

RESEARCH ARTICLE

Identification of insertion sites for the integrative and conjugative element Tn916 in the *Bacillus subtilis* chromosome

Emily L. Bean[‡], Janet L. Smith[‡], Alan D. Grossman^{*,}

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA.

[‡] These authors share first authorship on this work.

* adg@mit.edu



Abstract

Integrative and conjugative elements (ICEs) are found in many bacterial species and are mediators of horizontal gene transfer. Tn916 is an ICE found in several Gram-positive genera, including *Enterococcus*, *Staphylococcus*, *Streptococcus*, and *Clostridioides* (previously *Clostridium*). In contrast to the many ICEs that preferentially integrate into a single site, Tn916 can integrate into many sites in the host chromosome. The consensus integration motif for Tn916, based on analyses of approximately 200 independent insertions, is an approximately 16bp AT-rich sequence. Here, we describe the identification and mapping of approximately 10⁵ independent Tn916 insertions in the *Bacillus subtilis* chromosome. The insertions were distributed between 1,554 chromosomal sites, and approximately 99% of the insertions were in 303 sites and 65% were in only ten sites. One region, between *ykuC* and *ykyB* (*kre*), was a 'hotspot' for integration with ~22% of the insertions in that single location. In almost all of the top 99% of sites, Tn916 was found with similar frequencies in both orientations relative to the chromosome and relative to the direction of transcription, with a few notable exceptions. Using the sequences of all insertion regions, we determined a consensus motif which is similar to that previously identified for *C. difficile*. The insertion sites are largely AT-rich, and some sites overlap with regions bound by the nucleoid-associated protein Rok, a functional analog of H-NS of Gram-negative bacteria. Rok functions as a negative regulator of at least some horizontally acquired genes. We found that the presence or absence of Rok had little or no effect on insertion site specificity of Tn916.

Introduction

Integrative and conjugative elements (ICEs), also called conjugative transposons, are a type of self-transmissible mobile genetic element that facilitates horizontal gene transfer and contributes to bacterial evolution [1–5]. ICEs typically carry accessory (cargo) genes that benefit the host cell, including genes that confer antibiotic resistances, pathogenic or symbiotic determinants, or alternative metabolic pathways [3,6,7]. Typically, when an ICE is integrated in a host genome, the ICE genes needed for conjugation are not expressed. Either stochastically,

OPEN ACCESS

Citation: Bean EL, Smith JL, Grossman AD (2025) Identification of insertion sites for the integrative and conjugative element Tn916 in the *Bacillus subtilis* chromosome. PLoS One 20(5): e0318964. <https://doi.org/10.1371/journal.pone.0318964>

Editor: Arthur J. Lustig, Tulane University Health Sciences Center, UNITED STATES OF AMERICA

Received: January 25, 2025

Accepted: April 14, 2025

Published: May 16, 2025

Copyright: © 2025 Bean et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Data are available in dataset GSE286917 at NCBI Gene Expression Omnibus database (GEO, <http://www.ncbi.nlm.nih.gov/geo>).

Funding: Research reported here is based upon work supported, in part, by the National

Institute of General Medical Sciences of the National Institutes of Health under award numbers GM050895, GM122538, and GM148343 to ADG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."

Competing interests: The authors have declared that no competing interests exist.

or upon some environmental stimulus, an ICE can excise from the chromosome and the genes needed for conjugation are expressed. The element DNA can then be nicked and unwound to generate ssDNA that can then be transferred through the element-encoded conjugation machinery, a Type IV secretion system (T4SS), into a recipient cell. Once in the recipient, the ICE DNA eventually integrates into the chromosome of the new host to generate a stable transconjugant (for reviews, see [3,5]).

Tn916 was the first described ICE and was discovered because of its ability to transfer tetracycline resistance between isolates of *Enterococcus faecalis* [8,9]. Since its initial discovery, Tn916 and Tn916-like elements have been found in several Gram-positive host species, including *E. faecalis*, *C. difficile*, *Staphylococcus aureus*, and *Streptococcus pneumoniae* [4,10–15]. Additionally, Tn916 is functional and its biology has frequently been studied in *B. subtilis* (e.g., [16–19]).

ICEs encode an integrase (recombinase) that catalyzes integration into and excision out of a host chromosome. Many ICEs have a preferred integration (attachment) site in the chromosome, often in a conserved and essential gene [20]. If the preferred site is absent, secondary attachment sites can be used [21–24], analogous to temperate phages that can utilize secondary attachment sites [25] and non-conjugative transposons, e.g., Tn7 (reviewed in [26]).

Some ICEs, including Tn916, are promiscuous in target site selection and can integrate into many sites in a host chromosome [27–30]. Tn916 integrates into AT-rich regions in a host chromosome [27–30]. To date, the most extensive characterization of Tn916 integration site selection was completed in the context of *C. difficile* clinical isolates. Based on the sequences of approximately 200 independent Tn916 insertions the consensus insertion motif was defined as the AT-rich sequence 5'-TTTTTA[AT][AT][AT][AT]AAAA [30]. A similar site is used in *B. subtilis*, based on the sequences of two different insertions [31] and in *Butyrivibrio proteoclasticus* [29].

Here, we describe the sequence of sites of Tn916 integration in the *B. subtilis* chromosome from ~10⁵ independent Tn916 transconjugants. The ~10⁵ insertions were distributed between 1,554 chromosomal sites, with widely varying frequencies in each site. Ninety-nine percent of insertions were in 303 sites, 65% were in only ten sites, and 22% were in a single (hotspot) site. In almost all of the top 99% of sites, Tn916 was found with similar frequencies in both orientations relative to the direction of DNA replication and transcription (for insertions in open reading frames), with a few notable exceptions. Insertions were centered around an AT-rich consensus motif that is similar to that previously described for Tn916 insertions in *C. difficile* [30] and *B. proteoclasticus* [29].

We also determined the insertion site specificity of Tn916 in cells missing the nucleoid associated protein Rok. Rok binds to AT-rich regions of the chromosome [32,33] and is analogous to H-NS (reviewed in [34], found in many Gram-negative bacteria). Although *rok* itself is not conserved, functional analogs are found in other organisms [35]. We did not detect a global effect of *rok* on insertion site specificity of Tn916, especially at the more abundant insertion sites.

Results and discussion

Rationale and overview

We aimed to identify Tn916 integration sites in the *B. subtilis* chromosome and to determine if the nucleoid binding protein Rok substantially altered the integration site preferences. Based on previous analyses of Tn916 insertions in other organisms (e.g., [28–30]), we anticipated that insertions in *B. subtilis* would also be in AT-rich regions with a consensus similar (or identical) to that found in other species. These previous findings motivated our experiments to test the effects of Rok, a non-essential nucleoid-associated protein that preferentially binds AT-rich regions of the chromosome and can repress expression of some horizontally acquired genes [32,33], analogous to the roles of H-NS in Gram-negative bacteria [32,33,36–38]. We used high-throughput sequencing to identify Tn916 insertion sites in the *B. subtilis* genome and determined if Rok had any major effects on utilization of these sites.

Libraries of Tn916 insertions

We made libraries of approximately 10^5 independent Tn916 insertions in isogenic *B. subtilis* strains that contained a null mutation in *comK* (to prevent natural transformation) and that were otherwise *rok*⁺ or Δ *rok*. Briefly, a Tn916 donor (host) strain that was defective in competence development (*comK::mls*) and contained a copy of Tn916 integrated near the *yufK* start codon (strain ELC43) was crossed with two recipient strains that did not already contain a copy of Tn916 (*rok*⁺ Δ *comK::spc* strain ELC38 and Δ *rok* Δ *comK::spc* strain ELC42) and transconjugants that obtained Tn916 were selected (Methods). These *B. subtilis* strains are all cured of ICEBs1 and were derived from strain JMA222 (Methods). The conjugation efficiencies using wild type and Δ *rok* recipients were both $\sim 0.02\%$ transconjugants per donor. Transconjugants from multiple matings were recovered from petri plates, pooled, and then grown in LB medium to stationary phase before harvesting genomic DNA to map Tn916 insertion sites by high throughput sequencing.

Identification of insertion sites

We identified the Tn916 insertion (integration) sites in transconjugants using a transposon-directed insertion sequencing method (TraDIS) [30,39–41]. Briefly, genomic DNA from the pooled transconjugants (*rok*⁺ and Δ *rok*, separately) was randomly sheared, adapters were annealed to the ends of the fragments, and the junctions between the left end of Tn916 (as drawn in Fig 1A) and the chromosome were amplified by PCR. A Tn916-specific primer that hybridizes between *orf24* and *attL* was used for sequencing (Fig 1B, Methods). The relative number of sequence reads for a given site reflects the relative number of insertions in that site after growth of the cells, assuming that all other reactions are of equal efficiencies for each insertion. Approximately 87% of the sequence reads from each pool (*rok*⁺ and Δ *rok*) mapped to the *B. subtilis* chromosome and five percent of the total reads were from the excised circular form of Tn916. The remaining eight percent of sequences did not map to the *B. subtilis* genome, often due to poor sequence quality. We did not pursue these.

Insertions at any given genomic locus were often observed in both orientations and at multiple adjacent or nearby bases (e.g., Fig 1C). Because these nearby insertions are associated with the same insertion motif, we grouped them together into insertion sites (see Methods). For wild-type cells, insertions occurred at a total of 5377 nucleotide positions and clustered into 1554 insertion sites (Methods; S1 Table). The frequency of insertions in different sites varied over a wide range in both *rok*⁺ and Δ *rok* strains (Fig 2 and S1 Table).

We found that 99% of the Tn916 insertions in wild type cells were in 303 sites (S1 Table). Of these 303 sites, 59% were outside of annotated open reading frames (intergenic), although only $\sim 12\%$ of the genome is intergenic. Taking into account the frequency of insertions in each site, 75% of the insertions in these 303 sites were in intergenic regions. This bias is likely driven, in part, by the AT-rich nature of these regions (see below). A similar bias was observed for Tn916 insertions in *C. difficile* [30].

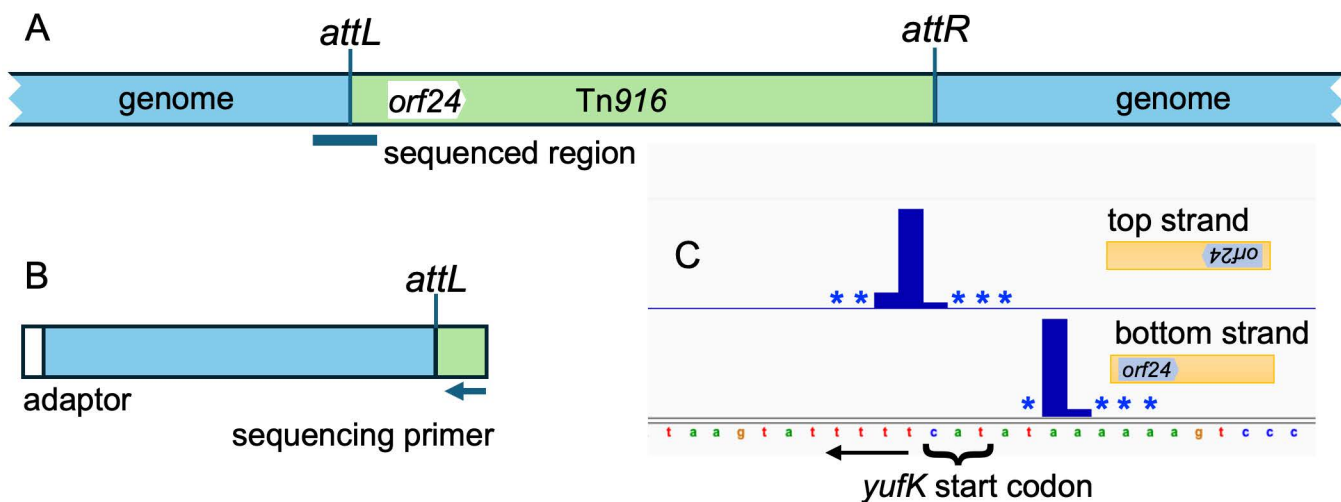


Fig 1. Schematic of Tn916 and identification of insertion sites. A. A generic insertion of Tn916 (green) in the chromosome (blue) is shown, including *attL*, *attR*, and *orf24* (the first open reading frame within Tn916). DNA for sequencing was amplified from the left end of Tn916 (as drawn) extending into the chromosome, as indicated below the insertion. B. Fragments that span the insertion site at the left end of Tn916 were amplified from the library of transconjugants, resulting in fragments approximately 300 bp in length. Amplification was carried out using an adaptor (shown in white) for the chromosomal end together with a primer that anneals within Tn916 (see Methods for details). The sequencing primer binds between *attL* and *orf24*. Reads that map to the bottom DNA strand correspond to insertions in the orientation drawn in panel A; reads mapping to the top strand are in the opposite orientation, i.e., with *orf24* the right. C. Tn916 insertions at *yufK*. Insertions occur in both orientations. Insertions with *orf24* to the left had DNA sequence reads on the bottom strand and insertions with *orf24* to the right had sequence reads on the top strand. Insertions were at multiple adjacent nucleotides and relative frequencies are indicated by the bar plots, with an asterisk indicating an insertion position used at a frequency too low to plot to scale.

<https://doi.org/10.1371/journal.pone.0318964.g001>

The Tn916 insertion site motif

We defined a consensus sequence motif for insertion of Tn916 in the *B. subtilis* chromosome using MEME [42]. We submitted 61-bp of DNA sequence surrounding the 303 most frequently used sites in wild type cells. The resulting consensus motif contained a palindromic region of T- and A-rich stretches separated by six base pairs that are AT-rich but less well conserved (Fig 3A). This motif has more sequence information, but is virtually the same as the consensus motif identified in other organisms [27–30].

Hotspots for Tn916 insertions

Remarkably, 65% of the independent Tn916 insertions were in ten sites (Table 1; Fig 2A), and 22% were in a site between *ykuC* and *ykyB* (*kre*). The *ykuC* gene product is likely a member of the major facilitator superfamily and null mutations increase the conjugation efficiency of ICEBs1 [43]. Null mutations in *ykyB* increase the stability of the *comK* mRNA [44]. The Tn916 insertions between these two genes are not known to disrupt either gene, and even if it happened to be polar on *ykyB*, there would be no effect on *comK* as this gene is missing in the strains used here.

The bias for insertions in the *ykuC/ykyB* site was not because the Tn916 from the donor strain had been in this site. The strain used as a donor (ELC43) had a single copy of Tn916 in a different site, the very 5'-end of *yufK* [19]. Tn916 integrated at the site in *yufK* in ~3% of transconjugants. We suspect that the bias for insertions in the *ykuC/ykyB* site could be due to the presence of a 17 bp inverted repeat that flanks the nucleotides where insertions occur (Fig 3B), and none of the other insertion sites had a 17 bp inverted repeat. We have not tested experimentally the importance of this inverted repeat on insertion frequency.

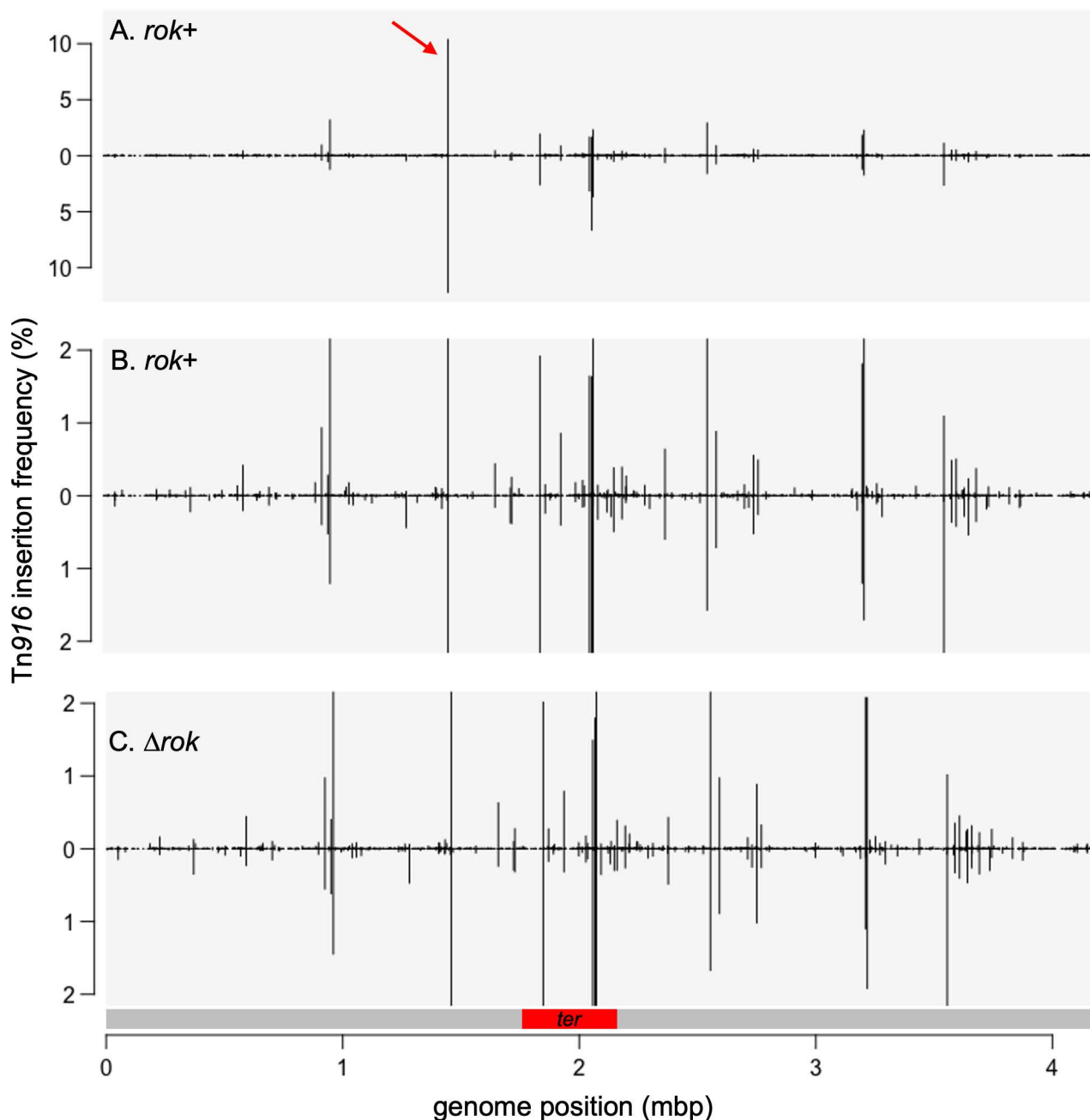


Fig 2. Relative frequencies of Tn916 insertions throughout the chromosome. The relative frequency of Tn916 insertions along the chromosome with the origin of replication at the left end and the terminus region near the middle are presented. Insertions on the top strand are shown above the line and those on the bottom strand below the line. The linear depictions of the chromosome begin and end at the origin of replication and the red bar below (C) indicates the *ter* region. A, B) Insertions from mating Tn916 into wild type strain (ELC38). Data are the same in each panel, but the scale in (B) is different to accentuate insertions in sites used at a low frequency. The red arrow in (A) indicates the insertion hotspot between *ykuC* and *ykyB* that contains ~22% of the total insertions. C) Insertions from mating Tn916 into the Δ *rok* strain (ELC42). Data are presented on the same scale as in B for comparison.

<https://doi.org/10.1371/journal.pone.0318964.g002>

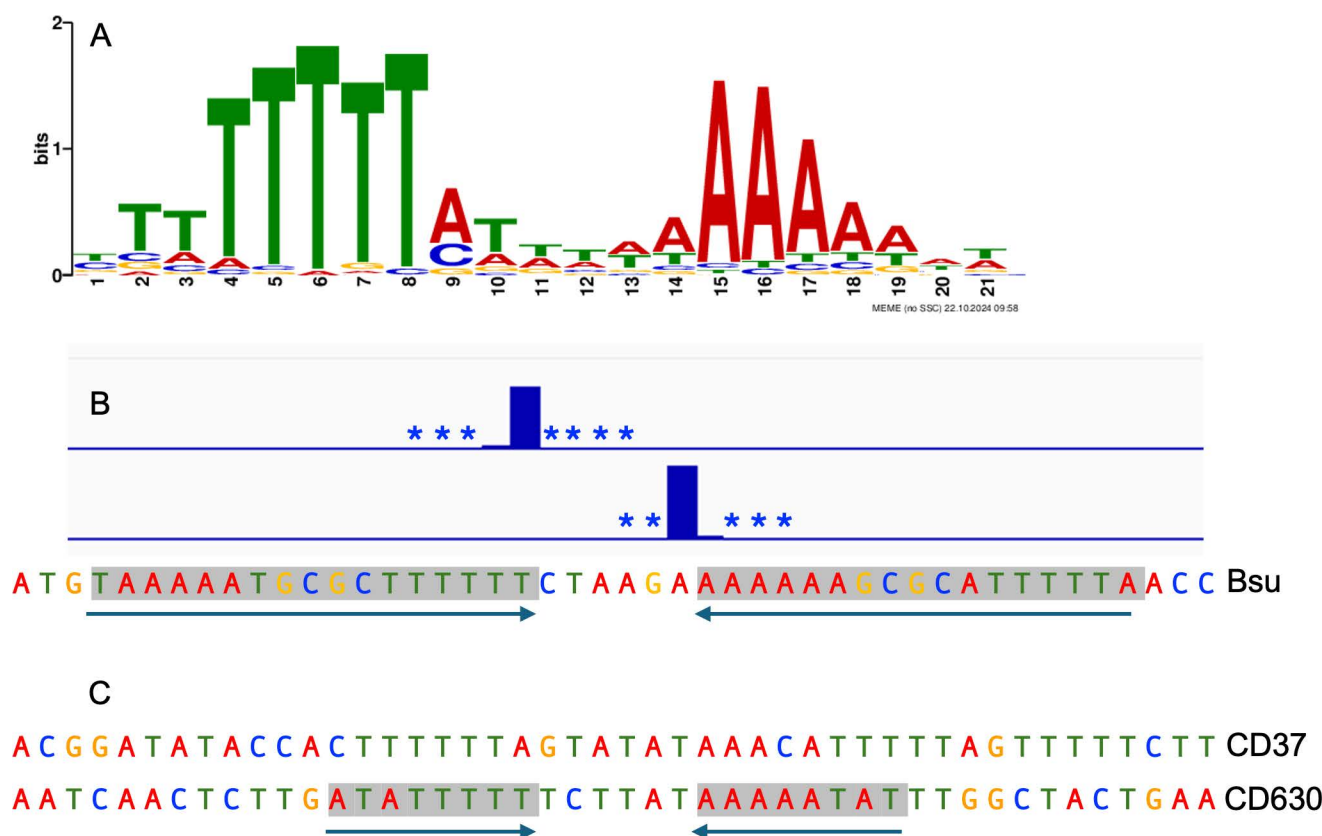


Fig 3. The consensus Tn916 insertion site motif. A) The consensus sequence logo for the top 303 Tn916 insertion sites in wild type cells. The consensus sequence is similar to that determined for Tn916 insertions in other organisms [29,30]. B) Tn916 insertions in the hotspot between *ykuC* and *ykyB* (*kre*), where 22% of insertions occur. Insertions in the reverse (top strand) or forward (bottom strand) are indicated by bar plots, with additional insertion positions present at frequencies too low to plot to scale indicated by asterisks. A 17-bp perfect inverted repeat is indicated by the arrows below the sequence and gray shading. C) Sequence of the Tn916 insertion hotspot in *C. difficile* strain CD37 [45,46] and in *C. difficile* strain 630 [30]. Shading and arrows indicate the 8 bp inverted repeat in the hotspot in CD630.

<https://doi.org/10.1371/journal.pone.0318964.g003>

The high frequency (~22%) of Tn916 insertions in a single site in the *B. subtilis* chromosome is intriguing when considered with the previously reported finding that Tn916 has one highly preferred integration site in the *C. difficile* strain CD37 (eight independent integrations in this site, one transconjugant from each of eight different conjugation experiments), despite the presence of multiple potential integration sites in the chromosome [45,46]. This hotspot is located downstream of *buk2* and does not have an inverted repeat. In *C. difficile* strain 630 there are a range of insertion frequencies at various sites, with insertions occurring most frequently near the 3' end of the *mgtA* coding sequence [30]. This insertion site is immediately flanked by a perfect 8bp inverted repeat. Although this is much shorter than the 17 bp inverted repeat at the *B. subtilis* hotspot, it is consistent with the possibility that inverted repeats can increase insertion frequency. There are many other possible explanations for the presence of the inverted repeat at the hotspot in one organism and not another. For example, the presence of the inverted repeat could be a random occurrence and not have any real significance. Additionally, features beyond the DNA sequence itself likely influence target site selection (see below).

Overall, our results are consistent with those previously reported [30,45,46], and together indicate that insertion site specificity is determined by a combination of sequence and other factors that influence DNA conformation [30,45,46]. Because the Tn916 recombinase (Int) has a preference for DNA sites with a static bend [47], the preferred sites in different organisms likely reflect that local DNA structure in combination with the sequence. Static bends are associated with

Table 1. Top ten Tn916 insertion sites in *B. subtilis*.

Nearest genes ^a	Genome position ^b	wt total frequency	rok total frequency	wt fraction on F strand	Notes ^c
downstream of <i>yhbH</i> ; upstream of <i>yhbI</i>	959987	4.4%	4.5%	0.72	
upstream of <i>ykyB</i> ; downstream of <i>ykuC</i>	1458673	22.6%	22.0%	0.46	
within <i>ymaC</i>	1847619	4.5%	5.0%	0.43	<i>ter</i>
within <i>yobW</i>	2057036	4.8%	4.4%	0.34	<i>ter</i>
upstream of <i>yocH</i> ; downstream of <i>yocI</i>	2066751	8.3%	8.0%	0.20	<i>ter</i>
downstream of <i>yocN</i> ; downstream of <i>yozO</i>	2072167	6.0%	6.1%	0.39	<i>ter</i>
upstream of <i>pstS</i> ; downstream of <i>pbpA</i>	2554501	4.5%	4.5%	0.65	
within <i>yufK</i>	3209746	3.0%	3.2%	0.60	
downstream of <i>nupQ</i> ; upstream of <i>maeN</i>	3217469	4.0%	4.0%	0.57	
upstream of <i>yvcA</i> ; downstream of <i>hisl</i>	3555502	3.7%	3.8%	0.29	

a The list is ordered by position in the chromosome. The *rok*⁺ and Δ *rok* strains had the same top ten insertion sites, but there are small differences in the order of the five sites with insertion frequencies of 4–5%.

b Genome position is the nucleotide number in the genome of strain AG174 [62].

c Sites in the *ter* region, where termination of DNA replication occurs, are indicated.

<https://doi.org/10.1371/journal.pone.0318964.t001>

polyA tracts [48], which are a feature in the Tn916 insertion motif, but additional sequence elements and DNA binding proteins may enhance the formation of static bends in these regions. These types of contributions beyond the obvious DNA sequence at the binding site have been collectively referred to as dynamic DNA-protein recognition [49]. Whatever the contributions beyond the specific DNA sequence, we believe that they are likely conserved between organisms and reflect the biochemical mechanisms required for the recombinase to function.

In addition to reflecting preferences for specific DNA sequences and structures, the frequency at which insertions in each site are recovered is also influenced by the selective pressures on the insertion mutants during growth. The vast majority of genes are not expected to have a significant effect on fitness under the rich media growth conditions used here [43,50–52], and the outgrowth used to prepare our library was likely not enough for small differences in fitness to have a large effect of the final population. Thus, we believe that the bulk of the differences we observe in the frequencies of insertion in different regions are due to DNA site preferences and not effects of the insertions on fitness.

Increased Tn916 insertion frequency in the terminus region of the chromosome

Four of the ten insertion sites that are utilized most frequently are in the ~10% of the chromosome that comprises the terminus (*ter*) region of replication (Fig 2A and 2B) [53]. As a result, this region (indicated by a red bar in Fig 2) has a 3.8-fold higher insertion frequency compared the rest of the chromosome. The matings were done during mid- and late-exponential phase in rich medium, and on average, there are approximately four copies of the origin region for each copy of the terminus region in these growth conditions. This indicates that the inherent preference for the sites in the *ter* region during exponential growth may be stronger than the observed frequencies indicate. These high frequency sites are not coincident with the any of the nine *ter* sites where termination actually occurs. Nor are they coincident with the *dif* site which is in the *ter* region and is used by the site-specific recombinases CodV and RipX to resolve chromosome dimers that can form during replication [54,55]. The *C. difficile* strain R20291 may also have an elevated frequency of insertions in the *ter* region [30], indicating that there might be a feature of this chromosomal region, separate from the *ter* and *dif* sites, that favors Tn916 insertions.

We considered the possibility that the bias for the *ter* region might be due to differences in the AT-content of this region compared to the rest of the chromosome. The *ter* region is ~59% AT and the remainder of the chromosome is 56% AT. We suspect that this difference has little if any effect on the insertion bias for the *ter* region in our experiments, but we have not further pursued this.

It is also possible that the bias for insertions in the *ter* region (and against more origin-proximal regions) has to do with selective pressures against having multiple copies of Tn916 in a cell. An active Tn916 can cause growth arrest and cell death [56]. During growth, and especially rapid growth, there is a higher copy of origin-proximal vs terminus-proximal regions, and there might be selective pressures that deplete strains with insertions in the origin-proximal region. We have not directly tested this hypothesis but think it unlikely as there did not appear to be a gradient of insertions that would reflect relative chromosomal copy number.

The bias for insertion in the *ter* region could also reflect a preference for non-replicating DNA. As with copy number considerations, in a population of growing cells, the origin-proximal regions are more actively undergoing replication than the terminus region. The differences could also be due to potential differences in DNA modification, if these affect insertion frequency. Any post-replicative DNA modification will be on only one strand of newly replicated DNA and then both strands sometime after replication. If any DNA modifications (e.g., methylation from restriction-modification systems) affect integration, either directly or by affecting DNA architecture, then these could conceivably affect site specificity.

Orientation bias at insertion sites

We analyzed the frequency of insertions on the forward versus reverse strand at each insertion site (Fig 4). For the top 99% of insertion sites (303 sites; black dots in Fig 4), 97% had insertions in both orientations. The insertion sites that had a strong orientation bias (≥ 4 -fold in one orientation) tended to have lower numbers of insertions. We suspect that in many cases, the apparent orientation bias is likely due to the relatively low frequency of insertions in these sites and is not an indication of the properties of the insertion site *per se*. In general, the insertion frequency in each orientation at individual sites was similar, especially for more abundant insertion sites. However, even among the 10 most abundant sites (Table 1) there were some that had more than a two-fold bias in one orientation, indicating that there might be features of individual sites that drive asymmetric insertion frequencies.

The orientation of Tn916 insertions was not strongly influenced by leading- vs lagging-strand DNA replication (Fig 2). In the region from the origin of replication (*oriC*) going clockwise to the beginning of the terminus (*ter*) region, insertions were equally distributed between forward (leading strand) and reverse (lagging strand) orientation, and in the region from *oriC* going counterclockwise to the start of the *ter* region, 52% of insertions had *orf24* oriented codirectionally with leading strand DNA synthesis (Fig 2A and 2B). The *ter* region was not included in this analysis because leading vs. lagging strand synthesis across this region is not clearly demarked.

Orientation of insertions relative to host transcription

To determine if there was a systematic orientation bias related to host transcription, we analyzed the orientation of Tn916 insertions in open reading frames. Of the top 303 insertion sites (99% of insertions), 124 were in an open reading frame (S1 Table). We found that 56% of the insertions in these 124 sites were oriented such that *orf24* was codirectional with the host ORF. Based on these results, we conclude that there appears to be no strong systematic bias for insertion orientation in transcribed portions of the chromosome.

Despite the absence of systematic bias, there appeared to be an orientation bias of more than four-fold in nine sites in open reading frames with an insertion frequency of at least 0.1% (Fig 5, Table 2). Three of the sites had a higher frequency of insertions with *orf24* codirectional with transcription of the host *orf*, and six sites had a higher frequency of insertions in the opposite orientation. These biases could be due to random fluctuations or might reflect inherent differences in DNA structure that influences integration into these sites.

There might also be selective pressures against insertions in a particular orientation that are driven by transcription, but that are specific to a given locus. For example, if the host gene is regulated by an antiterminator between the promoter and insertion site, there might be read-through of the terminator at the left end of Tn916, which would cause a competitive disadvantage due to expression of genes involved in unwinding and rolling circle replication [57]. In addition, transcription

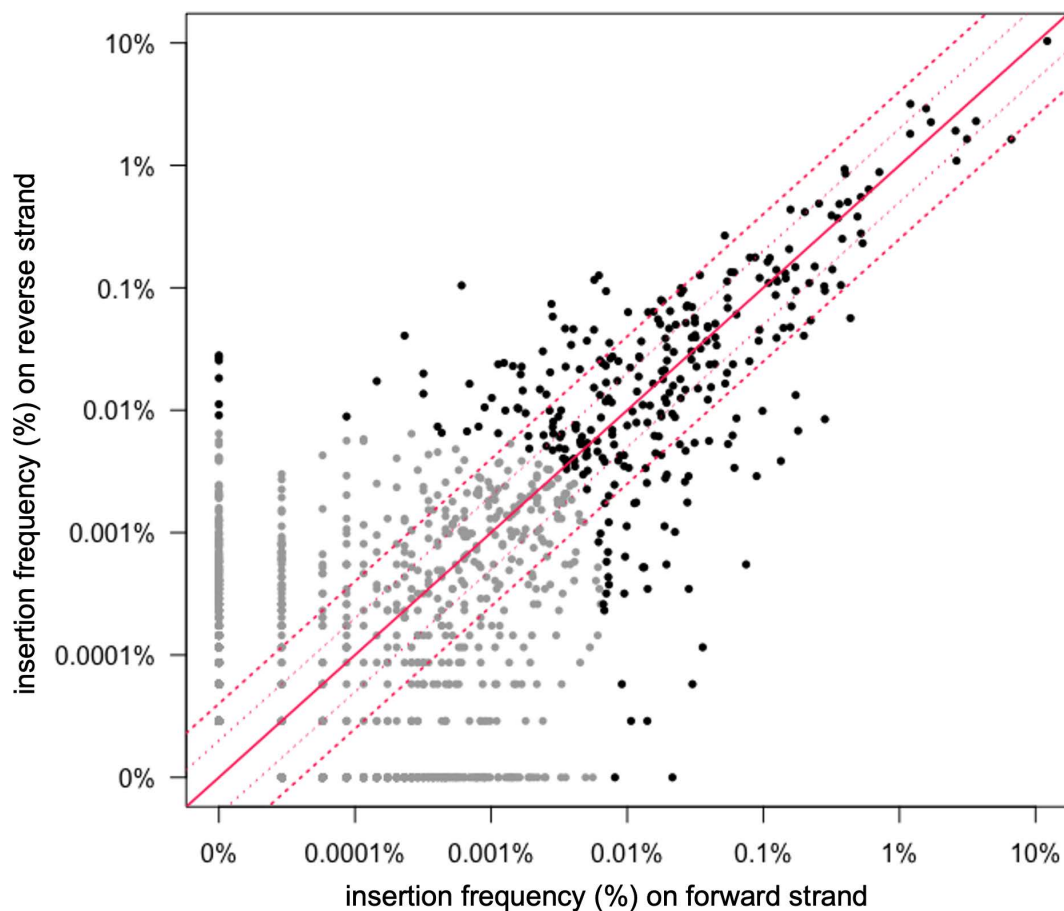


Fig 4. Orientation and frequency of insertions in each site. The frequency of insertions in each orientation in the reverse vs forward directions are plotted for all 1,554 insertion sites in wild type cells. The 303 most frequently used sites are shown in black and the rest are in gray. The solid red diagonal line indicates an equal number of insertions in each orientation. The dotted red lines indicate a 2- and 4-fold difference in abundance.

<https://doi.org/10.1371/journal.pone.0318964.g004>

from promoters at the right end of Tn916 (*Porf7*, *Pxis*, *Pint*) that are directed into the chromosome might cause expression of regions downstream of the insertion site (with *int* distal to the host promoter) or antisense transcription of regions upstream the insertion site (with *int* proximal to the host promoter). If such transcription is deleterious, it would be gene/insertion specific.

Evaluation of Rok effects on insertion site selection

Rok is a nucleoid-associated protein that binds AT-rich sequences [32,33] and could potentially affect the ability of Tn916 to integrate into specific sites. We compared the frequencies of Tn916 insertions in all sites in a *rok* null mutant vs. wild type (*rok*⁺), and also specifically analyzed insertions in sites that are in known Rok binding regions [33]. In general, there were similar insertion frequencies in most sites in *rok*⁺ and Δ *rok* strains (Fig 6), especially in sites with higher insertion frequencies. Thirteen of the insertion sites were in known Rok binding regions (Fig 6, red circles). A few of these had different insertion frequencies in Δ *rok* and *rok*⁺ cells, but no consistent pattern was observed as to whether there was a higher insertion frequency in Δ *rok* or *rok*⁺ cells. Since the sites that had a larger difference in the two strains were sites used at a relatively low frequency, they have a higher chance of differing due to random fluctuation. Based on these analyses, we

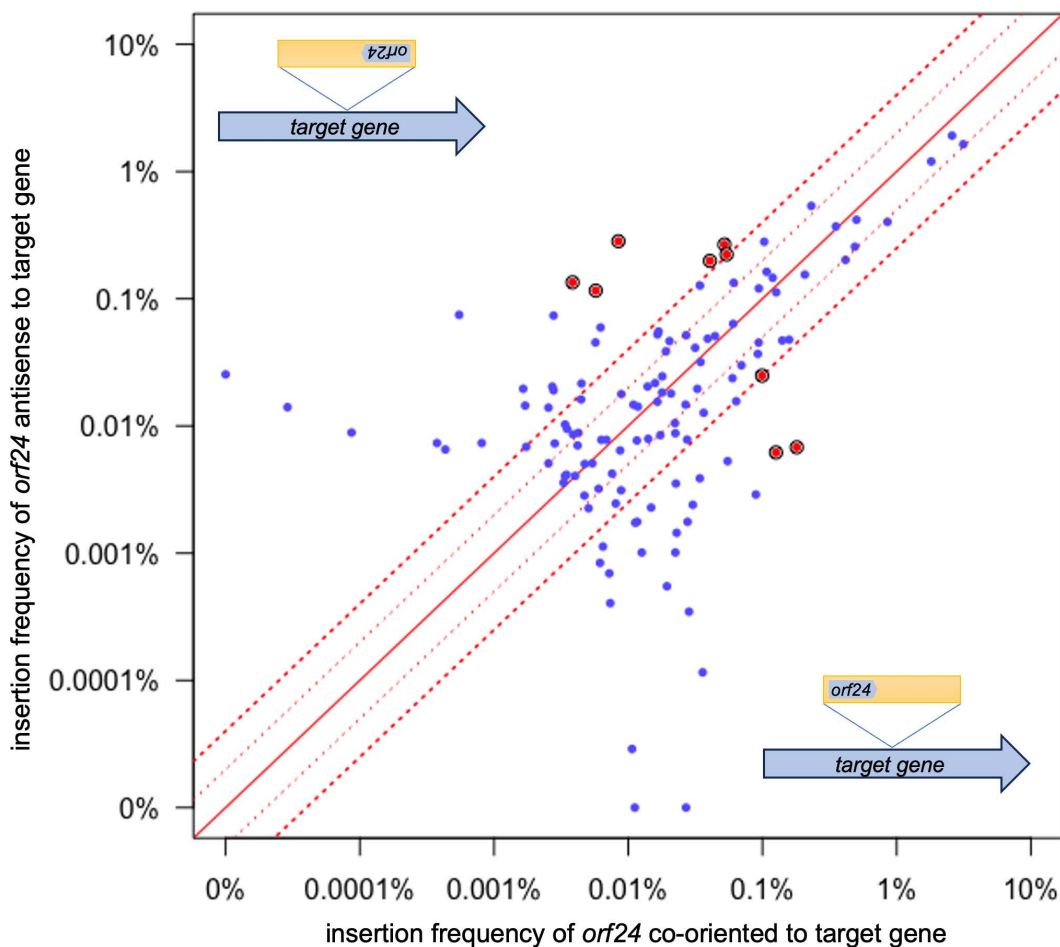


Fig 5. Insertion orientation in sites within ORFs. Of the top 303 insertion sites, 124 are within *orfs*. The frequencies of Tn916 insertions in each of these sites with *orf24* codirectional (x-axis) or antisense (y-axis) to transcription of the host gene are plotted. The solid red diagonal line indicates an equal number of insertions in each orientation. The dotted red lines indicate a 2- and 4-fold difference in abundance. The filled red circles indicate sites that had ≥ 4 -fold strand bias and an insertion frequency of $\geq 0.1\%$. These are also listed in Table 2.

<https://doi.org/10.1371/journal.pone.0318964.g005>

conclude that Rok does not substantially or systematically affect the insertion site utilization of Tn916. We have not determined if Rok affects excision of Tn916 from the chromosome. If there are any effects on excision, we suspect that they would occur at a small number of sites.

Conclusions and perspectives

Overall, we identified >1900 unique Tn916 insertion sites in the *B. subtilis* chromosome. We did not observe a systematic influence of Rok on integration site selection. Despite the relatively large number of independent insertions analyzed, the insertion sites were not saturated. It could prove interesting to evaluate the integration site distribution after removing the high-frequency insertion site from the chromosome (between *ykuC* and *ykyB*), and perhaps others, to favor the insertion into the lower-frequency sites. Additionally, the utility of Tn916 as a potential mutagen has been previously explored and the limitations noted [16,30,58,59]. The preference exhibited in target site selection in *C. difficile* and *B. subtilis* indicates that Tn916 insertion sites lack the desired randomness and diversity of target sites for random transposon mutagenesis across the chromosome.

Table 2. Insertion sites in open reading frames with at least 4-fold orientation bias.

Gene	Genome position	Gene strand	wt total frequency ^a	Percent with <i>orf24</i> codirectional ^b
<i>yheJ</i>	1028733	+	0.12%	5%
<i>mtnU</i>	1409015	–	0.12%	80%
<i>yosV</i>	2130878	–	0.28%	20%
<i>yorM</i>	2148167	–	0.29%	3%
<i>yomW</i>	2212241	+	0.32%	16%
<i>yugS</i>	3188952	–	0.24%	17%
<i>comP</i>	3228200	–	0.13%	95%
<i>clsA</i>	3735469	+	0.19%	96%
<i>sacT</i>	3879050	–	0.14%	3%

^a The sites included here have an insertion frequency of at least 0.1% of the total number of insertions and at least 80% of the insertions are in one orientation.

^b The percent of insertions in the indicated site with *orf24* from Tn916 codirectional with transcription of the host gene.

<https://doi.org/10.1371/journal.pone.0318964.t002>

Our analyses used only a single donor and a single variant of Tn916. That is, we analyzed transconjugants generated from Tn916 that came from only a single site in a single donor, and we analyzed only one isolate of Tn916 despite the natural variations observed in the family of elements [4,13,60]. It might be useful to compare results with transconjugants from a different donor site, and with different variants. However, most of the other studies used Tn916 from a different host and thus the donor element was in a site different from that in our work. Nonetheless, the consensus motif is virtually identical. Therefore, we believe that it is unlikely that analyses of transconjugants generated from elements originating in different sites will yield much, if any, new information.

One of the most interesting aspects of Tn916 integration is the presence of integration ‘hotspots’ in different organisms. The sequence of the hotspot in one organism is different from that in another (e.g., *B. subtilis* vs *C. difficile*), even though the same recombinase catalyzes the reaction and the consensus integration site is virtually the same in each organism. Consistent with other analyses, our findings indicate that there are likely multiple factors that contribute to the use of one site preferentially over others, perhaps including genome location and organization, local topology, and surrounding sequences.

Materials and methods

Growth conditions

Bacillus subtilis cells were grown in LB medium for strain constructions and experiments. The antibiotics tetracycline (12.5 µg/ml) and spectinomycin (100 µg/ml) were used as indicated. Growth was monitored by optical density at 600 nm (OD₆₀₀).

Strains and alleles

The *B. subtilis* strains used were all derived from JMA222 (*trpC2*, *pheA1*, ICEBs1⁰), a derivative of JH642 [61,62] that is cured of ICEBs1 [63]. ELC38 ($\Delta comK::spc$), ELC42 ($\Delta rok \Delta comK::spc$), and ELC43 {*att(yufKL)::Tn916* $\Delta comK::mIs$ } were made by natural transformation [64] and are described in more detail below. The *comK* mutants were used to prevent cells from acquiring DNA by transformation rather than conjugation during mating assays. Additionally, loss of *comK* suppresses the growth defect of *rok* mutants [65].

The Tn916 donor strain (ELC43) is derived from CMJ253 [43] and contains a single copy of Tn916 inserted at the beginning of *yufK*. CMJ253 was generated by natural transformation of JMA222 with genomic DNA from BS49 [19,31,43,66] and selecting for tetracycline resistance, indicating Tn916 acquisition. The *comK::mIs* allele in ELC43 is a deletion-insertion that was generated by replacing most of the *comK* open reading frame (from 47 bp upstream of *comK*

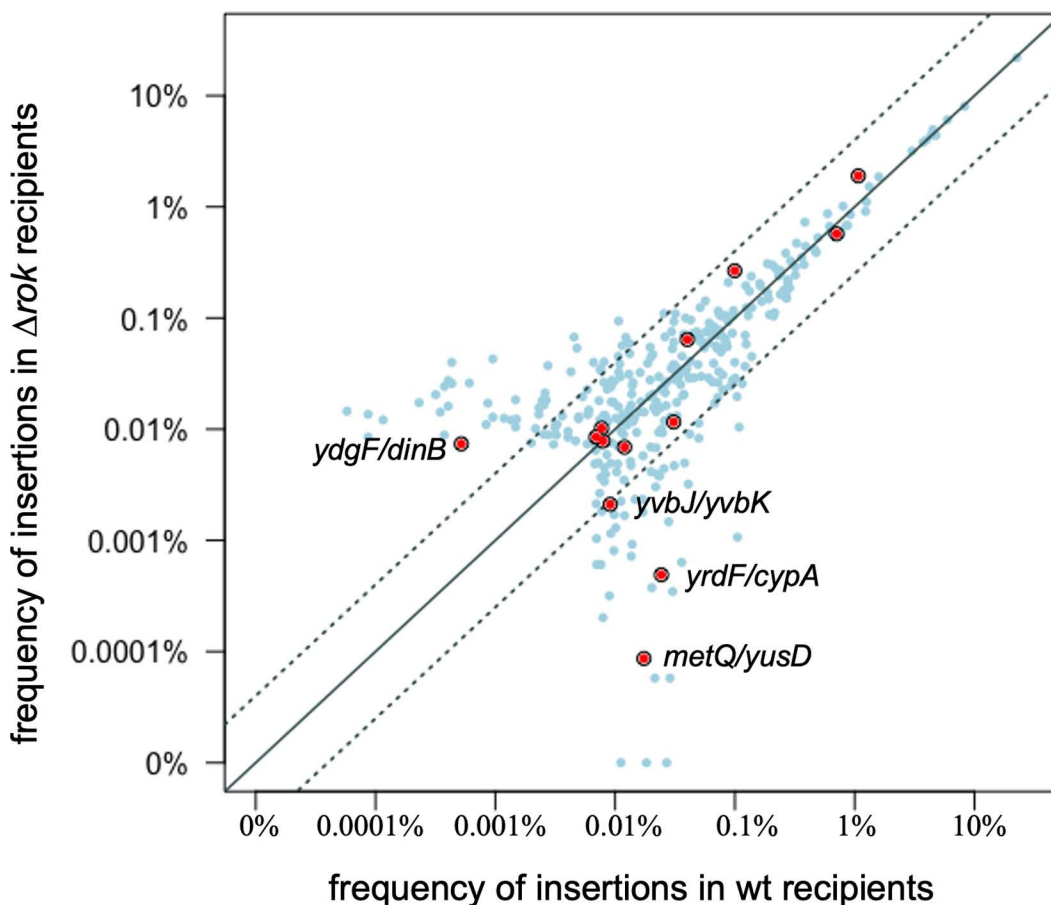


Fig 6. Comparison of insertion frequencies in wild type and Δrok strains. The frequency of insertions in each of the top 99% of sites in the Δrok mutant (302 sites) are plotted versus the frequency in wild type cells (303 sites). Filled red circles indicate Rok binding regions previously reported [33]. The solid diagonal line indicates an equal number of insertions in each strain. The dotted lines indicate a 4-fold difference in insertion frequency. The location of the four Rok binding sites that have >4-fold difference in insertion frequency in between the two strains is indicated with the gene names.

<https://doi.org/10.1371/journal.pone.0318964.g006>

to 19 bp upstream of its stop codon) with the resistance cassette from pDG795 (macrolide–lincosamide–streptogramin (MLS)). The cassette was combined with up- and downstream homology arms via isothermal assembly [67].

The $\Delta comK::spc$ allele in ELC38 and ELC42 was previously described [63,68]. The Δrok allele in ELC42 is an unmarked deletion of *rok* (from 1 bp upstream of *rok* to 11 bp upstream of its stop codon). First, *rok* was replaced with a chloramphenicol resistance cassette (*cat*) flanked by *lox* sites. *cat* was then removed from the chromosome by Cre-mediated recombination between the *lox* sites. *cre* was introduced by transforming cells with pDR244, a temperature-sensitive plasmid that expresses Cre, leaving an unmarked deletion with a small scar [43,69]. The plasmid was cured by growth at non-permissive temperature.

Generation of Tn916 insertion libraries and mating assays

Libraries of Tn916 insertions were made in both wild type (ELC38) and Δrok (ELC42) *B. subtilis* strains. Briefly, the Tn916 donor strain (ELC43; has Tn916 at the beginning of the *yufK* ORF) and recipients were grown in LB medium to mid- or late-exponential phase (OD_{600} of 0.5 and 1, respectively), then mixed at a 1:1 ratio (4 total OD_{600} units of cells) and applied

to a nitrocellulose filter. Filters were placed on a 1.5% agar plate with 1x Spizizen's salts [64] for 3 h at 37°C. Cells were harvested off the filters and plated onto 1.5% agar plates containing tetracycline and spectinomycin to select for transconjugants to be used for insertion sequence analysis. Transconjugants from three matings for each recipient were recovered from petri plates, pooled, and then grown in LB medium to stationary phase before harvesting genomic DNA to map Tn916 insertion sites by high throughput sequencing. In total we isolated approximately 10^5 independent transconjugants in each of *rok*⁺ and Δ *rok* strains. Conjugation efficiency (transconjugants per donor) is defined as the number of transconjugant CFUs obtained after mating per donor CFUs added to the mating mixture $\times 100\%$.

Harvesting transconjugant gDNA

Transconjugant colonies were resuspended from tetracycline/spectinomycin agar plates and used to inoculate a culture of LB medium containing 100 μ g/ml spectinomycin (to prevent growth of donors). Tetracycline was excluded during growth so as not to stimulate Tn916 excision from original integration sites [19,70–73]. Theoretically, any residual recipient cells that did not receive a copy of Tn916 could also grow in the absence of tetracycline, but these should not contribute to our Tn916-specific sequencing. Cells were grown to stationary phase ($OD_{600} \sim 2-3$) before harvesting by centrifugation. gDNA was prepared using a Qiagen 100 G tip purification kit. The isolated gDNA had an *oriC/ter* ratio of ~ 1 , determined as previously described [74], indicating that the measured frequencies of insertions in each site was not artificially affected by copy number variation due to ongoing DNA replication.

DNA sequencing

In general, the Transposon-Directed Insertion Site (TraDIS) sequencing pipeline was followed to prepare DNA for sequencing [30,39–41]. This strategy uses PCR to enrich for Tn916-chromosome junctions from randomly sheared gDNA. Approximately 5 μ g of gDNA from each library was sheared into ~ 300 bp fragments with a Covaris ultrasonicator. Sample sizes were confirmed by gel electrophoresis. Ampure SPRI cleanups were routinely used between each step of DNA preparation. The NEBNext End Repair kit was used for blunt-end repair, followed by the NEBNext dA-tailing kit.

A unique adapter (termed “splinkerette”) was generated to aid in the enrichment of transposon-chromosome junctions via PCR [39,41]. One strand of this adapter forms a stable hairpin to which a primer cannot anneal during PCR. Instead, a Tn916-specific primer must be used for the first round of amplification. In this way, there should be an enrichment of PCR products containing a Tn916-chromosome junction rather than chromosomal DNA alone. The so-called “splinkerette” DNA adapter was prepared by annealing oligonucleotides oELC97 (5'- P*CCACTAGTGT CGACACCAGT CTCTAATTTT TTTTTC AAAA AAAA; P* indicates that the 5' end of the primer is phosphorylated) and oELC98 (5'- CGAAGAGTAA CCGTTGCTAG GAGAGACCGT GGCTGAATGA GACTGGTGTGCGACACTAGTGG*T; *T indicates that the final T residue is attached by a phosphorothioate bond). Oligonucleotides were diluted to 200 μ M in 1mM tris-HCl pH 8.5 and mixed in equal volumes before placing in a 98°C heat block for 2 minutes, then cooled to room temperature for annealing to occur. T4 DNA ligase (NEB) was used to anneal the adapter to the dA-tailed DNA fragments.

Adapter-ligated DNA was PCR-amplified (19 cycles) with the Tn916-specific P5 primer oELC99 (5'- AATGATACGG CGACCACCGA GATCTACACT GAGTGGTTTT GACCTTGATAAAGTGTGATAAGTCCAGTTT TTATGCG) where the underlined portion anneals within Tn916, and the indexed TraDIS P7 primer oELC100.X (5'- CAAGCAGAAG ACGGCAT-ACG AGATXXXXXX CGAAGAGTAA CCGTTGCTAG GAGAGACCG) where XXXXXX is a barcode sequenced used to identify each library after pooling for sequencing. oELC101.X is unable to bind to its target sequence until after the first round of amplification performed by oELC99.

The resulting DNA was pooled and sequenced on a MiSeq (Illumina) to produce 47 bp reads spanning the Tn916-chromosome junction using custom sequencing primer oELC101.2 (5' - TGAGTGGTTT TGACCTTGAT AAAGTGTGAT AAGTCCAGTT TTTATGCGGA TAACTAGATT TTT). oELC102 (5' - CCACGGTCTC TCCTAGCAAC GGTTACTCTT CG) was used to sequence the attached barcode. The first 10 cycles were run “dark” because the transposon-specific

sequencing primer oELC101.2 anneals 10 bp from the end of Tn916, resulting in what would be a mono-template across all samples.

Analyses of DNA sequence reads

Mapping reads to the chromosome. In general, the “Bio::TraDIS pipeline” was followed for sequence analysis [39]. As described above, data were not collected for the first 10 rounds of sequencing because that portion is within Tn916 and would result in identical reads from all PCR products. The sequence reads thus begin with the coupling sequence (6 bp variable sequence at the left end of Tn916 that is derived from the donor) and therefore may not match the host target sequence [28,75,76]. We followed the approach of [30] and removed the first 10 bp of each sequencing read prior to mapping to eliminate potential mismatches with the recipient.

Reads were mapped to the AG174 (JH642) genome sequence [62]. The Tn916 host strains sequenced contained some engineered differences compared to AG174. Both sequenced strains are ICEBs1-cured and $\Delta comK::spc$, so these regions lack insertions because they were not present in our experiments. The *rok* allele differed in the two sequenced strains, but no insertions in *rok* were obtained in the *rok*⁺ strain (ELC38), so the different alleles per se did not alter any of the comparisons to the Δrok strain (ELC42).

The “bacteria_tradis” pipeline [39] was used to map reads, allowing up to 2 mismatches. Approximately 87% of reads in each condition mapped to the chromosome (~1.5 million for each pool), and ~5% of reads mapped to the circularized form of Tn916. For the combined data from both wild type and Δrok strains, insertions were observed at a total of 6977 nucleotide positions.

Clustering reads into insertion sites. The 6977 nucleotide positions where reads mapped tended to form clusters at discrete insertion motifs (Fig 1C). We often observed 1) reads mapping to both the top and bottom strand of the chromosome, which reflects insertions at that locus in both orientations, and 2) insertions at multiple closely spaced bases on the same strand, which is presumably because that the polyA and polyT arms of the insertion motif allow for some slippage in how the integrase binds.

We used custom R scripts to cluster reads that were associated with the same insertion site. We combined reads from the wild-type and Δrok samples for this. At insertion sites that had Tn916 inserted in both orientations, reads on the bottom strand mapped about 14 nucleotides to the left of reads that mapped to the top strand. This apparent 14 bp 3' overhang is consistent with a 6 bp 5' overhang when the mapped positions of the reads are corrected for the 10 bp that was removed from the 5' end prior to mapping to the chromosome. We did a first-pass assignment of reads into clusters by assigning any read that was more than 14 bp downstream of the previous read, regardless of strand, as a new insertion site, and then ran several diagnostics on this set. We found that for insertion sites that had insertions at more than one position on a strand, these insertions typically spanned 2–5 bps. Using these metrics we identified five chromosomal regions that contained closely spaced but distinct insertion sites that were not correctly processed in our automated approach, and resolved those manually, leading to a final set of 1944 insertion sites.

The frequency at which these 1944 insertion sites occur varied over six logs, from 2.9×10^{-7} (corresponding to sites with just one read) to 0.22 for the site between the *ykyB* and *ykuC* ORFs. The relative frequency of a Tn916 insertion in each identified site was estimated by the proportion of sequence reads mapping to that site, but it is not possible to accurately estimate insertion frequencies for sites with very few sequence reads. Although these low abundance insertion sites were not analyzed in detail, they are included in the full data set of insertions (Supplemental Table 1).

Annotation of insertion sites. We used the center of the insertion site to determine 1) the gene(s) which the insertion site was within or between, 2) whether the insertion site is within a Rok binding region as reported by [33], and 3) the DNA sequence surrounding the insertion site. For insertion sites with reads on both strands, we assigned the center as the midpoint between the position with the most reads on top strand and the position with the most reads on the bottom strand. For insertion sites with reads on only one strand, the center was determined by adding or subtracting seven (*i.e.*,

half the average distance between reads mapping to the top or bottom strand) from the position with the most inserts, depending on whether the reads were mapped to the bottom or top strand, respectively.

Binding motif and inverted repeats

Binding motifs were determined using MEME version 5.5.6 [42], which was run online at meme-suite.org. For the general insertion site motif (Fig 3A), we used 61 bp centered on the insertion site from the top 303 insertion sites. The sequences were unweighted, and both strands were searched, allowing for zero or one hit to be found per sequence. A first-order background set from [77] was used. To search for inverted repeats, we scanned 61 bp sequences centered on the insertion sites using an online version of Inverted Repeat Finder [78]. The default settings were used, except we reduced the minimum score to 20.

Supporting information

S1 Table. List of Tn916 insertion sites. Tn916 insertion sites are listed in an Excel file. The table includes the location of the insertion site, the gene(s) the insertion site is within or between, the relative frequency of insertions in both the wild type and the *rok* mutant, the fraction of insertions in each orientation, whether the site is in a known Rok binding region, and the DNA sequence of the insertion site.
(XLSX)

Acknowledgments

We thank Stuart Levine and the MIT BioMicro Center for technical assistance and high throughput DNA sequencing.

Author contributions

Conceptualization: Emily L. Bean, Janet L. Smith, Alan D. Grossman.

Data curation: Emily L. Bean, Janet L. Smith.

Formal analysis: Emily L. Bean, Janet L. Smith.

Funding acquisition: Alan D. Grossman.

Investigation: Emily L. Bean, Janet L. Smith.

Methodology: Emily L. Bean, Janet L. Smith.

Project administration: Alan D. Grossman.

Supervision: Alan D. Grossman.

Validation: Emily L. Bean, Janet L. Smith.

Visualization: Emily L. Bean, Janet L. Smith.

Writing – original draft: Emily L. Bean.

Writing – review & editing: Emily L. Bean, Janet L. Smith, Alan D. Grossman.

References

1. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: Evolution and diversity. *FEMS Microbiol Rev.* 2014;38(4):720–60. <https://doi.org/10.1111/1574-6976.12058> PMID: [24372381](https://pubmed.ncbi.nlm.nih.gov/24372381/)
2. Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev.* 2017;41(4):512–37. <https://doi.org/10.1093/femsre/fux008> PMID: [28369623](https://pubmed.ncbi.nlm.nih.gov/28369623/)
3. Johnson CM, Grossman AD. Integrative and Conjugative Elements (ICEs): What they do and how they work. *Annu Rev Genet.* 2015;49:577–601. <https://doi.org/10.1146/annurev-genet-112414-055018> PMID: [26473380](https://pubmed.ncbi.nlm.nih.gov/26473380/)

4. Roberts AP, Mullany P. A modular master on the move: the Tn916 family of mobile genetic elements. *Trends Microbiol.* 2009;17(6):251–8. <https://doi.org/10.1016/j.tim.2009.03.002> PMID: [19464182](#)
5. Wozniak RAF, Waldor MK. Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol.* 2010;8(8):552–63. <https://doi.org/10.1038/nrmicro2382> PMID: [20601965](#)
6. Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* 2005;3(9):722–32. <https://doi.org/10.1038/nrmicro1235> PMID: [16138100](#)
7. Treangen TJ, Rocha EPC. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 2011;7(1):e1001284. <https://doi.org/10.1371/journal.pgen.1001284> PMID: [21298028](#)
8. Franke AE, Clewell DB. Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of “conjugal” transfer in the absence of a conjugative plasmid. *J Bacteriol.* 1981;145(1):494–502. <https://doi.org/10.1128/jb.145.1.494-502.1981> PMID: [6257641](#)
9. Franke AE, Clewell DB. Evidence for conjugal transfer of a *Streptococcus faecalis* transposon (Tn916) from a chromosomal site in the absence of plasmid DNA. *Cold Spring Harb Symp Quant Biol.* 1981;45 Pt 1:77–80. <https://doi.org/10.1101/sqb.1981.045.01.014> PMID: [6271493](#)
10. Fitzgerald GF, Clewell DB. A conjugative transposon (Tn919) in *Streptococcus sanguis*. *Infect Immun.* 1985;47(2):415–20. <https://doi.org/10.1128/iai.47.2.415-420.1985> PMID: [2981772](#)
11. Clewell DB, An FY, White BA, Gawron-Burke C. *Streptococcus faecalis* sex pheromone (cAM373) also produced by *Staphylococcus aureus* and identification of a conjugative transposon (Tn918). *J Bacteriol.* 1985;162(3):1212–20. <https://doi.org/10.1128/jb.162.3.1212-1220.1985> PMID: [2987186](#)
12. Clewell DB, Flannagan SE. The Conjugative Transposons of Gram-Positive Bacteria. In: Clewell DB, editor. *Bacterial Conjugation*. Boston, MA: Springer US; 1993. pp. 369–393. [doi:10.1007/978-1-4757-9357-4_15](https://doi.org/10.1007/978-1-4757-9357-4_15)
13. Roberts AP, Mullany P. Tn916-like genetic elements: A diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev.* 2011;35(5):856–71. <https://doi.org/10.1111/j.1574-6976.2011.00283.x> PMID: [21658082](#)
14. Santoro F, Vianna ME, Roberts AP. Variation on a theme; An overview of the Tn916/Tn1545 family of mobile genetic elements in the oral and nasopharyngeal streptococci. *Front Microbiol.* 2014;5:535. <https://doi.org/10.3389/fmicb.2014.00535> PMID: [25368607](#)
15. Sansever EA, Robinson DA. Staphylococci on ICE: Overlooked agents of horizontal gene transfer. *Mob Genet Elements.* 2017;7(4):1–10. <https://doi.org/10.1080/2159256X.2017.1368433> PMID: [28932624](#)
16. Ivins BE, Welkos SL, Knudson GB, Leblanc DJ. Transposon Tn916 mutagenesis in *Bacillus anthracis*. *Infect Immun.* 1988;56(1):176–81. <https://doi.org/10.1128/iai.56.1.176-181.1988> PMID: [2826334](#)
17. Roberts AP, Hennequin C, Elmore M, Collignon A, Karjalainen T, Minton N, et al. Development of an integrative vector for the expression of anti-sense RNA in *Clostridium difficile*. *J Microbiol Methods.* 2003;55(3):617–24. [https://doi.org/10.1016/s0167-7012\(03\)00200-8](https://doi.org/10.1016/s0167-7012(03)00200-8) PMID: [14607405](#)
18. Scott JR, Kirchman PA, Caparon MG. An intermediate in transposition of the conjugative transposon Tn916. *Proc Natl Acad Sci U S A.* 1988;85(13):4809–13. <https://doi.org/10.1073/pnas.85.13.4809> PMID: [2838847](#)
19. Wright LD, Grossman AD. Autonomous replication of the conjugative transposon Tn916. *J Bacteriol.* 2016;198(24):3355–66. <https://doi.org/10.1128/JB.00639-16> PMID: [27698087](#)
20. Burrus V, Waldor MK. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol.* 2004;155(5):376–86. <https://doi.org/10.1016/j.resmic.2004.01.012> PMID: [15207870](#)
21. Wang H, Roberts AP, Lyras D, Rood JI, Wilks M, Mullany P. Characterization of the ends and target sites of the novel conjugative transposon Tn5397 from *Clostridium difficile*: Excision and circularization is mediated by the large resolvase, TndX. *J Bacteriol.* 2000;182(13):3775–83. <https://doi.org/10.1128/JB.182.13.3775-3783.2000> PMID: [10850994](#)
22. Wang H, Smith MCM, Mullany P. The conjugative transposon Tn5397 has a strong preference for integration into its *Clostridium difficile* target site. *J Bacteriol.* 2006;188(13):4871–8. <https://doi.org/10.1128/JB.00210-06> PMID: [16788196](#)
23. Lee CA, Auchtung JM, Monson RE, Grossman AD. Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *Mol Microbiol.* 2007;66(6):1356–69. <https://doi.org/10.1111/j.1365-2958.2007.06000.x> PMID: [18005101](#)
24. Menard KL, Grossman AD. Selective pressures to maintain attachment site specificity of integrative and conjugative elements. *PLoS Genet.* 2013;9(7):e1003623. <https://doi.org/10.1371/journal.pgen.1003623> PMID: [23874222](#)
25. Shimada K, Weisberg RA, Gottesman ME. Prophage lambda at unusual chromosomal locations. I. Location of the secondary attachment sites and the properties of the lysogens. *J Mol Biol.* 1972;63(3):483–503. [https://doi.org/10.1016/0022-2836\(72\)90443-3](https://doi.org/10.1016/0022-2836(72)90443-3) PMID: [4552408](#)
26. Craig NL. Tn7: A target site-specific transposon. *Mol Microbiol.* 1991;5(11):2569–73. <https://doi.org/10.1111/j.1365-2958.1991.tb01964.x> PMID: [1664019](#)
27. Clewell DB, Flannagan SE, Ike Y, Jones JM, Gawron-Burke C. Sequence analysis of termini of conjugative transposon Tn916. *J Bacteriol.* 1988;170(7):3046–52. <https://doi.org/10.1128/jb.170.7.3046-3052.1988> PMID: [2838457](#)
28. Scott JR, Bringel F, Marra D, Van Alstine G, Rudy CK. Conjugative transposition of Tn916: Preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. *Mol Microbiol.* 1994;11(6):1099–108. <https://doi.org/10.1111/j.1365-2958.1994.tb00386.x> PMID: [8022279](#)

29. Cookson AL, Noel S, Hussein H, Perry R, Sang C, Moon CD, et al. Transposition of Tn916 in the four replicons of the *Butyrivibrio proteoclasticus* B316(T) genome. FEMS Microbiol Lett. 2011;316(2):144–51. <https://doi.org/10.1111/j.1574-6968.2010.02204.x> PMID: [21204937](#)
30. Mullany P, Williams R, Langridge GC, Turner DJ, Whalan R, Clayton C, et al. Behavior and target site selection of conjugative transposon Tn916 in two different strains of toxigenic *Clostridium difficile*. Appl Environ Microbiol. 2012;78(7):2147–53. <https://doi.org/10.1128/AEM.06193-11> PMID: [22267673](#)
31. Browne HP, Anvar SY, Frank J, Lawley TD, Roberts AP, Smits WK. Complete genome sequence of BS49 and draft genome sequence of BS34A, *Bacillus subtilis* strains carrying Tn916. FEMS Microbiol Lett. 2015;362(3):1–4. <https://doi.org/10.1093/femsle/fnu050> PMID: [25673660](#)
32. Smits WK, Grossman AD. The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*. PLoS Genet. 2010;6(11):e1001207. <https://doi.org/10.1371/journal.pgen.1001207> PMID: [21085634](#)
33. Seid CA, Smith JL, Grossman AD. Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in *Bacillus subtilis*. Mol Microbiol. 2017;103(5):798–817. <https://doi.org/10.1111/mmi.13590> PMID: [27902860](#)
34. Grainger DC. Structure and function of bacterial H-NS protein. Biochem Soc Trans. 2016;44(6):1561–9. <https://doi.org/10.1042/BST20160190> PMID: [27913665](#)
35. Qin L, Erkelens AM, Ben Bdira F, Dame RT. The architects of bacterial DNA bridges: A structurally and functionally conserved family of proteins. Open Biol. 2019;9(12):190223. <https://doi.org/10.1098/rsob.190223> PMID: [31795918](#)
36. Dorman CJ. H-NS, the genome sentinel. Nat Rev Microbiol. 2007;5(2):157–61. <https://doi.org/10.1038/nrmicro1598> PMID: [17191074](#)
37. Fang FC, Rimsky S. New insights into transcriptional regulation by H-NS. Curr Opin Microbiol. 2008;11(2):113–20. <https://doi.org/10.1016/j.mib.2008.02.011> PMID: [18387844](#)
38. Navarre WW, McClelland M, Libby SJ, Fang FC. Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. Genes Dev. 2007;21(12):1456–71. <https://doi.org/10.1101/gad.1543107> PMID: [17575047](#)
39. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, et al. The TraDIS toolkit: Sequencing and analysis for dense transposon mutant libraries. Bioinformatics. 2016;32(7):1109–11. <https://doi.org/10.1093/bioinformatics/btw022> PMID: [26794317](#)
40. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. Genome Res. 2009;19(12):2308–16. <https://doi.org/10.1101/gr.097097.109> PMID: [19826075](#)
41. Potter CJ, Luo L. Splinkerette PCR for mapping transposable elements in *Drosophila*. PLoS One. 2010;5(4):e10168. <https://doi.org/10.1371/journal.pone.0010168> PMID: [20405015](#)
42. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28–36. PMID: [7584402](#)
43. Johnson CM, Grossman AD. Identification of host genes that affect acquisition of an integrative and conjugative element in *Bacillus subtilis*. Mol Microbiol. 2014;93(6):1284–301. <https://doi.org/10.1111/mmi.12736> PMID: [25069588](#)
44. Gamba P, Jonker MJ, Hamoen LW. A novel feedback loop that controls bimodal expression of genetic competence. PLoS Genet. 2015;11(6):e1005047. <https://doi.org/10.1371/journal.pgen.1005047> PMID: [26110430](#)
45. Mullany P, Wilks M, Tabaqchali S. Transfer of Tn916 and Tn916 delta E into *Clostridium difficile*: demonstration of a hot-spot for these elements in the C. difficile genome. FEMS Microbiol Lett. 1991;63(2–3):191–4. [https://doi.org/10.1016/0378-1097\(91\)90084-n](https://doi.org/10.1016/0378-1097(91)90084-n) PMID: [1647998](#)
46. Wang H, Roberts AP, Mullany P. DNA sequence of the insertional hot spot of Tn916 in the *Clostridium difficile* genome and discovery of a Tn916-like element in an environmental isolate integrated in the same hot spot. FEMS Microbiol Lett. 2000;192(1):15–20. <https://doi.org/10.1111/j.1574-6968.2000.tb09352.x> PMID: [11040422](#)
47. Lu F, Churchward G. Tn916 target DNA sequences bind the C-terminal domain of integrase protein with different affinities that correlate with transposon insertion frequency. J Bacteriol. 1995;177(8):1938–46. <https://doi.org/10.1128/jb.177.8.1938-1946.1995> PMID: [7721684](#)
48. Haran TE, Mohanty U. The unique structure of A-tracts and intrinsic DNA bending. Q Rev Biophys. 2009;42(1):41–81. <https://doi.org/10.1017/S0033583509004752> PMID: [19508739](#)
49. Fuxreiter M, Simon I, Bondos S. Dynamic protein-DNA recognition: beyond what can be seen. Trends Biochem Sci. 2011;36(8):415–23. <https://doi.org/10.1016/j.tibs.2011.04.006> PMID: [21620710](#)
50. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, et al. Essential *Bacillus subtilis* genes. Proc Natl Acad Sci U S A. 2003;100(8):4678–83. <https://doi.org/10.1073/pnas.0730515100> PMID: [12682299](#)
51. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, et al. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell. 2016;165(6):1493–506. <https://doi.org/10.1016/j.cell.2016.05.003> PMID: [27238023](#)
52. Koo B-M, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, et al. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. Cell Syst. 2017;4(3):291–305.e7. <https://doi.org/10.1016/j.cels.2016.12.013> PMID: [28189581](#)
53. Wake RG. Replication fork arrest and termination of chromosome replication in *Bacillus subtilis*. FEMS Microbiol Lett. 1997;153(2):247–54. <https://doi.org/10.1111/j.1574-6968.1997.tb12581.x> PMID: [9271849](#)
54. Sciochetti SA, Piggot PJ, Blakely GW. Identification and characterization of the dif Site from *Bacillus subtilis*. J Bacteriol. 2001;183(3):1058–68. <https://doi.org/10.1128/JB.183.3.1058-1068.2001> PMID: [11208805](#)

55. Gogou C, Japaridze A, Dekker C. Mechanisms for chromosome segregation in bacteria. *Front Microbiol.* 2021;12:685687. <https://doi.org/10.3389/fmicb.2021.685687> PMID: [34220773](#)
56. Bean EL, McLellan LK, Grossman AD. Activation of the integrative and conjugative element Tn916 causes growth arrest and death of host bacteria. *PLoS Genet.* 2022;18(10):e1010467. <https://doi.org/10.1371/journal.pgen.1010467> PMID: [36279314](#)
57. Wirachman ES, Grossman AD. Transcription termination and antitermination are critical for the fitness and function of the integrative and conjugative element Tn916. *PLoS Genet.* 2024;20(12):e1011417. <https://doi.org/10.1371/journal.pgen.1011417> PMID: [39652596](#)
58. Awad MM, Rood JJ. Isolation of alpha-toxin, theta-toxin and kappa-toxin mutants of *Clostridium perfringens* by Tn916 mutagenesis. *Microb Pathog.* 1997;22(5):275–84. <https://doi.org/10.1006/mpat.1996.0115> PMID: [9160297](#)
59. Lin WJ, Johnson EA. Transposon Tn916 mutagenesis in *Clostridium botulinum*. *Appl Environ Microbiol.* 1991;57(10):2946–50. <https://doi.org/10.1128/aem.57.10.2946-2950.1991> PMID: [1660696](#)
60. Ciric L, Jasni A, de Vries LE, Agersø Y, Mullany P, Roberts AP. The Tn916/Tn1545 family of conjugative transposons. *Bacterial Integrative Mobile Genetic Elements.* CRC Press. 2022. p. 153–70. <https://doi.org/10.1201/9780367813925-9>
61. Perego M, Spiegelman GB, Hoch JA. Structure of the gene for the transition state regulator, abrB: regulator synthesis is controlled by the spo0A sporulation gene in *Bacillus subtilis*. *Mol Microbiol.* 1988;2(6):689–99. <https://doi.org/10.1111/j.1365-2958.1988.tb00079.x> PMID: [3145384](#)
62. Smith JL, Goldberg JM, Grossman AD. Complete Genome Sequences of *Bacillus subtilis* subsp. *subtilis* Laboratory Strains JH642 (AG174) and AG1839. *Genome Announc.* 2014;2(4):e00663-14. <https://doi.org/10.1128/genomeA.00663-14> PMID: [24994804](#)
63. Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc Natl Acad Sci U S A.* 2005;102(35):12554–9. <https://doi.org/10.1073/pnas.0505835102> PMID: [16105942](#)
64. Harwood CR, Cutting SM. *Molecular Biological Methods For Bacillus.* Chichester, United Kingdom: John Wiley & Sons; 1990.
65. Hoa TT, Tortosa P, Albano M, Dubnau D. Rok (YkuW) regulates genetic competence in *Bacillus subtilis* by directly repressing comK. *Mol Microbiol.* 2002;43(1):15–26. <https://doi.org/10.1046/j.1365-2958.2002.02727.x> PMID: [11849533](#)
66. Haraldsen JD, Sonenshein AL. Efficient sporulation in *Clostridium difficile* requires disruption of the sigmaK gene. *Mol Microbiol.* 2003;48(3):811–21. <https://doi.org/10.1046/j.1365-2958.2003.03471.x> PMID: [12694623](#)
67. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* 2009;6(5):343–5. <https://doi.org/10.1038/nmeth.1318> PMID: [19363495](#)
68. Luttinger A, Hahn J, Dubnau D. Polynucleotide phosphorylase is necessary for competence development in *Bacillus subtilis*. *Mol Microbiol.* 1996;19(2):343–56. <https://doi.org/10.1046/j.1365-2958.1996.380907.x> PMID: [8825779](#)
69. Meisner J, Montero Llopis P, Sham L-T, Garner E, Bernhardt TG, Rudner DZ. FtsEX is required for CwO peptidoglycan hydrolase activity during cell wall elongation in *Bacillus subtilis*. *Mol Microbiol.* 2013;89(6):1069–83. <https://doi.org/10.1111/mmi.12330> PMID: [23855774](#)
70. Celli J, Trieu-Cuot P. Circularization of Tn916 is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol Microbiol.* 1998;28(1):103–17. <https://doi.org/10.1046/j.1365-2958.1998.00778.x> PMID: [9593300](#)
71. Manganelli R, Romano L, Ricci S, Zazzi M, Pozzi G. Dosage of Tn916 circular intermediates in *Enterococcus faecalis*. *Plasmid.* 1995;34(1):48–57. <https://doi.org/10.1006/plas.1995.1032> PMID: [7480170](#)
72. Showsh SA, Andrews RE Jr. Tetracycline enhances Tn916-mediated conjugal transfer. *Plasmid.* 1992;28(3):213–24. [https://doi.org/10.1016/0147-619x\(92\)90053-d](https://doi.org/10.1016/0147-619x(92)90053-d) PMID: [1334267](#)
73. Su YA, He P, Clewell DB. Characterization of the tet(M) determinant of Tn916: evidence for regulation by transcription attenuation. *Antimicrob Agents Chemother.* 1992;36(4):769–78. <https://doi.org/10.1128/AAC.36.4.769> PMID: [1323953](#)
74. Anderson ME, Smith JL, Grossman AD. Multiple mechanisms for overcoming lethal over-initiation of DNA replication. *Mol Microbiol.* 2022;118(4):426–42. <https://doi.org/10.1111/mmi.14976> PMID: [36053906](#)
75. Caparon MG, Scott JR. Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. *Cell.* 1989;59(6):1027–34. [https://doi.org/10.1016/0092-8674\(89\)90759-9](https://doi.org/10.1016/0092-8674(89)90759-9) PMID: [2557157](#)
76. Rudy CK, Scott JR. Length of the coupling sequence of Tn916. *J Bacteriol.* 1994;176(11):3386–8. <https://doi.org/10.1128/jb.176.11.3386-3388.1994> PMID: [8195096](#)
77. Zhang H, Li P, Zhong H-S, Zhang S-H. Conservation vs. variation of dinucleotide frequencies across bacterial and archaeal genomes: evolutionary implications. *Front Microbiol.* 2013;4:269. <https://doi.org/10.3389/fmicb.2013.00269> PMID: [24046767](#)
78. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.* 2004;14(10A):1861–9. <https://doi.org/10.1101/gr.2542904> PMID: [15466286](#)