# Tables

Table 1. The percentage of genomes that contain metadata for each of the metadata fields in the initial public data set of 1641 *E. coli* in the SuperPhy database.

|  |  |
| --- | --- |
| Metadata field | Percentage |
| Location | 85 |
| Host | 79 |
| Date of Isolation | 63 |
| Source | 52 |
| Serotype | 44 |
| Stx2 subtype | 23 |
| Stx1 subtype | 18 |
| Disease syndrome | 6 |

Table 2. The number of conserved core genomic regions present in 19 selected bacterial genomes, from the total 3598 conserved core genomic regions found in at least 70% of the 2324 *E. coli* genomes examined.

|  |  |
| --- | --- |
| Genome | No. ‘conserved core’ genes |
| E. coli O103:H2,12009 | 3563 |
| E. coli O157:H7, EDL933 | 3557 |
| E. coli K-12, MG1655 | 3550 |
| E. coli, UMN026 | 3483 |
| E. coli O7:K1, CE10 | 3448 |
| E. coli O83:H1, NRG 857C | 3289 |
| Shigella sonnei, 53G | 3259 |
| Shigella flexneri 2002017 | 3148 |
| Shigella boydii, CDC 3083-94 | 2965 |
| Shigella dysenteriae, 1617 | 2683 |
| Escherichia fergusonii ATCC 35469 | 1619 |
| Salmonella enterica subsp. Enterica serovar Typhimurium str. 14028S | 95 |
| Citrobacter rodentium, ICC168 | 77 |
| Klebsiella oxytoca, E718 | 50 |
| Klebsiella pneumoniae subsp. Pneumoniae, 1084 | 50 |
| Klebsiella variicola, At-22 | 46 |
| Escherichia blattae, DSM 4481 | 27 |
| Staphylococcus aureus, 04-02981 | 0 |
| Listeria monocytogenes, 07PF0776 | 0 |

Table 3. The ten *E. coli* species-specific genomic regions identified in this study based on a total sequence identity of 90%, their location in the K12 reference genome MG1655, the number out of 2324 *E. coli* genomes each region was found in, and their putative function based on the top scoring BLASTx hit.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Region ID | Start bp | End bp | No. Genomes | Putative function |
| 3160548 | 347258 | 346259 | 2238 | Propionate catabolism operon regulatory protein PrpR |
| 3160296 | 537566 | 536567 | 2256 | 2-hydroxy-3-oxopropionate reductase |
| 3160113 | 538566 | 537567 | 2248 | Allantoin permease |
| 3159571 | 541565 | 540567 | 2275 | Purine permease ybbY |
| 3159389 | 542566 | 541567 | 2268 | Glycerate kinase |
| 3158844 | 545665 | 544666 | 2261 | Allantoate amidohydrolase |
| 3158667 | 546665 | 545666 | 2272 | Ureidoglycolate dehydrogenase |
| 3159808 | 1588200 | 1587201 | 2171 | FimH protein |
| 3160196 | 4411062 | 4410063 | 2261 | Hypothetical protein |
| 3158082 | 4456632 | 4457631 | 2074 | Mur ligase family, glutamate ligase domain protein |