bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Full Title:

Combining *Shigella* Tn-seq data with Gold-standard *E. coli* Gene Deletion Data Suggests Rare Transitions between Essential and Non-essential Gene Functionality

Short Title:

Transitions between Essential and Non-essential Functionality

Nikki E. Freed[1, 2], Dirk Bumann[2], and Olin K. Silander[1, 3, *]

1 Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand

2 Infection Biology, Biozentrum, University of Basel, Basel, Switzerland

3 Computational and Systems Biology, Biozentrum, University of Basel, Basel, Switzerland

* Corresponding author: olinsilander@gmail.com (OKS)

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Gene essentiality - whether or not a gene is necessary for cell growth - is a fundamental component of gene function. It is not well established how quickly gene essentiality can change, as few studies have compared empirical measures of essentiality between closely related organisms. Here we present the results of a Tn-seq experiment designed to detect essential protein coding genes in the bacterial pathogen *Shigella flexneri* 2a 2457T on a genome-wide scale. Superficial analysis of this data suggested that 451 protein-coding genes in this *Shigella* strain are critical for robust cellular growth on rich media. Comparison of this set of genes with a gold-standard data set of essential genes in the closely related *Escherichia coli* K12 BW25113 suggested that an excessive number of genes appeared essential in *Shigella* but non-essential in *E. coli.* Importantly, and in converse to this comparison, we found no genes that were essential in *E. coli* and non-essential in *Shigella*, suggesting that many genes were artefactually inferred as essential in *Shigella.* Controlling for such artefacts resulted in a much smaller set of discrepant genes. Among these, we identified three sets of functionally related genes; two of which have previously been implicated as critical for *Shigella* growth, but which are dispensable for *E. coli* growth. The data presented here highlight the small number of protein coding genes for which we have strong evidence that their essentiality status differs between the closely related bacterial taxa *E. coli* and *Shigella.* A set of genes involved in acetate utilization provides a canonical example. These results leave open the possibility of developing strain-specific antibiotic treatments targeting such differentially essential genes, but suggest that such opportunities may be rare in closely related bacteria.

Author Summary

Essential genes are those that encode proteins required for growth and survival in a particular environment. We performed experiments using transposons, genetic elements that disrupt gene function, to determine the set of essential genes in the pathogenic bacteria *Shigella flexneri*. We then compared our results to the well-characterized set of essential genes in the closely related, yet non-pathogenic, bacteria *Escherichia coli*. We found only a small number of genes that are important for growth in *Shigella flexneri*, yet not in *Escherichia* coli. We believe these findings are interesting for several reasons; they help us better understand how quickly the functions of proteins change over time; they suggest possible targets for developing strain-specific antibiotic treatments; and they expand our basic understanding of this pathogen's metabolic processes.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

One general functional characteristic of a gene is essentiality - whether that gene is required for cellular viability and growth. In haploid (e.g. bacterial) genomes, this characteristic can be assessed by attempting to delete a specific gene from a genome. When such a deletion is not possible, this gene is frequently termed "essential" [1], implying that the gene is necessary for cell growth and viability. Gene disruption, although less precise, is more commonly used to infer essentiality using a similar criterion. For example, genes that cannot be disrupted by transposon insertion have been inferred as being essential (e.g. [2]).

One important question is how quickly essential functions change over evolutionary time. If orthologous protein coding genes in two bacterial strains differ in their essentiality classification, this suggests that either the biochemical nature of the protein has changed, or that the cellular context in which the protein acts has changed [3]. It has been experimentally established that such transitions can occur [4-6]. Here we examine how frequently proteins go from being essential to non-essential and vice versa in nature.

A recent study quantified changes in the essentiality classifications of protein coding genes between three alpha-proteobacteria: *Caulobacter crescentus*, *Brevundimonas subvibrioides*, and *Agrobacterium tumefaciens* [3]. The analysis showed that although orthologous cell components are well conserved, the essentiality of such components (e.g. those involved in the cell cycle) had changed considerably, with only 106 orthologous genes being essential in all three organisms, despite their relatively close evolutionary relationship (89%-93% identity in 16S RNA genes).

In this study we combine dense transposon mutagenesis with high-throughput sequencing (Tn-seq [7]) to quantify gene essentiality in *Shigella flexneri* 2a 2457T (hereafter referred to as *Shigella*). We compare the essentiality classifications of protein coding genes in *Shigella* with a gold-standard assessment of essentiality in the closely related strain *Escherichia coli* K12 BW25113 (hereafter referred to as *E. coli*) [1]. These two stains are 99.5% identical in their 16S RNA genes and share approximately 70% of their genomic content.

This proximity in evolutionary distance, and the use of a gold-standard data set, brings two unique advantages that have not been available in other studies that have used Tn-Seq or similar methods to quantify gene essentiality [3, 7-19]. First, by relying on the null hypothesis that protein coding genes do maintain their essentiality characteristics, we can objectively assess which quantitative features in the *Shigella* Tn-seq data best predict essentiality or non-essentiality of their orthologous counterparts in *E. coli*; such a comparison to a gold-standard has not yet been used to assess the quality and sensitivity of Tn-Seq data [20], although several studies have validated a small number of Tn-seq-inferred growth defects using clean

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

deletion methods (e.g. [10]). Second, the use of very closely related taxa allows us to quantify on a much shorter time scale the fraction of the essential gene complement that has changed, providing a fine scale window into the rate with which orthologues change in their essential functions.

The data presented here suggest that the essential gene complement of *Shigella* and *E. coli* overlap considerably. Indeed, we find no strong evidence that there are any protein-coding genes that are essential in *E. coli* but not *Shigella*. Conversely, we do find a small number of genes that play critical roles for *Shigella* growth, but which have dispensable roles in *E. coli*, or which are absent entirely from *E. coli.* This implies that the functional correspondence, in terms of essentiality, has changed for only a small number of protein-coding genes.

However, our analysis also suggests that some protein-coding genes that we observe as undisrupted by transposon insertions are in fact not essential for cell growth. Instead, they are either essential for transposon insertion to occur successfully, or their disruption (but not clean deletion) is detrimental to cell growth. This result emphasizes that in high throughput transposon mutagenesis studies, false positive inferences of essentiality may be common, and that simply increasing the resolution or precision of a dataset cannot necessarily solve this problem.

Taken together, our data suggests that the essential gene complement is relatively static over short time scales. However, when protein-coding genes do change from being non-essential to being essential, this appears more likely to occur in pathogenic organisms, perhaps because host environments absolve the organism from manufacturing its own nutrients, or because such organisms have smaller population sizes and are prone to the accumulation of deleterious mutations. It would be interesting to see if this pattern is observed when comparing other pathogens to their free-living sister taxa. If antibiotics can be directed against the function or expression of such differentially essential genes, this may allow targeting such treatments toward specific bacterial strains.

Results and Discussion

A Transposon Mutagenesis Library Provides Fine-scale Resolution of Gene Essentiality

We generated a transposon insertion library by transforming a *Shigella icsA* mutant [21] with a plasmid containing a mini-Tn10 transposase with decreased hotspot activity ([22]; **Fig. 1A**, inset) inducible by isopropyl-β-D-thiogalactopyranoside (IPTG) [23, 24]. After overnight growth on Tryptic Soy Broth (TSB) agar plates containing IPTG, we harvested approximately10^6 colonies carrying transposon insertions. We pooled and then split this

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

library of clones into six replicates. Three replicates were subject to additional growth step inside the cytoplasm of HeLa cells for four hours. The resulting cells were then harvested, the replicates were bar coded and libraries were prepared. We sequenced all six of these pools on a single Illumina HiSeq lane. For all the analyses presented in this study, we have pooled the data from all replicates and from both of these treatments, as we are focusing on *Shigella* genes that are essential across any permissive growth conditions.

From this pool, we mapped insertions at 131,670 unique positions on the *Shigella* chromosome (with many insertions occurring on both the forward and reverse strands but at the same position), and 12,552 unique positions on the large *Shigella* virulence plasmid (see **Methods**). The median distance between inserts on the chromosome was 17 base pairs (bp); on the plasmid this distance was 9 bp. 95% of all inter-insert distances on the chromosome were less than 107 bp; the corresponding figure for the plasmid was 59 bp (**Figs. 1A and B**).

Although the distribution of transposon inserts was relatively even across both the chromosome (**S1 Fig.**), at smaller scales we found many regions in which few or no insertions occurred. Quantitative analyses showed that regions containing no transposon insertions for 100bp or more were considerably enriched (see **Methods**; **Figs. 1C and D**). It is likely that many of these regions are critical for cellular growth in *Shigella*. Indeed, we found that for many of the protein-coding genes in these regions, the orthologous *E. coli* genes are known to be essential (**Figs. 2** and **3**).

In contrast to the *Shigella* chromosome, we found that few open reading frames on the virulence plasmid were devoid of insertions. Only six out of 263 plasmid ORFs had no inserts. Two of these were replication proteins (CP0258 and CP0259), and two (CP0217 and CP0218) were located within the plasmid stabilisation region. The remaining two, *mxiH* and *acp*, are both less than 250 bp in length (the cut-off used here to classify ORFs as essential; see below), and thus have a lower likelihood of being hit due simply to their smaller target size. The third replication protein of the plasmid, CP0260, contained a single insert in its 858 bp length. The absence of inserts in the plasmid replication or stabilisation regions is explained by the fact that if such insertions did occur, the plasmid would be lost; such insertions would thus never be sequenced. Thus, this data is consistent with the fact that the *Shigella* plasmid contains no essential genes [25], and suggested that our transposon library provided a fine-scale assessment of which *Shigella* chromosomal ORFs provide critical cellular functions.

Average Distance Between Inserts Clearly Delineates Essential and Non-essential ORFs

We next quantified which transposon insertion patterns in the chromosome were good predictors of the essentiality of open reading frames. To do so, we first identified 3,027

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

orthologous open reading frames present in both *E. coli* and *Shigella* for which we also had data on essentiality from both the Keio [1, 26] and the Profiling of the Escherichia coli Chromosome (PEC) [27] studies (**S1 Table**). We considered this combined gene set as a gold-standard of essentiality, for two reasons: it is not subject to artefacts that might exist in Tn-seq dataset, such as insertion biases or biases arising during sequencing library preparation (e.g. [28]); and combining both the Keio and PEC datasets should result in few false positive or false negative essentiality characterizations. This set consisted of 277 orthologues considered essential by both studies, 2,717 genes considered non-essential by both studies, and 33 genes for which the two studies disagree.

We next quantified several characteristics for each protein-coding gene in our Tn-seq data set, including the total number of inserts per ORF, the mean distance between inserts, the length of the 5' fraction of the ORF upstream of the first insertion, the largest uninterrupted region in the ORF, and others (**S2 Fig.**). We took as a null hypothesis that generally, genes have maintained their essentiality characteristics since the divergence of *E. coli* and *Shigella*. We then tested which of these characteristics best predicted the essentiality status of their orthologous counterparts in the gold-standard dataset of open reading frame essentiality in *E. coli* (the Keio and PEC datasets).

We found that the best predictor of essentiality status in *E. coli* was the mean distance between transposon insertions in their *Shigella* orthologues (Materials and Methods; **S2 Fig.**). For the *Shigella* orthologues of the 277 *E. coli* essential genes, only four had a mean distance between inserts of less than 150bp. 17 (6%) had a mean inter-insert distance less than 250bp. In contrast, only 6% of the orthologues of non-essential *E. coli* genes had a mean distance between inserts of greater than 250bp (**Fig. 3A**). We selected this mean inter-insert distance of 250bp as a cut-off for classifying *Shigella* ORFs as essential, as it provided a balance between protein coding genes classified as essential in *E. coli* but non-essential in *Shigella* (a 6% false negative rate) versus non-essential in *E. coli* and essential in *Shigella* (a 6% false positive rate). By extension, genes that are less than 250bp in length and in which we do not observe insertions were inferred as essential (26 ORFs in total, of which 12 were ribosomal proteins and five were leader peptides). We note, importantly, that almost all of the predictors we tested performed extremely well (**S2 Fig.**).

We next investigated in greater detail the disagreements in essentiality classification between *E. coli* and *Shigella* (**Fig. 3B**). Of the17 *E. coli*-essential genes that this metric identified as non-essential in *Shigella*, all are likely to be false negatives (i.e. in fact essential in *Shigella,* but not classified as such by our criterion). All 17 have a mean distance between inserts of greater than 100 bp (**Fig. 3A**), and nine are uninterrupted for more than 90% of their reading frame. This suggests, surprisingly, that there are no genes that are essential in *E. coli* but

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

whose *Shigella* orthologues are non-essential. This similarity in essentiality is not due to the fact that we use a characteristic that most closely predicts essentiality in the gold standard dataset – this result is robustly corroborated by any meaningful metric that we used (e.g. using other mean distances between inserts as cut-offs for essentiality, using the total number of inserts, the longest uninterrupted gene fraction, or others (**S2 Fig.**)). Overall, this data gives us a very strong prior that genes have maintained their essentiality status (or near-lethal effects on growth) since the divergence of *E. coli* and *Shigella*.

Many Non-essential *E. coli* orthologues of Essential *Shigella* Genes Exhibit Impaired Growth

As a result of this strong prior, we thus expect that many of the discrepancies in essentiality between *E. coli* and *Shigella* are false positives due simply to the *Shigella* mutants being non-essential, but having significantly impaired growth. Indeed, of the 160 discrepant genes classified as non-essential in *E. coli* but essential in *Shigella*, 34% of the orthologous *E. coli* deletion genotypes exhibit low growth yields (less than 0.5 OD600 after 22 hours of growth in LB [1]). This contrasts strongly with the 2557 ORFs we classified as non-essential in *Shigella*: only 3.7% of the orthologous *E. coli* deletion genotypes had low growth yields (**Fig. 4**). Similar but less striking patterns were observed for growth in glucose minimal MOPS media (**S3 Fig**.).

It is important to note that Tn-seq assays have only limited power to differentiate essential genes from those whose deletion results in severe growth deficiencies. During the course of preparing the library for sequencing, we estimate that at least 20 generations of growth occurred. If a mutant has a growth rate even 60% that of the wild type, we would expect it to undergo only 12 doublings in contrast to the 20 of the wild type. This would result in a greater than 200-fold underrepresentation of such a mutant ($2^{12}/2^{20}$). In addition, this calculation does not take into account any effects that the mutations have on the length of the lag time, which might also have significant effects on the relative frequency of some mutants.

In light of this limited resolution power; given that our prior expectation is that essentiality status changes only rarely; and because we are specifically interested in genes that may have changed in essentiality status, from this point on we focus our analysis on essential *Shigella* genes whose orthologous *E. coli* deletion genotype exhibits robust growth yields (OD600 greater than 0.75 after 22 hours growth in LB (**Fig. 4**)). For these genes, we have relatively high confidence that while their deletion in *E. coli* has few effects on growth, their disruption in *Shigella* is lethal or results in a severe growth deficiency.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Artefacts of the Transposon Screen Explain Some False Positive Discrepancies

35 essential *Shigella* genes have orthologous *E. coli* deletion genotypes with growth yields higher than 0.75 OD600 after 22 hours growth in LB [1] (**Fig. 4**). Careful inspection suggested that some genes were present in this set due to differences in the growth conditions between *E. coli* Keio and PEC collections and our own. For example, *fhuACD*, and *tonB* all appeared in this set of genes (**Table 1**). All four are involved in iron acquisition, and it is likely that iron was limiting in the solid agar media [29] used during the preparation of the Tn-seq library, as compared to the liquid LB used to measure growth in the Keio study.

Table 1. Genes artefactually inferred as essential in *Shigella.*

Gene(s)	Evidence for Artefactual Inference of *Shigella* Essentiality
acrAB, tolC, ybaB, ksgA, yebC, smpB (0.73)[1]	Affect kanamycin resistance [30-32]
dnaQ, holD, recD, *xseA*, *ruvA* (0.58), *ruvB* (0.60), *ruvC* (0.61), *recB* (0.58), *recC* (0.65)	Likely to affect the transposition process; *dnaQ, holD, ruvA,* and *ruvB* inferred as essential using Tn-seq in *Salmonella* [13]
priB	Deficient in plasmid maintenance [33, 34]
fhuACD, tonB	Involved in iron acquisition which is critical for growth in the iron limited media used in this study [29]
rpsT	S20 ribosomal subunit; new data indicates mutants have poor growth [35] (in conflict with Keio data)
miaA	tRNA dimethyl transferase; previous data indicates *E. coli* mutants have poor growth [36] (in conflict with Keio data)
ompA	Outer membrane porin; clean knockouts appear viable [37]; mutant forms are frequently lethal [38]
pitA	Metal phosphate transporter with ten transmembrane segments; transposon disruption of substrate transporters is three-fold more likely to be inferred as essential compared to clean deletion (see main text; **Fig**)
potB	Type I ABC transporter (Putrescine / spermidine transporter)
cysU	Type I ABC transporter (Sulfate / thiosulfate transporter)
sapB	Type I ABC transporter (unknown substrate)
ptsH	Short 306 bp reading frame
ydhR	Short 258 bp reading frame

[1] Gene deletions of orthologous *E. coli* genes with growth levels less than OD600 0.75 have these levels in parentheses

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

For a second set of genes, the discrepancies are likely due to the differences in methodology between the *E. coli* (precise gene deletions) and *Shigella* (transposon inactivation) studies (**Table 1**). We inferred *acrAB* and *tolC* as essential. These genes act together as an efflux pump, and mutations in these genes result in hypersensitivity to antibiotics [39]. Thus, clones with transposon insertions in these genes are unlikely to survive during library growth under kanamycin selection. A similar explanation likely underlies the fact that we inferred *ybaB, ksgA, yebC,* and *smpB* as essential: these four play role in aminoglycoside resistance [30-32].

We also inferred *priB*, *dnaQ*, *holD*, *xseA*, and *recD* as being essential in *Shigella*, although the *E. coli* deletion genotypes exhibit robust growth. All of these are involved in DNA replication, recombination, and double strand break repair, all of which are essential processes in the completion of the transposition process [40]. The related genes *recBC* and *ruvABC* contained a single insert between the five of them, while the *E. coli* deletion genotypes all exhibit only slightly impaired growth of 0.6 OD600 or more (**Table 1**). Certain *recBC* mutants can have considerable effects on the rate of Tn10 excision [41, 42] and we speculate that this may be one reason why we rarely observed insertions in these loci. It has also been speculated that *ruv* mutants inhibit transposition [43]. We propose that after transposition occurs, in order for the event to be successfully resolved, transcription of these genes is often required, and the transposition itself precludes the formation of a proper transcript.

Thus, the dispensability of these ten genes in *E. coli*, and the similarity in their function, suggests that they all affect successful transposon insertion rather than having critical effects on growth. Notably, *priB*, *dnaQ*, *holD, ruvA,* and *ruvB* were also inferred as essential in the closely related bacterium *Salmonella typhimurium* via a high-throughput transposon assay. In the same study *ruvC, recB, recC,* and *ybaB,* were inferred as extremely important for growth while *ksgA* and *yebC* were inferred as significantly impairing growth [13]. Again, the majority of these knockouts in *E. coli* exhibit very robust growth (greater than 0.75 OD600 after 22h growth in LB). Given the roles that these genes are known to play in transposition and antibiotic resistance, this suggests that the inference of essentiality may be due to artefacts of the transposon screen.

For a third set of genes, the literature presents conflicting information on the growth phenotypes, with studies that have individually assessed growth rates suggesting poor growth. These include *rpsT* [35], *miaA* [36], and *ompA* [38] (**Table 1**).

There were also two open reading frames that we inferred as differentially essential as they were completely uninterrupted in our data. However, these two open reading frames, *ydhR* and *ptsH*, are very small and less likely to be disrupted, being 306 bp and 258 bp long,

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

267 respectively. It is probable, then, that this discrepancy is not driven by different physiological
268 roles that they play in *E. coli* as compared to *Shigella*.

269 Finally, we tested for other possible artefactual patterns in the data based on gene function.
270 We asked whether there were specific functional categories in which genes were more likely
271 to be inferred as essential using the transposon mutagenesis screen in *Shigella* as opposed to
272 clean deletions in *E. coli*. We found two functional categories of genes that showed clear
273 enrichment: genes involved in substrate transport and / or active transport, which were 3- and
274 2.1-fold enriched, respectively (**Fig**. **5**). We hypothesize that one reason for this enrichment is
275 that truncated versions of these proteins disturb the operation of the *sec* machinery, thereby
276 decreasing or stopping growth. Thus, we propose that the four active transporters we infer as
277 essential in *Shigella* but not *E. coli* (**Table 1**) are artefacts due to the transposition process
278 resulting in truncated proteins.

279 ## Genes uniquely essential in *Shigella flexneri*

280 While many differences in essentiality classification between *Shigella* and *E. coli* are likely
281 due to (1) severe growth defects present in both *E. coli* and *Shigella* rather than strict
282 essentiality; and (2) differences in environmental conditions (e.g. iron) between the *E. coli*
283 and *Shigella* assays; and (3) artefacts of the *Shigella* transposon screen that do not occur in
284 the *E. coli* knockout screen, we do find a number of genes which we infer to be uniquely
285 essential to Shigella. We expect that the physiological differences between *E. coli* and
286 *Shigella* are driving these differences in gene essentiality (**Table 2**).

287 Among the set of genes essential in *Shigella* but dispensable in *E. coli* is *lysS*: this ORF has a
288 functional homologue in *E. coli* (*lysU* [45]), while in *Shigella flexneri* there is no homologue.
289 Also in this set of genes are *proA, proB, and proC*. These genes act in proline biosynthesis.
290 Given the rich media the cells were grown in, it is surprising that they would be essential. In
291 addition, as *proB* is involved in the first committed step of proline synthesis, its disruption
292 should not cause accumulation of toxic intermediates. However, the data provide strong
293 evidence that the disruption of any these three genes is either lethal or causes severe growth
294 defects (**Fig. 6**). Interestingly, the active proline transporter *putP* is absent from *Shigella* [46].
295 It is also known that in *Salmonella,* the cryptic proline transporter *proY* is silent [47], and we
296 hypothesize that this may also be true of this transporter in *Shigella.* Thus, inefficient proline
297 transport from the media might necessitate biosynthesis.

298 A suite of genes involved in acetate utilization (*aceE*, *aceF*, *ackA*, *pta*, and *pykF*) were all
299 inferred as essential in *Shigella* but dispensable in *E. coli.* The significantly detrimental effect
300 on growth that such mutants have has been noted previously using a completely different
301 approach [21]. The difference in essentiality between these two organisms is most likely due

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

to the absence of acetyl CoA synthetase from *Shigella,* and confirms the sensitivity and relevance of our transposon mutagenesis assay for assaying differences between *E. coli* and *Shigella* biology.

Table 2. Genes inferred as uniquely essential in *Shigella*. All gene deletions in homologous *E. coli* genes show robust growth in rich media after 22 hours (greater than 0.75 OD600), suggesting that these genes are uniquely essential in *Shigella* as compared to *E. coli*.

Gene(s)	Function [44]	Evidence for Different Physiological Roles in E. coli and Shigella
lysS	Aminoacyl tRNA synthetase, tRNA modification	The *lysU* functional homologue is absent in Shigella [45]
proABC	Proline biosynthesis	The active proline transporter *putP* is absent from Shigella [46]. The cryptic transporter *proY* may be silent, as observed in Salmonella [47], possibly necessitating proline biosynthesis
ackA	acetate kinase	All affect acetate accumulation [48] and utilization [21], which is required for robust growth (Shigella lacks the acetyl CoA synthetase present in *E. coli* K12 [49])
pta	phosphotransacetylase	
aceEF	pyruvate dehydrogenase	
pykF	pyruvate kinase	
rfbF, rfbG, rfc, and *rfbI*	sugar nucleotide biosynthesis for LPS	No *E. coli* K12 orthologues, as this locus has been replaced by the laterally transferred *wbb* locus [50]
spr	Murein DD-endopeptidase	None known
tufB	Elongation factor EF-Tu	None known

For only two other orthologous gene pairs is there strong evidence of discrepant essentiality status: *tufB* (two insert locations; **Fig. 7**) and *spr* (one insert at base pair 543 across 567 bp). For neither of these genes do we have a hypothetical causal explanation. Interestingly, we also found very few transposon insertions in the *tufB* paralogue *tufA* (three insert locations; **Fig. 7**), suggesting that this gene, too, is important for *Shigella* growth despite its relative dispensability in *E. coli* (0.72 OD600 after 22h in LB). We note that these two genes are nearly identical in their sequence, which creates ambiguities in mapping some reads. However, this does not explain the absence of reads mapping to either of them. Understanding the molecular mechanisms driving these apparent disparities in growth phenotypes between *Shigella* and *E. coli* is an important topic for future research.

Finally, the transposon insertion data indicated that a within single large operon, containing the ORFs *rfbACEFGI / rfc,* four genes completely lacked insertions (*rfbF, rfbG, rfc,* and *rfbI*) (**Fig. 8**). Only *rfbA* and *rfbC* in this operon have *E. coli* orthologues. The remaining genes lie within a commonly laterally transferred region of the *E. coli* chromosome containing *wbbHIJKL, wzxB* (*rfbX*), and *glf*. These were all laterally transferred into the K12 lineage [50],

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

replacing the *Shigella*-like *rfb* operon. The genes in this operon all play a role in sugar nucleotide biosynthesis necessary for O-antigen synthesis and production of the lipopolysaccharide component of the outer membrane [44]. This provides some evidence that specific aspects of this process have become essential in *Shigella*, despite these genes having been replaced by a laterally transferred set in *E. coli* K12.

Conclusions

By exploiting the extremely close evolutionary relationship of *Shigella flexneri* with *E. coli* K12, the bacterial strain that has been the most extensively and carefully characterized for its essential gene complement [1, 27], we were able to develop an objective metric to precisely quantify how the results of the Tn-seq data relate to essentiality.

A superficial analysis of our Tn-seq data suggested that a total of 451 ORFs in *Shigella* were essential for cellular growth in rich media. This is very much in line with what other Tn-Seq studies have found, with numbers ranging from 480 in *Caulobacter crescentus* [8] to 447 in *B. subvibrioides* to 372 in *Agrobacterium tumefaciens* [3]. However, it is considerably more than the number that had been found in *E. coli* using in-frame gene knockouts, which is on the order of 300 essential genes. In addition, we found that close to 100% of the reading frames that were classified as essential in *E. coli* K12 were also essential in *S. flexneri*, giving us a strong prior expectation that the essentiality classifications should match between these two taxa.

A more nuanced analysis suggested four explanations for artefactual discrepancies in essentiality between *E. coli* and *Shigella*: (1) many *Shigella* genes were not strictly essential but instead gene disruption caused severe growth impairment; (2) differences in experimental conditions (i.e. iron availability); (3) many of the genes we inferred as essential were important for antibiotic resistance or successful transposition, and are in fact dispensable for growth; and (4) transposon disruption of specific functional classes of genes may result in systematically different effects as compared to gene deletions, for example due to the production of truncated protein products. By carefully dissecting the functions of discrepant genes that do not appear to be artefactual, we were able to pinpoint several genes for which there is some evidence of differential physiological roles in *E. coli* and *Shigella*. Among others, these included *lysS,* three genes involved in proline biosynthesis, and a suite of genes involved in acetate utilization (**Table 2**). In addition to these, we found one large operon which appears to have an essential role in *Shigella* growth but which is missing completely in *E. coli.* Surprisingly, we found only two additional genes that are differentially essential (*tufB* and *spr*) (**Table 2**).

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Even after attempting to decrease false positive inferences of gene essentiality in *Shigella*, it appears to be considerably more common for genes to be dispensable for growth in *E. coli*, but critical for growth in *Shigella*. We suggests that one reason *Shigella* more may have a larger complement of essential genes than *E. coli* is that it frequently lives as an intracellular pathogen, and may have lost some of the functional redundancy that is present in *E. coli*. This may occur because host environments provide an abundance of nutrients, or because pathogens requiring a small infectious dose, such as *Shigella* [51], have inherently smaller population sizes and are more subject to genetic drift. A third possibility is that changes in gene function or redundancy may have occurred through selection for increased virulence, which has resulted in the inactivation of certain genes being selectively advantageous. Finally, we note that the discrepancies in essentiality between these two bacteria may be exploited to develop antibiotics that have strain-specific effects [21].

Methods

Strains

For all experiments, *Shigella flexneri* 2457T Δ*icsA* was provided by M. B. Goldberg was used as the parental strain. This strain is unable to exploit the host actin cytoskeleton for motility and spreading [52]. Bacterial cells were grown in Tryptic Soy Broth (TSB) media. For experiments using eukaryotic cells, HeLa cells were cultured in DMEM supplemented with 10 mM Hepes, 25 mM glucose, and 4 mM glutamine. *Shigella* were grown to exponential phase in tryptic soy broth, coated with poly-L-lysine, and added at a multiplicity of infection of 25, resulting in an infection rate of around 60%. *Shigella* was centrifuged onto HeLa cells (600 × g for 5 min). At 30 min postinfection, we added gentamicin (100 μg/mL) to kill extracellular bacteria. Bacterial cells were allowed to grow within HeLa cells for a total of 4 hours.

Transposon library

Using a Tn10 transposon with a T7 promoter [23, 24] we created a library consisting of approximately 10^6 clones. This library was created by mating *E. coli* strain BW20767 containing the pJA1 transposon plasmid with a spontaneous nalidixic acid resistant clone of *Shigella flexneri* 2457T Δ icsA for 5 hours. Transposase expression was induced by plating onto TSB plates containing 0.2 mM isopropyl-β-D-thiogalactoside (IPTG). Colonies were allowed to grow at 37°C for 18 hours on TSB agar plates. All colonies from these plates were then pooled and 100μl aliquots of the transposon library were stored at -80°C.

Three replicate experiments were carried out on different days in which an aliquot of the transposon library was grown for 18 hours in TSB to stationary phase, diluted 1:100 and

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

grown to exponential phase (0.7 OD600). This exponential phase culture was split into two: part of the bacterial culture was pelleted and saved and other was used for infecting HeLa cells (as described in [21]). After 3.5 hours, HeLa cells infected with the *Shigella* transposon library were trypsinized and pelleted. Uninfected HeLa cells were also collected and used to spike the original bacterial culture not used for HeLa infection in order to account for HeLa DNA. All resulting DNA was extracted using the Bacterial Genomic Miniprep Kit (Sigma).

Sequence library construction and sequencing

To amplify the transposon region from these pools, we used one top strand primer annealing to the transposon and a pool of three bottom strand primers each of which consisted of 10 random nucleotides followed by a pentamer of common nucleotides in *E. coli* [53]: N_{10}GGTGC, N_{10}GATAT, and N_{10}AGTAC, using Phusion pfu (**S4 Fig**). A nested PCR was then performed to add the P7 and P5 Illumina adapters, as well as a barcode. The products from this second PCR were then size selected for inserts between 200bp and 300bp, quantified using a Qubit, and sequenced on an Illumina HiSeq2000 at the D-BSSE Quantitative Genomics Facility resulting in 49bp single end reads. We used a custom sequencing primer on the P5 end of the molecule such that on both ends of the molecule, reads started directly on the chromosome.

Read mapping

In total, we obtained 198,682,954 reads. We found that the number of reads at each location in the genome varied by up to four orders of magnitude. For this reason, we considered only whether an insertion had occurred at a specific location, and not on the number of reads we obtained at a specific location, which is likely to be highly biased due to PCR artefacts. We thus first deduplicated the reads using *tally* [54], and then used bowtie2 [55] to align the reads to the *Shigella flexneri* 2a 2457T genome and the *Shigella flexneri* 2a str. 301 plasmid pCP301. The sequence of the *S. flexneri* 2457T plasmid is not available. However, the *S. flexneri* 2457T and 2a str. 301 plasmids are nearly identical in sequence (differing by 30 SNPs; see below). Sequence reads were not trimmed for quality as read quality is taken into account in bowtie2. We used the --sensitive-local option to allow soft clipping on the 3' end of the reads (so that reads that contained adapter sequences at the 3' end could map successfully), and required at least 22bp of matching sequence at the 5' end of the read.

We checked for single nucleotide polymorphisms (SNPs) on both the chromosome and the plasmid using the samtools mpileup and bcftools utilities [56, 57]. We retained as possible SNPs only those sites that fulfilled the following three criteria: (1) the SNP was inferred as homozygous (necessarily true, as *Shigella* is haploid); (2) the quality score was above 20; and (3) at least three reads on both the reverse and forward strands confirmed the SNP. We found

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

99 SNPs on the chromosome (as compared to the reference *Shigella flexneri* 2457T in NCBI) and 30 SNPs on the plasmid (as compared to the *Shigella flexneri* 2a str. 301 plasmid in NCBI (in addition to 12 and 2 small indels, respectively). These are detailed in **S2 Table** and **S3 Table,** respectively.

Within chromosomal protein coding regions, 44% of all SNPs were synonymous, while 32% fell outside of genic regions (i.e. protein coding or RNA genes). These fractions are greater than one would expect if such SNPs were randomly located on the genome. Only 24% of all mutations in chromosomal coding regions are expected to be synonymous (not accounting for mutational biases), and only 28% of the chromosome is annotated as nongenic (including repeat regions, although for many of these regions, the absence of an annotation may be erroneous). Additionally, only 2 of the 12 (17%) small chromosomal indels fell in coding regions. This suggests that there was some selection against nonsynonymous substitutions that occurred during the culturing and derivation of the *Shigella flexneri* 2a 2457T *virG* mutant. More importantly, the small number of SNPs that we found suggests that few, if any, reads remained unmapped due to sequence differences between the strain used in our experiments and the sequenced GenBank strain.

In total, the reads mapped to 89,028 unique locations on the forward strand and 83,074 on the reverse strand of the chromosome, for a total of 172,102 insertions. Some of these insertions occurred at identical positions but on opposite strands, so in total, insertions occurred at 131,670 unique sites in the chromosome. Correspondingly, the reads mapped to 8,208 unique locations on the forward strand and 8,585 unique locations on the reverse strand of the plasmid, for a total of 12,552 unique sites. During the insertion of the Tn10 transposon, a 9 bp target DNA sequence is duplicated [58]. We accounted for this duplication in calculating the distances between insertions (by moving the inferred site of insertion for one direction (we arbitrarily selected the antisense direction) backward by 9 bp). Similarly, this duplication was accounted for in calculating various statistics of insertions within genes: sense insertions that were inferred as occurring in the last 9 bp of a gene were ignored in calculating the mean number of insertions per gene (as these bp are duplicated upstream of the insertion). Antisense insertions occurring in the first 9 bp of a gene were ignored, as these bp are duplicated downstream of the insertion.

Using the read frequencies at all unique insert locations, we found that the transposon insertions occurred in a biased manner, integrating more often at sites similar to the known 9bp consensus NGCTNAGC [58], although this bias was relatively weak (**Figs. 1A** and **B**, insets). This low level of bias is likely due to our using a transposon with reduced hotspot activity [22]. In addition, we found that insertion frequency was slightly influenced by

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

nucleotides further downstream of this 9bp consensus (**Figs. 1A** and **B**, insets). Sequence logos for this analysis were visualized using the R package seqLogo [59].

The median distance between inserts was 17 bp in the chromosome and 9 bp in the plasmid (**Fig. 1B**), suggesting that the transposon libraries yielded a relatively fine-grained map of the essential genomic complement for both the chromosome and the plasmid.

As expected given the variation in insertion densities across the chromosome, we found high variance in the distribution of inter-insert distances. The total length of the *S. flexneri* genome is 4,599,354 bp in total. Given that we observed 131,670 inserts, under a model of random insertion, we would expect a median distance between inserts of 35bp, with 95% of all inter-insert distances being less than 107 bp (under the assumption that these distances are distributed in a geometric manner (i.e. a negative binomial with the number of successes set to one). For the plasmid, we observed 12,552 inserts over 221,618 bp, such that we expect a median distance of 18bp between inserts, and that 95% of all inter-insert distances are less than 59 bp. However, as noted above, we found that on average transposons insertions were separated by a median of 17bp on the chromosome and 9bp on the plasmid. Fitting a geometric distribution to the observed data over 99% of the range of the inter-insert distances (i.e. from 1 to 237 bp for the chromosome and from 1 to 78 bp for the plasmid) more exactly quantified this over-dispersion, and showed that uninterrupted regions in the chromosome greater than 100 bp were considerably enriched (**Fig. 1C**).

Paired end read mapping and inference of IS element dynamics

We used 100 bp paired end Illumina sequencing data from this same library to look for structural rearrangements due to IS elements in the genome. However, this analysis was complicated by the fact that many IS elements share close to 100% identity with others around the genome. During these analyses we thus restricted our searches to regions of the genome for which we had *a priori* expectations that they harboured a rearrangement (i.e. if there were no inserts and the orthologous *E. coli* locus was non-essential or absent). Specifically, we followed the following procedure: we extracted a 50 kilobase pair (Kbp) region from the genome surrounding each hypothesized rearrangement (in all cases, this was a deletion). We then used bowtie2 with the paired end option, allowing up to 10 Kbp inserts to map all reads from our 100 bp PE dataset. From these mapped reads, we retained only read pairs that had (1) mapping quality scores greater than 20; (2) at least one read that matched perfectly (i.e. at all 101 bases of the read) to the genome; and (3) were unique in their length at any specific location (thereby excluding artefacts such as PCR doublets). From these paired reads we then inferred the insert size, which is plotted in **S5 Fig**. The vast majority of insert sizes ranged between 100 and 400bp. However, some were much larger (e.g. up to 9,000 bp

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in **S5B Fig.**). We inferred that these surrounded regions of the genome that must have been deleted.

Such deletions would result in the set of genes contained within as being inferred as essential because of their lack of transposon insertions. However, in the vast majority of cases, we found that when large operons lacked insertions but had non-essential orthologous operons in *E. coli*, or were missing entirely from *E. coli*, these operons were in fact missing from the *Shigella* clone that we used, most likely due to the rapid dynamics of IS elements in this bacterium [60]. For example, no sequence reads we obtained mapped to the *yeaKLMNOP* operon, which spans a total of 9,240 bp. Upon further analysis using a paired end genomic data set, we found that this region was clearly missing from our *Shigella* clone (**S6B Fig**.). This was similarly true for several other operons, as well as for single genes. We did not consider any region in which we identified a deletion in our downstream analyses.

Essential open reading frames

We identified 3,027 unambiguously ORFs that were present in both *E. coli* and *Shigella flexneri* 2457T [61], and for which we had essentiality data. We used reciprocal shortest distance [62] to find orthologues, with the requirement that the alignment of the two hypothetical orthologues extend over at least 60% of the longer ORF. To establish a gold-standard set of essential genes we combined the data from two studies of the effects of gene deletion on growth in *E. coli* K12: the Keio collection [1] and the PEC study [27]. We retained only those ORFs which we had data on essentiality from both studies. We then quantified which transposon insertion patterns that most closely corresponded with the essentiality delineations in theses studies. Specifically, we selected the feature that maximized the number of true positive essential genes (maximizing the sensitivity) while minimizing the number of FP (maximizing specificity) (this metric is a receiver operator characteristic for which we quantified the area under the curve (AUC; **S2 Fig.**)). We selected from eleven non-independent features: (1) the total number of insertions; (2) the mean number of bp between insertions; (3) the median number of bp between insertions; (4) the number of bp in the 5' end preceding the first insertion; (5) the number of bp in the 5' end preceding the first insertion relative to the total bp in the gene; (6) the number of bp in the 5' end preceding the second insertion; (7) the number of bp in the 5' end preceding the second insertion relative to the total bp in the gene; (8) the number of bp in the longest uninterrupted stretch of the gene; (9) the number of bp in the longest uninterrupted stretch of the gene relative to the total length of the gene, and (10) the number of bp in the longest stretch of the gene interrupted by at most one insertion; (11) the number of bp in the longest stretch of the gene interrupted by at most one insertion, relative to the total length of the gene.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We found that for both the PEC dataset and the Keio dataset, the two best predictors of essentiality were the mean distance between inserts (AUC = 0.972 for the PEC dataset, 0.952 for the Keio dataset, and 0.973 for the genes on which both datasets agreed on the essentiality classifications); and the fraction of the gene that lay in the longest uninterrupted region (AUC = 0.969 for the PEC dataset, 0.955 for the Keio dataset, and 0.971 for the genes on which both datasets agreed on the essentiality classifications) (**S2 Fig.**). We selected mean distance as on average, it marginally outperformed the other statistic on the gold standard data set.

We note that for eight of the 14 genes classified as essential solely in the Keio dataset, the orthologous *Shigella* ORFs have mean distances less than 30 bp, suggesting that these genes may be falsely annotated as essential in the Keio study. In contrast, nine of the ten genes inferred as essential solely in the PEC dataset have mean distances greater than 200 bp; the tenth has a mean distance of 189 bp.

tRNA disruptions

We found insertions in 27 out of 99 tRNAs, with tRNAs for certain amino acids being considerably overrepresented (**S2 Table**).

Additional analyses of differentially classified essential genes

We also tested for the enrichment of certain functional categories in the set of genes that were classified as being essential in *Shigella* but not *E. coli*. This differs from the analysis present in **Fig. 5** in that we are asking whether across a broad set of functions, are specific categories enriched for *Shigella*-essential genes. In **Fig. 5** we ask whether *within* a single functional category, is there a much higher fraction of *Shigella* essential genes than we would expect, given the fraction of genes in that functional category that are essential in *E. coli.*

Thus, for this analysis, we separated genes by primary functional category and secondary subcategory using the MultiFun designations (e.g. the primary functional category cell processes divided into the secondary subcategories of cell division, SOS, stress, protection, and motility). We then calculated the fraction of *Shigella*-essential genes within each secondary subcategory and compared this to the total fraction of *Shigella*-essential genes within the primary category (e.g. we calculated the fraction of essential genes in Ribosomal Function (the secondary category) and the fraction of *Shigella*-essential genes in all other categories in Cell Structure (the primary category) (**S6 Fig**.). We tested for enrichment (depletion) using a Fisher exact test.

We also examined gene conservation. Highly conserved genes were considered to be those present in more than 50% of all gamma-proteobacteria [61]. We found that genes classified as uniquely essential in *Shigella* were much more conserved across gamma-proteobacteria (79%

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

566 highly conserved) compared to genes that were found non-essential in both *E. coli* and
567 *Shigella* (36% highly conserved; p=1.0e-33, Wilcox rank sum test).

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Availability of supporting data

All read data are in the SRA with accession numbers XXX.

List of abbreviations used

bp - base pairs; *Shigella* – *Shigella flexneri* 2a 2457T; *E. coli* – *Escherichia coli* BW25113; ORF – open reading frame; PEC - Profiling the E coli Chromosome database

Competing interests

The authors declare no competing interests.

Authors' contributions

NEF, DB, and OKS conceived and designed the transposon mutagenesis. NEF performed the mutagenesis and sequencing. OKS analysed the data with input from DB and NEF. NEF and OKS wrote the paper.

Acknowledgements

The authors thank Luise Wolf for input on designing the sequencing protocol.

Endnotes

References

1. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006 0008. Epub 2006/06/02. doi: 10.1038/msb4100050. PubMed PMID: 16738554; PubMed Central PMCID: PMC1681482.

2. Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, et al. Global transposon mutagenesis and a minimal mycoplasma genome. Science. 1999;286(5447):2165-9. doi: Doi 10.1126/Science.286.5447.2165. PubMed PMID: WOS:000084157300055.

3. Curtis PD, Brun YV. Identification of essential Alphaproteobacterial genes reveals operational variability in conserved developmental and cell cycle systems Mol Microbiol. 2014. doi: 10.1111/mmi.12686.

4. Blank D, Wolf L, Ackermann M, Silander OK. The predictability of molecular evolution during functional innovation. P Natl Acad Sci USA. 2014;111(8):3044-9. PubMed PMID: WOS:000332180900041.

5. Liu GW, Yong MYJ, Yurieva M, Srinivasan KG, Liu J, Lim JSY, et al. Gene Essentiality Is a Quantitative Property Linked to Cellular Evolvability. Cell. 2015;163(6):1388-99. PubMed PMID: WOS:000366044800003.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

599 6. Bergmiller T, Ackermann M, Silander OK. Patterns of Evolutionary Conservation of Essential
600 Genes Correlate with Their Compensability. Plos Genet. 2012;8(6). PubMed PMID:
601 WOS:000305961000055.
602 7. van Opijnen T, Camilli A. Transposon insertion sequencing: a new tool for systems-level
603 analysis of microorganisms. Nat Rev Microbiol. 2013;11(7). doi: Doi 10.1038/Nrmicro3033. PubMed
604 PMID: WOS:000320368400013.
605 8. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, et al. The essential
606 genome of a bacterium. Mol Syst Biol. 2011;7. doi: Artn 528
607 Doi 10.1038/Msb.2011.58. PubMed PMID: ISI:000294537800007.
608 9. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness
609 and genetic interaction studies in microorganisms. Nat Methods. 2009;6(10):767-U21. doi: Doi
610 10.1038/Nmeth.1377. PubMed PMID: ISI:000270355200023.
611 10. Kamp HD, Patimalla-Dipali B, Lazinski DW, Wallace-Gadsden F, Camilli A. Gene Fitness
612 Landscapes of Vibrio cholerae at Important Stages of Its Life Cycle. Plos Pathog. 2013;9(12). doi: Artn
613 E1003800
614 Doi 10.1371/Journal.Ppat.1003800. PubMed PMID: WOS:000330535400034.
615 11. Klein BA, Tenorio EL, Lazinski DW, Camilli A, Duncan MJ, Hu LDT. Identification of
616 essential genes of the periodontal pathogen Porphyromonas gingivalis. Bmc Genomics. 2012;13. doi:
617 Artn 578
618 Doi 10.1186/1471-2164-13-578. PubMed PMID: WOS:000314646600001.
619 12. van Opijnen T, Camilli A. A fine scale phenotype-genotype virulence map of a bacterial
620 pathogen. Genome Res. 2012;22(12):2541-51. doi: Doi 10.1101/Gr.137430.112. PubMed PMID:
621 WOS:000311895500022.
622 13. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay
623 of every Salmonella Typhi gene using one million transposon mutants. Genome Res.
624 2009;19(12):2308-16. doi: Doi 10.1101/Gr.097097.109. PubMed PMID: ISI:000272273400015.
625 14. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. Tracking insertion
626 mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes
627 required in the lung. P Natl Acad Sci USA. 2009;106(38):16422-7. doi: Doi 10.1073/Pnas.0906627106.
628 PubMed PMID: WOS:000270071600076.
629 15. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, et al. Identifying
630 Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. Cell Host Microbe.
631 2009;6(3):279-89. doi: Doi 10.1016/J.Chom.2009.08.003. PubMed PMID: WOS:000270290700011.
632 16. Lee SA, Gallagher LA, Thongdee M, Staudinger BJ, Lippman S, Singh PK, et al. General and
633 condition-specific essential functions of Pseudomonas aeruginosa. P Natl Acad Sci USA.
634 2015;112(16):5189-94. PubMed PMID: WOS:000353239100085.
635 17. Chao MC, Pritchard JR, Zhang YJJ, Rubin EJ, Livny J, Davis BM, et al. High-resolution
636 definition of the Vibrio cholerae essential gene set with hidden Markov model-based analyses of

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

637 transposon-insertion sequencing data. Nucleic Acids Res. 2013;41(19):9033-48. doi: Doi
638 10.1093/Nar/Gkt654. PubMed PMID: WOS:000326044700026.
639 18. Pritchard JR, Chao MC, Abel S, Davis BM, Baranowski C, Zhang YJJ, et al. ARTIST: High-
640 Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. Plos Genet.
641 2014;10(11). PubMed PMID: WOS:000345455200029.
642 19. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, et al. Comprehensive
643 transposon mutant library of Pseudomonas aeruginosa. P Natl Acad Sci USA. 2003;100(24):14339-44.
644 PubMed PMID: ISI:000186803800105.
645 20. Chao MC, Abel S, Davis BM, Waldor MK. The design and analysis of transposon insertion
646 sequencing experiments. Nat Rev Microbiol. 2016;14:119-28.
647 21. Kentner D, Martano G, Callon M, Chiquet P, Brodmann M, Burton O, et al. Shigella reroutes
648 host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth. P
649 Natl Acad Sci USA. 2014;111(27):9929-34. doi: Doi 10.1073/Pnas.1406694111. PubMed PMID:
650 WOS:000338514800054.
651 22. Kleckner N, Bender J, Gottesman S. Uses of Transposons with Emphasis on Tn10. Method
652 Enzymol. 1991;204:139-80. PubMed PMID: WOS:A1991GN46900007.
653 23. Badarinarayana V, Estep PW, Shendure J, Edwards J, Tavazoie S, Lam F, et al. Selection
654 analyses of insertional mutants using subgenic-resolution arrays. Nat Biotechnol. 2001;19(11):1060-5.
655 PubMed PMID: ISI:000172002600022.
656 24. Chan K, Kim CC, Falkow S. Microarray-based detection of Salmonella enterica serovar
657 typhimurium transposon mutants that cannot survive in macrophages and mice. Infect Immun.
658 2005;73(9):5438-49. doi: Doi 10.1128/Iai.73.9.5438-5449.2005. PubMed PMID:
659 WOS:000231460000017.
660 25. Sansonetti PJ, Kopecko DJ, Formal SB. Involvement of a Plasmid in the Invasive Ability of
661 Shigella-Flexneri. Infect Immun. 1982;35(3):852-60. PubMed PMID: WOS:A1982NE19800013.
662 26. Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, et al. Update on
663 the Keio collection of Escherichia coli single-gene deletion mutants. Mol Syst Biol. 2009;5. doi: Artn
664 335
665 Doi 10.1038/Msb.2009.92. PubMed PMID: WOS:000273359200008.
666 27. Kato JI, Hashimoto M. Construction of consecutive deletions of the Escherichia coli
667 chromosome. Mol Syst Biol. 2007;3:132. PubMed PMID: ISI:000249223700002.
668 28. Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, et al. A
669 postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: Application
670 to Mycobacterium tuberculosis. P Natl Acad Sci USA. 2003;100(12):7213-8. doi: Doi
671 10.1073/Pnas.1231432100. PubMed PMID: WOS:000183493500058.
672 29. Inoue T, Shingaki R, Hirose S, Waki K, Mori H, Fukui K. Genome-wide screening of genes
673 required for swarming motility in Escherichia coli K-12. J Bacteriol. 2007;189(3):950-7. PubMed
674 PMID: WOS:000244112100030.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

675 30. Skunca N, Bosnjak M, Krisko A, Panov P, Dzeroski S, Smuc T, et al. Phyletic Profiling with
676 Cliques of Orthologs Is Enhanced by Signatures of Paralogy Relationships. Plos Comput Biol.
677 2013;9(1). PubMed PMID: WOS:000314595600012.
678 31. Sparling PF. Kasugamycin Resistance . 30s Ribosomal Mutation with an Unusual Location on
679 Escherichia-Coli Chromosome. Science. 1970;167(3914):56-&. PubMed PMID:
680 WOS:A1970E985700019.
681 32. Corvaisier S, Bordeau V, Felden B. Inhibition of transfer messenger RNA aminoacylation and
682 trans-translation by aminoglycoside antibiotics. J Biol Chem. 2003;278(17):14788-97. PubMed PMID:
683 WOS:000182516100028.
684 33. Kim PD, Banack T, Lerman DM, Tracy JC, Camara JE, Crooke E, et al. Identification of a
685 novel membrane-associated gene product that suppresses toxicity of a TrfA peptide from plasmid RK2
686 and its relationship to the DnaA host initiation protein. J Bacteriol. 2003;185(6):1817-24. PubMed
687 PMID: ISI:000181448900008.
688 34. Berges H, Oreglia J, JosephLiauzun E, Fayet O. Isolation and characterization of a priB
689 mutant of Escherichia coli influencing plasmid copy number of Delta rop ColE1-type plasmids. J
690 Bacteriol. 1997;179(3):956-8. PubMed PMID: WOS:A1997WE44000050.
691 35. Bubunenko M, Baker T, Court DL. Essentiality of ribosomal and transcription antitermination
692 proteins analyzed by systematic gene replacement in Escherichia coli. J Bacteriol. 2007;189(7):2844-
693 53. PubMed PMID: WOS:000245842000030.
694 36. Diaz I, Pedersen S, Kurland CG. Effects of Miaa on Translation and Growth-Rates. Mol Gen
695 Genet. 1987;208(3):373-6. PubMed PMID: WOS:A1987J026300002.
696 37. Ambrosi C, Pompili M, Scribano D, Zagaglia C, Ripa S, Nicoletti M. Outer Membrane
697 Protein A (OmpA): A New Player in Shigella flexneri Protrusion Formation and Inter-Cellular
698 Spreading. Plos One. 2012;7(11). PubMed PMID: WOS:000311151900175.
699 38. Freudl R, Braun G, Hindennach I, Henning U. Lethal Mutations in the Structural Gene of an
700 Outer-Membrane Protein (Ompa) of Escherichia-Coli-K12. Mol Gen Genet. 1985;201(1):76-81.
701 PubMed PMID: WOS:A1985ARK1200013.
702 39. Ma D, Cook DN, Alberti M, Pon NG, Nikaido H, Hearst JE. Genes Acra and Acrb Encode a
703 Stress-Induced Efflux System of Escherichia-Coli. Mol Microbiol. 1995;16(1):45-55. PubMed PMID:
704 WOS:A1995QU56400005.
705 40. Kleckner N. Transposable elements in prokaryotes. Annu Rev Genet. 1981;15:341-404.
706 PubMed PMID: Medline:6279020.
707 41. Lundblad V, Taylor AF, Smith GR, Kleckner N. Unusual alleles of recB and recC stimulate
708 excision of inverted repeat transposons Tn10 and Tn5. P Natl Acad Sci USA. 1984;81(3):824-8.
709 PubMed PMID: Medline:6322169.
710 42. Chan SH, Lau A, Lei V, Woo J. Effects of recB, recC and recF mutations on Tn10
711 Transposition in Escherichia coli. ournal of Experimental Microbiology and Immunology. 2006;9:75-
712 80.
713 43. Attfield PV, Benson FE, Lloyd RG. Analysis of the ruv locus of Escherichia coli K-12 and
714 identification of the gene product. J Bacteriol. 1985;164(1):276-81. PubMed PMID: Medline:2995311.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

715 44. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-
716 Martinez C, et al. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res.
717 2013;41(D1):D605-D12. PubMed PMID: WOS:000312893300086.
718 45. Kawakami K, Ito K, Nakamura Y. Differential Regulation of 2 Genes Encoding Lysyl-
719 Transfer Rna-Synthetases in Escherichia-Coli - Lysu-Constitutive Mutations Compensate for a Lyss
720 Null Mutation. Mol Microbiol. 1992;6(13):1739-45. PubMed PMID: WOS:A1992JC20400003.
721 46. Zhang XB, Liu H, Yang F, Yang J, Xue Y, Dong J, et al. Comparative genome analysis of
722 deleted genes in Shigella flexneri 2a strain 301. Chinese Sci Bull. 2003;48(9):846-52. PubMed PMID:
723 WOS:000183995400002.
724 47. Liao MK, Gort S, Maloy S. A cryptic proline permease in Salmonella typhimurium.
725 Microbiol-Uk. 1997;143:2903-11. PubMed PMID: WOS:A1997XW07300007.
726 48. Wong MS, Wu S, Causey TB, Bennett GN, San KY. Reduction of acetate accumulation in
727 Escherichia coli cultures for increased recombinant protein production. Metab Eng. 2008;10(2):97-108.
728 PubMed PMID: WOS:000261595200004.
729 49. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. Genome-scale
730 metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to
731 nutritional environments. P Natl Acad Sci USA. 2013;110(50):20338-43. PubMed PMID:
732 WOS:000328061700082.
733 50. Hooper SD, Berg OG. Gene import or deletion: A study of the different genes in Escherichia
734 coli strains K12 and O157 : H7. J Mol Evol. 2002;55(6):734-44. PubMed PMID:
735 WOS:000179503100012.
736 51. Kothary MH, Babu US. Infective dose of foodborne pathogens in volunteers: A review. J
737 Food Safety. 2001;21(1):49-73. PubMed PMID: WOS:000169172400004.
738 52. Goldberg MB, Theriot JA. Shigella-Flexneri Surface Protein Icsa Is Sufficient to Direct Actin-
739 Based Motility. P Natl Acad Sci USA. 1995;92(14):6572-6. PubMed PMID: WOS:A1995RG73600073.
740 53. Caetanoanolles G. Amplifying DNA with Arbitrary Oligonucleotide Primers. Pcr Meth Appl.
741 1993;3(2):85-94. PubMed PMID: WOS:A1993MD19100001.
742 54. Davis MPA, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. Kraken: A set of
743 tools for quality control and analysis of high-throughput sequence data. Methods. 2013;63(1):41-9.
744 55. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
745 2012;9(4):357-U54. PubMed PMID: WOS:000302218500017.
746 56. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
747 population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987-93.
748 doi: Doi 10.1093/Bioinformatics/Btr509. PubMed PMID: WOS:000296099300009.
749 57. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
750 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi: Doi
751 10.1093/Bioinformatics/Btp352. PubMed PMID: ISI:000268808600014.
752 58. Halling SM, Kleckner N. A Symmetrical 6-Base-Pair Target Site Sequence Determines Tn10
753 Insertion Specificity. Cell. 1982;28(1):155-63. doi: Doi 10.1016/0092-8674(82)90385-3. PubMed
754 PMID: WOS:A1982MY42100020.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

59. Bembom O. seqLogo: An R package for plotting DNA sequence logos. 2007.

60. Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, et al. Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T (vol 71, pg 2775, 2003). Infect Immun. 2003;71(7):4223-. PubMed PMID: ISI:000183797200078.

61. Silander OK, Ackermann M. The constancy of gene conservation across divergent bacterial orders. BMC Research Notes. 2009;2:2.

62. Wall DP, Fraser HB, Hirsh AE. Detecting putative orthologs. Bioinformatics. 2003;19(13):1710-1. PubMed PMID: ISI:000185310600016.

63. Serres MH, Riley M. MultiFun, a multifunctional classification scheme for Escherichia coli K-12 gene products. Microb Comp Genomics. 2000;5(4):205-22. PubMed PMID: 11471834.

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table and Figure legends

Table 1. Genes artefactually inferred as essential in *Shigella*. Listed here are those genes that were likely inferred as essential largely due to the gene disruptions having direct effects on (1) antibiotic resistance, (2) successful transposition events, (3) differences between the growth conditions of the *E. coli* and *Shigella* essentiality studies, or (4) systematically different effects of gene disruption versus gene deletion.

Table 2. Genes inferred as uniquely essential in *Shigella*. Listed here are genes inferred as essential in *Shigella*, but which have orthologous *E. coli* deletion genotypes that exhibit robust growth (greater than 0.75 OD600 after 22h growth in LB). The genes in the *rfb* operon have no orthologues in *E. coli* K12 (see main text).

Fig 1. **Histograms of distances between inserts on the chromosome for (A) the *Shigella* chromosome and (B) the *Shigella* virulence plasmid**. The median distance between inserts is indicated by the dotted line. The insets in (A) and (B) indicate the slight but detectable biases in transposon insert location using a weight matrix motif. The reverse cumulative plots show the observed fractions of distances between inserts for the chromosome (C) and the plasmid (D). In blue, the observed frequencies are plotted. In black, the expected frequencies are plotted, given a geometric distribution (negative binomial with the number of successes set to one) of inter-insert distances (see main text). For both the chromosome and the plasmid, there are considerably more large regions uninterrupted by transposons than one would expect given the geometric null model, observed as a shift of the curve to the right.

Fig 2. **Orthologous genes known to be essential in *E. coli* are also essential in *Shigella.*** A region of the *Shigella* chromosome is shown, with genes whose orthologues are known to be essential for growth in *E. coli* (coloured in white) [1, 27], or non-essential (coloured in grey). The unique locations of transposon insertions are plotted as vertical black segments. In the genome region shown here, none of the genes essential in *E. coli* have orthologues that are interrupted in *Shigella*.

Fig 3. Differences in essentiality classification between *E. coli* K12 and *Shigella.* (A) Cumulative distributions showing the mean distances between inserts for ORFs depending on whether their orthologues are known to be non-essential (black curve) or essential (blue curve) in *E. coli*. All ORFs that are completely uninterrupted by transposons have been plotted at the very right end of the x-axis. The dotted vertical line indicates the cut-off that we used to delineate essentiality in *Shigella* (a mean distance between transposons of 250 bp or more). The 17 blue points to the left of the dotted vertical line indicate ORFs that are essential in *E. coli* but not *Shigella* by our metric. These are likely to be false positives (i.e. non-essential in both *Shigella* and *E. coli*), as all have inter-insert distances greater than 100 bp

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(see main text). Black points to the right of the dotted vertical line indicate ORFS that we classify as essential in *Shigella* but not in *E. coli*. Many of these ORFs have *E. coli* orthologues whose deletion genotypes exhibit robust growth, suggesting that their essentiality status has changed. (B) A Venn diagram showing the overlap between essential orthologous ORFs in *E. coli* and *Shigella*.

Fig 4. **Orthologous gene pairs that are non-essential in *E. coli* but inferred as essential in *Shigella* (blue) tend to exhibit low growth yields in *E. coli*.** ORFs that we infer to be uniquely essential in *Shigella* consistently have *E. coli* orthologues with low growth phenotypes in LB media after 22 hours (apparent as a strong leftward shift in the cumulative curve). For genes inferred as uniquely essential in *Shigella,* 34% of the orthologous *E. coli* deletion genotypes exhibit low growth yields (less than 0.5 OD600 after 22 hours of growth in LB). For genes we classified as non-essential in *Shigella* and *E. coli* only 3.7% exhibit low growth yields. Thus, some genes we infer as essential in *Shigella* may not be strictly essential, but instead be required for robust growth. Despite this enrichment for low-growth phenotypes, there are many genes which we infer as essential in *Shigella*, but which have *E. coli* orthologues whose deletion genotypes exhibit robust growth (OD600 greater than 0.75 after 22 hours growth in LB).

Fig 5. **Transposon disruption of *Shigella* genes with transport-related functions are more likely to be inferred as essential compared to clean deletions of similarly functioning genes in *E. coli*.** We classified genes according to function using the MultiFun functional classification system [63]. For any category containing more than ten essential *E. coli* genes, we also calculated the number of *Shigella*-essential genes. As expected, most categories show a relative excess of *Shigella*-essential genes, as we inferred approximately 50% more genes as being essential in *Shigella* versus *E. coli* (**Fig. 3B**). However, two functional categories show a clear excess above this level: substrate transport and active transport, showing a 3- and 2.1-fold increased probability of inferring a gene as being essential in *Shigella* as opposed to *E. coli*. This provides evidence that genes in these functional categories may be more likely to be inferred as artefactually essential. For each functional category (y-axis), we show the number of genes in that category (to the right of each bar); the number of genes found to be essential in *E. coli* (within each bar); and the level of enrichment of essential genes in *Shigella* (x-axis).

Fig 6. Three genes involved in proline biosynthesis (*proABC*) appear uniquely essential in *Shigella*. The orthologous *E. coli* deletion strains exhibit robust growth (OD600 greater than 0.75 after 22 hours growth in LB), but are essential by our criteria**.** *proA* and *proC* completely lack transposon insertions, while *proB* contains only two insertions near the 3'end,

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

836 which leaves approximately 70% of the gene intact, including the entire kinase and substrate-
837 binding domain.

838 **Fig 7. Both elongation factor paralogues *tufA* and *tufB* appear differentially essential in**
839 ***Shigella* as compared to *E. coli***. The orthologous *E. coli* deletion strains of *tufA* and *tufB*
840 exhibit robust growth (OD600 of 0.72 and 0.78 after 22 hours in LB), but are essential by our
841 criteria**.** Both genes contain insertions only at the 5' or 3' ends of the genes. Genes that are
842 essential in both *E. coli* and *Shigella* are coloured in white. Those inferred as being essential
843 in *Shigella* but for which the orthologous deletion genotypes exhibit robust growth in *E. coli*
844 are indicated in blue. Genes inferred as essential in *Shigella* and which do not exhibit robust
845 growth in *E. coli* are coloured in light blue. tRNA genes are indicated in dark grey.

846 **Fig 8. The region of the genome containing the *rfb* operon is largely uninterrupted by**
847 **transposon insertions.** *rfbI*, *rfc*, *rfbG*, and *rfbF* are completely uninterrupted by transposon
848 insertions; *rfbE* is uninterrupted over 90% of its length. None of these genes have orthologous
849 counterparts in *E. coli* K12 due to a lateral transfer event that occurred at this locus (see main
850 text). This operon encodes genes active in O-antigen biosynthesis.

851 **S1 Table**. Full table of gene characteristics and orthologue relationships used in the analyses.

852 **S2 Table.** List of chromosomal SNPs and indels observed in the *Shigella* strain used here that
853 differ from the GenBank sequence NC004741.

854 **S3 Table.** List of plasmid SNPs and indels observed in the *Shigella* strain used here that are
855 different from the GenBank sequence of the *Shigella flexneri* 2a strain 301 virulence plasmid
856 pCP301 (NC_004851).

857 **S4 Table**. Table listing tRNA genes and the number of insertions in each.

858 **S1 Fig. Distribution of transposon insertions across the genome.** We observed little bias
859 on the chromosomal level of insert locations.

860 **S2 Fig. ROC curves showing the predictive power of various features.** To select a feature
861 that was the best predictor of essentiality in *E. coli* orthologues, used only ORFs that we had
862 data on essentiality from both the Keio and PEC studies. We then selected transposon
863 insertion patterns that most closely match the essentiality delineations in theses studies.
864 Specifically, we selected the feature that maximized the number of true positive "essential"
865 genes (maximizing the sensitivity) while minimizing the number of FP (maximizing
866 specificity). We selected from eleven (non-independent) features shown here: (1) the total
867 number of insertions; (2) the mean number of bp between insertions; (3) the median number
868 of bp between insertions; (4) the number of bp in the 5' end preceding the first insertion; (5)
869 the number of bp in the 5' end preceding the first insertion relative to the total bp in the gene;

bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(6) the number of bp in the 5' end preceding the second insertion; (7) the number of bp in the 5' end preceding the second insertion relative to the total bp in the gene; (8) the number of bp in the longest uninterrupted stretch of the gene; (9) the number of bp in the longest uninterrupted stretch of the gene relative to the total length of the gene, and (10) the number of bp in the longest stretch of the gene interrupted by at most one insertion; (11) the number of bp in the longest stretch of the gene interrupted by at most one insertion, relative to the total length of the gene. See the **Methods** section for more details of this analysis.

S3 Fig. **Analogous plots to that shown in Fig. 3, for growth in minimal glucose MOPS media after (A) 24 and (B) 48 hours.** In both cases, we find that the shift is less pronounced than that observed for LB.

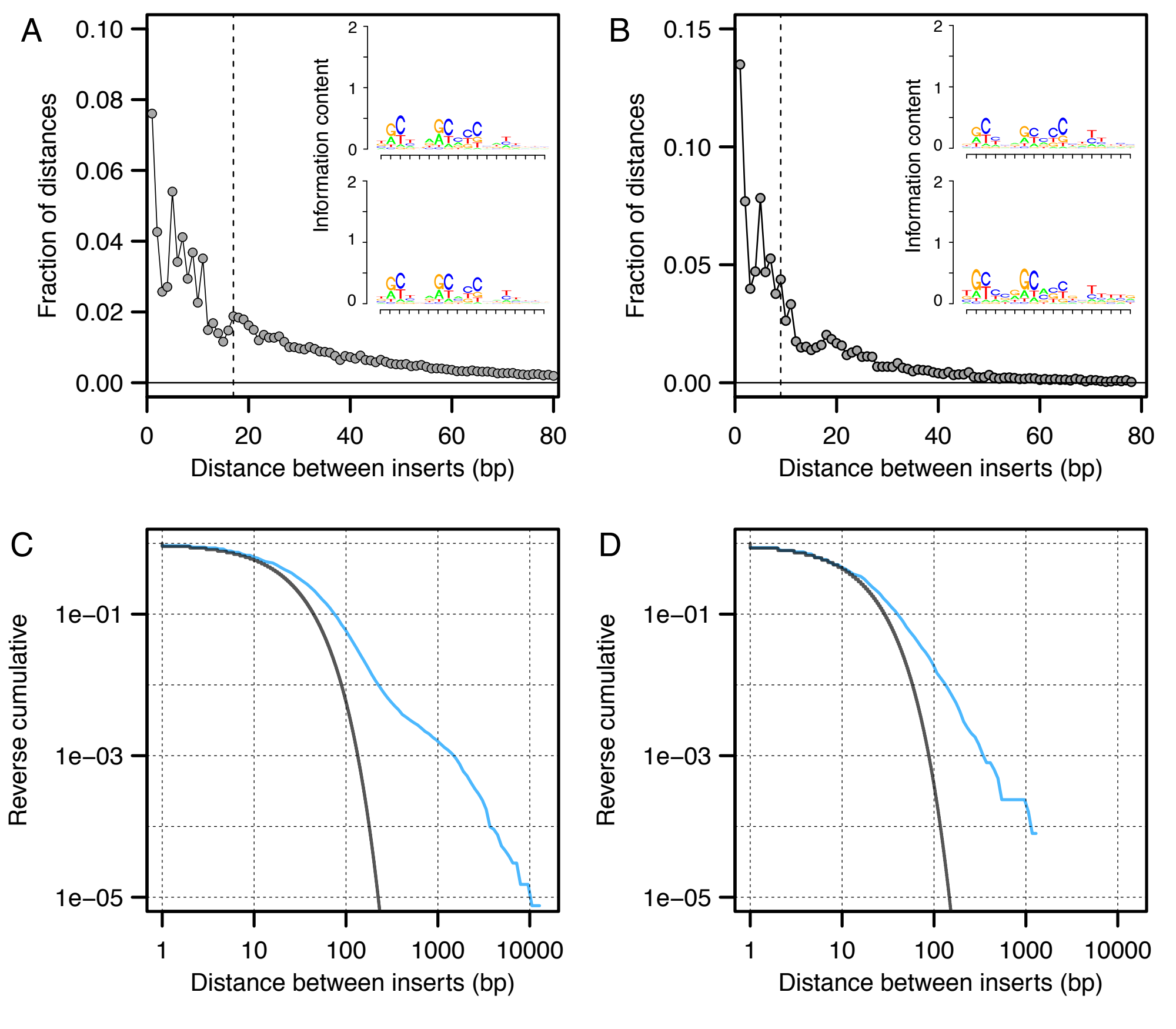
S4 Fig. **Schematic of the primer positions used for Illumina sequencing of transposon insertions**.

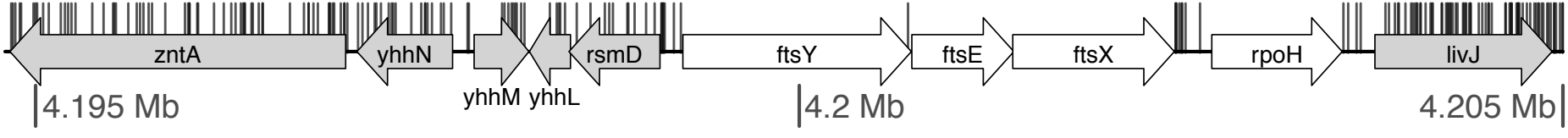
S5 Fig. Inferred fragment lengths of perfectly mapped reads across several genomic regions. For each plot, the inferred fragment lengths are arranged by increasing length (ranked on the y-axis). Thus, very long fragments are present at the top of the y-axis. Most fragments have lengths between 100bp and 400bp; a small number have lengths over 1000bp or more. It is very likely that these are not the true insert sizes, but appear that way because of large scale deletions in our *Shigella* clone compared to the clone present in the NCBI genome database; see **Methods** for more details. (A) A region of the chromosome in which a complicated series of rearrangements has occurred, leading to paired end reads perfectly matching to different locations in this region. 45 mapped read pairs span more than 1.5 Kbp, a size that is not concordant with the majority of insert sizes. (B) A genomic region where an approximately 10Kbp deletion occurred, removing a region containing the *yeaKLMNOP* operon. 92 mapped read pairs span more than 8.5Kbp. This region is flanked by two IS elements. (C) A region where an approximately 4Kbp deletion occurred, removing two genes with no *E. coli* K12 orthologues. 68 mapped read pairs span more than 4 Kbp, and again this region is flanked by two IS elements. (D) A genomic region where an approximately 2Kbp deletion occurred, removing *yhdW*. 244 mapped read pairs span more than 2Kbp, and the region is flanked by two IS elements. (E) A deletion in the region of the chromosome containing *S4145 (yiaN)*. 232 mapped read pairs spanned more than 1.8 Kbp, and this region is also flanked by two IS elements. (F) A region of the chromosome containing the *rfb* operon. Most of the genes within this operon are uninterrupted by transposons. However, we find no evidence that this is due to a deletion of this region in our *Shigella* clone, as we find no reads mapping across the region; a small number of reads map within the region; and the closest IS elements are 15 Kb upstream of *rfbJ* and 20 Kb downstream of *rfbA*. The genes in this operon have no orthologues in *E. coli* K12.

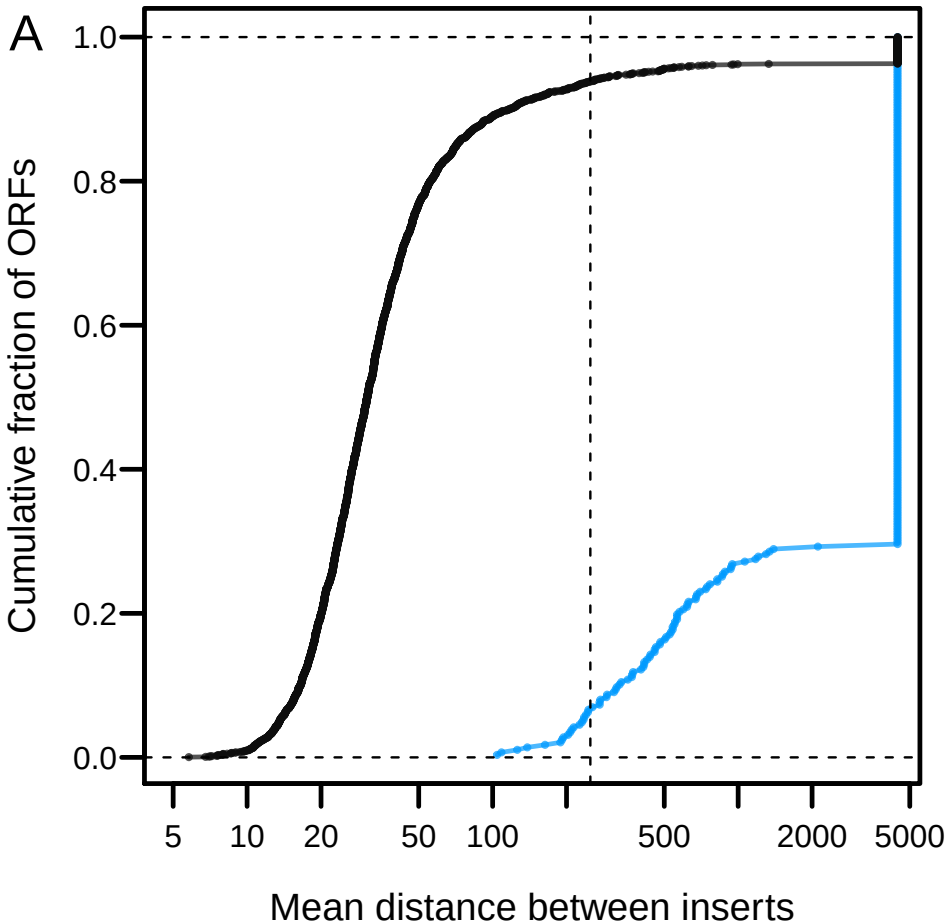
bioRxiv preprint doi: https://doi.org/10.1101/038869; this version posted February 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

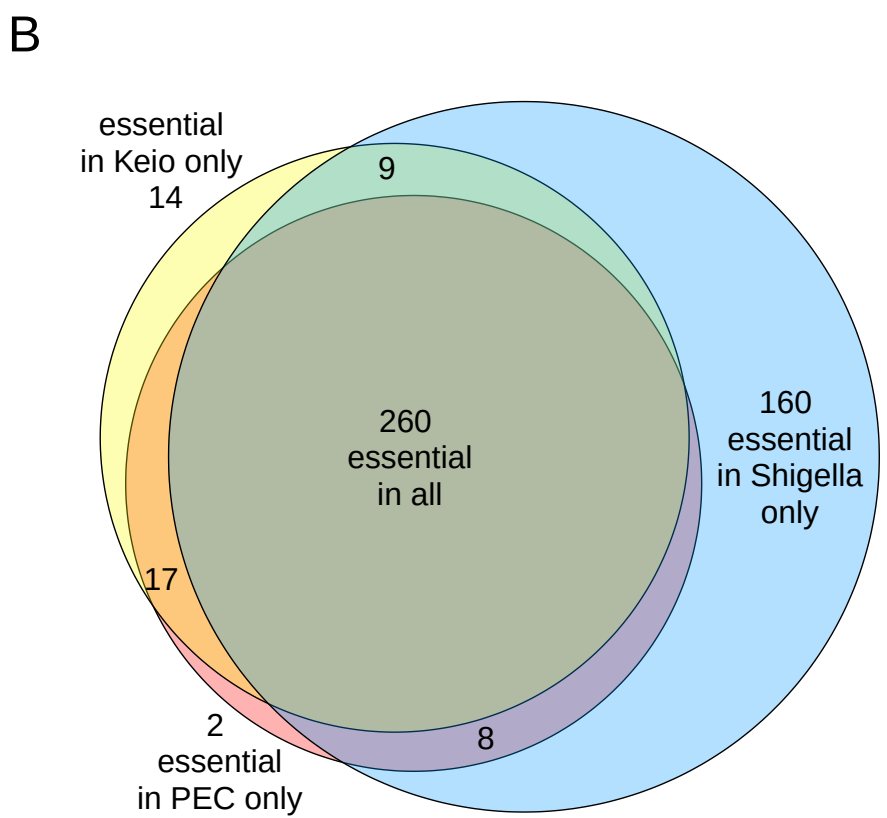
S6 Fig. **Functional categories of non-essential *E. coli* genes that are enriched (or depleted) for essential *Shigella* genes.** We classified genes according to function using the MultiFun classification system [63]. For this analysis we considered only genes that are non-essential in *E. coli*. We find that genes uniquely essential in *Shigella* are enriched in some functional categories. For example, of the 27 ribosomal proteins identified as non-essential in *E. coli*, we identify approximately 45% as being essential in *Shigella.* In contrast, of the 565 membrane proteins identified as non-essential in *E. coli*, we find that less than 10% are essential in *Shigella.* Thus, ORFs uniquely essential in *Shigella* are far more likely to function in the ribosome than one would expect. The number of non-essential *E. coli* genes is indicated above each bar; the probability of finding the level of enrichment (or depletion) that we observe in each secondary category is indicated for cases in which this probability is less than 0.025, using a Fisher exact test. (A) All non-essential genes in *E. coli*. (B) An identical analysis excluding all non-essential genes in *E. coli* that exhibit very low growth yields (OD600 less than 0.5 after 22 hours of growth LB). In both cases, the only subcategory notably enriched for essential genes is that containing ribosomal proteins. The only categories appreciably depleted for genes with essential function are genes with function in RNA processing and to some extent, energy production.

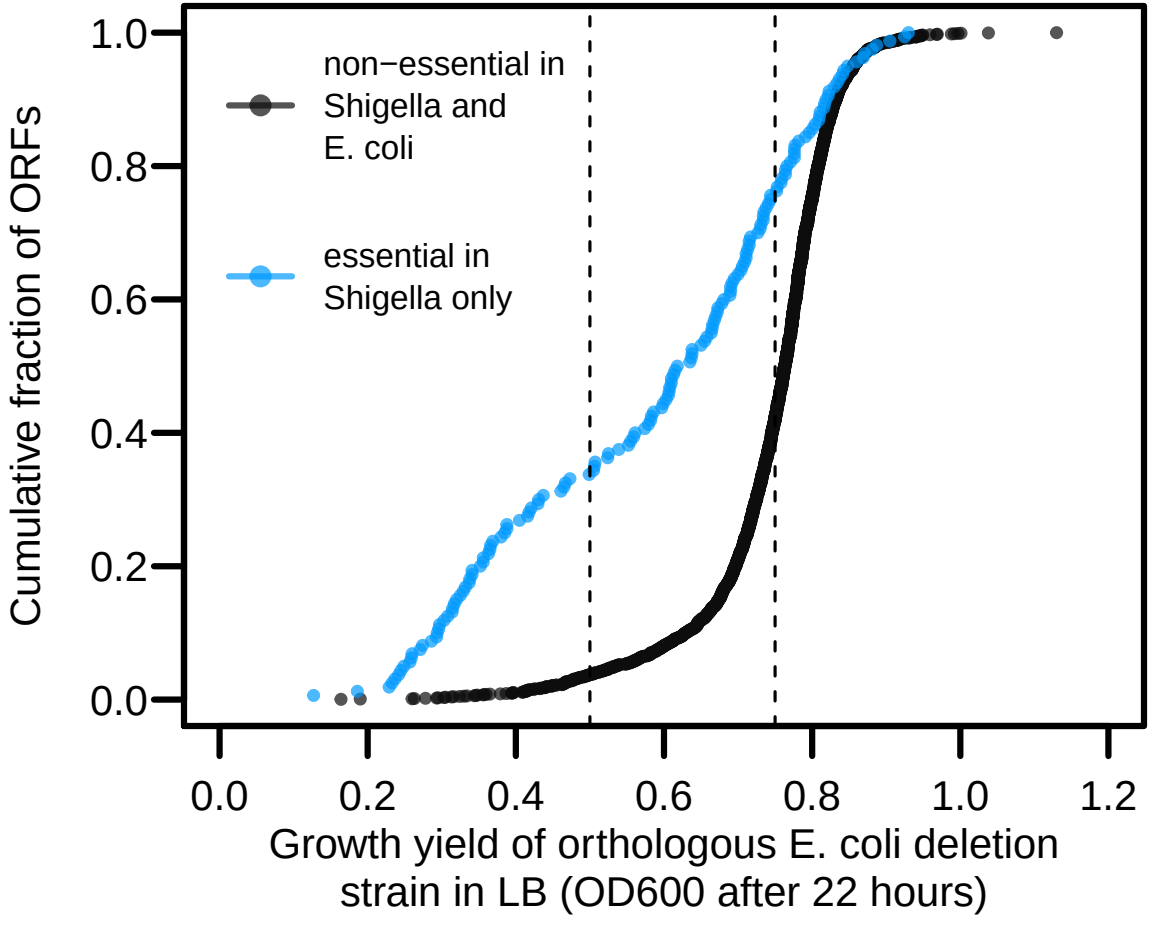
S7 Fig. Transposon insertion locations across the entire *Shigella* chromosome. Each insertion site is indicated by a vertical red line. ORFs are indicated in light green; rRNAs in dark green; and tRNAs in dark blue. The annotation is taken from the GenBank sequence of *Shigella flexneri* 2a 2457T.

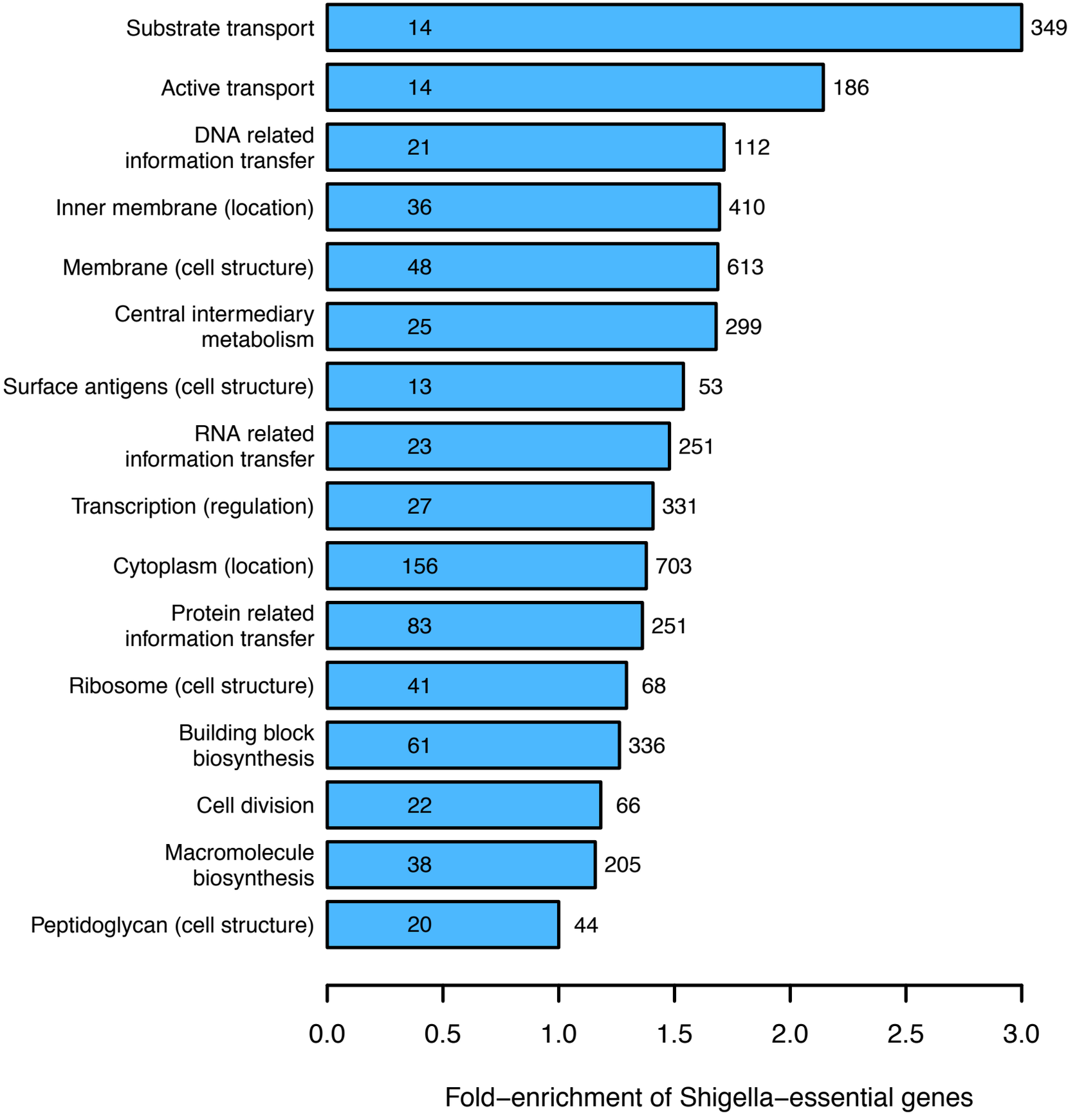

A
B
C
D
Fraction of distances
Information content
Distance between inserts (bp)
Reverse cumulative
0.10
0.08
0.06
0.04
0.02
0.00
0.15
0.05
0
20
40
60
80
1e−01
1e−03
1e−05
1
10
100
1000
10000

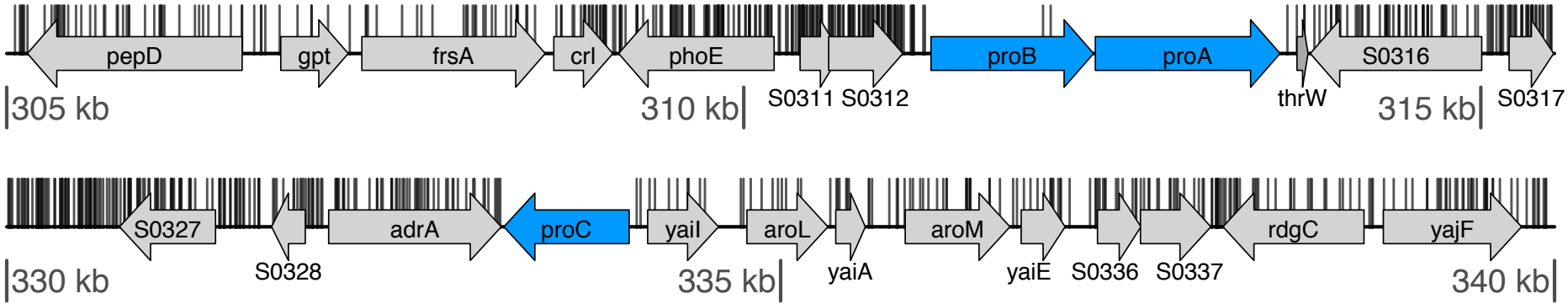

zntA
yhhN
yhhM
yhhL
rsmD
ftsY
ftsE
ftsX
rpoH
livJ
4.195 Mb
4.2 Mb
4.205 Mb

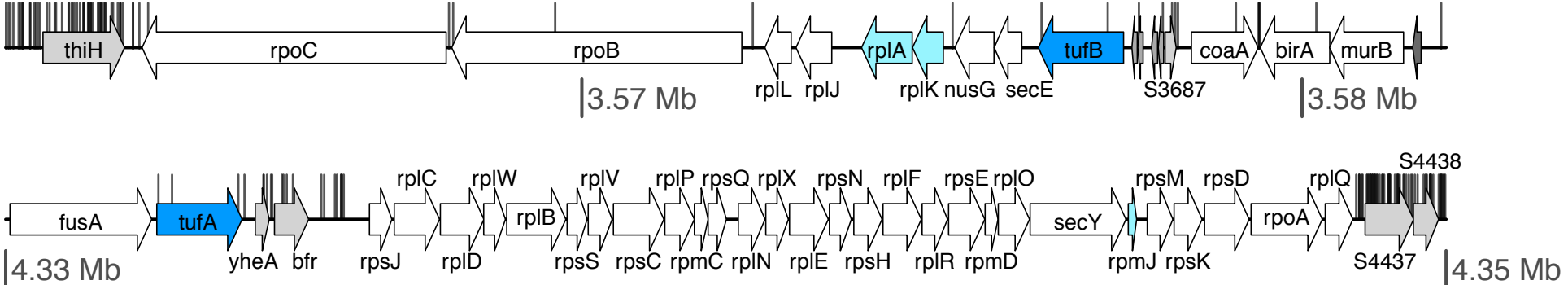

A
1.0
0.8
0.6
0.4
0.2
0.0
Cumulative fraction of ORFs
5
10
20
50
100
500
2000
5000
Mean distance between inserts

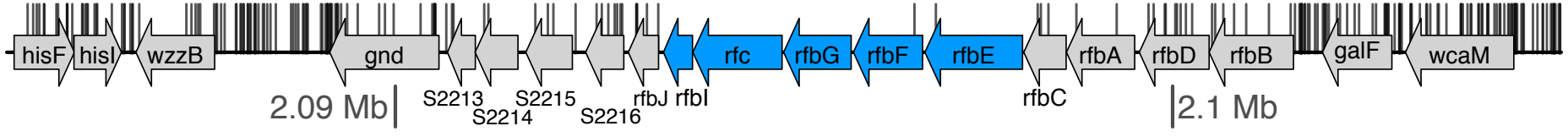

B
essential
in Keio only
14
9
260
essential
in all
160
essential
in Shigella
only
17
2
essential
in PEC only
8

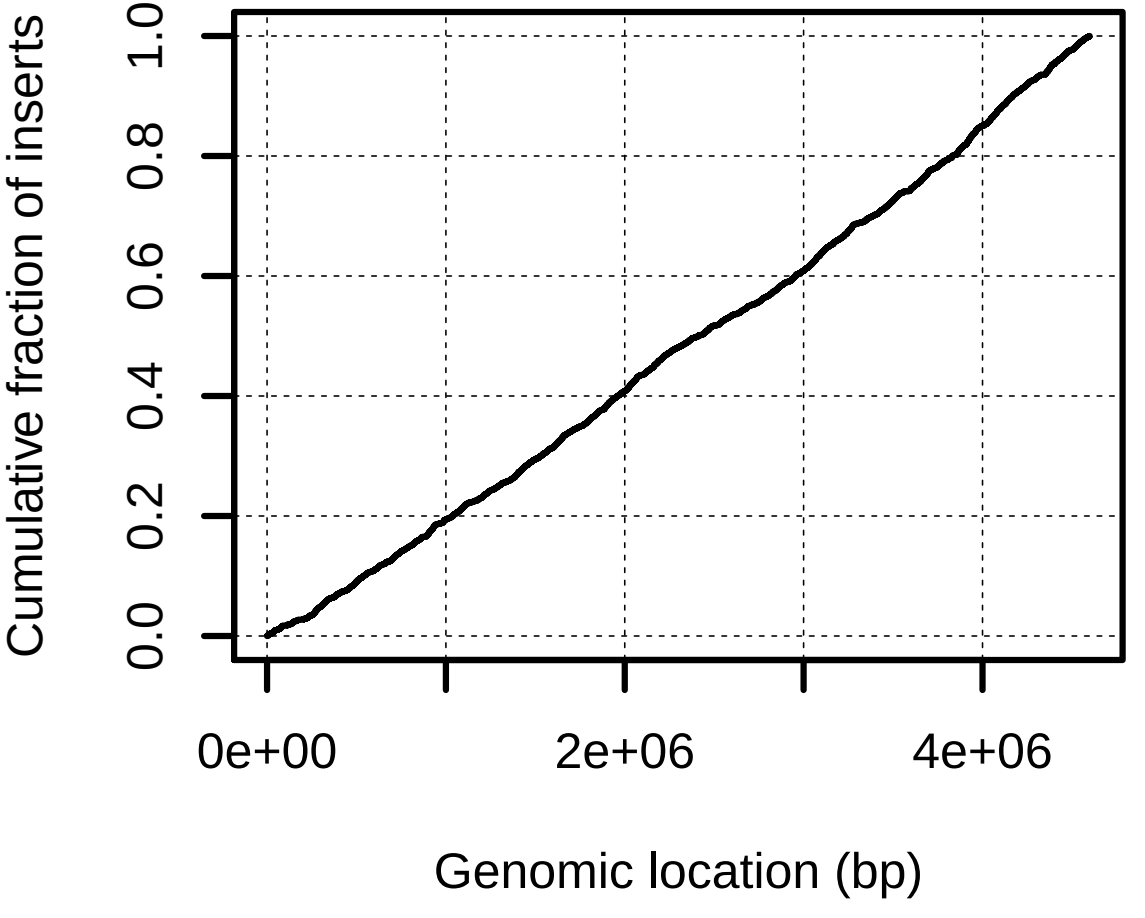

Cumulative fraction of ORFs
1.0
0.8
0.6
0.4
0.2
0.0
non−essential in Shigella and E. coli
essential in Shigella only
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Growth yield of orthologous E. coli deletion strain in LB (OD600 after 22 hours)

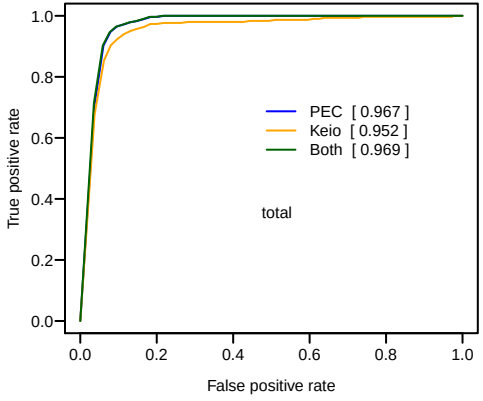

Substrate transport
14
349
Active transport
14
186
DNA related information transfer
21
112
Inner membrane (location)
36
410
Membrane (cell structure)
48
613
Central intermediary metabolism
25
299
Surface antigens (cell structure)
13
53
RNA related information transfer
23
251
Transcription (regulation)
27
331
Cytoplasm (location)
156
703
Protein related information transfer
83
251
Ribosome (cell structure)
41
68
Building block biosynthesis
61
336
Cell division
22
66
Macromolecule biosynthesis
38
205
Peptidoglycan (cell structure)
20
44
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Fold−enrichment of Shigella−essential genes

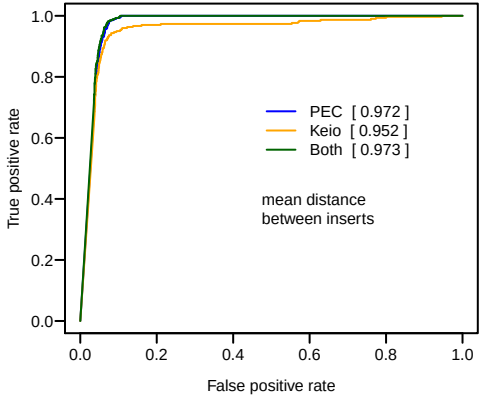

pepD
gpt
frsA
crl
phoE
S0311
S0312
proB
proA
thrW
S0316
S0317
305 kb
310 kb
315 kb
S0327
S0328
adrA
proC
yaiI
aroL
yaiA
aroM
yaiE
S0336
S0337
rdgC
yajF
330 kb
335 kb
340 kb

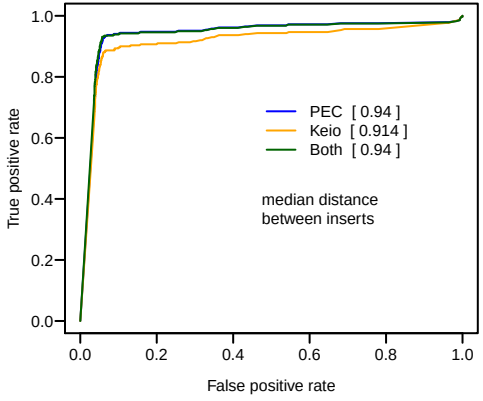

thiH
rpoC
rpoB
3.57 Mb
rplL
rplJ
rplA
rplK
nusG
secE
tufB
S3687
coaA
birA
murB
3.58 Mb
fusA
tufA
yheA
bfr
rpsJ
rplC
rplD
rplW
rplB
rplV
rpsS
rplP
rpsC
rpsQ
rpmC
rplX
rplN
rpsN
rplE
rplF
rpsH
rplR
rpsE
rpmD
rplO
secY
rpmJ
rpsM
rpsK
rpsD
rpoA
rplQ
S4438
S4437
4.33 Mb
4.35 Mb

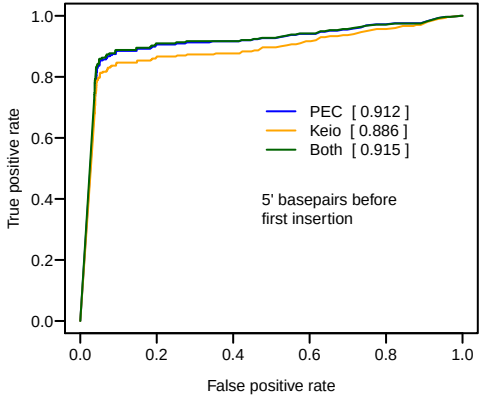

hisF
hisI
wzzB
gnd
S2213
S2214
S2215
S2216
rfbJ
rfbI
rfc
rfbG
rfbF
rfbE
rfbC
rfbA
rfbD
rfbB
galF
wcaM
2.09 Mb
2.1 Mb

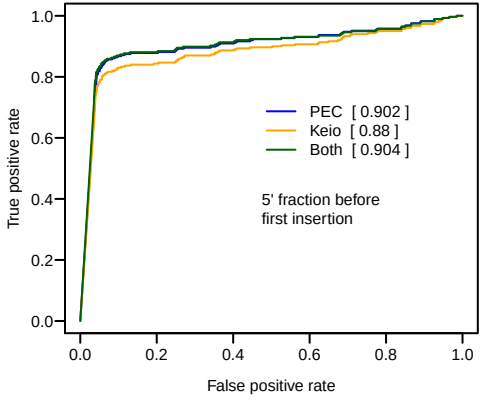

Cumulative fraction of inserts
0.0
0.2
0.4
0.6
0.8
1.0
0e+00
2e+06
4e+06
Genomic location (bp)

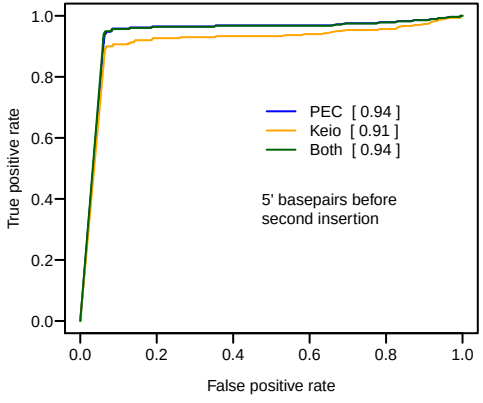

True positive rate
False positive rate
PEC [0.967]
Keio [0.952]
Both [0.969]
total

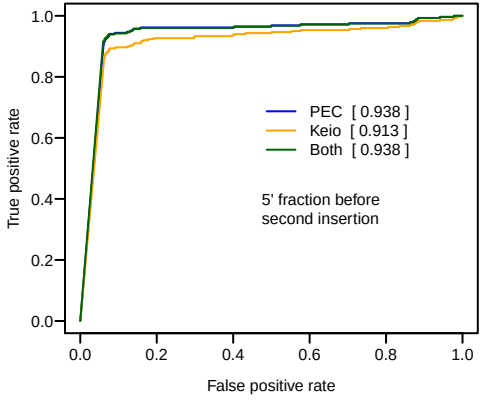

True positive rate
False positive rate
PEC [0.972]
Keio [0.952]
Both [0.973]
mean distance
between inserts

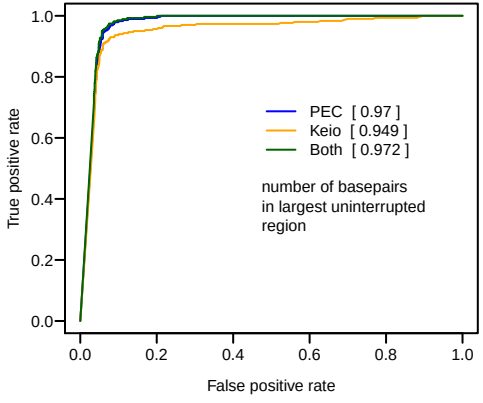

True positive rate
False positive rate
PEC [0.94]
Keio [0.914]
Both [0.94]
median distance
between inserts

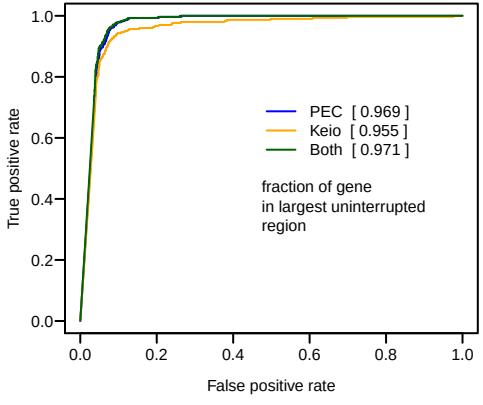

True positive rate
False positive rate
PEC [0.912]
Keio [0.886]
Both [0.915]
5' basepairs before
first insertion

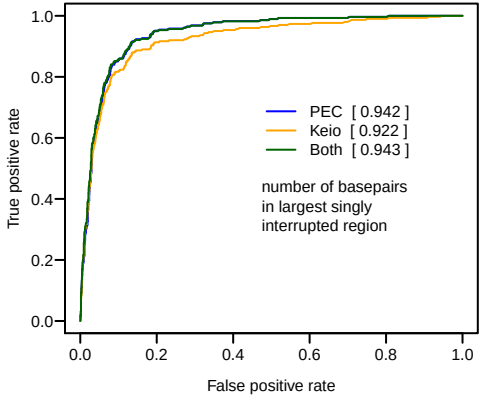

True positive rate
False positive rate
PEC [0.902]
Keio [0.88]
Both [0.904]
5' fraction before
first insertion

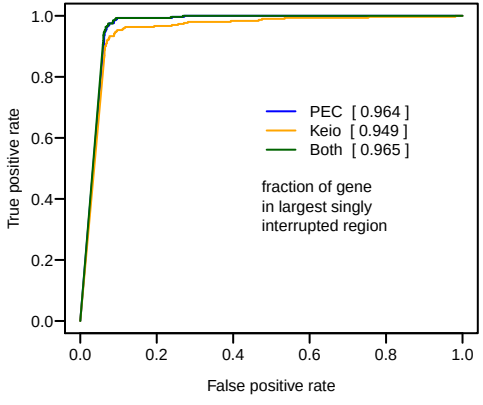

True positive rate
False positive rate
PEC [0.94]
Keio [0.91]
Both [0.94]
5' basepairs before
second insertion

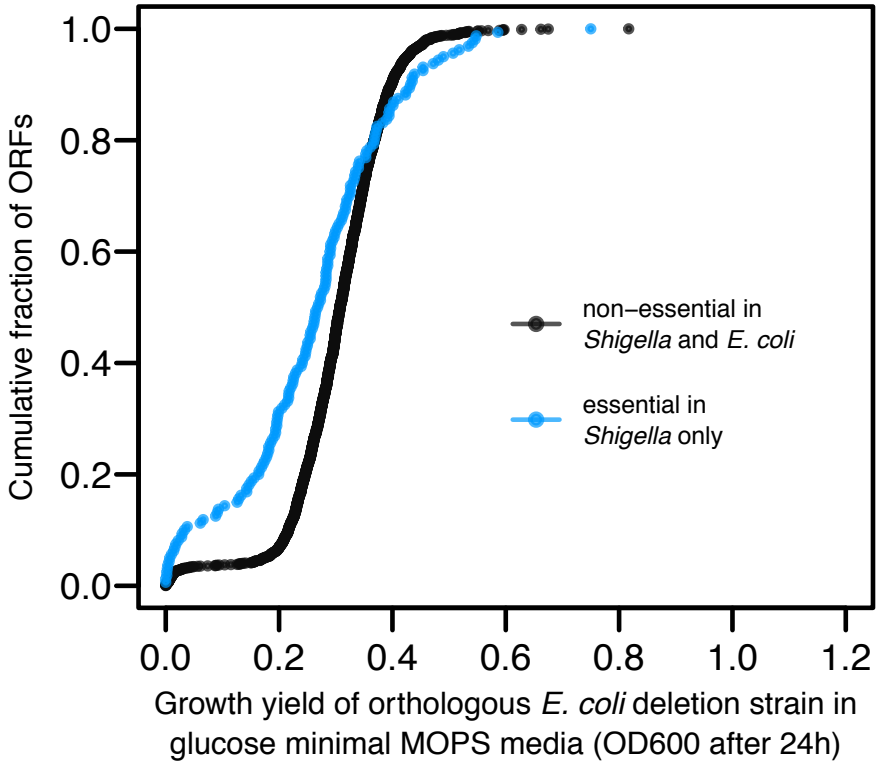

True positive rate
False positive rate
PEC [0.938]
Keio [0.913]
Both [0.938]
5' fraction before
second insertion

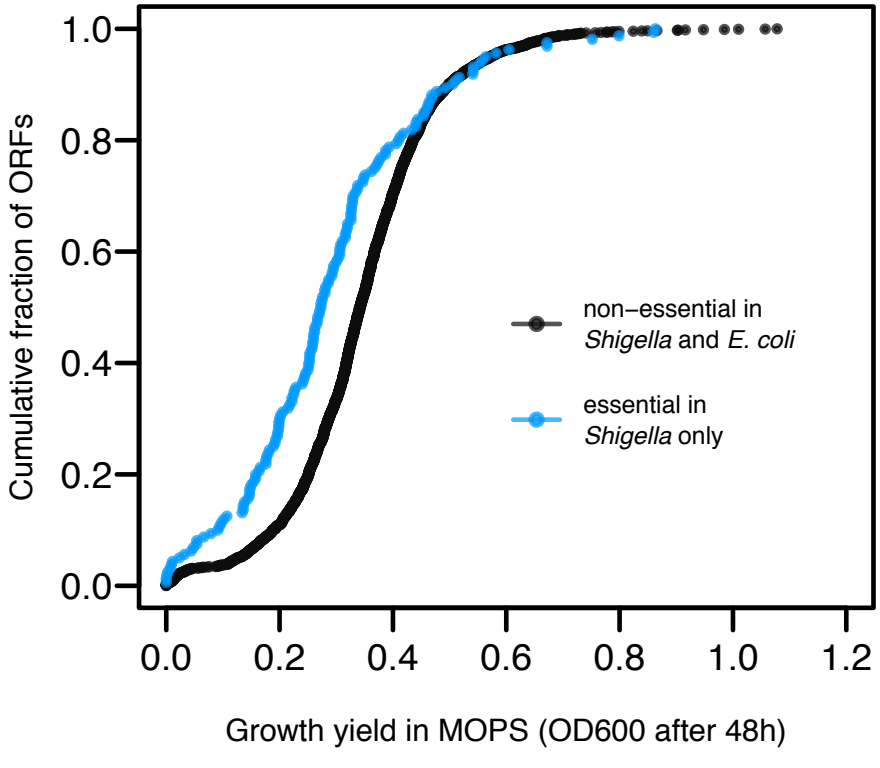

True positive rate
False positive rate
PEC [0.97]
Keio [0.949]
Both [0.972]
number of basepairs
in largest uninterrupted
region

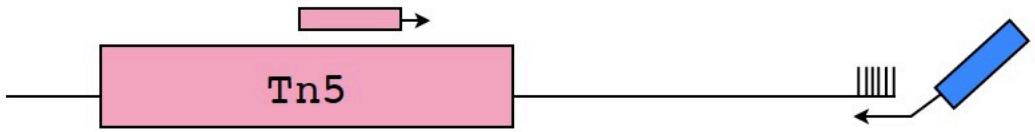

True positive rate
False positive rate
PEC [0.969]
Keio [0.955]
Both [0.971]
fraction of gene
in largest uninterrupted
region

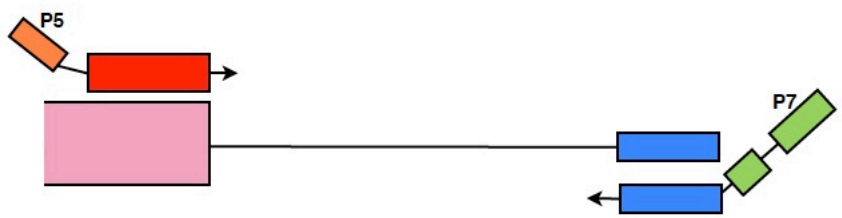

True positive rate
False positive rate
PEC [0.942]
Keio [0.922]
Both [0.943]
number of basepairs
in largest singly
interrupted region

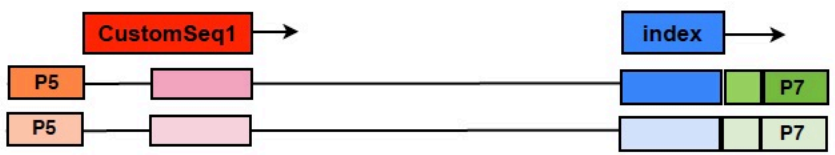

True positive rate
False positive rate
PEC [0.964]
Keio [0.949]
Both [0.965]
fraction of gene
in largest singly
interrupted region


Cumulative fraction of ORFs
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
non−essential in
Shigella and E. coli
essential in
Shigella only
Growth yield of orthologous E. coli deletion strain in
glucose minimal MOPS media (OD600 after 24h)


Cumulative fraction of ORFs
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
non−essential in
Shigella and E. coli
essential in
Shigella only
Growth yield in MOPS (OD600 after 48h)







double stranded PCR product,
gel purify band/smear at (300-400bp)
(insert at 200-300bp)

1: Arbitrary PCR

using phusion PCR enzyme, so that no A's added at 3' end

```
1_transp_deep_seq_F1: TCTATCGCCTTCTTGACGAG
```

```
1_transp_deep_seq_R2: GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT NNNNNNNNNNggtgc
1_transp_deep_seq_R3: GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT NNNNNNNNNNgatat
1_transp_deep_seq_R4: GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT NNNNNNNNNNagtac
```

2: Nested PCR (addition of P7 and P5)

using phusion PCR enzyme, so that no A's added at 3' end

```
2_transp_deep_seq_ FP5: AATGATACGGCGACCACCGAGATCT actggtcggcg CATTAGGGGATTCATCAG
```

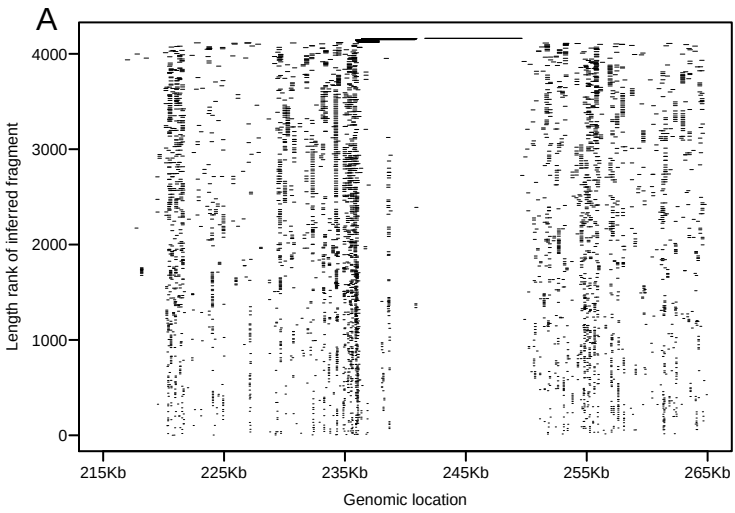
```
Rev_P7 6_02
CAAGCAGAAGACGGCATACGAGAT agtctt GTGACTGGAGTTCAGACGTGT
CAAGCAGAAGACGGCATACGAGAT gtatct GTGACTGGAGTTCAGACGTGT
CAAGCAGAAGACGGCATACGAGAT tcatgg GTGACTGGAGTTCAGACGTGT
CAAGCAGAAGACGGCATACGAGAT cgcgac GTGACTGGAGTTCAGACGTGT
CAAGCAGAAGACGGCATACGAGAT acgata GTGACTGGAGTTCAGACGTGT
CAAGCAGAAGACGGCATACGAGAT tataga GTGACTGGAGTTCAGACGTGT
```

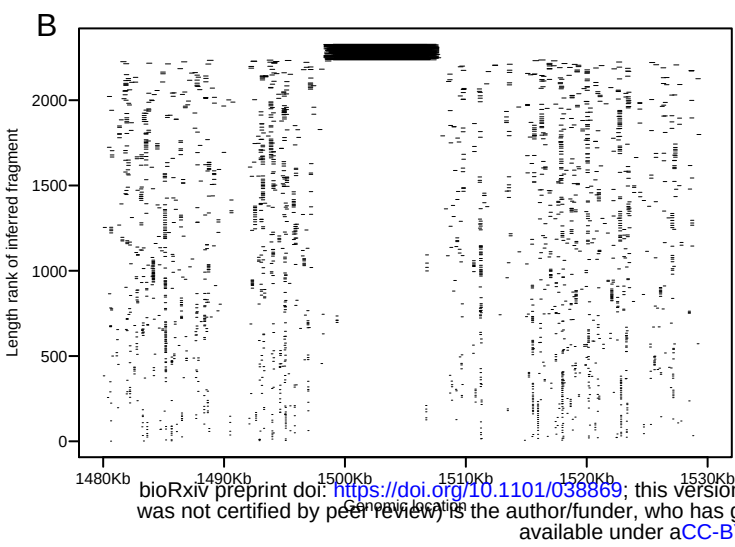
3: Sequencing on Illlumina chip

```
transp_deep_seq_Custom seq 1: actggtcggcgCATTAGGGGATTCATCAG
```

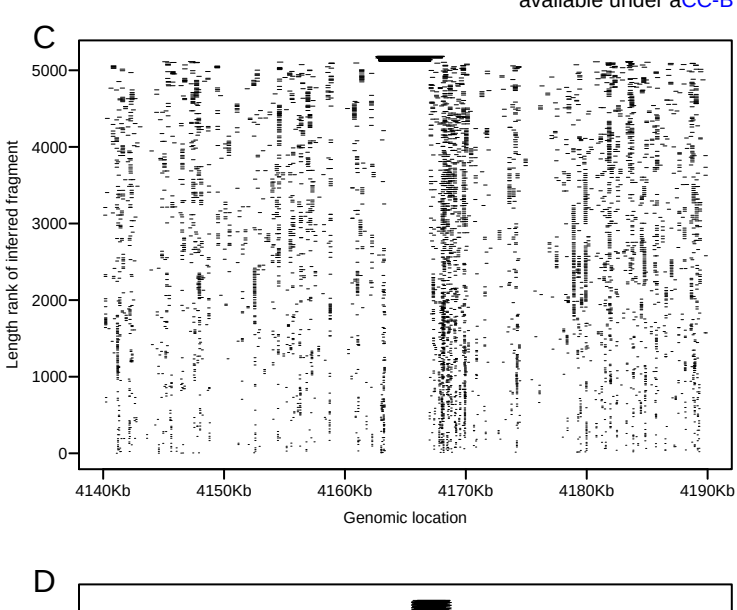
4: Barcode is read on Illlumina chip

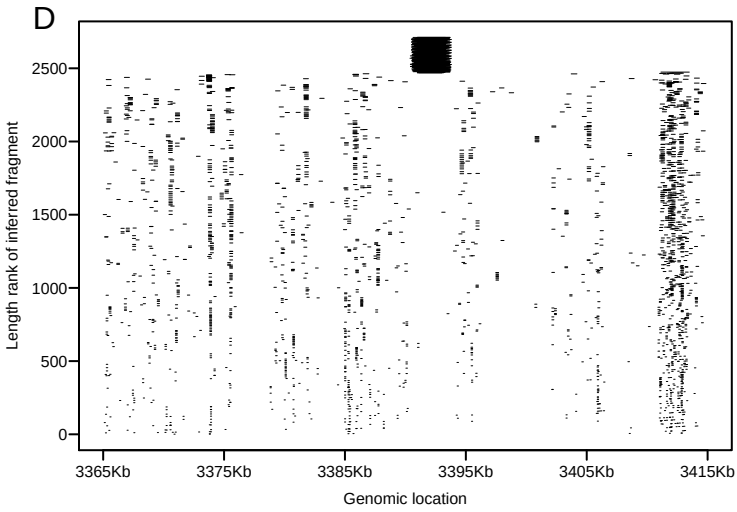
```
transp_deep_seq_Index seq: GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
```

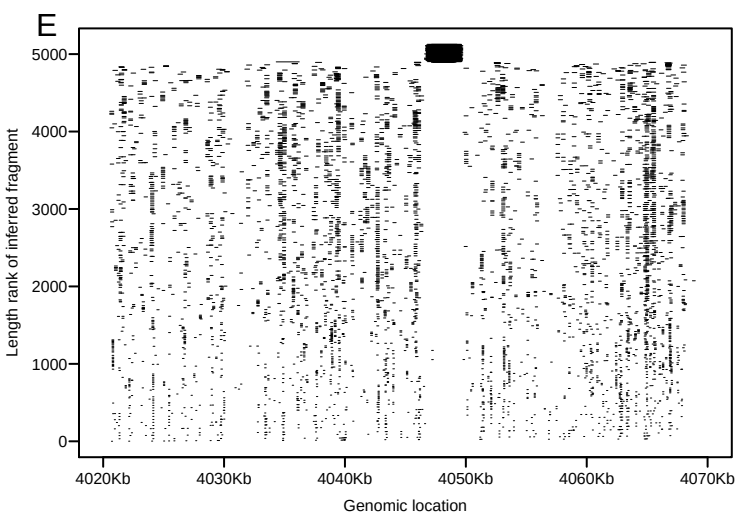


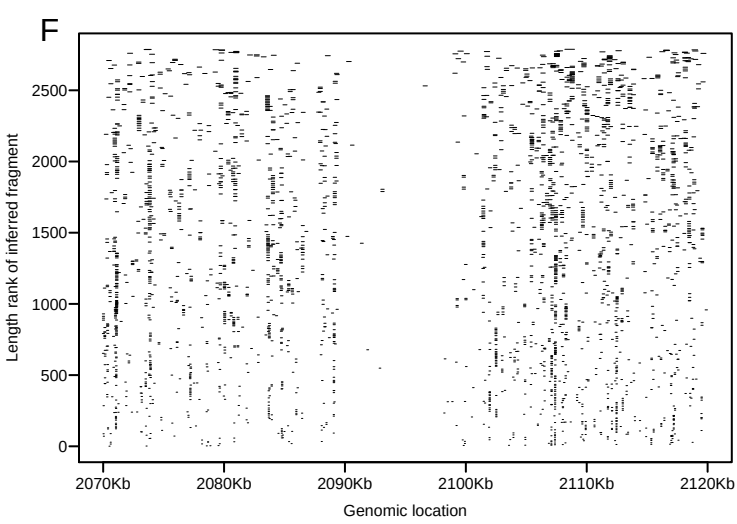


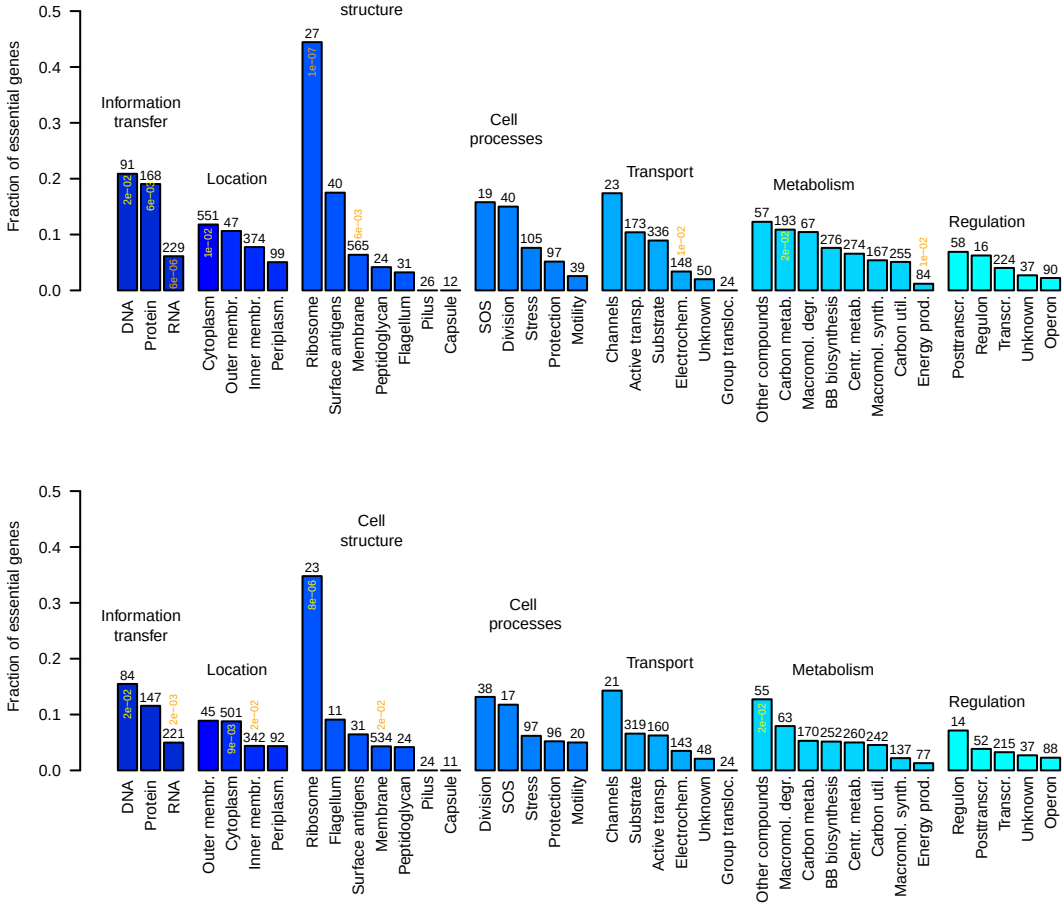
bioRxiv preprint doi: https://doi.org/10.1101/038869; this version
was not certified by peer review) is the author/funder, who has gr
available under aCC-BY-



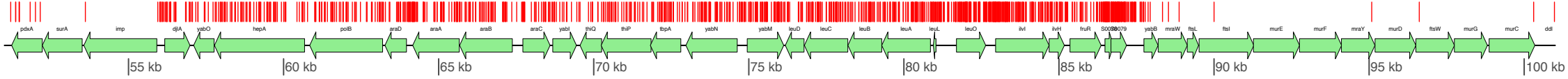






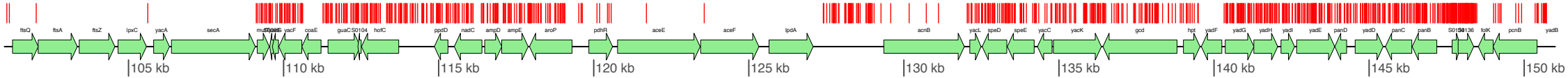

Fraction of essential genes
0.0
0.1
0.2
0.3
0.4
0.5
Information transfer
Location
Cell structure
Cell processes
Transport
Metabolism
Regulation
DNA
Protein
RNA
Cytoplasm
Outer membr.
Inner membr.
Periplasm.
Ribosome
Surface antigens
Membrane
Peptidoglycan
Flagellum
Pilus
Capsule
SOS
Division
Stress
Protection
Motility
Channels
Active transp.
Substrate
Electrochem.
Unknown
Group transloc.
Other compounds
Carbon metab.
Macromol. degr.
BB biosynthesis
Centr. metab.
Macromol. synth.
Carbon util.
Energy prod.
Posttranscr.
Regulon
Transcr.
Unknown
Operon
91
168
229
551
47
374
99
27
40
565
24
31
26
12
19
40
105
97
39
23
173
336
148
50
24
57
193
67
276
274
167
255
84
58
16
224
37
90
84
147
221
45
501
342
92
23
11
31
534
24
24
11
38
17
97
96
20
21
319
160
143
48
24
55
63
170
252
260
242
137
77
14
52
215
37
88

51 – 101 Kb

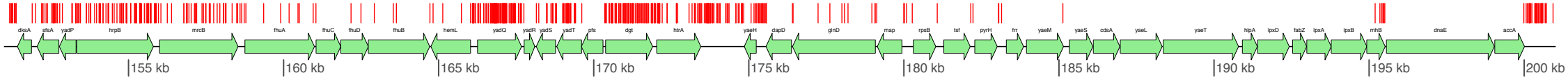

pdxA
surA
imp
djlA
yabO
hepA
polB
araD
araA
araB
araC
yabI
thiQ
thiP
tbpA
yabN
yabM
leuD
leuC
leuB
leuA
leuL
leuO
ilvI
ilvH
fruR
S0078
S0079
yabB
mraW
ftsL
ftsI
murE
murF
mraY
murD
ftsW
murG
murC
ddl
55 kb
60 kb
65 kb
70 kb
75 kb
80 kb
85 kb
90 kb
95 kb
100 kb

5 kb

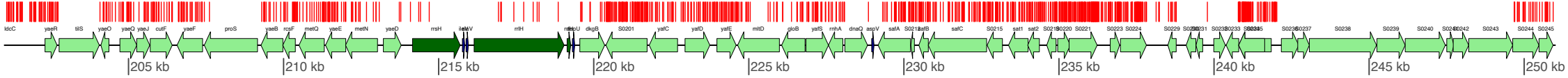
101 – 151 Kb


ftsQ
ftsA
ftsZ
lpxC
yacA
secA
yacF
coaE
guaC
S0104
hofC
ppdD
nadC
ampD
ampE
aroP
pdhR
aceE
aceF
lpdA
acnB
yacL
speD
speE
yacC
yacK
gcd
hpt
yadF
yadG
yadH
yadI
yadE
panD
yadD
panC
panB
S0136
folK
pcnB
yadB
105 kb
110 kb
115 kb
120 kb
125 kb
130 kb
135 kb
140 kb
145 kb
150 kb
5 kb

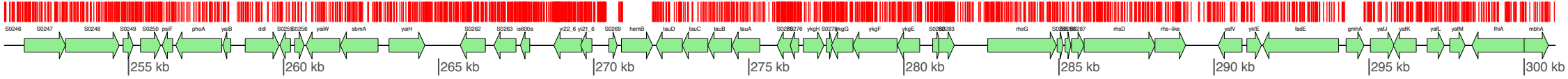
151 – 201 Kb


dksA
sfsA
yadP
hrpB
mrcB
fhuA
fhuC
fhuD
fhuB
hemL
yadQ
yadR
yadS
yadT
pfs
dgt
htrA
yaeH
dapD
glnD
map
rpsB
tsf
pyrH
frr
yaeM
yaeS
cdsA
yaeL
yaeT
hlpA
lpxD
fabZ
lpxA
lpxB
rnhB
dnaE
accA
155 kb
160 kb
165 kb
170 kb
175 kb
180 kb
185 kb
190 kb
195 kb
200 kb
5 kb

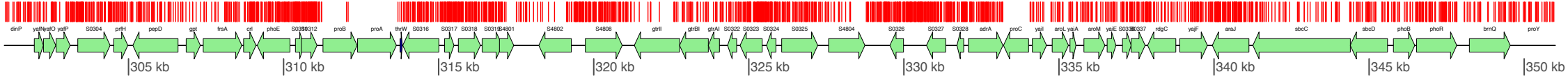
201 – 251 Kb


ldcC
yaeR
tilS
yaeO
yaeQ
yaeJ
cutF
yaeF
proS
yaeB
rcsF
metQ
yaeE
metN
yaeD
rrsH
rrlH
dkgB
S0201
yafC
yafD
yafE
mltD
gloB
yafS
rnhA
dnaQ
aspV
safA
S0218
safB
safC
S0215
sat1
sat2
S0219
S0220
S0221
S0223
S0224
S0229
S0231
S0233
S0237
S0238
S0239
S0240
S0242
S0243
S0244
S0245
205 kb
210 kb
215 kb
220 kb
225 kb
230 kb
235 kb
240 kb
245 kb
250 kb
5 kb

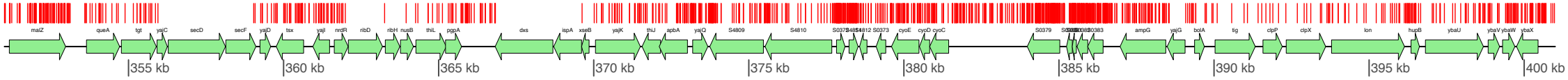
251 – 301 Kb


S0246
S0247
S0248
S0249
S0250
psiF
phoA
yaiB
ddl
S0255
S0256
yaiW
sbmA
yaiH
S0262
S0263
is600a
yi22_6
yi21_6
S0269
hemB
tauD
tauC
tauB
tauA
S0276
ykgH
S0278
ykgG
ykgF
ykgE
S0283
rhsG
S0287
rhsD
rhs–like
yafV
yafE
fadE
gmhA
yafJ
yafK
yafL
yafM
fhiA
mbhA
255 kb
260 kb
265 kb
270 kb
275 kb
280 kb
285 kb
290 kb
295 kb
300 kb
5 kb

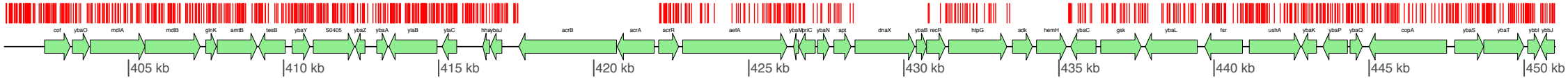
301 – 351 Kb


dinP
yafN
yafO
yafP
S0304
prfH
pepD
gpt
frsA
crl
phoE
S0312
proB
proA
thrW
S0316
S0317
S0318
S4801
S4802
S4808
gtrII
gtrBI
gtrAI
S0322
S0323
S0324
S0325
S4804
S0326
S0327
S0328
adrA
proC
yaiI
aroL
yaiA
aroM
yaiE
S0337
rdgC
yajF
araJ
sbcC
sbcD
phoB
phoR
brnQ
proY
305 kb
310 kb
315 kb
320 kb
325 kb
330 kb
335 kb
340 kb
345 kb
350 kb
5 kb

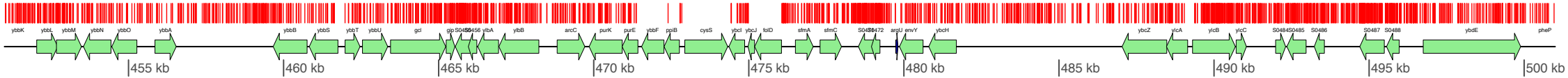
351 – 401 Kb


malZ
queA
tgt
yajC
secD
secF
yajD
tsx
yajI
nrdR
ribD
ribH
nusB
thiL
pgpA
dxs
ispA
xseB
yajK
thiJ
apbA
yajQ
S4809
S4810
S0373
cyoE
cyoD
cyoC
S0379
ampG
yajG
bolA
tig
clpP
clpX
lon
hupB
ybaU
ybaV
ybaW
ybaX
355 kb
360 kb
365 kb
370 kb
375 kb
380 kb
385 kb
390 kb
395 kb
400 kb
5 kb

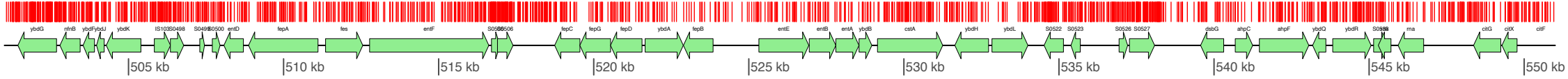
401 – 451 Kb


cof
ybaO
mdlA
mdlB
glnK
amtB
tesB
ybaY
S0405
ybaZ
ybaA
ylaB
ylaC
hhaybaJ
acrB
acrA
acrR
aefA
ybaM
yriC
ybaN
apt
dnaX
ybaB
recR
htpG
adk
hemH
ybaC
gsk
ybaL
fsr
ushA
ybaK
ybaP
ybaQ
copA
ybaS
ybaT
ybbI
ybbJ
405 kb
410 kb
415 kb
420 kb
425 kb
430 kb
435 kb
440 kb
445 kb
450 kb
5 kb

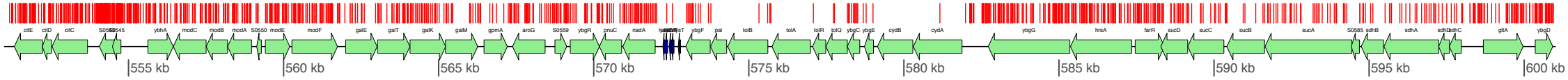
451 – 501 Kb


ybbK
ybbL
ybbM
ybbN
ybbO
ybbA
ybbB
ybbS
ybbT
ybbU
gcl
glp
S0456
ybbA
ybbB
arcC
purK
purE
ybbF
ppiB
cysS
ybcI
ybcJ
folD
sfmA
sfmC
S0472
argU
envY
ybcH
ybcZ
ylcA
ylcB
ylcC
S0484
S0485
S0486
S0487
S0488
ybdE
pheP
455 kb
460 kb
465 kb
470 kb
475 kb
480 kb
485 kb
490 kb
495 kb
500 kb
5 kb

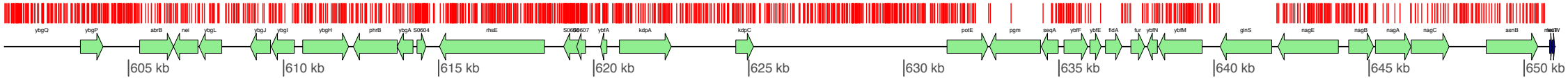
501 – 551 Kb


ybdG
nfnB
ybdF
ybdJ
ybdK
IS1203
S0498
S0499
S0500
entD
fepA
fes
entF
S0505
S0506
fepC
fepG
fepD
ybdA
fepB
entE
entB
entA
ybdB
cstA
ybdH
ybdL
S0522
S0523
S0526
S0527
dsbG
ahpC
ahpF
ybdQ
ybdR
S0544
S0545
rna
citG
citX
citF
505 kb
510 kb
515 kb
520 kb
525 kb
530 kb
535 kb
540 kb
545 kb
550 kb
5 kb

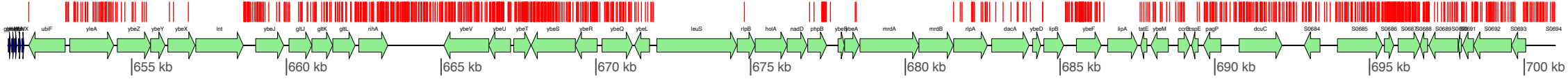
551 – 601 Kb


citE
citD
citC
S0545
ybhA
modC
modB
modA
S0550
modE
modF
galE
galT
galK
galM
gpmA
aroG
S0559
ybgR
pnuC
nadA
ybgF
pal
tolB
tolA
tolR
tolQ
ybgC
ybgE
cydB
cydA
ybgG
hrsA
farR
sucD
sucC
sucB
sucA
S0585
sdhB
sdhA
sdhD
dhC
gltA
ybgD
555 kb
560 kb
565 kb
570 kb
575 kb
580 kb
585 kb
590 kb
595 kb
600 kb
5 kb

601 – 651 Kb

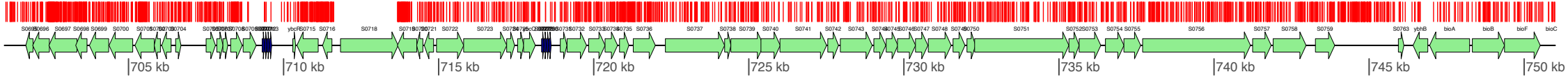

ybgQ
ybgP
abrB
nei
ybgL
ybgJ
ybgI
ybgH
phrB
ybgA
S0604
rhsE
S0606
S0607
ybfA
kdpA
kdpC
potE
pgm
segA
ybfF
ybfE
fldA
fur
ybfN
ybfM
glnS
nagE
nagB
nagA
nagC
asnB
605 kb
610 kb
615 kb
620 kb
625 kb
630 kb
635 kb
640 kb
645 kb
650 kb
5 kb

651 – 701 Kb

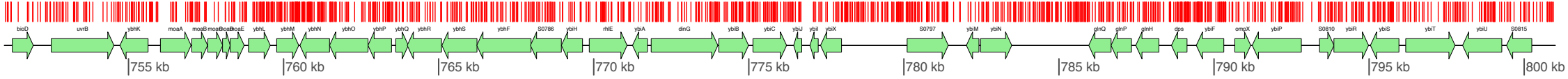


5 kb

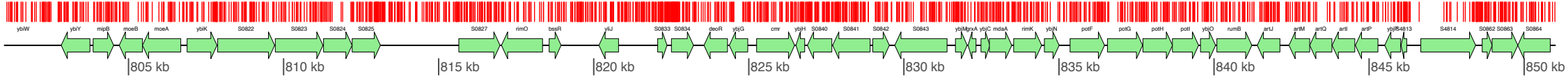
701 – 751 Kb


S0695 S0696 S0697 S0698 S0699 S0700 S0704 ybcR S0715 S0716 S0718 S0721 S0722 S0723 S0732 S0735 S0736 S0737 S0738 S0739 S0740 S0741 S0742 S0743 S0747 S0748 S0750 S0751 S0752 S0753 S0754 S0755 S0756 S0757 S0758 S0759 S0763 ybhB bioA bioB bioF bioC
705 kb
710 kb
715 kb
720 kb
725 kb
730 kb
735 kb
740 kb
745 kb
750 kb
5 kb

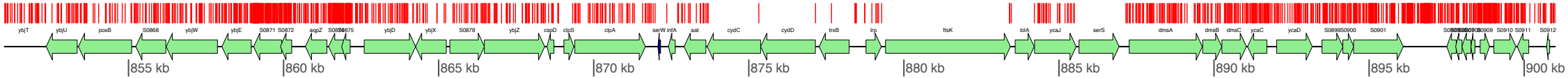
751 – 801 Kb


bioD
uvrB
ybhK
moaA
moaB
moaC
moaD
moaE
ybhL
ybhM
ybhN
ybhO
ybhP
ybhQ
ybhR
ybhS
ybhF
S0786
ybiH
rhlE
ybiA
dinG
ybiB
ybiC
ybiJ
ybiI
ybiX
S0797
ybiM
ybiN
glnQ
glnP
glnH
dps
ybiF
ompX
ybiP
S0810
ybiR
ybiS
ybiT
ybiU
S0815
755 kb
760 kb
765 kb
770 kb
775 kb
780 kb
785 kb
790 kb
795 kb
800 kb
5 kb

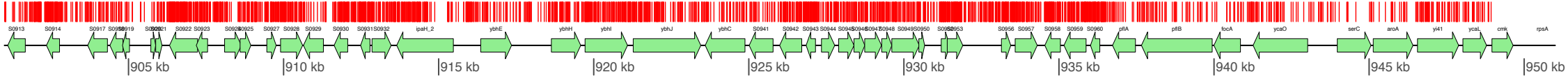
801 – 851 Kb

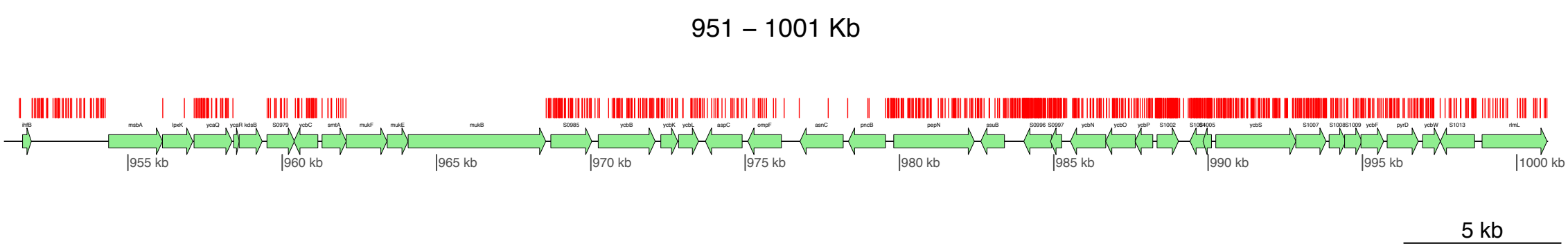

ybiW
ybiY
mipB
moeB
moeA
ybiK
S0822
S0823
S0824
S0825
S0827
rimO
bssR
yfiJ
S0833
S0834
deoR
ybjG
cmr
ybjH
S0840
S0841
S0842
S0843
ybjM
mrxA
ybjC
mdaA
rimK
ybjN
potF
potG
potH
potI
ybiO
rumB
artJ
artM
artQ
artI
artP
ybiR
S4813
S4814
S0862
S0863
S0864
805 kb
810 kb
815 kb
820 kb
825 kb
830 kb
835 kb
840 kb
845 kb
850 kb
5 kb

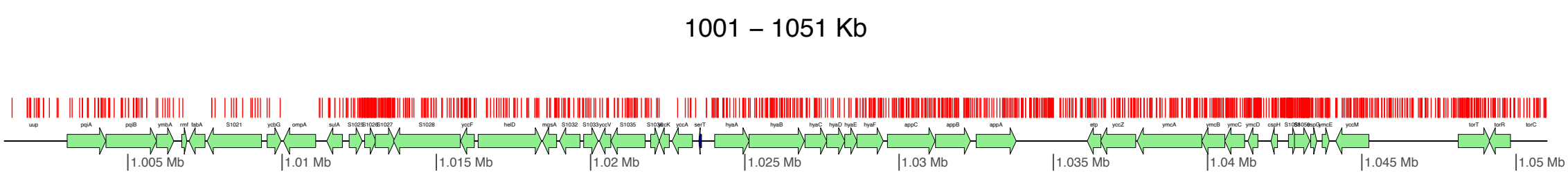
851 – 901 Kb

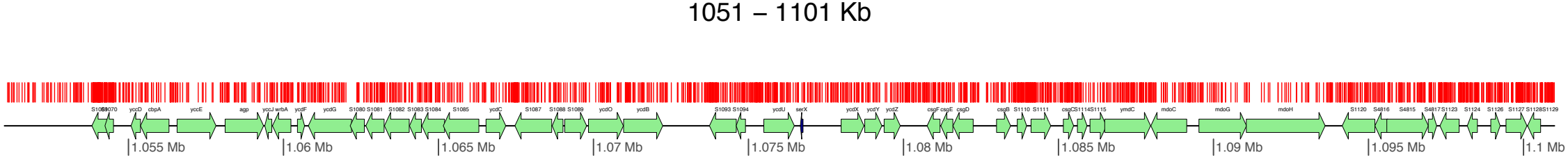

ybjT
ybjU
poxB
S0868
ybjW
ybjE
S0871
S0872
acpZ
S0875
ybjD
ybjX
S0878
ybjZ
cspD
clpS
clpA
serW
infA
aat
cydC
cydD
trxB
lrp
ftsK
lolA
ycaJ
serS
dmsA
dmsB
dmsC
ycaC
ycaD
S0899
S0900
S0901
S0909
S0910
S0911
S0912
855 kb
860 kb
865 kb
870 kb
875 kb
880 kb
885 kb
890 kb
895 kb
900 kb
5 kb

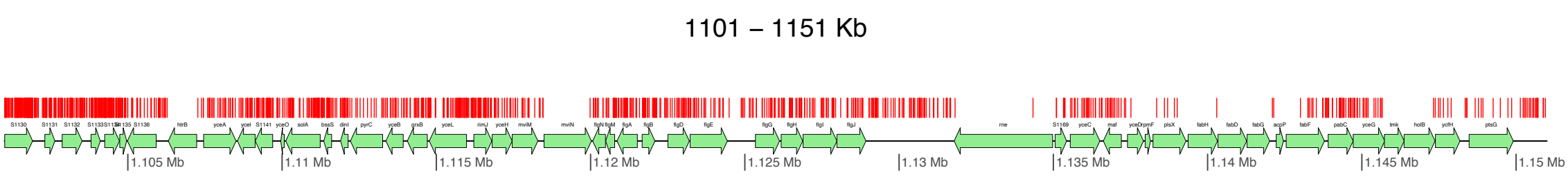
901 – 951 Kb


S0913
S0914
S0917
S0919
S0921
S0922
S0923
S0925
S0927
S0928
S0929
S0930
S0931
S0932
ipaH_2
ybhE
ybhH
ybhI
ybhJ
ybhC
S0941
S0942
S0943
S0944
S0945
S0947
S0948
S0949
S0950
S0953
S0956
S0957
S0958
S0959
S0960
pflA
pflB
focA
ycaO
serC
aroA
yi41
ycaL
cmk
rpsA
905 kb
910 kb
915 kb
920 kb
925 kb
930 kb
935 kb
940 kb
945 kb
950 kb
5 kb


951 – 1001 Kb
ihfB
msbA
lpxK
ycaQ
ycaR
kdsB
S0979
ycbC
smtA
mukF
mukE
mukB
S0985
ycbB
ycbK
ycbL
aspC
ompF
asnC
pncB
pepN
ssuB
S0996
S0997
ycbN
ycbO
ycbP
S1002
S1004
S1005
ycbS
S1007
S1008
S1009
ycbF
pyrD
ycbW
S1013
rlmL
955 kb
960 kb
965 kb
970 kb
975 kb
980 kb
985 kb
990 kb
995 kb
1000 kb
5 kb

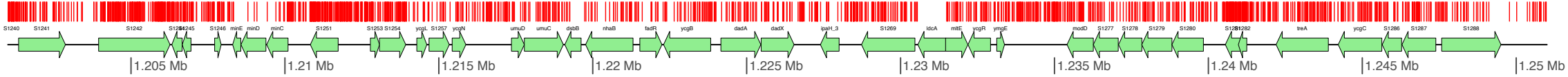

1001 – 1051 Kb
uup
pqiA
pqiB
ymbA
rmf
fabA
S1021
ycbG
ompA
sulA
S1025
S1026
S1027
S1028
yccF
helD
mgsA
S1032
S1033
yccV
S1035
S1036
ecK
yccA
serT
hyaA
hyaB
hyaC
hyaD
hyaE
hyaF
appC
appB
appA
etp
yccZ
ymcA
ymcB
ymcC
ymcD
cspH
S1058
espG
ymcE
yccM
torT
torR
torC
1.005 Mb
1.01 Mb
1.015 Mb
1.02 Mb
1.025 Mb
1.03 Mb
1.035 Mb
1.04 Mb
1.045 Mb
1.05 Mb
5 kb


1051 – 1101 Kb
S1069
S1070
yccD
cbpA
yccE
agp
yccJ
wrbA
ycdF
ycdG
S1080
S1081
S1082
S1083
S1084
S1085
ycdC
S1087
S1088
S1089
ycdO
ycdB
S1093
S1094
ycdU
serX
ycdX
ycdY
ycdZ
csgF
csgE
csgD
csgB
S1110
S1111
csgC
S1114
S1115
ymdC
mdoC
mdoG
mdoH
S1120
S4816
S4815
S4817
S1123
S1124
S1126
S1127
S1128
S1129
1.055 Mb
1.06 Mb
1.065 Mb
1.07 Mb
1.075 Mb
1.08 Mb
1.085 Mb
1.09 Mb
1.095 Mb
1.1 Mb
5 kb

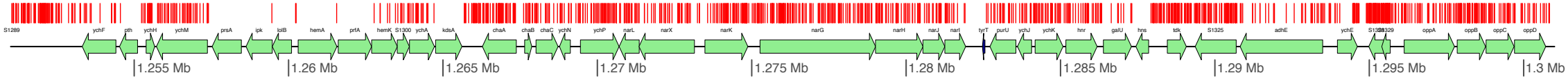

1101 – 1151 Kb
S1130
S1131
S1132
S1133
S1134
S1135
S1136
htrB
yceA
yceI
S1141
yceO
solA
bssS
dinI
pyrC
yceB
grxB
yceL
rimJ
yceH
mviM
mviN
flgN
flgM
flgA
flgB
flgD
flgE
flgG
flgH
flgI
flgJ
rne
S1169
yceC
maf
yceD
rpmF
plsX
fabH
fabD
fabG
acpP
fabF
pabC
yceG
tmk
holB
ycfH
ptsG
1.105 Mb
1.11 Mb
1.115 Mb
1.12 Mb
1.125 Mb
1.13 Mb
1.135 Mb
1.14 Mb
1.145 Mb
1.15 Mb
5 kb

1151 – 1201 Kb
ycfF ycfL ycfM thiK ycfO ycfP ndh ycfJ ycfQ ycfR ycfS mfd ycfT ycfU lolD ycfW ycfX cobB potD potC potB potA pepT ycfD S1211 S1212 S1214 S1215 S1219 S1222 S1223 S1227 S1228 S1229 phoQ phoP purB ycfC mnmA S1236 ymfC icdA S1239
1.155 Mb
1.16 Mb
1.165 Mb
1.17 Mb
1.175 Mb
1.18 Mb
1.185 Mb
1.19 Mb
1.195 Mb
1.2 Mb
5 kb

1201 – 1251 Kb


S1240
S1241
S1242
S1243
S1244
S1245
S1246
minE
minD
minC
S1251
S1253
S1254
ycgL
S1257
ycgN
umuD
umuC
dsbB
nhaB
fadR
ycgB
dadA
dadX
ipaH_3
S1269
ldcA
mltE
ycgR
ymgE
modD
S1277
S1278
S1279
S1280
S1281
S1282
treA
ycgC
S1286
S1287
S1288
1.205 Mb
1.21 Mb
1.215 Mb
1.22 Mb
1.225 Mb
1.23 Mb
1.235 Mb
1.24 Mb
1.245 Mb
1.25 Mb
5 kb

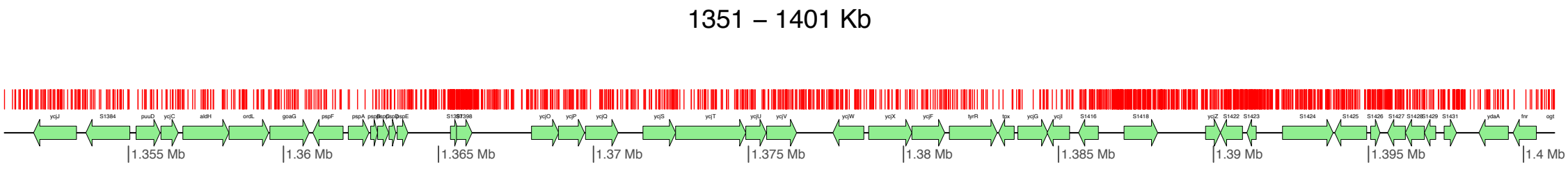
1251 – 1301 Kb

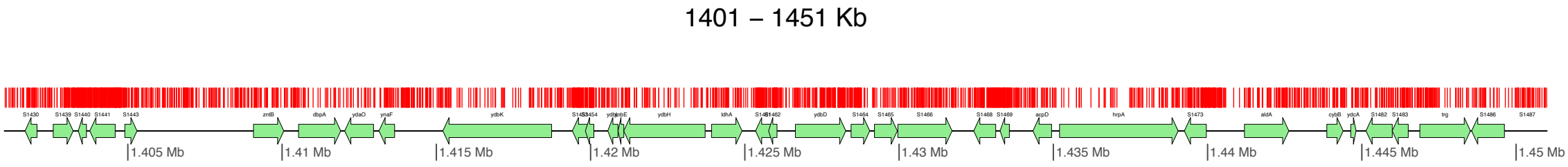


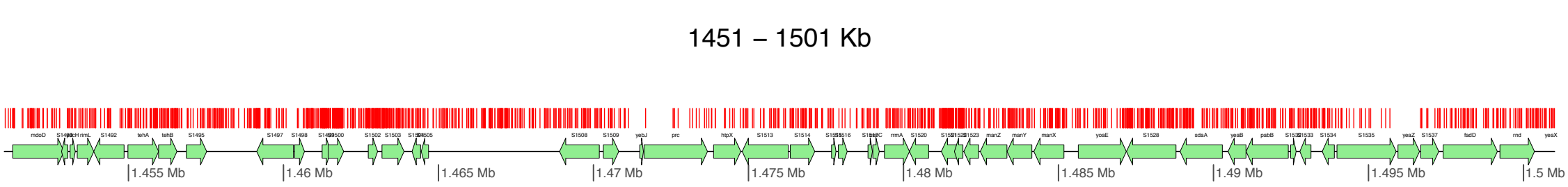
5 kb

1301 – 1351 Kb

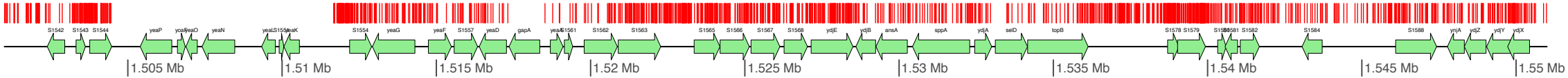
oppF
S1335
cls
kch
S1338
S1339
yciI
tonB
yciA
yciB
yciC
yciD
yciE
yciF
yciG
trpA
trpB
trpC
trpD
trpE
trpL
yciV
yciO
yciL
btuR
yciK
sohB
yciN
topA
cysB
acnA
rlbA
pgpB
yciS
yciM
pyrF
yciH
osmB
S1371
rnb
yciW
fabI
yciD
sapF
sapD
sapC
sapB
sapA
1.305 Mb
1.31 Mb
1.315 Mb
1.32 Mb
1.325 Mb
1.33 Mb
1.335 Mb
1.34 Mb
1.345 Mb
1.35 Mb
5 kb


1351 – 1401 Kb
ycjJ
S1384
puuD
ycjC
aldH
ordL
goaG
pspF
pspA
pspB
pspC
pspD
pspE
S1397
S1398
ycjO
ycjP
ycjQ
ycjS
ycjT
ycjU
ycjV
ycjW
ycjX
ycjF
tyrR
tpx
ycjG
ycjI
S1416
S1418
ycjZ
S1422
S1423
S1424
S1425
S1426
S1427
S1428
S1429
S1431
ydaA
fnr
ogt
1.355 Mb
1.36 Mb
1.365 Mb
1.37 Mb
1.375 Mb
1.38 Mb
1.385 Mb
1.39 Mb
1.395 Mb
1.4 Mb
5 kb


1401 – 1451 Kb
S1430
S1439
S1440
S1441
S1443
zntB
dbpA
ydaO
ynaF
ydbK
S1453
S1454
ydbH
ldhA
S1462
ydbD
S1464
S1465
S1466
S1468
S1469
acpD
hrpA
S1473
aldA
cybB
ydcA
S1482
S1483
trg
S1486
S1487
1.405 Mb
1.41 Mb
1.415 Mb
1.42 Mb
1.425 Mb
1.43 Mb
1.435 Mb
1.44 Mb
1.445 Mb
1.45 Mb
5 kb


1451 – 1501 Kb
mdoD
S1492
tehA
tehB
S1495
S1497
S1498
S1502
S1503
S1508
S1509
yebJ
prc
htpX
S1513
S1514
rrmA
S1520
manZ
manY
manX
yoaE
S1528
sdaA
yeaB
pabB
S1533
S1534
S1535
yeaZ
S1537
fadD
rnd
yeaX
1.455 Mb
1.46 Mb
1.465 Mb
1.47 Mb
1.475 Mb
1.48 Mb
1.485 Mb
1.49 Mb
1.495 Mb
1.5 Mb
5 kb

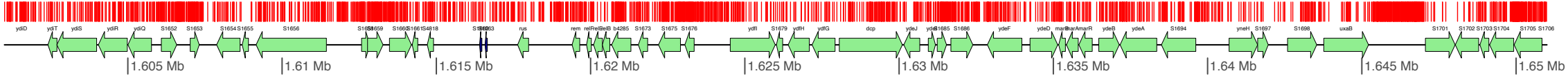
1501 – 1551 Kb


S1542
S1543
S1544
yeaP
yeaO
yeaN
yeaL
yeaK
S1554
yeaG
yeaF
S1557
yeaD
gapA
S1561
S1562
S1563
S1565
S1566
S1567
S1568
ydjE
ydjB
ansA
sppA
ydjA
selD
topB
S1578
S1579
S1581
S1582
S1584
S1588
ynjA
ydjZ
ydjY
ydjX
1.505 Mb
1.51 Mb
1.515 Mb
1.52 Mb
1.525 Mb
1.53 Mb
1.535 Mb
1.54 Mb
1.545 Mb
1.55 Mb
5 kb

1551 – 1601 Kb

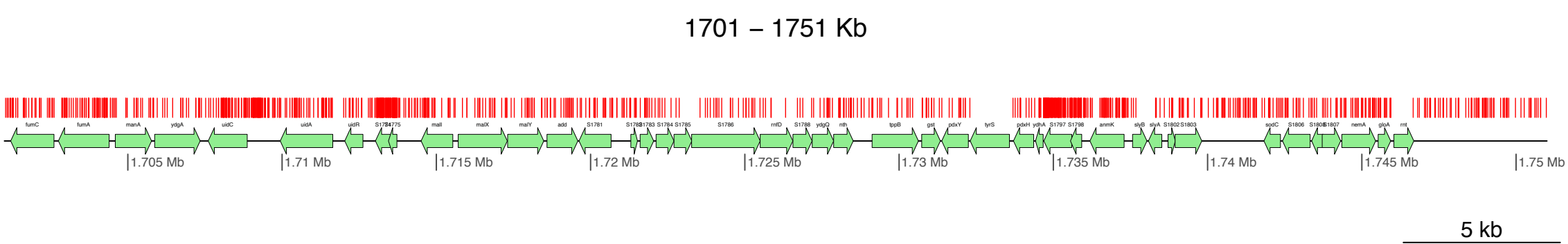
xthA
S1596
astD
ydjS
spy
S1601
S1602
nadE
osmE
celA
celC
celF
ydjC
katE
S1614
yniC
S1621
S1622
S1623
plkB
S1625
thrS
infC
rpmI
rpsT
pheM
pheS
pheT
ihfA
btuC
btuE
btuD
nlpC
S1641
S1642
ydiE
aroH
ydiA
ppsA
1.555 Mb
1.56 Mb
1.565 Mb
1.57 Mb
1.575 Mb
1.58 Mb
1.585 Mb
1.59 Mb
1.595 Mb
1.6 Mb
5 kb

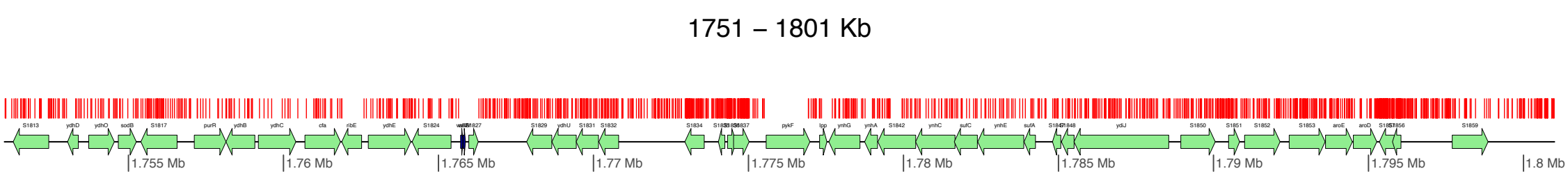
1601 – 1651 Kb


ydiD
ydiT
ydiS
ydiR
ydiQ
S1652
S1653
S1654
S1655
S1656
rus
rem
rem
rem
B
B
b4285
S1673
S1675
S1676
ydfI
S1679
ydfH
ydfG
dcp
ydeJ
S1685
S1686
ydeF
ydeD
ydeB
ydeA
S1694
yneH
S1697
S1698
uxaB
S1701
S1702
S1703
S1704
S1705
S1706
1.605 Mb
1.61 Mb
1.615 Mb
1.62 Mb
1.625 Mb
1.63 Mb
1.635 Mb
1.64 Mb
1.645 Mb
1.65 Mb
5 kb

1651 – 1701 Kb

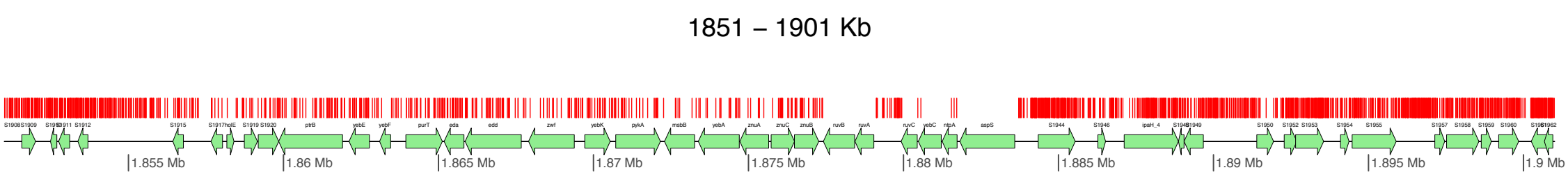
ydeW
S1709
ydeY
ydeZ
S1714
S1715
S1716
S1717
rspA
S1719
S1720
ymfC
S1723
S1727
S1728
S1730
S1731
S1732
S1736
S1738
S1739
S1742
S1743
S1744
S1745
bioD
mlc
ymfL
asr
S1751
S1752
S1753
tqsA
pntB
pntA
S1757
S1758
ydgB
ydgC
rstA
S1762
rstB
tus
1.655 Mb
1.66 Mb
1.665 Mb
1.67 Mb
1.675 Mb
1.68 Mb
1.685 Mb
1.69 Mb
1.695 Mb
1.7 Mb
5 kb

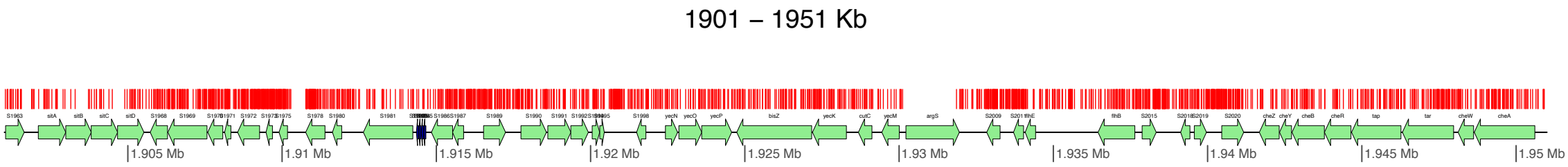

1701 – 1751 Kb
fumC
fumA
manA
ydgA
uidC
uidA
uidR
S1774
S1775
malI
malX
malY
add
S1781
S1782
S1783
S1784
S1785
S1786
rnfD
S1788
ydgQ
nth
tppB
gst
pdxY
tyrS
pdxH
ydhA
S1797
S1798
anmK
slyB
slyA
S1802
S1803
sodC
S1806
S1807
nemA
gloA
rnt
1.705 Mb
1.71 Mb
1.715 Mb
1.72 Mb
1.725 Mb
1.73 Mb
1.735 Mb
1.74 Mb
1.745 Mb
1.75 Mb
5 kb


1751 – 1801 Kb
S1813
ydhD
ydhO
sodB
S1817
purR
ydhB
ydhC
cfa
ribE
ydhE
S1824
S1827
S1829
ydhU
S1831
S1832
S1834
pykF
lpp
ynhG
ynhA
S1842
ynhC
sufC
ynhE
sufA
S1848
ydiJ
S1850
S1851
S1852
S1853
aroE
aroD
S1859
1.755 Mb
1.76 Mb
1.765 Mb
1.77 Mb
1.775 Mb
1.78 Mb
1.785 Mb
1.79 Mb
1.795 Mb
1.8 Mb
5 kb

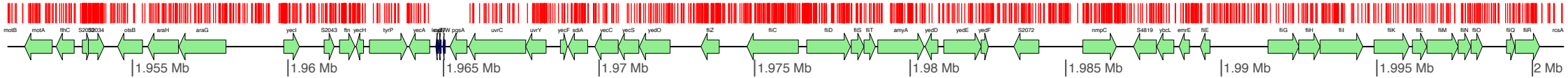
1801 – 1851 Kb

yddB
gadB
xasA
osmC
sfcA
adhP
yddM
fdnI
fdnH
fdnG
yddG
narU
yncB
1.805 Mb
1.81 Mb
1.815 Mb
1.82 Mb
1.825 Mb
1.83 Mb
1.835 Mb
1.84 Mb
1.845 Mb
1.85 Mb
5 kb


1851 – 1901 Kb
S1908
S1909
S1911
S1912
S1915
S1917
holE
S1919
S1920
ptrB
yebE
yebF
purT
eda
edd
zwf
yebK
pykA
msbB
yebA
znuA
znuC
znuB
ruvB
ruvA
ruvC
yebC
ntpA
aspS
S1944
S1946
ipaH_4
S1948
S1949
S1950
S1952
S1953
S1954
S1955
S1957
S1958
S1959
S1960
S1962
1.855 Mb
1.86 Mb
1.865 Mb
1.87 Mb
1.875 Mb
1.88 Mb
1.885 Mb
1.89 Mb
1.895 Mb
1.9 Mb
5 kb

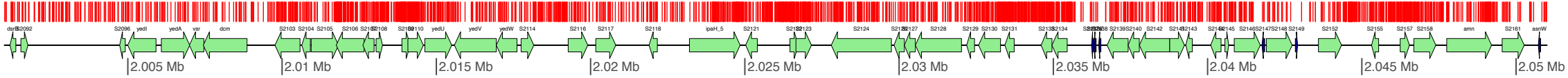

1901 – 1951 Kb
S1963
sitA
sitB
sitC
sitD
S1968
S1969
S1970
S1971
S1972
S1973
S1975
S1978
S1980
S1981
S1986
S1987
S1989
S1990
S1991
S1998
yecN
yecO
yecP
bisZ
yecK
cutC
yecM
argS
S2009
S2011
flhE
flhB
S2015
S2018
S2019
S2020
cheZ
cheY
cheB
cheR
tap
tar
cheW
cheA
1.905 Mb
1.91 Mb
1.915 Mb
1.92 Mb
1.925 Mb
1.93 Mb
1.935 Mb
1.94 Mb
1.945 Mb
1.95 Mb
5 kb

1951 – 2001 Kb



5 kb

2001 – 2051 Kb

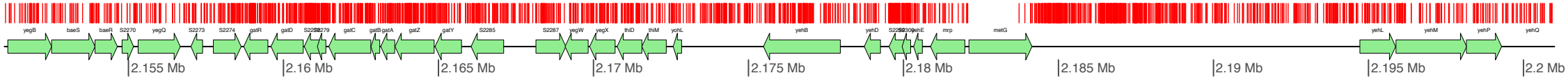

dsrB
S2092
S2096
yedI
yedA
vsr
dcm
S2103
S2104
S2105
S2106
S2110
yedU
yedV
yedW
S2114
S2116
S2117
S2118
ipaH_5
S2121
S2123
S2124
S2127
S2128
S2129
S2130
S2131
S2134
S2139
S2140
S2142
S2143
S2145
S2146
S2147
S2148
S2149
S2152
S2155
S2157
S2158
amn
S2161
asnW
2.005 Mb
2.01 Mb
2.015 Mb
2.02 Mb
2.025 Mb
2.03 Mb
2.035 Mb
2.04 Mb
2.045 Mb
2.05 Mb
5 kb

2051 – 2101 Kb

asnU
cbl
nac
asnV
erfK
cobT
cobS
S2172
S2173
is629a
is629b
yeeX
S2178
S2181
S2182
S2184
S2185
S2187
S2188
S2191
sbcB
yeeD
yeeE
yeeF
yeeY
S2198
S2199
S2200
hisG
hisD
hisC
hisB
hisH
hisA
hisF
hisI
wzzB
gnd
S2213
S2214
S2215
S2216
rfbJ
rfbI
rfc
rfbG
rfbF
rfbE
rfbC
rfbA
rfbD
2.055 Mb
2.06 Mb
2.065 Mb
2.07 Mb
2.075 Mb
2.08 Mb
2.085 Mb
2.09 Mb
2.095 Mb
2.1 Mb
5 kb

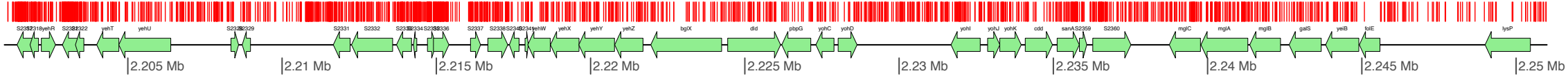
2101 – 2151 Kb
rfbB
galF
wcaM
wcaL
wcaK
wzxC
wcaJ
cpsG
cpsB
wcaI
wcaH
wcaG
gmd
wcaF
S2243
wcaB
S2248
wzb
wza
yegH
asmA
dcd
udk
S2256
alkA
S2261
S2262
issfl4
yegO
2.105 Mb
2.11 Mb
2.115 Mb
2.12 Mb
2.125 Mb
2.13 Mb
2.135 Mb
2.14 Mb
2.145 Mb
2.15 Mb
5 kb

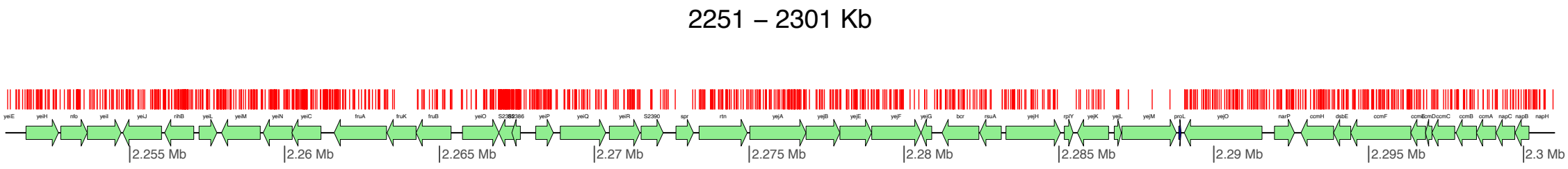
2151 – 2201 Kb



5 kb

2201 – 2251 Kb

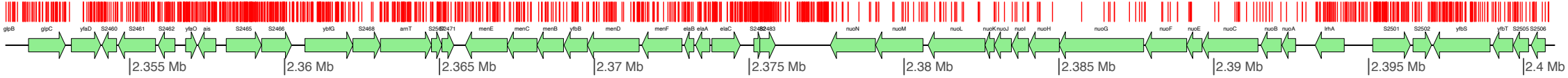

yehR
yehT
yehU
S2331
S2332
yehW
yehX
yehY
yehZ
bglX
dld
pbpG
yohC
yohD
yohI
yohJ
yohK
cdd
sanA
S2359
S2360
mglC
mglA
mglB
galS
yeiB
folE
lysP
2.205 Mb
2.21 Mb
2.215 Mb
2.22 Mb
2.225 Mb
2.23 Mb
2.235 Mb
2.24 Mb
2.245 Mb
2.25 Mb
5 kb


2251 – 2301 Kb
yeiE
yeiH
nfo
yeiI
yeiJ
rihB
yeiL
yeiM
yeiN
yeiC
fruA
fruK
fruB
yeiO
S2385
S2386
yeiP
yeiQ
yeiR
S2390
spr
rtn
yejA
yejB
yejE
yejF
yejG
bcr
rsuA
yejH
rplY
yejK
yejL
yejM
proL
yejO
narP
ccmH
dsbE
ccmF
ccmE
ccmD
ccmC
ccmB
ccmA
nagC
napB
napH
2.255 Mb
2.26 Mb
2.265 Mb
2.27 Mb
2.275 Mb
2.28 Mb
2.285 Mb
2.29 Mb
2.295 Mb
2.3 Mb
5 kb

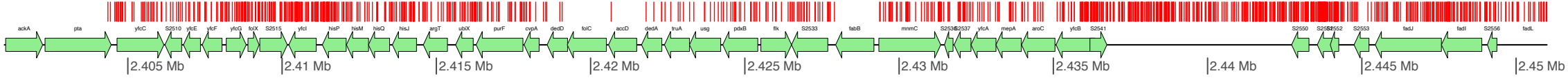
2301 – 2351 Kb

napG
napA
napD
napF
eco
yojI
alkB
ada
yojL
ompC
yojN
rcsB
rcsC
S2434
S2435
S2436
S2438
S2440
S2442
gyrA
ubiG
S2447
S2448
nrdA
nrdB
yfaE
inaA
yfaH
glpQ
glpT
glpA
2.305 Mb
2.31 Mb
2.315 Mb
2.32 Mb
2.325 Mb
2.33 Mb
2.335 Mb
2.34 Mb
2.345 Mb
2.35 Mb
5 kb

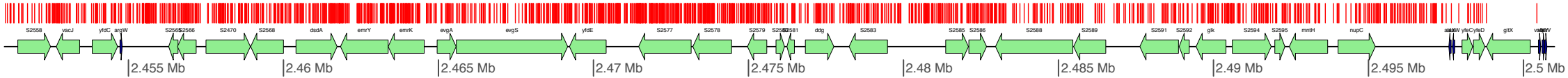
2351 – 2401 Kb


glpB
glpC
yfaD
S2460
S2461
S2462
yfaO
ais
S2465
S2466
ybfG
S2468
arnT
S2470
S2471
menE
menC
menB
yfbB
menD
menF
elaB
elaA
elaC
S2482
S2483
nuoN
nuoM
nuoL
nuoK
nuoJ
nuoI
nuoH
nuoG
nuoF
nuoE
nuoC
nuoB
nuoA
lrhA
S2501
S2502
yfbS
yfbT
S2505
S2506
2.355 Mb
2.36 Mb
2.365 Mb
2.37 Mb
2.375 Mb
2.38 Mb
2.385 Mb
2.39 Mb
2.395 Mb
2.4 Mb
5 kb

2401 – 2451 Kb


ackA
pta
yfcC
S2510
yfcE
yfcF
yfcG
folX
S2515
yfcI
hisP
hisM
hisQ
hisJ
argT
ubiX
purF
cvpA
dedD
folC
accD
dedA
truA
usg
pdxB
flk
S2533
fabB
mnmC
S2536
S2537
yfcA
mepA
aroC
yfcB
S2541
S2550
S2552
S2553
fadJ
fadI
S2556
fadL
2.405 Mb
2.41 Mb
2.415 Mb
2.42 Mb
2.425 Mb
2.43 Mb
2.435 Mb
2.44 Mb
2.445 Mb
2.45 Mb
5 kb

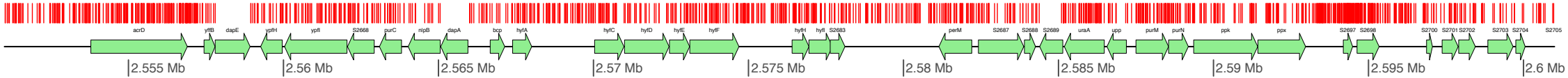
2451 – 2501 Kb


S2558
vacJ
yfdC
argW
S2565
S2566
S2470
S2568
dsdA
emrY
emrK
evgA
evgS
yfdE
S2577
S2578
S2579
S2581
ddg
S2583
S2585
S2586
S2588
S2589
S2591
S2592
glk
S2594
S2595
mntH
nupC
yfeC
yfeD
gltX
2.455 Mb
2.46 Mb
2.465 Mb
2.47 Mb
2.475 Mb
2.48 Mb
2.485 Mb
2.49 Mb
2.495 Mb
2.5 Mb
5 kb

2501 – 2551 Kb

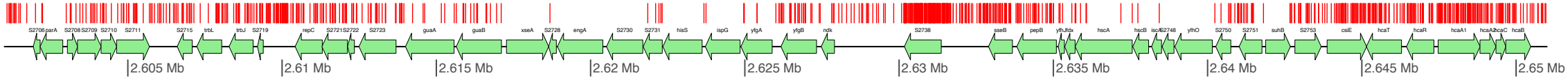
fixA
yfeR
yfeH
ligA
zipA
cysZ
cysK
ptsH
ptsI
crr
pdxK
yfeK
cysM
cysA
cysW
cysU
cysP
ucpA
yfeT
murQ
murP
S2632
S2633
S2634
S2635
amiA
yfeG
S2639
S2641
S2642
S2648
S2649
S2651
S2654
S2655
S2656
talA
lktB
S2659
yfiH
yfiG
2.505 Mb
2.51 Mb
2.515 Mb
2.52 Mb
2.525 Mb
2.53 Mb
2.535 Mb
2.54 Mb
2.545 Mb
2.55 Mb
5 kb

2551 – 2601 Kb



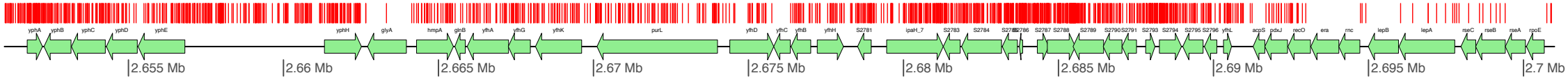
5 kb

2601 – 2651 Kb



5 kb

2651 – 2701 Kb


yphA
yphB
yphC
yphD
yphE
yphH
glyA
hmpA
glnB
yfhA
yfhG
yfhK
purL
yfhD
yfhC
yfhB
yfhH
S2781
ipaH_7
S2783
S2784
S2785
S2786
S2787
S2788
S2789
S2790
S2791
S2793
S2794
S2795
S2796
yfhL
acpS
pdxJ
recO
era
rnc
lepB
lepA
rseC
rseB
rseA
rpoE
2.655 Mb
2.66 Mb
2.665 Mb
2.67 Mb
2.675 Mb
2.68 Mb
2.685 Mb
2.69 Mb
2.695 Mb
2.7 Mb
5 kb

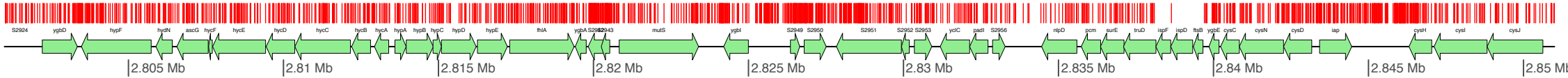
2701 – 2751 Kb

yfiC
srmB
yfiE
yfiK
yfiD
ung
yfiF
trxC
yfiP
yfiQ
pssA
yfiM
kgtP
rrfG
rrfG
gltW
rrsG
S2827
S2828
rluD
yfiH
clpB
S2832
pheL
pheA
tyrA
aroF
yfiL
S2840
yfiB
rplS
trmD
rimM
rpsP
ffh
S2848
yfjD
grpE
ppnK
recN
smpB
S2859
S2860
2.705 Mb
2.71 Mb
2.715 Mb
2.72 Mb
2.725 Mb
2.73 Mb
2.735 Mb
2.74 Mb
2.745 Mb
2.75 Mb
5 kb

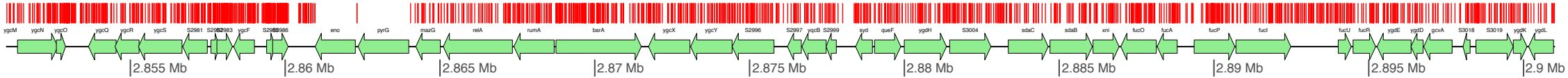
2751 – 2801 Kb

S2863
ileY
S2870
S2871
ygaF
gabT
gabP
ygaE
ygaS
S2878
S2879
S2880
S2881
ygaP
stpA
S2884
ygaC
nrdH
nrdI
nrdE
nrdF
proV
proW
proX
S2894
S2896
ygaH
emrR
emrB
ygaG
gshA
S2903
ygaB
csrA
alaS
recX
recA
ygaD
miaB
srlE
srlB
srlD
gutM
srlR
gutQ
ygaA
2.755 Mb
2.76 Mb
2.765 Mb
2.77 Mb
2.775 Mb
2.78 Mb
2.785 Mb
2.79 Mb
2.795 Mb
2.8 Mb
5 kb

2801 – 2851 Kb


S2924
ygbD
hypF
hydN
ascG
hycF
hycE
hycD
hycC
hycB
hycA
hypA
hypB
hypC
hypD
hypE
fhlA
ygbA
S2943
mutS
ygbI
S2949
S2950
S2951
S2952
S2953
yciC
padI
S2956
nlpD
pcm
surE
truD
ispF
ispD
ftsB
ygbE
cysC
cysN
cysD
iap
cysH
cysI
cysJ
2.805 Mb
2.81 Mb
2.815 Mb
2.82 Mb
2.825 Mb
2.83 Mb
2.835 Mb
2.84 Mb
2.845 Mb
2.85 Mb
5 kb

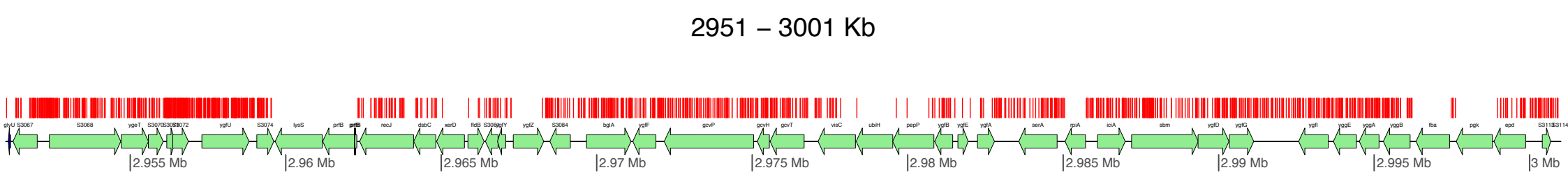
2851 – 2901 Kb



5 kb

2901 – 2951 Kb

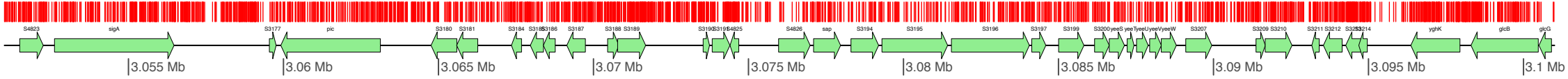
mltA
S3025
argA
recD
recB
ptr
recC
ppdD
ppdB
thyA
lgt
ptsP
ygdP
mutH
S3040
S3041
tas
ygeD
aas
galR
lysA
lysR
ygeA
araE
kduD
kdul
yqeF
S3053
S3055
S3056
S3059
S3060
S3062
S3063
S3064
2.905 Mb
2.91 Mb
2.915 Mb
2.92 Mb
2.925 Mb
2.93 Mb
2.935 Mb
2.94 Mb
2.945 Mb
2.95 Mb
5 kb


2951 – 3001 Kb
glyU
S3067
S3068
ygeT
S3070
S3071
S3072
ygfU
S3074
lysS
prfB
recJ
dsbC
xerD
fldB
S3083
ygfY
ygfZ
S3084
bglA
yglF
gcvP
gcvH
gcvT
visC
ubiH
pepP
ygfB
ygfE
ygfA
serA
rpiA
iciA
sbm
ygfD
ygfG
ygfI
yggE
yggA
yggB
fba
pgk
epd
S3113
S3114
2.955 Mb
2.96 Mb
2.965 Mb
2.97 Mb
2.975 Mb
2.98 Mb
2.985 Mb
2.99 Mb
2.995 Mb
3 Mb
5 kb

3001 – 3051 Kb

S3116
S3117
S3118
yggD
yggF
S3122
S3125
S3126
tktA
yggG
speB
S3132
speA
metK
galP
sprT
endA
yggJ
gshB
yggE
yggF
yggR
yggS
yggT
yggU
yggV
yggW
S3151
S3152
S3153
S3154
S3155
yggM
ansB
yggN
yggL
trmB
mutY
yggX
miaC
nupG
speC
yggA
pheV
S4822
S3169
3.005 Mb
3.01 Mb
3.015 Mb
3.02 Mb
3.025 Mb
3.03 Mb
3.035 Mb
3.04 Mb
3.045 Mb
3.05 Mb
5 kb

3051 – 3101 Kb


S4823
sigA
S3177
pic
S3180
S3181
S3184
S3185
S3186
S3187
S3188
S3189
S3190
S3191
S4825
S4826
sap
S3194
S3195
S3196
S3197
S3199
S3200
yeeS
yeeT
yeeU
yeeV
yeeW
S3207
S3209
S3210
S3211
S3212
S3213
S3214
yghK
glcB
glcG
3.055 Mb
3.06 Mb
3.065 Mb
3.07 Mb
3.075 Mb
3.08 Mb
3.085 Mb
3.09 Mb
3.095 Mb
3.1 Mb
5 kb

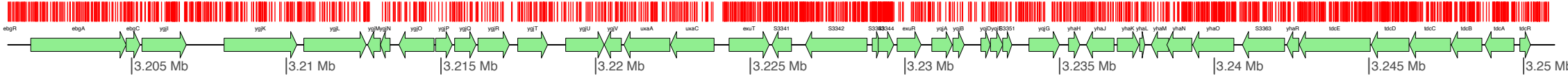
3101 – 3151 Kb

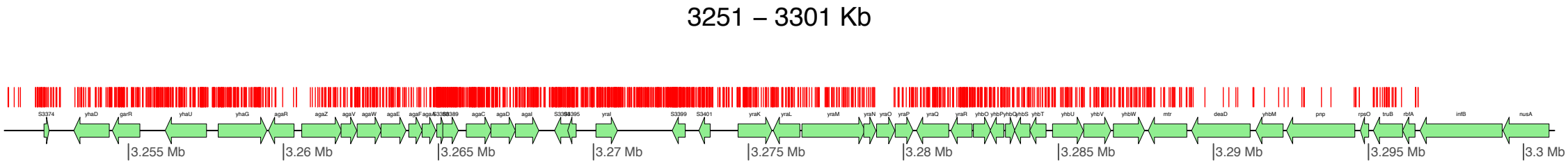
glcC
S3222
S3223
S3225
S3226
S3227
S3228
S3229
S3230
yghQ
yghR
yghS
yghT
pitB
gsp
S3237
hybG
hypA
hybE
hybD
hybC
hybB
hybA
S3245
S3246
S3247
S3249
yghA
yghA
S3253
exbB
S3255
metC
yghB
yghC
yghD
dkgA
yghG
yghH
S3263
sufI
plsC
parC
S3268
3.105 Mb
3.11 Mb
3.115 Mb
3.12 Mb
3.125 Mb
3.13 Mb
3.135 Mb
3.14 Mb
3.145 Mb
3.15 Mb
5 kb

3151 – 3201 Kb

ygiW
ygiX
ygiY
S3272
mdaB
ygiN
parE
ygiA
icc
ygiB
nudF
tolC
ygiA
ygiB
ygiC
ygiD
ygiE
ribB
S3287
S3288
S3289
S3292
glgS
S3295
S3298
glnE
ygiF
ygiM
cca
uppP
folB
ygiH
ygiP
ttdA
ttdB
ygjE
ygjD
rpsU
dnaG
rpoD
S3314
ileX
ygjI
ygjG
3.155 Mb
3.16 Mb
3.165 Mb
3.17 Mb
3.175 Mb
3.18 Mb
3.185 Mb
3.19 Mb
3.195 Mb
3.2 Mb
5 kb

3201 – 3251 Kb

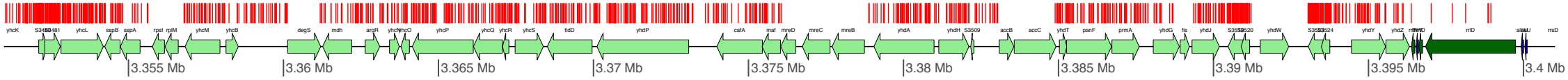

ebgR
ebgA
ebgC
ygjI
ygjK
ygjL
ygjM
ygjN
ygjO
ygjP
ygjQ
ygjR
ygjT
ygjU
ygjV
uxaA
uxaC
exuT
S3341
S3342
S3343
S3344
exuR
ygjA
ygjB
ygjD
ygjE
S3351
ygjG
yhaH
yhaJ
yhaK
yhaL
yhaM
yhaN
yhaO
S3363
yhaR
tdcE
tdcD
tdcC
tdcB
tdcA
tdcR
3.205 Mb
3.21 Mb
3.215 Mb
3.22 Mb
3.225 Mb
3.23 Mb
3.235 Mb
3.24 Mb
3.245 Mb
3.25 Mb
5 kb


3251 – 3301 Kb
S3374
yhaD
garR
yhaU
yhaG
agaR
agaZ
agaV
agaW
agaE
agaF
agaA
S3388
S3389
agaC
agaD
agaI
S3394
S3395
yraI
S3399
S3401
yraK
yraL
yraM
yraN
yraO
yraP
yraQ
yraR
yhbO
yhbP
yhbQ
yhbS
yhbT
yhbU
yhbV
yhbW
mtr
deaD
yhbM
pnp
rpsO
truB
rbfA
infB
nusA
3.255 Mb
3.26 Mb
3.265 Mb
3.27 Mb
3.275 Mb
3.28 Mb
3.285 Mb
3.29 Mb
3.295 Mb
3.3 Mb
5 kb

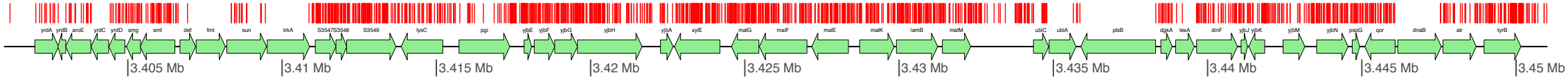
3301 – 3351 Kb

yhbC
metY
argG
yhbX
leuA
leuCG
glmM
folP
hflB
rrmJ
yhbY
greA
dacB
obgE
yhbE
rpmA
rplU
ispB
nlp
murA
yrbA
yrbB
yrbC
vpsC
vpsB
vpsA
yrbG
yrbH
yrbI
yrbK
yhbN
yhbG
rpoN
yhbH
ptsN
yhbJ
ptsO
yrbL
yhbL
arcB
yhcC
gltB
gltD
S3473
yhcH
yhcI
yhcJ
nanT
nanA
3.305 Mb
3.31 Mb
3.315 Mb
3.32 Mb
3.325 Mb
3.33 Mb
3.335 Mb
3.34 Mb
3.345 Mb
3.35 Mb
5 kb

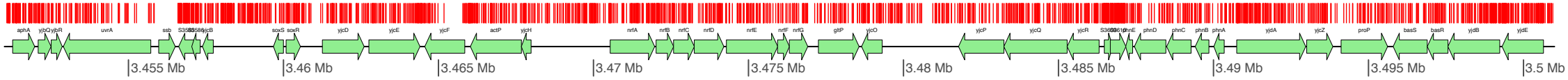
3351 – 3401 Kb


yhcK
yhcL
sspB
sspA
rpsI
rplM
yhcM
yhcB
degS
mdh
argR
yhcO
yhcP
yhcQ
yhcR
yhcS
tldD
yhdP
cafA
maf
mreD
mreC
mreB
yhdA
yhdH
accB
accC
yhdT
panF
prmA
yhdG
fis
yhdJ
yhdW
yhdY
yhdZ
rrlD
3.355 Mb
3.36 Mb
3.365 Mb
3.37 Mb
3.375 Mb
3.38 Mb
3.385 Mb
3.39 Mb
3.395 Mb
3.4 Mb
5 kb

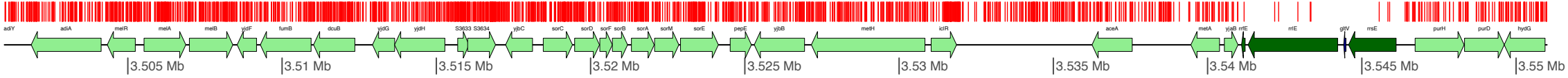
3401 – 3451 Kb


yrdA
yrdB
arcE
yrdC
yrdD
smg
smf
def
fmt
sun
trkA
S3547
S3548
S3549
lysC
pgi
yjbE
yjbF
yjbG
yjbH
yjbA
xylE
malG
malF
malE
malK
lamB
malM
ubiC
ubiA
plsB
dgkA
lexA
dinF
yjbJ
yjbK
yjbM
yjbN
pspG
qor
dnaB
air
tyrB
3.405 Mb
3.41 Mb
3.415 Mb
3.42 Mb
3.425 Mb
3.43 Mb
3.435 Mb
3.44 Mb
3.445 Mb
3.45 Mb
5 kb

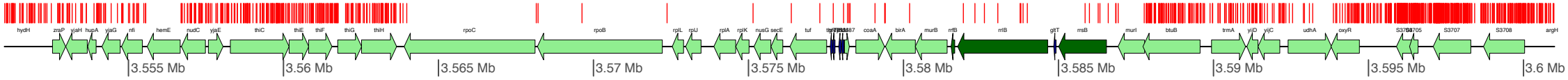
3451 – 3501 Kb


aphA
yjbQ
yjbR
uvrA
ssb
yjcB
soxS
soxR
yjcD
yjcE
yjcF
actP
yjcH
nrfA
nrfB
nrfC
nrfD
nrfE
nrfF
nrfG
gltP
yjcO
yjcP
yjcQ
yjcR
phnE
phnD
phnC
phnB
phnA
yjdA
yjcZ
proP
basS
basR
yjdB
yjdE
3.455 Mb
3.46 Mb
3.465 Mb
3.47 Mb
3.475 Mb
3.48 Mb
3.485 Mb
3.49 Mb
3.495 Mb
3.5 Mb
5 kb

3501 – 3551 Kb


adiY
adiA
melR
metA
melB
yidF
fumB
dcuB
yidG
yjdH
S3633
S3634
yjbC
sorC
sorD
sorF
sorB
sorA
sorM
sorE
pepE
yjbB
metH
iclR
aceA
metA
yiaB
rrlE
rrlE
gltV
rrsE
purH
purD
hydG
3.505 Mb
3.51 Mb
3.515 Mb
3.52 Mb
3.525 Mb
3.53 Mb
3.535 Mb
3.54 Mb
3.545 Mb
3.55 Mb
5 kb

3551 – 3601 Kb


hydH
zraP
ylaH
hupA
yjaG
nfi
hemE
nudC
yjaE
thiC
thiE
thiF
thiG
thiH
rpoC
rpoB
rplL
rplJ
rplA
rplK
nusG
secE
tuf
coaA
birA
murB
rrfB
rrlB
glyT
rrsB
murI
btuB
trmA
yijD
yijC
udhA
oxyR
S3705
S3707
S3708
argH
3.555 Mb
3.56 Mb
3.565 Mb
3.57 Mb
3.575 Mb
3.58 Mb
3.585 Mb
3.59 Mb
3.595 Mb
3.6 Mb
5 kb

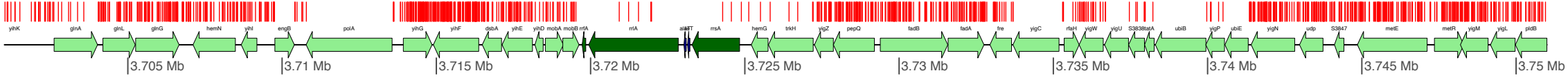
3601 – 3651 Kb

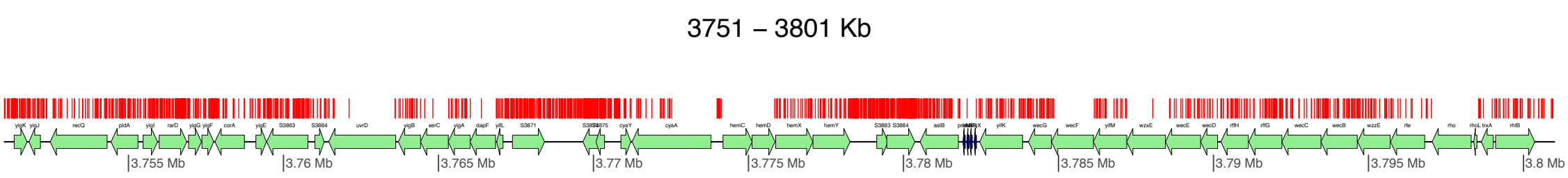
argB
argC
argE
ppc
yijP
yijO
frwD
pflC
pflD
frwB
frwC
katG
metF
metL
metB
metJ
yiiX
rpmE
priA
cytR
ftsN
hslV
hslU
menA
menG
yiiU
glpF
glpK
glpX
fpr
yiiT
yiiS
yiiQ
tpiA
cdh
sbp
pfkA
3.605 Mb
3.61 Mb
3.615 Mb
3.62 Mb
3.625 Mb
3.63 Mb
3.635 Mb
3.64 Mb
3.645 Mb
3.65 Mb
5 kb

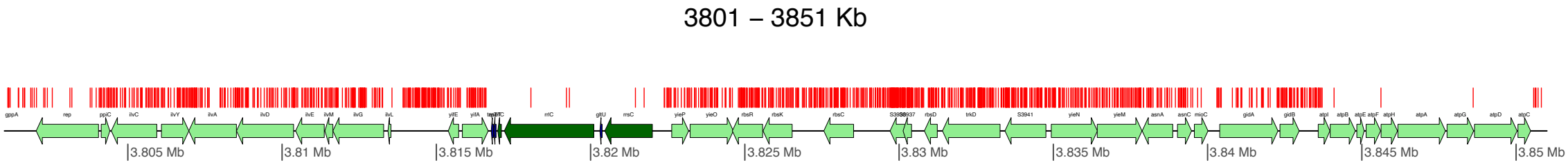
3651 – 3701 Kb

fieF
cpxP
cpxR
cpxA
yiiM
S3760
S3761
kdgT
sodA
rhaR
rhaS
rhaB
rhaA
rhaD
yiiL
frvA
frvB
frvX
frvR
fdhD
fdoG
fdoH
fdoI
fdhE
yiiF
yiiE
S4827
yiiD
yihZ
rbn
yihX
yihW
yihV
yihU
yihT
yihS
yihR
yihQ
yihO
yihN
yihM
yihL
3.655 Mb
3.66 Mb
3.665 Mb
3.67 Mb
3.675 Mb
3.68 Mb
3.685 Mb
3.69 Mb
3.695 Mb
3.7 Mb
5 kb

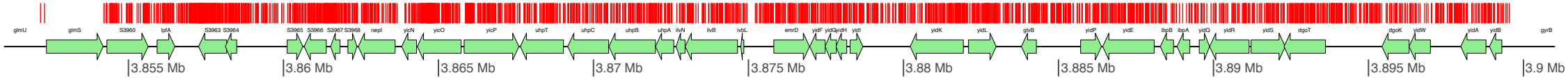
3701 – 3751 Kb


yihK
glnA
glnL
glnG
hemN
yihI
engB
polA
yihG
yihF
dsbA
yihE
yihD
mobA
mobB
rrfA
rrlA
alaT
rrsA
hemG
trkH
yigZ
pepQ
fadB
fadA
fre
yigC
rlsH
yigW
yigU
S3838
tatA
ubiB
yigP
ubiE
yigN
udp
S3847
metE
metR
yigM
yigL
pldB
3.705 Mb
3.71 Mb
3.715 Mb
3.72 Mb
3.725 Mb
3.73 Mb
3.735 Mb
3.74 Mb
3.745 Mb
3.75 Mb
5 kb


3751 – 3801 Kb
yigK
yigJ
recQ
pldA
yigI
rarD
yigG
yigF
corA
yigE
S3863
S3864
uvrD
yigB
xerC
yigA
dapF
yifL
S3871
S3875
cyaY
cyaA
hemC
hemD
hemX
hemY
S3883
S3884
atsB
yifK
wecG
wecF
yifM
wzxE
wecE
wecD
rffH
rffG
wecC
wecB
wzzE
rfe
rho
rhoL
trxA
rhlB
3.755 Mb
3.76 Mb
3.765 Mb
3.77 Mb
3.775 Mb
3.78 Mb
3.785 Mb
3.79 Mb
3.795 Mb
3.8 Mb
5 kb


3801 – 3851 Kb
gppA
rep
ppiC
ilvC
ilvY
ilvA
ilvD
ilvE
ilvM
ilvG
ilvL
yifE
yifA
rrlC
gltU
rrsC
yieP
yieO
rbsR
rbsK
rbsC
S3938
S3937
rbsD
trkD
S3941
yieN
yieM
asnA
asnC
miaC
gidA
gidB
atpI
atpB
atpE
atpF
atpH
atpA
atpG
atpD
atpC
3.805 Mb
3.81 Mb
3.815 Mb
3.82 Mb
3.825 Mb
3.83 Mb
3.835 Mb
3.84 Mb
3.845 Mb
3.85 Mb
5 kb

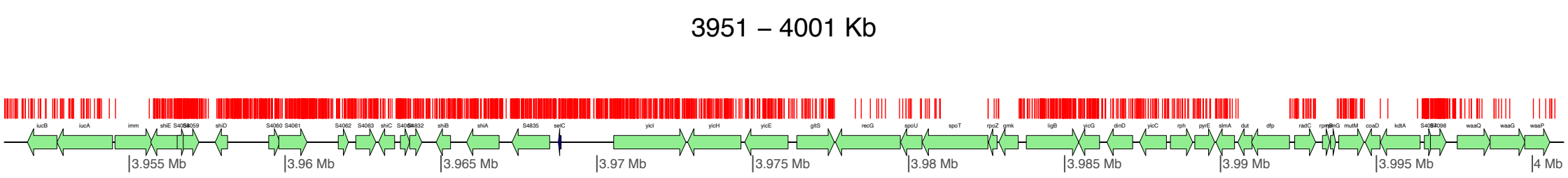
3851 – 3901 Kb

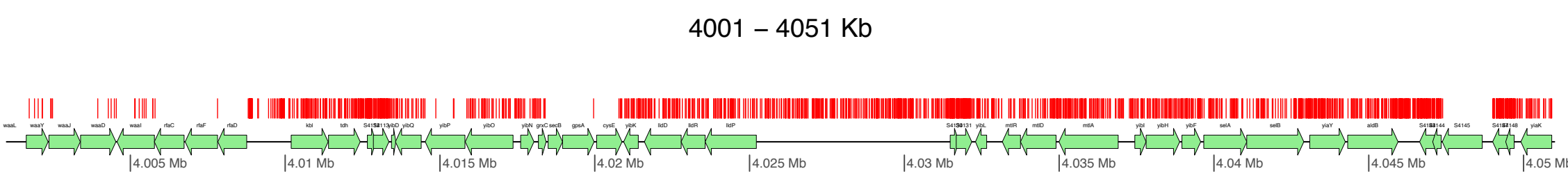


5 kb

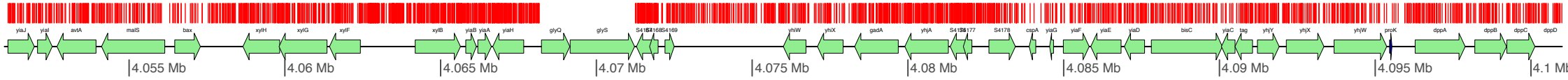
3901 – 3951 Kb

recF
dnaN
dnaA
rpmH
hpA
yidC
trmE
S4014
S4015
tnaL
tnaA
tnaB
yidY
S4021
S4022
yieE
yieF
yieG
yieH
yieI
S4029
S4030
S4031
yieK
yieL
yieC
S4036
S4037
bglB
bglF
bglG
phoU
pstB
pstA
pstC
pstS
S4046
lplC
S4049
S4050
iutA
iucD
iucC
3.905 Mb
3.91 Mb
3.915 Mb
3.92 Mb
3.925 Mb
3.93 Mb
3.935 Mb
3.94 Mb
3.945 Mb
3.95 Mb
5 kb

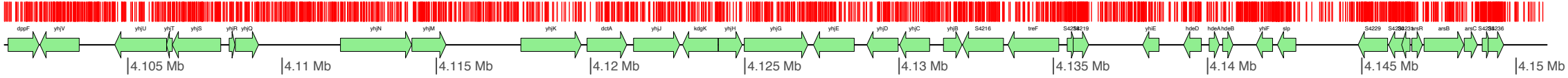

3951 – 4001 Kb
iucB
iucA
imm
shiE
S4058
S4059
shiD
S4060
S4061
S4062
S4063
shiC
S4064
S4832
shiB
shiA
S4835
selC
yicI
yicH
yicE
gltS
recG
spoU
spoT
rpoZ
gmk
ligB
yicG
dinD
yicC
rph
pyrE
smA
dut
dfp
radC
rpmH
rnpA
rnpB
mutM
coaD
kdtA
S4097
S4098
waaQ
waaG
waaP
3.955 Mb
3.96 Mb
3.965 Mb
3.97 Mb
3.975 Mb
3.98 Mb
3.985 Mb
3.99 Mb
3.995 Mb
4 Mb
5 kb


4001 – 4051 Kb
waaL
waaY
waaJ
waaD
waaI
rfaC
rfaF
rfaD
kbl
tdh
ybiD
ybiQ
ybiP
ybiO
ybiN
gnxC
secB
gpsA
cysE
ybiK
lldD
lldR
lldP
ybiL
mtlR
mtlD
mtlA
ybiI
ybiH
ybiF
selA
selB
ylaY
aldB
S4145
yiaK
4.005 Mb
4.01 Mb
4.015 Mb
4.02 Mb
4.025 Mb
4.03 Mb
4.035 Mb
4.04 Mb
4.045 Mb
4.05 Mb
5 kb

4051 – 4101 Kb


yiaJ
yiaI
avtA
malS
bax
xylH
xylG
xylF
xylB
yiaB
yiaA
yiaH
glyQ
glyS
S4168
S4169
yhiW
yhiX
gadA
yhjA
S4177
S4178
cspA
yiaG
yiaF
yiaE
yiaD
bisC
yiaC
tag
yhjY
yhjX
yhjW
proK
dppA
dppB
dppC
dppD
4.055 Mb
4.06 Mb
4.065 Mb
4.07 Mb
4.075 Mb
4.08 Mb
4.085 Mb
4.09 Mb
4.095 Mb
4.1 Mb
5 kb

4101 – 4151 Kb


dppF
yhjV
yhjU
yhjT
yhjS
yhjR
yhjQ
yhjN
yhjM
yhjK
dctA
yhjJ
kdgK
yhjH
yhjG
yhjE
yhjD
yhjC
yhjB
S4216
treF
S4218
S4219
yhjE
hdeD
hdeA
hdeB
yhjF
slp
S4229
ansR
arsB
arsC
4.105 Mb
4.11 Mb
4.115 Mb
4.12 Mb
4.125 Mb
4.13 Mb
4.135 Mb
4.14 Mb
4.145 Mb
4.15 Mb
5 kb

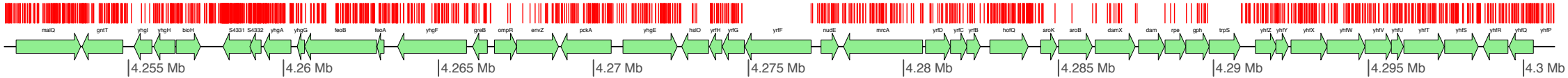
4151 – 4201 Kb

gor
yhiR
prlC
yhiQ
uspA
yhiO
pitA
yhiN
S4248
S4249
S4250
S4254
S4252
S4253
S4255
S4836
S4837
S4838
yhiJ
yhiI
yhiH
yhhJ
S4263
yhhG
nikE
nikD
nikC
nikB
nikA
yhhU
yhhT
yhhS
S4273
yhhQ
yhhP
zntA
yhhN
yhhM
yhhL
rsmD
ftsY
4.155 Mb
4.16 Mb
4.165 Mb
4.17 Mb
4.175 Mb
4.18 Mb
4.185 Mb
4.19 Mb
4.195 Mb
4.2 Mb
5 kb

4201 – 4251 Kb

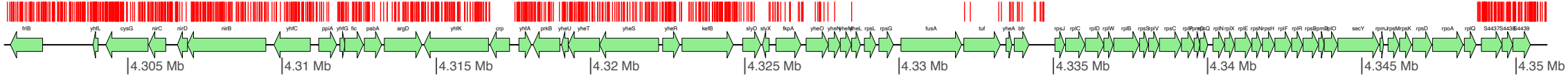
ftsE
ftsX
rpoH
livJ
yhhK
livK
livH
livM
livG
livF
ugpB
ugpA
ugpE
ugpQ
yhhA
ggt
yhhY
yhhX
yhhW
gntR
gntK
gntU
yhgN
asd
glgB
glgX
glgC
glgA
glgP
S4313
S4314
glgD
glgE
glgG
glgR
rtcR
rtcB
yhgJ
malP
4.205 Mb
4.21 Mb
4.215 Mb
4.22 Mb
4.225 Mb
4.23 Mb
4.235 Mb
4.24 Mb
4.245 Mb
4.25 Mb
5 kb

4251 – 4301 Kb

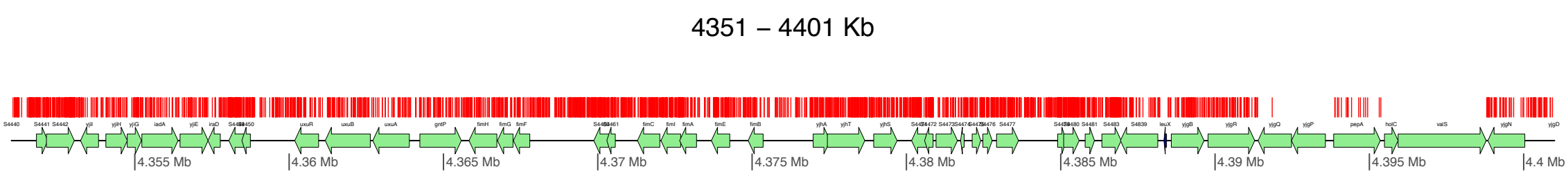


5 kb

4301 – 4351 Kb



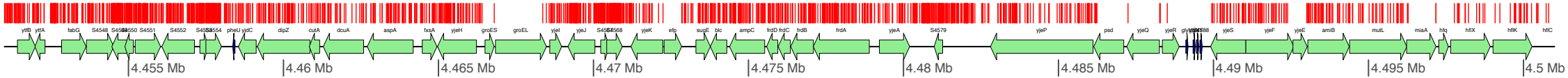
5 kb


4351 – 4401 Kb
S4440
S4441
S4442
yjiI
yjiH
yjiG
iadA
yjiE
iraD
uxuR
uxuB
uxuA
gntP
fimH
fimG
fimF
fimC
fimI
fimA
fimE
fimB
yjhA
yjhT
yjhS
S4477
S4483
S4839
leuX
yjgB
yjgR
yjgQ
yjgP
pepA
holC
valS
yjgN
yjgD
4.355 Mb
4.36 Mb
4.365 Mb
4.37 Mb
4.375 Mb
4.38 Mb
4.385 Mb
4.39 Mb
4.395 Mb
4.4 Mb
5 kb

4401 – 4451 Kb

argI
S4497
S4498
yjgK
yjgJ
S4502
yjgI
yjgH
yjgG
yjgL
pyrB
pyrI
yjgF
mgtA
treR
treB
treC
nrdD
nrdG
pmbA
yjgA
yjfG
fbp
S4521
yttT
yttR
ppa
yttP
yttN
yttM
msrA
yttL
yttK
yttJ
cysQ
cpdB
yttH
yttG
yttF
yttE
cycA
4.405 Mb
4.41 Mb
4.415 Mb
4.42 Mb
4.425 Mb
4.43 Mb
4.435 Mb
4.44 Mb
4.445 Mb
4.45 Mb
5 kb

4451 – 4501 Kb


ytfB ytfA fabG S4548 S4549 S4550 S4551 S4552 S4553 S4554 pheU yjdC dipZ cutA dcuA aspA fxsA yjeH groES groEL yjeI yjeJ S4567 S4568 yjeK efp sugE blc ampC frdD frdC frdB frdA yjeA S4579 yjeP psd yjeQ yjeR S4588 yjeS yjeF yjeE amiB mutL miaA hfq hflX hflK hflC
4.455 Mb
4.46 Mb
4.465 Mb
4.47 Mb
4.475 Mb
4.48 Mb
4.485 Mb
4.49 Mb
4.495 Mb
4.5 Mb
5 kb

4501 – 4551 Kb

yjeT
purA
yjeB
vacB
yjfH
yjfM
yjfC
yjfN
yjfO
yjfP
yjfQ
yjfR
ulaA
sgaB
ptxA
ulaD
sgaU
sgaE
yjfY
rpsF
priB
rsR
rplI
yjiA
yjiX
yjiY
hpaC
hpaA
hpaX
hpaI
hpaH
hpaF
hpaD
hpaE
hpaG
hpcR
tsr
4.505 Mb
4.51 Mb
4.515 Mb
4.52 Mb
4.525 Mb
4.53 Mb
4.535 Mb
4.54 Mb
4.545 Mb
4.55 Mb
5 kb

4551 – 4601 Kb

yjiZ
yjjM
mdoB
yjjA
dnaC
dnaT
yjjB
S4667
bglJ
fhuF
rsmC
holD
rimI
yjjG
prfC
osmY
yjjU
yjjV
yjjW
yjjI
deoC
deoA
deoB
deoD
S4688
smp
serB
sms
nadR
yjjK
slt
trpR
gpmB
rob
creA
creB
creC
creD
arcA
yjjY
lasT
4.555 Mb
4.56 Mb
4.565 Mb
4.57 Mb
4.575 Mb
4.58 Mb
4.585 Mb
4.59 Mb
4.595 Mb
4.6 Mb
5 kb