster_324754	22	15	2	8	92	65	LI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_326026	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		fimbrial family protein
Cluster_326414	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		dinI-like family protein
Cluster_326791	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		ogr/Delta-like zinc finger family protein
Cluster_326910	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		hydantoinase/oxoprolinase family protein
Cluster_327168	18	7	6	16	75	30	LI	7.66	0.0056	0.0034		helix-turn-helix domain protein
Cluster_330511	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		phage major tail tube protein
Cluster_331430	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_332291	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		phage small terminase subunit
Cluster_333715	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*dinI*	DNA-damage-inducible protein I
Cluster_335585	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		ybgS-like family protein
Cluster_337809	14	6	10	17	58	26	LI	3.76	0.0524	0.0392		putative antigen 43 domain protein
Cluster_34074	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		lysis S family protein
Cluster_341403	24	19	0	4	100	83	LI	2.60	0.1067	0.0496	*lyx*	L-xylulose/3-keto-L-gulonate kinase
Cluster_34159	13	2	11	21	54	9	LI	9.18	0.0024	0.0013		conserved hypothetical protein
Cluster_342277	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		ogr/Delta-like zinc finger family protein
Cluster_342342	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		conserved hypothetical protein
Cluster_342343	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		conserved hypothetical protein
Cluster_342345	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein
Cluster_342430	21	13	3	10	88	57	LI	4.19	0.0406	0.0243		conserved hypothetical protein

Cluster_342620	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		putative predicted protein
Cluster_342998	17	6	7	17	71	26	LI	7.71	0.0055	0.0034		phage tail protein I
Cluster_343399	23	14	1	9	96	61	LI	6.61	0.0101	0.0044		phage P2 GpU family protein
Cluster_345268	18	23	6	0	75	100	AI	4.54	0.0332	0.0219	*fryC*	fructose-like permease IIC component 1
Cluster_345283	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		carbon-nitrogen hydrolase family protein
Cluster_345362	24	19	0	4	100	83	LI	2.60	0.1067	0.0496	*araD*	L-ribulose-5-phosphate 4-epimerase
Cluster_346478	21	11	3	12	88	48	LI	6.78	0.0092	0.0050		conserved hypothetical protein
Cluster_346834	22	15	2	8	92	65	LI	3.45	0.0631	0.0363		conserved hypothetical protein
Cluster_347183	20	11	4	12	83	48	LI	5.11	0.0238	0.0145		small Multidrug Resistance family protein
Cluster_347409	23	13	1	10	96	57	LI	8.05	0.0045	0.0018		phage tail sheath family protein
Cluster_347413	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		hok/gef family protein
Cluster_347784	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		phage holin family protein
Cluster_348249	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		neutral zinc metallopeptidase family protein
Cluster_348353	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*yajR*	inner membrane transport protein YajR
Cluster_348774	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		D-alanyl-D-alanine carboxypeptidase family protein
Cluster_3723	0	6	24	17	0	26	AI	5.03	0.0250	0.0094		conserved hypothetical protein
Cluster_3745	4	11	20	12	17	48	AI	3.91	0.0479	0.0305		aminotransferase class I and II family protein
Cluster_38037	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*hpaD*	3,4-dihydroxyphenylacetate 2,3-dioxygenase
Cluster_38989	0	5	24	18	0	22	AI	3.78	0.0520	0.0219		conserved hypothetical protein
Cluster_40177	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		major tail V domain protein
Cluster_43613	15	7	9	16	63	30	LI	3.65	0.0561	0.0415		phage gp6-like head-tail connector family protein
Cluster_43808	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		type II/IV secretion system family protein
Cluster_44635	17	23	7	0	71	100	AI	5.75	0.0165	0.0094		type III secretion , HrpO family protein
Cluster_4527	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		tail attachment protein
Cluster_46213	3	9	21	14	13	39	AI	3.09	0.0787	0.0490		putative transmembrane protein
Cluster_46843	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		dnaB-like helicase N terminal domain protein
Cluster_47784	4	11	20	12	17	48	AI	3.91	0.0479	0.0305		bacterial regulatory, luxR family protein
Cluster_48960	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		conserved hypothetical protein
Cluster_5054	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		uracil-xanthine permease family protein
Cluster_51806	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*ptsP*	phosphoenolpyruvate-protein phosphotransferase
Cluster_52793	19	11	5	12	79	48	LI	3.73	0.0534	0.0355	*mngB*	mannosylglycerate hydrolase
Cluster_5286	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*atpF*	ATP synthase F0, B subunit
Cluster_53395	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		haemolysin expression modulating family protein
Cluster_53578	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		putative protein
Cluster_53910	1	6	23	17	4	26	AI	2.89	0.0891	0.0479	*intA*	prophage CP4-57 integrase
Cluster_55403	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		helix-turn-helix domain protein
Cluster_57003	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		putative phage immunity repressor protein
Cluster_58271	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		conserved hypothetical protein
Cluster_59004	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		phage prohead protease, HK97 family
Cluster_59324	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		putative acetyltransferase
Cluster_59689	12	3	12	20	50	13	LI	5.78	0.0162	0.0114		putative incFII replication protein RepA3
Cluster_66669	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		conserved hypothetical protein
Cluster_68405	10	18	14	5	42	78	AI	5.10	0.0239	0.0171	*fimA*	type-1 fimbrial protein, A chain
Cluster_68745	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		phage portal protein, PBSX family
Cluster_69484	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		DNA N-6-adenine-methyltransferase family protein
Cluster_69487	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_69489	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_70068	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		conserved hypothetical protein
Cluster_71993	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		putative membrane protein
Cluster_71994	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 11 family protein
Cluster_71995	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain III family protein
Cluster_71996	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase
Cluster_72088	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		glycosyl transferases group 1 family protein
Cluster_72092	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		transposase DDE domain protein
Cluster_72094	11	3	13	20	46	13	LI	4.57	0.0325	0.0243		phage integrase family protein
Cluster_72173	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		xylose isomerase-like TIM barrel family protein
Cluster_72685	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		hypothetical protein
Cluster_73386	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		caudovirales tail fibre assembly family protein
Cluster_73491	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		PIN domain protein
Cluster_73497	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein
Cluster_73598	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		conserved hypothetical protein
Cluster_73840	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_73967	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		phage Tail Protein X family protein
Cluster_74186	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		helix-turn-helix domain protein
Cluster_74390	12	2	12	21	50	9	LI	7.71	0.0055	0.0034		hypothetical protein
Cluster_74420	19	6	5	17	79	26	LI	11.24	0.0008	0.0004		conserved hypothetical protein
Cluster_78310	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		peptidase M23 family protein
Cluster_78687	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		minor tail protein U
Cluster_79151	22	14	2	9	92	61	LI	4.61	0.0317	0.0173		bundlin family protein
Cluster_79549	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*prpR*	propionate catabolism operon regulatory protein PrpR
Cluster_8046	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		hypothetical protein
Cluster_81331	18	23	6	0	75	100	AI	4.54	0.0332	0.0219		lppC lipofamily protein
Cluster_81842	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		MATE efflux family protein
Cluster_81862	19	23	5	0	79	100	AI	3.39	0.0654	0.0496		conserved hypothetical protein
Cluster_82353	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		phage late control gene D family protein
Cluster_82835	16	6	8	17	67	26	LI	6.22	0.0126	0.0084		conserved hypothetical protein
Cluster_83084	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*purD*	phosphoribosylamine--glycine ligase
Cluster_83254	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		glycosyl transferases group 1 family protein
Cluster_83927	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase
Cluster_83929	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*rfbB*	dTDP-glucose 4,6-dehydratase
Cluster_84211	15	7	9	16	63	30	LI	3.65	0.0561	0.0415		conserved hypothetical protein
Cluster_87728	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*phrB*	deoxyribodipyrimidine photo-lyase
Cluster_88468	24	19	0	4	100	83	LI	2.60	0.1067	0.0496	*yqiJ*	inner membrane protein YqiJ
Cluster_88571	18	7	6	16	75	30	LI	7.66	0.0056	0.0034	*cro*	regulatory protein cro
Cluster_89657	24	19	0	4	100	83	LI	2.60	0.1067	0.0496		bacterial extracellular solute-binding , 7 family protein
Cluster_902	9	16	15	7	38	70	AI	3.65	0.0561	0.0415		outer membrane transport family protein
Cluster_92542	0	5	24	18	0	22	AI	3.78	0.0520	0.0219	*espS*	putative t3SS effector protein EspS
Cluster_9334	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein
Cluster_93472	1	6	23	17	4	26	AI	2.89	0.0891	0.0479	*intA*	prophage CP4-57 integrase
Cluster_93726	19	23	5	0	79	100	AI	3.39	0.0654	0.0496	*tolA*	protein TolA
Cluster_94384	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		transposase, Mutator family protein
Cluster_94431	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		hypothetical protein
Cluster_99113	16	7	8	16	67	30	LI	4.81	0.0284	0.0199		putative membrane protein
Cluster_99362	14	5	10	18	58	22	LI	5.10	0.0239	0.0171		resolvase, N terminal domain protein

Non-lethal Symptomatic (NSI) vs. Asymptomatic (AI)

No. of Genomes				
Present (≥0.9)[a,b]	Absent (<0.9)[a,b]	Percentage of Genomes Containing Cluster	Chi-squared test	Fisher's exact test

Cluster ID	NSI	AI	NSI	AI	% NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_103416	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		conserved hypothetical protein
Cluster_108598	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		DNA-packaging protein FI
Cluster_108599	21	14	2	9	91	61	NSI	4.30	0.0381	0.0351		tail attachment protein
Cluster_109753	8	16	15	7	35	70	AI	4.27	0.0388	0.0377		outer membrane transport family protein
Cluster_11588	1	8	22	15	4	35	AI	4.97	0.0257	0.0220		conserved hypothetical protein
Cluster_12327	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		putative prophage protein
Cluster_127817	23	17	0	6	100	74	NSI	4.79	0.0286	0.0216	*repA*	incFII family plasmid replication initiator RepA
Cluster_128486	11	3	12	20	48	13	NSI	5.03	0.0249	0.0230		phage antitermination Q family protein
Cluster_128567	20	13	3	10	87	57	NSI	3.86	0.0494	0.0472		phage portal protein, lambda family
Cluster_137907	4	12	19	11	17	52	AI	4.70	0.0302	0.0287		hypothetical protein
Cluster_138051	18	7	5	16	78	30	NSI	8.76	0.0031	0.0027		lysis S family protein
Cluster_147010	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		MULE transposase domain protein
Cluster_147011	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		conserved hypothetical protein
Cluster_147543	6	0	17	23	26	0	NSI	4.79	0.0286	0.0216		putative yfdA
Cluster_147907	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		HNH endonuclease family protein
Cluster_152550	10	2	13	21	43	9	NSI	5.52	0.0188	0.0165		mbeD/MobD like family protein
Cluster_153717	3	10	20	13	13	43	AI	3.86	0.0494	0.0472	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_153719	2	10	21	13	9	43	AI	5.52	0.0188	0.0165		glycosyl transferase 8 family protein
Cluster_165317	14	6	9	17	61	26	NSI	4.33	0.0373	0.0361		conserved hypothetical protein
Cluster_169965	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		ogr/Delta-like zinc finger family protein
Cluster_173601	21	14	2	9	91	61	NSI	4.30	0.0381	0.0351	*folP*	dihydropteroate synthase
Cluster_173657	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein
Cluster_178274	23	17	0	6	100	74	NSI	4.79	0.0286	0.0216		conserved hypothetical protein
Cluster_17854	8	1	15	22	35	4	NSI	4.97	0.0257	0.0220		conserved hypothetical protein
Cluster_183342	7	17	16	6	30	74	AI	7.06	0.0079	0.0072		hypothetical protein
Cluster_185869	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		transposase family protein
Cluster_188375	19	10	4	13	83	43	NSI	5.97	0.0145	0.0134		phage major capsid E family protein
Cluster_191931	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		lysR substrate binding domain protein
Cluster_191933	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-oxoacid CoA-transferase, B subunit
Cluster_191934	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		acetyl-CoA C-acetyltransferase family protein
Cluster_191935	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-hydroxyacyl-CoA dehydrogenase, NAD binding domain protein
Cluster_193295	11	19	12	4	48	83	AI	4.70	0.0302	0.0287		putative membrane protein
Cluster_198519	2	9	21	14	9	39	AI	4.30	0.0381	0.0351		transposase family protein
Cluster_200453	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		putative derepression protein
Cluster_20103	14	6	9	17	61	26	NSI	4.33	0.0373	0.0361		outer membrane transport family protein
Cluster_203385	3	12	20	11	13	52	AI	6.33	0.0119	0.0106		conserved hypothetical family protein
Cluster_2127	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-hydroxybutyrate dehydrogenase family protein
Cluster_221391	9	2	14	21	39	9	NSI	4.30	0.0381	0.0351		conserved hypothetical protein
Cluster_225286	1	8	22	15	4	35	AI	4.97	0.0257	0.0220		RHS Repeat family protein
Cluster_227577	2	9	21	14	9	39	AI	4.30	0.0381	0.0351	*ssb*	single-stranded DNA-binding protein
Cluster_230986	0	6	23	17	0	26	AI	4.79	0.0286	0.0216	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_232396	3	11	20	12	13	48	AI	5.03	0.0249	0.0230		putative dNA-binding helix-turn-helix protein
Cluster_234924	23	18	0	5	100	78	NSI	3.59	0.0581	0.0491		orn/Lys/Arg decarboxylase, C-terminal domain protein
Cluster_237554	7	0	16	23	30	0	NSI	6.07	0.0138	0.0092		conserved hypothetical protein
Cluster_237572	21	14	2	9	91	61	NSI	4.30	0.0381	0.0351		head decoration protein
Cluster_24422	3	10	20	13	13	43	AI	3.86	0.0494	0.0472		conserved hypothetical protein
Cluster_247805	23	18	0	5	100	78	NSI	3.59	0.0581	0.0491		glucosamine-6-phosphate isomerases/6-phosphogluconolactonase family protein
Cluster_248097	12	4	11	19	52	17	NSI	4.70	0.0302	0.0287		conserved hypothetical protein
Cluster_248119	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		dnaB-like helicase N terminal domain protein
Cluster_253514	23	17	0	6	100	74	NSI	4.79	0.0286	0.0216	*wcaF*	colanic acid biosynthesis acetyltransferase WcaF
Cluster_261591	15	7	8	16	65	30	NSI	4.27	0.0388	0.0377		methyltransferase domain protein
Cluster_262069	9	2	14	21	39	9	NSI	4.30	0.0381	0.0351		istB-like ATP binding family protein
Cluster_266373	18	7	5	16	78	30	NSI	8.76	0.0031	0.0027		lysis S family protein
Cluster_267342	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		putative bacteriophage protein
Cluster_271578	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		conserved hypothetical protein
Cluster_275878	19	10	4	13	83	43	NSI	5.97	0.0145	0.0134		lysR substrate binding domain protein
Cluster_281680	11	3	12	20	48	13	NSI	5.03	0.0249	0.0230		phage antitermination Q family protein
Cluster_289981	9	2	14	21	39	9	NSI	4.30	0.0381	0.0351		integrase core domain protein
Cluster_291605	7	1	16	22	30	4	NSI	3.78	0.0518	0.0470		putative upf87.6
Cluster_294187	22	14	1	9	96	61	NSI	6.26	0.0123	0.0098		conserved hypothetical protein
Cluster_29592	10	18	13	5	43	78	AI	4.47	0.0345	0.0331		conserved hypothetical protein
Cluster_296966	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein
Cluster_297448	7	19	16	4	30	83	AI	10.70	0.0011	0.0008		conserved hypothetical protein
Cluster_302692	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		hypothetical protein
Cluster_302972	22	16	1	7	96	70	NSI	3.78	0.0518	0.0470		DNA methylase family protein
Cluster_307827	10	2	13	21	43	9	NSI	5.52	0.0188	0.0165	*mbeC*	mobilization protein MbeC
Cluster_307965	13	3	10	20	57	13	NSI	7.76	0.0053	0.0045		traP family protein
Cluster_310837	0	5	23	18	0	22	AI	3.59	0.0581	0.0491	*kilR*	killing protein KilR
Cluster_316814	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		AAA domain protein
Cluster_318890	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		nucleoside transporter family protein
Cluster_318894	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		lysR substrate binding domain protein
Cluster_31976	3	10	20	13	13	43	AI	3.86	0.0494	0.0472	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_322029	15	7	8	16	65	30	NSI	4.27	0.0388	0.0377		hypothetical protein
Cluster_322922	12	4	11	19	52	17	NSI	4.70	0.0302	0.0287		lysis S family protein
Cluster_324271	0	6	23	17	0	26	AI	4.79	0.0286	0.0216	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_325003	8	18	15	5	35	78	AI	7.17	0.0074	0.0067		conserved hypothetical protein
Cluster_326414	0	6	23	17	0	26	AI	4.79	0.0286	0.0216		dinI-like family protein
Cluster_327168	15	7	8	16	65	30	NSI	4.27	0.0388	0.0377		helix-turn-helix domain protein
Cluster_332075	10	3	13	20	43	13	NSI	3.86	0.0494	0.0472	*espN*	putative t3SS effector protein EspN2-2
Cluster_337590	12	3	11	20	52	13	NSI	6.33	0.0119	0.0106	*rop*	regulatory protein rop
Cluster_34074	18	7	5	16	78	30	NSI	8.76	0.0031	0.0027		lysis S family protein
Cluster_34159	12	2	11	21	52	9	NSI	8.32	0.0039	0.0031		conserved hypothetical protein
Cluster_341964	9	17	14	6	39	74	AI	4.33	0.0373	0.0361		putative exonuclease VIII
Cluster_34197	14	6	9	17	61	26	NSI	4.33	0.0373	0.0361		HTH-like domain protein
Cluster_342342	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		conserved hypothetical protein
Cluster_342343	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		conserved hypothetical protein
Cluster_342345	13	4	10	19	57	17	NSI	5.97	0.0145	0.0134		conserved hypothetical protein
Cluster_40177	20	13	3	10	87	57	NSI	3.86	0.0494	0.0472		major tail V domain protein
Cluster_4527	21	13	2	10	91	57	NSI	5.52	0.0188	0.0165		tail attachment protein
Cluster_46843	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		dnaB-like helicase N terminal domain protein
Cluster_46983	15	7	8	16	65	30	NSI	4.27	0.0388	0.0377		conserved hypothetical protein
Cluster_4778	3	11	20	12	13	48	AI	5.03	0.0249	0.0230		conserved hypothetical protein
Cluster_4832	17	8	6	15	74	35	NSI	5.61	0.0179	0.0169		DNA methylase family protein
Cluster_49147	6	0	17	23	26	0	NSI	4.79	0.0286	0.0216		putative phage protein
Cluster_49492	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		acetyltransferase family protein
Cluster_50317	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein
Cluster_51720	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-oxoacid CoA-transferase, A subunit
Cluster_51724	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		citrate transporter family protein

Cluster_55403	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		helix-turn-helix domain protein
Cluster_57003	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		putative phage immunity repressor protein
Cluster_59342	7	1	16	22	30	4	NSI	3.78	0.0518	0.0470		conserved hypothetical protein
Cluster_59365	3	10	20	13	13	43	AI	3.86	0.0494	0.0472		phage antitermination Q family protein
Cluster_59689	10	3	13	20	43	13	NSI	3.86	0.0494	0.0472		putative incFII replication protein RepA3
Cluster_64612	12	3	11	20	52	13	NSI	6.33	0.0119	0.0106	*rop*	regulatory protein rop
Cluster_68405	9	18	14	5	39	78	AI	5.74	0.0166	0.0156	*fimA*	type-1 fimbrial protein, A chain
Cluster_70068	14	5	9	18	61	22	NSI	5.74	0.0166	0.0156		conserved hypothetical protein
Cluster_7049	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein
Cluster_73925	9	2	14	21	39	9	NSI	4.30	0.0381	0.0351	*mbeA*	DNA relaxase MbeA
Cluster_74390	11	2	12	21	48	9	NSI	6.86	0.0088	0.0074		hypothetical protein
Cluster_78687	20	13	3	10	87	57	NSI	3.86	0.0494	0.0472		minor tail protein U
Cluster_79228	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein
Cluster_7938	20	12	3	11	87	52	NSI	5.03	0.0249	0.0230		DNA methylase family protein
Cluster_82579	6	17	17	6	26	74	AI	8.70	0.0032	0.0028	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_83701	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		traP family protein
Cluster_88571	16	7	7	16	70	30	NSI	5.57	0.0183	0.0174	*cro*	regulatory protein cro
Cluster_8918	1	7	22	16	4	30	AI	3.78	0.0518	0.0470	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_902	8	16	15	7	35	70	AI	4.27	0.0388	0.0377		outer membrane transport family protein
Cluster_94384	22	16	1	7	96	70	NSI	3.78	0.0518	0.0470		transposase, Mutator family protein

[a] Genes were grouped into gene clusters using a "stringent" clustering threshold of ≥90% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.9, while those that are "absent" have a TBLASTN LS-BSR <0.9.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

Table S3. Genes identified by LS-BSR that are significantly correlated with LI, NSI, or AI clinical outcomes for the 61 typical EPEC genomes analyzed using a stringent gene clustering threshold

Lethal (LI) vs. Non-lethal Symptomatic (NSI)

	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI	NSI	LI	NSI	% LI	% NSI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_121796	0	4	24	16	0	20	NSI	3.14	0.0765	0.0357		caudovirales tail fibre assembly family protein
Cluster_123776	0	5	24	15	0	25	NSI	4.51	0.0336	0.0143		conserved hypothetical protein
Cluster_137178	4	11	20	9	17	55	NSI	5.53	0.0187	0.0112		hypothetical protein
Cluster_140609	2	8	22	12	8	40	NSI	4.56	0.0328	0.0270		conserved hypothetical protein
Cluster_163539	0	4	24	16	0	20	NSI	3.14	0.0765	0.0357		hypothetical protein
Cluster_168638	9	2	15	18	38	10	LI	3.06	0.0805	0.0445		putative tniA transposase domain protein
Cluster_175605	6	0	18	20	25	0	LI	3.86	0.0494	0.0246	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_196644	17	20	7	0	71	100	NSI	4.93	0.0264	0.0111	*hscA*	fe-S protein assembly chaperone HscA
Cluster_208031	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246		substrate binding domain of ABC-type glycine betaine transport system family protein
Cluster_21640	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246	*nagE*	PTS system, N-acetylglucosamine-specific IIBC component
Cluster_216712	1	6	23	14	4	30	NSI	3.68	0.0550	0.0353		putative membrane protein
Cluster_217459	0	5	24	15	0	25	NSI	4.51	0.0336	0.0143		HTH domain protein
Cluster_24348	5	12	19	8	21	60	NSI	5.50	0.0190	0.0128		phage tail tape measure protein, lambda family
Cluster_245108	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246		electron transport complex, RnfABCDGE type, B subunit
Cluster_247837	6	0	18	20	25	0	LI	3.86	0.0494	0.0246		putative domain protein
Cluster_257914	24	16	0	4	100	80	LI	3.14	0.0765	0.0357	*cadC*	transcriptional activator CadC
Cluster_265225	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246		PRD domain protein
Cluster_274418	4	10	20	10	17	50	NSI	4.16	0.0415	0.0253		mu-like prophage FluMu gp41 family protein
Cluster_280171	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246	*tatD*	tat-linked quality control protein TatD
Cluster_282502	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246	*yphB*	aldose 1-epimerase
Cluster_283931	24	16	0	4	100	80	LI	3.14	0.0765	0.0357		pyridine nucleotide-disulfide oxidoreductase family protein
Cluster_290242	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246	*yphF*	ABC transporter periplasmic-binding protein YphF
Cluster_301785	4	10	20	10	17	50	NSI	4.16	0.0415	0.0253		phage tail tape measure protein, TP901 family, core region
Cluster_306818	0	4	24	16	0	20	NSI	3.14	0.0765	0.0357		conserved hypothetical protein
Cluster_313761	9	2	15	18	38	10	LI	3.06	0.0805	0.0445		HNH endonuclease family protein
Cluster_315801	2	9	22	11	8	45	NSI	5.99	0.0144	0.0121		phage tail protein I
Cluster_317131	1	6	23	14	4	30	NSI	3.68	0.0550	0.0353		conserved hypothetical protein
Cluster_32291	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246	*dcd*	deoxycytidine triphosphate deaminase
Cluster_324641	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246	*yphB*	aldose 1-epimerase
Cluster_332075	4	10	20	10	17	50	NSI	4.16	0.0415	0.0253	*espN*	putative t3SS effector protein EspN2-2
Cluster_346896	2	8	22	12	8	40	NSI	4.56	0.0328	0.0270		phage tail tape measure protein, lambda family
Cluster_46213	3	10	21	10	13	50	NSI	5.68	0.0172	0.0092		putative transmembrane protein
Cluster_4962	15	18	9	2	63	90	NSI	3.06	0.0805	0.0445		conserved hypothetical protein
Cluster_64413	0	4	24	16	0	20	NSI	3.14	0.0765	0.0357		phosphoadenosine phosphosulfate reductase family protein
Cluster_67345	7	0	17	20	29	0	LI	4.93	0.0264	0.0111		conserved hypothetical protein
Cluster_81331	18	20	6	0	75	100	NSI	3.86	0.0494	0.0246		lppC lipofamily protein
Cluster_83927	6	0	18	20	25	0	LI	3.86	0.0494	0.0246	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase
Cluster_83929	6	0	18	20	25	0	LI	3.86	0.0494	0.0246	*rfbB*	dTDP-glucose 4,6-dehydratase
Cluster_99077	4	10	20	10	17	50	NSI	4.16	0.0415	0.0253		cysteine protease , YopT-type domain protein

Lethal (LI) vs. Asymptomatic (AI)

	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI	AI	LI	AI	% LI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_102911	2	6	22	11	8	35	AI	3.05	0.0808	0.0486		phage integrase family protein
Cluster_103685	24	13	0	4	100	76	LI	3.87	0.0491	0.0235		hypothetical protein
Cluster_10467	17	6	7	11	71	35	LI	3.76	0.0524	0.0308		phage Terminase family protein
Cluster_108011	13	2	11	15	54	12	LI	5.99	0.0144	0.0082	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_108599	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		tail attachment protein
Cluster_108754	17	6	7	11	71	35	LI	3.76	0.0524	0.0308		phage terminase, small subunit, P27 family
Cluster_109753	9	13	15	4	38	76	AI	4.61	0.0318	0.0252		outer membrane transport family protein
Cluster_11079	11	14	13	3	46	82	AI	4.15	0.0417	0.0251		transglycosylase SLT domain protein
Cluster_11588	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		conserved hypothetical protein
Cluster_12200	7	11	17	6	29	65	AI	3.76	0.0524	0.0308		glycosyltransferase Family 4 family protein
Cluster_12327	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		putative prophage protein
Cluster_125122	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		putative predicted protein
Cluster_127340	16	4	8	13	67	24	LI	5.79	0.0162	0.0109		DDE_Tnp_1-associated family protein
Cluster_128567	22	10	2	7	92	59	LI	4.49	0.0340	0.0208		phage portal protein, lambda family
Cluster_131841	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*fryC*	fructose-like permease IIC component 1
Cluster_132559	12	15	12	2	50	88	AI	4.88	0.0272	0.0183		phoH-like family protein
Cluster_133256	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		conserved hypothetical protein
Cluster_133289	23	11	1	6	96	65	LI	4.79	0.0286	0.0141		bacterial Ig-like domain family protein
Cluster_1339	5	9	19	8	21	53	AI	3.25	0.0716	0.0476		bacteriophage replication gene A family protein
Cluster_138051	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		lysis S family protein
Cluster_142846	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein
Cluster_14417	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		putative predicted protein
Cluster_145816	4	9	20	8	17	53	AI	4.49	0.0341	0.0198		putative glyco3, capsid size determination protein Sid
Cluster_14627	5	11	19	6	21	65	AI	6.31	0.0120	0.0086		cysteine protease , YopT-type domain protein
Cluster_147263	20	9	4	8	83	53	LI	3.09	0.0786	0.0454		methyltransferase domain protein
Cluster_147543	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		putative yfdA
Cluster_148358	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		putative exported protein
Cluster_14894	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		conserved hypothetical protein
Cluster_149480	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		conserved hypothetical protein
Cluster_14998	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		bacteriophage replication gene A family protein
Cluster_151023	6	11	18	6	25	65	AI	4.93	0.0264	0.0230		conserved hypothetical protein
Cluster_151647	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		phage major capsid E family protein
Cluster_152085	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_153041	16	5	8	12	67	29	LI	4.14	0.0419	0.0278		conserved hypothetical protein
Cluster_153717	5	10	19	7	21	59	AI	4.66	0.0309	0.0212	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_157040	18	6	6	11	75	35	LI	4.93	0.0264	0.0230		methyltransferase domain protein
Cluster_166190	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		phage integrase family protein
Cluster_168917	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		nucleoside 2-deoxyribosyltransferase family protein
Cluster_169965	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		ogr/Delta-like zinc finger family protein
Cluster_174674	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein
Cluster_175131	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		bacterial transferase hexapeptide family protein
Cluster_175605	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_175612	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		transposase DDE domain protein
Cluster_175618	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		polysaccharide biosynthesis/export family protein
Cluster_176078	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		outer membrane autotransporter barrel domain protein
Cluster_176094	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		NIF3 family protein

Cluster_176531	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		hypothetical protein
Cluster_176970	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_177334	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_177488	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*fliC*	flagellin
Cluster_178120	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_178144	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative dsORF-h1
Cluster_178244	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_178274	23	11	1	6	96	65	LI	4.79	0.0286	0.0141		conserved hypothetical protein
Cluster_178488	12	2	12	15	50	12	LI	4.88	0.0272	0.0183	*nleE*	putative nleE
Cluster_178542	8	0	16	17	33	0	LI	5.08	0.0242	0.0128	*epaR*	type III secretion apparatus protein SpaR/YscT/HrcT
Cluster_178543	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		type III secretion , HrpO family protein
Cluster_180437	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		putative transposase
Cluster_181608	20	9	4	8	83	53	LI	3.09	0.0786	0.0454		transposase family protein
Cluster_182662	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		putative membrane protein
Cluster_183172	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		DDE domain protein
Cluster_183173	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		transposase family protein
Cluster_183278	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		caudovirales tail fibre assembly family protein
Cluster_185869	1	7	23	10	4	41	AI	6.48	0.0109	0.0051		transposase family protein
Cluster_18601	15	16	9	1	63	94	AI	3.82	0.0508	0.0281		ribbon-helix-helix , copG family protein
Cluster_187872	2	6	22	11	8	35	AI	3.05	0.0808	0.0486		conserved hypothetical protein
Cluster_18923	7	11	17	6	29	65	AI	3.76	0.0524	0.0308		glycosyltransferase Family 4 family protein
Cluster_196644	17	17	7	0	71	100	AI	4.10	0.0430	0.0295	*hscA*	fe-S protein assembly chaperone HscA
Cluster_197463	15	16	9	1	63	94	AI	3.82	0.0508	0.0281	*hipA*	serine/threonine-protein kinase HipA
Cluster_198519	4	9	20	8	17	53	AI	4.49	0.0341	0.0198		transposase family protein
Cluster_200453	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		putative derepression protein
Cluster_20103	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		outer membrane transport family protein
Cluster_201168	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*eutK*	ethanolamine utilization protein EutK
Cluster_201403	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein
Cluster_201731	10	1	14	16	42	6	LI	4.80	0.0285	0.0136		helix-turn-helix, Psq domain protein
Cluster_201738	10	1	14	16	42	6	LI	4.80	0.0285	0.0136		conserved hypothetical protein
Cluster_207867	7	12	17	5	29	71	AI	5.30	0.0213	0.0124		putative invasin
Cluster_208660	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*dnaX*	DNA polymerase III, subunit gamma and tau
Cluster_209025	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		putative epaO1 protein
Cluster_209146	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		shET2 enterotoxin, N-terminal region family protein
Cluster_210354	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_211131	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		insA N-terminal domain protein
Cluster_21153	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_21640	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*nagE*	PTS system, N-acetylglucosamine-specific IIBC component
Cluster_217662	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		conserved hypothetical protein
Cluster_218232	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		conserved hypothetical protein
Cluster_221116	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		hypothetical protein
Cluster_221391	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		conserved hypothetical protein
Cluster_221708	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein
Cluster_221813	19	8	5	9	79	47	LI	3.25	0.0716	0.0476		conserved hypothetical protein
Cluster_221927	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		phage Tail Collar domain protein
Cluster_222331	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		phage portal protein, HK97 family
Cluster_222333	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		phage major capsid protein, HK97 family
Cluster_222761	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122
Cluster_222808	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		polysaccharide biosynthesis/export family protein
Cluster_225286	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		RHS Repeat family protein
Cluster_225861	3	8	21	9	13	47	AI	4.42	0.0355	0.0292	*psu*	polarity suppression protein
Cluster_23022	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein
Cluster_230986	2	6	22	11	8	35	AI	3.05	0.0808	0.0486	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_232988	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		putative molybdate metabolism regulator
Cluster_234042	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		3-octaprenyl-4-hydroxybenzoate carboxy-lyase family protein
Cluster_234325	10	1	14	16	42	6	LI	4.80	0.0285	0.0136		bacterial extracellular solute-binding, 5 Middle family protein
Cluster_234872	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		lysR substrate binding domain protein
Cluster_236134	16	5	8	12	67	29	LI	4.14	0.0419	0.0278	*dhfrI*	dihydrofolate reductase type 1
Cluster_236502	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		binding-protein-dependent transport system inner membrane component family protein
Cluster_237200	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		RHS repeat-associated core domain protein
Cluster_237298	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_23755	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		alpha amylase, N-terminal ig-like domain protein
Cluster_237554	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein
Cluster_23757	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		binding-protein-dependent transport system inner membrane component family protein
Cluster_237572	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		head decoration protein
Cluster_23758	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		bacterial extracellular solute-binding family protein
Cluster_23840	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		glycosyl transferase family 2 domain protein
Cluster_239028	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		phage portal protein, HK97 family
Cluster_239410	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		bacteriophage replication gene A family protein
Cluster_240829	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		hypothetical protein
Cluster_243374	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		esterase family protein
Cluster_243375	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		tatD related DNase family protein
Cluster_244976	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*ynjE*	thiosulfate sulfurtransferase YnjE domain protein
Cluster_246425	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_246778	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein
Cluster_252976	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		phage prohead protease, HK97 family
Cluster_257914	24	12	0	5	100	71	LI	5.53	0.0187	0.0083	*cadC*	transcriptional activator CadC
Cluster_258248	6	12	18	5	25	71	AI	6.65	0.0099	0.0052		conserved hypothetical protein
Cluster_259563	18	17	6	0	75	100	AI	3.18	0.0746	0.0327		MASE1 family protein
Cluster_261591	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		methyltransferase domain protein
Cluster_261646	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		phage/conjugal plasmid C-4 type zinc finger , TraR family protein
Cluster_261814	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_262069	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		istB-like ATP binding family protein
Cluster_263937	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein
Cluster_263945	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein
Cluster_263946	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		phage major capsid E family protein
Cluster_263948	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein
Cluster_264826	19	7	5	10	79	41	LI	4.66	0.0309	0.0212		yagB/YeeU/YfjZ family protein
Cluster_265225	18	17	6	0	75	100	AI	3.18	0.0746	0.0327		PRD domain protein
Cluster_266373	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		lysis S family protein
Cluster_267179	20	9	4	8	83	53	LI	3.09	0.0786	0.0454		conserved hypothetical protein
Cluster_267890	5	9	19	8	21	53	AI	3.25	0.0716	0.0476		bacteriophage replication gene A family protein
Cluster_268703	3	9	21	8	13	53	AI	6.03	0.0141	0.0127	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_271134	2	6	22	11	8	35	AI	3.05	0.0808	0.0486		outer membrane autotransporter barrel domain protein
Cluster_271379	15	16	9	1	63	94	AI	3.82	0.0508	0.0281		glycosyl hydrolase family 65, N-terminal domain protein
Cluster_271578	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		conserved hypothetical protein
Cluster_271588	1	6	23	11	4	35	AI	4.79	0.0286	0.0141	*intA*	prophage CP4-57 integrase

Cluster_271719	6	11	18	6	25	65	AI	4.93	0.0264	0.0230		conserved hypothetical protein
Cluster_272941	19	6	5	11	79	35	LI	6.31	0.0120	0.0086		phage tail E family protein
Cluster_273556	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		glycosyltransferase like 2 family protein
Cluster_274418	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		mu-like prophage FluMu gp41 family protein
Cluster_275815	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*gmm*	GDP-mannose mannosyl hydrolase
Cluster_276264	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		replication regulatory RepB family protein
Cluster_280171	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*tatD*	tat-linked quality control protein TatD
Cluster_283175	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		bacterial regulatory, luxR family protein
Cluster_284077	17	6	7	11	71	35	LI	3.76	0.0524	0.0308	*fepE*	ferric enterobactin (Enterochelin) transport
Cluster_286879	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_289072	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		transposase DDE domain protein
Cluster_289981	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		integrase core domain protein
Cluster_290717	11	14	13	3	46	82	AI	4.15	0.0417	0.0251		istB-like ATP binding family protein
Cluster_295885	18	17	6	0	75	100	AI	3.18	0.0746	0.0327		acetyltransferase domain protein
Cluster_296966	0	5	24	12	0	29	AI	5.53	0.0187	0.0083		conserved hypothetical protein
Cluster_301785	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		phage tail tape measure protein, TP901 family, core region
Cluster_302107	14	2	10	15	58	12	LI	7.22	0.0072	0.0035		conserved hypothetical protein
Cluster_308000	7	12	17	5	29	71	AI	5.30	0.0213	0.0124	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_310837	0	5	24	12	0	29	AI	5.53	0.0187	0.0083	*kilR*	killing protein KilR
Cluster_312941	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		putative antirepressor protein
Cluster_313541	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_313758	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		caudovirales tail fibre assembly family protein
Cluster_313761	9	0	15	17	38	0	LI	6.13	0.0133	0.0055		HNH endonuclease family protein
Cluster_314109	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*casB*	CRISPR type I-E/ECOLI-associated protein CasB/Cse2
Cluster_314206	7	0	17	17	29	0	LI	4.10	0.0430	0.0295	*eprH*	type III secretion apparatus protein PrgH/EprH
Cluster_315364	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		archaeal transcriptional regulator TrmB family protein
Cluster_315365	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		ABC transporter family protein
Cluster_315366	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative glycosidase
Cluster_315444	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		papC C-terminal domain protein
Cluster_316058	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		phospholipase_D-nuclease N-terminal family protein
Cluster_316833	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_317010	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein
Cluster_317549	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		ead/Ea22-like family protein
Cluster_317550	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_317551	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein
Cluster_318890	0	5	24	12	0	29	AI	5.53	0.0187	0.0083		nucleoside transporter family protein
Cluster_318894	0	5	24	12	0	29	AI	5.53	0.0187	0.0083		lysR substrate binding domain protein
Cluster_31976	5	10	19	7	21	59	AI	4.66	0.0309	0.0212	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_321224	2	8	22	9	8	47	AI	6.13	0.0133	0.0083		conserved hypothetical protein
Cluster_321885	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		conserved hypothetical protein
Cluster_321886	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		hNH endonuclease family protein
Cluster_32291	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*dcd*	deoxycytidine triphosphate deaminase
Cluster_322922	15	3	9	14	63	18	LI	6.41	0.0114	0.0096		lysis S family protein
Cluster_326026	2	8	22	9	8	47	AI	6.13	0.0133	0.0083		fimbrial family protein
Cluster_326910	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		hydantoinase/oxoprolinase family protein
Cluster_331430	24	13	0	4	100	76	LI	3.87	0.0491	0.0235		conserved hypothetical protein
Cluster_34074	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		lysis S family protein
Cluster_34159	13	2	11	15	54	12	LI	5.99	0.0144	0.0082		conserved hypothetical protein
Cluster_342998	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		phage tail protein I
Cluster_345268	18	17	6	0	75	100	AI	3.18	0.0746	0.0327	*fryC*	fructose-like permease IIC component 1
Cluster_346478	21	8	3	9	88	47	LI	6.03	0.0141	0.0127		conserved hypothetical protein
Cluster_347476	20	9	4	8	83	53	LI	3.09	0.0786	0.0454		conserved hypothetical protein
Cluster_3745	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		aminotransferase class I and II family protein
Cluster_40177	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		major tail V domain protein
Cluster_43613	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		phage gp6-like head-tail connector family protein
Cluster_44635	17	17	7	0	71	100	AI	4.10	0.0430	0.0295		type III secretion , HrpO family protein
Cluster_4527	22	10	2	7	92	59	LI	4.49	0.0340	0.0208		tail attachment protein
Cluster_4740	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		marR family protein
Cluster_47784	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		bacterial regulatory, luxR family protein
Cluster_52793	19	8	5	9	79	47	LI	3.25	0.0716	0.0476	*mngB*	mannosylglycerate hydrolase
Cluster_55403	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		helix-turn-helix domain protein
Cluster_56967	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		RNA 2'-phosphotransferase, Tpt1 / KptA family protein
Cluster_57003	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		putative phage immunity repressor protein
Cluster_58271	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein
Cluster_59004	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		phage prohead protease, HK97 family
Cluster_59760	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		putative variable tail fiber protein
Cluster_72088	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		glycosyl transferases group 1 family protein
Cluster_72092	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		transposase DDE domain protein
Cluster_72094	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		phage integrase family protein
Cluster_73598	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_74186	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		helix-turn-helix domain protein
Cluster_74390	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		hypothetical protein
Cluster_74420	19	6	5	11	79	35	LI	6.31	0.0120	0.0086		conserved hypothetical protein
Cluster_7621	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		putative type III secretion EprI protein
Cluster_78687	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		minor tail protein U
Cluster_79564	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative oxidoreductase subunit domain protein
Cluster_80022	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		alpha amylase, catalytic domain protein
Cluster_80973	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		phage integrase family protein
Cluster_81331	18	17	6	0	75	100	AI	3.18	0.0746	0.0327		lppC lipofamily protein
Cluster_82167	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein
Cluster_82835	16	5	8	12	67	29	LI	4.14	0.0419	0.0278		conserved hypothetical protein
Cluster_83254	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		glycosyl transferases group 1 family protein
Cluster_83927	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase
Cluster_83929	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*rfbB*	dTDP-glucose 4,6-dehydratase
Cluster_84140	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative cymA protein
Cluster_84211	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		conserved hypothetical protein
Cluster_902	9	13	15	4	38	76	AI	4.61	0.0318	0.0252		outer membrane transport family protein
Cluster_94431	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		hypothetical protein
Cluster_95057	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		P22AR C-terminal domain protein
Cluster_95290	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		invH outer membrane lipofamily protein
Cluster_95942	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein
Cluster_98381	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		molybdopterin guanine dinucleotide synthesis B family protein
Cluster_99113	16	4	8	13	67	24	LI	5.79	0.0162	0.0109		putative membrane protein
Cluster_99362	14	2	10	15	58	12	LI	7.22	0.0072	0.0035		resolvase, N terminal domain protein
Cluster_99439	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative dsORF-e4

Non-lethal Symptomatic (NSI) vs. Asymptomatic (AI)

	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	NSI	AI	NSI	AI	% NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_103416	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		conserved hypothetical protein
Cluster_104527	9	2	11	15	45	12	NSI	3.40	0.0653	0.0365	*fimA*	type-1 fimbrial protein, A chain
Cluster_104881	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		putative toxin higB-2
Cluster_108598	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		DNA-packaging protein FI
Cluster_108599	19	11	1	6	95	65	NSI	3.70	0.0544	0.0335		tail attachment protein
Cluster_109753	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		outer membrane transport family protein
Cluster_111596	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		ethanolamine utilization EutN/carboxysome family protein
Cluster_111600	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		electron transport complex, RnfABCDGE type, C subunit
Cluster_11588	1	8	19	9	5	47	AI	6.69	0.0097	0.0055		conserved hypothetical protein
Cluster_12200	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		glycosyltransferase Family 4 family protein
Cluster_1225	16	8	4	9	80	47	NSI	3.05	0.0808	0.0470	*cbeA*	cytoskeleton bundling-enhancing protein CbeA
Cluster_12327	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		putative prophage protein
Cluster_127340	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		DDE_Tnp_1-associated family protein
Cluster_128422	15	6	5	11	75	35	NSI	4.40	0.0360	0.0220		dynamin family protein
Cluster_128486	11	3	9	14	55	18	NSI	3.98	0.0461	0.0397		phage antitermination Q family protein
Cluster_128654	16	8	4	9	80	47	NSI	3.05	0.0808	0.0470		conserved hypothetical protein
Cluster_131700	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*xseA*	exodeoxyribonuclease VII, large subunit
Cluster_137907	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		hypothetical protein
Cluster_138051	15	4	5	13	75	24	NSI	7.79	0.0052	0.0029		lysis S family protein
Cluster_141364	11	3	9	14	55	18	NSI	3.98	0.0461	0.0397	*ompD*	outer membrane porin protein OmpD
Cluster_142846	5	0	15	17	25	0	NSI	3.01	0.0829	0.0498		conserved hypothetical protein
Cluster_145816	4	9	16	8	20	53	AI	3.05	0.0808	0.0470		putative glyco3, capsid size determination protein Sid
Cluster_147543	6	0	14	17	30	0	NSI	4.08	0.0434	0.0220		putative yfdA
Cluster_147907	1	7	19	10	5	41	AI	5.12	0.0236	0.0140		HNH endonuclease family protein
Cluster_148358	2	8	18	9	10	47	AI	4.66	0.0309	0.0234		putative exported protein
Cluster_151023	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		conserved hypothetical protein
Cluster_151647	19	11	1	6	95	65	NSI	3.70	0.0544	0.0335		phage major capsid E family protein
Cluster_152550	10	2	10	15	50	12	NSI	4.51	0.0337	0.0173		mbeD/MobD like family protein
Cluster_153716	2	9	18	8	10	53	AI	6.19	0.0129	0.0097		conserved hypothetical protein
Cluster_153717	2	10	18	7	10	59	AI	7.89	0.0050	0.0037	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_153719	2	10	18	7	10	59	AI	7.89	0.0050	0.0037		glycosyl transferase 8 family protein
Cluster_167315	19	11	1	6	95	65	NSI	3.70	0.0544	0.0335		conserved hypothetical protein
Cluster_167317	17	8	3	9	85	47	NSI	4.43	0.0353	0.0323		intergenic-region protein
Cluster_167318	19	11	1	6	95	65	NSI	3.70	0.0544	0.0335		DNA repair RadC family protein
Cluster_168494	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		conserved hypothetical protein
Cluster_168917	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		nucleoside 2-deoxyribosyltransferase family protein
Cluster_169965	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		ogr/Delta-like zinc finger family protein
Cluster_173912	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481	*espZ*	sepZ family protein
Cluster_178274	20	11	0	6	100	65	NSI	6.03	0.0141	0.0053		conserved hypothetical protein
Cluster_178309	9	2	11	15	45	12	NSI	3.40	0.0653	0.0365		conserved hypothetical protein
Cluster_17854	8	1	12	16	40	6	NSI	4.11	0.0428	0.0227		conserved hypothetical protein
Cluster_178542	6	0	14	17	30	0	NSI	4.08	0.0434	0.0220	*epaR*	type III secretion apparatus protein SpaR/YscT/HrcT
Cluster_178543	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		type III secretion , HrpO family protein
Cluster_180288	16	7	4	10	80	41	NSI	4.35	0.0369	0.0210	*xseA*	exodeoxyribonuclease VII, large subunit
Cluster_182223	11	3	9	14	55	18	NSI	3.98	0.0461	0.0397		putative outer membrane fimbrial usher protein
Cluster_183305	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		putative type III secretion protein EprJ
Cluster_183342	7	13	13	4	35	76	AI	4.80	0.0284	0.0202		hypothetical protein
Cluster_185869	1	7	19	10	5	41	AI	5.12	0.0236	0.0140		transposase family protein
Cluster_18601	13	16	7	1	65	94	AI	3.04	0.0813	0.0481		ribbon-helix-helix , copG family protein
Cluster_188375	17	7	3	10	85	41	NSI	5.94	0.0148	0.0075		phage major capsid E family protein
Cluster_18923	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		glycosyltransferase Family 4 family protein
Cluster_192298	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		transposase IS116/IS110/IS902 family protein
Cluster_196208	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*eutP*	ethanolamine utilization protein, EutP
Cluster_196216	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		putative propanediol utilization protein
Cluster_196229	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		putative cobalamin biosynthesis protein CbiG
Cluster_197488	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		putative dNA-binding transcriptional regulator
Cluster_198519	2	9	18	8	10	53	AI	6.19	0.0129	0.0097		transposase family protein
Cluster_200453	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		putative derepression protein
Cluster_20103	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		outer membrane transport family protein
Cluster_203385	3	9	17	8	15	53	AI	4.43	0.0353	0.0323		conserved hypothetical family protein
Cluster_212386	20	13	0	4	100	76	NSI	3.12	0.0774	0.0360		sigma-54 interaction domain protein
Cluster_218232	5	0	15	17	25	0	NSI	3.01	0.0829	0.0498		conserved hypothetical protein
Cluster_218915	20	13	0	4	100	76	NSI	3.12	0.0774	0.0360		parB-like nuclease domain protein
Cluster_221767	11	2	9	15	55	12	NSI	5.76	0.0164	0.0140	*nleA*	putative nleA1 protein
Cluster_222895	8	1	12	16	40	6	NSI	4.11	0.0428	0.0227		conserved hypothetical protein
Cluster_225286	1	8	19	9	5	47	AI	6.69	0.0097	0.0055		RHS Repeat family protein
Cluster_225773	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduD*	propanediol dehydratase medium subunit
Cluster_225778	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		BMC domain protein
Cluster_227577	2	8	18	9	10	47	AI	4.66	0.0309	0.0234	*ssb*	single-stranded DNA-binding protein
Cluster_23022	5	0	15	17	25	0	NSI	3.01	0.0829	0.0498		conserved hypothetical protein
Cluster_230774	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduL*	phosphate propanoyltransferase
Cluster_230986	0	6	20	11	0	35	AI	6.03	0.0141	0.0053	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_232396	1	8	19	9	5	47	AI	6.69	0.0097	0.0055		putative dNA-binding helix-turn-helix protein
Cluster_233786	11	2	9	15	55	12	NSI	5.76	0.0164	0.0140	*eae*	intimin
Cluster_234325	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		bacterial extracellular solute-binding, 5 Middle family protein
Cluster_23668	20	13	0	4	100	76	NSI	3.12	0.0774	0.0360	*allC*	allantoate amidohydrolase
Cluster_237554	7	0	13	17	35	0	NSI	5.23	0.0221	0.0094		conserved hypothetical protein
Cluster_237572	19	11	1	6	95	65	NSI	3.70	0.0544	0.0335		head decoration protein
Cluster_237633	19	11	1	6	95	65	NSI	3.70	0.0544	0.0335		transposase, IS605 OrfB family
Cluster_248119	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		dnaB-like helicase N terminal domain protein
Cluster_253514	20	11	0	6	100	65	NSI	6.03	0.0141	0.0053	*wcaF*	colanic acid biosynthesis acetyltransferase WcaF
Cluster_259786	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduC*	propanediol dehydratase large subunit
Cluster_259788	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduE*	propanediol dehydratase small subunit
Cluster_26108	11	3	9	14	55	18	NSI	3.98	0.0461	0.0397	*ompD*	outer membrane porin protein OmpD
Cluster_261591	14	4	6	13	70	24	NSI	6.19	0.0128	0.0081		methyltransferase domain protein
Cluster_265851	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		BMC domain protein
Cluster_266373	15	4	5	13	75	24	NSI	7.79	0.0052	0.0029		lysis S family protein
Cluster_267342	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		putative bacteriophage protein
Cluster_268703	3	9	17	8	15	53	AI	4.43	0.0353	0.0323	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_271578	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		conserved hypothetical protein
Cluster_271719	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		conserved hypothetical protein
Cluster_273174	20	13	0	4	100	76	NSI	3.12	0.0774	0.0360	*cobC*	alpha-ribazole phosphatase
Cluster_275755	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduU*	propanediol utilization protein PduU

Cluster_275766	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		dehydratase medium subunit
Cluster_275878	17	8	3	9	85	47	NSI	4.43	0.0353	0.0323		lysR substrate binding domain protein
Cluster_281680	11	3	9	14	55	18	NSI	3.98	0.0461	0.0397		phage antitermination Q family protein
Cluster_283607	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		methyltransferase domain protein
Cluster_285341	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		H+ symporter family protein
Cluster_291265	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		iron-containing alcohol dehydrogenase family protein
Cluster_291267	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduT*	shell protein
Cluster_294187	19	10	1	7	95	59	NSI	5.12	0.0236	0.0140		conserved hypothetical protein
Cluster_29592	10	15	10	2	50	88	AI	4.51	0.0337	0.0173		conserved hypothetical protein
Cluster_296966	0	5	20	12	0	29	AI	4.52	0.0335	0.0142		conserved hypothetical protein
Cluster_297448	7	14	13	3	35	82	AI	6.58	0.0103	0.0071		conserved hypothetical protein
Cluster_301551	1	6	19	11	5	35	AI	3.70	0.0544	0.0335		phage tail protein I
Cluster_302270	15	7	5	10	75	41	NSI	3.07	0.0797	0.0498		bacterial Ig-like domain family protein
Cluster_302692	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		hypothetical protein
Cluster_302715	19	11	1	6	95	65	NSI	3.70	0.0544	0.0335		conserved hypothetical protein
Cluster_303767	5	10	15	7	25	59	AI	3.07	0.0797	0.0498		repressor protein CI
Cluster_307827	10	2	10	15	50	12	NSI	4.51	0.0337	0.0173	*mbeC*	mobilization protein MbeC
Cluster_307926	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489	*traY*	relaxosome protein TraY
Cluster_307965	12	3	8	14	60	18	NSI	5.19	0.0227	0.0176		traP family protein
Cluster_309716	5	10	15	7	25	59	AI	3.07	0.0797	0.0498	*espF*	espF repeat family protein
Cluster_309725	1	6	19	11	5	35	AI	3.70	0.0544	0.0335	*eae*	intimin
Cluster_310837	0	5	20	12	0	29	AI	4.52	0.0335	0.0142	*kilR*	killing protein KilR
Cluster_313775	15	7	5	10	75	41	NSI	3.07	0.0797	0.0498		outer membrane autotransporter barrel domain protein
Cluster_317369	15	7	5	10	75	41	NSI	3.07	0.0797	0.0498		tellurite resistance TerB family protein
Cluster_318890	0	5	20	12	0	29	AI	4.52	0.0335	0.0142		nucleoside transporter family protein
Cluster_318894	0	5	20	12	0	29	AI	4.52	0.0335	0.0142		lysR substrate binding domain protein
Cluster_319379	2	8	18	9	10	47	AI	4.66	0.0309	0.0234	*rfaJ*	lipopolysaccharide 1,2-glucosyltransferase
Cluster_31976	2	10	18	7	10	59	AI	7.89	0.0050	0.0037	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_31977	2	8	18	9	10	47	AI	4.66	0.0309	0.0234	*rfaJ*	lipopolysaccharide 1,2-glucosyltransferase
Cluster_321135	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		helix-turn-helix domain protein
Cluster_321136	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		MIP channel family protein
Cluster_321224	2	8	18	9	10	47	AI	4.66	0.0309	0.0234		conserved hypothetical protein
Cluster_321570	2	8	18	9	10	47	AI	4.66	0.0309	0.0234		transposase family protein
Cluster_321885	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		conserved hypothetical protein
Cluster_321886	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		hNH endonuclease family protein
Cluster_322029	13	4	7	13	65	24	NSI	4.80	0.0284	0.0202		hypothetical protein
Cluster_324271	0	4	20	13	0	24	AI	3.12	0.0774	0.0360	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_325619	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		aldehyde dehydrogenase family protein
Cluster_325632	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduA*	propanediol utilization protein PduA
Cluster_326585	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		putative protein TraJ
Cluster_331430	20	13	0	4	100	76	NSI	3.12	0.0774	0.0360		conserved hypothetical protein
Cluster_337590	11	3	9	14	55	18	NSI	3.98	0.0461	0.0397	*rop*	regulatory protein rop
Cluster_34074	15	4	5	13	75	24	NSI	7.79	0.0052	0.0029		lysis S family protein
Cluster_34159	10	2	10	15	50	12	NSI	4.51	0.0337	0.0173		conserved hypothetical protein
Cluster_341964	9	14	11	3	45	82	AI	3.98	0.0461	0.0397		putative exonuclease VIII
Cluster_342342	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		conserved hypothetical protein
Cluster_342343	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		conserved hypothetical protein
Cluster_342345	13	4	7	13	65	24	NSI	4.80	0.0284	0.0202		conserved hypothetical protein
Cluster_346896	8	1	12	16	40	6	NSI	4.11	0.0428	0.0227		phage tail tape measure protein, lambda family
Cluster_348887	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		hypothetical protein
Cluster_39298	11	16	9	1	55	94	AI	5.28	0.0215	0.0102		endodeoxyribonuclease RusA family protein
Cluster_40501	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		helix-turn-helix domain protein
Cluster_42758	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		cob(I)alamin adenosyltransferase
Cluster_42763	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduA*	propanediol utilization protein PduA
Cluster_4527	19	10	1	7	95	59	NSI	5.12	0.0236	0.0140		tail attachment protein
Cluster_46843	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		dnaB-like helicase N terminal domain protein
Cluster_46983	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		conserved hypothetical protein
Cluster_4740	5	0	15	17	25	0	NSI	3.01	0.0829	0.0498		marR family protein
Cluster_4778	1	8	19	9	5	47	AI	6.69	0.0097	0.0055		conserved hypothetical protein
Cluster_4832	16	5	4	12	80	29	NSI	7.63	0.0057	0.0030		DNA methylase family protein
Cluster_49147	6	0	14	17	30	0	NSI	4.08	0.0434	0.0220		putative phage protein
Cluster_4962	18	9	2	8	90	53	NSI	4.66	0.0309	0.0234		conserved hypothetical protein
Cluster_54451	5	10	15	7	25	59	AI	3.07	0.0797	0.0498		conserved hypothetical protein
Cluster_55403	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		helix-turn-helix domain protein
Cluster_57003	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		putative phage immunity repressor protein
Cluster_59342	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		conserved hypothetical protein
Cluster_64612	11	3	9	14	55	18	NSI	3.98	0.0461	0.0397	*rop*	regulatory protein rop
Cluster_6530	4	9	16	8	20	53	AI	3.05	0.0808	0.0470		putative exonuclease VIII
Cluster_70068	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		conserved hypothetical protein
Cluster_73925	9	2	11	15	45	12	NSI	3.40	0.0653	0.0365	*mbeA*	DNA relaxase MbeA
Cluster_74390	9	2	11	15	45	12	NSI	3.40	0.0653	0.0365		hypothetical protein
Cluster_7621	6	0	14	17	30	0	NSI	4.08	0.0434	0.0220		putative type III secretion EprI protein
Cluster_79228	0	5	20	12	0	29	AI	4.52	0.0335	0.0142		conserved hypothetical protein
Cluster_7938	18	9	2	8	90	53	NSI	4.66	0.0309	0.0234		DNA methylase family protein
Cluster_82167	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		conserved hypothetical protein
Cluster_82579	6	13	14	4	30	76	AI	6.19	0.0128	0.0081	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_83701	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		traP family protein
Cluster_88753	14	5	6	12	70	29	NSI	4.54	0.0330	0.0217		conserved hypothetical protein
Cluster_8918	0	4	20	13	0	24	AI	3.12	0.0774	0.0360	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_89529	5	10	15	7	25	59	AI	3.07	0.0797	0.0498		hypothetical protein
Cluster_902	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		outer membrane transport family protein
Cluster_9782	5	10	15	7	25	59	AI	3.07	0.0797	0.0498		putative cytoplasmic protein
Cluster_99342	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		putative aec62

---

[a] Genes were grouped into gene clusters using a "stringent" clustering threshold of ≥90% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.9, while those that are "absent" have a TBLASTN LS-BSR <0.9.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

Table S4. Genes identified by LS-BSR that are significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for all 70 EPEC genomes analyzed using a stringent gene clustering threshold

Symptomatic (LI+NSI) vs. Asymptomatic (AI)												
	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI+NSI	AI	LI+NSI	AI	% LI+NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_34159	25	2	22	21	53	9	LI+NSI	11.09	0.0009	0.0003		conserved hypothetical protein
Cluster_138051	35	7	12	16	74	30	LI+NSI	10.71	0.0011	0.0006		lysis S family protein
Cluster_266373	35	7	12	16	74	30	LI+NSI	10.71	0.0011	0.0006		lysis S family protein
Cluster_34074	35	7	12	16	74	30	LI+NSI	10.71	0.0011	0.0006		lysis S family protein
Cluster_74390	23	2	24	21	49	9	LI+NSI	9.21	0.0024	0.0011		hypothetical protein
Cluster_4527	43	13	4	10	91	57	LI+NSI	9.72	0.0018	0.0012		tail attachment protein
Cluster_307965	25	3	22	20	53	13	LI+NSI	8.77	0.0031	0.0016		traP family protein
Cluster_88571	34	7	13	16	72	30	LI+NSI	9.52	0.0020	0.0016		regulatory protein cro
Cluster_322922	27	4	20	19	57	17	LI+NSI	8.48	0.0036	0.0020		lysis S family protein
Cluster_327168	33	7	14	16	70	30	LI+NSI	8.42	0.0037	0.0022		helix-turn-helix domain protein
Cluster_272941	31	6	16	17	66	26	LI+NSI	8.32	0.0039	0.0023		phage tail E family protein
Cluster_74420	31	6	16	17	66	26	LI+NSI	8.32	0.0039	0.0023		conserved hypothetical protein
Cluster_12327	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		putative prophage protein
Cluster_169965	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		ogr/Delta-like zinc finger family protein
Cluster_200453	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		putative derepression protein
Cluster_271578	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		conserved hypothetical protein
Cluster_55403	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		helix-turn-helix domain protein
Cluster_57003	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		putative phage immunity repressor protein
Cluster_262069	21	2	26	21	45	9	LI+NSI	7.51	0.0061	0.0027		istB-like ATP binding family protein
Cluster_289981	21	2	26	21	45	9	LI+NSI	7.51	0.0061	0.0027		integrase core domain protein
Cluster_296966	0	5	47	18	0	22	AI	7.97	0.0048	0.0028		conserved hypothetical protein
Cluster_310837	0	5	47	18	0	22	AI	7.97	0.0048	0.0028	*kilR*	killing protein KilR
Cluster_318890	0	5	47	18	0	22	AI	7.97	0.0048	0.0028		nucleoside transporter family protein
Cluster_318894	0	5	47	18	0	22	AI	7.97	0.0048	0.0028		lysR substrate binding domain protein
Cluster_147543	14	0	33	23	30	0	LI+NSI	6.80	0.0091	0.0029		putative yfdA
Cluster_237554	14	0	33	23	30	0	LI+NSI	6.80	0.0091	0.0029		conserved hypothetical protein
Cluster_128567	42	13	5	10	89	57	LI+NSI	8.04	0.0046	0.0038		phage portal protein, lambda family
Cluster_40177	42	13	5	10	89	57	LI+NSI	8.04	0.0046	0.0038		major tail V domain protein
Cluster_78687	42	13	5	10	89	57	LI+NSI	8.04	0.0046	0.0038		minor tail protein U
Cluster_297448	21	19	26	4	45	83	AI	7.59	0.0059	0.0042		conserved hypothetical protein
Cluster_127817	46	17	1	6	98	74	LI+NSI	7.37	0.0066	0.0042	*repA*	incFII family plasmid replication initiator RepA
Cluster_178274	46	17	1	6	98	74	LI+NSI	7.37	0.0066	0.0042		conserved hypothetical protein
Cluster_326414	1	6	46	17	2	26	AI	7.37	0.0066	0.0042		dinI-like family protein
Cluster_103685	45	16	2	7	96	70	LI+NSI	7.25	0.0071	0.0044		hypothetical protein
Cluster_185869	2	7	45	16	4	30	AI	7.25	0.0071	0.0044		transposase family protein
Cluster_316814	2	7	45	16	4	30	AI	7.25	0.0071	0.0044		AAA domain protein
Cluster_94384	45	16	2	7	96	70	LI+NSI	7.25	0.0071	0.0044		transposase, Mutator family protein
Cluster_261591	32	7	15	16	68	30	LI+NSI	7.41	0.0065	0.0045		methyltransferase domain protein
Cluster_68405	19	18	28	5	40	78	AI	7.42	0.0065	0.0046	*fimA*	type-1 fimbrial protein, A chain
Cluster_342345	25	4	22	19	53	17	LI+NSI	6.75	0.0094	0.0049		conserved hypothetical protein
Cluster_221391	20	2	27	21	43	9	LI+NSI	6.72	0.0095	0.0054		conserved hypothetical protein
Cluster_108598	27	5	20	18	57	22	LI+NSI	6.56	0.0104	0.0056		DNA-packaging protein FI
Cluster_14998	27	5	20	18	57	22	LI+NSI	6.56	0.0104	0.0056		bacteriophage replication gene A family protein
Cluster_70068	27	5	20	18	57	22	LI+NSI	6.56	0.0104	0.0056		conserved hypothetical protein
Cluster_188375	37	10	10	13	79	43	LI+NSI	7.17	0.0074	0.0060		phage major capsid E family protein
Cluster_276264	19	2	28	21	40	9	LI+NSI	5.97	0.0146	0.0063		replication regulatory RepB family protein
Cluster_108599	43	14	4	9	91	61	LI+NSI	7.66	0.0057	0.0065		tail attachment protein
Cluster_237572	43	14	4	9	91	61	LI+NSI	7.66	0.0057	0.0065		head decoration protein
Cluster_347409	41	13	6	10	87	57	LI+NSI	6.61	0.0101	0.0065		phage tail sheath family protein
Cluster_59689	22	3	25	20	47	13	LI+NSI	6.27	0.0123	0.0074		putative incFII replication protein RepA3
Cluster_166870	36	10	11	13	77	43	LI+NSI	6.12	0.0134	0.0083		integron integrase family protein
Cluster_248097	24	4	23	19	51	17	LI+NSI	5.96	0.0146	0.0092		conserved hypothetical protein
Cluster_151647	42	14	5	9	89	61	LI+NSI	6.16	0.0131	0.0094		phage major capsid E family protein
Cluster_294187	42	14	5	9	89	61	LI+NSI	6.16	0.0131	0.0094		conserved hypothetical protein
Cluster_117195	47	19	0	4	100	83	LI+NSI	5.74	0.0166	0.0097		conserved hypothetical protein
Cluster_168917	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		nucleoside 2-deoxyribosyltransferase family protein
Cluster_171515	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		putative lipoprotein
Cluster_171516	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		methyltransferase domain protein
Cluster_171517	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		beta-ketoacyl synthase, N-terminal domain protein
Cluster_171518	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		acyltransferase family protein
Cluster_171521	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		putative membrane protein
Cluster_171522	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		AMP-binding enzyme family protein
Cluster_171524	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		glycosyl transferase 2 family protein
Cluster_171525	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		thioesterase-like superfamily protein
Cluster_171526	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		outer membrane lipocarrier LolA family protein
Cluster_171527	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		MMPL family protein
Cluster_171529	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		beta-ketoacyl synthase, N-terminal domain protein
Cluster_172243	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		phage integrase family protein
Cluster_193190	47	19	0	4	100	83	LI+NSI	5.74	0.0166	0.0097		nlpC/P60 family protein
Cluster_206025	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		phosphopantetheine attachment site family protein
Cluster_206026	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		phosphopantetheine attachment site family protein
Cluster_206029	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein
Cluster_206034	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein
Cluster_206036	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		fabA-like domain protein
Cluster_206037	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		KR domain protein
Cluster_206038	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		beta-ketoacyl synthase, N-terminal domain protein
Cluster_206099	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein
Cluster_207740	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		caudovirales tail fibre assembly family protein
Cluster_321885	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein
Cluster_321886	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		hNH endonuclease family protein
Cluster_331430	47	19	0	4	100	83	LI+NSI	5.74	0.0166	0.0097		conserved hypothetical protein
Cluster_341403	47	19	0	4	100	83	LI+NSI	5.74	0.0166	0.0097	*lyx*	L-xylulose/3-keto-L-gulonate kinase
Cluster_345362	47	19	0	4	100	83	LI+NSI	5.74	0.0166	0.0097	*araD*	L-ribulose-5-phosphate 4-epimerase
Cluster_72173	47	19	0	4	100	83	LI+NSI	5.74	0.0166	0.0097		xylose isomerase-like TIM barrel family protein
Cluster_78310	47	19	0	4	100	83	LI+NSI	5.74	0.0166	0.0097		peptidase M23 family protein
Cluster_221813	34	9	13	14	72	39	LI+NSI	5.86	0.0155	0.0098		conserved hypothetical protein
Cluster_264826	34	9	13	14	72	39	LI+NSI	5.86	0.0155	0.0098		yagB/YeeU/YfjZ family protein
Cluster_225861	8	11	39	12	17	48	AI	5.93	0.0148	0.0100	*psu*	polarity suppression protein
Cluster_20103	29	6	18	17	62	26	LI+NSI	6.48	0.0109	0.0101		outer membrane transport family protein
Cluster_342998	29	6	18	17	62	26	LI+NSI	6.48	0.0109	0.0101		phage tail protein I
Cluster_103416	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		conserved hypothetical protein
Cluster_248119	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		dnaB-like helicase N terminal domain protein
Cluster_302692	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		hypothetical protein
Cluster_325003	21	18	26	5	45	78	AI	5.76	0.0164	0.0105		conserved hypothetical protein
Cluster_342342	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		conserved hypothetical protein
Cluster_342343	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		conserved hypothetical protein
Cluster_46843	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		dnaB-like helicase N terminal domain protein
Cluster_145816	9	12	38	11	19	52	AI	6.52	0.0106	0.0111		putative glyco3, capsid size determination protein Sid

Cluster_203385	9	12	38	11	19	52	AI	6.52	0.0106	0.0111		conserved hypothetical family protein
Cluster_267179	38	11	9	12	81	48	LI+NSI	6.52	0.0106	0.0111		conserved hypothetical protein
Cluster_109753	17	16	30	7	36	70	AI	5.64	0.0176	0.0113		outer membrane transport family protein
Cluster_902	17	16	30	7	36	70	AI	5.64	0.0176	0.0113		outer membrane transport family protein
Cluster_118881	44	16	3	7	94	70	LI+NSI	5.46	0.0194	0.0115		transposase, Mutator family protein
Cluster_8918	3	7	44	16	6	30	AI	5.46	0.0194	0.0115	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_44635	36	23	11	0	77	100	AI	4.74	0.0294	0.0122		type III secretion , HrpO family protein
Cluster_173657	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein
Cluster_191931	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		lysR substrate binding domain protein
Cluster_191933	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-oxoacid CoA-transferase, B subunit
Cluster_191934	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		acetyl-CoA C-acetyltransferase family protein
Cluster_191935	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-hydroxyacyl-CoA dehydrogenase, NAD binding domain protein
Cluster_210976	2	6	45	17	4	26	AI	5.27	0.0216	0.0128		putative prophage protein
Cluster_2127	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-hydroxybutyrate dehydrogenase family protein
Cluster_230986	2	6	45	17	4	26	AI	5.27	0.0216	0.0128	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_234924	46	18	1	5	98	78	LI+NSI	5.28	0.0215	0.0128		orn/Lys/Arg decarboxylase, C-terminal domain protein
Cluster_247805	46	18	1	5	98	78	LI+NSI	5.28	0.0215	0.0128		glucosamine-6-phosphate isomerases/6-phosphogluconolactonase family protein
Cluster_316170	2	6	45	17	4	26	AI	5.27	0.0216	0.0128		putative predicted protein
Cluster_316818	2	6	45	17	4	26	AI	5.27	0.0216	0.0128		putative predicted protein
Cluster_324271	2	6	45	17	4	26	AI	5.27	0.0216	0.0128	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_3723	2	6	45	17	4	26	AI	5.27	0.0216	0.0128		conserved hypothetical protein
Cluster_38989	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein
Cluster_49492	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		acetyltransferase family protein
Cluster_50317	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein
Cluster_51720	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-oxoacid CoA-transferase, A subunit
Cluster_51724	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		citrate transporter family protein
Cluster_7049	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein
Cluster_79228	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein
Cluster_271379	32	22	15	1	68	96	AI	5.18	0.0228	0.0135		glycosyl hydrolase family 65, N-terminal domain protein
Cluster_137907	10	12	37	11	21	52	AI	5.48	0.0192	0.0135		hypothetical protein
Cluster_321224	4	8	43	15	9	35	AI	5.77	0.0163	0.0145		conserved hypothetical protein
Cluster_118990	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein
Cluster_128383	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein
Cluster_14537	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		viral (Super1) RNA helicase family protein
Cluster_222761	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122
Cluster_261666	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		AAA domain protein
Cluster_276697	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein
Cluster_292258	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein
Cluster_59324	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		putative acetyltransferase
Cluster_73497	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein
Cluster_226889	7	10	40	13	15	43	AI	5.40	0.0202	0.0160		putative P4-specific DNA primase
Cluster_330511	40	13	7	10	85	57	LI+NSI	5.40	0.0202	0.0160		phage major tail tube protein
Cluster_53578	40	13	7	10	85	57	LI+NSI	5.40	0.0202	0.0160		putative protein
Cluster_275878	35	10	12	13	74	43	LI+NSI	5.18	0.0228	0.0167		lysR substrate binding domain protein
Cluster_103347	23	4	24	19	49	17	LI+NSI	5.22	0.0223	0.0176		conserved hypothetical protein
Cluster_309396	23	4	24	19	49	17	LI+NSI	5.22	0.0223	0.0176		conserved hypothetical protein
Cluster_157040	33	9	14	14	70	39	LI+NSI	4.99	0.0255	0.0191		methyltransferase domain protein
Cluster_266630	22	4	25	19	47	17	LI+NSI	4.53	0.0332	0.0194		conserved hypothetical protein
Cluster_202509	25	5	22	18	53	22	LI+NSI	5.02	0.0251	0.0198		transposase C of IS166 homeodomain protein
Cluster_267342	25	5	22	18	53	22	LI+NSI	5.02	0.0251	0.0198		putative bacteriophage protein
Cluster_281686	25	5	22	18	53	22	LI+NSI	5.02	0.0251	0.0198		conserved hypothetical protein
Cluster_82579	20	17	27	6	43	74	AI	4.90	0.0268	0.0211	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_82835	27	6	20	17	57	26	LI+NSI	4.90	0.0268	0.0211		conserved hypothetical protein
Cluster_11588	5	8	42	15	11	35	AI	4.46	0.0346	0.0223		conserved hypothetical protein
Cluster_147672	42	15	5	8	89	65	LI+NSI	4.46	0.0346	0.0223		terminase small subunit
Cluster_210977	5	8	42	15	11	35	AI	4.46	0.0346	0.0223		conserved hypothetical protein
Cluster_210978	5	8	42	15	11	35	AI	4.46	0.0346	0.0223		conserved hypothetical protein
Cluster_21749	42	15	5	8	89	65	LI+NSI	4.46	0.0346	0.0223		conserved hypothetical protein
Cluster_225286	5	8	42	15	11	35	AI	4.46	0.0346	0.0223		RHS Repeat family protein
Cluster_326026	5	8	42	15	11	35	AI	4.46	0.0346	0.0223		fimbrial family protein
Cluster_39211	42	15	5	8	89	65	LI+NSI	4.46	0.0346	0.0223		transposase family protein
Cluster_217504	24	5	23	18	51	22	LI+NSI	4.33	0.0374	0.0224		istB-like ATP binding family protein
Cluster_232396	9	11	38	12	19	48	AI	4.90	0.0269	0.0227		putative dNA-binding helix-turn-helix protein
Cluster_4778	9	11	38	12	19	48	AI	4.90	0.0269	0.0227		conserved hypothetical protein
Cluster_7938	38	12	9	11	81	52	LI+NSI	4.90	0.0269	0.0227		DNA methylase family protein
Cluster_153717	8	10	39	13	17	43	AI	4.36	0.0368	0.0228	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_221865	39	13	8	10	83	57	LI+NSI	4.36	0.0368	0.0228		bacterial type II and III secretion system family protein
Cluster_261646	39	13	8	10	83	57	LI+NSI	4.36	0.0368	0.0228		phage/conjugal plasmid C-4 type zinc finger , TraR family protein
Cluster_297203	8	10	39	13	17	43	AI	4.36	0.0368	0.0228		conserved hypothetical protein
Cluster_31976	8	10	39	13	17	43	AI	4.36	0.0368	0.0228	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_82353	39	13	8	10	83	57	LI+NSI	4.36	0.0368	0.0228		phage late control gene D family protein
Cluster_104053	26	6	21	17	55	26	LI+NSI	4.20	0.0403	0.0244		conserved hypothetical protein
Cluster_183342	21	17	26	6	45	74	AI	4.20	0.0403	0.0244		hypothetical protein
Cluster_242117	26	6	21	17	55	26	LI+NSI	4.20	0.0403	0.0244		antidote-toxin recognition MazE family protein
Cluster_261668	26	6	21	17	55	26	LI+NSI	4.20	0.0403	0.0244		antidote-toxin recognition MazE family protein
Cluster_347632	26	6	21	17	55	26	LI+NSI	4.20	0.0403	0.0244		antidote-toxin recognition MazE family protein
Cluster_108649	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_108651	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_114691	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_119046	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative tail fiber protein
Cluster_125122	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative predicted protein
Cluster_127445	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_131841	38	23	9	0	81	100	AI	3.49	0.0618	0.0254	*fryC*	fructose-like permease IIC component 1
Cluster_152085	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_163907	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		glycosyltransferase Family 4 family protein
Cluster_163983	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_164950	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		UTRA domain protein
Cluster_166190	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		phage integrase family protein
Cluster_239410	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		bacteriophage replication gene A family protein
Cluster_261875	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		hypothetical protein
Cluster_262139	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_263945	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_263946	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		phage major capsid E family protein
Cluster_263948	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_266287	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		hypothetical protein
Cluster_272862	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		helix-turn-helix domain protein
Cluster_291165	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		phage integrase family protein
Cluster_291297	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		glycosyl transferase 2 family protein
Cluster_345268	38	23	9	0	81	100	AI	3.49	0.0618	0.0254	*fryC*	fructose-like permease IIC component 1
Cluster_49147	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative phage protein
Cluster_69484	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		DNA N-6-adenine-methyltransferase family protein
Cluster_69487	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_69489	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_71993	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative membrane protein

Cluster_71994	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		glycosyl transferase 11 family protein
Cluster_71995	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain III family protein
Cluster_71996	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase
Cluster_72685	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		hypothetical protein
Cluster_73386	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		caudovirales tail fibre assembly family protein
Cluster_73840	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein
Cluster_103412	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		repressor protein C2
Cluster_17854	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		conserved hypothetical protein
Cluster_201731	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		helix-turn-helix, Psq domain protein
Cluster_201738	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		conserved hypothetical protein
Cluster_267010	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		putative pcar
Cluster_281878	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		helix-turn-helix domain protein
Cluster_48960	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		conserved hypothetical protein
Cluster_173601	41	14	6	9	87	61	LI+NSI	4.91	0.0268	0.0269	*folP*	dihydropteroate synthase
Cluster_198519	6	9	41	14	13	39	AI	4.91	0.0268	0.0269		transposase family protein
Cluster_268703	6	9	41	14	13	39	AI	4.91	0.0268	0.0269	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_276834	41	14	6	9	87	61	LI+NSI	4.91	0.0268	0.0269		clp protease family protein
Cluster_343399	41	14	6	9	87	61	LI+NSI	4.91	0.0268	0.0269		phage P2 GpU family protein
Cluster_108011	19	3	28	20	40	13	LI+NSI	4.18	0.0410	0.0279	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_11079	28	20	19	3	60	87	AI	4.18	0.0410	0.0279		transglycosylase SLT domain protein
Cluster_178309	19	3	28	20	40	13	LI+NSI	4.18	0.0410	0.0279		conserved hypothetical protein
Cluster_72094	19	3	28	20	40	13	LI+NSI	4.18	0.0410	0.0279		phage integrase family protein
Cluster_292597	36	11	11	12	77	48	LI+NSI	4.56	0.0327	0.0288		bacteriophage lambda tail assembly I family protein
Cluster_347183	36	11	11	12	77	48	LI+NSI	4.56	0.0327	0.0288		small Multidrug Resistance family protein
Cluster_151023	10	11	37	12	21	48	AI	4.00	0.0456	0.0292		conserved hypothetical protein
Cluster_271719	10	11	37	12	21	48	AI	4.00	0.0456	0.0292		conserved hypothetical protein
Cluster_3745	10	11	37	12	21	48	AI	4.00	0.0456	0.0292		aminotransferase class I and II family protein
Cluster_115550	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*iciA*	chromosome initiation inhibitor
Cluster_1253	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		glycosyl transferases group 1 family protein
Cluster_130922	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*cptA*	toxin CptA
Cluster_1317	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		phage tail-collar fibre family protein
Cluster_133750	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		VWA domain containing CoxE-like family protein
Cluster_134672	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*prfB*	peptide chain release factor 2
Cluster_141879	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		PKD domain protein
Cluster_151507	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_152658	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*mtlA*	PTS system, mannitol-specific IIC component family protein
Cluster_170088	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_171551	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		bacterial extracellular solute-binding family protein
Cluster_171552	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		binding-protein-dependent transport system inner membrane component family protein
Cluster_171555	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		glycerophosphoryl diester phosphodiesterase family protein
Cluster_171556	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		ROK family protein
Cluster_171708	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_1719	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		glycosyl transferase 2 family protein
Cluster_173284	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_173286	0	3	47	20	0	13	AI	3.62	0.0571	0.0324	*pifC*	transcriptional repressor PifC
Cluster_186687	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*serA*	D-3-phosphoglycerate dehydrogenase
Cluster_191111	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_191116	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_191126	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		phage portal protein, lambda family
Cluster_191129	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_191132	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_191136	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		hypothetical protein
Cluster_206060	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		ABC transporter family protein
Cluster_206339	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		putative predicted protein
Cluster_207555	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		recF/RecN/SMC N terminal domain protein
Cluster_207556	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_207558	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_2076	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		recombinase family protein
Cluster_20847	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_22912	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		putative hydroxypyruvate isomerase YgbM
Cluster_233858	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*ygfI*	DNA-binding transcriptional regulator domain protein
Cluster_23603	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		putative acyltransferase YihG
Cluster_245044	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*ydiN*	inner membrane transport protein YdiN
Cluster_25913	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*ygfZ*	tRNA-modifying protein YgfZ
Cluster_262370	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*mscS*	small-conductance mechanosensitive channel
Cluster_264237	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*argO*	arginine exporter protein ArgO
Cluster_266695	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		glycosyl hydrolases 25 family protein
Cluster_272440	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		putative transcriptional regulator
Cluster_279315	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		flavodoxin
Cluster_283246	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*pgk*	phosphoglycerate kinase
Cluster_313114	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		histidine phosphatase super family protein
Cluster_315927	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*yiaJ*	HTH-type transcriptional regulator YiaJ
Cluster_321371	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		binding-protein-dependent transport system inner membrane component family protein
Cluster_327487	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		class II Aldolase and Adducin N-terminal domain protein
Cluster_33461	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*epd*	erythrose-4-phosphate dehydrogenase
Cluster_340496	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*xerD*	tyrosine recombinase XerD
Cluster_36336	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		cupin fold metallo, WbuC family protein
Cluster_36339	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		rmlD substrate binding domain protein
Cluster_36341	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		dihydrouridine synthase family protein
Cluster_36450	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		type III restriction enzyme, res subunit
Cluster_36837	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_37187	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		O-antigen ligase like membrane family protein
Cluster_38770	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_38775	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_3913	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_3917	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		putative cytochrome P450 51
Cluster_39213	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		polysaccharide deacetylase family protein
Cluster_3922	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_3924	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		caudovirus prohead protease family protein
Cluster_39244	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_39245	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_3926	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_39264	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		ead/Ea22-like family protein
Cluster_3934	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_39356	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		P63C domain protein
Cluster_39363	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_4140	0	3	47	20	0	13	AI	3.62	0.0571	0.0324	*espN*	putative t3SS effector protein EspN2-2
Cluster_44306	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*cptB*	antitoxin CptB
Cluster_4584	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_4585	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		ipaB/EvcA family protein
Cluster_4591	0	3	47	20	0	13	AI	3.62	0.0571	0.0324	*msbB*	lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
Cluster_4592	0	3	47	20	0	13	AI	3.62	0.0571	0.0324	*eptC*	phosphoethanolamine transferase EptC
Cluster_52846	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		3-hydroxyacyl-CoA dehydrogenase, NAD binding domain protein
Cluster_62079	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		HTH domain protein

Cluster_66109	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		AAA domain family protein
Cluster_66189	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		histidine biosynthesis family protein
Cluster_6886	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		transposase family protein
Cluster_6888	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		transposase IS66 family protein
Cluster_69114	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_69143	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_82015	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_84327	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_8866	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_8868	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		glycosyltransferase Family 4 family protein
Cluster_94403	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein
Cluster_96049	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324		DSBA-like thioredoxin domain protein
Cluster_96081	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*fbaA*	fructose-bisphosphate aldolase, class II
Cluster_97991	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*ghrA*	glyoxylate/hydroxypyruvate reductase A
Cluster_99789	47	20	0	3	100	87	LI+NSI	3.62	0.0571	0.0324	*rpiA*	ribose 5-phosphate isomerase A
Cluster_14843	43	16	4	7	91	70	LI+NSI	4.07	0.0436	0.0324		head-to-tail joining protein W
Cluster_179213	43	16	4	7	91	70	LI+NSI	4.07	0.0436	0.0324		conserved hypothetical protein
Cluster_183211	43	16	4	7	91	70	LI+NSI	4.07	0.0436	0.0324		DNA methylase family protein
Cluster_271588	4	7	43	16	9	30	AI	4.07	0.0436	0.0324	*intA*	prophage CP4-57 integrase
Cluster_43808	43	16	4	7	91	70	LI+NSI	4.07	0.0436	0.0324		type II/IV secretion system family protein
Cluster_64148	43	16	4	7	91	70	LI+NSI	4.07	0.0436	0.0324		hypothetical protein
Cluster_107631	34	10	13	13	72	43	LI+NSI	4.34	0.0372	0.0338		phage tail tape measure protein, TP901 family, core region
Cluster_175201	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339		fimbrial family protein
Cluster_175778	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339		homeo-like domain protein
Cluster_221767	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339	*nleA*	putative nleA1 protein
Cluster_233786	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339	*eae*	intimin
Cluster_309397	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339		winged helix-turn-helix DNA-binding family protein
Cluster_342620	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339		putative predicted protein
Cluster_53395	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339		haemolysin expression modulating family protein
Cluster_103321	40	14	7	9	85	61	LI+NSI	3.86	0.0494	0.0342		tonB family C-terminal domain protein
Cluster_227577	7	9	40	14	15	39	AI	3.86	0.0494	0.0342	*ssb*	single-stranded DNA-binding protein
Cluster_117516	45	18	2	5	96	78	LI+NSI	3.48	0.0620	0.0345		lysR substrate binding domain protein
Cluster_210145	2	5	45	18	4	22	AI	3.48	0.0620	0.0345		conserved hypothetical protein
Cluster_210210	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaC*	4-hydroxyphenylacetate 3-monooxygenase, reductase component
Cluster_210211	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaB*	4-hydroxyphenylacetate 3-monooxygenase, oxygenase component
Cluster_210215	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaH*	2-oxo-hepta-3-ene-1,7-dioic acid hydratase
Cluster_244107	45	18	2	5	96	78	LI+NSI	3.48	0.0620	0.0345		putative purine permease YgfU
Cluster_284907	45	18	2	5	96	78	LI+NSI	3.48	0.0620	0.0345	*bglH*	cryptic outer membrane porin BglH
Cluster_315512	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaA*	4-hydroxyphenylacetate catabolism regulatory protein HpaA
Cluster_315513	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaX*	4-hydroxyphenylacetate permease
Cluster_315514	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaI*	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase
Cluster_315518	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaE*	5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase
Cluster_315519	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpcE*	homoprotocatechuate catabolism bifunctional isomerase/decarboxylase
Cluster_315520	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaR*	homoprotocatechuate degradation operon regulator, HpaR
Cluster_322310	45	18	2	5	96	78	LI+NSI	3.48	0.0620	0.0345		conserved hypothetical protein
Cluster_336901	45	18	2	5	96	78	LI+NSI	3.48	0.0620	0.0345	*yiaO*	2,3-diketo-L-gulonate-binding periplasmic protein YiaO
Cluster_38037	2	5	45	18	4	22	AI	3.48	0.0620	0.0345	*hpaD*	3,4-dihydroxyphenylacetate 2,3-dioxygenase
Cluster_94313	2	5	45	18	4	22	AI	3.48	0.0620	0.0345		conserved hypothetical protein
Cluster_227923	35	11	12	12	74	48	LI+NSI	3.75	0.0527	0.0348		conserved hypothetical protein
Cluster_331872	35	11	12	12	74	48	LI+NSI	3.75	0.0527	0.0348		putative transposase
Cluster_346478	35	11	12	12	74	48	LI+NSI	3.75	0.0527	0.0348		conserved hypothetical protein
Cluster_150583	46	19	1	4	98	83	LI+NSI	3.37	0.0665	0.0372		transposase family protein
Cluster_212386	46	19	1	4	98	83	LI+NSI	3.37	0.0665	0.0372		sigma-54 interaction domain protein
Cluster_23668	46	19	1	4	98	83	LI+NSI	3.37	0.0665	0.0372	*allC*	allantoate amidohydrolase
Cluster_244058	46	19	1	4	98	83	LI+NSI	3.37	0.0665	0.0372		zinc-binding dehydrogenase family protein
Cluster_273174	46	19	1	4	98	83	LI+NSI	3.37	0.0665	0.0372	*cobC*	alpha-ribazole phosphatase
Cluster_315018	1	4	46	19	2	17	AI	3.37	0.0665	0.0372		conserved hypothetical protein
Cluster_81324	46	19	1	4	98	83	LI+NSI	3.37	0.0665	0.0372		PTS system, mannose/fructose/sorbose , IID component family protein
Cluster_89657	46	19	1	4	98	83	LI+NSI	3.37	0.0665	0.0372		bacterial extracellular solute-binding , 7 family protein
Cluster_83620	32	9	15	14	68	39	LI+NSI	4.21	0.0402	0.0376		bacteriophage lambda tail assembly I family protein
Cluster_195887	23	5	24	18	49	22	LI+NSI	3.69	0.0546	0.0385		conserved hypothetical protein
Cluster_211780	24	18	23	5	51	78	AI	3.69	0.0546	0.0385		conserved hypothetical protein
Cluster_228673	23	5	24	18	49	22	LI+NSI	3.69	0.0546	0.0385		phage regulatory Rha family protein
Cluster_337356	33	10	14	13	70	43	LI+NSI	3.60	0.0578	0.0391		aminoglycoside/hydroxyurea antibiotic resistance kinase family protein
Cluster_127340	30	8	17	15	64	35	LI+NSI	4.15	0.0418	0.0398		DDE_Tnp_1-associated family protein
Cluster_188245	30	8	17	15	64	35	LI+NSI	4.15	0.0418	0.0398		conserved hypothetical protein
Cluster_4832	30	8	17	15	64	35	LI+NSI	4.15	0.0418	0.0398		DNA methylase family protein
Cluster_88753	30	8	17	15	64	35	LI+NSI	4.15	0.0418	0.0398		conserved hypothetical protein
Cluster_152550	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		mbeD/MobD like family protein
Cluster_246778	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		conserved hypothetical protein
Cluster_289072	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		transposase DDE domain protein
Cluster_307827	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401	*mbeC*	mobilization protein MbeC
Cluster_72088	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		glycosyl transferases group 1 family protein
Cluster_72092	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		transposase DDE domain protein
Cluster_73491	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		PIN domain protein
Cluster_135838	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		PIN domain protein
Cluster_165317	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		conserved hypothetical protein
Cluster_266376	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		conserved hypothetical protein
Cluster_266615	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		putative yubO
Cluster_276693	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		PIN domain protein
Cluster_302107	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		conserved hypothetical protein
Cluster_337809	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		putative antigen 43 domain protein
Cluster_24448	31	9	16	14	66	39	LI+NSI	3.51	0.0610	0.0420		prophage minor tail Z family protein
Cluster_222333	27	7	20	16	57	30	LI+NSI	3.49	0.0616	0.0434		phage major capsid protein, HK97 family
Cluster_252976	27	7	20	16	57	30	LI+NSI	3.49	0.0616	0.0434		phage prohead protease, HK97 family
Cluster_59004	27	7	20	16	57	30	LI+NSI	3.49	0.0616	0.0434		phage prohead protease, HK97 family
Cluster_183184	29	8	18	15	62	35	LI+NSI	3.48	0.0623	0.0435		phage minor tail protein G
Cluster_147263	38	13	9	10	81	57	LI+NSI	3.47	0.0623	0.0455		methyltransferase domain protein
Cluster_153719	9	10	38	13	19	43	AI	3.47	0.0623	0.0455		glycosyl transferase 8 family protein
Cluster_326910	38	13	9	10	81	57	LI+NSI	3.47	0.0623	0.0455		hydantoinase/oxoprolinase family protein
Cluster_107544	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		tape measure domain protein
Cluster_107547	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		conserved hypothetical protein
Cluster_163913	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		polysaccharide biosynthesis family protein
Cluster_177488	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461	*fliC*	flagellin
Cluster_240829	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		hypothetical protein
Cluster_246425	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		conserved hypothetical protein
Cluster_261814	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		conserved hypothetical protein
Cluster_272852	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		conserved hypothetical protein
Cluster_275815	39	23	8	0	83	100	AI	2.90	0.0887	0.0461	*gmm*	GDP-mannose mannosyl hydrolase
Cluster_281736	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		putative orf4 protein
Cluster_291294	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		putative pyridoxamine 5-phosphate-dependent dehydrase
Cluster_326791	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		ogr/Delta-like zinc finger family protein
Cluster_348249	39	23	8	0	83	100	AI	2.90	0.0887	0.0461		neutral zinc metallopeptidase family protein
Cluster_73598	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		conserved hypothetical protein

Cluster_73829	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461	rmlD substrate binding domain protein
Cluster_73987	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461	hypothetical protein
Cluster_74186	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461	helix-turn-helix domain protein
Cluster_120672	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	conserved hypothetical protein
Cluster_14417	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	putative predicted protein
Cluster_201403	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	conserved hypothetical protein
Cluster_244238	35	22	12	1	74	96	AI	3.29	0.0697	0.0476	conserved hypothetical protein
Cluster_263937	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	conserved hypothetical protein
Cluster_278863	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	conserved hypothetical protein
Cluster_300133	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	phage integrase family protein
Cluster_4740	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	marR family protein
Cluster_56342	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	conserved hypothetical protein
Cluster_58271	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476	conserved hypothetical protein
Cluster_105805	42	16	5	7	89	70	LI+NSI	2.98	0.0842	0.0498	DNA methylase family protein
Cluster_147010	5	7	42	16	11	30	AI	2.98	0.0842	0.0498	MULE transposase domain protein
Cluster_147011	5	7	42	16	11	30	AI	2.98	0.0842	0.0498	conserved hypothetical protein
Cluster_147907	5	7	42	16	11	30	AI	2.98	0.0842	0.0498	HNH endonuclease family protein
Cluster_302972	42	16	5	7	89	70	LI+NSI	2.98	0.0842	0.0498	DNA methylase family protein

Lethal (LI) vs. Non-Lethal (NSI+AI)												
	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI	NSI+AI	LI	NSI+AI	% LI	% NSI+AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_101055	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*prc*	tail-specific protease
Cluster_103685	24	37	0	9	100	80	LI	3.78	0.0518	0.0230		hypothetical protein
Cluster_10467	17	20	7	26	71	43	LI	3.70	0.0543	0.0434		phage Terminase family protein
Cluster_104831	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		binding-protein-dependent transport system inner membrane component family protein
Cluster_104965	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		transposase family protein
Cluster_105254	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*aceF*	dihydrolipoyllysine-residue acetyltransferase
Cluster_105805	23	35	1	11	96	76	LI	3.05	0.0807	0.0471		DNA methylase family protein
Cluster_108011	13	9	11	37	54	20	LI	7.23	0.0072	0.0059	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_108037	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*baeS*	signal transduction histidine-protein kinase BaeS
Cluster_108754	17	20	7	26	71	43	LI	3.70	0.0543	0.0434		phage terminase, small subunit, P27 family
Cluster_11079	11	37	13	9	46	80	NSI+AI	7.23	0.0072	0.0059		transglycosylase SLT domain protein
Cluster_113457	14	39	10	7	58	85	NSI+AI	4.65	0.0311	0.0201	*fimF*	protein FimF
Cluster_113721	7	30	17	16	29	65	NSI+AI	6.84	0.0089	0.0056		phage tail tape measure protein, lambda family
Cluster_115306	7	4	17	42	29	9	LI	3.56	0.0590	0.0381		putative membrane protein
Cluster_11780	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*ntrC*	nitrogen regulation protein NR
Cluster_118562	8	5	16	41	33	11	LI	3.88	0.0488	0.0482		istB-like ATP binding family protein
Cluster_120204	13	13	11	33	54	28	LI	3.49	0.0617	0.0406		conserved hypothetical protein
Cluster_120557	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*surE*	5'/3'-nucleotidase SurE
Cluster_125122	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		putative predicted protein
Cluster_12646	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*rstB*	sensor protein RstB
Cluster_12875	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*narU*	nitrite extrusion 2 domain protein
Cluster_13363	5	2	19	44	21	4	LI	3.11	0.0780	0.0422	*cdtB*	cytolethal distending toxin subunit B
Cluster_135012	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		gene 25-like lysozyme family protein
Cluster_135282	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*rplQ*	ribosomal protein L17
Cluster_136944	17	42	7	4	71	91	NSI+AI	3.56	0.0590	0.0381		phage terminase large subunit family protein
Cluster_136953	21	28	3	18	88	61	LI	4.13	0.0420	0.0279		haemolysin expression modulating family protein
Cluster_137178	4	21	20	25	17	46	NSI+AI	4.58	0.0324	0.0193		hypothetical protein
Cluster_13962	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		PLD-like domain protein
Cluster_14417	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		putative predicted protein
Cluster_14423	10	5	14	41	42	11	LI	7.15	0.0075	0.0051		conserved hypothetical protein
Cluster_147686	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein
Cluster_147687	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		putative membrane protein
Cluster_14793	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_14794	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_14795	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_14796	3	0	21	46	13	0	LI	3.35	0.0673	0.0370	*dam*	DNA adenine methylase family protein
Cluster_14797	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		helix-turn-helix family protein
Cluster_14800	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_148189	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		cytolethal distending toxin A/C family protein
Cluster_148976	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_148979	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative gp16
Cluster_14998	17	15	7	31	71	33	LI	7.81	0.0052	0.0049		bacteriophage replication gene A family protein
Cluster_15006	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		phage major capsid protein, P2 family
Cluster_15064	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_152420	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		phage major capsid E family protein
Cluster_153994	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		conserved hypothetical protein
Cluster_155747	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*phnH*	phosphonate C-P lyase system protein PhnH
Cluster_163350	2	14	22	32	8	30	NSI+AI	3.21	0.0734	0.0410		conserved hypothetical protein
Cluster_166190	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		phage integrase family protein
Cluster_173822	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein
Cluster_174674	7	4	17	42	29	9	LI	3.56	0.0590	0.0381		conserved hypothetical protein
Cluster_175599	5	2	19	44	21	4	LI	3.11	0.0780	0.0422	*rfbD*	dTDP-4-dehydrorhamnose reductase
Cluster_175605	6	0	18	46	25	0	LI	9.59	0.0020	0.0010	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_175612	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		transposase DDE domain protein
Cluster_175618	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		polysaccharide biosynthesis/export family protein
Cluster_177488	6	2	18	44	25	4	LI	4.76	0.0291	0.0165	*fliC*	flagellin
Cluster_181608	20	26	4	20	83	57	LI	3.91	0.0479	0.0339		transposase family protein
Cluster_182116	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein
Cluster_183020	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		hypothetical protein
Cluster_183172	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		DDE domain protein
Cluster_183173	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		transposase family protein
Cluster_183381	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		subtilase family protein
Cluster_183988	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		putative transposase
Cluster_184733	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*entS*	enterobactin exporter EntS
Cluster_188245	18	20	6	26	75	43	LI	5.11	0.0238	0.0219		conserved hypothetical protein
Cluster_188256	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		conserved hypothetical protein
Cluster_195799	1	11	23	35	4	24	NSI+AI	3.05	0.0807	0.0471		putative icd-like protein
Cluster_195951	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370		histidine kinase family protein
Cluster_196644	17	46	7	0	71	100	NSI+AI	11.84	0.0006	0.0003	*hscA*	fe-S protein assembly chaperone HscA
Cluster_196649	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*sseA*	3-mercaptopyruvate sulfurtransferase
Cluster_197373	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*emrB*	multidrug resistance protein B
Cluster_201168	18	45	6	1	75	98	NSI+AI	6.77	0.0093	0.0055	*eutK*	ethanolamine utilization protein EutK
Cluster_201403	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		conserved hypothetical protein
Cluster_201731	10	4	14	42	42	9	LI	8.75	0.0031	0.0031		helix-turn-helix, Psq domain protein
Cluster_201738	10	4	14	42	42	9	LI	8.75	0.0031	0.0031		conserved hypothetical protein
Cluster_20292	19	44	5	2	79	96	NSI+AI	3.11	0.0780	0.0422		acetyl xylan esterase family protein
Cluster_208031	18	44	6	2	75	96	NSI+AI	4.76	0.0291	0.0165		substrate binding domain of ABC-type glycine betaine transport system family protein
Cluster_208660	18	45	6	1	75	98	NSI+AI	6.77	0.0093	0.0055	*dnaX*	DNA polymerase III, subunit gamma and tau
Cluster_208864	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*hypF*	carbamoyltransferase HypF
Cluster_209349	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*tnaA*	tryptophanase

Cluster_209403	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*selB*	selenocysteine-specific translation elongation factor
Cluster_209404	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116		iron-containing alcohol dehydrogenase family protein
Cluster_211001	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		phage virion morphogenesis protein
Cluster_21139	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*caiD*	carnitinyl-CoA dehydratase
Cluster_211547	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*dnaE*	DNA polymerase III, alpha subunit
Cluster_21586	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*citG*	triphosphoribosyl-dephospho-CoA synthase CitG
Cluster_21640	18	46	6	0	75	100	NSI+AI	9.59	0.0020	0.0010	*nagE*	PTS system, N-acetylglucosamine-specific IIBC component
Cluster_21650	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*pgm*	phosphoglucomutase, alpha-D-glucose phosphate-specific
Cluster_21657	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*kdpB*	K+-transporting ATPase, B subunit
Cluster_217459	0	10	24	36	0	22	NSI+AI	4.44	0.0351	0.0124		HTH domain protein
Cluster_21749	23	34	1	12	96	74	LI	3.67	0.0555	0.0270		conserved hypothetical protein
Cluster_218595	8	5	16	41	33	11	LI	3.88	0.0488	0.0482	*traJ*	protein TraJ
Cluster_218615	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		alpha-2-macroglobulin family protein
Cluster_21870	19	44	5	2	79	96	NSI+AI	3.11	0.0780	0.0422	*ssb*	single-stranded DNA-binding protein
Cluster_218768	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*dacB*	D-alanyl-D-alanine carboxypeptidase/D-alanyl-D-alanine-endopeptidase
Cluster_219846	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*rutD*	pyrimidine utilization protein D
Cluster_220312	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		baseplate J-like family protein
Cluster_220402	19	44	5	2	79	96	NSI+AI	3.11	0.0780	0.0422	*alr*	alanine racemase
Cluster_220923	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*nuoN*	NADH-quinone oxidoreductase subunit N
Cluster_22096	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035		AAA-like domain protein
Cluster_221263	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein
Cluster_221350	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*ftsY*	signal recognition particle-docking protein FtsY
Cluster_221365	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*ugpC*	sn-glycerol-3-phosphate import ATP-binding protein UgpC
Cluster_221813	19	24	5	22	79	52	LI	3.78	0.0519	0.0386		conserved hypothetical protein
Cluster_221865	22	30	2	16	92	65	LI	4.47	0.0344	0.0207		bacterial type II and III secretion system family protein
Cluster_222292	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein
Cluster_222331	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		phage portal protein, HK97 family
Cluster_222333	17	17	7	29	71	37	LI	5.95	0.0147	0.0113		phage major capsid protein, HK97 family
Cluster_222669	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_222670	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_222671	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_222677	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_222678	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		hypothetical protein
Cluster_222679	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative yahA protein
Cluster_222748	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative ccgAII
Cluster_222761	9	6	15	40	38	13	LI	4.24	0.0394	0.0300		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122
Cluster_222808	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		polysaccharide biosynthesis/export family protein
Cluster_224504	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*rodZ*	cytoskeleton protein RodZ
Cluster_230143	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_230280	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		taurine catabolism dioxygenase TauD, TfdA family protein
Cluster_234123	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*tonB*	protein TonB
Cluster_23468	18	21	6	25	75	46	LI	4.38	0.0364	0.0238		phage capsid scaffolding (GPO) serine peptidase family protein
Cluster_23480	15	17	9	29	63	37	LI	3.18	0.0745	0.0482		conserved hypothetical protein
Cluster_236094	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		insA N-terminal domain protein
Cluster_236134	16	14	8	32	67	30	LI	7.04	0.0080	0.0052	*dhfrI*	dihydrofolate reductase type 1
Cluster_236296	23	32	1	14	96	70	LI	5.00	0.0254	0.0130		ATPase subunit of terminase family protein
Cluster_236458	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*uvrB*	excinuclease ABC subunit B
Cluster_237374	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*proS*	proline--tRNA ligase
Cluster_237804	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035		nitrate reductase, alpha subunit
Cluster_23840	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		glycosyl transferase family 2 domain protein
Cluster_239025	7	3	17	43	29	7	LI	4.88	0.0271	0.0256		phage terminase, small subunit, P27 family
Cluster_239026	7	3	17	43	29	7	LI	4.88	0.0271	0.0256		phage Terminase family protein
Cluster_239028	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		phage portal protein, HK97 family
Cluster_240829	7	1	17	45	29	2	LI	8.84	0.0029	0.0018		hypothetical protein
Cluster_24188	21	29	3	17	88	63	LI	3.50	0.0613	0.0496		conserved hypothetical protein
Cluster_242402	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative membrane protein
Cluster_24313	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		polysaccharide biosynthesis family protein
Cluster_24314	5	0	19	46	21	0	LI	7.42	0.0065	0.0035		degT/DnrJ/EryC1/StrS aminotransferase family protein
Cluster_24337	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		phage portal protein, lambda family
Cluster_24348	5	29	19	17	21	63	NSI+AI	9.62	0.0019	0.0010		phage tail tape measure protein, lambda family
Cluster_245108	18	45	6	1	75	98	NSI+AI	6.77	0.0093	0.0055		electron transport complex, RnfABCDGE type, B subunit
Cluster_245260	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*cheA*	chemotaxis protein CheA
Cluster_246425	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		conserved hypothetical protein
Cluster_246762	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*ligB*	DNA ligase B
Cluster_246821	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		helix-turn-helix domain protein
Cluster_247242	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_247243	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		cobQ/CobB/MinD/ParA nucleotide binding domain protein
Cluster_247315	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_247317	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		toxin-antitoxin system, antitoxin component, Xre family
Cluster_247403	5	0	19	46	21	0	LI	7.42	0.0065	0.0035		glycosyl transferase 2 family protein
Cluster_247837	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		putative domain protein
Cluster_24791	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*yjiY*	inner membrane protein YjiY
Cluster_248983	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*pepB*	peptidase B
Cluster_249336	13	13	11	33	54	28	LI	3.49	0.0617	0.0406		zonular occludens toxin family protein
Cluster_252396	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		tail attachment protein
Cluster_252971	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		HNH endonuclease family protein
Cluster_252976	17	17	7	29	71	37	LI	5.95	0.0147	0.0113		phage prohead protease, HK97 family
Cluster_254547	17	19	7	27	71	41	LI	4.39	0.0362	0.0245		bacteriophage lysis family protein
Cluster_257914	24	35	0	11	100	76	LI	5.12	0.0236	0.0120	*cadC*	transcriptional activator CadC
Cluster_258318	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370		MFS transporter, phenyl propionate permease family protein
Cluster_258452	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*malP*	maltodextrin phosphorylase
Cluster_259070	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*mfd*	transcription-repair coupling factor
Cluster_259563	18	45	6	1	75	98	NSI+AI	6.77	0.0093	0.0055		MASE1 family protein
Cluster_261646	22	30	2	16	92	65	LI	4.47	0.0344	0.0207		phage/conjugal plasmid C-4 type zinc finger , TraR family protein
Cluster_261814	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		conserved hypothetical protein
Cluster_262069	12	11	12	35	50	24	LI	3.75	0.0527	0.0348		istB-like ATP binding family protein
Cluster_263937	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		conserved hypothetical protein
Cluster_263945	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		conserved hypothetical protein
Cluster_263946	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		phage major capsid E family protein
Cluster_263948	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		conserved hypothetical protein
Cluster_264521	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035		conserved hypothetical protein
Cluster_264826	19	24	5	22	79	52	LI	3.78	0.0519	0.0386		yagB/YeeU/YfjZ family protein
Cluster_265225	18	46	6	0	75	100	NSI+AI	9.59	0.0020	0.0010		PRD domain protein
Cluster_265241	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*purL*	phosphoribosylformylglycinamidine synthase
Cluster_265334	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*mdfA*	multidrug transporter MdfA
Cluster_265879	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*hisG*	ATP phosphoribosyltransferase
Cluster_266191	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*yecS*	inner membrane amino-acid ABC transporter permease protein YecS
Cluster_266615	16	15	8	31	67	33	LI	6.10	0.0135	0.0107		putative yubO
Cluster_269540	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*can*	carbonic anhydrase 2
Cluster_271746	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		KR domain protein
Cluster_272941	19	18	5	28	79	39	LI	8.60	0.0034	0.0022		phage tail E family protein
Cluster_273680	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*menD*	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase
Cluster_274418	4	22	20	24	17	48	NSI+AI	5.29	0.0214	0.0179		mu-like prophage FluMu gp41 family protein
Cluster_275430	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*adiC*	arginine/agmatine antiporter

Cluster_275815	18	44	6	2	75	96	NSI+AI	4.76	0.0291	0.0165	*gmm*	GDP-mannose mannosyl hydrolase
Cluster_276292	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159		stress-induced bacterial acidophilic repeat motif family protein
Cluster_277010	20	26	4	20	83	57	LI	3.91	0.0479	0.0339		transposase DDE domain protein
Cluster_277146	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*rsmH*	16S rRNA (cytosine(1402)-N(4))-methyltransferase
Cluster_280171	18	46	6	0	75	100	NSI+AI	9.59	0.0020	0.0010	*tatD*	tat-linked quality control protein TatD
Cluster_282548	8	5	16	41	33	11	LI	3.88	0.0488	0.0482		putative predicted protein
Cluster_283709	19	44	5	2	79	96	NSI+AI	3.11	0.0780	0.0422	*acnA*	aconitate hydratase 1
Cluster_284853	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		AAA domain protein
Cluster_284866	7	4	17	42	29	9	LI	3.56	0.0590	0.0381		bacterial regulatory, tetR family protein
Cluster_286129	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		P2 phage tail completion R family protein
Cluster_286226	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*cybH*	ni/Fe-hydrogenase, b-type cytochrome subunit
Cluster_286879	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		conserved hypothetical protein
Cluster_286880	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		hypothetical protein
Cluster_286883	5	0	19	46	21	0	LI	7.42	0.0065	0.0035		acetyltransferase domain protein
Cluster_287268	11	9	13	37	46	20	LI	4.12	0.0423	0.0278		conserved hypothetical protein
Cluster_287304	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370		branched-chain amino acid transport system / permease component family protein
Cluster_28951	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		hypothetical protein
Cluster_289981	12	11	12	35	50	24	LI	3.75	0.0527	0.0348		integrase core domain protein
Cluster_290242	18	44	6	2	75	96	NSI+AI	4.76	0.0291	0.0165	*yphF*	ABC transporter periplasmic-binding protein YphF
Cluster_290560	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035		AAA-like domain protein
Cluster_291813	12	10	12	36	50	22	LI	4.61	0.0318	0.0285		conserved hypothetical protein
Cluster_294535	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*serB*	phosphoserine phosphatase SerB
Cluster_29582	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		HNH endonuclease family protein
Cluster_295885	18	45	6	1	75	98	NSI+AI	6.77	0.0093	0.0055		acetyltransferase domain protein
Cluster_296132	5	22	19	24	21	48	NSI+AI	3.78	0.0519	0.0386		ogr/Delta-like zinc finger family protein
Cluster_301782	5	22	19	24	21	48	NSI+AI	3.78	0.0519	0.0386		major tail sheath protein
Cluster_301785	4	22	20	24	17	48	NSI+AI	5.29	0.0214	0.0179		phage tail tape measure protein, TP901 family, core region
Cluster_305186	5	22	19	24	21	48	NSI+AI	3.78	0.0519	0.0386		phage P2 GpU family protein
Cluster_307627	2	14	22	32	8	30	NSI+AI	3.21	0.0734	0.0410		conserved hypothetical protein
Cluster_312709	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		amino acid permease family protein
Cluster_313189	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		putative predicted protein
Cluster_313761	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		HNH endonuclease family protein
Cluster_31447	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		phosphotransferase enzyme family protein
Cluster_314938	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*murE*	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase
Cluster_315743	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*ycdY*	chaperone protein YcdY
Cluster_315801	2	16	22	30	8	35	NSI+AI	4.47	0.0344	0.0207		phage tail protein I
Cluster_316170	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		putative predicted protein
Cluster_317131	1	13	23	33	4	28	NSI+AI	4.32	0.0378	0.0250		conserved hypothetical protein
Cluster_317219	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*shp*	head decoration protein
Cluster_32291	18	46	6	0	75	100	NSI+AI	9.59	0.0020	0.0010	*dcd*	deoxycytidine triphosphate deaminase
Cluster_322922	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		lysis S family protein
Cluster_324957	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		cytolethal distending toxin A/C family protein
Cluster_326910	22	29	2	17	92	63	LI	5.17	0.0230	0.0114		hydantoinase/oxoprolinase family protein
Cluster_327168	18	22	6	24	75	48	LI	3.71	0.0541	0.0418		helix-turn-helix domain protein
Cluster_328365	21	28	3	18	88	61	LI	4.13	0.0420	0.0279		hypothetical protein
Cluster_329513	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370		macB-like periplasmic core domain protein
Cluster_330511	22	31	2	15	92	67	LI	3.82	0.0506	0.0380		phage major tail tube protein
Cluster_332291	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		phage small terminase subunit
Cluster_333715	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*dinI*	DNA-damage-inducible protein I
Cluster_335571	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*cydA*	cytochrome d ubiquinol oxidase subunit 1
Cluster_335585	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035		ybgS-like family protein
Cluster_342277	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		ogr/Delta-like zinc finger family protein
Cluster_342430	21	28	3	18	88	61	LI	4.13	0.0420	0.0279		conserved hypothetical protein
Cluster_342998	17	18	7	28	71	39	LI	5.14	0.0234	0.0224		phage tail protein I
Cluster_343372	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*leuB*	3-isopropylmalate dehydrogenase
Cluster_343399	23	32	1	14	96	70	LI	5.00	0.0254	0.0130		phage P2 GpU family protein
Cluster_345283	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159		carbon-nitrogen hydrolase family protein
Cluster_346239	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		HNH endonuclease family protein
Cluster_346478	21	25	3	21	88	54	LI	6.29	0.0121	0.0074		conserved hypothetical protein
Cluster_347409	23	31	1	15	96	67	LI	5.71	0.0168	0.0069		phage tail sheath family protein
Cluster_347413	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		hok/gef family protein
Cluster_347784	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		phage holin family protein
Cluster_348353	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*yajR*	inner membrane transport protein YajR
Cluster_348774	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		D-alanyl-D-alanine carboxypeptidase family protein
Cluster_36312	2	15	22	31	8	33	NSI+AI	3.82	0.0506	0.0380		ethanolamine utilization - propanediol utilization family protein
Cluster_3723	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		conserved hypothetical protein
Cluster_43613	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		phage gp6-like head-tail connector family protein
Cluster_44635	17	42	7	4	71	91	NSI+AI	3.56	0.0590	0.0381		type III secretion , HrpO family protein
Cluster_46213	3	19	21	27	13	41	NSI+AI	4.81	0.0283	0.0157		putative transmembrane protein
Cluster_4706	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		minor tail protein U
Cluster_5054	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159		uracil-xanthine permease family protein
Cluster_51806	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*ptsP*	phosphoenolpyruvate-protein phosphotransferase
Cluster_52075	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		type I restriction modification DNA specificity domain protein
Cluster_5286	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*atpF*	ATP synthase F0, B subunit
Cluster_53578	22	31	2	15	92	67	LI	3.82	0.0506	0.0380		putative protein
Cluster_56841	13	13	11	33	54	28	LI	3.49	0.0617	0.0406		putative membrane protein
Cluster_57580	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370		tonB-dependent siderophore receptor family protein
Cluster_58271	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		conserved hypothetical protein
Cluster_59004	17	17	7	29	71	37	LI	5.95	0.0147	0.0113		phage prohead protease, HK97 family
Cluster_61005	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116		conserved hypothetical protein
Cluster_64413	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		phosphoadenosine phosphosulfate reductase family protein
Cluster_66669	7	4	17	42	29	9	LI	3.56	0.0590	0.0381		conserved hypothetical protein
Cluster_68745	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		phage portal protein, PBSX family
Cluster_70295	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein
Cluster_70300	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein
Cluster_73598	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		conserved hypothetical protein
Cluster_73967	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		phage Tail Protein X family protein
Cluster_74186	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		helix-turn-helix domain protein
Cluster_74247	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		dnaJ domain protein
Cluster_74249	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative yaiB protein
Cluster_74250	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative yajA protein
Cluster_74254	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_74420	19	18	5	28	79	39	LI	8.60	0.0034	0.0022		conserved hypothetical protein
Cluster_75095	5	22	19	24	21	48	NSI+AI	3.78	0.0519	0.0386		phage major tail tube protein
Cluster_75099	5	22	19	24	21	48	NSI+AI	3.78	0.0519	0.0386		phage late control gene D family protein
Cluster_77548	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		hsp70 family protein
Cluster_79151	22	31	2	15	92	67	LI	3.82	0.0506	0.0380		bundlin family protein
Cluster_79549	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*prpR*	propionate catabolism operon regulatory protein PrpR
Cluster_8046	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		hypothetical protein
Cluster_81331	18	46	6	0	75	100	NSI+AI	9.59	0.0020	0.0010		lppC lipofamily protein
Cluster_81348	20	46	4	0	83	100	NSI+AI	5.33	0.0209	0.0116	*pnp*	polyribonucleotide nucleotidyltransferase
Cluster_81842	19	44	5	2	79	96	NSI+AI	3.11	0.0780	0.0422		MATE efflux family protein
Cluster_81862	19	44	5	2	79	96	NSI+AI	3.11	0.0780	0.0422		conserved hypothetical protein
Cluster_82331	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*gsiC*	glutathione transport system permease protein GsiC

Cluster_82353	22	30	2	16	92	65	LI	4.47	0.0344	0.0207		phage late control gene D family protein
Cluster_82835	16	17	8	29	67	37	LI	4.46	0.0347	0.0240		conserved hypothetical protein
Cluster_83084	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*purD*	phosphoribosylamine--glycine ligase
Cluster_83254	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		glycosyl transferases group 1 family protein
Cluster_83414	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein
Cluster_83418	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		hypothetical protein
Cluster_83927	6	0	18	46	25	0	LI	9.59	0.0020	0.0010	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase
Cluster_83929	6	0	18	46	25	0	LI	9.59	0.0020	0.0010	*rfbB*	dTDP-glucose 4,6-dehydratase
Cluster_84211	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		conserved hypothetical protein
Cluster_84374	19	44	5	2	79	96	NSI+AI	3.11	0.0780	0.0422		transglycosylase SLT domain protein
Cluster_87728	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*phrB*	deoxyribodipyrimidine photo-lyase
Cluster_89114	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370	*trpD*	anthranilate synthase component II
Cluster_89392	21	46	3	0	88	100	NSI+AI	3.35	0.0673	0.0370		NADPH-dependent FMN reductase family protein
Cluster_90723	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*grxB*	glutaredoxin, GrxB family
Cluster_92542	0	9	24	37	0	20	NSI+AI	3.78	0.0518	0.0230	*espS*	putative t3SS effector protein EspS
Cluster_93726	19	46	5	0	79	100	NSI+AI	7.42	0.0065	0.0035	*tolA*	protein TolA
Cluster_97263	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*cadA*	cadmium-translocating P-type ATPase
Cluster_97497	19	45	5	1	79	98	NSI+AI	4.83	0.0280	0.0159	*tolA*	protein TolA
Cluster_99113	16	16	8	30	67	35	LI	5.24	0.0221	0.0133		putative membrane protein
Cluster_99362	14	13	10	33	58	28	LI	4.82	0.0282	0.0202		resolvase, N terminal domain protein

[a] Genes were grouped into gene clusters using a "stringent" clustering threshold of ≥90% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.9, while those that are "absent" have a TBLASTN LS-BSR <0.9.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

Table S5. Genes identified by LS-BSR that are significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for the 61 typical EPEC genomes analyzed using a stringent clustering threshold

Symptomatic (LI+NSI) vs. Asymptomatic (AI)

	No. of Genomes											
	Present (≥0.9)[a,b]		Absent (<0.9)[a,b]		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI+NSI	AI	LI+NSI	AI	% LI+NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_138051	32	4	12	13	73	24	LI+NSI	10.32	0.0013	0.0010		lysis S family protein
Cluster_266373	32	4	12	13	73	24	LI+NSI	10.32	0.0013	0.0010		lysis S family protein
Cluster_34074	32	4	12	13	73	24	LI+NSI	10.32	0.0013	0.0010		lysis S family protein
Cluster_296966	0	5	44	12	0	29	AI	10.46	0.0012	0.0010		conserved hypothetical protein
Cluster_310837	0	5	44	12	0	29	AI	10.46	0.0012	0.0010	*kilR*	killing protein KilR
Cluster_318890	0	5	44	12	0	29	AI	10.46	0.0012	0.0010		nucleoside transporter family protein
Cluster_318894	0	5	44	12	0	29	AI	10.46	0.0012	0.0010		lysR substrate binding domain protein
Cluster_185869	2	7	42	10	5	41	AI	10.33	0.0013	0.0011		transposase family protein
Cluster_178274	43	11	1	6	98	65	LI+NSI	10.11	0.0015	0.0013		conserved hypothetical protein
Cluster_261591	31	4	13	13	70	24	LI+NSI	9.21	0.0024	0.0013		methyltransferase domain protein
Cluster_321224	4	8	40	9	9	47	AI	8.91	0.0028	0.0021		conserved hypothetical protein
Cluster_198519	6	9	38	8	14	53	AI	8.21	0.0042	0.0028		transposase family protein
Cluster_268703	6	9	38	8	14	53	AI	8.21	0.0042	0.0028	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_153717	7	10	37	7	16	59	AI	9.20	0.0024	0.0028	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_31976	7	10	37	7	16	59	AI	9.20	0.0024	0.0028	*rfaY*	lipopolysaccharide core heptose(II) kinase RfaY
Cluster_4527	41	10	3	7	93	59	LI+NSI	8.20	0.0042	0.0031		tail attachment protein
Cluster_12327	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		putative prophage protein
Cluster_169965	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		ogr/Delta-like zinc finger family protein
Cluster_200453	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		putative derepression protein
Cluster_271578	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		conserved hypothetical protein
Cluster_55403	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		helix-turn-helix domain protein
Cluster_57003	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		putative phage immunity repressor protein
Cluster_34159	23	2	21	15	52	12	LI+NSI	6.73	0.0095	0.0041		conserved hypothetical protein
Cluster_230986	2	6	42	11	5	35	AI	7.66	0.0057	0.0043	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_168917	0	4	44	13	0	24	AI	7.57	0.0059	0.0046		nucleoside 2-deoxyribosyltransferase family protein
Cluster_321885	0	4	44	13	0	24	AI	7.57	0.0059	0.0046		conserved hypothetical protein
Cluster_321886	0	4	44	13	0	24	AI	7.57	0.0059	0.0046		hNH endonuclease family protein
Cluster_331430	44	13	0	4	100	76	LI+NSI	7.57	0.0059	0.0046		conserved hypothetical protein
Cluster_11588	5	8	39	9	11	47	AI	7.31	0.0069	0.0046		conserved hypothetical protein
Cluster_225286	5	8	39	9	11	47	AI	7.31	0.0069	0.0046		RHS Repeat family protein
Cluster_326026	5	8	39	9	11	47	AI	7.31	0.0069	0.0046		fimbrial family protein
Cluster_79228	1	5	43	12	2	29	AI	7.35	0.0067	0.0051		conserved hypothetical protein
Cluster_151023	10	11	34	6	23	65	AI	7.80	0.0052	0.0055		conserved hypothetical protein
Cluster_271719	10	11	34	6	23	65	AI	7.80	0.0052	0.0055		conserved hypothetical protein
Cluster_153719	9	10	35	7	20	59	AI	6.72	0.0095	0.0059		glycosyl transferase 8 family protein
Cluster_188375	35	7	9	10	80	41	LI+NSI	6.72	0.0095	0.0059		phage major capsid E family protein
Cluster_147543	14	0	30	17	32	0	LI+NSI	5.34	0.0209	0.0064		putative yfdA
Cluster_178542	14	0	30	17	32	0	LI+NSI	5.34	0.0209	0.0064	*epaR*	type III secretion apparatus protein SpaR/YscT/HrcT
Cluster_237554	14	0	30	17	32	0	LI+NSI	5.34	0.0209	0.0064		conserved hypothetical protein
Cluster_7621	14	0	30	17	32	0	LI+NSI	5.34	0.0209	0.0064		putative type III secretion EprI protein
Cluster_12200	11	11	33	6	25	65	AI	6.75	0.0094	0.0067		glycosyltransferase Family 4 family protein
Cluster_18923	11	11	33	6	25	65	AI	6.75	0.0094	0.0067		glycosyltransferase Family 4 family protein
Cluster_128567	40	10	4	7	91	59	LI+NSI	6.51	0.0107	0.0071		phage portal protein, lambda family
Cluster_302107	22	2	22	15	50	12	LI+NSI	6.00	0.0143	0.0080		conserved hypothetical protein
Cluster_99362	22	2	22	15	50	12	LI+NSI	6.00	0.0143	0.0080		resolvase, N terminal domain protein
Cluster_127340	28	4	16	13	64	24	LI+NSI	6.38	0.0115	0.0090		DDE_Tnp_1-associated family protein
Cluster_342998	28	4	16	13	64	24	LI+NSI	6.38	0.0115	0.0090		phage tail protein I
Cluster_322922	25	3	19	14	57	18	LI+NSI	6.08	0.0137	0.0091		lysis S family protein
Cluster_157040	32	6	12	11	73	35	LI+NSI	5.81	0.0159	0.0094		methyltransferase domain protein
Cluster_108599	41	11	3	6	93	65	LI+NSI	5.80	0.0160	0.0106		tail attachment protein
Cluster_151647	41	11	3	6	93	65	LI+NSI	5.80	0.0160	0.0106		phage major capsid E family protein
Cluster_237572	41	11	3	6	93	65	LI+NSI	5.80	0.0160	0.0106		head decoration protein
Cluster_253514	41	11	3	6	93	65	LI+NSI	5.80	0.0160	0.0106	*wcaF*	colanic acid biosynthesis acetyltransferase WcaF
Cluster_109753	17	13	27	4	39	76	AI	5.59	0.0181	0.0106		outer membrane transport family protein
Cluster_20103	27	4	17	13	61	24	LI+NSI	5.59	0.0181	0.0106		outer membrane transport family protein
Cluster_902	17	13	27	4	39	76	AI	5.59	0.0181	0.0106		outer membrane transport family protein
Cluster_145816	8	9	36	8	18	53	AI	5.74	0.0166	0.0109		putative glyco3, capsid size determination protein Sid
Cluster_111596	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		ethanolamine utilization EutN/carboxysome family protein
Cluster_111600	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		electron transport complex, RnfABCDGE type, C subunit
Cluster_131700	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*xseA*	exodeoxyribonuclease VII, large subunit
Cluster_196208	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*eutP*	ethanolamine utilization protein, EutP
Cluster_196229	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		putative cobalamin biosynthesis protein CbiG
Cluster_225773	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduD*	propanediol dehydratase medium subunit
Cluster_225778	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		BMC domain protein
Cluster_230774	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduL*	phosphate propanoyltransferase
Cluster_259786	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduC*	propanediol dehydratase large subunit
Cluster_259788	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduE*	propanediol dehydratase small subunit
Cluster_265851	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		BMC domain protein
Cluster_275755	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduU*	propanediol utilization protein PduU
Cluster_275766	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		dehydratase medium subunit
Cluster_291265	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		iron-containing alcohol dehydrogenase family protein
Cluster_291267	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduT*	shell protein
Cluster_307965	24	3	20	14	55	18	LI+NSI	5.35	0.0207	0.0109		traP family protein
Cluster_321135	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		helix-turn-helix domain protein
Cluster_321136	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		MIP channel family protein
Cluster_325632	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduA*	propanediol utilization protein PduA
Cluster_42758	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		cob(I)alamin adenosyltransferase
Cluster_42763	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduA*	propanediol utilization protein PduA
Cluster_218232	13	0	31	17	30	0	LI+NSI	4.74	0.0294	0.0125		conserved hypothetical protein
Cluster_234325	17	1	27	16	39	6	LI+NSI	4.85	0.0277	0.0127		bacterial extracellular solute-binding, 5 Middle family protein
Cluster_137907	10	10	34	7	23	59	AI	5.70	0.0169	0.0135		hypothetical protein
Cluster_148358	6	8	38	9	14	47	AI	5.97	0.0145	0.0140		putative exported protein
Cluster_142846	12	0	32	17	27	0	LI+NSI	4.17	0.0410	0.0142		conserved hypothetical protein
Cluster_182662	12	0	32	17	27	0	LI+NSI	4.17	0.0410	0.0142		putative membrane protein
Cluster_23022	12	0	32	17	27	0	LI+NSI	4.17	0.0410	0.0142		conserved hypothetical protein
Cluster_95290	12	0	32	17	27	0	LI+NSI	4.17	0.0410	0.0142		invH outer membrane lipofamily protein
Cluster_117516	42	12	2	5	95	71	LI+NSI	5.22	0.0224	0.0147		lysR substrate binding domain protein
Cluster_94313	2	5	42	12	5	29	AI	5.22	0.0224	0.0147		conserved hypothetical protein
Cluster_221767	21	2	23	15	48	12	LI+NSI	5.31	0.0212	0.0167	*nleA*	putative nleA1 protein
Cluster_233786	21	2	23	15	48	12	LI+NSI	5.31	0.0212	0.0167	*eae*	intimin
Cluster_74390	21	2	23	15	48	12	LI+NSI	5.31	0.0212	0.0167		hypothetical protein
Cluster_178488	20	2	24	15	45	12	LI+NSI	4.66	0.0308	0.0175	*nleE*	putative nleE
Cluster_262069	20	2	24	15	45	12	LI+NSI	4.66	0.0308	0.0175		istB-like ATP binding family protein
Cluster_289981	20	2	24	15	45	12	LI+NSI	4.66	0.0308	0.0175		integrase core domain protein
Cluster_313775	33	7	11	10	75	41	LI+NSI	4.81	0.0284	0.0178		outer membrane autotransporter barrel domain protein
Cluster_103685	43	13	1	4	98	76	LI+NSI	4.81	0.0283	0.0186		hypothetical protein
Cluster_212386	43	13	1	4	98	76	LI+NSI	4.81	0.0283	0.0186		sigma-54 interaction domain protein
Cluster_23668	43	13	1	4	98	76	LI+NSI	4.81	0.0283	0.0186	*allC*	allantoate amidohydrolase
Cluster_273174	43	13	1	4	98	76	LI+NSI	4.81	0.0283	0.0186	*cobC*	alpha-ribazole phosphatase
Cluster_117195	44	14	0	3	100	82	LI+NSI	4.83	0.0280	0.0189		conserved hypothetical protein
Cluster_171551	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		bacterial extracellular solute-binding family protein

Cluster_171552	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		binding-protein-dependent transport system inner membrane component family protein
Cluster_171555	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		glycerophosphoryl diester phosphodiesterase family protein
Cluster_171556	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		ROK family protein
Cluster_171708	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		conserved hypothetical protein
Cluster_193190	44	14	0	3	100	82	LI+NSI	4.83	0.0280	0.0189		nlpC/P60 family protein
Cluster_206060	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		ABC transporter family protein
Cluster_206339	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		putative predicted protein
Cluster_207740	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		caudovirales tail fibre assembly family protein
Cluster_23603	44	14	0	3	100	82	LI+NSI	4.83	0.0280	0.0189		putative acyltransferase YihG
Cluster_245044	44	14	0	3	100	82	LI+NSI	4.83	0.0280	0.0189	*ydiN*	inner membrane transport protein YdiN
Cluster_321371	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		binding-protein-dependent transport system inner membrane component family protein
Cluster_90557	44	14	0	3	100	82	LI+NSI	4.83	0.0280	0.0189		conserved hypothetical protein
Cluster_94403	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		conserved hypothetical protein
Cluster_225861	7	8	37	9	16	47	AI	4.85	0.0277	0.0192	*psu*	polarity suppression protein
Cluster_227577	7	8	37	9	16	47	AI	4.85	0.0277	0.0192	*ssb*	single-stranded DNA-binding protein
Cluster_232396	7	8	37	9	16	47	AI	4.85	0.0277	0.0192		putative dNA-binding helix-turn-helix protein
Cluster_31977	7	8	37	9	16	47	AI	4.85	0.0277	0.0192	*rfaJ*	lipopolysaccharide 1,2-glucosyltransferase
Cluster_4778	7	8	37	9	16	47	AI	4.85	0.0277	0.0192		conserved hypothetical protein
Cluster_207867	15	12	29	5	34	71	AI	5.22	0.0223	0.0199		putative invasin
Cluster_308000	15	12	29	5	34	71	AI	5.22	0.0223	0.0199	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_4832	29	5	15	12	66	29	LI+NSI	5.22	0.0223	0.0199		DNA methylase family protein
Cluster_88753	29	5	15	12	66	29	LI+NSI	5.22	0.0223	0.0199		conserved hypothetical protein
Cluster_297448	21	14	23	3	48	82	AI	4.68	0.0305	0.0204		conserved hypothetical protein
Cluster_14998	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		bacteriophage replication gene A family protein
Cluster_222333	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		phage major capsid protein, HK97 family
Cluster_252976	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		phage prohead protease, HK97 family
Cluster_59004	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		phage prohead protease, HK97 family
Cluster_80973	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		phage integrase family protein
Cluster_133289	40	11	4	6	91	65	LI+NSI	4.38	0.0364	0.0218		bacterial Ig-like domain family protein
Cluster_167315	40	11	4	6	91	65	LI+NSI	4.38	0.0364	0.0218		conserved hypothetical protein
Cluster_271588	4	6	40	11	9	35	AI	4.38	0.0364	0.0218	*intA*	prophage CP4-57 integrase
Cluster_302715	40	11	4	6	91	65	LI+NSI	4.38	0.0364	0.0218		conserved hypothetical protein
Cluster_40177	40	11	4	6	91	65	LI+NSI	4.38	0.0364	0.0218		major tail V domain protein
Cluster_56967	40	11	4	6	91	65	LI+NSI	4.38	0.0364	0.0218		RNA 2'-phosphotransferase, Tpt1 / KptA family protein
Cluster_78687	40	11	4	6	91	65	LI+NSI	4.38	0.0364	0.0218		minor tail protein U
Cluster_258248	16	12	28	5	36	71	AI	4.49	0.0341	0.0225		conserved hypothetical protein
Cluster_128422	30	6	14	11	68	35	LI+NSI	4.21	0.0402	0.0239		dynamin family protein
Cluster_14627	14	11	30	6	32	65	AI	4.21	0.0402	0.0239		cysteine protease , YopT-type domain protein
Cluster_272941	30	6	14	11	68	35	LI+NSI	4.21	0.0402	0.0239		phage tail E family protein
Cluster_284077	30	6	14	11	68	35	LI+NSI	4.21	0.0402	0.0239	*fepE*	ferric enterobactin (Enterochelin) transport
Cluster_74420	30	6	14	11	68	35	LI+NSI	4.21	0.0402	0.0239		conserved hypothetical protein
Cluster_178543	16	1	28	16	36	6	LI+NSI	4.25	0.0392	0.0241		type III secretion , HrpO family protein
Cluster_18601	28	16	16	1	64	94	AI	4.25	0.0392	0.0241		ribbon-helix-helix , copG family protein
Cluster_222895	16	1	28	16	36	6	LI+NSI	4.25	0.0392	0.0241		conserved hypothetical protein
Cluster_39298	28	16	16	1	64	94	AI	4.25	0.0392	0.0241		endodeoxyribonuclease RusA family protein
Cluster_82167	16	1	28	16	36	6	LI+NSI	4.25	0.0392	0.0241		conserved hypothetical protein
Cluster_242712	25	4	19	13	57	24	LI+NSI	4.20	0.0405	0.0244		DDE_Tnp_1-associated family protein
Cluster_285341	25	4	19	13	57	24	LI+NSI	4.20	0.0405	0.0244		H+ symporter family protein
Cluster_342345	25	4	19	13	57	24	LI+NSI	4.20	0.0405	0.0244		conserved hypothetical protein
Cluster_99113	25	4	19	13	57	24	LI+NSI	4.20	0.0405	0.0244		putative membrane protein
Cluster_283175	11	0	33	17	25	0	LI+NSI	3.63	0.0567	0.0255		bacterial regulatory, luxR family protein
Cluster_313761	11	0	33	17	25	0	LI+NSI	3.63	0.0567	0.0255		HNH endonuclease family protein
Cluster_314206	11	0	33	17	25	0	LI+NSI	3.63	0.0567	0.0255	*eprH*	type III secretion apparatus protein PrgH/EprH
Cluster_4740	11	0	33	17	25	0	LI+NSI	3.63	0.0567	0.0255		marR family protein
Cluster_59760	11	0	33	17	25	0	LI+NSI	3.63	0.0567	0.0255		putative variable tail fiber protein
Cluster_153716	9	9	35	8	20	53	AI	4.76	0.0292	0.0258		conserved hypothetical protein
Cluster_203385	9	9	35	8	20	53	AI	4.76	0.0292	0.0258		conserved hypothetical family protein
Cluster_147907	5	7	39	10	11	41	AI	5.14	0.0234	0.0263		HNH endonuclease family protein
Cluster_294187	39	10	5	7	89	59	LI+NSI	5.14	0.0234	0.0263		conserved hypothetical protein
Cluster_173912	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267	*espZ*	sepZ family protein
Cluster_180437	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		putative transposase
Cluster_183305	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		putative type III secretion protein EprJ
Cluster_197488	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		putative dNA-binding transcriptional regulator
Cluster_209025	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		putative epaO1 protein
Cluster_209146	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		shET2 enterotoxin, N-terminal region family protein
Cluster_211131	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		insA N-terminal domain protein
Cluster_221708	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		conserved hypothetical protein
Cluster_246778	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		conserved hypothetical protein
Cluster_289072	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		transposase DDE domain protein
Cluster_317010	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		conserved hypothetical protein
Cluster_348887	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		hypothetical protein
Cluster_72088	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		glycosyl transferases group 1 family protein
Cluster_72092	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		transposase DDE domain protein
Cluster_179213	41	12	3	5	93	71	LI+NSI	3.69	0.0548	0.0321		conserved hypothetical protein
Cluster_78756	3	5	41	12	7	29	AI	3.69	0.0548	0.0321		prophage CP4-57 regulatory family protein
Cluster_167317	34	8	10	9	77	47	LI+NSI	3.91	0.0481	0.0321		intergenic-region protein
Cluster_346478	34	8	10	9	77	47	LI+NSI	3.91	0.0481	0.0321		conserved hypothetical protein
Cluster_226889	6	7	38	10	14	41	AI	4.03	0.0448	0.0334		putative P4-specific DNA primase
Cluster_108011	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_178309	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335		conserved hypothetical protein
Cluster_232988	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335		putative molybdate metabolism regulator
Cluster_234872	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335		lysR substrate binding domain protein
Cluster_72094	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335		phage integrase family protein
Cluster_132559	26	15	18	2	59	88	AI	3.50	0.0615	0.0359		phoH-like family protein
Cluster_14894	18	2	26	15	41	12	LI+NSI	3.50	0.0615	0.0359		conserved hypothetical protein
Cluster_175201	18	2	26	15	41	12	LI+NSI	3.50	0.0615	0.0359		fimbrial family protein
Cluster_175778	18	2	26	15	41	12	LI+NSI	3.50	0.0615	0.0359		homeo-like domain protein
Cluster_221391	18	2	26	15	41	12	LI+NSI	3.50	0.0615	0.0359		conserved hypothetical protein
Cluster_300651	18	2	26	15	41	12	LI+NSI	3.50	0.0615	0.0359		conserved hypothetical protein
Cluster_322858	26	15	18	2	59	88	AI	3.50	0.0615	0.0359		conserved hypothetical protein
Cluster_180288	32	7	12	10	73	41	LI+NSI	4.01	0.0451	0.0360	*xseA*	exodeoxyribonuclease VII, large subunit
Cluster_264826	32	7	12	10	73	41	LI+NSI	4.01	0.0451	0.0360		yagB/YeeU/YfjZ family protein
Cluster_327168	32	7	12	10	73	41	LI+NSI	4.01	0.0451	0.0360		helix-turn-helix domain protein
Cluster_88571	32	7	12	10	73	41	LI+NSI	4.01	0.0451	0.0360	*cro*	regulatory protein cro
Cluster_54451	12	10	32	7	27	59	AI	4.01	0.0451	0.0360		conserved hypothetical protein
Cluster_9782	12	10	32	7	27	59	AI	4.01	0.0451	0.0360		putative cytoplasmic protein
Cluster_182223	22	3	22	14	50	18	LI+NSI	4.05	0.0441	0.0402		putative outer membrane fimbrial usher protein
Cluster_141364	21	3	23	14	48	18	LI+NSI	3.47	0.0623	0.0418	*ompD*	outer membrane porin protein OmpD
Cluster_20553	21	3	23	14	48	18	LI+NSI	3.47	0.0623	0.0418		conserved hypothetical protein
Cluster_26108	21	3	23	14	48	18	LI+NSI	3.47	0.0623	0.0418	*ompD*	outer membrane porin protein OmpD
Cluster_290717	23	14	21	3	52	82	AI	3.47	0.0623	0.0418		istB-like ATP binding family protein
Cluster_192298	13	10	31	7	30	59	AI	3.32	0.0686	0.0436		transposase IS116/IS110/IS902 family protein
Cluster_303767	13	10	31	7	30	59	AI	3.32	0.0686	0.0436		repressor protein CI
Cluster_317369	31	7	13	10	70	41	LI+NSI	3.32	0.0686	0.0436		tellurite resistance TerB family protein
Cluster_1235	29	6	15	11	66	35	LI+NSI	3.53	0.0602	0.0437		conserved hypothetical protein
Cluster_83238	29	6	15	11	66	35	LI+NSI	3.53	0.0602	0.0437		sirA-like family protein
Cluster_108598	27	5	17	12	61	29	LI+NSI	3.82	0.0506	0.0438		DNA-packaging protein FI

Cluster_104881	20	13	24	4	45	76	AI	3.58	0.0584	0.0444		putative toxin higB-2
Cluster_137958	20	13	24	4	45	76	AI	3.58	0.0584	0.0444	*symE*	toxic protein SymE
Cluster_196216	20	13	24	4	45	76	AI	3.58	0.0584	0.0444		putative propanediol utilization protein
Cluster_283607	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		methyltransferase domain protein
Cluster_325619	20	13	24	4	45	76	AI	3.58	0.0584	0.0444		aldehyde dehydrogenase family protein
Cluster_337809	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		putative antigen 43 domain protein
Cluster_40501	20	13	24	4	45	76	AI	3.58	0.0584	0.0444		helix-turn-helix domain protein
Cluster_43613	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		phage gp6-like head-tail connector family protein
Cluster_82579	20	13	24	4	45	76	AI	3.58	0.0584	0.0444	*espJ*	putative non-LEE-encoded effector EspJ
Cluster_84211	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		conserved hypothetical protein
Cluster_324271	2	4	42	13	5	24	AI	3.07	0.0796	0.0457	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_8918	2	4	42	13	5	24	AI	3.07	0.0796	0.0457	*ssbF*	plasmid-derived single-stranded DNA-binding protein
Cluster_218915	42	13	2	4	95	76	LI+NSI	3.07	0.0796	0.0457		parB-like nuclease domain protein
Cluster_118990	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein
Cluster_128383	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein
Cluster_14537	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		viral (Super1) RNA helicase family protein
Cluster_183369	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein
Cluster_197463	30	16	14	1	68	94	AI	3.16	0.0755	0.0469	*hipA*	serine/threonine-protein kinase HipA
Cluster_261666	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		AAA domain protein
Cluster_271379	30	16	14	1	68	94	AI	3.16	0.0755	0.0469		glycosyl hydrolase family 65, N-terminal domain protein
Cluster_276697	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein
Cluster_291791	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		plasmid conjugative transfer entry exclusion TraS family protein
Cluster_292258	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein
Cluster_59324	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		putative acetyltransferase
Cluster_73497	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein
Cluster_171049	7	7	37	10	16	41	AI	3.11	0.0776	0.0471	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_281011	7	7	37	10	16	41	AI	3.11	0.0776	0.0471		conserved hypothetical protein
Cluster_147263	36	9	8	8	82	53	LI+NSI	3.90	0.0484	0.0479		methyltransferase domain protein
Cluster_267179	36	9	8	8	82	53	LI+NSI	3.90	0.0484	0.0479		conserved hypothetical protein
Cluster_319379	8	8	36	9	18	47	AI	3.90	0.0484	0.0479	*rfaJ*	lipopolysaccharide 1,2-glucosyltransferase
Cluster_321570	8	8	36	9	18	47	AI	3.90	0.0484	0.0479		transposase family protein
Cluster_7938	36	9	8	8	82	53	LI+NSI	3.90	0.0484	0.0479		DNA methylase family protein
Cluster_103416	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein
Cluster_153041	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein
Cluster_208887	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		putative escherichia coli IMT2125 genomic chromosome, IMT2125
Cluster_248119	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		dnaB-like helicase N terminal domain protein
Cluster_302692	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		hypothetical protein
Cluster_342342	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein
Cluster_342343	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein
Cluster_46843	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		dnaB-like helicase N terminal domain protein
Cluster_70068	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein
Cluster_82835	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein
Cluster_68405	18	12	26	5	41	71	AI	3.22	0.0729	0.0486	*fimA*	type-1 fimbrial protein, A chain
Cluster_73146	18	12	26	5	41	71	AI	3.22	0.0729	0.0486		conserved hypothetical protein
Cluster_221116	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		hypothetical protein
Cluster_312941	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		putative antirepressor protein
Cluster_317549	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		ead/Ea22-like family protein
Cluster_317551	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		conserved hypothetical protein
Cluster_44635	34	17	10	0	77	100	AI	3.11	0.0777	0.0493		type III secretion , HrpO family protein
Cluster_95057	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		P22AR C-terminal domain protein

Lethal (LI) vs. Non-Lethal (NSI+AI)

	No. of Genomes											
	Present (≥0.9)^{a,b}		Absent (<0.9)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test		
Cluster ID	LI	NSI+AI	LI	NSI+AI	% LI	% NSI+AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function
Cluster_101055	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*prc*	tail-specific protease
Cluster_104965	4	0	20	37	17	0	LI	4.16	0.0414	0.0204		transposase family protein
Cluster_105254	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*aceF*	dihydrolipoyllysine-residue acetyltransferase
Cluster_108011	13	8	11	29	54	22	LI	5.46	0.0194	0.0132	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN
Cluster_108780	18	35	6	2	75	95	NSI+AI	3.34	0.0678	0.0481		phage terminase large subunit family protein
Cluster_11079	11	29	13	8	46	78	NSI+AI	5.46	0.0194	0.0132		transglycosylase SLT domain protein
Cluster_113457	14	31	10	6	58	84	NSI+AI	3.65	0.0562	0.0383	*fimF*	protein FimF
Cluster_113721	7	22	17	15	29	59	NSI+AI	4.21	0.0402	0.0350		phage tail tape measure protein, lambda family
Cluster_120557	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*surE*	5'/3'-nucleotidase SurE
Cluster_125122	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		putative predicted protein
Cluster_12875	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*narU*	nitrite extrusion 2 domain protein
Cluster_132559	12	29	12	8	50	78	NSI+AI	4.11	0.0426	0.0277		phoH-like family protein
Cluster_133256	0	7	24	30	0	19	NSI+AI	3.44	0.0638	0.0358		conserved hypothetical protein
Cluster_135282	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*rplQ*	ribosomal protein L17
Cluster_136944	17	35	7	2	71	95	NSI+AI	4.78	0.0288	0.0221		phage terminase large subunit family protein
Cluster_137178	4	18	20	19	17	49	NSI+AI	5.15	0.0233	0.0144		hypothetical protein
Cluster_14417	9	4	15	33	38	11	LI	4.69	0.0303	0.0231		putative predicted protein
Cluster_14423	10	5	14	32	42	14	LI	4.80	0.0285	0.0170		conserved hypothetical protein
Cluster_14627	5	20	19	17	21	54	NSI+AI	5.34	0.0208	0.0159		cysteine protease , YopT-type domain protein
Cluster_14998	17	13	7	24	71	35	LI	6.06	0.0138	0.0090		bacteriophage replication gene A family protein
Cluster_152085	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_155747	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*phnH*	phosphonate C-P lyase system protein PhnH
Cluster_166190	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage integrase family protein
Cluster_174674	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		conserved hypothetical protein
Cluster_175055	8	4	16	33	33	11	LI	3.36	0.0669	0.0474	*trbJ*	P-type conjugative transfer protein TrbJ
Cluster_175131	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		bacterial transferase hexapeptide family protein
Cluster_175605	6	0	18	37	25	0	LI	7.63	0.0057	0.0024	*rfbA*	glucose-1-phosphate thymidylyltransferase
Cluster_175612	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		transposase DDE domain protein
Cluster_175618	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		polysaccharide biosynthesis/export family protein
Cluster_176078	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		outer membrane autotransporter barrel domain protein
Cluster_176094	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		NIF3 family protein
Cluster_176970	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_177334	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		conserved hypothetical protein
Cluster_177488	6	2	18	35	25	5	LI	3.34	0.0678	0.0481	*fliC*	flagellin
Cluster_178120	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_178144	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		putative dsORF-h1
Cluster_178244	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_182662	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		putative membrane protein
Cluster_182949	8	4	16	33	33	11	LI	3.36	0.0669	0.0474	*repA*	regulatory protein RepA
Cluster_183172	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		DDE domain protein
Cluster_183173	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		transposase family protein
Cluster_184733	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*entS*	enterobactin exporter EntS
Cluster_196644	17	37	7	0	71	100	NSI+AI	9.49	0.0021	0.0008	*hscA*	fe-S protein assembly chaperone HscA
Cluster_196649	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*sseA*	3-mercaptopyruvate sulfurtransferase
Cluster_197373	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*emrB*	multidrug resistance protein B
Cluster_201168	18	36	6	1	75	97	NSI+AI	5.10	0.0239	0.0122	*eutK*	ethanolamine utilization protein EutK
Cluster_201403	9	4	15	33	38	11	LI	4.69	0.0303	0.0231		conserved hypothetical protein
Cluster_201731	10	4	14	33	42	11	LI	6.19	0.0128	0.0108		helix-turn-helix, Psq domain protein
Cluster_201738	10	4	14	33	42	11	LI	6.19	0.0128	0.0108		conserved hypothetical protein
Cluster_204545	16	33	8	4	67	89	NSI+AI	3.36	0.0669	0.0474		conserved hypothetical protein
Cluster_208031	18	36	6	1	75	97	NSI+AI	5.10	0.0239	0.0122		substrate binding domain of ABC-type glycine betaine transport system family protein

Cluster_208392	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		conserved hypothetical protein
Cluster_208660	18	36	6	1	75	97	NSI+AI	5.10	0.0239	0.0122	*dnaX*	DNA polymerase III, subunit gamma and tau
Cluster_209349	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*tnaA*	tryptophanase
Cluster_209403	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*selB*	selenocysteine-specific translation elongation factor
Cluster_209404	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204		iron-containing alcohol dehydrogenase family protein
Cluster_211547	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*dnaE*	DNA polymerase III, alpha subunit
Cluster_21586	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*citG*	triphosphoribosyl-dephospho-CoA synthase CitG
Cluster_21640	18	37	6	0	75	100	NSI+AI	7.63	0.0057	0.0024	*nagE*	PTS system, N-acetylglucosamine-specific IIBC component
Cluster_21650	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*pgm*	phosphoglucomutase, alpha-D-glucose phosphate-specific
Cluster_217662	0	7	24	30	0	19	NSI+AI	3.44	0.0638	0.0358		conserved hypothetical protein
Cluster_218595	8	3	16	34	33	8	LI	4.68	0.0306	0.0180	*traJ*	protein TraJ
Cluster_21870	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*ssb*	single-stranded DNA-binding protein
Cluster_218768	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*dacB*	D-alanyl-D-alanine carboxypeptidase/D-alanyl-D-alanine-endopeptidase
Cluster_220402	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*alr*	alanine racemase
Cluster_220923	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*nuoN*	NADH-quinone oxidoreductase subunit N
Cluster_22096	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071		AAA-like domain protein
Cluster_221116	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		hypothetical protein
Cluster_221350	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*ftsY*	signal recognition particle-docking protein FtsY
Cluster_222331	15	12	9	25	63	32	LI	4.19	0.0408	0.0342		phage portal protein, HK97 family
Cluster_222333	17	13	7	24	71	35	LI	6.06	0.0138	0.0090		phage major capsid protein, HK97 family
Cluster_222808	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		polysaccharide biosynthesis/export family protein
Cluster_224504	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*rodZ*	cytoskeleton protein RodZ
Cluster_236134	16	13	8	24	67	35	LI	4.61	0.0318	0.0201	*dhfrI*	dihydrofolate reductase type 1
Cluster_237298	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_23757	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		binding-protein-dependent transport system inner membrane component family protein
Cluster_237589	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		RHS repeat-associated core domain protein
Cluster_237804	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071		nitrate reductase, alpha subunit
Cluster_23840	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		glycosyl transferase family 2 domain protein
Cluster_239025	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage terminase, small subunit, P27 family
Cluster_239026	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage Terminase family protein
Cluster_239028	15	12	9	25	63	32	LI	4.19	0.0408	0.0342		phage portal protein, HK97 family
Cluster_239410	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		bacteriophage replication gene A family protein
Cluster_240829	7	1	17	36	29	3	LI	6.78	0.0092	0.0046		hypothetical protein
Cluster_24145	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_24314	5	0	19	37	21	0	LI	5.86	0.0155	0.0071		degT/DnrJ/EryC1/StrS aminotransferase family protein
Cluster_243375	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		tatD related DNase family protein
Cluster_24348	5	20	19	17	21	54	NSI+AI	5.34	0.0208	0.0159		phage tail tape measure protein, lambda family
Cluster_245108	18	36	6	1	75	97	NSI+AI	5.10	0.0239	0.0122		electron transport complex, RnfABCDGE type, B subunit
Cluster_245260	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*cheA*	chemotaxis protein CheA
Cluster_246425	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_247403	5	0	19	37	21	0	LI	5.86	0.0155	0.0071		glycosyl transferase 2 family protein
Cluster_247837	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		putative domain protein
Cluster_252971	5	1	19	36	21	3	LI	3.55	0.0597	0.0307		HNH endonuclease family protein
Cluster_252976	17	13	7	24	71	35	LI	6.06	0.0138	0.0090		phage prohead protease, HK97 family
Cluster_254547	17	15	7	22	71	41	LI	4.21	0.0402	0.0350		bacteriophage lysis family protein
Cluster_254674	7	21	17	16	29	57	NSI+AI	3.42	0.0644	0.0403		phage integrase family protein
Cluster_256982	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		hypothetical protein
Cluster_257914	24	28	0	9	100	76	LI	5.05	0.0246	0.0088	*cadC*	transcriptional activator CadC
Cluster_258248	6	22	18	15	25	59	NSI+AI	5.64	0.0175	0.0100		conserved hypothetical protein
Cluster_258452	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*malP*	maltodextrin phosphorylase
Cluster_259070	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*mfd*	transcription-repair coupling factor
Cluster_259563	18	36	6	1	75	97	NSI+AI	5.10	0.0239	0.0122		MASE1 family protein
Cluster_261814	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_263937	9	4	15	33	38	11	LI	4.69	0.0303	0.0231		conserved hypothetical protein
Cluster_263945	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		conserved hypothetical protein
Cluster_263946	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage major capsid E family protein
Cluster_263948	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		conserved hypothetical protein
Cluster_264521	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071		conserved hypothetical protein
Cluster_265225	18	37	6	0	75	100	NSI+AI	7.63	0.0057	0.0024		PRD domain protein
Cluster_265334	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*mdfA*	multidrug transporter MdfA
Cluster_266191	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*yecS*	inner membrane amino-acid ABC transporter permease protein YecS
Cluster_2667	8	3	16	34	33	8	LI	4.68	0.0306	0.0180		hypothetical protein
Cluster_269540	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*can*	carbonic anhydrase 2
Cluster_272941	19	17	5	20	79	46	LI	5.34	0.0208	0.0159		phage tail E family protein
Cluster_273556	4	16	20	21	17	43	NSI+AI	3.54	0.0600	0.0496		glycosyltransferase like 2 family protein
Cluster_273680	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*menD*	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase
Cluster_274418	4	18	20	19	17	49	NSI+AI	5.15	0.0233	0.0144		mu-like prophage FluMu gp41 family protein
Cluster_275430	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*adiC*	arginine/agmatine antiporter
Cluster_275815	18	36	6	1	75	97	NSI+AI	5.10	0.0239	0.0122	*gmm*	GDP-mannose mannosyl hydrolase
Cluster_276292	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307		stress-induced bacterial acidophilic repeat motif family protein
Cluster_276357	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307		type VII secretion system (T7SS), usher family protein
Cluster_280171	18	37	6	0	75	100	NSI+AI	7.63	0.0057	0.0024	*tatD*	tat-linked quality control protein TatD
Cluster_282502	18	35	6	2	75	95	NSI+AI	3.34	0.0678	0.0481	*yphB*	aldose 1-epimerase
Cluster_282548	8	3	16	34	33	8	LI	4.68	0.0306	0.0180		putative predicted protein
Cluster_286226	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*cybH*	ni/Fe-hydrogenase, b-type cytochrome subunit
Cluster_286879	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		conserved hypothetical protein
Cluster_286883	5	0	19	37	21	0	LI	5.86	0.0155	0.0071		acetyltransferase domain protein
Cluster_28951	4	0	20	37	17	0	LI	4.16	0.0414	0.0204		hypothetical protein
Cluster_290242	18	35	6	2	75	95	NSI+AI	3.34	0.0678	0.0481	*yphF*	ABC transporter periplasmic-binding protein YphF
Cluster_290560	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071		AAA-like domain protein
Cluster_29582	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		HNH endonuclease family protein
Cluster_295885	18	36	6	1	75	97	NSI+AI	5.10	0.0239	0.0122		acetyltransferase domain protein
Cluster_296132	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		ogr/Delta-like zinc finger family protein
Cluster_301782	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		major tail sheath protein
Cluster_301785	4	18	20	19	17	49	NSI+AI	5.15	0.0233	0.0144		phage tail tape measure protein, TP901 family, core region
Cluster_302107	14	10	10	27	58	27	LI	4.74	0.0295	0.0181		conserved hypothetical protein
Cluster_305186	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		phage P2 GpU family protein
Cluster_307747	12	29	12	8	50	78	NSI+AI	4.11	0.0426	0.0277		conserved hypothetical protein
Cluster_312709	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		amino acid permease family protein
Cluster_312941	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		putative antirepressor protein
Cluster_313405	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		luxR family transcriptional regulator
Cluster_313761	9	2	15	35	38	5	LI	8.09	0.0045	0.0043		HNH endonuclease family protein
Cluster_315364	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		archaeal transcriptional regulator TrmB family protein
Cluster_315365	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		ABC transporter family protein
Cluster_315743	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*ycdY*	chaperone protein YcdY
Cluster_315801	2	12	22	25	8	32	NSI+AI	3.52	0.0608	0.0335		phage tail protein I
Cluster_317549	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		ead/Ea22-like family protein
Cluster_317551	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		conserved hypothetical protein
Cluster_32291	18	37	6	0	75	100	NSI+AI	7.63	0.0057	0.0024	*dcd*	deoxycytidine triphosphate deaminase
Cluster_324641	18	35	6	2	75	95	NSI+AI	3.34	0.0678	0.0481	*yphB*	aldose 1-epimerase
Cluster_333715	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*dinI*	DNA-damage-inducible protein I
Cluster_335585	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071		ybgS-like family protein
Cluster_342998	17	15	7	22	71	41	LI	4.21	0.0402	0.0350		phage tail protein I
Cluster_343372	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*leuB*	3-isopropylmalate dehydrogenase
Cluster_345283	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307		carbon-nitrogen hydrolase family protein
Cluster_346239	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		HNH endonuclease family protein
Cluster_346478	21	21	3	16	88	57	LI	5.06	0.0245	0.0127		conserved hypothetical protein

Cluster_347476	20	20	4	17	83	54	LI	4.31	0.0379	0.0271		conserved hypothetical protein
Cluster_348353	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*yajR*	inner membrane transport protein YajR
Cluster_44635	17	34	7	3	71	92	NSI+AI	3.30	0.0693	0.0397		type III secretion , HrpO family protein
Cluster_46213	3	16	21	21	13	43	NSI+AI	5.06	0.0245	0.0127		putative transmembrane protein
Cluster_5054	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307		uracil-xanthine permease family protein
Cluster_51806	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*ptsP*	phosphoenolpyruvate-protein phosphotransferase
Cluster_52793	19	19	5	18	79	51	LI	3.68	0.0549	0.0340	*mngB*	mannosylglycerate hydrolase
Cluster_5286	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*atpF*	ATP synthase F0, B subunit
Cluster_58271	9	4	15	33	38	11	LI	4.69	0.0303	0.0231		conserved hypothetical protein
Cluster_59004	17	13	7	24	71	35	LI	6.06	0.0138	0.0090		phage prohead protease, HK97 family
Cluster_61005	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204		conserved hypothetical protein
Cluster_67345	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		conserved hypothetical protein
Cluster_73598	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein
Cluster_74186	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		helix-turn-helix domain protein
Cluster_74420	19	17	5	20	79	46	LI	5.34	0.0208	0.0159		conserved hypothetical protein
Cluster_75095	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		phage major tail tube protein
Cluster_75099	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		phage late control gene D family protein
Cluster_76816	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		putative dsORF-f3
Cluster_79549	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*prpR*	propionate catabolism operon regulatory protein PrpR
Cluster_79564	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		putative oxidoreductase subunit domain protein
Cluster_79937	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		prophage CP4-57 regulatory family protein
Cluster_80022	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		alpha amylase, catalytic domain protein
Cluster_81331	18	37	6	0	75	100	NSI+AI	7.63	0.0057	0.0024		lppC lipofamily protein
Cluster_81348	20	37	4	0	83	100	NSI+AI	4.16	0.0414	0.0204	*pnp*	polyribonucleotide nucleotidyltransferase
Cluster_81842	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307		MATE efflux family protein
Cluster_81862	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307		conserved hypothetical protein
Cluster_83084	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*purD*	phosphoribosylamine--glycine ligase
Cluster_83254	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		glycosyl transferases group 1 family protein
Cluster_83903	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		conserved hypothetical protein
Cluster_83927	6	0	18	37	25	0	LI	7.63	0.0057	0.0024	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase
Cluster_83929	6	0	18	37	25	0	LI	7.63	0.0057	0.0024	*rfbB*	dTDP-glucose 4,6-dehydratase
Cluster_84140	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		putative cymA protein
Cluster_87728	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*phrB*	deoxyribodipyrimidine photo-lyase
Cluster_93726	19	37	5	0	79	100	NSI+AI	5.86	0.0155	0.0071	*tolA*	protein TolA
Cluster_94431	0	7	24	30	0	19	NSI+AI	3.44	0.0638	0.0358		hypothetical protein
Cluster_95057	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		P22AR C-terminal domain protein
Cluster_95290	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		invH outer membrane lipofamily protein
Cluster_97263	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*cadA*	cadmium-translocating P-type ATPase
Cluster_97497	19	36	5	1	79	97	NSI+AI	3.55	0.0597	0.0307	*tolA*	protein TolA
Cluster_99077	4	16	20	21	17	43	NSI+AI	3.54	0.0600	0.0496		cysteine protease , YopT-type domain protein
Cluster_99113	16	13	8	24	67	35	LI	4.61	0.0318	0.0201		putative membrane protein
Cluster_99274	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		putative t3SS effector protein EspS
Cluster_99362	14	10	10	27	58	27	LI	4.74	0.0295	0.0181		resolvase, N terminal domain protein
Cluster_99439	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		putative dsORF-e4

[a] Genes were grouped into gene clusters using a "stringent" clustering threshold of ≥90% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.9, while those that are "absent" have a TBLASTN LS-BSR <0.9.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

Table S6. Genes identified by LS-BSR that are significantly correlated with one clinical outcome over another for all 70 EPEC genomes analyzed using an inclusive clustering threshold

Lethal (LI) vs. Non-lethal Symptomatic (NSI)

	No. of Genomes												
	Present (≥0.8)[a,b]		Absent (<0.8)[a,b]		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI	NSI	LI	NSI	% LI	% NSI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_175606	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase	N
Cluster_222636	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*rfbA*	glucose-1-phosphate thymidylyltransferase	N
Cluster_147640	11	3	13	20	46	13	LI	4.57	0.0325	0.0243		conserved hypothetical protein	Y
Cluster_112366	4	11	20	12	17	48	NSI	3.91	0.0479	0.0305		mu-like prophage FluMu gp41 family protein	N
Cluster_252971	20	12	4	11	83	52	LI	3.91	0.0479	0.0305		HNH endonuclease family protein	N
Cluster_196142	19	11	5	12	79	48	LI	3.73	0.0534	0.0355		conserved hypothetical protein	N
Cluster_140609	2	8	22	15	8	35	NSI	3.45	0.0631	0.0363		conserved hypothetical protein	N
Cluster_243755	15	7	9	16	63	30	LI	3.65	0.0561	0.0415		ASCH domain protein	N
Cluster_98923	17	9	7	14	71	39	LI	3.58	0.0585	0.0415		bacteriophage lysis family protein	Y
Cluster_236134	16	8	8	15	67	35	LI	3.59	0.0582	0.0422	*dhfrI*	dihydrofolate reductase type 1	N
Cluster_123546	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		phosphoadenosine phosphosulfate reductase family protein	N
Cluster_123571	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		ABC transporter family protein	Y
Cluster_160926	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		helix-turn-helix domain protein	N
Cluster_214112	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		phage tail-collar fibre family protein	Y
Cluster_216712	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		putative membrane protein	Y
Cluster_217182	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479	*fyuA*	pesticin receptor	Y
Cluster_217185	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479	*ybtU*	yersiniabactin biosynthetic protein YbtU	N
Cluster_217187	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		amino acid adenylation domain protein	N
Cluster_217190	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		ABC transporter family protein	Y
Cluster_217193	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		phage integrase family protein	N
Cluster_217459	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		HTH domain protein	N
Cluster_284866	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		bacterial regulatory, tetR family protein	N
Cluster_305198	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		conserved hypothetical protein	Y
Cluster_307421	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		alpha/beta hydrolase family protein	N
Cluster_307423	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		methyltransferase family protein	N
Cluster_307429	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		salicylate synthase	N
Cluster_48358	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		major Facilitator Superfamily protein	Y
Cluster_99052	1	6	23	17	4	26	NSI	2.89	0.0891	0.0479		conserved hypothetical protein	Y
Cluster_14414	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		conserved hypothetical protein	N
Cluster_14423	10	3	14	20	42	13	LI	3.48	0.0619	0.0490		conserved hypothetical protein	N
Cluster_175608	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		degT/DnrJ/EryC1/StrS aminotransferase family protein	N
Cluster_247403	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 2 family protein	N
Cluster_83926	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		acetyltransferase domain protein	N
Cluster_121796	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		caudovirales tail fibre assembly family protein	N
Cluster_121973	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		conserved hypothetical protein	N
Cluster_163539	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		hypothetical protein	N
Cluster_188256	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		conserved hypothetical protein	Y
Cluster_217181	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		conserved hypothetical protein	N
Cluster_8046	0	4	24	19	0	17	NSI	2.60	0.1067	0.0496		hypothetical protein	N

Lethal (LI) vs. Asymptomatic (AI)

	No. of Genomes												
	Present (≥0.8)[a,b]		Absent (<0.8)[a,b]		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI	AI	LI	AI	% LI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_34159	13	2	11	21	54	9	LI	9.18	0.0024	0.0013		conserved hypothetical protein	N
Cluster_262069	12	2	12	21	50	9	LI	7.71	0.0055	0.0034		istB-like ATP binding family protein	N
Cluster_267235	12	2	12	21	50	9	LI	7.71	0.0055	0.0034		hypothetical protein	N
Cluster_289981	12	2	12	21	50	9	LI	7.71	0.0055	0.0034		integrase core domain protein	N
Cluster_341980	22	12	2	11	92	52	LI	7.29	0.0069	0.0034		conserved hypothetical protein	Y
Cluster_307754	18	7	6	16	75	30	LI	7.66	0.0056	0.0034	*cro*	regulatory protein cro	N
Cluster_327168	18	7	6	16	75	30	LI	7.66	0.0056	0.0034		helix-turn-helix domain protein	N
Cluster_147543	8	0	16	23	33	0	LI	7.03	0.0080	0.0039		putative yfdA	Y
Cluster_157448	1	9	23	14	4	39	AI	6.61	0.0101	0.0044	*espS*	putative T3SS effector protein EspS	N
Cluster_14414	10	1	14	22	42	4	LI	7.16	0.0074	0.0044		conserved hypothetical protein	N
Cluster_58277	10	1	14	22	42	4	LI	7.16	0.0074	0.0044		helix-turn-helix, Psq domain protein	N
Cluster_260984	13	3	11	20	54	13	LI	7.11	0.0077	0.0050	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_169965	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		ogr/Delta-like zinc finger family protein	N
Cluster_200452	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		putative prophage protein	N
Cluster_200454	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		putative phage immunity repressor protein	N
Cluster_271578	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		conserved hypothetical protein	N
Cluster_316955	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		helix-turn-helix domain protein	N
Cluster_63459	4	13	20	10	17	57	AI	6.45	0.0111	0.0065		putative derepression protein	N
Cluster_252971	20	10	4	13	83	43	LI	6.45	0.0111	0.0065		HNH endonuclease family protein	N
Cluster_83828	19	9	5	14	79	39	LI	6.24	0.0125	0.0077		conserved hypothetical protein	N
Cluster_56014	22	13	2	10	92	57	LI	5.89	0.0152	0.0078		hydantoinase/oxoprolinase family protein	N
Cluster_107351	11	2	13	21	46	9	LI	6.35	0.0118	0.0078		conserved hypothetical protein	N
Cluster_183593	11	2	13	21	46	9	LI	6.35	0.0118	0.0078		conserved hypothetical protein	N
Cluster_276264	11	2	13	21	46	9	LI	6.35	0.0118	0.0078		replication regulatory RepB family protein	N
Cluster_312794	11	2	13	21	46	9	LI	6.35	0.0118	0.0078		transposase DDE domain protein	N
Cluster_236134	16	6	8	17	67	26	LI	6.22	0.0126	0.0084	*dhfrI*	dihydrofolate reductase type 1	N
Cluster_346831	16	6	8	17	67	26	LI	6.22	0.0126	0.0084		conserved hypothetical protein	N
Cluster_43705	18	8	6	15	75	35	LI	6.14	0.0132	0.0084		conserved hypothetical protein	Y
Cluster_222333	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		phage major capsid protein, HK97 family	N
Cluster_59004	17	7	7	16	71	30	LI	6.14	0.0132	0.0087		phage prohead protease, HK97 family	N
Cluster_106122	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		conserved hypothetical protein	N
Cluster_115298	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		conserved hypothetical protein	N
Cluster_166190	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		phage integrase family protein	N
Cluster_240829	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		hypothetical protein	N
Cluster_261817	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		putative predicted protein	N
Cluster_262291	17	23	7	0	71	100	AI	5.75	0.0165	0.0094		type III secretion , HrpO family protein	Y
Cluster_263945	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		conserved hypothetical protein	N
Cluster_74193	7	0	17	23	29	0	LI	5.75	0.0165	0.0094		phage major capsid E family protein	N
Cluster_3723	0	6	24	17	0	26	AI	5.03	0.0250	0.0094		conserved hypothetical protein	N
Cluster_13383	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		conserved hypothetical protein	N
Cluster_183140	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122	Y
Cluster_227359	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		conserved hypothetical protein	N
Cluster_227360	9	1	15	22	38	4	LI	5.85	0.0155	0.0102		putative membrane protein	Y
Cluster_276738	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		putative protein	Y

Cluster_340919	23	15	1	8	96	65	LI	5.27	0.0217	0.0102		winged helix-turn-helix DNA-binding family protein	N
Cluster_90939	21	12	3	11	88	52	LI	5.42	0.0199	0.0114		plasmid stability family protein	N
Cluster_225861	3	11	21	12	13	48	AI	5.42	0.0199	0.0114	*psu*	polarity suppression protein	N
Cluster_225859	4	12	20	11	17	52	AI	5.11	0.0238	0.0145		putative glyco3, capsid size determination protein Sid	N
Cluster_88073	20	11	4	12	83	48	LI	5.11	0.0238	0.0145		small Multidrug Resistance family protein	Y
Cluster_166870	19	10	5	13	79	43	LI	4.91	0.0267	0.0171		integron integrase family protein	N
Cluster_14423	10	2	14	21	42	9	LI	5.09	0.0240	0.0173		conserved hypothetical protein	N
Cluster_178309	22	14	2	9	92	61	LI	4.61	0.0317	0.0173		conserved hypothetical protein	N
Cluster_79151	22	14	2	9	92	61	LI	4.61	0.0317	0.0173		bundlin family protein	Y
Cluster_18413	18	9	6	14	75	39	LI	4.80	0.0284	0.0189		resolvase, N terminal domain protein	N
Cluster_312629	18	9	6	14	75	39	LI	4.80	0.0284	0.0189		resolvase, N terminal domain protein	N
Cluster_43605	18	9	6	14	75	39	LI	4.80	0.0284	0.0189		HNH endonuclease family protein	N
Cluster_222331	16	7	8	16	67	30	LI	4.81	0.0284	0.0199		phage portal protein, HK97 family	N
Cluster_266615	16	7	8	16	67	30	LI	4.81	0.0284	0.0199		putative yubO	N
Cluster_106126	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		conserved hypothetical protein	N
Cluster_115308	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		helix-turn-helix domain protein	Y
Cluster_139135	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		putative pyridoxamine 5-phosphate-dependent dehydrase	N
Cluster_152085	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		conserved hypothetical protein	N
Cluster_168916	0	5	24	18	0	22	AI	3.78	0.0520	0.0219		nucleoside transporter family protein	Y
Cluster_175606	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase	N
Cluster_221922	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		glycosyl transferase family 2 domain protein	N
Cluster_222636	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*rfbA*	glucose-1-phosphate thymidylyltransferase	N
Cluster_247401	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		polysaccharide biosynthesis/export family protein	N
Cluster_25983	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		bacteriophage replication gene A family protein	N
Cluster_266659	24	18	0	5	100	78	LI	3.78	0.0520	0.0219	*repA*	incFII family plasmid replication initiator RepA	N
Cluster_291294	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		putative pyridoxamine 5-phosphate-dependent dehydrase	Y
Cluster_316809	0	5	24	18	0	22	AI	3.78	0.0520	0.0219		conserved hypothetical protein	N
Cluster_82368	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*fliC*	flagellin	N
Cluster_83254	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		glycosyl transferases group 1 family protein	N
Cluster_83929	6	0	18	23	25	0	LI	4.54	0.0332	0.0219	*rfbB*	dTDP-glucose 4,6-dehydratase	N
Cluster_84063	6	0	18	23	25	0	LI	4.54	0.0332	0.0219		polysaccharide biosynthesis/export family protein	Y
Cluster_124501	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		DNA N-6-adenine-methyltransferase family protein	N
Cluster_128383	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein	N
Cluster_128390	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		viral (Super1) RNA helicase family protein	N
Cluster_164940	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		AAA domain protein	N
Cluster_276697	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein	N
Cluster_291795	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		putative acetyltransferase	Y
Cluster_297253	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		transposase, Mutator family protein	Y
Cluster_316814	1	7	23	16	4	30	AI	4.03	0.0447	0.0226		AAA domain protein	N
Cluster_321659	23	16	1	7	96	70	LI	4.03	0.0447	0.0226	*bfpA*	major structural subunit of bundle-forming pilus	Y
Cluster_336502	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		type II/IV secretion system family protein	N
Cluster_53910	1	7	23	16	4	30	AI	4.03	0.0447	0.0226	*intA*	prophage CP4-57 integrase	N
Cluster_59092	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein	N
Cluster_67359	1	7	23	16	4	30	AI	4.03	0.0447	0.0226	*intA*	prophage CP4-57 integrase	N
Cluster_73497	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein	N
Cluster_94384	23	16	1	7	96	70	LI	4.03	0.0447	0.0226		transposase, Mutator family protein	N
Cluster_98944	8	1	16	22	33	4	LI	4.64	0.0313	0.0226		conserved hypothetical protein	N
Cluster_99052	1	7	23	16	4	30	AI	4.03	0.0447	0.0226		conserved hypothetical protein	Y
Cluster_316740	21	13	3	10	88	57	LI	4.19	0.0406	0.0243		conserved hypothetical protein	Y
Cluster_342430	21	13	3	10	88	57	LI	4.19	0.0406	0.0243		conserved hypothetical protein	N
Cluster_99069	21	13	3	10	88	57	LI	4.19	0.0406	0.0243		haemolysin expression modulating family protein	N
Cluster_177188	11	3	13	20	46	13	LI	4.57	0.0325	0.0243		phage integrase family protein	N
Cluster_316843	13	20	11	3	54	87	AI	4.57	0.0325	0.0243		transglycosylase SLT domain protein	Y
Cluster_112366	4	11	20	12	17	48	AI	3.91	0.0479	0.0305		mu-like prophage FluMu gp41 family protein	N
Cluster_237277	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein	N
Cluster_277010	20	12	4	11	83	52	LI	3.91	0.0479	0.0305		transposase DDE domain protein	Y
Cluster_303065	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein	Y
Cluster_337159	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein	N
Cluster_342585	20	12	4	11	83	52	LI	3.91	0.0479	0.0305	*folP*	dihydropteroate synthase	N
Cluster_48952	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		conserved hypothetical protein	Y
Cluster_74298	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		traP family protein	Y
Cluster_78370	12	4	12	19	50	17	LI	4.20	0.0403	0.0305		putative predicted protein	N
Cluster_302692	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		hypothetical protein	N
Cluster_56633	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		conserved hypothetical protein	N
Cluster_84849	13	5	11	18	54	22	LI	3.94	0.0470	0.0355		conserved hypothetical protein	N
Cluster_210977	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		conserved hypothetical protein	N
Cluster_261646	22	15	2	8	92	65	LI	3.45	0.0631	0.0363		phage/conjugal plasmid C-4 type zinc finger , TraR family protein	N
Cluster_263620	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		transposase DDE domain protein	N
Cluster_263937	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		conserved hypothetical protein	N
Cluster_297449	2	8	22	15	8	35	AI	3.45	0.0631	0.0363		conserved hypothetical protein	N
Cluster_341683	22	15	2	8	92	65	LI	3.45	0.0631	0.0363		bacterial type II and III secretion system family protein	Y
Cluster_59081	22	15	2	8	92	65	LI	3.45	0.0631	0.0363		type II secretion system (T2SS), F family protein	Y
Cluster_72088	9	2	15	21	38	9	LI	3.95	0.0469	0.0363		glycosyl transferases group 1 family protein	Y
Cluster_301670	15	21	9	2	63	91	AI	3.95	0.0469	0.0363	*escU*	escU	Y
Cluster_132520	6	13	18	10	25	57	AI	3.63	0.0569	0.0392		phage capsid scaffolding (GPO) serine peptidase family protein	N
Cluster_296985	6	13	18	10	25	57	AI	3.63	0.0569	0.0392		phage tail family protein	N
Cluster_108754	17	9	7	14	71	39	LI	3.58	0.0585	0.0415		phage terminase, small subunit, P27 family	N
Cluster_214595	15	7	9	16	63	30	LI	3.65	0.0561	0.0415		phage gp6-like head-tail connector family protein	N
Cluster_222329	17	9	7	14	71	39	LI	3.58	0.0585	0.0415		phage Terminase family protein	N
Cluster_309171	9	16	15	7	38	70	AI	3.65	0.0561	0.0415		outer membrane transport family protein	Y
Cluster_317314	17	9	7	14	71	39	LI	3.58	0.0585	0.0415		phage head-tail joining family protein	N
Cluster_317316	15	7	9	16	63	30	LI	3.65	0.0561	0.0415		putative bacteriophage protein	N
Cluster_346859	8	15	16	8	33	65	AI	3.59	0.0582	0.0422		fibronectin type III family protein	N
Cluster_128386	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		putative pcar	Y
Cluster_128751	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		HNH endonuclease family protein	N
Cluster_167317	23	17	1	6	96	74	LI	2.89	0.0891	0.0479		intergenic-region protein	N
Cluster_189418	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		conserved hypothetical protein	N
Cluster_235400	7	1	17	22	29	4	LI	3.52	0.0608	0.0479	*umuD*	protein UmuD	N
Cluster_239025	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		phage terminase, small subunit, P27 family	N
Cluster_287883	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		conserved hypothetical protein	N
Cluster_29580	7	1	17	22	29	4	LI	3.52	0.0608	0.0479		phage Terminase family protein	N
Cluster_307808	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		conserved hypothetical protein	N
Cluster_316818	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		putative predicted protein	N
Cluster_78613	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		putative prophage protein	N
Cluster_9132	1	6	23	17	4	26	AI	2.89	0.0891	0.0479		dinI-like family protein	N
Cluster_285169	14	20	10	3	58	87	AI	3.48	0.0619	0.0490		putative type III secretion system, LEE associated	N
Cluster_297448	14	20	10	3	58	87	AI	3.48	0.0619	0.0490		conserved hypothetical protein	N
Cluster_54644	14	20	10	3	58	87	AI	3.48	0.0619	0.0490		conserved hypothetical protein	N
Cluster_128404	21	14	3	9	88	61	LI	3.09	0.0787	0.0490		type II/IV secretion system family protein	N

Cluster_4127	3	9	21	14	13	39	AI	3.09	0.0787	0.0490		conserved hypothetical protein	N
Cluster_107544	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		tape measure domain protein	N
Cluster_107548	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_107738	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		putative membrane protein	Y
Cluster_108649	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_119046	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		putative tail fiber protein	N
Cluster_127350	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase	N
Cluster_163907	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyltransferase Family 4 family protein	N
Cluster_163915	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 2 family protein	N
Cluster_163986	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		hypothetical protein	N
Cluster_175608	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		degT/DnrJ/EryC1/StrS aminotransferase family protein	N
Cluster_183987	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		insA N-terminal domain protein	N
Cluster_183988	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		putative transposase	N
Cluster_247403	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 2 family protein	N
Cluster_261602	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_265998	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_266287	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		hypothetical protein	Y
Cluster_272862	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		helix-turn-helix domain protein	N
Cluster_276634	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		caudovirales tail fibre assembly family protein	N
Cluster_287882	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_291165	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		phage integrase family protein	N
Cluster_69486	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_69487	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_71994	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		glycosyl transferase 11 family protein	N
Cluster_72685	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		conserved hypothetical protein	N
Cluster_83926	5	0	19	23	21	0	LI	3.39	0.0654	0.0496		acetyltransferase domain protein	N
Cluster_171515	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		putative lipoprotein	Y
Cluster_171516	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		methyltransferase domain protein	N
Cluster_171517	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		beta-ketoacyl synthase, N-terminal domain protein	N
Cluster_171518	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		acyltransferase family protein	Y
Cluster_171520	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		phosphopantetheine attachment site family protein	N
Cluster_171523	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein	N
Cluster_171524	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		glycosyl transferase 2 family protein	N
Cluster_171527	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		MMPL family protein	Y
Cluster_171531	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		KR domain protein	N
Cluster_171532	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		beta-ketoacyl synthase, N-terminal domain protein	N
Cluster_17185	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein	Y
Cluster_172125	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein	Y
Cluster_188256	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein	Y
Cluster_203531	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		nucleoside 2-deoxyribosyltransferase family protein	N
Cluster_206025	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		phosphopantetheine attachment site family protein	N
Cluster_206027	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		putative membrane protein	Y
Cluster_206028	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		AMP-binding enzyme family protein	N
Cluster_206031	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		thioesterase-like superfamily protein	N
Cluster_206032	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		outer membrane lipocarrier LolA family protein	Y
Cluster_206034	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein	Y
Cluster_206035	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		beta-ketoacyl synthase, N-terminal domain protein	N
Cluster_206036	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		fabA-like domain protein	N
Cluster_206099	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein	N
Cluster_206540	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		hNH endonuclease family protein	N
Cluster_207740	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		caudovirales tail fibre assembly family protein	N
Cluster_297348	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		hypothetical protein	N
Cluster_307935	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		conserved hypothetical protein	N
Cluster_38836	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		phage integrase family protein	N
Cluster_8046	0	4	24	19	0	17	AI	2.60	0.1067	0.0496		hypothetical protein	N

Non-lethal Symptomatic (NSI) vs. Asymptomatic (AI)

	No. of Genomes												
	Present (≥0.8)[a,b]		Absent (<0.8)[a,b]		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	NSI	AI	NSI	AI	% NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_297448	7	20	16	3	30	87	AI	12.91	0.0003	0.0002		conserved hypothetical protein	N
Cluster_54644	8	20	15	3	35	87	AI	11.04	0.0009	0.0007		conserved hypothetical protein	N
Cluster_82579	7	18	16	5	30	78	AI	8.76	0.0031	0.0027	*espJ*	putative non-LEE-encoded effector EspJ	N
Cluster_34159	12	2	11	21	52	9	NSI	8.32	0.0039	0.0031		conserved hypothetical protein	N
Cluster_116275	5	15	18	8	22	65	AI	7.17	0.0074	0.0067		espF repeat family protein	Y
Cluster_183342	7	17	16	6	30	74	AI	7.06	0.0079	0.0072		hypothetical protein	Y
Cluster_267235	11	2	12	21	48	9	NSI	6.86	0.0088	0.0074		hypothetical protein	N
Cluster_132984	1	9	22	14	4	39	AI	6.26	0.0123	0.0098		traP family protein	Y
Cluster_337159	13	4	10	19	57	17	NSI	5.97	0.0145	0.0134		conserved hypothetical protein	N
Cluster_74298	13	4	10	19	57	17	NSI	5.97	0.0145	0.0134		traP family protein	Y
Cluster_307754	16	7	7	16	70	30	NSI	5.57	0.0183	0.0174		regulatory protein cro	N
Cluster_103192	6	0	17	23	26	0	NSI	4.79	0.0286	0.0216		putative phage protein	N
Cluster_147543	6	0	17	23	26	0	NSI	4.79	0.0286	0.0216		putative yfdA	Y
Cluster_167315	23	17	0	6	100	74	NSI	4.79	0.0286	0.0216		conserved hypothetical protein	N
Cluster_167317	23	17	0	6	100	74	NSI	4.79	0.0286	0.0216		intergenic-region protein	N
Cluster_189418	0	6	23	17	0	26	AI	4.79	0.0286	0.0216		conserved hypothetical protein	N
Cluster_9132	0	6	23	17	0	26	AI	4.79	0.0286	0.0216		dinI-like family protein	N
Cluster_128485	8	1	15	22	35	4	NSI	4.97	0.0257	0.0220		conserved hypothetical protein	Y
Cluster_147776	3	11	20	12	13	48	AI	5.03	0.0249	0.0230		conserved hypothetical protein	N
Cluster_232396	3	11	20	12	13	48	AI	5.03	0.0249	0.0230		putative dNA-binding helix-turn-helix protein	N
Cluster_169965	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		ogr/Delta-like zinc finger family protein	N
Cluster_200452	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		putative prophage protein	N
Cluster_200454	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		putative phage immunity repressor protein	N
Cluster_271578	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		conserved hypothetical protein	N
Cluster_302692	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		hypothetical protein	N
Cluster_316955	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		helix-turn-helix domain protein	N
Cluster_342331	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		putative bacteriophage protein	N
Cluster_62281	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		putative membrane protein	Y
Cluster_63459	5	13	18	10	22	57	AI	4.47	0.0345	0.0331		putative derepression protein	N
Cluster_84849	13	5	10	18	57	22	NSI	4.47	0.0345	0.0331		conserved hypothetical protein	N
Cluster_11587	2	9	21	14	9	39	AI	4.30	0.0381	0.0351		transposase family protein	Y
Cluster_183593	9	2	14	21	39	9	NSI	4.30	0.0381	0.0351		conserved hypothetical protein	N
Cluster_227577	2	9	21	14	9	39	AI	4.30	0.0381	0.0351	*ssb*	single-stranded DNA-binding protein	N
Cluster_262069	9	2	14	21	39	9	NSI	4.30	0.0381	0.0351		istB-like ATP binding family protein	N
Cluster_289981	9	2	14	21	39	9	NSI	4.30	0.0381	0.0351		integrase core domain protein	N
Cluster_64291	21	14	2	9	91	61	NSI	4.30	0.0381	0.0351	*folP*	dihydropteroate synthase	N

Cluster_281770	14	6	9	17	61	26	NSI	4.33	0.0373	0.0361		putative replication protein	N
Cluster_346882	14	6	9	17	61	26	NSI	4.33	0.0373	0.0361		conserved hypothetical protein	N
Cluster_309171	8	16	15	7	35	70	AI	4.27	0.0388	0.0377		outer membrane transport family protein	Y
Cluster_322029	15	7	8	16	65	30	NSI	4.27	0.0388	0.0377		hypothetical protein	N
Cluster_327168	15	7	8	16	65	30	NSI	4.27	0.0388	0.0377		helix-turn-helix domain protein	N
Cluster_147010	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		MULE transposase domain protein	N
Cluster_147193	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		resolvase, N terminal domain protein	N
Cluster_297043	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		hypothetical protein	N
Cluster_312558	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		conserved hypothetical protein	Y
Cluster_312602	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		HNH endonuclease family protein	N
Cluster_316814	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		AAA domain protein	N
Cluster_320753	1	7	22	16	4	30	AI	3.78	0.0518	0.0470		conserved hypothetical protein	N
Cluster_59342	7	1	16	22	30	4	NSI	3.78	0.0518	0.0470		conserved hypothetical protein	N
Cluster_69383	22	16	1	7	96	70	NSI	3.78	0.0518	0.0470		prophage minor tail Z family protein	N
Cluster_94384	22	16	1	7	96	70	NSI	3.78	0.0518	0.0470		transposase, Mutator family protein	N
Cluster_102885	20	13	3	10	87	57	NSI	3.86	0.0494	0.0472		integrase core domain protein	N
Cluster_24422	3	10	20	13	13	43	AI	3.86	0.0494	0.0472		conserved hypothetical protein	N
Cluster_168916	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		nucleoside transporter family protein	Y
Cluster_189417	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		acetyltransferase family protein	N
Cluster_191408	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein	N
Cluster_191933	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-oxoacid CoA-transferase, B subunit	N
Cluster_191937	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-hydroxybutyrate dehydrogenase family protein	Y
Cluster_207859	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein	N
Cluster_322942	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein	N
Cluster_51720	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-oxoacid CoA-transferase, A subunit	N
Cluster_51724	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		citrate transporter family protein	Y
Cluster_76780	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		lysR substrate binding domain protein	Y
Cluster_76783	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		acetyl-CoA C-acetyltransferase family protein	N
Cluster_7789	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		3-hydroxyacyl-CoA dehydrogenase, NAD binding domain protein	N
Cluster_79228	0	5	23	18	0	22	AI	3.59	0.0581	0.0491		conserved hypothetical protein	N
Cluster_266659	23	18	0	5	100	78	NSI	3.59	0.0581	0.0491	*repA*	incFII family plasmid replication initiator RepA	N

---

[a] Genes were grouped into gene clusters using an "inclusive" clustering threshold of ≥80% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.8, while those that are "absent" have a TBLASTN LS-BSR <0.8.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

Table S7. Genes identified by LS-BSR that are significantly correlated with one clinical outcome over another for the 61 typical EPEC genomes analyzed using an inclusive clustering threshold

Lethal (LI) vs. Non-lethal Symptomatic (NSI)

	No. of Genomes												
	Present (≥0.8)^{a,b}		Absent (<0.8)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI	NSI	LI	NSI	% LI	% NSI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_121796	0	4	24	16	0	20	NSI	2.94	0.0864	0.0402		caudovirales tail fibre assembly family protein	N
Cluster_121973	0	4	24	16	0	20	NSI	2.60	0.1067	0.0496		conserved hypothetical protein	N
Cluster_163539	0	4	24	16	0	20	NSI	2.76	0.0964	0.0448		hypothetical protein	N
Cluster_175606	6	0	18	20	25	0	LI	3.41	0.0649	0.0292	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase	N
Cluster_216712	1	6	23	14	4	30	NSI	2.89	0.0891	0.0479		putative membrane protein	Y
Cluster_222636	6	0	18	20	25	0	LI	3.41	0.0649	0.0292	*rfbA*	glucose-1-phosphate thymidylyltransferase	N
Cluster_99052	1	6	23	14	4	30	NSI	2.89	0.0891	0.0479		conserved hypothetical protein	Y

Lethal (LI) vs. Asymptomatic (AI)

	No. of Genomes												
	Present (≥0.8)^{a,b}		Absent (<0.8)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI	AI	LI	AI	% LI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_106122	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein	N
Cluster_106126	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_107351	11	1	13	16	46	6	LI	5.86	0.0155	0.0064		conserved hypothetical protein	N
Cluster_107804	20	9	4	8	83	53	LI	3.09	0.0786	0.0454		methyltransferase domain protein	N
Cluster_108754	17	6	7	11	71	35	LI	3.76	0.0524	0.0308		phage terminase, small subunit, P27 family	N
Cluster_112366	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		mu-like prophage FluMu gp41 family protein	N
Cluster_11511	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		putative exported protein	Y
Cluster_115298	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein	N
Cluster_115308	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		helix-turn-helix domain protein	Y
Cluster_11587	4	9	20	8	17	53	AI	4.49	0.0341	0.0198		transposase family protein	Y
Cluster_127453	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_13383	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein	N
Cluster_139135	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative pyridoxamine 5-phosphate-dependent dehydrase	N
Cluster_14414	10	1	14	16	42	6	LI	4.80	0.0285	0.0136		conserved hypothetical protein	N
Cluster_14627	5	11	19	6	21	65	AI	6.31	0.0120	0.0086		cysteine protease , YopT-type domain protein	N
Cluster_147543	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		putative yfdA	Y
Cluster_151023	6	11	18	6	25	65	AI	4.93	0.0264	0.0230		conserved hypothetical protein	Y
Cluster_152085	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_153090	2	6	22	11	8	35	AI	3.05	0.0808	0.0486		type II intron maturase family protein	N
Cluster_166190	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		phage integrase family protein	N
Cluster_167315	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		conserved hypothetical protein	N
Cluster_167317	23	11	1	6	96	65	LI	4.79	0.0286	0.0141		intergenic-region protein	N
Cluster_168916	0	5	24	12	0	29	AI	5.53	0.0187	0.0083		nucleoside transporter family protein	Y
Cluster_169965	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		ogr/Delta-like zinc finger family protein	N
Cluster_172125	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		conserved hypothetical protein	Y
Cluster_174672	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	Y
Cluster_175128	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		binding-protein-dependent transport system inner membrane component family protein	Y
Cluster_175131	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		bacterial transferase hexapeptide family protein	N
Cluster_175132	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative glycosidase	Y
Cluster_175134	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		archaeal transcriptional regulator TrmB family protein	N
Cluster_175606	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase	N
Cluster_176094	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		NIF3 family protein	N
Cluster_176098	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		alpha amylase, catalytic domain protein	N
Cluster_176426	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_176531	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		hypothetical protein	N
Cluster_177188	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		phage integrase family protein	N
Cluster_178309	22	9	2	8	92	53	LI	6.13	0.0133	0.0083		conserved hypothetical protein	N
Cluster_182012	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		phospholipase_D-nuclease N-terminal family protein	Y
Cluster_183140	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122	Y
Cluster_183593	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		conserved hypothetical protein	N
Cluster_183787	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		3-octaprenyl-4-hydroxybenzoate carboxy-lyase family protein	N
Cluster_200452	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		putative prophage protein	N
Cluster_200454	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		putative phage immunity repressor protein	N
Cluster_203531	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		nucleoside 2-deoxyribosyltransferase family protein	N
Cluster_206540	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		hNH endonuclease family protein	N
Cluster_207773	2	6	22	11	8	35	AI	3.05	0.0808	0.0486		collagen triple helix repeat family protein	N
Cluster_209146	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		shET2 enterotoxin, N-terminal region family protein	N
Cluster_214595	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		phage gp6-like head-tail connector family protein	N
Cluster_221116	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		hypothetical protein	N
Cluster_221922	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		glycosyl transferase family 2 domain protein	N
Cluster_221927	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		phage Tail Collar domain protein	Y
Cluster_222329	17	6	7	11	71	35	LI	3.76	0.0524	0.0308		phage Terminase family protein	N
Cluster_222331	16	4	8	13	67	24	LI	5.79	0.0162	0.0109		phage portal protein, HK97 family	N
Cluster_222333	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		phage major capsid protein, HK97 family	N
Cluster_222426	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		putative membrane protein	Y
Cluster_222636	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*rfbA*	glucose-1-phosphate thymidylyltransferase	N
Cluster_225859	4	9	20	8	17	53	AI	4.49	0.0341	0.0198		putative glyco3, capsid size determination protein Sid	N
Cluster_225861	3	8	21	9	13	47	AI	4.42	0.0355	0.0292	*psu*	polarity suppression protein	N
Cluster_227359	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		conserved hypothetical protein	N
Cluster_227360	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		putative membrane protein	Y
Cluster_232949	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		P22AR C-terminal domain protein	N
Cluster_236134	16	5	8	12	67	29	LI	4.14	0.0419	0.0278	*dhfrI*	dihydrofolate reductase type 1	N
Cluster_237277	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		conserved hypothetical protein	N
Cluster_237298	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_237299	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_23758	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		bacterial extracellular solute-binding family protein	Y
Cluster_240829	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		hypothetical protein	N
Cluster_247401	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		polysaccharide biosynthesis/export family protein	N
Cluster_251045	8	12	16	5	33	71	AI	4.14	0.0419	0.0278		phage tail tape measure protein, TP901 family, core region	Y
Cluster_252971	20	7	4	10	83	41	LI	6.10	0.0135	0.0079		HNH endonuclease family protein	N
Cluster_25983	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		bacteriophage replication gene A family protein	N
Cluster_260984	13	2	11	15	54	12	LI	5.99	0.0144	0.0082	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_261817	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		putative predicted protein	N
Cluster_262069	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		istB-like ATP binding family protein	N
Cluster_262291	17	17	7	0	71	100	AI	4.10	0.0430	0.0295		type III secretion , HrpO family protein	Y
Cluster_263620	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		transposase DDE domain protein	N
Cluster_263945	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein	N
Cluster_267235	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		hypothetical protein	N
Cluster_271134	2	6	22	11	8	35	AI	3.05	0.0808	0.0486		outer membrane autotransporter barrel domain protein	Y
Cluster_271578	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		conserved hypothetical protein	N
Cluster_276264	11	2	13	15	46	12	LI	3.88	0.0490	0.0391		replication regulatory RepB family protein	N
Cluster_289981	12	2	12	15	50	12	LI	4.88	0.0272	0.0183		integrase core domain protein	N
Cluster_291294	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative pyridoxamine 5-phosphate-dependent dehydrase	Y
Cluster_296985	6	11	18	6	25	65	AI	4.93	0.0264	0.0230		phage tail family protein	N

Cluster_297348	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		hypothetical protein	N
Cluster_297449	2	8	22	9	8	47	AI	6.13	0.0133	0.0083		conserved hypothetical protein	N
Cluster_307808	1	6	23	11	4	35	AI	4.79	0.0286	0.0141		conserved hypothetical protein	N
Cluster_307935	0	4	24	13	0	24	AI	3.87	0.0491	0.0235		conserved hypothetical protein	N
Cluster_309171	9	13	15	4	38	76	AI	4.61	0.0318	0.0252		outer membrane transport family protein	Y
Cluster_312794	11	1	13	16	46	6	LI	5.86	0.0155	0.0064		transposase DDE domain protein	N
Cluster_314972	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_315365	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		ABC transporter family protein	Y
Cluster_315371	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		binding-protein-dependent transport system inner membrane component family protein	Y
Cluster_315372	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		alpha amylase, N-terminal ig-like domain protein	N
Cluster_316214	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		tatD related DNase family protein	N
Cluster_316955	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		helix-turn-helix domain protein	N
Cluster_317314	17	6	7	11	71	35	LI	3.76	0.0524	0.0308		phage head-tail joining family protein	N
Cluster_317316	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		putative bacteriophage protein	N
Cluster_317552	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein	N
Cluster_317698	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		caudovirales tail fibre assembly family protein	N
Cluster_320510	6	11	18	6	25	65	AI	4.93	0.0264	0.0230		conserved hypothetical protein	Y
Cluster_322330	6	11	18	6	25	65	AI	4.93	0.0264	0.0230		conserved hypothetical protein	N
Cluster_322352	7	11	17	6	29	65	AI	3.76	0.0524	0.0308		conserved hypothetical protein	N
Cluster_334902	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		conserved hypothetical protein	N
Cluster_340924	15	16	9	1	63	94	AI	3.82	0.0508	0.0281		ribbon-helix-helix , copG family protein	N
Cluster_34159	13	2	11	15	54	12	LI	5.99	0.0144	0.0082		conserved hypothetical protein	N
Cluster_341980	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		conserved hypothetical protein	Y
Cluster_346831	16	5	8	12	67	29	LI	4.14	0.0419	0.0278		conserved hypothetical protein	N
Cluster_39077	17	6	7	11	71	35	LI	3.76	0.0524	0.0308		conserved hypothetical protein	N
Cluster_40179	18	17	6	0	75	100	AI	3.18	0.0746	0.0327		prophage minor tail Z family protein	N
Cluster_42814	4	8	20	9	17	47	AI	3.09	0.0786	0.0454		glycosyltransferase like 2 family protein	N
Cluster_43605	18	6	6	11	75	35	LI	4.93	0.0264	0.0230		HNH endonuclease family protein	N
Cluster_53910	1	6	23	11	4	35	AI	4.79	0.0286	0.0141	*intA*	prophage CP4-57 integrase	N
Cluster_56014	22	11	2	6	92	65	LI	3.05	0.0808	0.0486		hydantoinase/oxoprolinase family protein	N
Cluster_57681	6	11	18	6	25	65	AI	4.93	0.0264	0.0230		conserved hypothetical protein	N
Cluster_58277	10	1	14	16	42	6	LI	4.80	0.0285	0.0136		helix-turn-helix, Psq domain protein	N
Cluster_59004	17	4	7	13	71	24	LI	7.12	0.0076	0.0044		phage prohead protease, HK97 family	N
Cluster_63459	4	10	20	7	17	59	AI	6.10	0.0135	0.0079		putative derepression protein	N
Cluster_67359	1	6	23	11	4	35	AI	4.79	0.0286	0.0141	*intA*	prophage CP4-57 integrase	N
Cluster_72088	9	1	15	16	38	6	LI	3.82	0.0508	0.0281		glycosyl transferases group 1 family protein	Y
Cluster_74193	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		phage major capsid E family protein	N
Cluster_80024	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		esterase family protein	N
Cluster_80973	15	4	9	13	63	24	LI	4.61	0.0318	0.0252		phage integrase family protein	N
Cluster_82368	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*fliC*	flagellin	N
Cluster_83254	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		glycosyl transferases group 1 family protein	N
Cluster_83718	7	12	17	5	29	71	AI	5.30	0.0213	0.0124	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_83792	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		conserved hypothetical protein	N
Cluster_83828	19	8	5	9	79	47	LI	3.25	0.0716	0.0476		conserved hypothetical protein	N
Cluster_83929	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*rfbB*	dTDP-glucose 4,6-dehydratase	N
Cluster_84063	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		polysaccharide biosynthesis/export family protein	Y
Cluster_84140	6	0	18	17	25	0	LI	3.18	0.0746	0.0327		putative cymA protein	Y
Cluster_95058	7	0	17	17	29	0	LI	4.10	0.0430	0.0295		putative antirepressor protein	N
Cluster_97528	8	0	16	17	33	0	LI	5.08	0.0242	0.0128		ead/Ea22-like family protein	N
Cluster_97572	6	0	18	17	25	0	LI	3.18	0.0746	0.0327	*ynjE*	thiosulfate sulfurtransferase YnjE domain protein	Y

Non-lethal Symptomatic (NSI) vs. Asymptomatic (AI)

	No. of Genomes												
	Present (≥0.8)[a,b]		Absent (<0.8)[a,b]		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	NSI	AI	NSI	AI	% NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_103192	6	0	14	17	30	0	NSI	4.08	0.0434	0.0220		putative phage protein	N
Cluster_111596	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		ethanolamine utilization EutN/carboxysome family protein	N
Cluster_11511	2	8	18	9	10	47	AI	4.66	0.0309	0.0234		putative exported protein	Y
Cluster_11587	2	9	18	8	10	53	AI	6.19	0.0129	0.0097		transposase family protein	Y
Cluster_116275	5	11	15	6	25	65	AI	4.40	0.0360	0.0220		espF repeat family protein	Y
Cluster_118076	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduT*	shell protein	N
Cluster_118816	12	3	8	14	60	18	NSI	5.19	0.0227	0.0176		conserved hypothetical protein	Y
Cluster_127742	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		dehydratase medium subunit	N
Cluster_127750	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		iron-containing alcohol dehydrogenase family protein	N
Cluster_128485	8	1	12	16	40	6	NSI	4.11	0.0428	0.0227		conserved hypothetical protein	Y
Cluster_128654	16	8	4	9	80	47	NSI	3.05	0.0808	0.0470		conserved hypothetical protein	Y
Cluster_132984	1	7	19	10	5	41	AI	5.12	0.0236	0.0140		traP family protein	Y
Cluster_147193	1	6	19	11	5	35	AI	3.70	0.0544	0.0335		resolvase, N terminal domain protein	N
Cluster_147543	6	0	14	17	30	0	NSI	4.08	0.0434	0.0220		putative yfdA	Y
Cluster_147776	1	8	19	9	5	47	AI	6.69	0.0097	0.0055		conserved hypothetical protein	N
Cluster_151023	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		conserved hypothetical protein	Y
Cluster_159210	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		conserved hypothetical protein	N
Cluster_166468	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		cob(I)alamin adenosyltransferase	N
Cluster_166473	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduA*	propanediol utilization protein PduA	N
Cluster_167315	20	11	0	6	100	65	NSI	6.03	0.0141	0.0053		conserved hypothetical protein	N
Cluster_167317	20	11	0	6	100	65	NSI	6.03	0.0141	0.0053		intergenic-region protein	N
Cluster_168916	0	5	20	12	0	29	AI	4.52	0.0335	0.0142		nucleoside transporter family protein	Y
Cluster_169965	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		ogr/Delta-like zinc finger family protein	N
Cluster_172125	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		conserved hypothetical protein	Y
Cluster_183342	7	13	13	4	35	76	AI	4.80	0.0284	0.0202		hypothetical protein	Y
Cluster_186173	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		BMC domain protein	Y
Cluster_196208	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*eutP*	ethanolamine utilization protein, EutP	N
Cluster_196209	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduU*	propanediol utilization protein PduU	N
Cluster_196213	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		aldehyde dehydrogenase family protein	N
Cluster_196229	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		putative cobalamin biosynthesis protein CbiG	Y
Cluster_197986	16	8	4	9	80	47	NSI	3.05	0.0808	0.0470		tellurite resistance TerB family protein	N
Cluster_200452	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		putative prophage protein	N
Cluster_200454	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		putative phage immunity repressor protein	N
Cluster_203531	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		nucleoside 2-deoxyribosyltransferase family protein	N
Cluster_204051	5	10	15	7	25	59	AI	3.07	0.0797	0.0498		putative valyl-tRNA synthetase	N
Cluster_206540	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		hNH endonuclease family protein	N
Cluster_221927	6	0	14	17	30	0	NSI	4.08	0.0434	0.0220		phage Tail Collar domain protein	Y
Cluster_225773	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduD*	propanediol dehydratase medium subunit	N
Cluster_225859	4	9	16	8	20	53	AI	3.05	0.0808	0.0470		putative glyco3, capsid size determination protein Sid	N
Cluster_227577	2	8	18	9	10	47	AI	4.66	0.0309	0.0234	*ssb*	single-stranded DNA-binding protein	N
Cluster_230774	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduL*	phosphate propanoyltransferase	N
Cluster_232396	1	8	19	9	5	47	AI	6.69	0.0097	0.0055		putative dNA-binding helix-turn-helix protein	N
Cluster_236053	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		conserved hypothetical protein	N
Cluster_252493	1	6	19	11	5	35	AI	3.70	0.0544	0.0335		putative arginyl-tRNA synthetase	N
Cluster_259784	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduA*	propanediol utilization protein PduA	N
Cluster_259786	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduC*	propanediol dehydratase large subunit	N
Cluster_259788	8	14	12	3	40	82	AI	5.19	0.0227	0.0176	*pduE*	propanediol dehydratase small subunit	N
Cluster_261617	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		conserved hypothetical protein	Y
Cluster_265851	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		BMC domain protein	N

Cluster_265865	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		electron transport complex, RnfABCDGE type, C subunit	N
Cluster_267235	9	2	11	15	45	12	NSI	3.40	0.0653	0.0365		hypothetical protein	N
Cluster_271578	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		conserved hypothetical protein	N
Cluster_285156	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481	*espZ*	sepZ family protein	Y
Cluster_297043	1	6	19	11	5	35	AI	3.70	0.0544	0.0335		hypothetical protein	N
Cluster_297448	7	14	13	3	35	82	AI	6.58	0.0103	0.0071		conserved hypothetical protein	N
Cluster_297449	2	8	18	9	10	47	AI	4.66	0.0309	0.0234		conserved hypothetical protein	N
Cluster_302692	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		hypothetical protein	N
Cluster_303767	5	10	15	7	25	59	AI	3.07	0.0797	0.0498	*CI*	repressor protein CI	N
Cluster_309066	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		conserved hypothetical protein	N
Cluster_309171	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		outer membrane transport family protein	Y
Cluster_310627	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		helix-turn-helix domain protein	N
Cluster_312602	1	7	19	10	5	41	AI	5.12	0.0236	0.0140		HNH endonuclease family protein	N
Cluster_316955	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		helix-turn-helix domain protein	N
Cluster_320510	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		conserved hypothetical protein	Y
Cluster_321136	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		MIP channel family protein	Y
Cluster_322029	13	4	7	13	65	24	NSI	4.80	0.0284	0.0202		hypothetical protein	N
Cluster_322330	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		conserved hypothetical protein	N
Cluster_322352	5	11	15	6	25	65	AI	4.40	0.0360	0.0220		conserved hypothetical protein	N
Cluster_326585	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		putative protein TraJ	N
Cluster_334902	5	0	15	17	25	0	NSI	3.01	0.0829	0.0498		conserved hypothetical protein	N
Cluster_337159	13	4	7	13	65	24	NSI	4.80	0.0284	0.0202		conserved hypothetical protein	N
Cluster_340924	13	16	7	1	65	94	AI	3.04	0.0813	0.0481		ribbon-helix-helix , copG family protein	N
Cluster_34159	10	2	10	15	50	12	NSI	4.51	0.0337	0.0173		conserved hypothetical protein	N
Cluster_342331	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		putative bacteriophage protein	N
Cluster_54451	5	10	15	7	25	59	AI	3.07	0.0797	0.0498		conserved hypothetical protein	N
Cluster_54644	8	14	12	3	40	82	AI	5.19	0.0227	0.0176		conserved hypothetical protein	N
Cluster_57681	4	11	16	6	20	65	AI	5.88	0.0153	0.0084		conserved hypothetical protein	N
Cluster_59342	7	1	13	16	35	6	NSI	3.04	0.0813	0.0481		conserved hypothetical protein	N
Cluster_60983	0	4	20	13	0	24	AI	3.12	0.0774	0.0360		conserved hypothetical protein	N
Cluster_63459	4	10	16	7	20	59	AI	4.35	0.0369	0.0210		putative derepression protein	N
Cluster_72658	8	13	12	4	40	76	AI	3.60	0.0576	0.0453		putative propanediol utilization protein	N
Cluster_74298	12	4	8	13	60	24	NSI	3.60	0.0576	0.0453		traP family protein	Y
Cluster_74308	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489	*traY*	relaxosome protein TraY	N
Cluster_79228	0	5	20	12	0	29	AI	4.52	0.0335	0.0142		conserved hypothetical protein	N
Cluster_82579	7	14	13	3	35	82	AI	6.58	0.0103	0.0071	*espJ*	putative non-LEE-encoded effector EspJ	N
Cluster_84849	13	5	7	12	65	29	NSI	3.34	0.0675	0.0489		conserved hypothetical protein	N
Cluster_89529	5	10	15	7	25	59	AI	3.07	0.0797	0.0498		hypothetical protein	N

[a] Genes were grouped into gene clusters using an "inclusive" clustering threshold of ≥80% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.8, while those that are "absent" have a TBLASTN LS-BSR <0.8.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

DOI: 10.1038/NMICROBIOL.2015.14

Table S8. Genes identified by LS-BSR that are significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for all 70 EPEC genomes analyzed using an inclusive clustering threshold

Symptomatic (LI+NSI) vs. Asymptomatic (AI)													
	No. of Genomes												
	Present (≥0.8)^{a,b}		Absent (<0.8)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI+NSI	AI	LI+NSI	AI	% LI+NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_34159	25	2	22	21	53	9	LI+NSI	11.09	0.0009	0.0003		conserved hypothetical protein	N
Cluster_297448	21	20	26	3	45	87	AI	9.70	0.0018	0.0008		conserved hypothetical protein	N
Cluster_267235	23	2	24	21	49	9	LI+NSI	9.21	0.0024	0.0011		hypothetical protein	N
Cluster_54644	22	20	25	3	47	87	AI	8.77	0.0031	0.0016		conserved hypothetical protein	N
Cluster_307754	34	7	13	16	72	30	LI+NSI	9.52	0.0020	0.0016		regulatory protein cro	N
Cluster_327168	33	7	14	16	70	30	LI+NSI	8.42	0.0037	0.0022		helix-turn-helix domain protein	N
Cluster_169965	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		ogr/Delta-like zinc finger family protein	N
Cluster_200452	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		putative prophage protein	N
Cluster_200454	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		putative phage immunity repressor protein	N
Cluster_271578	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		conserved hypothetical protein	N
Cluster_316955	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		helix-turn-helix domain protein	N
Cluster_63459	9	13	38	10	19	57	AI	8.35	0.0039	0.0025		putative derepression protein	N
Cluster_262069	21	2	26	21	45	9	LI+NSI	7.51	0.0061	0.0027		istB-like ATP binding family protein	N
Cluster_289981	21	2	26	21	45	9	LI+NSI	7.51	0.0061	0.0027		integrase core domain protein	N
Cluster_168916	0	5	47	18	0	22	AI	7.97	0.0048	0.0028		nucleoside transporter family protein	Y
Cluster_266659	47	18	0	5	100	78	LI+NSI	7.97	0.0048	0.0028	*repA*	incFII family plasmid replication initiator RepA	N
Cluster_147543	14	0	33	23	30	0	LI+NSI	6.80	0.0091	0.0029		putative yfdA	Y
Cluster_167317	46	17	1	6	98	74	LI+NSI	7.37	0.0066	0.0042		intergenic-region protein	N
Cluster_189418	1	6	46	17	2	26	AI	7.37	0.0066	0.0042		conserved hypothetical protein	N
Cluster_9132	1	6	46	17	2	26	AI	7.37	0.0066	0.0042		dinI-like family protein	N
Cluster_316814	2	7	45	16	4	30	AI	7.25	0.0071	0.0044		AAA domain protein	N
Cluster_94384	45	16	2	7	96	70	LI+NSI	7.25	0.0071	0.0044		transposase, Mutator family protein	N
Cluster_337159	25	4	22	19	53	17	LI+NSI	6.75	0.0094	0.0049		conserved hypothetical protein	N
Cluster_74298	25	4	22	19	53	17	LI+NSI	6.75	0.0094	0.0049		traP family protein	Y
Cluster_183593	20	2	27	21	43	9	LI+NSI	6.72	0.0095	0.0054		conserved hypothetical protein	N
Cluster_276264	19	2	28	21	40	9	LI+NSI	5.97	0.0146	0.0063		replication regulatory RepB family protein	N
Cluster_341980	40	12	7	11	85	52	LI+NSI	7.13	0.0076	0.0073		conserved hypothetical protein	Y
Cluster_166870	36	10	11	13	77	43	LI+NSI	6.12	0.0134	0.0083		integron integrase family protein	N
Cluster_157448	5	9	42	14	11	39	AI	6.16	0.0131	0.0094	*espS*	putative T3SS effector protein EspS	N
Cluster_171515	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		putative lipoprotein	Y
Cluster_171516	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		methyltransferase domain protein	N
Cluster_171517	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		beta-ketoacyl synthase, N-terminal domain protein	N
Cluster_171518	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		acyltransferase family protein	Y
Cluster_171520	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		phosphopantetheine attachment site family protein	N
Cluster_171523	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein	N
Cluster_171524	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		glycosyl transferase 2 family protein	N
Cluster_171527	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		MMPL family protein	Y
Cluster_171531	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		KR domain protein	N
Cluster_171532	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		beta-ketoacyl synthase, N-terminal domain protein	N
Cluster_172125	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein	Y
Cluster_203531	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		nucleoside 2-deoxyribosyltransferase family protein	N
Cluster_206025	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		phosphopantetheine attachment site family protein	N
Cluster_206027	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		putative membrane protein	Y
Cluster_206028	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		AMP-binding enzyme family protein	N
Cluster_206031	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		thioesterase-like superfamily protein	N
Cluster_206032	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		outer membrane lipocarrier LolA family protein	Y
Cluster_206034	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein	Y
Cluster_206035	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		beta-ketoacyl synthase, N-terminal domain protein	N
Cluster_206036	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		fabA-like domain protein	N
Cluster_206099	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		conserved hypothetical protein	N
Cluster_206540	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		hNH endonuclease family protein	N
Cluster_207740	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		caudovirales tail fibre assembly family protein	N
Cluster_38836	0	4	47	19	0	17	AI	5.74	0.0166	0.0097		phage integrase family protein	N
Cluster_83828	34	9	13	14	72	39	LI+NSI	5.86	0.0155	0.0098		conserved hypothetical protein	N
Cluster_225861	8	11	39	12	17	48	AI	5.93	0.0148	0.0100	*psu*	polarity suppression protein	N
Cluster_302692	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		hypothetical protein	N
Cluster_82579	21	18	26	5	45	78	AI	5.76	0.0164	0.0105	*espJ*	putative non-LEE-encoded effector EspJ	N
Cluster_84849	26	5	21	18	55	22	LI+NSI	5.76	0.0164	0.0105		conserved hypothetical protein	N
Cluster_116275	15	15	32	8	32	65	AI	5.70	0.0170	0.0108		espF repeat family protein	Y
Cluster_225859	9	12	38	11	19	52	AI	6.52	0.0106	0.0111		putative glyco3, capsid size determination protein Sid	N
Cluster_309171	17	16	30	7	36	70	AI	5.64	0.0176	0.0113		outer membrane transport family protein	Y
Cluster_297253	44	16	3	7	94	70	LI+NSI	5.46	0.0194	0.0115		transposase, Mutator family protein	Y
Cluster_262291	36	23	11	0	77	100	AI	4.74	0.0294	0.0122		type III secretion, HrpO family protein	Y
Cluster_167315	45	17	2	6	96	74	LI+NSI	5.27	0.0216	0.0128		conserved hypothetical protein	N
Cluster_189417	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		acetyltransferase family protein	N
Cluster_191408	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein	N
Cluster_191933	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-oxoacid CoA-transferase, B subunit	N
Cluster_191937	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-hydroxybutyrate dehydrogenase family protein	Y
Cluster_207859	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein	N
Cluster_316809	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein	N
Cluster_316818	2	6	45	17	4	26	AI	5.27	0.0216	0.0128		putative predicted protein	N
Cluster_3723	2	6	45	17	4	26	AI	5.27	0.0216	0.0128		conserved hypothetical protein	N
Cluster_51720	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-oxoacid CoA-transferase, A subunit	N
Cluster_51724	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		citrate transporter family protein	Y
Cluster_76780	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		lysR substrate binding domain protein	Y
Cluster_76783	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		acetyl-CoA C-acetyltransferase family protein	N
Cluster_7789	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		3-hydroxyacyl-CoA dehydrogenase, NAD binding domain protein	N
Cluster_78613	2	6	45	17	4	26	AI	5.27	0.0216	0.0128		putative prophage protein	N
Cluster_79228	1	5	46	18	2	22	AI	5.28	0.0215	0.0128		conserved hypothetical protein	N
Cluster_297449	4	8	43	15	9	35	AI	5.77	0.0163	0.0145		conserved hypothetical protein	N
Cluster_128383	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein	N
Cluster_128390	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		viral (Super1) RNA helicase family protein	N
Cluster_164940	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		AAA domain protein	N
Cluster_183140	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122	Y
Cluster_276697	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein	N
Cluster_291795	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		putative acetyltransferase	Y
Cluster_58277	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		helix-turn-helix, Psq domain protein	N
Cluster_59092	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein	N
Cluster_73497	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein	N
Cluster_98944	14	1	33	22	30	4	LI+NSI	4.52	0.0335	0.0147		conserved hypothetical protein	N
Cluster_303065	23	4	24	19	49	17	LI+NSI	5.22	0.0223	0.0176		conserved hypothetical protein	N
Cluster_48952	23	4	24	19	49	17	LI+NSI	5.22	0.0223	0.0176		conserved hypothetical protein	Y
Cluster_342331	25	5	22	18	53	22	LI+NSI	5.02	0.0251	0.0198		putative bacteriophage protein	N
Cluster_56633	25	5	22	18	53	22	LI+NSI	5.02	0.0251	0.0198		conserved hypothetical protein	N
Cluster_107351	17	2	30	21	36	9	LI+NSI	4.59	0.0322	0.0209		conserved hypothetical protein	N
Cluster_312794	17	2	30	21	36	9	LI+NSI	4.59	0.0322	0.0209		transposase DDE domain protein	N
Cluster_346831	27	6	20	17	57	26	LI+NSI	4.90	0.0268	0.0211		conserved hypothetical protein	N
Cluster_210977	5	8	42	15	11	35	AI	4.46	0.0346	0.0223		conserved hypothetical protein	N
Cluster_147776	9	11	38	12	19	48	AI	4.90	0.0269	0.0227		conserved hypothetical protein	N

Cluster_232396	9	11	38	12	19	48	AI	4.90	0.0269	0.0227		putative dNA-binding helix-turn-helix protein	N
Cluster_102885	39	13	8	10	83	57	LI+NSI	4.36	0.0368	0.0228		integrase core domain protein	N
Cluster_56999	8	10	39	13	17	43	AI	4.36	0.0368	0.0228		putative P4-specific DNA primase	N
Cluster_183342	21	17	26	6	45	74	AI	4.20	0.0403	0.0244		hypothetical protein	Y
Cluster_47222	26	6	21	17	55	26	LI+NSI	4.20	0.0403	0.0244		conserved hypothetical protein	N
Cluster_139135	10	0	37	23	21	0	LI+NSI	4.10	0.0428	0.0245		putative pyridoxamine 5-phosphate-dependent dehydrase	N
Cluster_166190	10	0	37	23	21	0	LI+NSI	4.10	0.0428	0.0245		phage integrase family protein	N
Cluster_291294	10	0	37	23	21	0	LI+NSI	4.10	0.0428	0.0245		putative pyridoxamine 5-phosphate-dependent dehydrase	Y
Cluster_103192	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative phage protein	N
Cluster_106122	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_107548	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_107738	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative membrane protein	Y
Cluster_108649	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_115298	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_119046	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative tail fiber protein	N
Cluster_127350	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase	N
Cluster_152085	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_163907	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		glycosyltransferase Family 4 family protein	N
Cluster_163915	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		glycosyl transferase 2 family protein	N
Cluster_163986	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		hypothetical protein	N
Cluster_240829	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		hypothetical protein	N
Cluster_25983	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		bacteriophage replication gene A family protein	N
Cluster_261602	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_261817	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		putative predicted protein	N
Cluster_263945	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_265998	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_266287	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		hypothetical protein	Y
Cluster_272862	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		helix-turn-helix domain protein	N
Cluster_276634	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		caudovirales tail fibre assembly family protein	N
Cluster_287882	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_291165	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		phage integrase family protein	N
Cluster_69486	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_69487	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_71994	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		glycosyl transferase 11 family protein	N
Cluster_72685	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		conserved hypothetical protein	N
Cluster_74193	9	0	38	23	19	0	LI+NSI	3.49	0.0618	0.0254		phage major capsid E family protein	N
Cluster_128386	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		putative pcar	Y
Cluster_128485	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		conserved hypothetical protein	Y
Cluster_14414	13	1	34	22	28	4	LI+NSI	3.89	0.0486	0.0262		conserved hypothetical protein	N
Cluster_11587	6	9	41	14	13	39	AI	4.91	0.0268	0.0269		transposase family protein	Y
Cluster_132984	6	9	41	14	13	39	AI	4.91	0.0268	0.0269		traP family protein	Y
Cluster_178309	41	14	6	9	87	61	LI+NSI	4.91	0.0268	0.0269		conserved hypothetical protein	N
Cluster_64291	41	14	6	9	87	61	LI+NSI	4.91	0.0268	0.0269	*folP*	dihydropteroate synthase	N
Cluster_177188	19	3	28	20	40	13	LI+NSI	4.18	0.0410	0.0279		phage integrase family protein	N
Cluster_260984	19	3	28	20	40	13	LI+NSI	4.18	0.0410	0.0279	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_88073	36	11	11	12	77	48	LI+NSI	4.56	0.0327	0.0288		small Multidrug Resistance family protein	Y
Cluster_151023	10	11	37	12	21	48	AI	4.00	0.0456	0.0292		conserved hypothetical protein	Y
Cluster_320510	10	11	37	12	21	48	AI	4.00	0.0456	0.0292		conserved hypothetical protein	Y
Cluster_322330	10	11	37	12	21	48	AI	4.00	0.0456	0.0292		conserved hypothetical protein	N
Cluster_57681	10	11	37	12	21	48	AI	4.00	0.0456	0.0292		conserved hypothetical protein	N
Cluster_90939	37	12	10	11	79	52	LI+NSI	4.00	0.0456	0.0292		plasmid stability family protein	N
Cluster_1253	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		glycosyl transferases group 1 family protein	N
Cluster_1259	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		N-acetyl sugar amidotransferase family protein	N
Cluster_1317	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		phage tail-collar fibre family protein	Y
Cluster_170088	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_171495	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_171552	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		binding-protein-dependent transport system inner membrane component family protein	Y
Cluster_171553	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		binding-protein-dependent transport system inner membrane component family protein	Y
Cluster_171556	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		ROK family protein	N
Cluster_191111	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_191112	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_191123	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	Y
Cluster_191127	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_191132	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_191136	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		hypothetical protein	N
Cluster_206057	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		bacterial extracellular solute-binding family protein	Y
Cluster_206061	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		glycerophosphoryl diester phosphodiesterase family protein	N
Cluster_207555	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		recF/RecN/SMC N terminal domain protein	N
Cluster_207556	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_207558	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	Y
Cluster_207848	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_2357	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_297042	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		integrase core domain protein	N
Cluster_297563	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_319815	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		putative predicted protein	N
Cluster_319816	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	Y
Cluster_321372	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		ABC transporter family protein	N
Cluster_36336	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		cupin fold metallo, WbuC family protein	N
Cluster_36339	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		rmlD substrate binding domain protein	N
Cluster_36341	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		dihydrouridine synthase family protein	N
Cluster_36342	0	3	47	20	0	13	AI	3.62	0.0571	0.0324	*hisH*	imidazole glycerol phosphate synthase, glutamine amidotransferase subunit	N
Cluster_36450	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		type III restriction enzyme, res subunit	N
Cluster_36837	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_38630	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_38770	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_38775	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_38777	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_3917	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		putative cytochrome P450 51	N
Cluster_39213	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		polysaccharide deacetylase family protein	N
Cluster_3924	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		caudovirus prohead protease family protein	N
Cluster_39245	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_3925	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		phage portal protein, lambda family	N
Cluster_39264	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		ead/Ea22-like family protein	N
Cluster_3934	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_39356	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		P63C domain protein	N
Cluster_39363	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_4140	0	3	47	20	0	13	AI	3.62	0.0571	0.0324	*espN*	putative t3SS effector protein EspN2-2	N
Cluster_4584	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_4591	0	3	47	20	0	13	AI	3.62	0.0571	0.0324	*msbB*	lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase	Y
Cluster_6512	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		phage antitermination Q family protein	N
Cluster_66187	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		N-acetyl sugar amidotransferase family protein	N
Cluster_6886	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		transposase family protein	N
Cluster_69143	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N
Cluster_7205	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		phage tail fibre repeat family protein	N
Cluster_8866	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	Y
Cluster_8868	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		glycosyltransferase Family 4 family protein	N
Cluster_94403	0	3	47	20	0	13	AI	3.62	0.0571	0.0324		conserved hypothetical protein	N

Cluster_321659	43	16	4	7	91	70	LI+NSI	4.07	0.0436	0.0324	*bfpA*	major structural subunit of bundle-forming pilus	Y
Cluster_336502	43	16	4	7	91	70	LI+NSI	4.07	0.0436	0.0324		type II/IV secretion system family protein	N
Cluster_53910	4	7	43	16	9	30	AI	4.07	0.0436	0.0324	*intA*	prophage CP4-57 integrase	N
Cluster_67359	4	7	43	16	9	30	AI	4.07	0.0436	0.0324	*intA*	prophage CP4-57 integrase	N
Cluster_238652	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339		winged helix-turn-helix DNA-binding family protein	N
Cluster_78370	21	4	26	19	45	17	LI+NSI	3.89	0.0485	0.0339		putative predicted protein	N
Cluster_322942	2	5	45	18	4	22	AI	3.48	0.0620	0.0345		conserved hypothetical protein	N
Cluster_119706	23	5	24	18	49	22	LI+NSI	3.69	0.0546	0.0385		conserved hypothetical protein	N
Cluster_132520	14	13	33	10	30	57	AI	3.60	0.0578	0.0391		phage capsid scaffolding (GPO) serine peptidase family protein	N
Cluster_43705	30	8	17	15	64	35	LI+NSI	4.15	0.0418	0.0398		conserved hypothetical protein	Y
Cluster_301670	32	21	15	2	68	91	AI	3.35	0.0671	0.0401	*escU*	*escU*	Y
Cluster_263620	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		transposase DDE domain protein	N
Cluster_72088	15	2	32	21	32	9	LI+NSI	3.35	0.0671	0.0401		glycosyl transferases group 1 family protein	Y
Cluster_346882	25	6	22	17	53	26	LI+NSI	3.57	0.0590	0.0417		conserved hypothetical protein	N
Cluster_159210	27	7	20	16	57	30	LI+NSI	3.49	0.0616	0.0434		conserved hypothetical protein	N
Cluster_222333	27	7	20	16	57	30	LI+NSI	3.49	0.0616	0.0434		phage major capsid protein, HK97 family	N
Cluster_59004	27	7	20	16	57	30	LI+NSI	3.49	0.0616	0.0434		phage prohead protease, HK97 family	N
Cluster_107804	38	13	9	10	81	57	LI+NSI	3.47	0.0623	0.0455		methyltransferase domain protein	N
Cluster_56014	38	13	9	10	81	57	LI+NSI	3.47	0.0623	0.0455		hydantoinase/oxoprolinase family protein	N
Cluster_73988	38	13	9	10	81	57	LI+NSI	3.47	0.0623	0.0455		aminoglycoside/hydroxyurea antibiotic resistance kinase family protein	N
Cluster_106126	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		conserved hypothetical protein	N
Cluster_107544	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		tape measure domain protein	N
Cluster_115308	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		helix-turn-helix domain protein	Y
Cluster_117466	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		glycosyl transferase 2 family protein	N
Cluster_124512	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		conserved hypothetical protein	N
Cluster_281736	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		putative orf4 protein	N
Cluster_291295	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		rmlD substrate binding domain protein	N
Cluster_71992	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461		polysaccharide biosynthesis family protein	Y
Cluster_82368	8	0	39	23	17	0	LI+NSI	2.90	0.0887	0.0461	*fliC*	flagellin	N
Cluster_120672	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476		conserved hypothetical protein	N
Cluster_124501	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476		DNA N-6-adenine-methyltransferase family protein	N
Cluster_13383	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476		conserved hypothetical protein	N
Cluster_227359	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476		conserved hypothetical protein	N
Cluster_227360	12	1	35	22	26	4	LI+NSI	3.29	0.0697	0.0476		putative membrane protein	Y
Cluster_147010	5	7	42	16	11	30	AI	2.98	0.0842	0.0498		MULE transposase domain protein	N
Cluster_147193	5	7	42	16	11	30	AI	2.98	0.0842	0.0498		resolvase, N terminal domain protein	N
Cluster_297043	5	7	42	16	11	30	AI	2.98	0.0842	0.0498		hypothetical protein	N
Cluster_312558	5	7	42	16	11	30	AI	2.98	0.0842	0.0498		conserved hypothetical protein	Y
Cluster_312602	5	7	42	16	11	30	AI	2.98	0.0842	0.0498		HNH endonuclease family protein	N
Cluster_320753	5	7	42	16	11	30	AI	2.98	0.0842	0.0498		conserved hypothetical protein	N

Lethal (LI) vs. Non-Lethal (NSI+AI)

	No. of Genomes												
	Present (≥0.8)[a,b]		Absent (<0.8)[a,b]		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI	NSI+AI	LI	NSI+AI	% LI	% NSI+AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_175606	6	0	18	46	25	0	LI	9.59	0.0020	0.0010	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase	N
Cluster_222636	6	0	18	46	25	0	LI	9.59	0.0020	0.0010	*rfbA*	glucose-1-phosphate thymidylyltransferase	N
Cluster_14414	10	4	14	42	42	9	LI	8.75	0.0031	0.0031		conserved hypothetical protein	N
Cluster_175608	5	0	19	46	21	0	LI	7.42	0.0065	0.0035		degT/DnrJ/EryC1/StrS aminotransferase family protein	N
Cluster_247403	5	0	19	46	21	0	LI	7.42	0.0065	0.0035		glycosyl transferase 2 family protein	N
Cluster_83926	5	0	19	46	21	0	LI	7.42	0.0065	0.0035		acetyltransferase domain protein	N
Cluster_252971	20	22	4	24	83	48	LI	6.87	0.0088	0.0048		HNH endonuclease family protein	N
Cluster_14423	10	5	14	41	42	11	LI	7.15	0.0075	0.0051		conserved hypothetical protein	N
Cluster_58277	10	5	14	41	42	11	LI	7.15	0.0075	0.0051		helix-turn-helix, Psq domain protein	N
Cluster_236134	16	14	8	32	67	30	LI	7.04	0.0080	0.0052	*dhfrI*	dihydrofolate reductase type 1	N
Cluster_221922	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		glycosyl transferase family 2 domain protein	N
Cluster_247401	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		polysaccharide biosynthesis/export family protein	N
Cluster_83254	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		glycosyl transferases group 1 family protein	N
Cluster_83929	6	1	18	45	25	2	LI	6.77	0.0093	0.0055	*rfbB*	dTDP-glucose 4,6-dehydratase	N
Cluster_84063	6	1	18	45	25	2	LI	6.77	0.0093	0.0055		polysaccharide biosynthesis/export family protein	Y
Cluster_260984	13	9	11	37	54	20	LI	7.23	0.0072	0.0059	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_106122	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		conserved hypothetical protein	N
Cluster_115298	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		conserved hypothetical protein	N
Cluster_240829	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		hypothetical protein	N
Cluster_261817	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		putative predicted protein	N
Cluster_263945	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		conserved hypothetical protein	N
Cluster_74193	7	2	17	44	29	4	LI	6.60	0.0102	0.0060		phage major capsid E family protein	N
Cluster_13383	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		conserved hypothetical protein	N
Cluster_227359	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		conserved hypothetical protein	N
Cluster_227360	9	4	15	42	38	9	LI	6.85	0.0088	0.0073		putative membrane protein	Y
Cluster_222333	17	17	7	29	71	37	LI	5.95	0.0147	0.0113		phage major capsid protein, HK97 family	N
Cluster_59004	17	17	7	29	71	37	LI	5.95	0.0147	0.0113		phage prohead protease, HK97 family	N
Cluster_56014	22	29	2	17	92	63	LI	5.17	0.0230	0.0114		hydantoinase/oxoprolinase family protein	N
Cluster_108475	4	0	20	46	17	0	LI	5.33	0.0209	0.0116		PLD-like domain protein	N
Cluster_284853	4	0	20	46	17	0	LI	5.33	0.0209	0.0116		AAA domain protein	N
Cluster_54306	4	0	20	46	17	0	LI	5.33	0.0209	0.0116		putative yahA protein	N
Cluster_222331	16	16	8	30	67	35	LI	5.24	0.0221	0.0133		phage portal protein, HK97 family	N
Cluster_266615	16	16	8	30	67	35	LI	5.24	0.0221	0.0133		putative yubO	N
Cluster_106126	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		conserved hypothetical protein	N
Cluster_115308	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		helix-turn-helix domain protein	Y
Cluster_312709	6	2	18	44	25	4	LI	4.76	0.0291	0.0165		amino acid permease family protein	Y
Cluster_82368	6	2	18	44	25	4	LI	4.76	0.0291	0.0165	*fliC*	flagellin	N
Cluster_242060	8	4	16	42	33	9	LI	5.12	0.0237	0.0170		hypothetical protein	N
Cluster_112366	4	22	20	24	17	48	NSI+AI	5.29	0.0214	0.0179		mu-like prophage FluMu gp41 family protein	N
Cluster_341980	22	30	2	16	92	65	LI	4.47	0.0344	0.0207		conserved hypothetical protein	Y
Cluster_107351	11	8	13	38	46	17	LI	5.09	0.0240	0.0217		conserved hypothetical protein	N
Cluster_312794	11	8	13	38	46	17	LI	5.09	0.0240	0.0217		transposase DDE domain protein	N
Cluster_43705	18	20	6	26	75	43	LI	5.11	0.0238	0.0219		conserved hypothetical protein	Y
Cluster_18413	18	21	6	25	75	46	LI	4.38	0.0364	0.0238		resolvase, N terminal domain protein	N
Cluster_312629	18	21	6	25	75	46	LI	4.38	0.0364	0.0238		resolvase, N terminal domain protein	N
Cluster_43605	18	21	6	25	75	46	LI	4.38	0.0364	0.0238		HNH endonuclease family protein	N
Cluster_346831	16	17	8	29	67	37	LI	4.46	0.0347	0.0240		conserved hypothetical protein	N
Cluster_98923	17	19	7	27	71	41	LI	4.39	0.0362	0.0245		bacteriophage lysis family protein	Y
Cluster_157448	1	13	23	33	4	28	NSI+AI	4.32	0.0378	0.0250	*espS*	putative t3SS effector protein EspS	N
Cluster_276738	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		putative protein	Y
Cluster_340919	23	33	1	13	96	72	LI	4.32	0.0378	0.0250		winged helix-turn-helix DNA-binding family protein	N
Cluster_99052	1	13	23	33	4	28	NSI+AI	4.32	0.0378	0.0250		conserved hypothetical protein	Y
Cluster_128751	7	3	17	43	29	7	LI	4.88	0.0271	0.0256		HNH endonuclease family protein	N
Cluster_166190	7	3	17	43	29	7	LI	4.88	0.0271	0.0256		phage integrase family protein	N
Cluster_239025	7	3	17	43	29	7	LI	4.88	0.0271	0.0256		phage terminase, small subunit, P27 family	N
Cluster_29580	7	3	17	43	29	7	LI	4.88	0.0271	0.0256		phage Terminase family protein	N
Cluster_316843	13	37	11	9	54	80	NSI+AI	4.12	0.0423	0.0278		transglycosylase SLT domain protein	Y
Cluster_342430	21	28	3	18	88	61	LI	4.13	0.0420	0.0279		conserved hypothetical protein	N
Cluster_90939	21	28	3	18	88	61	LI	4.13	0.0420	0.0279		plasmid stability family protein	N

Cluster_183140	9	6	15	40	38	13	LI	4.24	0.0394	0.0300		putative escherichia coli chi7122 chi7122 genomic chromosome, chi7122	Y
Cluster_277010	20	26	4	20	83	57	LI	3.91	0.0479	0.0339		transposase DDE domain protein	Y
Cluster_237277	12	11	12	35	50	24	LI	3.75	0.0527	0.0348		conserved hypothetical protein	N
Cluster_262069	12	11	12	35	50	24	LI	3.75	0.0527	0.0348		istB-like ATP binding family protein	N
Cluster_289981	12	11	12	35	50	24	LI	3.75	0.0527	0.0348		integrase core domain protein	N
Cluster_14793	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_14794	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_14795	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_148543	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_148976	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_175609	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		polysaccharide biosynthesis family protein	Y
Cluster_199286	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		helix-turn-helix family protein	N
Cluster_199287	3	0	21	46	13	0	LI	3.35	0.0673	0.0370	*dam*	DNA adenine methylase family protein	N
Cluster_222670	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	Y
Cluster_222748	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative ccgAII	N
Cluster_24020	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		DEAD/DEAH box helicase family protein	N
Cluster_242402	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative membrane protein	Y
Cluster_24312	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		hypothetical protein	N
Cluster_247314	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative yaiB protein	N
Cluster_247315	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_247316	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		dnaJ domain protein	N
Cluster_74241	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_74242	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		cobQ/CobB/MinD/ParA nucleotide binding domain protein	N
Cluster_74244	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		hypothetical protein	N
Cluster_74245	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_74246	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		toxin-antitoxin system, antitoxin component, Xre family	N
Cluster_74250	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		putative yajA protein	N
Cluster_74251	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_74253	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_74254	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_83414	3	0	21	46	13	0	LI	3.35	0.0673	0.0370		conserved hypothetical protein	N
Cluster_79151	22	31	2	15	92	67	LI	3.82	0.0506	0.0380		bundlin family protein	Y
Cluster_262291	17	42	7	4	71	91	NSI+AI	3.56	0.0590	0.0381		type III secretion, HrpO family protein	Y
Cluster_284866	7	4	17	42	29	9	LI	3.56	0.0590	0.0381		bacterial regulatory, tetR family protein	N
Cluster_317552	7	4	17	42	29	9	LI	3.56	0.0590	0.0381		conserved hypothetical protein	N
Cluster_290675	5	22	19	24	21	48	NSI+AI	3.78	0.0519	0.0386		phage late control gene D family protein	N
Cluster_305186	5	22	19	24	21	48	NSI+AI	3.78	0.0519	0.0386		phage P2 GpU family protein	Y
Cluster_83828	19	24	5	22	79	52	LI	3.78	0.0519	0.0386		conserved hypothetical protein	N
Cluster_276547	13	13	11	33	54	28	LI	3.49	0.0617	0.0406		conserved hypothetical protein	Y
Cluster_56841	13	13	11	33	54	28	LI	3.49	0.0617	0.0406		putative membrane protein	Y
Cluster_207773	2	14	22	32	8	30	NSI+AI	3.21	0.0734	0.0410		collagen triple helix repeat family protein	N
Cluster_261646	22	32	2	14	92	70	LI	3.21	0.0734	0.0410		phage/conjugal plasmid C-4 type zinc finger , TraR family protein	N
Cluster_327168	18	22	6	24	75	48	LI	3.71	0.0541	0.0418		helix-turn-helix domain protein	N
Cluster_214595	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		phage gp6-like head-tail connector family protein	N
Cluster_243755	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		ASCH domain protein	N
Cluster_317316	15	16	9	30	63	35	LI	3.85	0.0497	0.0419		putative bacteriophage protein	N
Cluster_13363	5	2	19	44	21	4	LI	3.11	0.0780	0.0422	*cdtB*	cytolethal distending toxin subunit B	Y
Cluster_148189	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		cytolethal distending toxin A/C family protein	Y
Cluster_183987	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		insA N-terminal domain protein	N
Cluster_183988	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		putative transposase	N
Cluster_52075	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		type I restriction modification DNA specificity domain protein	N
Cluster_57949	5	2	19	44	21	4	LI	3.11	0.0780	0.0422		cytolethal distending toxin A/C family protein	Y
Cluster_108754	17	20	7	26	71	43	LI	3.70	0.0543	0.0434		phage terminase, small subunit, P27 family	N
Cluster_222329	17	20	7	26	71	43	LI	3.70	0.0543	0.0434		phage Terminase family protein	N
Cluster_317314	17	20	7	26	71	43	LI	3.70	0.0543	0.0434		phage head-tail joining family protein	N
Cluster_104965	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		transposase family protein	N
Cluster_125142	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		putative dNA transfer protein gp20	N
Cluster_147687	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		putative membrane protein	Y
Cluster_182116	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein	N
Cluster_182709	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		helix-turn-helix domain protein	N
Cluster_183020	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		hypothetical protein	N
Cluster_183381	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		subtilase family protein	N
Cluster_200360	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein	N
Cluster_242714	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein	Y
Cluster_285246	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein	N
Cluster_28951	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		hypothetical protein	N
Cluster_70300	4	1	20	45	17	2	LI	3.05	0.0808	0.0439		conserved hypothetical protein	N
Cluster_167436	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439		tail attachment protein	N
Cluster_217327	20	45	4	1	83	98	NSI+AI	3.05	0.0808	0.0439	*shp*	head decoration protein	N
Cluster_188256	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		conserved hypothetical protein	Y
Cluster_3723	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		conserved hypothetical protein	N
Cluster_8046	0	8	24	38	0	17	NSI+AI	3.15	0.0759	0.0442		hypothetical protein	N
Cluster_118562	8	5	16	41	33	11	LI	3.88	0.0488	0.0482		istB-like ATP binding family protein	N
Cluster_124501	8	5	16	41	33	11	LI	3.88	0.0488	0.0482		DNA N-6-adenine-methyltransferase family protein	N
Cluster_316740	21	29	3	17	88	63	LI	3.50	0.0613	0.0496		conserved hypothetical protein	Y
Cluster_99069	21	29	3	17	88	63	LI	3.50	0.0613	0.0496		haemolysin expression modulating family protein	N

[a] Genes were grouped into gene clusters using an "inclusive" clustering threshold of ≥80% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.8, while those that are "absent" have a TBLASTN LS-BSR <0.8.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

Table S9. Genes identified by LS-BSR that are significantly correlated with symptomatic, asymptomatic, lethal, or non-lethal clinical outcomes for the 61 typical EPEC genomes analyzed using an inclusive clustering threshold

Symptomatic (LI+NSI) vs. Asymptomatic (AI)

	No. of Genomes												
	Present (≥0.8)[a,b]		Absent (<0.8)[a,b]		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI+NSI	AI	LI+NSI	AI	% LI+NSI	% AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_168916	0	5	44	12	0	29	AI	10.46	0.0012	0.0010		nucleoside transporter family protein	Y
Cluster_167317	43	11	1	6	98	65	LI+NSI	10.11	0.0015	0.0013		intergenic-region protein	N
Cluster_297449	4	8	40	9	9	47	AI	8.91	0.0028	0.0021		conserved hypothetical protein	N
Cluster_11587	6	9	38	8	14	53	AI	8.21	0.0042	0.0028		transposase family protein	Y
Cluster_169965	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		ogr/Delta-like zinc finger family protein	N
Cluster_200452	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		putative prophage protein	N
Cluster_200454	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		putative phage immunity repressor protein	N
Cluster_271578	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		conserved hypothetical protein	N
Cluster_316955	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		helix-turn-helix domain protein	N
Cluster_63459	8	10	36	7	18	59	AI	7.88	0.0050	0.0038		putative derepression protein	N
Cluster_34159	23	2	21	15	52	12	LI+NSI	6.73	0.0095	0.0041		conserved hypothetical protein	N
Cluster_167315	42	11	2	6	95	65	LI+NSI	7.66	0.0057	0.0043		conserved hypothetical protein	N
Cluster_172125	0	4	44	13	0	24	AI	7.57	0.0059	0.0046		conserved hypothetical protein	Y
Cluster_203531	0	4	44	13	0	24	AI	7.57	0.0059	0.0046		nucleoside 2-deoxyribosyltransferase family protein	N
Cluster_206540	0	4	44	13	0	24	AI	7.57	0.0059	0.0046		hNH endonuclease family protein	N
Cluster_79228	1	5	43	12	2	29	AI	7.35	0.0067	0.0051		conserved hypothetical protein	N
Cluster_151023	10	11	34	6	23	65	AI	7.80	0.0052	0.0055		conserved hypothetical protein	Y
Cluster_320510	10	11	34	6	23	65	AI	7.80	0.0052	0.0055		conserved hypothetical protein	Y
Cluster_322330	10	11	34	6	23	65	AI	7.80	0.0052	0.0055		conserved hypothetical protein	N
Cluster_57681	10	11	34	6	23	65	AI	7.80	0.0052	0.0055		conserved hypothetical protein	N
Cluster_147543	14	0	30	17	32	0	LI+NSI	5.34	0.0209	0.0064		putative yfdA	Y
Cluster_322352	12	11	32	6	27	65	AI	5.81	0.0159	0.0094		conserved hypothetical protein	N
Cluster_307808	3	6	41	11	7	35	AI	5.80	0.0160	0.0106		conserved hypothetical protein	N
Cluster_309171	17	13	27	4	39	76	AI	5.59	0.0181	0.0106		outer membrane transport family protein	Y
Cluster_225859	8	9	36	8	18	53	AI	5.74	0.0166	0.0109		putative glyco3, capsid size determination protein Sid	N
Cluster_111596	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		ethanolamine utilization EutN/carboxysome family protein	N
Cluster_118076	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduT*	shell protein	N
Cluster_127742	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		dehydratase medium subunit	N
Cluster_127750	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		iron-containing alcohol dehydrogenase family protein	N
Cluster_166468	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		cob(I)alamin adenosyltransferase	N
Cluster_166473	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduA*	propanediol utilization protein PduA	N
Cluster_186173	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		BMC domain protein	Y
Cluster_196208	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*eutP*	ethanolamine utilization protein, EutP	N
Cluster_196209	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduU*	propanediol utilization protein PduU	N
Cluster_196229	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		putative cobalamin biosynthesis protein CbiG	Y
Cluster_225773	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduD*	propanediol dehydratase medium subunit	N
Cluster_230774	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduL*	phosphate propanoyltransferase	N
Cluster_259784	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduA*	propanediol utilization protein PduA	N
Cluster_259786	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduC*	propanediol dehydratase large subunit	N
Cluster_259788	20	14	24	3	45	82	AI	5.35	0.0207	0.0109	*pduE*	propanediol dehydratase small subunit	N
Cluster_265851	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		BMC domain protein	N
Cluster_265865	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		electron transport complex, RnfABCDGE type, C subunit	N
Cluster_310627	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		helix-turn-helix domain protein	N
Cluster_321136	20	14	24	3	45	82	AI	5.35	0.0207	0.0109		MIP channel family protein	Y
Cluster_221927	13	0	31	17	30	0	LI+NSI	4.74	0.0294	0.0125		phage Tail Collar domain protein	Y
Cluster_107351	17	1	27	16	39	6	LI+NSI	4.85	0.0277	0.0127		conserved hypothetical protein	N
Cluster_312794	17	1	27	16	39	6	LI+NSI	4.85	0.0277	0.0127		transposase DDE domain protein	N
Cluster_11511	6	8	38	9	14	47	AI	5.97	0.0145	0.0140		putative exported protein	Y
Cluster_178309	38	9	6	8	86	53	LI+NSI	5.97	0.0145	0.0140		conserved hypothetical protein	N
Cluster_222426	12	0	32	17	27	0	LI+NSI	4.17	0.0410	0.0142		putative membrane protein	Y
Cluster_334902	12	0	32	17	27	0	LI+NSI	4.17	0.0410	0.0142		conserved hypothetical protein	N
Cluster_97528	12	0	32	17	27	0	LI+NSI	4.17	0.0410	0.0142		ead/Ea22-like family protein	N
Cluster_267235	21	2	23	15	48	12	LI+NSI	5.31	0.0212	0.0167		hypothetical protein	N
Cluster_262069	20	2	24	15	45	12	LI+NSI	4.66	0.0308	0.0175		istB-like ATP binding family protein	N
Cluster_289981	20	2	24	15	45	12	LI+NSI	4.66	0.0308	0.0175		integrase core domain protein	N
Cluster_296985	13	11	31	6	30	65	AI	4.96	0.0259	0.0188		phage tail family protein	N
Cluster_171495	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		conserved hypothetical protein	N
Cluster_171552	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		binding-protein-dependent transport system inner membrane component family protein	Y
Cluster_171553	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		binding-protein-dependent transport system inner membrane component family protein	Y
Cluster_171556	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		ROK family protein	N
Cluster_206057	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		bacterial extracellular solute-binding family protein	Y
Cluster_206061	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		glycerophosphoryl diester phosphodiesterase family protein	N
Cluster_207740	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		caudovirales tail fibre assembly family protein	N
Cluster_207848	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		conserved hypothetical protein	N
Cluster_297563	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		conserved hypothetical protein	N
Cluster_319815	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		putative predicted protein	N
Cluster_319816	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		conserved hypothetical protein	Y
Cluster_321372	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		ABC transporter family protein	N
Cluster_94403	0	3	44	14	0	18	AI	4.83	0.0280	0.0189		conserved hypothetical protein	N
Cluster_147776	7	8	37	9	16	47	AI	4.85	0.0277	0.0192		conserved hypothetical protein	N
Cluster_225861	7	8	37	9	16	47	AI	4.85	0.0277	0.0192	*psu*	polarity suppression protein	N
Cluster_232396	7	8	37	9	16	47	AI	4.85	0.0277	0.0192		putative dNA-binding helix-turn-helix protein	N
Cluster_83718	15	12	29	5	34	71	AI	5.22	0.0223	0.0199	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_297448	21	14	23	3	48	82	AI	4.68	0.0305	0.0204		conserved hypothetical protein	N
Cluster_82579	21	14	23	3	48	82	AI	4.68	0.0305	0.0204	*espJ*	putative non-LEE-encoded effector EspJ	N
Cluster_222333	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		phage major capsid protein, HK97 family	N
Cluster_59004	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		phage prohead protease, HK97 family	N
Cluster_80973	26	4	18	13	59	24	LI+NSI	4.86	0.0274	0.0212		phage integrase family protein	N
Cluster_53910	4	6	40	11	9	35	AI	4.38	0.0364	0.0218	*intA*	prophage CP4-57 integrase	N
Cluster_67359	4	6	40	11	9	35	AI	4.38	0.0364	0.0218	*intA*	prophage CP4-57 integrase	N
Cluster_14627	14	11	30	6	32	65	AI	4.21	0.0402	0.0239		cysteine protease , YopT-type domain protein	N
Cluster_340924	28	16	16	1	64	94	AI	4.25	0.0392	0.0241		ribbon-helix-helix , copG family protein	N
Cluster_337159	25	4	19	13	57	24	LI+NSI	4.20	0.0405	0.0244		conserved hypothetical protein	N
Cluster_312602	5	7	39	10	11	41	AI	5.14	0.0234	0.0263		HNH endonuclease family protein	N
Cluster_209146	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267	*espL*	shET2 enterotoxin, N-terminal region family protein	N
Cluster_263620	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		transposase DDE domain protein	N
Cluster_285156	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267	*espZ*	sepZ family protein	Y
Cluster_72088	15	1	29	16	34	6	LI+NSI	3.69	0.0547	0.0267		glycosyl transferases group 1 family protein	Y
Cluster_157448	3	5	41	12	7	29	AI	3.69	0.0548	0.0321	*espS*	putative t3SS effector protein EspS	N
Cluster_303768	3	5	41	12	7	29	AI	3.69	0.0548	0.0321		putative regulatory protein N	N
Cluster_78756	3	5	41	12	7	29	AI	3.69	0.0548	0.0321		prophage CP4-57 regulatory family protein	N
Cluster_132984	6	7	38	10	14	41	AI	4.03	0.0448	0.0334		traP family protein	Y
Cluster_14370	6	7	38	10	14	41	AI	4.03	0.0448	0.0334		insertion element 4 transposase N-terminal family protein	Y
Cluster_177188	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335		phage integrase family protein	N
Cluster_237277	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335		conserved hypothetical protein	N
Cluster_260984	19	2	25	15	43	12	LI+NSI	4.06	0.0439	0.0335	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_183593	18	2	26	15	41	12	LI+NSI	3.50	0.0615	0.0359		conserved hypothetical protein	N
Cluster_260684	26	15	18	2	59	88	AI	3.50	0.0615	0.0359		conserved hypothetical protein	Y
Cluster_307754	32	7	12	10	73	41	LI+NSI	4.01	0.0451	0.0360		regulatory protein cro	N

Cluster_327168	32	7	12	10	73	41	LI+NSI	4.01	0.0451	0.0360		helix-turn-helix domain protein	N
Cluster_54451	12	10	32	7	27	59	AI	4.01	0.0451	0.0360		conserved hypothetical protein	N
Cluster_54644	22	14	22	3	50	82	AI	4.05	0.0441	0.0402		conserved hypothetical protein	N
Cluster_118816	21	3	23	14	48	18	LI+NSI	3.47	0.0623	0.0418		conserved hypothetical protein	Y
Cluster_132520	13	10	31	7	30	59	AI	3.32	0.0686	0.0436		phage capsid scaffolding (GPO) serine peptidase family protein	N
Cluster_252971	31	7	13	10	70	41	LI+NSI	3.32	0.0686	0.0436		HNH endonuclease family protein	N
Cluster_303767	13	10	31	7	30	59	AI	3.32	0.0686	0.0436		repressor protein CI	N
Cluster_116275	15	11	29	6	34	65	AI	3.53	0.0602	0.0437		espF repeat family protein	Y
Cluster_39077	29	6	15	11	66	35	LI+NSI	3.53	0.0602	0.0437		conserved hypothetical protein	N
Cluster_43605	29	6	15	11	66	35	LI+NSI	3.53	0.0602	0.0437		HNH endonuclease family protein	N
Cluster_196213	20	13	24	4	45	76	AI	3.58	0.0584	0.0444		aldehyde dehydrogenase family protein	N
Cluster_214595	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		phage gp6-like head-tail connector family protein	N
Cluster_222331	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		phage portal protein, HK97 family	N
Cluster_236053	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		conserved hypothetical protein	N
Cluster_317316	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		putative bacteriophage protein	N
Cluster_72658	20	13	24	4	45	76	AI	3.58	0.0584	0.0444		putative propanediol utilization protein	N
Cluster_74298	24	4	20	13	55	24	LI+NSI	3.58	0.0584	0.0444		traP family protein	Y
Cluster_60983	2	4	42	13	5	24	AI	3.07	0.0796	0.0457		conserved hypothetical protein	N
Cluster_128383	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein	N
Cluster_128390	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		viral (Super1) RNA helicase family protein	N
Cluster_164940	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		AAA domain protein	N
Cluster_276697	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein	N
Cluster_291791	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		plasmid conjugative transfer entry exclusion TraS family protein	Y
Cluster_291795	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		putative acetyltransferase	Y
Cluster_59092	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein	N
Cluster_73497	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein	N
Cluster_98944	14	1	30	16	32	6	LI+NSI	3.16	0.0755	0.0469		conserved hypothetical protein	N
Cluster_56999	7	7	37	10	16	41	AI	3.11	0.0776	0.0471		putative P4-specific DNA primase	N
Cluster_107804	36	9	8	8	82	53	LI+NSI	3.90	0.0484	0.0479		methyltransferase domain protein	N
Cluster_227577	8	8	36	9	18	47	AI	3.90	0.0484	0.0479	*ssb*	single-stranded DNA-binding protein	N
Cluster_159210	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein	N
Cluster_252361	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		putative antirestriction protein	N
Cluster_302692	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		hypothetical protein	N
Cluster_346831	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein	N
Cluster_84849	26	5	18	12	59	29	LI+NSI	3.22	0.0729	0.0486		conserved hypothetical protein	N
Cluster_127453	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		conserved hypothetical protein	N
Cluster_139135	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		putative pyridoxamine 5-phosphate-dependent dehydrase	N
Cluster_221116	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		hypothetical protein	N
Cluster_232949	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		P22AR C-terminal domain protein	N
Cluster_262291	34	17	10	0	77	100	AI	3.11	0.0777	0.0493		type III secretion , HrpO family protein	Y
Cluster_291294	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		putative pyridoxamine 5-phosphate-dependent dehydrase	Y
Cluster_95058	10	0	34	17	23	0	LI+NSI	3.11	0.0777	0.0493		putative antirepressor protein	N

Lethal (LI) vs. Non-Lethal (NSI+AI)													
	No. of Genomes												
	Present (≥0.8)^{a,b}		Absent (<0.8)^{a,b}		Percentage of Genomes Containing Cluster			Chi-squared test		Fisher's exact test			
Cluster ID	LI	NSI+AI	LI	NSI+AI	% LI	% NSI+AI	Greater %	statistic	p-value	p-value	Gene ID	Predicted Protein Function	Membrane or Secretion Domains
Cluster_104965	4	0	20	37	17	0	LI	4.16	0.0414	0.0204		transposase family protein	N
Cluster_106122	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		conserved hypothetical protein	N
Cluster_106126	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein	N
Cluster_107351	11	7	13	30	46	19	LI	3.86	0.0495	0.0428		conserved hypothetical protein	N
Cluster_108475	4	0	20	37	17	0	LI	4.16	0.0414	0.0204		PLD-like domain protein	N
Cluster_112366	4	18	20	19	17	49	NSI+AI	5.15	0.0233	0.0144		mu-like prophage FluMu gp41 family protein	N
Cluster_11448	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		conserved hypothetical protein	Y
Cluster_115298	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		conserved hypothetical protein	N
Cluster_115308	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		helix-turn-helix domain protein	Y
Cluster_128751	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		HNH endonuclease family protein	N
Cluster_132534	7	21	17	16	29	57	NSI+AI	3.42	0.0644	0.0403		baseplate J-like family protein	N
Cluster_13383	9	4	15	33	38	11	LI	4.69	0.0303	0.0231		conserved hypothetical protein	N
Cluster_14414	10	4	14	33	42	11	LI	6.19	0.0128	0.0108		conserved hypothetical protein	N
Cluster_14423	10	5	14	32	42	14	LI	4.80	0.0285	0.0170		conserved hypothetical protein	N
Cluster_14627	5	20	19	17	21	54	NSI+AI	5.34	0.0208	0.0159		cysteine protease , YopT-type domain protein	N
Cluster_147640	11	7	13	30	46	19	LI	3.86	0.0495	0.0428		conserved hypothetical protein	Y
Cluster_152085	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein	N
Cluster_166190	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage integrase family protein	N
Cluster_175131	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		bacterial transferase hexapeptide family protein	N
Cluster_175134	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		archaeal transcriptional regulator TrmB family protein	N
Cluster_175606	6	0	18	37	25	0	LI	7.63	0.0057	0.0024	*fdtA*	dTDP-6-deoxy-3,4-keto-hexulose isomerase	N
Cluster_175608	5	0	19	37	21	0	LI	5.86	0.0155	0.0071		degT/DnrJ/EryC1/StrS aminotransferase family protein	N
Cluster_176094	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		NIF3 family protein	N
Cluster_176098	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		alpha amylase, catalytic domain protein	N
Cluster_178309	22	25	2	12	92	68	LI	3.52	0.0608	0.0335		conserved hypothetical protein	N
Cluster_19453	7	21	17	16	29	57	NSI+AI	3.42	0.0644	0.0403		ATPase subunit of terminase family protein	N
Cluster_221116	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		hypothetical protein	N
Cluster_221922	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		glycosyl transferase family 2 domain protein	N
Cluster_222331	16	12	8	25	67	32	LI	5.56	0.0184	0.0171		phage portal protein, HK97 family	N
Cluster_222333	17	13	7	24	71	35	LI	6.06	0.0138	0.0090		phage major capsid protein, HK97 family	N
Cluster_222426	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		putative membrane protein	Y
Cluster_222636	6	0	18	37	25	0	LI	7.63	0.0057	0.0024	*rfbA*	glucose-1-phosphate thymidylyltransferase	N
Cluster_226605	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		conserved hypothetical protein	N
Cluster_227359	9	4	15	33	38	11	LI	4.69	0.0303	0.0231		conserved hypothetical protein	N
Cluster_227360	9	4	15	33	38	11	LI	4.69	0.0303	0.0231		putative membrane protein	Y
Cluster_232670	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		cysteine protease, YopT-type domain protein	N
Cluster_232949	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		P22AR C-terminal domain protein	N
Cluster_236134	16	13	8	24	67	35	LI	4.61	0.0318	0.0201	*dhfrI*	dihydrofolate reductase type 1	N
Cluster_237298	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein	N
Cluster_237299	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		conserved hypothetical protein	N
Cluster_239025	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage terminase, small subunit, P27 family	N
Cluster_240829	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		hypothetical protein	N
Cluster_242060	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		hypothetical protein	N
Cluster_247401	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		polysaccharide biosynthesis/export family protein	N
Cluster_247403	5	0	19	37	21	0	LI	5.86	0.0155	0.0071		glycosyl transferase 2 family protein	N
Cluster_251045	8	23	16	14	33	62	NSI+AI	3.76	0.0526	0.0374		phage tail tape measure protein, TP901 family, core region	Y
Cluster_252971	20	18	4	19	83	49	LI	6.05	0.0139	0.0075		HNH endonuclease family protein	N
Cluster_25983	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		bacteriophage replication gene A family protein	N
Cluster_260984	13	8	11	29	54	22	LI	5.46	0.0194	0.0132	*traN*	type-F conjugative transfer system mating-pair stabilization protein TraN	Y
Cluster_261817	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		putative predicted protein	N
Cluster_262291	17	34	7	3	71	92	NSI+AI	3.30	0.0693	0.0397		type III secretion , HrpO family protein	Y
Cluster_263945	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		conserved hypothetical protein	N
Cluster_264852	8	23	16	14	33	62	NSI+AI	3.76	0.0526	0.0374		phage tail tape measure protein, TP901 family, core region	Y
Cluster_282545	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		conserved hypothetical protein	N
Cluster_282548	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		putative predicted protein	Y
Cluster_284853	4	0	20	37	17	0	LI	4.16	0.0414	0.0204		AAA domain protein	N
Cluster_28951	4	0	20	37	17	0	LI	4.16	0.0414	0.0204		hypothetical protein	N

Cluster_290675	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		phage late control gene D family protein	N
Cluster_29580	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage Terminase family protein	N
Cluster_297348	0	7	24	30	0	19	NSI+AI	3.44	0.0638	0.0358		hypothetical protein	N
Cluster_305186	5	18	19	19	21	49	NSI+AI	3.68	0.0549	0.0340		phage P2 GpU family protein	Y
Cluster_307935	0	7	24	30	0	19	NSI+AI	3.44	0.0638	0.0358		conserved hypothetical protein	N
Cluster_312709	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		amino acid permease family protein	Y
Cluster_312794	11	7	13	30	46	19	LI	3.86	0.0495	0.0428		transposase DDE domain protein	N
Cluster_313403	8	4	16	33	33	11	LI	3.36	0.0669	0.0474	*repA*	regulatory protein RepA	Y
Cluster_313404	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		prophage CP4-57 regulatory family protein	N
Cluster_313405	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		luxR family transcriptional regulator	N
Cluster_315365	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		ABC transporter family protein	Y
Cluster_316214	6	2	18	35	25	5	LI	3.34	0.0678	0.0481		tatD related DNase family protein	N
Cluster_317552	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		conserved hypothetical protein	N
Cluster_42814	4	16	20	21	17	43	NSI+AI	3.54	0.0600	0.0496		glycosyltransferase like 2 family protein	N
Cluster_43605	18	17	6	20	75	46	LI	3.91	0.0481	0.0347		HNH endonuclease family protein	N
Cluster_51447	7	21	17	16	29	57	NSI+AI	3.42	0.0644	0.0403		ATPase subunit of terminase family protein	N
Cluster_54306	4	0	20	37	17	0	LI	4.16	0.0414	0.0204		putative yahA protein	N
Cluster_58277	10	4	14	33	42	11	LI	6.19	0.0128	0.0108		helix-turn-helix, Psq domain protein	N
Cluster_59004	17	13	7	24	71	35	LI	6.06	0.0138	0.0090		phage prohead protease, HK97 family	N
Cluster_74193	7	2	17	35	29	5	LI	4.78	0.0288	0.0221		phage major capsid E family protein	N
Cluster_82368	6	2	18	35	25	5	LI	3.34	0.0678	0.0481	*fliC*	flagellin	N
Cluster_83254	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		glycosyl transferases group 1 family protein	N
Cluster_83926	5	0	19	37	21	0	LI	5.86	0.0155	0.0071		acetyltransferase domain protein	N
Cluster_83929	6	1	18	36	25	3	LI	5.10	0.0239	0.0122	*rfbB*	dTDP-glucose 4,6-dehydratase	N
Cluster_84063	6	1	18	36	25	3	LI	5.10	0.0239	0.0122		polysaccharide biosynthesis/export family protein	Y
Cluster_95058	7	3	17	34	29	8	LI	3.30	0.0693	0.0397		putative antirepressor protein	N
Cluster_97528	8	4	16	33	33	11	LI	3.36	0.0669	0.0474		ead/Ea22-like family protein	N
Cluster_98923	17	15	7	22	71	41	LI	4.21	0.0402	0.0350		bacteriophage lysis family protein	Y
Cluster_99077	4	16	20	21	17	43	NSI+AI	3.54	0.0600	0.0496		cysteine protease , YopT-type domain protein	N

[a] Genes were grouped into gene clusters using an "inclusive" clustering threshold of ≥80% nucleotide identity.

[b] Gene clusters identified as "present" have a TBLASTN LS-BSR ≥0.8, while those that are "absent" have a TBLASTN LS-BSR <0.8.

[c] Clinical outcomes are classified as lethal (LI), non-lethal symptomatic (NSI), and asymptomatic (AI).

Table S10. The number of LS-BSR gene clusters with a predicted protein that contained a protein domain of surface-associated or secreted proteins

	No. of Gene Clusters[a]		
Clinical Outcomes[b]	All EPEC Genomes[c]	Significant in Both Comparisons[d]	TEPEC Genomes[e]
LI vs. AI	40	13	31
LI vs. NSI	11	2	2
NSI vs. AI	14	9	21
Symptomatic vs. Asymptomatic	50	19	31
Lethal vs. Non-Lethal	28	8	16

[a] The number of gene clusters with one or more predicted surface-associated or secreted protein domains that were significantly ($p<0.05$) associated with genomes of one clinical outcome over another that were not present (LS-BSR value <0.8) in any of the three commensal *E. coli* genomes.

[b] The clinical outcomes are lethal (LI), non-lethal symptomatic (NSI), asymptomatic (AI), symptomatic (LI+NSI), or non-lethal (NSI+AI).

[c] The number of gene clusters that were significantly associated with symptomatic or asymptomatic genomes when all 70 genomes were compared, which included both typical EPEC and atypical EPEC.

[d] The number of gene clusters that were significantly associated with symptomatic or asymptomatic genomes in the comparison with all 70 genomes, and also the comparison with only the 61 typical EPEC genomes.

[e] The number of gene clusters that were significantly associated with symptomatic or asymptomatic genomes when comparing the 61 typical EPEC genomes only.