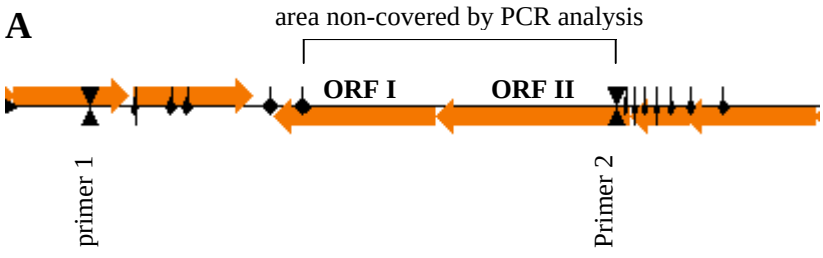
Assessment of conditional gene essentiality based on genetic footprinting data.

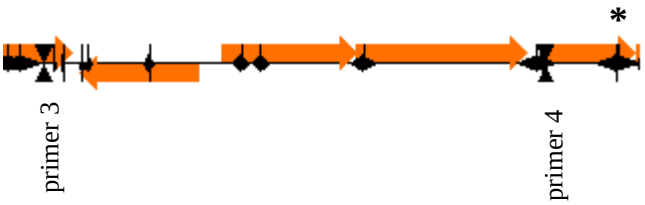
Gene essentiality conclusions were based on a semi-automatic analysis of the number and relative positions of inserts within each gene after selective outgrowth for 23 population doublings. Retention of multiple inserts generally identifies a gene as dispensable under these conditions. Failure to recover inserts, or the presence of only a limited number of inserts at the very ends of a coding sequence, suggests that the gene is essential under these growth conditions. Initial essentiality assertions were made automatically based on the number of transposon insertions detected within an ORF, and on the relative intensity of electrophoretic bands corresponding to each transposition event. The automatic calls were manually confirmed or corrected. Each ORF was categorized according to the criteria described below.

ORFs were asserted as undetermined (X) if no reliable PCR data could be obtained for the corresponding region of the *E. coli* chromosome for technical reasons, such as PCR failure, nonspecific primer annealing in areas of DNA repeats, or insufficient length of generated PCR products. The later problem is often due to simultaneous synthesis of a large number of PCR products in a particular reaction. This is illustrated below with examples of genetic footprints visualized withChromosomal Viewer software.

Legend: In each illustration the large horizontal orange arrows indicate the length and direction of each gene. Positions of landmark PCR primers are shown by bows crossing the genes. Black diamonds represent transposon insertions. The width of each diamond corresponds to mapping error introduced by gel electrophoresis (generally equal to ~4.5% of the size of each PCR product). The vertical line associated with each diamond shows the relative intensity of the corresponding electrophoretic band.



B



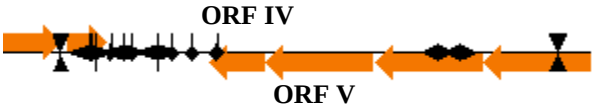
A: ORFs II and I were asserted as undetermined due to insufficient length of PCR products obtained with landmark primer 1. The longest PCR product originating from primer 1 is significantly shorter then the distance between landmark primers 1 and 2. **B:** The region between primers 3 and 4 was "covered" in its entirety, since the longest PCR product (marked by an asterisk) originating from primer 3 is longer then the distance between primers 3 and 4.

Potential sources of erroneous assertions. The absence of insertions in a chromosomal region distant from a landmark primer (example A) in some cases could be due to essentiality of ORF II. However, for consistency we chose to assert all genes exemplified by ORFs I and II as non-determined since we cannot be confident that they are essential.

ORFs asserted as essential (E). All ORFs longer then 240 base pairs in length (>80 AA) free from inserts were asserted as essential, if located in regions for which reliable PCR data were obtained. In addition, genes with only a few insertions within the 3'-most 20% or 5'-most 5% of the gene were asserted as essential. These insertions were considered non-disruptive. Examples of essential genes are shown below (ORFs III through VIII).

C

D



E

F

Potential sources of erroneous assertions.

1. Lack of insertions in relatively short ORFs (240 to 350 bp long), consistently interpreted as an indication of gene essentiality, can in fact be accidental, since the average density of detectable transposon insertion was 1 per 300 bp. To account for data interpretation errors of this kind, a numerical measure of confidence in an assertion was calculated for each essential ORF ("assertion error" in Table S1). Assertion error shows the probability of missing an ORF by chance if insert locations were completely random. It was calculated as follows: assuming that a Poisson process underlies transposition, the probability of missing an ORF is given by exp(-rL), where r is the local insertion density and L is the length of the ORF in base pairs. In our case, r was determined by counting the number of inserts within a 10 kb-long region centered on each ORF, excluding its coding sequence, and all essential ORFs and unanalyzed regions (gaps).

2. Lack of insertions within an ORF can be due to reasons other than essentiality. For example "cold spots" for transposition, or polar effects, in which insertions into a dispensable gene are selected against due to their disruptive effect on the transcription of a downstream essential gene (also see the main article and (Gerdes et al., 2002) would appear to be essential.

ORFs asserted as non-essential (N). Genes with 1 or more insertions located within 5% to 80% of it's coding length were considered non-essential, except for relatively long ORFs (>1000 bp). These were asserted "ambiguous" if insertion density within the coding sequence was significantly below average (3.2 inserts per 1000 bp)

Potential sources of erroneous assertions.

1. Selected essential genes have been shown to tolerate transposon inserts within certain restricted loci without detrimental effect on the corresponding gene products in genetic footprinting experiments in both *E. coli* (Hare et al., 2001) and *H. influenzae* (Akerley et al., 1998). These may be erroneously asserted as non-essential in Table S1.

2. For proteins consisting of two or more independently functioning domains, inserts may be tolerated within the 3'-portion of the gene if the C-terminal domain of the protein it encodes is associated with a dispensable function. This can occur even when a function associated with the N-terminal domain (from the 5'-region of the gene) is genuinely essential (as with *ftsX* (Gerdes et al., 2002)).

ORFs were asserted as ambiguous (?), if the experimental evidence was insufficient to make specific conclusions about essentiality. ORFs shorter than 240 bp (<80 aa) with no inserts were asserted as "ambiguous", since the average insertion density obtained was 1 per 300 bp (exemplified by ORF IX in panel F). Relatively long ORFs with a single insertion (as ORFs X and XI in panel G). ORFs with a single insert further within the coding sequence then the 3'-most 20% or 5'-most 5% of the gene (as in case of ORF XII). ORFs containing only inserts corresponding to PCR products of low intensity (ORFs XIII).

F

G

H

I

References:

Akerley, B. J., Rubin, E. J., Camilli, A., Lampe, D. J., Robertson, H. M., and Mekalanos, J. J. (1998). Systematic identification of essential genes by in vitro mariner mutagenesis. Proc Natl Acad Sci USA *95*, 8927-8932.

Gerdes, S., Scholle, M., D'Souza, M., Bernal, A., Baev, M., Farrell, M., Kurnasov, O., Daugherty, M., Mseeh, F., Polanuyer, B., *et al.* (2002). From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. J Bacteriol *184*, 4555-4572.

Hare, R. S., Walker, S. S., Dorman, T. E., Greene, J. R., Guzman, L. M., Kenney, T. J., Sulavik, M. C., Baradaran, K., Houseweart, C., Yu, H. Y., *et al.* (2001). Genetic footprinting in bacteria. J Bacteriol *183*, 1694-1706.