1

# Autonomous replication of the conjugative transposon Tn*916*

3

Laurel D. Wright and Alan D. Grossman*

Department of Biology

Massachusetts Institute of Technology

Cambridge, MA 02139

8

9

Short title: Autonomous replication of Tn*916*

11

12

*Corresponding author:
Department of Biology
Building 68-530
Massachusetts Institute of Technology
Cambridge, MA 02139
phone: 617-253-1515
email: adg@mit.edu

26      **Abstract**

27      Integrative and conjugative elements (ICEs), also known as conjugative transposons, are self-

28      transferable elements that are widely distributed among bacterial phyla and are important drivers

29      of horizontal gene transfer. Many ICEs carry genes that confer antibiotic resistances to their host

30      cells and are involved in the dissemination of these resistance genes. ICEs reside in host

31      chromosomes, but under certain conditions can excise to form a plasmid that is typically the

32      substrate for transfer. A few ICEs are known to undergo autonomous replication following

33      activation.  However, it is not clear if autonomous replication is a general property of many

34      ICEs.  We found that Tn*916,* the first conjugative transposon identified, replicates autonomously

35      via a rolling circle mechanism. Replication of Tn*916* was dependent on the relaxase encoded by

36      *orf20* of Tn*916*. The origin of transfer of Tn*916*, *oriT*(*916*), also functioned as an origin of

37      replication.  Using immunoprecipitation and mass spectrometry, we found that the relaxase

38      (Orf20) and the two putative helicase processivity factors (Orf22 and Orf23) encoded by Tn*916*

39      likely interact in a complex and that the Tn*916* relaxase contains a previously unidentified

40      conserved helix-turn-helix domain in its N-terminal region that is required for relaxase function

41      and replication.  Lastly, we identified a functional single strand origin of replication (*sso*) in

42      Tn*916* that we predict primes second strand synthesis during rolling circle replication._Together

43      these results add to the emerging data that show that several ICEs replicate via a conserved,

44      rolling circle mechanism.

45

46

47    **Importance**

48       Integrative and conjugative elements (ICEs) drive horizontal gene transfer and the spread of

49    antibiotic resistances in bacteria. ICEs reside integrated in a host genome, but can excise to

50    create a plasmid that is the substrate for transfer to other cells. Here we show that Tn*916,* an ICE

51    with broad host range, undergoes autonomous rolling circle replication when in the plasmid

52    form. We found that the origin of transfer functions as a double-stranded origin of replication

53    and identified a single stranded origin of replication.  It was long thought that ICEs do not

54    undergo autonomous replication.  Our work adds to the evidence that ICEs replicate

55    autonomously as part of their normal lifecycle, and indicates that diverse ICEs use the same

56    replicative mechanism.

57

58

59

60    **Introduction**

61    Integrative and conjugative elements (ICEs), also called conjugative transposons, are mobile

62    genetic elements that encode proteins that mediate transfer of the element from the host cell

63    (donor) to a recipient by conjugation.  ICEs often contain additional (cargo) genes that can

64    provide a selective advantage to the host cells (reviewed in 14, 40).  Most ICEs have been

65    identified based on their cargo genes and the phenotypes conferred.  For example, many ICEs

66    carry genes encoding antibiotic resistances.  The horizontal dissemination of ICEs and their

67    associated cargo genes is a major driver of bacterial genome plasticity and evolution and the

68    spread of antibiotic resistances (e.g., 1, 56, 57, 67).

69    ICEs are typically found integrated in a host chromosome and are passively inherited by

70    vertical transmission via chromosomal replication and partitioning. When integrated, most ICE

71    genes are repressed.  However, under certain conditions, or stochastically, ICE genes required

72    for excision and transfer are expressed and the ICE can excise from the genome. A site-specific

73    recombinase (integrase) catalyzes this excision and formation of a circular plasmid species that is

74    a substrate for conjugative transfer. All functional ICEs that use a type IV secretion system {i.e.,

75    ICEs outside of actinomyces (75)} encode an origin of transfer *oriT* and a cognate relaxase. The

76    relaxase nicks at a site in the *oriT* and becomes covalently attached to the 5' end of the DNA.

77    The nicked dsDNA is unwound and the relaxase attached to the ssDNA is transferred via a type

78    IV secretion system out of the donor and into a recipient cell to generate a transconjugant

79    (reviewed in 40, 82).  In the transconjugant, the relaxase catalyzes re-circularization of the

80    ssDNA, releasing a ssDNA circle and a free relaxase.

81    The DNA processing steps accompanying conjugative transfer are similar to the steps

82    underlying rolling circle replication of some plasmids and phages (reviewed in 41). Plasmid

83    rolling circle replication initiates when a relaxase encoded by the plasmid nicks at the origin of

84    replication in the double stranded DNA (the double strand origin) and covalently attaches to the

85    5' end of the nicked strand. The nicked dsDNA is unwound and unidirectional replication

86    proceeds around the circle from the free 3'-end. The relaxase re-circularizes the nicked strand,

87    releasing a ssDNA circle.  The ssDNA circle typically contains a single strand origin of

88    replication (*sso*) that enables priming for DNA synthesis that converts the ssDNA to dsDNA.

89        ICEs that transfer ssDNA through a type IV secretion system were long thought to lack the

90    ability to replicate autonomously (12, 14, 82). ICEs can be maintained as integrated

91    chromosomal elements, and it appeared that ICEs could rely exclusively on vertical transmission

92    for inheritance. Furthermore, it is difficult to detect ICE replication because activation and

93    excision of most ICEs occur in a small fraction of donor cells.

94        However, there is compelling evidence that at least two ICEs undergo autonomous

95    replication.  When de-repressed, ICE*Bs1* from *Bacillus subtilis* undergoes autonomous rolling

96    circle replication (46, 76, 83).  This replication initiates at the origin of transfer (*oriT*) after

97    nicking by the ICE*Bs1*-encoded relaxase NicK (46). Processive unwinding of the nicked DNA is

98    dependent on the host translocase PcrA and the ICE*Bs1*-encoded helicase processivity factor

99    HelP (76). ICE*Bs1* contains a single strand origin of replication (*sso*) that enables second strand

100   synthesis (83).  Recently, the ICE R391, a member of the SXT/R391 family of ICEs from *Vibrio*

101   *cholera* and *Providencia rettgeri,* was found to replicate autonomously in *Escherichia coli*, and

102   the relaxase and *oriT* of R391 are important for R391 replication (18).  The copy number of

103   circularized SXT is also greater than the number of chromosomal sites from which it excised

104   (13), indicating that SXT undergoes autonomous replication and that replication is a conserved

105   feature of the SXT/R391 family of ICEs. R391 also encodes a conserved, functional plasmid

106    partitioning system (18), also consistent with autonomous replication and segregation.  Other

107    ICEs and ICE-like elements may also be capable of autonomous, plasmid-like replication (16,

108    59, 77). However most ICEs activate and excise in a small fraction of cells (13, 19, 49, 51),

109    thereby hindering detection of replicating intermediates in population-based assays.

110         We were interested in determining if other ICEs found in Gram-positive bacteria are capable

111    of autonomous replication.  We focused on Tn*916* (Fig. 1A), the first conjugative transposon

112    identified (28, 29), and one of the most widely studied ICEs (62, 82). Tn*916* and related elements

113    (e.g., Tn*1545*) contain a gene conferring resistance to tetracycline, exhibit a broad host range

114    (23), are found in many clinical isolates of *Enterococcus faecalis, Clostridium difficile,* and

115    *Streptococcus pneumoniae* (reviewed in 63), and can function in *B. subtilis* (22, 66). Unlike

116    many ICEs (e.g., ICE*Bs1*, R391) that have a specific integration site, Tn*916* can integrate into

117    multiple sites in a host genome, with a preference for AT-rich regions (52, 65).

118         We found that Tn*916* is capable of autonomous replication in *B. subtilis*. Replication was

119    dependent on the relaxase encoded by Tn*916 orf20*.  In addition, we found that the conjugative

120    origin of transfer of Tn*916, oriT*(*916*), could also function as an origin of replication, and we

121    identified a functional *sso* in Tn*916.* Our results demonstrate that Tn*916* replicates autonomously

122    by rolling circle replication. These findings strengthen the model that many, and perhaps all,

123    functional ICEs undergo autonomous replication as part of their normal lifecycle.

124

125

126    **Materials and Methods**

127    **Media and growth conditions**

128    *Bacillus subtilis* cells were grown in LB medium or the MOPs-buffered S7$_{50}$ defined minimal

129    medium (37) as indicated.  Cultures of cells containing pHP13-derived plasmids were grown for

130    imaging and ChIP-qPCR in medium containing 2.5 µg/ml chloramphenicol to select for the

131    plasmid as described (83). Cells containing *myc*-tagged *orf20* alleles were grown in medium

132    containing 50 µg/ml spectinomycin to maintain the single-cross-over integrations. Where

133    indicated, tetracycline (2.5 µg/ml) was added to Tn*916*-containing cells to increase gene

134    expression and excision (20). Antibiotics were otherwise used at the following concentrations:

135    kanamycin (5 µg/ml), chloramphenicol (5 µg/ml), spectinomycin (100 µg/ml), tetracycline (10

136    µg/ml), and a combination of erythromycin (0.5 µg/ml) and lincomycin (12.5 µg/ml) to select for

137    macrolide-lincosamide-streptogramin (*mls*) resistance.

138    **Strains**

139    *E. coli* strains used for plasmid construction were AG1111 (MC1061 F' *lacI*$^q$ *lacZ*M15 Tn*10*)

140    (36) and TP611 (*thi thr leuB6 lacY1 tonA21 supE44 hsdR hsdM recBC lop-11 cya-610 pcnB80*

141    *zad*::Tn*10*) (31).

142    *B*. *subtilis* strains were derived from JH642 (*pheA1 trpC2*) (58, 73) and are listed in Table 1.

143    Most strains were derived from JMA222, a derivative of JH642 that was cured of ICE*Bs1* (2).

144    The *ssb-mgfpmut2* fusion is expressed from the *rpsF* promoter and PrpsF-*ssb-mgfpmut2* was

145    inserted by double crossover at *lacA*, as described previously (5). Strains were constructed by

146    natural transformation (34).

147    B. subtilis strains containing Tn*916*.  Tn*916* host strain LDW173 was generated by natural

148    transformation of strain JMA222 with genomic DNA from strain BS49 (33) and selecting for

149    resistance to tetracycline as previously described (39).  The Tn*916* genomic integration site was

150    mapped by inverse PCR essentially as described (45). As in the parental strain (10), Tn*916* is

151    integrated between chromosomal genes *yufK* and *yufL* at coordinate 3,209,748 (73). Tn*916* is

152    oriented such that transcription of *orf24* and *int* of Tn*916* is co-directional with that of *yufL*. This

153    insertion site is identical to one of the two Tn*916* insertion sites found in strain BS49 (10). The

154    second Tn*916* insertion in BS49 between *ykyB* and *ykuC* is not present in LDW173 and likely

155    was not transferred during transformation.

156        *orf20.*  We constructed two mutations in the *orf20* (relaxase) gene. 1) Δ*orf20*-631 is a

157    markerless deletion that fuses the first 90 codons of *orf20* with the *orf20* stop codon, deleting the

158    intervening 306 codons and preserving *oriT*(*916*)  (38). Two ~1 kb fragments containing DNA

159    flanking the deletion endpoints were PCR amplified and inserted into the BamHI and EcoRI sites

160    of pCAL1422 (a plasmid that contains *E. coli lacZ*) via isothermal assembly (30) as previously

161    described (76, 83). The resulting plasmid, pLW625, was integrated into the chromosome of

162    LDW173 (WT Tn*916*) via single-cross-over recombination. Transformants were screened for

163    loss of *lacZ,* indicating loss of the integrated plasmid, and PCR was used to identify a Δ*orf20*

164    clone.  2) The *orf20-3UAA* nonsense mutation replaces the third codon of *orf20,* GAA, with the

165    stop codon UAA. *orf20-3UAA* was also constructed by allelic replacement using essentially the

166    same strategy as for the Δ*orf20*-631 allele. Approximately 1 kb fragments containing DNA

167    flanking the point mutation site were PCR amplified using primers containing the G to T

168    mutation or its reverse complement. The two PCR products were inserted into the BamHI and

169    EcoRI sites of pCAL1422 via isothermal assembly, and the isothermal assembly product was

170    transformed directly into LDW173 (contains wild type Tn*916*) cells. Transformants were

171    screened for loss of *lacZ,* and mutants containing the G to T point mutation were identified by

172    sequencing. One such mutant (strain LDW853) was then used for experiments.

173         We fused various alleles of *orf20* to the LacI-repressible-IPTG-inducible promoter

174    Pspank(hy) and used these to test *orf20* function.  Constructs included:  Pspank(hy)-*orf20* (wild

175    type), present in strains LDW929 and LDW931;  Pspank(hy)-*orf20∆hth-myc* (missing the N-

176    terminal helix-turn-helix domain), present in strains LDW930 and LDW931. The wild type *orf20*

177    coding sequence begins at the CUG codon, whereas the *orf20∆hth* coding sequence begins at the

178    annotated AUG start codon (see below). Both the *orf20* and *orf20∆hth* expression constructs

179    include the 24 bases upstream of the presumed CUG start codon, including the putative ribosome

180    binding site ATTGGAGG. Both *orf20* and *orf20∆hth* were PCR amplified from LDW173

181    genomic DNA, and the PCR fragments were inserted into the SphI and SacI sites of pCJ80 by

182    isothermal assembly. pCJ80 contains Pspank(hy) and the repressor *lacI;* an *mls* cassette marker;

183    and flanking homology for insertion by double-crossover into the chromosome at *lacA*. The

184    alleles were inserted into the chromosome by double-crossover, producing *lacA*::{ (Pspank(hy)-

185    *orf20*) *mls*} or *lacA*::{ (Pspank(hy)-*orf20∆hth orf20*) *mls*}.

186         *orf20* and *orf20∆hth* were tagged with three *myc* epitopes at the C-terminus, producing

187    *lacA*::Pspank(hy)-*orf20-myc* alleles, by single-crosssover integration of a pCAL812-derived

188    plasmid as previously described (74). Briefly, ~1 kb of *orf20* encoding the C terminal end of the

189    protein that is common to both the wild type and *orf20∆hth* alleles was PCR amplified and

190    inserted into the XhoI and EcoRI sites of pCAL812 by isothermal assembly, resulting in plasmid

191    pLW920.  pLW920 was transformed into *lacA*::{Pspank(hy)-*orf20*} or *lacA*::{Pspank(hy)-

192    *orf20∆hth*} by selecting for the spectinomycin resistance gene on pLW920.

193      *att1*Tn*916.*  A copy of *att1*Tn*916* was inserted at *amyE* to make *amyE*::{(*att1*Tn*916*) *spc*} in

194      LDW737, a control strain for qPCR. *att1*Tn*916* was PCR amplified from DNA from LDW173

195      cells, which contain a small amount of excised Tn*916* circles. The PCR product was inserted via

196      isothermal assembly into the EcoRI and HindIII sites of pAS24, an *amyE* insertion vector that

197      contains a *spc* cassette.

198      Plasmids. Plasmids pLW805 and pLW859 are pUS19-based and contain *oriT*(*916*) and

199      *orf23, orf22* and *orf20* driven by Pspank. In pLW859, *orf20* is tagged with six histidine residues

200      at the 3' end. Pspank, *lacI* and an intervening multi-cloning site from pDR110 were inserted into

201      the HindIII and EcoRI sites in pUS19 to make pCAL799 (pDR110 is from D. Rudner, Harvard

202      Medical School) (4). *orf23* and *orf22* were PCR amplified from LDW173 and inserted into the

203      NheI and SphI sites in pCAL799 to make pLW521 (pUS19, Pspank-*orf23-orf22, lacI*). A

204      fragment encompassing *oriT*(*916*) (38) and *orf20* was PCR amplified from LDW173 and

205      inserted into the SphI site downstream of *orf22* in pLW521 to make pLW805 and pLW859. In

206      pLW859, the *his* tag was added to *orf20* with the downstream PCR primer.

207      We constructed pHP13 derivatives to test for SSO activity as previously described (83).

208      pLW868 (pHP13sso916) and pLW890 (pHP13sso916R) contain *sso916* in the functional and

209      reverse orientation, respectively, relative to the direction of leading strand DNA synthesis of

210      pHP13 (32). We PCR amplified a 663 bp fragment from 89 bp upstream of the 3' end of *orf19*

211      through the first 458 bp of *orf18,* and inserted the fragment into the BamHI and EcoRI sites in

212      pHP13.  *ssiA,* which forms a primosome assembly site, was PCR amplified from pHV1610-1

213      (11) to make pLW862 (pHP13ssiA).

214     **Tn*916* excision and replication**

215     We used qPCR to measure Tn*916* excision and replication (Fig. 1B). Excision was measured

216     using primers oLW542 (5'-GCAATGCGAT TAATACAACG ATAC) and oLW543 (5'-

217     TCGAGCATTC CATCATACAT TC) (Fig. 1B, primers a and d) to quantitate the vacant

218     insertion site *att1*. *att1* amplification was normalized to a control chromosomal region in *mrpG,*

219     which is 15 kb downstream of *att1*. *mrpG* was amplified with primers oLW544 (5'-

220     CCTGCTTGGG ATTCTCTTTA TC) and oLW545 (5'-GTCATCTTGC ACACTTCTCT C).

221     The copy number of the Tn*916* circle was measured with primer pair oLW526 (5'-

222     AAACGTGAAG TATCTTCCTA CAG) and oLW527 (5'-TCGTCGTATC AAAGCTCATT C)

223     (Fig. 1B, primers b and c) to quantitate the unique *att*Tn*916* junction formed via site-specific

224     recombination, and the average copy number of circular Tn*916* per cell was calculated by

225     normalizing *att*Tn*916* amplification to *mrpG*. To determine if Tn*916* was replicating, we

226     determined the ratio of the copies of circular Tn*916* to the copies of the excision site.

227     Copy numbers of *att*Tn*916* and *att1* were determined by the standard curve method (15).

228     Standard curves for *att*Tn*916, att1* and control chromosome locus *mrpG* were generated from

229     genomic DNA of LDW737, which contains one copy of each amplicon in the chromosome.

230     LDW737 contains an ectopic copy of *att*Tn*916* inserted at *amyE.* LDW737 does not contain

231     Tn*916* and therefore contains a copy of the unoccupied chromosome site *att1*. DNA for standard

232     curves was prepared from stationary-phase LDW737 with an *oriC/terC* ratio of 1.3, verifying

233     that *amyE*::*att*Tn*916, att1* and *mrpG* were represented in ~1:1:1 ratios.

234     **Determination of copy number of an *oriT*(916) plasmid**

235     Copy number of plasmid pLW805 {*oriT*(*916*), Pspank-*orf20-orf22-orf23*} was determined

236     essentially as described (76). We used primers oLW128 (5'-ATGGAGAAGA TTCAGCCACT

237    GC) and oLW129 (5'-GCCATTATGG ATTCGTCAGA GG) that are specific to *spcE* in the

238    plasmid and normalized the amount of *spcE* to that of the chromosomal locus *mrpG*. Strain

239    LDW818 contains pLW805 inserted into Tn*916* by single cross over and was used to represent a

240    plasmid copy number of one, and to generate standard curves to calculate plasmid copy number.

241        **Identification of the XRE-like helix-turn-helix domain**

242        We searched for conserved domains within *orf20* using the CUG start codon upstream of the

243    annotated AUG (Fig. 3A) using the NCBI Conserved Domain Database

244    (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The helix-turn-helix domain cd00093

245    was identified (E-value = 3.50e-10) along with the expected relaxase domain pfam02486. A

246    search for Orf20 homologues within ICEs using HMMER3::phmmer and the ICEberg database

247    yielded 101 hits, including relaxases containing the XRE-like helix-turn-helix domain (6).  A

248    HMMR3::phmmer search of reference proteomes yielded additional relaxase homologues that

249    encoded the XRE-like helix-turn-helix domain and are present in a variety of sequenced Gram-

250    positive species (http://www.ebi.ac.uk/Tools/hmmer/).

251        **Orf20-His6 purification and mass spectrometry**

252        Orf20-His6 was purified from *B. subtilis* cells containing pLW859 (*oriT*(*916*),  Pspank-*orf23-*

253    *orf22-orf20-his*$_6$, *spc*). Cells were grown to mid-exponential phase in LB medium containing 1

254    mM IPTG and 100 µg/ml spectinomycin. Cells from 250 ml of culture were pelleted, washed

255    with 1X phosphate-buffered saline (PBS), re-centrifuged and stored at -80°C.  Pellets were

256    thawed on ice, and lysed in 25 ml binding buffer (300 mM NaCl, sodium phosphate buffer 50

257    mM, pH=7.4) containing 0.1 mg/ml lysozyme and the protease inhibitor 4-(2-aminoethyl)

258    benzenesulfonyl fluoride hydrochloride (AEBSF) at 1 mM at 37° C for 15 minutes. Lysates were

259    sonicated for 5 minutes total (0.3 sec. pulses, 15% output), followed by centrifugation to remove

260    insoluble material.

261         Lysates were pre-incubated with 0.5 ml TALON Superflow cobalt resin (GE Healthcare) on

262    a rotating platform for 1 hr at 4°C. Resin and lystate were loaded onto a Poly Prep column

263    (BioRad), and Orf20-His6 was purified according to GE Healthcare's protocol for batch, gravity-

264    flow purification under native conditions. Elution fractions were exchanged into buffer 20 mM

265    Hepes 150 mM KCl pH 7.5 using PD-10 desalting columns (GE Healthcare).

266         Purified protein was precipitated using trichloroacetic acid (TCA) and solubilized in 8 M

267    urea, 100 mM Tris, pH 8.5. Cysteines were reduced and alkylated using Tris 2-carboxyethyl

268    phosphine HCL (TCEP) and iodoacetamide respectively. The sample was then digested for four

269    hours at 37°C with endopeptidase Lys-C (Roche). The sample was then diluted to 2M Urea and

270    digested with trypsin (Promega). Peptides were identified using the LTQ XL Ion trap mass

271    spectrometer (Thermo Fisher) using MudPIT and SEQUEST software as previously described

272    (78). Tandem mass spectra were searched against a database of predicted ORFs from the genome

273    of *B. subtilis str. 168* (NCBI accession ASM904v1) and from Tn*916* (accession U09422.1)*,*

274    including the predicted full length Orf20 sequence.

275         **Measurement of single strand origin activity**

276         We used live cell microscopy and ChIP-qPCR to analyze association of Ssb-GFP with

277    pHP13-derived plasmids containing DNA fragments with candidate *sso*s (83). ChIP-qPCR was

278    carried out as described (83). For imaging, cells were placed on a slice of agarose dissolved in

279    1X Spizizen's salts (2 g/l $(NH_4)SO_4$, 14 g/l $K_2HPO_4$, 6 g/l $KH_2PO_4$, 1 g/l $Na_3$ citrate-$2H_2O$, 0.2

280    g/l $MgSO_4$-$7H_20$) (34) essentially as described (3). Images were acquired on a Nikon Ti-E

281    inverted microscope using a CoolSnap HQ camera (Photometrics). GFP fluorescence was

282    generated using a Nikon Intensilight mercury illuminator through Chroma filter set 49002.

283    Image processing was performed using ImageJ software.

284    **Results**

285    **Rationale**

286    Excision of Tn*916* from a site in the chromosome yields two products (Fig. 1B):  1) the

287    vacated (empty) chromosomal attachment (integration) site, and 2) the circular plasmid form of

288    Tn*916*. We reasoned that if Tn*916* replicates autonomously, then the copy number of the circular

289    form following excision would be greater than that of the chromosomal location that was vacated

290    upon excision.  We used quantitative PCR (qPCR) to measure the amount of the empty

291    chromosomal integration site (called here *att1*) relative to a nearby gene (*mrpG*) and the amount

292    of the Tn*916* circle relative to the amount of *att1* (Materials and Methods).  We also reasoned

293    that if Tn*916* replicates by rolling circle replication, then the Tn*916* relaxase encoded by *orf20*

294    (Fig. 1A) would be required for replication.

295    **Excision of Tn*916* from a site in the *B. subtilis* chromosome**

296    We measured the excision frequency of Tn*916* from a chromosomal site in *B. subtilis* during

297    exponential growth and entry into stationary phase. *B. subtilis* strain LDW173 contains a copy of

298    Tn*916* inserted between chromosomal genes *yufK* and *yufL* (*malK*) (Materials and Methods). We

299    used quantitative PCR to monitor excision from the *yufK-yufL* integration site, *att1*, and

300    normalized *att1* amplification to that of a nearby chromosomal locus, *mrpG,* that should be

301    unaffected by excision. We considered a strain without Tn*916* to represent 100% excision (all

302    cells had *att1*) and we used this to generate standard curves to calculate excision frequencies.

303    (Materials and Methods).

304    There was a basal level of excision of Tn*916* in cells growing exponentially in rich (LB)

305    medium (Fig. 2A).  *B. subtilis* cells (strain LDW173) growing exponentially in LB medium were

306    diluted to a low density (OD600 ~0.05) in the absence or presence of tetracycline (see below). In

307    the absence of tetracycline we detected approximately 0.003 empty *att1* sites per *mrpG*, and

308    excision increased to approximately 0.01 *att1* sites per *mrpG* during exponential growth (Fig.

309    2A).  After cells entered stationary phase, the amount of *att1* per *mrpG* appeared to decline (Fig.

310    2A).  Based on these results, we infer that Tn*916* had excised from *att1* in ~0.3-1% of cells

311    during exponential growth.

312    **Excision of Tn*916* from a site in the *B. subtilis* chromosome is stimulated by tetracycline**

313    Excision of Tn*916* requires the expression of *int* and *xis* (49) and these genes are downstream

314    of *tetM* (encoding tetracycline resistance) (Fig. 1A).  The presence of tetracycline enhances

315    transcription of the genes needed for excision (20) and enhances excision of Tn*916* from the

316    chromosome of *Enterococcus faecalis* (48). Conjugative transfer of Tn*916* also increases in the

317    presence of tetracycline (20, 25, 48, 70), perhaps as a result of increased excision and/or

318    increased expression of the conjugation genes that become fused to the promoters driving *int* and

319    *xis* after excision and circularization.  Interestingly, the amount of excision and conjugation and

320    the amount of stimulation by tetracycline is different for different insertion sites (48).

321    To determine the effects of tetracycline on excision of Tn*916* from *att1* in *B. subtilis*,

322    exponentially growing cells were diluted to a low density into tetracycline (2.5 µg/ml), as

323    described above.  Growth in the presence of tetracycline caused an increase in the amount of *att1*

324    per *mrpG* compared to cells grown in the absence of tetracycline (Fig. 2A), indicating that the

325    presence of tetracycline caused an increase in the excision frequency.  This increase was most

326    apparent during entry into and during early stationary phase (Fig. 2A).  There was a decline in

327   the amount of *att1* later in stationary phase.  This decline could be indicative of death of cells

328   from which Tn*916* had excised, thereby causing a decrease in the amount of *att1* (present in ~1-

329   3% of the cells) with relatively little overall affect on the amount of *mrpG* (present in almost all

330   cells). Alternatively, the decrease in *att1* could indicate that Tn*916* reintegrated into *att1*.

331       **Autonomous replication of Tn*916***

332       To measure the amount of the circular Tn*916* after excision, we used a control that would

333   mimic 100% excision and a copy number of one Tn*916* circle per chromosome.  To generate this

334   control, we cloned the circular junction generated by excision of Tn*916* from *att1* (*att1*Tn*916*,

335   Fig. 1B) and integrated a single copy of this DNA into the *B. subtilis* chromosome. We

336   considered a strain with *att1*Tn*916* in single copy (and without any copies of Tn*916*; strain

337   LDW737) to represent 100% excision and a copy number of one.  This strain was used in the

338   determination of the copy number of the circular form of Tn*916* (Materials and Methods).

339       We used qPCR to measure the amount of the circular form of Tn*916* relative to the

340   chromosomal locus *mrpG* from the same samples used to determine the amount of *att1* (above).

341   We found that, similar to the amount of *att1*, the amount of the circular form of Tn*916* increased

342   during growth and reached a maximum as cells approached and entered stationary phase (Fig.

343   2B).  The increase in copy number was most dramatic early in stationary phase where there was

344   a peak of ~0.2 copies of the circular form of Tn*916* per copy of *mrpG* (Fig. 2B). This increase in

345   stationary was apparent following growth in the presence or absence of tetracycline, although the

346   copy number appeared to be ~2-fold higher in the presence of tetracycline (Fig. 2B).

347       We found that the copy number of the circular form of Tn*916* was greater than that of the

348   vacated chromosomal site (*att1*) (Fig. 2C), indicating that there might be autonomous replication

349   of the excised form of Tn*916*.  During exponential growth, this ratio was approximately 2-3 and

350    increased to ~10 circles per *att1* in stationary phase (Fig. 2C).  These results indicate that the

351    copy number of the circle was greater than that of the empty chromosomal site and that the

352    excised Tn*916* was most likely replicating autonomously.  The increased copy number in

353    stationary phase compared to exponential growth is probably due to continued replication and

354    decreased cell growth.

355        **Increased copy number of the Tn*916* circle was dependent on the relaxase encoded by**

356    **Tn*916* orf20**

357        Plasmids that use rolling circle replication require a plasmid-encoded relaxase (sometimes

358    called the initiator or replicase) that nicks DNA in the origin of replication (the double strand

359    origin or *dso*) to initiate unidirectional replication (41). Replicative relaxases are homologous to

360    conjugative relaxases (21, 26, 72). Some replicative relaxases can function in conjugation (47,

361    53, 69), and some conjugative relaxases can function in replication (46), thereby blurring the

362    distinction between replicative and conjugative relaxases.

363        We found that the conjugative relaxase of Tn*916* (encoded by *orf20*) was needed for

364    autonomous replication of Tn*916*. The copy number of the Tn*916* circle was significantly

365    reduced in the *orf20* mutant (Fig. 2B, C). There appeared to be a similar amount of excision in

366    the relaxase (Δ*orf20*) mutant as judged by the amount of *att1* compared to *mrpG* (Fig. 2A),

367    indicating that loss of the relaxase did not affect excision. However, the average copy number of

368    the circular form of Tn*916* Δ*orf20* relative to the excision site was 0.39 ± 0.06, and the copy

369    number remained relatively constant during the entire time course of the experiment (Fig. 2B,

370    C). This is considerably different from the ratio for the wild type Tn*916* (*orf20*+).  Together,

371    these results indicate that Tn*916* normally undergoes autonomous rolling circle replication after

372    excision from the chromosome and the Tn*916*-encoded relaxase is required for this replication.

373    **The Tn*916* origin of transfer *oriT* can function as an origin of replication**

374    If Tn*916* uses its relaxase for rolling circle replication from *oriT*, then we expected that *oriT*

375    and *orf20* could function to support replication of a heterologous plasmid that is otherwise

376    missing an origin of replication.  By analogy to ICE*Bs1* (76), we also expected that this

377    replication would also require the homologs of the ICE*Bs1* helicase processivity factor HelP that

378    are encoded by Tn*916 orf22* and *orf23*.

379    We found that the Tn*916* origin of transfer *oriT*(*916*) could function as an origin of

380    replication.  We cloned *oriT*(*916*) into a plasmid that is otherwise incapable of autonomous

381    replication in *B. subtilis*. The parent plasmid, pUS19 (4), contains a pUC origin that is not

382    functional in *B. subtilis* but is functional in *E. coli.* pUS19 also contains *spcE*, conferring

383    resistance to spectinomycin in *B. subtilis*. In addition to *oriT*(*916*), we cloned the genes *orf20*

384    (relaxase), and *orf22* and *orf23* (*helP* homologues) from Tn*916* to generate plasmid pLW805.  In

385    this plasmid, transcription of *orf20, orf22, and orf23* is controlled by the LacI-repressible-IPTG-

386    inducible promoter Pspank (Pspank-*orf23-orf22-orf20*), making expression of these genes

387    dependent on IPTG.

388    We transformed cells lacking Tn*916* with pLW805 {*oriT*(916), Pspank-*orf23-orf22-orf20*}

389    DNA (that had been isolated from *E. coli*) and selected for spectinomycin-resistant

390    transformants. Transformants were obtained in the presence of IPTG (enabling expression of

391    *orf20*, *orf22*, and *orf23*), but no transformants were obtained in the absence of IPTG.  These

392    results indicate that *oriT* was capable of supporting replication and that replication was likely

393    dependent on expression of the relaxase and perhaps the *helP* homologues. The plasmid copy

394    number was approximately $4 \pm 1$ per cell as determined by qPCR of *spcE* plasmid DNA relative

395    to the chromosomal locus *mrpG.*

396    The *oriT*(*916*) plasmid (pLW805) was unstable, even when cells were grown in IPTG and

397    spectinomycin.  The fraction of plasmid-containing cells was determined by counting colony

398    forming units (CFUs) on LB agar plates containing IPTG and spectinomycin or on LB agar

399    without additives. After 7-8 generations of exponential growth in liquid LB medium with IPTG

400    and spectinomycin, approximately $17 \pm 6\%$ of cells were able to form colonies on LB plates with

401    IPTG and spectinomycin (the total number of cells was determined by CFUs on LB plates with

402    neither spectinomycin nor IPTG), indicating that the plasmid had been lost from ~83% of the

403    cells growing in culture.  This is not surprising for a plasmid that has a relatively low copy

404    number, replicates by rolling circle replication, lacks a single strand origin of replication, and

405    lacks partitioning functions.

406    The *oriT*(*916*) plasmid (pLW805) was even more unstable when cells were grown non-

407    selectively and in the absence of IPTG (causing decreased expression of *orf20, orf22,* and *orf23*).

408    Plasmid-containing cells were transferred to medium lacking IPTG and spectinomycin, and after

409    7-8 generations of exponential growth without inducer or selection, only 0.2% of cells were

410    resistant to spectinomycin, indicating that the *oriT* plasmid was lost in >99.5% of cells. These

411    results indicate that expression of the relaxase and perhaps the predicted helicase processivity

412    factors was needed for plasmid propagation.  Based on what is known about rolling circle

413    replication and the functions of the relaxase and helicase processivity factors (e.g., 76), and the

414    finding that the relaxase was needed for replication of Tn*916* (Fig. 2C), we conclude that

415    replication from *oriT*(*916*) was dependent on the relaxase (*orf20*) and probably at least one of the

416    HelP homologues (*orf22, orf23*). Results below indicate that the relaxase and both HelP

417    homologues are associated with the plasmid replicating from *oriT*(*916*).

418   **Analysis of Orf20 reveals a conserved N-terminal helix-turn-helix domain**

419   Tn*916 orf20* (relaxase) is annotated to start with an AUG codon (Fig. 3A, B) (64). We

420   noticed that *orf20* lacks an obvious ribosome binding site (RBS) upstream of the putative start

421   codon. However, there is a potential ribosome binding site and CUG start codon upstream of the

422   annotated AUG start (Fig. 3B). The predicted protein generated using this CUG start includes a

423   helix-turn-helix (HTH) domain that is found in the xenobiotic response element (XRE)-like

424   family of DNA-binding proteins (e.g., the repressor Xre of the *B. subtilis* defective phage PBSX;

425   lambda cI and Cro) (NCBI accession cd00093) (50, 55, 80, 81).

426   The XRE-like helix-turn-helix domain is conserved in many homologues of the *orf20-*

427   encoded relaxase (see Materials and Methods), including the relaxases of Tn*916*-related elements

428   present in multi-drug-resistant *C. difficile* strain 630 (CTn*7*) and in pathogenic strains of *S.*

429   *pneumonia* (Tn*5253*), and relaxases present in putative mobile elements from several Gram-

430   positive species (35, 63, 67) (Fig. 4). We suspect that some *orf20* homologues were misannotated

431   based on the initial annotation of Tn*916 orf20*.  In the reference genomes (e.g., relaxase *orf26* in

432   CTn*1*; Fig. 4) there are sequences encoding a potential HTH motif in or immediately upstream of

433   the annotated start codon (8, 9, 67), consistent with the notion that the actual relaxase is larger

434   than that originally annotated. Other *orf20* homologues, including the relaxase NicK from

435   ICE*Bs1,* do not contain an XRE-like helix-turn-helix domain.

436   We postulated that the Tn*916* relaxase was larger than previously predicted and contained a

437   conserved helix-turn-helix domain.  To test this, we analyzed peptide fragments from purified

438   relaxase.  We fused a hexahistidine tag to the C-terminus of the relaxase (Orf20-his) in the

439   *oriT*(*916*) plasmid (generating pLW859). Like the parent plasmid, the plasmid with *orf20-his*

440   was capable of autonomous replication in *B. subtilis* (strain LDW879)*,* indicating that Orf20-his

441   was functional. We purified Orf20-his from *B. subtilis* and analyzed the protein by mass

442   spectrometry.  We identified peptides from both the N-terminal helix-turn-helix and C-terminal

443   relaxase regions (Fig. 3A, Table 2). These results indicate that cells produce Orf20 starting with

444   the CUG codon and containing the helix-turn-helix region.

445        To verify that the helix-turn-helix region was part of the relaxase, we made a nonsense

446   mutation in this region of *orf20*. We mutated the third codon downstream of the presumed CUG

447   start codon to a stop codon (*orf20-3UAA*) (Fig. 3B). Like the *orf20* deletion, the *orf20-3UAA*

448   nonsense mutation abolished replication of Tn*916* (Table 3). Replication was restored to both

449   Tn*916Δorf20* and Tn*916orf20-3UAA* when full-length *orf20* (*orf20-myc*), starting with the CUG

450   codon and containing a C terminal *myc* tag, was expressed from Pspank(hy) (Table 3).

451   Abrogation of relaxase function with the *orf20-3UAA* mutation indicates that the annotated AUG

452   start codon in Tn*916* (Fig. 3B) does not initiate translation of a functional relaxase and that the

453   start codon is most likely upstream of the position of the nonsense mutation.

454        We also overexpressed the previously annotated *orf20,* missing the helix-turn-helix domain

455   (*orf20Δhth-myc*), from Pspank(hy). The Pspank(hy)-*orf20Δhth-myc* allele was unable to

456   complement the replication defects of relaxase-deficient Tn*916* (Table 3)*.* The simplest

457   interpretation of these results is that the helix-turn-helix domain of Orf20 is required for

458   replication.

459        Based on results above, we conclude that the actual *orf20* open reading frame contains the

460   helix-turn-helix motif found in many XRE-like proteins.  Furthermore, the open reading frame

461   most likely begins at the CUG codon that is proceeded by a potential ribosome binding site (Fig.

462   3). It seems reasonable to retain the name *orf20* for the Tn*916* gene encoding the conjugative

463   (and replicative) relaxase, recognizing that in some of the literature, this refers to the shorter

464   open reading frame, but in many cases, the exact coding sequence is not so important for the

465   genetic analyses.  It would also be reasonable to change the name of the Tn*916* gene for the

466   relaxase, perhaps calling it *nicK*, adopting the name used for the ICE*Bs1*-encoded relaxase.

467   Here, we continue to refer to the full length relaxase gene as *orf20*.

468   **Association of HelP homologues Orf23 and Orf22 with relaxase Orf20-his**

469   The plasmid replicating from *oriT*(*916*) with *orf20-his* (pLW859) also contained *orf23* and

470   *orf22* from Tn*916*, the homologues of ICE*Bs1 helP*. Mass spectrometry of affinity-purified

471   Orf20-his identified peptides from both Orf23 and Orf22 (Table 2). Co-purification of the HelP

472   homologues with functional relaxase indicates that the HelPs are part of the relaxosome and are

473   likely important for replication from *oriT*. These data are consistent with the model that Tn*916*

474   replicates by a rolling circle mechanism and uses helicase processivity factors to facilitate

475   unwinding of the DNA strands after relaxase nicking, analogous to autonomous replication of

476   ICE*Bs1*.

477   **Identification of a single strand origin in Tn*916***

478   Because Tn*916* replicates by rolling circle replication, we expected it would have a single

479   strand origin of replication. Rolling circle replicating plasmids and phages contain an *sso* or

480   encode a primase that enables conversion of ssDNA to dsDNA (42-44, 79). ICE*Bs1* has a single

481   strand origin that enables second strand synthesis (83).

482   We tested parts of Tn*916* for *sso* activity using a plasmid-based assay. pHP13 is a rolling

483   circle replicating plasmid that lacks an *sso* and accumulates ssDNA (7). In cells expressing a

484   fusion of the host single stranded DNA binding protein to GFP (Ssb-GFP), accumulation of

485   ssDNA can be visualized as large fluorescent foci of Ssb-GFP in most pHP13-containing cells

486   (Fig. 5A and B, strains CMJ118 without plasmid and CMJ129 with pHP13). We previously

487    found that when pHP13 contains *sso1* from ICE*Bs1*, there is a reduction in the size and intensity

488    of Ssb-GFP foci (83). Similarly, *ssiA* from pAMß, which is a primosome assembly site (11),

489    reduces pHP13 ssDNA (Fig 5C, pLW862, strain LDW872), showing that this fluorescent-

490    microscopy based assay, in conjunction with the pHP13 vector can be used to rapidly screen

491    single strand origins of multiple types (RNAP- vs. primase-dependent).

492        Most *sso*'s in RCR plasmids are found in intergenic regions and are orientation-specific (42).

493    Therefore, we cloned several intergenic regions from Tn*916* into pHP13 and screened for a

494    reduction in ssDNA accumulation as visualized by a reduction in the size of Ssb-GFP foci. One

495    of the regions we screened, encompassing the intergenic region between *orf19* and *orf18* (Fig.

496    1A)*,* reduced ssDNA (pLW858, strain LDW878) (Fig. 5D).  In addition, we found that the Sso

497    activity of the '*orf19-orf18*' region was orientation-specific. That is, the fragment cloned in the

498    opposite orientation into pHP13 (pLW890, strain LDW894) did not reduce the size or intensity

499    of Ssb-GFP foci (Fig. 5E). The predicted secondary structure of the sequence in this region did

500    not appear to resemble any of the three common types of *sso*'s, $sso_A$, $sso_U$, or $sso_T$ (42) whereas

501    *sso1* from ICE*Bs1* resembles that from pTA1060 (83) and belongs to the $sso_T$ family.

502        We quantified Sso activity of the '*orf19-orf18*' fragment (referred to as *sso916*) by

503    immunoprecipitating Ssb-GFP and determining the amount of plasmid DNA bound to Ssb

504    (ssDNA) using qPCR **(**Fig.6).  *sso916* (present in pLW858) reduced the amount of Ssb-bound

505    plasmid DNA ~30-fold, similar to *sso1* from ICE*Bs1* (83). In contrast, in cells containing

506    pLW890 (pHP13 with *sso916R*, *sso916* in the opposite orientation), the amount of plasmid DNA

507    bound to Ssb-GFP was similar to that of the parent plasmid (pHP13) without an *sso* (Fig. 6). The

508    differences in the amount of Ssb-GFP bound to each of the plasmids was not due to differences

509    in plasmid copy number.  The copy number of pLW858 (pHP13 with *sso916*) was approximately

510    the same as that of pHP13, and that of pLW890 (pHP13 with *sso916R*) was approximately 1.4-

511    fold that of pHP13.

512          We found that *sso916* is conserved in other Tn*916*-like ICEs. We searched for sequences

513    similar to the 116 bp intergenic region that contains *sso916* and found that 15 additional ICEs

514    contained a region with 94-100% sequence identity (Table 4). The conjugation genes from all of

515    these ICEs are highly similar (≥80% identity) to those in Tn*916*, but each contains accessory

516    (cargo) genes different from those in Tn*916*. Based on these results, we conclude that sso*916* is a

517    functional single strand origin of replication in Tn*916* that has been conserved during genetic

518    diversification of Tn*916*-like elements.

519

520

521          **Discussion**

522          **Tn*916* replicates autonomously**

523          We found that the broad-host range conjugative transposon Tn*916* undergoes autonomous

524    rolling-circle replication. The excised circular form of Tn*916* is multi-copy, and replication is

525    dependent on the relaxase encoded by *orf20*. The Tn*916* origin of transfer *oriT*(*916*) supports

526    replication of a plasmid that is otherwise incapable of replication in *B. subtilis*. Replication

527    appears to be dependent on the relaxase and at least one and perhaps both of the helicase

528    processivity factor homologues Orf23 and Orf22. Co-purification of the relaxase and both HelP

529    homologues indicates that both of the HelP homologues are likely functioning in DNA

530    unwinding.  We do not know if one or the other or both are required, nor if they are redundant.

531    Lastly, Tn*916* contains a functional *sso, sso916.* Our results support a model in which relaxase

532    Orf20 initiates rolling circle replication from *oriT*(*916*) and the HelP homologues facilitate

533   processive unwinding of the nicked strand, analogous to the role of the relaxase and HelP in

534   ICE*Bs1* (46, 76). After recircularization of the unwound strand, Sso activity would be used to

535   initiate priming of lagging strand DNA synthesis.

536       Tn*916* was thought to be incapable of autonomous replication. Previous studies may have

537   failed to detect replicating Tn*916* because, like many ICEs, Tn*916* excises in a small fraction of

538   host cells (19, 49). The circular form of Tn*916* was detected using Southern blotting when *xis*

539   and *int* were overexpressed (49). The ratio of Tn*916* circles per excision event was reported to be

540   1.8 Tn*916* circles per excision site. We observed a similar ratio when nutrients were non-limiting

541   (~1-3 circles per excision sites during exponential growth, Fig. 2C).

542       Tn*916* excision and copy number peaked during early stationary phase. Likewise, maximal

543   excision of Tn*916* in *E. faecalis* and *Listeria monocytogenes* occurs during late exponential

544   phase (19), consistent with the notion that activation of Tn*916* is dependent, in part, on cell

545   growth phase. Other ICEs also have growth-phase-dependent excision due to nutrient limitation,

546   in response to cell-cell signaling, or both (2, 17, 59, 60, 68).

547       **Identification of an N-terminal helix-turn-helix domain in the Tn*916*-encoded relaxase**

548       We identified a conserved helix-turn-helix domain in the N-terminal region of the relaxase

549   Orf20. This domain is conserved in many relaxase homologues and our results indicate that this

550   region is essential for relaxase function.

551       A purified form of Orf20 from Tn*916* nicks *oriT*(*916*) non-specifically *in vitro*, and co-

552   incubation with the recombinase Int then generated strand- and sequence-specific nicking (64).

553   However, because *orf20* was misannotated, Orf20 used in these experiments was purified

554   without the N-terminal helix-turn-helix domain. Our results indicate that Orf20 contains an N-

555   terminal helix-turn-helix domain.  Because *oriT*(*916*) functions as an origin of replication in the

556    absence of *int,* we suggest that Int is not involved in nicking *oriT*(*916*) and that the helix-turn-

557    helix domain in Orf20 likely facilitates recognition of *oriT*(*916*).

558        A class of conjugative relaxases from plasmids has recently been described that contain an

559    N-terminal helix-turn-helix motif (27), although the domain is not a member of the XRE-like

560    family present in Orf20. Mutation of a highly conserved glutamate residue in the helix-turn-helix

561    domain of representative relaxase TraX prevented relaxase binding to *oriT* (27), consistent with

562    the model that the N-terminal helix-turn-helix domain in Orf20 is needed for proper recognition

563    and nicking of *oriT*(*916*).

564        **Replication and maintenance**

565        Replication of ICE*Bs1* and R391 is required for maintenance of the elements in dividing host

566    cells (18, 46), and studies with other ICEs found that the relaxase is required for stability of the

567    cognate ICE (e.g., 59). However, we did not observe a significant loss of a Tn*916* Δ*orf20* mutant

568    (missing the relaxase).  This apparent stability could indicate that the circular form of Tn*916*

569    might cause growth arrest or possibly cell death. No genes in Tn*916* have been identified that

570    cause such a phenotype, but there are several genes with unknown function.  We also observed a

571    decrease in signal of *att1* during stationary phase. This is consistent with Tn*916* reintegration

572    into *att1* or death of cells in which Tn*916* has excised. We do not favor the first hypothesis

573    because Tn*916* can integrate into multiple sites (52, 54, 61), and we have observed that Tn*916*

574    does not have a preference for reintegration into *att1* in transconjugants (Wright and Grossman,

575    unpublished results).

576        Some ICEs are known to cause growth arrest and/or cell death.  For example, when activated,

577    ICE*clc,* an ICE active in *Pseudomonas* species, can inhibit host cell growth (~50% of activated

578    cells stop dividing) and cause cell lysis (24, 60). Despite the damage incurred by host cells, 75%

579   of donors with excised ICE*clc* that contact a recipient cell successfully transfer the element (24).

580   Single-cell microscopy studies, such as those conducted on ICE*clc,* are required to assess the

581   affect of Tn*916* induction on host cell fate.

582   **Autonomous replication of integrative and conjugative elements is conserved**

583   Growing evidence indicates that several ICEs replicate autonomously by a common

584   mechanism. ICE*Bs1* and Tn*916* both replicate by a rolling circle mechanism using similar

585   machinery. However, ICE*Bs1* and Tn*916* are very different elements, with different regulatory

586   mechanisms, different modes of integration, and different cargo genes. The ICE R391 also

587   replicates autonomously in Gram-negative *E. coli*, and its relaxase and *oriT* are important for

588   replication, indicating that R391 likely also uses rolling circle replication. Since all functional

589   ICEs that use a type IV secretion system have an origin of transfer and a cognate relaxase, the

590   accumulating findings support the notion that many, and perhaps all, ICEs are capable of

591   autonomous rolling circle replication.

592

593

602     this report are those of the authors and do not necessarily reflect the views of the National

603     Institutes of Health.

604

605
606     **References**
607

608     1.    **Amita, S. R. Chowdhury, M. Thungapathra, T. Ramamurthy, G. B. Nair, and A.**
609           **Ghosh.** 2003. Class I integrons and SXT elements in El Tor strains isolated before and
610           after 1992 *Vibrio cholerae* O139 outbreak, Calcutta, India. Emerg Infect Dis **9:**500-502.
611     2.    **Auchtung, J. M., C. A. Lee, R. E. Monson, A. P. Lehman, and A. D. Grossman.** 2005.
612           Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the
613           global DNA damage response. Proc Natl Acad Sci U S A **102:**12554-12559.
614     3.    **Babic, A., M. B. Berkmen, C. A. Lee, and A. D. Grossman.** 2011. Efficient gene transfer
615           in bacterial cell chains. mBio **2:**00027-00011.
616     4.    **Benson, A. K., and W. G. Haldenwang.** 1993. Regulation of sigma-B levels and activity
617           in *Bacillus subtilis*. J Bacteriol **175:**2347-2356.
618     5.    **Berkmen, M. B., and A. D. Grossman.** 2006. Spatial and temporal organization of the
619           *Bacillus subtilis* replication cycle. Mol Microbiol **62:**57-71.
620     6.    **Bi, D., Z. Xu, E. M. Harrison, C. Tai, Y. Wei, X. He, S. Jia, Z. Deng, K. Rajakumar,**
621           **and H. Y. Ou.** 2012. ICEberg: a web-based resource for integrative and conjugative
622           elements found in Bacteria. Nucleic Acids Res **40:**D621-626.
623     7.    **Bron, S., W. Meijer, S. Holsappel, and P. Haima.** 1991. Plasmid instability and
624           molecular cloning in *Bacillus subtilis*. Res Microbiol **142:**875-883.
625     8.    **Brouwer, M. S., A. P. Roberts, P. Mullany, and E. Allan.** 2012. In silico analysis of
626           sequenced strains of *Clostridium difficile* reveals a related set of conjugative transposons
627           carrying a variety of accessory genes. Mob Genet Elements **2:**8-12.
628     9.    **Brouwer, M. S., P. J. Warburton, A. P. Roberts, P. Mullany, and E. Allan.** 2011.
629           Genetic organisation, mobility and predicted functions of genes on integrated, mobile
630           genetic elements in sequenced strains of *Clostridium difficile*. PLoS One **6:**e23014.
631     10.   **Browne, H. P., S. Y. Anvar, J. Frank, T. D. Lawley, A. P. Roberts, and W. K. Smits.**
632           2015. Complete genome sequence of BS49 and draft genome sequence of BS34A, *Bacillus*
633           *subtilis* strains carrying Tn*916*. FEMS Microbiol Lett **362:**1-4.
634     11.   **Bruand, C., S. D. Ehrlich, and L. Janniere.** 1995. Primosome assembly site in *Bacillus*
635           *subtilis*. EMBO J **14:**2642-2650.
636     12.   **Burrus, V., G. Pavlovic, B. Decaris, and G. Guedon.** 2002. Conjugative transposons: the
637           tip of the iceberg. Mol Microbiol **46:**601-610.
638     13.   **Burrus, V., and M. K. Waldor.** 2003. Control of SXT integration and excision. J
639           Bacteriol **185:**5045-5054.
640     14.   **Burrus, V., and M. K. Waldor.** 2004. Shaping bacterial genomes with integrative and
641           conjugative elements. Res Microbiol **155:**376-386.
642     15.   **Bustin, S. A., and T. Nolan.** 2004. Data analysis and interpretation, p. 439-492. *In* S. A.
643           Bustin (ed.), A-Z of quantitative PCR. http://www.fivephoton.com.

644   16.   **Carraro, N., V. Libante, C. Morel, F. Charron-Bourgoin, P. Leblond, and G. Guedon.**
645         2016. Plasmid-like replication of a minimal Streptococcal Integrative and Conjugative
646         Element. Microbiology **162:**622-632.
647   17.   **Carraro, N., V. Libante, C. Morel, B. Decaris, F. Charron-Bourgoin, P. Leblond, and**
648         **G. Guedon.** 2011. Differential regulation of two closely related integrative and conjugative
649         elements from *Streptococcus thermophilus*. BMC Microbiol **11:**238.
650   18.   **Carraro, N., D. Poulin, and V. Burrus.** 2015. Replication and Active Partition of
651         Integrative and Conjugative Elements (ICEs) of the SXT/R391 Family: The Line between
652         ICEs and Conjugative Plasmids Is Getting Thinner. PLoS Genet **11:**e1005298.
653   19.   **Celli, J., C. Poyart, and P. Trieu-Cuot.** 1997. Use of an excision reporter plasmid to
654         study the intracellular mobility of the conjugative transposon Tn*916* in gram-positive
655         bacteria. Microbiology **143 ( Pt 4):**1253-1261.
656   20.   **Celli, J., and P. Trieu-Cuot.** 1998. Circularization of Tn*916* is required for expression of
657         the transposon-encoded transfer functions: characterization of long tetracycline-inducible
658         transcripts reading through the attachment site. Mol Microbiol **28:**103-117.
659   21.   **Chandler, M., F. De La Cruz, F. Dyda, A. B. Hickman, G. Moncalian, and B. Ton-**
660         **Hoang.** 2013. Breaking and joining single-stranded DNA: the HUH endonuclease
661         superfamily. Nat Rev Microbiol **11:**525-538.
662   22.   **Christie, P. J., R. Z. Korman, S. A. Zahler, J. C. Adsit, and G. M. Dunny.** 1987. Two
663         conjugation systems associated with *Streptococcus faecalis* plasmid pCF10: identification
664         of a conjugative transposon that transfers between *S. faecalis* and *Bacillus subtilis*. J
665         Bacteriol **169:**2529-2536.
666   23.   **Clewell, D. B., and S. E. Flannagan.** 1993. The Conjugative Transposons of Gram-
667         Positive Bacteria, p. 369-393. *In* D. B. Clewell (ed.), Bacterial Conjugation. Plenum Press,
668         New York, NY.
669   24.   **Delavat, F., S. Mitri, S. Pelet, and J. R. Van Der Meer.** 2016. Highly variable individual
670         donor cell fates characterize robust horizontal gene transfer of an integrative and
671         conjugative element. Proc Natl Acad Sci U S A **113:**E3375-3383.
672   25.   **Doucet-Populaire, F., P. Trieu-Cuot, I. Dosbaa, A. Andremont, and P. Courvalin.**
673         1991. Inducible transfer of conjugative transposon Tn1545 from Enterococcus faecalis to
674         Listeria monocytogenes in the digestive tracts of gnotobiotic mice. Antimicrobial Agents
675         and Chemotherapy **35:**185-187.
676   26.   **Fernández-López, C., A. Bravo, S. Ruiz-Cruz, V. Solano-Collado, D. A. Garsin, F.**
677         **Lorenzo-Diaz, and M. Espinosa.** 2014. Mobilizable Rolling-Circle Replicating Plasmids
678         from Gram-Positive Bacteria: A Low-Cost Conjugative Transfer. Microbiol Spectrum **2:**8.
679   27.   **Francia, M. V., D. B. Clewell, F. De La Cruz, and G. Moncalian.** 2013. Catalytic
680         domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction
681         endonucleases. Proc Natl Acad Sci U S A **110:**13606-13611.
682   28.   **Franke, A. E., and D. B. Clewell.** 1981. Evidence for a chromosome-borne resistance
683         transposon (Tn*916*) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the
684         absence of a conjugative plasmid. J Bacteriol **145:**494-502.
685   29.   **Gawron-Burke, C., and D. B. Clewell.** 1982. A transposon in *Streptococcus faecalis* with
686         fertility properties. Nature **300:**281-284.
687   30.   **Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd, and H. O.**
688         **Smith.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat
689         Methods **6:**343-345.

690   31.   **Glaser, P., F. Kunst, M. Arnaud, M. P. Coudart, W. Gonzales, M. F. Hullo, M.**
691         **Ionescu, B. Lubochinsky, L. Marcelino, I. Moszer, and Et Al.** 1993. Bacillus subtilis
692         genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333
693         degrees. Mol Microbiol **10:**371-384.
694   32.   **Haima, P., S. Bron, and G. Venema.** 1987. The effect of restriction on shotgun cloning
695         and plasmid stability in *Bacillus subtilis* Marburg. Mol Gen Genet **209:**335-342.
696   33.   **Haraldsen, J. D., and A. L. Sonenshein.** 2003. Efficient sporulation in *Clostridium*
697         *difficile* requires disruption of the sigmaK gene. Mol Microbiol **48:**811-821.
698   34.   **Harwood, C. R., and S. M. Cutting.** 1990. Molecular Biological Methods for *Bacillus*.
699         John Wiley & Sons, Chichester.
700   35.   **Henderson-Begg, S. K., A. P. Roberts, and L. M. Hall.** 2009. Diversity of putative
701         Tn*5253*-like elements in *Streptococcus pneumoniae*. Int J Antimicrob Agents **33:**364-367.
702   36.   **Ireton, K., and A. D. Grossman.** 1992. Interactions among mutations that cause altered
703         timing of gene expression during sporulation in *Bacillus subtilis*. J Bacteriol **174:**3185-
704         3195.
705   37.   **Jaacks, K. J., J. Healy, R. Losick, and A. D. Grossman.** 1989. Identification and
706         characterization of genes controlled by the sporulation-regulatory gene *spo0H* in *Bacillus*
707         *subtilis*. J Bacteriol **171:**4121-4129.
708   38.   **Jaworski, D. D., and D. B. Clewell.** 1995. A functional origin of transfer (*oriT*) on the
709         conjugative transposon Tn*916*. J Bacteriol **177:**6644-6651.
710   39.   **Johnson, C. M., and A. D. Grossman.** 2014. Identification of host genes that affect
711         acquisition of an integrative and conjugative element in *Bacillus subtilis*. Mol Microbiol
712         **93:**1284-1301.
713   40.   **Johnson, C. M., and A. D. Grossman.** 2015. Integrative and Conjugative Elements
714         (ICEs): What They Do and How They Work. Annu Rev Genet **49:**577-601.
715   41.   **Khan, S. A.** 2005. Plasmid rolling-circle replication: highlights of two decades of research.
716         Plasmid **53:**126-136.
717   42.   **Khan, S. A.** 1997. Rolling-circle replication of bacterial plasmids. Microbiol Mol Biol Rev
718         **61:**442-455.
719   43.   **Kramer, M. G., M. Espinosa, T. K. Misra, and S. A. Khan.** 1999. Characterization of a
720         single-strand origin, *ssoU*, required for broad host range replication of rolling-circle
721         plasmids. Mol Microbiol **33:**466-475.
722   44.   **Kramer, M. G., S. A. Khan, and M. Espinosa.** 1997. Plasmid rolling circle replication:
723         identification of the RNA polymerase-directed primer RNA and requirement for DNA
724         polymerase I for lagging strand synthesis. EMBO J **16:**5784-5795.
725   45.   **Lee, C. A., J. M. Auchtung, R. E. Monson, and A. D. Grossman.** 2007. Identification
726         and characterization of *int* (integrase), *xis* (excisionase) and chromosomal attachment sites
727         of the integrative and conjugative element ICE*Bs1* of *Bacillus subtilis*. Mol Microbiol
728         **66:**1356-1369.
729   46.   **Lee, C. A., A. Babic, and A. D. Grossman.** 2010. Autonomous plasmid-like replication of
730         a conjugative transposon. Mol Microbiol **75:**268-279.
731   47.   **Lee, C. A., J. Thomas, and A. D. Grossman.** 2012. The *Bacillus subtilis* conjugative
732         transposon ICE*Bs1* mobilizes plasmids lacking dedicated mobilization functions. J
733         Bacteriol **194:**3165-3172.
734   48.   **Manganelli, R., L. Romano, S. Ricci, M. Zazzi, and G. Pozzi.** 1995. Dosage of Tn*916*
735         circular intermediates in Enterococcus faecalis. Plasmid **34:**48-57.

736    49.    **Marra, D., and J. R. Scott.** 1999. Regulation of excision of the conjugative transposon
737           Tn*916*. Mol Microbiol **31:**609-621.
738    50.    **Mcdonnell, G. E., and D. J. Mcconnell.** 1994. Overproduction, isolation, and DNA-
739           binding characteristics of Xre, the repressor protein from the *Bacillus subtilis* defective
740           prophage PBSX. J Bacteriol **176:**5831-5834.
741    51.    **Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. Van Der**
742           **Meer.** 2008. Stochasticity and bistability in horizontal transfer control of a genomic island
743           in *Pseudomonas*. Proc. Natl. Acad. Sci. U.S.A. **105:**20792-20797.
744    52.    **Mullany, P., R. Williams, G. C. Langridge, D. J. Turner, R. Whalan, C. Clayton, T.**
745           **Lawley, H. Hussain, K. Mccurrie, N. Morden, E. Allan, and A. P. Roberts.** 2012.
746           Behavior and target site selection of conjugative transposon Tn*916* in two different strains
747           of toxigenic *Clostridium difficile*. Appl Environ Microbiol **78:**2147-2153.
748    53.    **Naglich, J. G., and R. E. Andrews, Jr.** 1988. Tn*916*-dependent conjugal transfer of
749           pC194 and pUB110 from *Bacillus subtilis* into *Bacillus thuringiensis* subsp. *israelensis*.
750           Plasmid **20:**113-126.
751    54.    **Norgren, M., and J. R. Scott.** 1991. The presence of conjugative transposon Tn*916* in the
752           recipient strain does not impede transfer of a second copy of the element. J Bacteriol
753           **173:**319-324.
754    55.    **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu Rev Biochem
755           **53:**293-321.
756    56.    **Palmer, K. L., V. N. Kos, and M. S. Gilmore.** 2010. Horizontal gene transfer and the
757           genomics of enterococcal antibiotic resistance. Curr Opin Microbiol **13:**632-639.
758    57.    **Paulsen, I. T., L. Banerjei, G. S. Myers, K. E. Nelson, R. Seshadri, T. D. Read, D. E.**
759           **Fouts, J. A. Eisen, S. R. Gill, J. F. Heidelberg, H. Tettelin, R. J. Dodson, L. Umayam,**
760           **L. Brinkac, M. Beanan, S. Daugherty, R. T. Deboy, S. Durkin, J. Kolonay, R.**
761           **Madupu, W. Nelson, J. Vamathevan, B. Tran, J. Upton, T. Hansen, J. Shetty, H.**
762           **Khouri, T. Utterback, D. Radune, K. A. Ketchum, B. A. Dougherty, and C. M. Fraser.**
763           2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*.
764           Science **299:**2071-2074.
765    58.    **Perego, M., G. B. Spiegelman, and J. A. Hoch.** 1988. Structure of the gene for the
766           transition state regulator, *abrB*: regulator synthesis is controlled by the *spo0A* sporulation
767           gene in *Bacillus subtilis*. Mol Microbiol **2:**689-699.
768    59.    **Ramsay, J. P., J. T. Sullivan, G. S. Stuart, I. L. Lamont, and C. W. Ronson.** 2006.
769           Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase
770           IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. Mol
771           Microbiol **62:**723-734.
772    60.    **Reinhard, F., R. Miyazaki, N. Pradervand, and J. R. Van Der Meer.** 2013. Cell
773           differentiation to "mating bodies" induced by an integrating and conjugative element in
774           free-living bacteria. Curr Biol **23:**255-259.
775    61.    **Rice, L. B., L. L. Carias, S. Rudin, R. A. Hutton, and S. Marshall.** 2010. Multiple
776           copies of functional, Tet(M)-encoding Tn*916*-like elements in a clinical *Enterococcus*
777           *faecium* isolate. Plasmid **64:**150-155.
778    62.    **Roberts, A. P., and P. Mullany.** 2009. A modular master on the move: the Tn*916* family
779           of mobile genetic elements. Trends Microbiol **17:**251-258.

780   63.   **Roberts, A. P., and P. Mullany.** 2011. Tn*916*-like genetic elements: a diverse group of
781         modular mobile elements conferring antibiotic resistance. FEMS Microbiol Rev **35:**856-
782         871.
783   64.   **Rocco, J. M., and G. Churchward.** 2006. The integrase of the conjugative transposon
784         Tn*916* directs strand- and sequence-specific cleavage of the origin of conjugal transfer,
785         *oriT*, by the endonuclease Orf20. J Bacteriol **188:**2207-2213.
786   65.   **Scott, J. F.** 1993. Conjugative Transposons, p. 597-614. *In* A. L. Sonenshein, J. A. Hoch,
787         and R. Losick (ed.), Bacillus subtilis and other Gram-positive bacteria: Biochemistry,
788         physiology, and molecular genetics. American Society of Microbiology, Washington, D.C.
789   66.   **Scott, J. R., P. A. Kirchman, and M. G. Caparon.** 1988. An intermediate in transposition
790         of the conjugative transposon Tn916. Proc Natl Acad Sci U S A **85:**4809-4813.
791   67.   **Sebaihia, M., B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R.**
792         **Thomson, A. P. Roberts, A. M. Cerdeno-Tarraga, H. Wang, M. T. Holden, A. Wright,**
793         **C. Churcher, M. A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A.**
794         **Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S.**
795         **Moule, K. Mungall, C. Price, E. Rabbinowitsch, S. Sharp, M. Simmonds, K. Stevens,**
796         **L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell, and J. Parkhill.** 2006. The
797         multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic
798         genome. Nat Genet **38:**779-786.
799   68.   **Sentchilo, V., R. Ravatn, C. Werlen, A. J. B. Zehnder, and J. R. Van Der Meer.** 2003.
800         Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. strain
801         B13. J. Bacteriol. **185:**4530-4538.
802   69.   **Showsh, S. A., and R. E. Andrews, Jr.** 1999. Analysis of the requirement for a pUB110
803         mob region during Tn*916*-dependent mobilization. Plasmid **41:**179-186.
804   70.   **Showsh, S. A., and R. E. Andrews, Jr.** 1992. Tetracycline enhances Tn*916*-mediated
805         conjugal transfer. Plasmid **28:**213-224.
806   71.   **Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H.**
807         **Mcwilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins.** 2011. Fast,
808         scalable generation of high-quality protein multiple sequence alignments using Clustal
809         Omega. Mol Syst Biol **7:**539.
810   72.   **Smillie, C., M. P. Garcillan-Barcia, M. V. Francia, E. P. Rocha, and F. De La Cruz.**
811         2010. Mobility of plasmids. Microbiol Mol Biol Rev **74:**434-452.
812   73.   **Smith, J. L., J. M. Goldberg, and A. D. Grossman.** 2014. Complete genome sequences
813         of *Bacillus subtilis* subsp. *subtilis* laboratory strains JH642 (AG174) and AG1839. Genome
814         Announc **2**.
815   74.   **Smits, W. K., and A. D. Grossman.** 2010. The transcriptional regulator Rok binds A+T-
816         rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*.
817         PLoS Genet **6:**e1001207.
818   75.   **Te Poele, E. M., H. Bolhuis, and L. Dijkhuizen.** 2008. Actinomycete integrative and
819         conjugative elements. Antonie Van Leeuwenhoek **94:**127-143.
820   76.   **Thomas, J., C. A. Lee, and A. D. Grossman.** 2013. A conserved helicase processivity
821         factor is needed for conjugation and replication of an integrative and conjugative element.
822         PLoS Genet **9:**e103198.
823   77.   **Wang, J., G. R. Wang, N. B. Shoemaker, and A. A. Salyers.** 2001. Production of two
824         proteins encoded by the Bacteroides mobilizable transposon NBU1 correlates with time-
825         dependent accumulation of the excised NBu1 circular form. J Bacteriol **183:**6335-6343.

826   78.   **Washburn, M. P., D. Wolters, and J. R. Yates, 3rd.** 2001. Large-scale analysis of the
827           yeast proteome by multidimensional protein identification technology. Nat Biotechnol
828           **19:**242-247.
829   79.   **Wilkins, B., and E. Lanka.** 1993. DNA Processing and Replication during Plasmid
830           Transfer between Gram-Negative Bacteria, p. 105-136. *In* D. B. Clewell (ed.), Bacterial
831           Conjugation. Springer US.
832   80.   **Wood, H. E., M. T. Dawson, K. M. Devine, and D. J. Mcconnell.** 1990. Characterization
833           of PBSX, a defective prophage of *Bacillus subtilis*. J Bacteriol **172:**2667-2674.
834   81.   **Wood, H. E., K. M. Devine, and D. J. Mcconnell.** 1990. Characterisation of a repressor
835           gene (*xre*) and a temperature-sensitive allele from the *Bacillus subtilis* prophage, PBSX.
836           Gene **96:**83-88.
837   82.   **Wozniak, R. A., and M. K. Waldor.** 2010. Integrative and conjugative elements: mosaic
838           mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol **8:**552-
839           563.
840   83.   **Wright, L. D., C. M. Johnson, and A. D. Grossman.** 2015. Identification of a Single
841           Strand Origin of Replication in the Integrative and Conjugative Element ICE*Bs1* of
842           *Bacillus subtilis*. PLoS Genet **11:**e1005556.
843
844
845

846    **Table 1.** *Bacillus subtilis* **strains used.**

| Strain | Relevant genotype[a] (reference) |
|--------|-----------------------------------|
| BS49 | *metB5, hisA1, thr-5, att*(*yufKL*)[b]::Tn*916, att*(*ykyB-ykuC*)::Tn*916* (10, 22, 33) |
| JMA222 | ICE*Bs1*[0] (2) |
| CMJ129 | pHP13, *lacA*::{(PrpsF-*rpsF-ssb-mgfpmut2*) *tet*} (83) |
| LDW173 | *att1*::Tn*916*  {same as *att*(*yufKL*)::Tn*916*} |
| LDW631 | *att1*::Tn*916* Δ*orf20*-631;  deletes most of *orf20,* leaves a functional *oriT* |
| LDW737 | *amyE*::{(*att*Tn9*16*) *spc*} |
| LDW815 | pLW805 {*oriT*(*916*), Pspank-*orf23-orf22-orf20, spc, lacI*} |
| LDW818 | *att1*::Tn*916*::pLW805 {*oriT*(*916*), Pspank-*orf23-orf22-orf20, spc, lacI*} |
| LDW853 | *att1*::Tn*916 orf20-3UAA* |
| LDW872 | pLW862 (pHP13ssiA *cat mls*),  *lacA*::{PrpsF-*rpsF-ssb-mgfpmut2, tet*} |
| LDW878 | pLW868 (pHP13sso916 *cat mls*),  *lacA*::{PrpsF-*rpsF-ssb-mgfpmut2, tet*} |
| LDW879 | pLW859 {*oriT*(*916*), Pspank-*orf23-orf22-orf20-his*6, *spc, lacI*} |
| LDW894 | pLW890 (pHP13sso916R *cat mls*), *lacA*::{PrpsF-*rpsF-ssb-mgfpmut2, tet*} |
| LDW929 | *att1*::Tn*916* Δ*orf20*-631, *lacA*::{*mls*, Pspank(hy)-*orf20-myc,* pLW920 *spc*};  pLW920 is integrated into *orf20* to generate *orf20-myc* |
| LDW930 | *att1*::Tn*916* Δ*orf20*-631, *lacA*::{*mls*, Pspank(hy)-*orf20*Δ*hth-myc,* pLW920 *spc*}  pLW920 is integrated into *orf20*Δ*hth* to generate *orf20*Δ*hth-myc* |
| LDW931 | *att1*::Tn*916 orf20-3UAA*,  *lacA*::{*mls*, Pspank(hy)-*orf20-myc,* pLW920 *spc*} |
| LDW932 | *att1*::Tn*916 orf20-3UAA*, *lacA*::{*mls* (Pspank(hy)-*orf20*Δ*hth-myc,* pLW920 *spc* } |

847

848    [a] All strains except BS49 are derived from JH642 and contain the *trpC2 pheA1* alleles (58,

849    73).  Strains do not contain Tn*916* unless Tn*916* is specifically indicated.

850    [b] *att1* is the same as *att*(*yufKL*) and is located between *yufK* and *yufL*.

851

852

853   **Table 2. Mass spectrometry of affinity-purified Orf20-his shows that the HelP homologues**

854   **are associated with the relaxase.**

855

| Protein | % Sequence coverage[a] | # Peptides[b] | MW (kDa)[c] |
|---|---|---|---|
| Orf20 relaxase | 51.5 | 84 | 46.8 |
| Orf22 HelP | 67.2 | 33 | 14.1 |
| Orf23 HelP | 69.6 | 18 | 11.8 |

856

857   [a] Percentage of the protein sequence detected by mass spectrometry. Amino acid sequences were

858   based on Tn*916* genes (GenBank U09422.1) except Orf20, which was based on the re-annotated

859   gene starting at CUG and containing the N-terminal helix-turn-helix region (Fig. 3B).

860

861   [b] Number of total peptides detected.

862

863   [c] Predicted molecular weight in kilodaltons.

864

865

866

**Table 3. Complementation of the Tn*916* replication defects of relaxase mutants Δ*orf20* and *orf20-3UAA*.**

| Line | Tn*916* genotype | Pspank(hy)[a] | Circles per excision[b] |
|------|------------------|---------------|-------------------------|
| 1 | wild type | --- | 6.05 (±0.43) |
| 2 | Δ*orf20* | --- | 0.43 (±0.12) |
| 3 | *orf20-3UAA* | --- | 0.61 (±0.26) |
| 4 | Δ*orf20* | *orf20-myc* | 8.03 (±2.94) |
| 5 | *orf20-3UAA* | *orf20-myc* | 6.98 (±1.60) |
| 6 | Δ*orf20* | *orf20Δhth-myc* | 0.55 (±0.08) |
| 7 | *orf20-3UAA* | *orf20Δhth-myc* | 0.46 (±0.17) |

[a]*orf20* complementation alleles were expressed in trans at the *lacA* locus and driven by promoter Pspank(hy).

[b] Circles per excision were quantified by qPCR by measuring the amount of the circular junction *att*Tn*916* relative to the vacant chromosome site *att1* (Fig. 1A). Data are averages from 3 independent experiments (± standard deviation). Strains were LDW173, LDW631, LDW853, LDW929, LDW931, LDW930 and LDW932 (lines 1-7).

880    **Table 4. *sso916* is conserved in other Tn*916*-like ICEs.**

| ICE[a] | Organism[b] |
|---|---|
| Tn*6009* | *Klebsiella pneumoniae* 41 |
| ICE*Spn*H034800032 | *Streptococcus pneumoniae* H034800032 |
| CTn*6009* | *Streptococcus cristatus* |
| ICE*Spn*9409 | *S. pneumoniae* 9409 |
| ICE*Spn*11928 | *S. pneumoniae* 11928 |
| ICE*Spn*MalM6 | *S. pneumoniae* Mal M6 |
| ICE*Spn*11930-2 | *S. pneumoniae* 11930 |
| ICE*Spn*23771 | *S. pneumoniae* 23771 |
| ICE*Spn*11876 | *S. pneumoniae* 11876 |
| ICE*Spn*11930 | *S. pneumoniae* 11930 |
| ICE*Ssu*(BM407)1 | *Streptococcus suis* BM407 |
| ICE*Ssu*(BM407)2 | *S. suis* BM407 |
| ICESp23FST81 | *S. pneumoniae* ATCC 700669 |
| ICE*Ssu*(SC84) | *S. suis* SC84 |
| Tn*5397* | *Clostridium difficile* 630 |
| Tn*1545* | *S. pneumoniae* BM4200 |

881

882    [a]ICEs with regions similar to the intergenic sequence containing *sso916* were identified using

883    WU-BLAST 2.0 and searching the ICEberg v1.0 database of ICE nucleotide sequences (6). The

884    search identified *sso*916 in Tn*5251,* which is essentially identical to Tn*916* (>95% identity at the

885    nucleotide level) and is not included in the table. ICE*Spn*H034800032 and CTn*6002* are listed

886    separately in ICEberg and were identified in different organisms, but are essentially identical

887    elements (>95% identity at the nucleotide level).  Conservation of each putative *sso* with *sso916*

888    is 100% except for Tn*5397* and Tn*1545*, which have 94% and 96% identity, respectively.

889

890    [b] Each ICE was initially identified in the indicated species and strain.

891

892     **Figure legends**

893     **Figure 1. Genetic map of Tn*916* and schematic for detecting excision products.** The ends

894     of Tn*916* are indicated by black rectangles.  **(A)** Map of Tn*916*. Rectangular gray arrows

895     represent the genes in Tn*916* and the direction of transcription. *orf23* and *orf22* encode

896     homologues of the helicase processivity factor (HelP from ICE*Bs1*) and *orf20* encodes the

897     relaxase needed for DNA transfer and replication. All other gene names (numbers) are indicated

898     above the corresponding gene. *oriT*(*916*) (38) and *sso*916 are indicated with thick black lines.

899     **(B)** Cartoon of the qPCR strategy to measure Tn*916* replication. Excision of Tn*916* from the

900     chromosome of strain LDW173 produces 1) circular Tn*916* containing site *att*Tn*916* and 2)

901     vacated chromosome site *att1*. The products are detected via qPCR using primers b and c

902     (primers oLW526 and oLW527), or a and d (primers oLW542 and oLW543), respectively,

903     which are represented as small black arrows.

904

905     **Figure 2. Products generated following excision of Tn*916* from the chromosome.** Cells

906     containing Tn*916* (strain LDW173, triangles) or Tn*916* Δ*orf20* (LDW631, circles) were grown in

907     the presence (filled symbols) or absence (open symbols) of 2.5 µg/ml tetracycline. Quantitative

908     PCR data (black lines, left axes) and growth phase as determined by $OD_{600}$ (gray lines, right

909     axes) are shown. Strain LDW737 contains one copy of each qPCR amplicon (*att*Tn*916; att1;*

910     and *mrpG*) and was used to generate standard curves for qPCR (see Materials and Methods). **(A)**

911     Excision of Tn*916*.  The vacated chromosomal site *att1* was amplified by qPCR, and *att1* signal

912     was normalized to the signal of an unrelated chromosomal locus, *mrpG*. **(B)** The average copy

913     number of circular Tn*916* per host chromosome. The *att*Tn*916* junction present in the circular

914     (excised) form of Tn*916* was amplified by qPCR, and the qPCR signal was normalized to that of

915     chromosomal gene *mrpG*. **(C)** Tn*916* circles per excision event. Circular Tn*916* (*att1*Tn*916*) was

916     amplified by qPCR, and the *att1*Tn*916* signal was normalized to that of the *att1* site formed from

917     excision. Data are means ± standard deviation of ≥4 independent experiments.

918

919

920     **Figure 3. The relaxase encoded by *orf20* contains a conserved N-terminal helix-turn-**

921     **helix region**. **(A)** The revised amino acid sequence of the Tn*916* relaxase (Orf20). The

922     methionine previously thought to be the start of the protein is indicated in bold (**M**). Peptides

923     identified by mass spectrometry were mapped to the amino acid sequence of Orf20 and are

924     underlined. The bold underline indicates peptides that overlapped the junction between the N

925     terminal region and what was previously thought to be the initiating methionine. **(B-D)** Maps of

926     *orf20* and *orf20* overexpression alleles used here. Gray arrows correspond to the *orf20* coding

927     sequence and the direction of transcription. The dark gray rectangles represent the conserved

928     XRE-like helix-turn-helix (HTH) and relaxase regions in *orf20* as determined using the NCBI

929     Conserved Domain Database (see Materials and Methods). The black rectangles represent the

930     putative ribosome binding site (RBS) that is present in Tn*916* and preserved in the myc-tagged

931     alleles. **(B)** *orf20* in Tn*916* and upstream sequence. The hatched black arrow represents the C

932     terminus of *orf21*. The CUG codon and previously proposed AUG start codons are indicated.

933     The relative location of the UAA nonsense mutation in *orf20-3UAA* is marked with an asterisk.

934     **(C-D)** Myc-tagged *orf20* alleles with (**C**; WT *orf20*) or without (**D**; *orf20Δhth*) the N terminal

935     helix-turn-helix region. The C-terminal myc tags are not shown. *orf20*-myc alleles were driven

936     by promoter Pspank(hy) (black arrow).  **(C)** WT*orf20-myc* contains the entire coding sequence as

937     depicted in (A) and non-coding DNA upstream of the CUG start. **(D)** *orf20Δhth-myc* contains the

938     *orf20* coding sequence starting at the previously proposed AUG start codon as depicted in (A)

939     and non-coding DNA upstream of the CUG start.

940

941     **Figure 4. Alignment of several relaxase homologues.** The amino acid sequences of the

942     relaxases from Tn*916,* Tn*5253*, CTn*7*, CTn*1* and ICE*Bs1* were aligned with the Clustal Omega

943     algorithm (http://www.ebi.ac.uk/Tools/msa/clustalo/) (71). Black-shaded residues are identical in

944     all five relaxases, and gray-shaded residues are similar in all relaxases. The helix-turn-helix

945     region present in four of the five relaxases is boxed. Previously proposed N-terminal methionines

946     in *orf20* of Tn*916* and *orf26* of CTn*1* are bolded and boxed. The output alignment was shaded

947     using BoxShade (http://www.ch.embnet.org/software/BOX_form.html). The order of sequences

948     (Tn*916* Orf20 to ICE*Bs1* NicK) reflects the order of the original input queries; closely related

949     sequences were not computationally grouped in the final alignment.

950

951

952      **Figure 5. Ssb-GFP to visualize ssDNA and single strand origin activity.** All cells contain

953      ssDNA-binding protein Ssb fused to GFP. Phase contrast (top panels) and GFP fluorescence

954      (bottom panels) are shown. **(A)** No plasmid, strain CMJ118. In cells without a plasmid, Ssb-GFP

955      forms small foci at the replication forks (white arrow) (5). **(B)** pHP13, strain CMJ129. Ssb-GFP

956      forms large foci (white arrowhead) in cells containing pHP13, which does not encode a

957      functional *sso* and accumulates ssDNA (83). **(C)** pHP13ssiA, strain LDW872. The primosome

958      binding site *ssiA* from pAMß, which can function as an *sso* in rolling circle replicating plasmids

959      and reduce ssDNA (11), was cloned into pHP13 to make pHP13ssiA (pLW862). Cells

960      containing pHP13ssiA had small foci of Ssb-GFP (white arrow) **(D)** pHP13sso916, strain

961      LDW878. Cells containing pHP13 with the *sso* from Tn*916* (pLW868) did not accumulate large

962      Ssb-GFP foci (white arrow), indicating reduced ssDNA. **(E)** pHP13sso916R, strain LDW894.

963      Cells containing pHP13 with *sso*916 cloned in the reverse orientation (pLW890) had large Ssb-

964      GFP foci (white arrowhead), indicating accumulation of ssDNA. Data are representative images

965      from ≥ 3 independent experiments.

966

967

968      **Figure 6. *sso*916 reduces the amount of Ssb-GFP bound to plasmid DNA.** Plasmid DNA

969      associated with Ssb-GFP. Ssb-GFP was immunoprecipitated following crosslinking with

970      formaldehyde.  The amount of plasmid DNA that was co-immunoprecipitated (with Ssb-GFP)

971      was amplified with qPCR using primers to the *cat* gene in pHP13.  The amounts of PCR

972      products were normalized to the amount of plasmid DNA in total lysates, essentially as described

973      (83). Data are means ± standard deviation of biological triplicates.

974
975

**A**

*24* *21* *19* *18* *17* *16* *15* *14* *13* *12* *tet(M)* *6* *9 10* *7* *8* *5* *xis* *int*

*attL* *attR*

*oriT* *sso*

**orf23; orf22 helPs** **orf20 relaxase**

**B**

Integrated Tn*916* → Excision → Tn*916* circle   Vacated chromosome

*att*Tn*916*   *att1*

**A** tet

Excision

*att1* per *mrpG*

+tet

uninduc.

OD600

**B** Circles per cell

*attTn916* per *mrpG*

WT

OD600

Δ*orf20*

**C** Circles per excision events

*att*Tn*916* per *att1*

WT

OD600

Δ*orf20*

time (minutes)
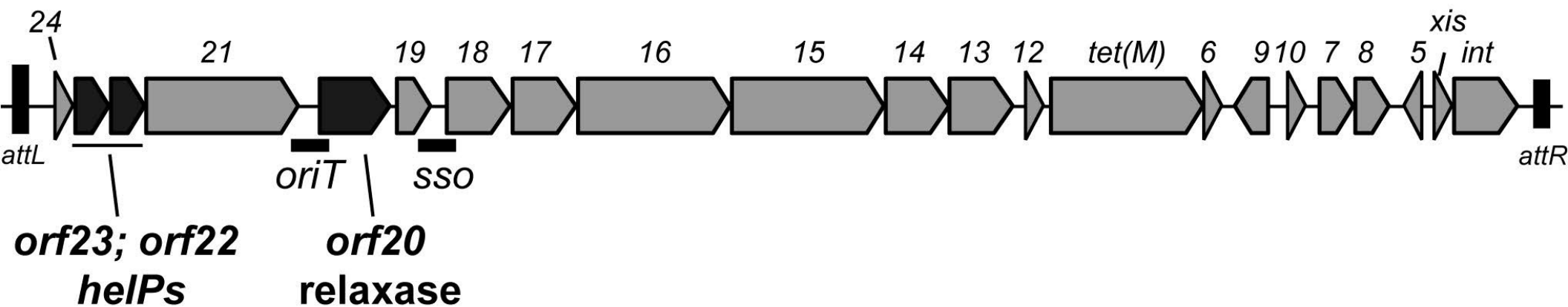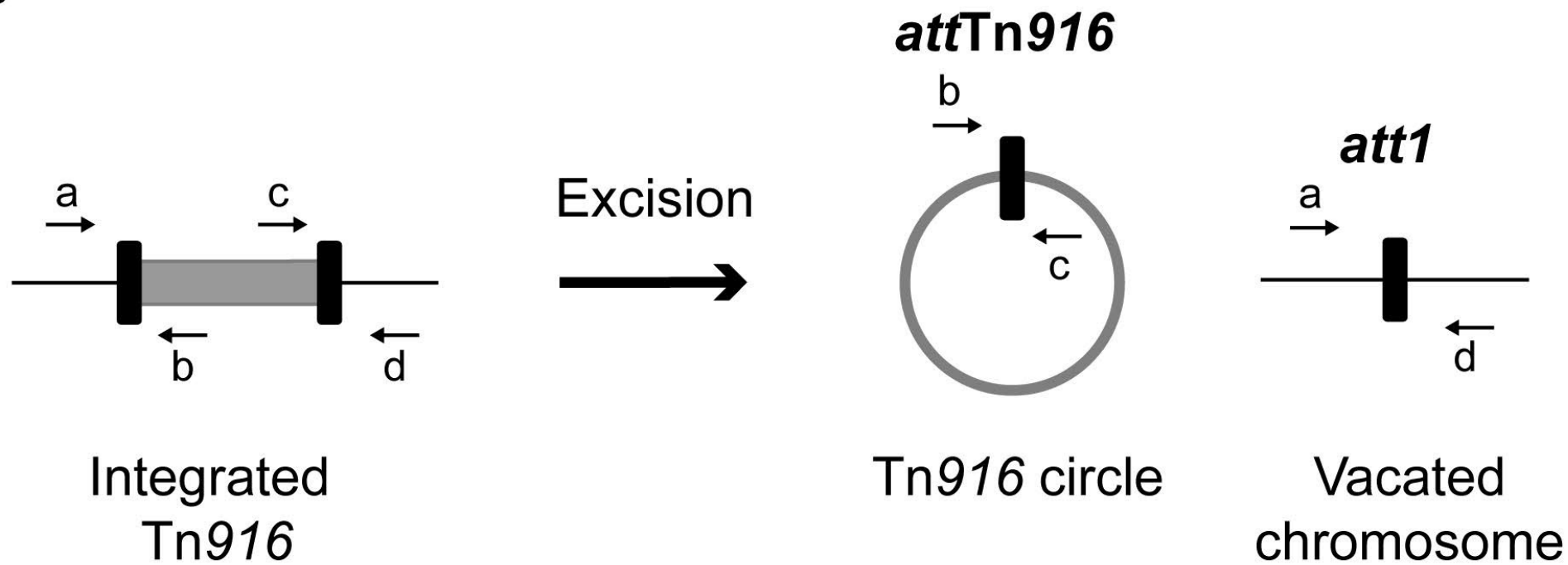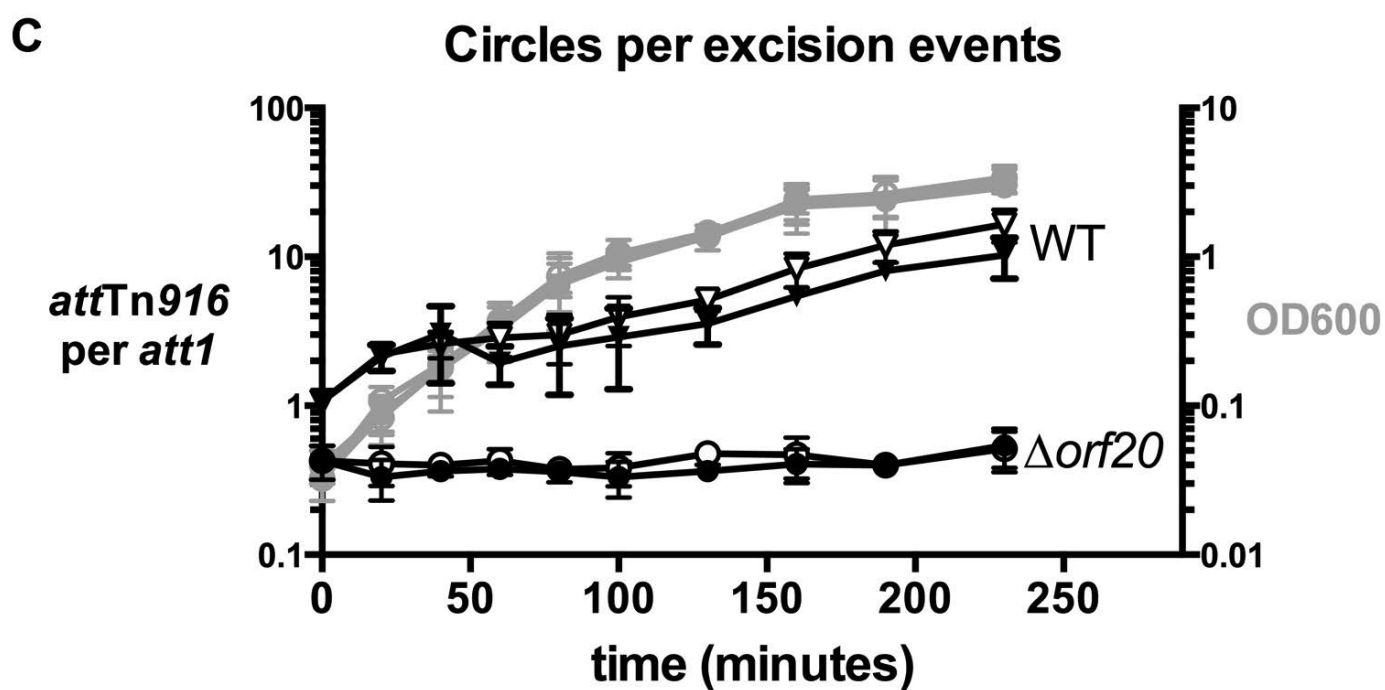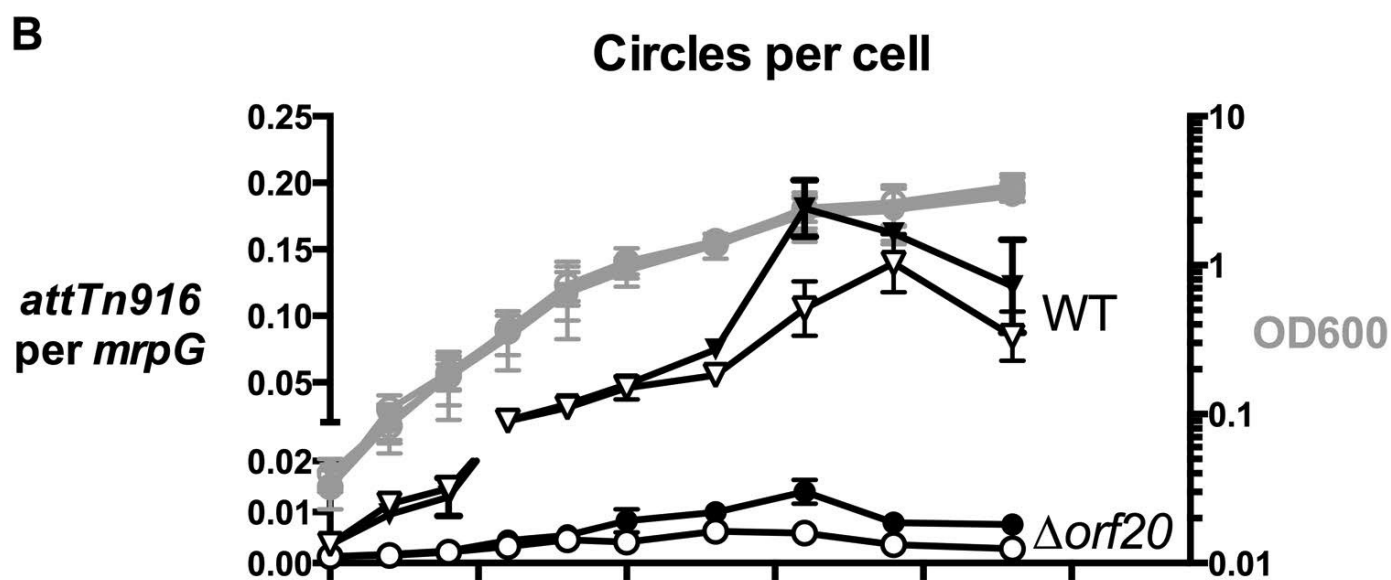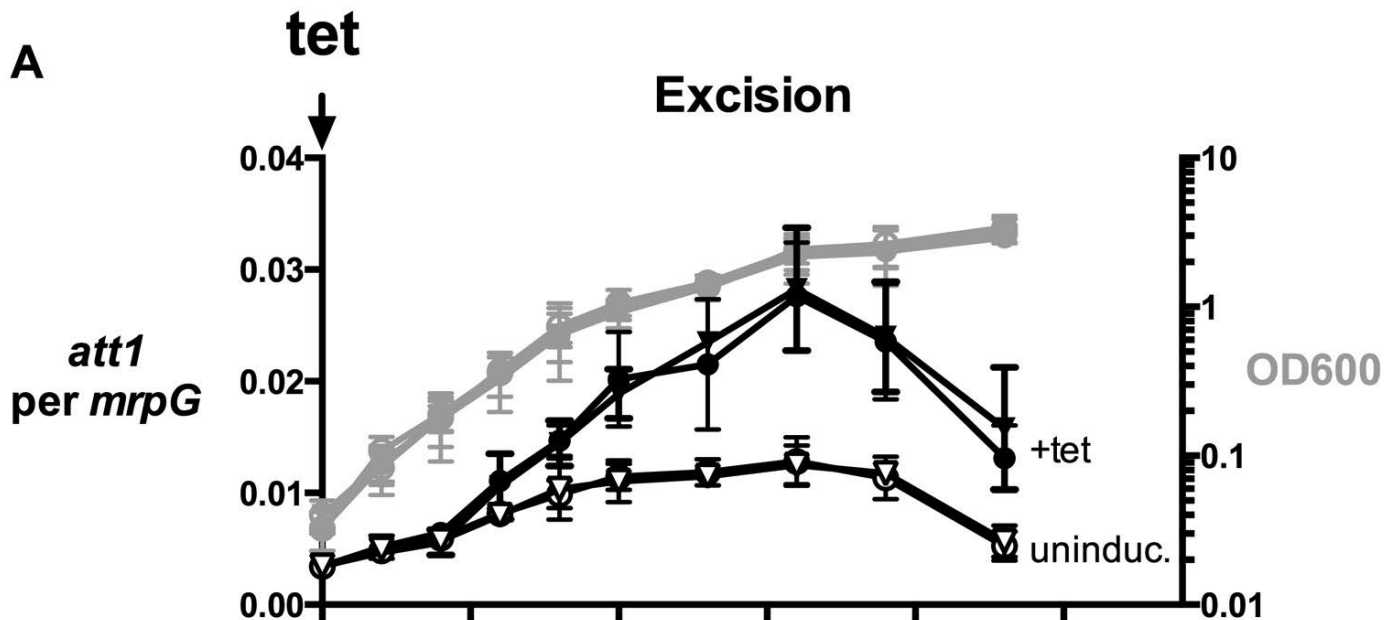
## A

**Mass spectrometry peptide coverage of Orf20 relaxase**

```
MNEQTWLQHLKEKRLAYGLSQNRLAVATGITRQYLSDIETGKVKPSEDLQQSLWEALERFNPDAPL
EMLFDYVRIRFPTTDVQQVVENILQLKLSYFLHEDYGFYSYSEHYALGDIFVLCSHELDKGVLVEL
KGRGCRQFESYLLAQQRSWYEFFMDVLVAGGVMKRLDLAINDKTGILNIPVLTEKCQQEECISVFR
SFKSYRSGELVRKEEKECMGNTLYIGSLQSEVYFCIYEKDYEQYKKNDIPIEDAEVKNRFEIRLKN
ERAYYAVRDLLVYDNPEHTAFKIINRYIRFVDKDDSKPRSDWKLNEEWAWFIGNNRERLKLTTKPE
PYSFQRTLNWLSHQVAPTLKVAIKLDEINQTQVVKDILDHAKLTDRHKQILKQQSVKEQDVITTKK
```

## B

**orf20 re-annotated**



orf21

CUG    AUG

rbs

HTH    relaxase

*
UAA

## C

orf20-myc

## D

orf20Δhth-myc

**Helix-turn-helix**

```
Tn916_Orf20     1   -----MNEQTWLQHLKEKRLAYGLSQNRLAVATGITRQYLSDIETGKVKPSEDLQQSLWEALERFNPDAP
Tn5253_Orf20    1   MEGFLLNEQTWLQHLKEKRLAYGLSQNRLAVATGITRQYLSDIETGKVKPSEDLQQSLWEALERFNPDAP
CTn7_Orf27      1   ---MVLNEEQWIKELREKRIAYGISQGRLAVASGITREYLNKIESGKMKPSKELLHTLHKELARFNPETP
CTn1_Orf26      1   -----MNEKDFLTALKEKRCDYGVSQTRLAIMAGISREHLSRIEAGKVTLTEDMKHKLLEAVEKFNPDNP
ICEBs1_NicK     1   ------------------------------------------------MDELKQPPHA-NRGVVIVKEKNEAVESP
```

```
Tn916_Orf20    66   LEMLFDYVRIRFPTTDVQQVVENILQLKLSYFLHEDYGFYSYSEHYALGDIFVLCSHELDKGVLVELKGR
Tn5253_Orf20   71   LEMLFDYVRIRFPTTDVQQVVENILQLKLSYFLHEDYGFYSYSEHYALGDIFVLCSHELDKGVLVELKGR
CTn7_Orf27     68   LTMLFDYVKIRFPTLDIQHIIKDILKLNINYMLHEDYGHYSYTEHYSLGDIFIYTSADEEKGVLLELKGR
CTn1_Orf26     66   MFLLFDYVRIRFPTTDIQHIIKDILKLNINYMLHEDYGHYKYTEHYHIGEVFVYVSQDEEKGVLLELKGK
ICEBs1_NicK    28   LVSMVDYIRVSFKTHDVDRIIEEVLHLSKDFMTEKQSGFYGYVGTYELDYIKVFYSAPDDRGVLIEMSGQ
```

```
Tn916_Orf20   136   GCRQFESYLLAQQRSWYEFFMDVLVAGGVMKRLDLAINDKTGILNIPVLTEKCQQEECISVFRSFKSYRS
Tn5253_Orf20  141   GCRQFESYLLAQQRSWYEFFMDVLVAGGVMKRLDLAINDKTGILNIPVLTEKCQQEECISVFRSFKSYRS
CTn7_Orf27    138   GCRQFESYLLAQQRSWYDFLMDALVDGGVMKRIDLAINDHTGILDIPELAEKCRKREYIGKSRSYKFYQS
CTn1_Orf26    136   GCRQFESYLLAQERSWYDFFMDALVEGGVMKRIDLAINDRTGLLDIPELIQKCENEECISKFRSFKNYGS
ICEBs1_NicK    98   GCRQFESFLECRKKTWYDFFQDCMQQGGSFTRFDLAIDDKKTYFSIPELLKKAQKGECISRFRKSDFNGS
```

```
Tn916_Orf20   206   GELVRKEE--KECMGNTLYIGSLQSEVYFCIYEKDYEQYKKNDIPIEDAEVKNRFEIRLKNERAYYAVRD
Tn5253_Orf20  211   GELVRKEE--KECMGNTLYIGSLQSEVYFCIYEKDYEQYKKNDIPIEDAEVKNRFEIRLKNERAYYAVRD
CTn7_Orf27    208   GELIKHRED--EYMGRTLYLGSLKSDVYFCIYEKDYEQYVKLGIPLEEADIINRFEIRLNERAYYAVRD
CTn1_Orf26    206   GELVKHNETDKGGMGHTLYIGSFSSEVYFCCYEKNYEQYAKLGIPIEEVPIKNRFEIRLKNERAYYAVRE
ICEBs1_NicK   168   FDLSD-----GITGGTTIYFGSKKSEAYLCFYEKNYEQAEKYNIPLEELGDWNRYELRLKNERAQVAIDA
```

```
Tn916_Orf20   274   LLVYDNPEHTAFKIINRYIRFVDKDDSKPRSDWKLNEEWAWFIGNNRERLKLTTKPEPYSFQRTLNWLSH
Tn5253_Orf20  279   LLVYDNPEHTAFKIINRYIRFVDKDDSKPRSDWKLNEEWAWFIGNNRERLKLTTKPEPYSFQRTLNWLSH
CTn7_Orf27    276   LLTYYDAEQTAFSVINQYVRFVDEEPDKRKNDWKLNDRWAWFIGDNRQSLKLTTKPEPYTLDRTLRWVQR
CTn1_Orf26    276   LLTNYDAELTAFSIINQYIRFADKEPDKRKSDWKTNARWSWFIGEGRPPIKLTTKPEPFTMERTMKWLQR
ICEBs1_NicK   233   LLKTKDLTLIAMQIINNYVRFVDADENITREHWKTSLFWSDFIGDV-GRLPLYVKPQKDFYQKSRNWLRN
```

```
Tn916_Orf20   344   QVAPTLKVAIKLDEINQTQVVKDILDHAKLTDRHKQILKQQSVKEQDVITTKK
Tn5253_Orf20  349   QVAPTLKVAIKLDEINQTQVVKDILDHAKLTDRHKQILKQQSVKEQDVITTKK
CTn7_Orf27    346   QVAPTLKMLKKIDKGNGTDYMETIEQQAKLTEKHEMIIKQQTTPAKDLVKS--
CTn1_Orf26    346   QVAPTLKMMKKIDKGNGTDYMETIEQQAKLTEKHEMIIKQQTTPAKDLVES--
ICEBs1_NicK   302   SCAPTMKMVLEADEHLGKTDLSDMIAEAELADKHKKMLDVYMADVADMVV---
```

|   | A | B | C | D | E |
|---|---|---|---|---|---|
| Phase contrast | | | | | |
| ssDNA (Ssb-GFP) | | 2 µm | | | |
| | No plasmid | pHP13 | pHP13ssiA | pHP13sso916 | pHP13sso916R |
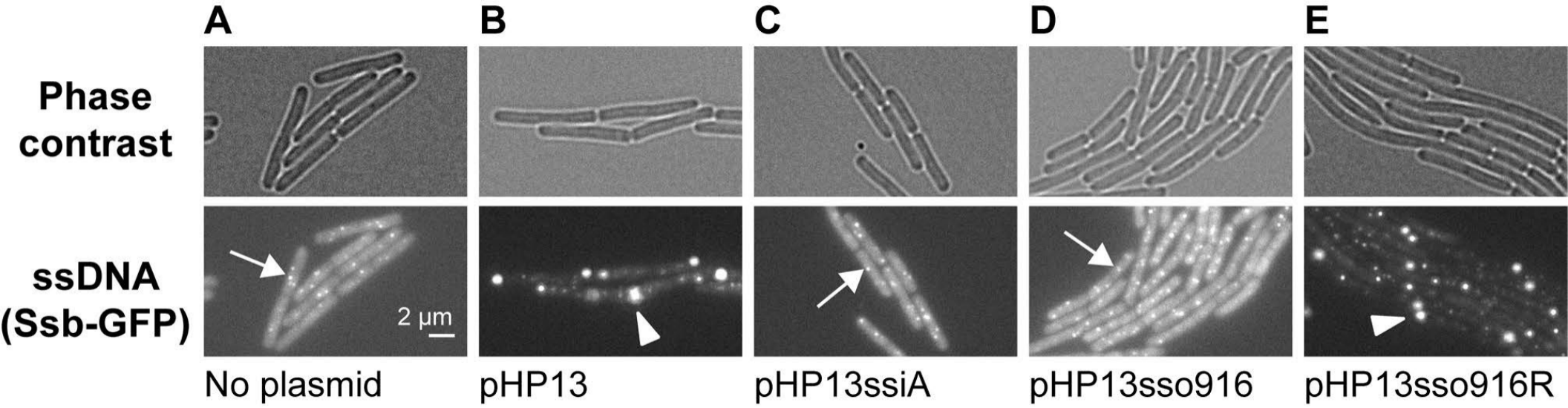
**ssDNA**

% of input
amount of plasmid DNA bound to Ssb

y-axis: 10, 1, 0.1, 0.01

x-axis: pHP13, phP13sso916, pHP13sso916R