Communications Biology

https://doi.org/10.1038/s42003-026-09874-7

Article in Press

Dual regulatory role of IS*91*-encoded Orf121 in IS*91* transposition

Received: 17 July 2025

Accepted: 6 March 2026

Cite this article as: Fauconnier, A., Da Re, S., Gaschet, M. *et al.* Dual regulatory role of IS*91*-encoded Orf121 in IS*91* transposition. *Commun Biol* (2026). https://doi.org/10.1038/s42003-026-09874-7

Aurélien Fauconnier, Sandra Da Re, Margaux Gaschet, Thomas Jové, Marie-Cécile Ploy & Cécile Pasternak

Dual regulatory role of IS*91*-encoded Orf121 in IS*91* transposition

Aurélien Fauconnier[1#a], Sandra Da Re[¶1], Margaux Gaschet[1], Thomas Jové[1#b], Marie-Cécile Ploy[1¶*], Cécile Pasternak[1¶†]

[1]Univ. Limoges, INSERM, CHU Limoges, RESINFIT, U 1092, F-87000 Limoges, France

[#a] Current address: Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada.

[#b] Present address: no longer employed by INSERM

[*]Corresponding author

E-mail: marie-cecile.ploy@unilim.fr

[¶] These authors jointly supervised this work

Keywords: DNA transposition / bacterial insertion sequences / IS*91* family / Regulation / Orf121

Abstract

Insertion sequences (IS) are key players in bacterial genome plasticity and horizontal gene transfer. IS*91* family members, belonging to the HUH superfamily of single-strand nucleases, are often linked with antibiotic resistance genes. Among these, the element IS*91* is unique as it also carries a sequence called *orf121*, whose stop codon overlaps with the start codon of *tnpA*, a highly conserved feature of IS*91* isoforms. We show that Orf121 serves as a dual regulator of IS*91* transposition: Orf121 inhibits transposition activity of TnpA while facilitating accurate excision of IS*91* single-strand circular intermediates. This accurate excision reduces one-ended transposition events, i.e., events arising when proper termination fails, leading to the co-mobilization of adjacent DNA. We also provide evidence that the bottom-stranded ssDNA circular intermediate is the functional substrate for IS*91*. These findings highlight a sophisticated regulatory strategy balancing IS*91* mobility and genetic stability.

Introduction

Mobile genetic elements (MGEs) in bacteria are crucial in the propagation of antibiotic resistance genes, presenting a formidable threat to global health across human, animal, and environmental domains (for recent reviews, see [1,2]). These elements encompass genetic entities that facilitate intercellular transfer (e.g., transmissible plasmids and integrative conjugative elements) as well as transposable elements (TE) that drive intracellular mobility, including insertion sequences (IS), transposons, and gene cassettes from integrons. TEs not only enhance dissemination but can also activate proximal antibiotic resistance genes (ARGs).

Among TEs, the IS*91* family is particularly atypical, with only four members confirmed to be actively transposing: the canonical IS*91*[3,4], IS*801*[5], IS*1294*[6], and IS*1294b*[7] closely related to IS*1294*. They encode transposases (TnpA) that belong to the HUH superfamily of single-strand nucleases[8], which are presumed to transpose via a rolling-circle (RC) mechanism[9].

Unlike traditional IS that feature terminal inverted repeats (IR) flanking the transposase gene *tnpA*, IS*91* family members possess two functionally distinct ends[10]: the *ori*IS and the *ter*IS sequences, respectively initiating and terminating transposition. However, in IS*91*[10] and IS*1294*[6]/IS*1294b*[7], termination often fails at frequencies ranging from 1% to 10% and 1% to 14%, respectively, leading to the mobilization of adjacent DNA fragments, a process known as one-ended transposition (OET). Consequently, as IS*91*-like elements are often linked with virulence and antibiotic resistance genes, they could significantly influence the spread of such genes[10–12]. Yet their exact role in the dissemination of resistance genes remains elusive.

IS*91* family members do not generate target site duplications (TSD) upon insertion but display target sequence specificity. The recognized target for the insert is a tetranucleotide whose sequence is closely related for the four IS active in transposition, 5'-CTTG for IS*91*[13], 5'-GTTC for IS*801*[5], 5'-GTTC for IS*1294*[6] and 5'-GTCC for IS*1294b*[7]. The *ori*IS region of IS*801* and IS*91* contains two

imperfect subterminal palindromes, whereas IS*1294* and IS*1294b* feature a single presumed palindrome (Fig. 1). The *ter*IS region contains a single imperfect subterminal palindrome adjacent to the cleavage site, differing between IS*91* (5'-CTCG) and the other three members (5'-GTTC). Previous investigations have documented the *in vivo* formation of both single-stranded (ss) and double-stranded (ds) IS*91* circular intermediates during transposition, with both forms bearing a junction where *ori*IS and *ter*IS are juxtaposed[14]. However, it remains to be determined which intermediates play a functional role in the IS*91* transposition pathway.

Moreover, unlike other members of the IS*91* family, IS*91* itself features an additional open reading frame upstream of the *tnpA* transposase gene, called *orf121*, which encodes a 121 amino acid polypeptide (Fig.1). The termination codon of *orf121* overlaps with the initiation codon of *tnpA*, implying potential coupling in their expression. Prior work using a single-plasmid carrying an IS*91*-derived transposable cassette of the type IR_L-*kan*-IR_R and a transposase gene (*tnpA*) under the control of the P_{lac} promoter reported that induction of *tnpA* increases IS*91* transposition in an intramolecular assay by ~100-fold and that insertion mutants of *orf121* alter insertion-site choice[15]. Moreover, an earlier study reported that Tn*1732* insertions within *orf121* did not to affect the overall TnpA transposition frequency[15,16]. To our knowledge, no data about the distribution or sequence conservation of *orf121* across IS*91* elements have been published or are available.

Here we aimed to determine the involvement of Orf121 in IS*91* transposition within *Escherichia coli* and which intermediate is functional for transposition. We showed the crucial role of the overlapping region in *orf121* and *tnpA* co-expression. Using a genetic system based on the mating-out assay, we also showed that Orf121 has an inhibitory impact on IS*91* transposition *in vivo* and is necessary for the precise recognition and cleavage of the *ter*IS end. Furthermore, we established that only the bottom single-stranded DNA circular intermediates are inserted into new target sequences, and that Orf121 negatively influences this step.

Results

orf121 is a conserved feature in the IS*91* element

We first investigated whether *orf121* was conserved in the different isoforms of IS*91*. To retrieve a collection of IS*91* sequences to be screened for *orf121* content, we performed an *in silico* analysis using the amino acid sequence of the TnpA transposase encoded by IS*91* (accession number CAD30457.1) as a reference for the first described isoform. Sequences were filtered to retain those with a complete TnpA sequence with an amino acid identity of 95% or higher. Next, we applied an additional filter, retaining only nucleotide sequences covering the entire IS with an identity of 95% or higher relative to the IS*91* V1 reference (X17114.5). We identified 134 distinct isoforms of IS*91* among the 924 accession numbers analyzed, mainly in Enterobacterales, and as single copy on plasmids, with over half of the identified plasmids belonging to the IncFII type (Supplementary Table 1; Supplementary Data 1 and Supplementary Data 2). Orf121 encodes a full-length protein of 121 amino acids (aa) in 108 out of the 134 identified isoforms (Supplementary Data 2). Truncated versions of Orf121 were observed for 8 isoforms of variable lengths (26 aa, 29 aa, 75 aa or 82 aa). These truncated variants reflect point mutations introducing premature stop codons and/or single-nucleotide indels that shift the reading frame. Conversely, for 17 isoforms, Orf121 was found to be longer, including a 133 aa form (consistent with stop-loss) and long chimeric ORFs (516-517 aa) caused by frameshift/stop-loss at the *orf121-tnpA* junction that places both ORFs in the same reading frame by single/or two-nucleotide indels. The one base overlap between the *orf121* and *tnpA* genes was highly conserved, in 80.6 % of cases. In isoforms lacking the canonical −1 nt *orf121-tnpA* overlap, (TGA|ATG), the absence is attributable to small indels (± 1-2 nt) within homopolymeric runs (e.g., poly-A/T) inside *orf121*, which shift the stop-start junction and abolish the overlap. The target insertion sites showed a consistent preference for 5'-GTTC or 5'-CTTG, followed by 5'-CTCG, with a T consistently in the second position

(Supplementary Table 1; Supplementary Data 2). The predominant cleavage site tetranucleotide was 5'-CTCG (Supplementary Table 1; Supplementary Data 2), which is in line with previous reports[10,13]. For subsequent experimental assays, we used the canonical IS*91* V1 isoform (NCBI X17114.5), originally described from an *Escherichia coli* α-hemolysin plasmid[13].

P*orf121* drives expression of both *orf121* and *tnpA* genes

The overlap of the termination codon of *orf121* with the initiation codon of the *tnpA* gene in IS*91*, suggests that the two genes may be expressed from the same promoter.

Using BPROM (http://www.softberry.com)[17], we identified putative σ^{70} promoters and Shine-Dalgarno (SD) sequences for both genes, the putative *tnpA* promoter P*tnpA* being located within the coding sequence of *orf121* and P*orf121* in the *ter*IS region (Fig. 2A). The activity of the predicted promoters P*tnpA* and P*orf121* was assessed by generating transcriptional fusions with the *lacZ* gene, including the *lacZ* ribosome binding site (RBS) (Fig. 2B). β-galactosidase assays revealed that both promoters are functional, with P*orf121* exhibiting significantly higher activity than P*tnpA* (62.8-fold; Fig. 2B).

We thus estimated the transcriptional expression level of *orf121* and *tnpA* in their natural configuration within the IS*91* element to assess the significance of the overlap between the *orf121* stop codon and the *tnpA* start codon. To investigate this expression, we made two constructs in which the *tnpA* gene was replaced by the *lacZ* gene in the native organization (i.e. with the −1 overlap) or without the overlap, by insertion of one nucleotide between the TGA stop of *orf121* and the ATG start of *lacZ* (called Δ−1, no-overlap, +1 nt) leading to P*orf121*-*orf121*(P*tnpA*)::*lacZ* and P*orf121*-*orf121*(P*tnpA*)::*lacZ* Δ−1, respectively (Table 1, Fig. 2C). Additionally, we created two new *lacZ* transcriptional fusions with P*tnpA*, including its native RBS, either retaining or removing the -1 overlap leading to P*tnpA*::*lacZ* and P*tnpA*::*lacZ* Δ−1, respectively (Table 1, Fig. 2C). Globally, the β-galactosidase activities were much lower in constructs in their native configuration (Fig. 2C).

The removal of the -1 overlap between *orf121* and *tnpA* resulted in a 4.1-fold increase in P_{tnpA} activity (Fig. 2C; P_{tnpA}::*lacZ* Δ−1 vs P_{tnpA}::*lacZ*) and in a significant 2.9-fold decrease in P_{orf121} β-galactosidase activity (Fig. 2C; (P_{orf121}-*orf121*$_{(PtnpA)}$)::*lacZ* Δ−1 vs (P_{orf121}-*orf121*$_{(PtnpA)}$)::*lacZ*). When under the control of P_{tnpA} alone, the β-galactosidase activity was minimal (<1 MU), but it significantly increased when the P_{orf121} promoter was present upstream (P_{orf121}-*orf121*$_{(PtnpA)}$) with and without the -1 overlap (Fig. 2C). Moreover, the overlap (**TGATG**) yields higher β-galactosidase activity than the +1-nt spacer/no-overlap (Δ−1) version for the P_{orf121}-*orf121*$_{(PtnpA)}$::*lacZ* construct (Fig.2C, different letters in the compact-letter display). To further investigate the contributions of the P_{orf121} and P_{tnpA} promoters to *tnpA* expression, we engineered mutants of these promoters within the P_{orf121}-*orf121*$_{(PtnpA)}$::*lacZ* construct. Specifically, we mutated P_{orf121} in the -10 box (TATAAA to CGCGAA) and P_{tnpA} in the -35 box (TTGCCG to TGGCGG) without altering the amino acid sequence of Orf121. The activities of the mutated promoters showed significant reductions, with a 3-fold decrease with the mutant P_{tnpA*} and an 18.1-fold decrease with the mutant $P_{orf121*}$ (Fig. 2D).

Altogether, these results suggest that *tnpA* is predominantly expressed from the P_{orf121} promoter, and that its expression is enhanced by the overlap between *orf121* and *tnpA*.

Inhibitory role of Orf121 in IS*91* transposition

To further explore the potential regulatory role of *orf121* in IS*91* transposition *in vivo*, we performed mating-out assays in *E. coli* using a genetic system composed of three plasmids to assess the mobility of IS*91* derivatives into the pOX38Km conjugative plasmid upon expression of *tnpA*, alone or with *orf121 in trans* (relative to IS*91*) (see Fig. 3A, and Materials and Methods). Two different IS*91* derivatives were cloned into a high-copy number plasmid (pUC18). Both derivatives include a sequence of 89 bp from the left-end of IS*91* at the *ter*IS cleavage site (5'-CTCG) and the final 82 bp of the right-end, shown to be sufficient for functional transposition[9], encompassing the

*ori*IS as well as the essential 5'-CTTG tetranucleotide adjacent to the *ori*IS end. The *ter*IS and *ori*IS sequences flanked a Cm^R cassette (for selection of transposition events) in the IS*91* derivative termed mini-IS *ter*IS-*cm*^{R}. The other derivative, *ter*IS::*orf121-cm*^{R}, also contains compared to *ter*IS-*cm*^{R}, the *orf121* gene in its native configuration with *ter*IS, allowing it expression *in cis* relative to IS*91* (Table 1). *tnpA* was expressed *in trans* relative to the IS derivative from a compatible plasmid under the control of the inducible P_{lac} promoter, either alone, in P_{lac}::*tnpA*, or with *orf121* in P_{lac}::*orf121-tnpA* (with the native configuration with overlapping *orf121*-stop/*tnpA*-start) (Table 1). Transposition frequencies of IS*91* derivatives into pOX38Km were evaluated as illustrated in Fig. 3A.

In cells carrying the mini-IS *ter*IS-*cm*^{R}, a basal transposition activity was detected due to P_{lac}::*tnpA* leakage (in absence of IPTG). Upon IPTG induction, the transposition frequency increased significantly by a 70-fold factor (4.68×10^{-5} vs 3.28×10^{-3}; Fig. 3B). Under native conditions, when *tnpA* and *orf121* were co-expressed (P_{lac}::*orf121-tnpA*) *in trans* relative to *ter*IS-*cm*^{R}, we observed a strong reduction in transposition frequency compared to *tnpA* expression alone (P_{lac}::*tnpA*), 20.7-fold in the absence of IPTG and 8045-fold upon IPTG induction (Fig. 3B), indicating that Orf121 acts as a potent inhibitor of IS*91* transposition.

The significant 11.1-fold reduction in transposition frequency of the mini-IS *ter*IS::*orf121-cm*^{R} (*orf121* expressed *in cis*) compared to *ter*IS-*cm*^{R} in conditions of basal expression of *tnpA* (P_{lac}::*tnpA* without IPTG induction), appears to support the inhibitory role of Orf121 (Fig. 3B). However, in the presence of IPTG, no significant difference in transposition frequency was observed between *ter*IS::*orf121-cm*^{R} and *ter*IS-*cm*^{R}, suggesting that the inhibitory effect of Orf121 may be counterbalanced when *tnpA* expression is fully induced (P_{lac}::*tnpA*) (Fig. 3B).

We then examined whether elevated Orf121 levels could inhibit transposition even when *tnpA* expression is fully induced by IPTG, thereby reinforcing the idea that Orf121 actively regulates

transposition frequency. We assessed the transposition frequency of the *ter*IS-*cm*^{R} derivative in *E. coli*, where *tnpA* and/or *orf121* were expressed independently from the inducible promoters P_{lac} and $P_{LtetO-1}$, respectively, designated as *orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA* (Fig. 3C). In the absence of IPTG and anhydrotetracycline (aTc) inducers, we observed the same basal transposition frequency due to P_{lac}::*tnpA* leakage. When *orf121* was expressed upon aTc induction, we observed a decrease of the basal transposition frequency by 2.25-fold. When *tnpA* expression was induced with IPTG alone, transposition frequency increased significantly by 45.1-fold relative to the basal level. However, co-expression of *orf121* with *tnpA* in presence of both inducers aTc and IPTG, led to an 8.6-fold reduction in transposition frequency relative to IPTG alone. This result reveals that, aside from the inherent configuration of the *orf121* and *tnpA* genes impacting element mobility, the protein Orf121 modulates transposition frequency. However, the strongest inhibitory effect on the transposition frequency of *ter*IS-*cm*^{R} was observed when *orf121* and *tnpA* were co-expressed in their native configuration with the -1 frameshift (P_{lac}::*orf121-tnpA*) upon IPTG induction of the P_{lac} promoter (Fig. 3B) suggesting that besides a direct inhibition of Orf121 on TnpA activity, there was also a potential decrease in *tnpA* expression, and/or TnpA protein amounts in this condition.

We thus quantified the transcripts of *orf121* and *tnpA* in the mating-out assays. Transcript levels were quantified in log-phase donor cultures just before mating. Our results revealed that co-expressing *orf121 in trans* with *tnpA*, whether from the same or independent promoters (mating-out with *ter*IS-*cm*^{R}), resulted in a significant reduction in the mRNA copy number of *tnpA* by 6.7-fold and 1.8-fold, respectively (Fig. 4A). However, co-expression of *orf121* and *tnpA* from the same promoter (P_{lac}::*orf121-tnpA*) resulted in a significant reduction in the mRNA copy number of *tnpA* compared to when the two genes were expressed from independent promoters by 8.5-fold (Fig. 4A). Similar trends were observed when *orf121* was co-expressed *in cis* (*ter*IS::*orf121-cm*^{R})

with *tnpA* (P_{lac}::*tnpA*), with a significant 1.9-fold decrease in *tnpA* transcript levels compared to *ter*IS-cm^R ($p=6.56e^{-5}$). Although, there was no impact of *tnpA* expression on the mRNA copy number of *orf121 in cis* (P_{orf121}) or *in trans* ($P_{LtetO-1}$), the transcript number of *orf121* when co-expressed with *tnpA* in the native configuration was 3052.8-fold higher than *tnpA* transcript number (P_{lac}) (Fig. 4B). Our results suggest that, in the native configuration with the co-expression of *orf121* and *tnpA* with the -1 frameshift, *orf121* expression negatively impacts *tnpA* transcription. These observations also raise the possibility that Orf121 plays a role in stabilizing *tnpA* mRNA.

To assess the protein levels of Orf121 and TnpA, we performed Western-immunoblotting on bacterial lysate of mating-out assays done with C-terminal His-tagged TnpA (*tnpA*^{6his}) and N-terminal His-tagged Orf121 (^{6his}*orf121*) proteins (Table 1). Gel loading was equalized across conditions based on Bradford-quantified total protein. His-tags did not impact transposition frequencies (Fig. S1). Orf121 was exclusively detected in the soluble fraction of the bacterial lysate, whereas TnpA was present in the pellet as inclusion bodies. The identities of ^{6his}Orf121 and TnpA^{6his} were verified via mass spectrometry. As shown in Fig. 4C, when *orf121* and *tnpA* were expressed independently (*orf121*^{6his}::$P_{LtetO-1}$-P_{lac}::*tnpA*^{6his}), the quantity of TnpA was similar, around 20 AU (Arbitrary Units), regardless of whether *tnpA* was expressed alone or with *orf121*. However, when *tnpA* and *orf121* were expressed from the same promoter (P_{lac}::*orf121-tnpA*^{6his}), the quantity of TnpA was 4.1-fold lower. Orf121 production was similar across all the tested conditions (Fig. 4D). These findings correlate with the transcript results and with the reduction of transposition frequency (Fig. 3B).

Lack of Orf121 influence on IS*91* target site insertion

Previous research suggested that Orf121 might influence insertion specificity, as *orf121* mutations affected target selection without impacting transposition[15,16]. To confirm that the observed Cm^R exconjugants were due to transposition of the IS*91* derivative into the pOX38Km plasmid, rather

than spontaneous mutations, arbitrarily primed PCR (AP-PCR) analysis was conducted on numerous transposition events to identify target insertion sites (Table 2). We performed a χ^2 test to evaluate whether the presence or absence of *orf121*, as well as its *cis* and *trans* expression, affected the insertion target site preferences at 5'-GTTC and 5'-CTTG. The data revealed that the target site specificity was consistent across conditions, with *orf121* expressed either *in cis* (*ter*IS::*orf121-cm*^{R}) or *trans* (P_{lac}::*orf121-tnpA*) and was similar to that observed with *tnpA* alone (P_{lac}::*tnpA*) (Table 2 and Fig. S2A) (χ^2 value = 22.20; p-value = 0.45). Additionally, no significant differences in target specificity were detected when *orf121* and *tnpA* were controlled by distinct promoters (*orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA*) (Table 2 and Fig. S2B) (χ^2 value = 3.77; p-value = 0.29). These results suggest that the expression of Orf121, whether in *cis* or *trans*, does not alter the specificity of target selection or insertion sites under our experimental conditions.

Orf121 is a key factor in precise *ter*IS end recognition and cleavage

We explored the involvement of Orf121 in the specific recognition and cleavage of the *ter*IS site by analyzing the tetranucleotide cleavage patterns at the *ter*IS end across multiple transposition events (Table 3). AP-PCR analysis consistently identified a single cleavage site (5'-CTCG) in all instances, whether the transposase was expressed independently (P_{lac}::*tnpA* with *ter*IS-*cm*^{R}; 43 events) or alongside *orf121 in trans* (P_{lac}::*orf121-tnpA* with *ter*IS-*cm*^{R}; 68 events).

Subsequently, we investigated the proportion of one-ended transposition (OET) in different mating-out assays, given the known inefficiency of the IS*91 ter*IS end. As detailed in Table 3, expressing *tnpA* alone (P_{lac}::*tnpA* or *orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA* with IPTG only) led to approximately half of OET events. In contrast, the co-expression of *tnpA* and *orf121 in trans* either from P_{lac}::*orf121-tnpA* or *orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA* with IPTG and aTc significantly lowered OET rates by 2.94- and 1.5-fold, respectively (relative to the conditions with *tnpA* only; Table 3). Similarly, *orf121* expression *in cis* (P_{lac}::*tnpA* with *ter*IS::*orf121-cm*^{R}) reduced OET incidence to

29%, marking a substantial 1.89-fold decrease compared to the condition without *orf121* (P_{lac}::*tnpA* with *ter*IS-cm^R). Notably, even without *tnpA* induction, the Orf121 inhibitory effect on OET was manifest, as shown by a 2.64-fold reduction in OET when the $P_{LtetO-1}$ promoter (*orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA*) was derepressed (11% with aTc; 29% without aTc, Table 3). These results highlight that Orf121 improves the precision of recognition and cleavage at the *ter*IS boundary.

Bottom single-stranded IS*91* yields a functional circular intermediate

It has been previously established that two distinct DNA species, single-stranded (ss) and double-stranded (ds) IS*91* circular intermediates, are independently formed *in vivo* post-induction of transposase expression, with *ori*IS and *ter*IS fusion[14]. Nevertheless, the functional intermediates facilitating IS*91* insertion have yet to be clarified. To address this, we employed a suicide mating assay of the pSW23T plasmid carrying the *ter*IS-*ori*IS (*ter91*-*ori91*) or *ori*IS-*ter*IS (*ori91*-*ter91*) junctions (Fig. 5A; Table 1) into *E. coli* DH5α strains harboring various *tnpA* or *tnpA*/*orf121* expression plasmids (P_{lac}::*tnpA* or P_{lac}::*orf121*-*tnpA*). The ds or ss nature of the IS*91* circular intermediate was assessed respectively via transformation or conjugation as presented in Fig. S3. In these assays, the plasmid carrying the functional junctions being non replicative in the recipient strain, will be inserted in the bacterial chromosome through the action of TnpA, and Cm^R transformants or ex-conjugants will be selected.

For experiments with dsDNA, apart from the positive transformation control, pUC18 plasmid, which yielded the expected transformation frequency, no transformants were detected with any of the ds-circular junction intermediate *ter*IS-*ori*IS or *ori*IS-*ter*IS (Fig. 5C).

For ssDNA experiments, no Cm^R exconjugants were observed as expected with donor cells carrying the empty plasmid pSW23T (lacking *ter*IS-*ori*IS or *ori*IS-*ter*IS junction) regardless of the recipient cells, P_{lac}::*tnpA* or P_{lac}::*orf121*-*tnpA* (Fig. 5C). Likewise, no Cm^R exconjugants could be

detected when the recipient cells received the ssDNA junction in the *ter*IS-*ori*IS orientation. Conversely, when donor cells carried the IS*91* junction in the *ori*IS-*ter*IS orientation, we observed a basal insertion activity with *tnpA* alone, which increased upon inducer addition. A significant reduction in basal insertion frequency (absence of IPTG) was detected when *orf121* and *tnpA* were co-expressed from the same promoter (P_{lac}::*orf121-tnpA*) compared to *tnpA* alone (P_{lac}::*tnpA*) (Fig. 5C). Upon IPTG induction, the transposition frequency of the *ori*IS-*ter*IS junction was detectable in cells expressing *tnpA* alone but became undetectable in cells co-expressing *orf121* and *tnpA*, with a substantial 12347-fold reduction. Inhibitory action of Orf121 on transposition was confirmed when *orf121* and *tnpA* were expressed from different promoters (*orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA*) (Fig. 5D).

Molecular analysis via arbitrarily primed PCR (AP-PCR) of numerous exconjugant clones confirmed all Cm^R exconjugants originated from *ori*IS-*ter*IS junction insertion into the recipient cell's chromosome (Supplementary Table 2).

Our results indicate that within the sensitivity limit of our assay conditions, only the bottom-strand ssDNA circle is engaged in the process of initiating insertion. They also confirm that Orf121 acts as an inhibitor, reducing the insertion efficiency of the circular intermediate mediated by TnpA.

Discussion

IS*91*, as the prototype for both the IS*91* family and the related IS*CR*s family, has often been associated with or located near antibiotic resistance genes[18–22], suggesting that these IS elements might play a role in the mobilization of antibiotic resistance. Despite its biological and clinical relevance, this family remains poorly characterized. In this study, we mechanistically explored IS*91* transposition and showed the crucial role of Orf121 in TnpA transposition efficacy and accurate excision during transposition.

The *orf121* and *tnpA* genes overlap by one nucleotide between the stop codon of *orf121* and the start codon of *tnpA* (**TGA**TG; Fig. 2A). This co-directional overlap is preceded by a Shine-Dalgarno sequence. We showed that *tnpA* is mainly co-expressed with orf*121* from the promoter P_{orf121} and that the overlap allows a higher expression of *tnpA* (Fig. 2C). Using RNAfold[23], we identified highly complex conformation in the *orf121-tnpA* mRNA overlap region, indicating that *tnpA* translation signals might be sequestered within a secondary structure (Fig. S4A), as observed in IS*10*[24] and IS*3*[25]. Such RNA secondary structures have been proposed to facilitate ribosome pausing and recycling, thereby promoting translation reinitiation[26–28]. Through scanning of the sequence by the ribosomes, the presence of a start ATG codon close to a stop codon redirects initiation from the start codon. Such a reinitiation mechanism has been proposed in bacterial systems where overlapping ORFs are separated by motifs such as A**TGA** or **TGA**TG[26,29]. This translational coupling might play a role in regulating *tnpA* gene expression. As translational coupling can also occur between adjacent stop and start codons without overlap, the modification of the spacing between the Shine-Dalgarno (SD) sequence and the ATG may also contribute to the reduced expression of *lacZ* in constructs lacking the overlap (P_{orf121}-*orf121*(P_{tnpA})::*lacZ* Δ−1)[30]. We also found that both *tnpA* and *orf121* promoters showed drastically reduced activity when they drive the expression of their native sequences when compared with their activity when placed

upstream of the SD-*lacZ* cassette (Fig. 2). Additionally, we showed that expression of *orf121 in cis* or *in trans* has a negative effect on the number of *tnpA* transcripts (Fig. 4A), and that the production of the transposase TnpA is reduced in the native *orf121-tnpA* configuration (Western blot analysis Fig. 4C). Therefore, we can hypothesize mRNA$_{orf121}$-mRNA$_{tnpA}$, mRNA$_{orf121}$-TnpA or Orf121-mRNA$_{tnpA}$ interactions, or a role of Orf121 in stabilizing *tnpA* mRNA. Overall, our results argue for a complex regulatory network of *tnpA* expression.

For Orf121, we observed that, despite abundant *orf121* transcripts upon induction of P$_{lac}$::*orf121-tnpA*, there was a lack of proportional Orf121 accumulation (Fig. 4D). This mRNA-protein decoupling likely reflects post-transcriptional constraints (e.g., mRNA structure/RBS accessibility, ribosome allocation *in cis* arrangement).

We also showed that Orf121 has a regulatory role in the transposition of IS*91*. Our *in vivo* mating-out assays showed that Orf121 exerts a strong negative effect on TnpA-mediated transposition frequency. In both conditions of Orf121 expression, *cis*-encoded Orf121 and *trans*-encoded Orf121(Fig. 3B-C), we observed a decreased transposition frequency under basal P$_{lac}$::*tnpA* expression (absence of IPTG, no *tnpA* induction). However, this effect disappeared when P$_{lac}$::*tnpA* was induced (+IPTG) with *cis*-encoded Orf121 (TnpA saturating conditions), but not when trans-encoded Orf121 was induced from an independent promoter (important amounts of both proteins; Fig. 3C) or in the native configuration (greater amount of Orf121 than TnpA; Figs. 3B and 4C-D). These results strongly suggest that the stoichiometry between the two proteins plays a role in the regulation of TnpA activity. It should be noted that the difference in length (371 bp) between the two constructs *ter*IS::*orf121-cm*^{R} and *ter*IS-*cm*^{R} may also affect the transposition frequency (as for Tn7-like[31]).

In a previous study, Mendiola *et al.* reported slight differences in transposition frequencies with insertion mutants in *orf121* and considered that these differences were not significant. They

therefore concluded that Orf121 did not affect TnpA-mediated transposition frequency[16]. This discrepancy with our findings can be explained by differences in methodological assays, plausibly accounting for the differing readouts: i) we used a different transposition reporter plasmid for mating-out assay (IncF pOX38[32] vs IncW R388[33]) which have distinct replication and conjugative-transfer systems ii) we used wild-type Orf121 and various Orf121:TnpA stoichiometry *in cis* [P_{orf121}] and *in trans* [P_{lac} or $P_{LtetO-1}$] when Mendiola *et al.* used *orf121* insertion mutants within a native IS*91*, iii) we used a mini-IS whereas Mendiola *et al.* used full length IS*91* with insertion in *orf121*, these differences in length might affect transposition frequency.

The negative effect of Orf121 on transposition could be explained by a direct interaction between the Orf121 and TnpA proteins, or alternatively, by a competition of Orf121 with TnpA for binding to the *ori*IS and/or *ter*IS. Such a regulatory mechanism by competition for binding has been shown for InsA, a protein encoded by IS*1*, which specifically binds to the terminal inverted repeats (IRs) of this IS, thereby negatively regulating IS*1* transposition activity[34]. This has also been shown in the IS*3* family, where OrfA, encoded by the IS, inhibits transposase-mediated transposition events by binding to the IRs[25]. We identified in the Orf121 peptide sequence potential leucine zipper motifs (V49-X_5-L55-X_5-V61 or V61-X_6-L68-X_6-V75) that could be involved in DNA binding[35], along with a classic zinc finger motif (C105-X_2-C108), which might facilitate the interaction with the TnpA transposase or Orf121 dimerization[36] (Fig. S4B).

We also demonstrated that the transposase recognizes specifically the bottom single-strand circular intermediate and that Orf121 impacts the transposition process negatively by reducing the insertion efficiency of the bottom strand. Mechanistically, this inhibition may involve Orf121 competing with TnpA for binding to the insertion machinery or interfering with the proper recognition of target sites by forming specific protein-DNA or protein-protein interactions. Additionally, the interaction between Orf121 and the transposition complex could restrict the availability or

conformation of the bottom-strand circular intermediate, thereby limiting its ability to integrate into the recipient DNA. Under our assay conditions, the bottom-strand ssDNA circle yielded detectable transposition compared to the top-strand ssDNA circle, whereas ds circles were not detected. These data also support a negative impact of Orf121 on the insertion efficiency of the circular intermediate.

In addition to its inhibitory effect on transposition, our data also showed that Orf121 enhances the accurate recognition and cleavage of *ter*IS, thereby reducing the occurrence of OET. It has been previously shown that IS*91*-type elements can mobilize up to 40 kb[37]. However, it should be noted that our OET analysis method differs from previous studies. Indeed, in studies involving IS*91*[9] or IS*1294*[6], all genes were on the same plasmid, and for IS*1294b*, the analysis was based on inactivation of the *sacB* gene[7], which provides only a limited number of accessible insertion sites. This restricts the assessment of transposition dynamics and may underestimate insertion frequency and target site diversity. In our study, we used three plasmids, the pOX38Km plasmid as a transposition reporter plasmid with more than 1000 possible insertion sites, and IS*91*-derived inserted into a high copy number plasmid (pUC18) and the transposase expressed from a p15A-derived plasmid; therefore, the IS*91*-derivative is not limiting, unlike TnpA, enabling more transposition events to be captured. Moreover, our *ter*IS cleavage analysis clearly indicates that *orf121* expression does not affect target-site selection, and that cleavage consistently occurs at 5′-CTCG. These findings align with the *ter*IS conserved features identified in our *in silico* analysis, which showed that the predominant cleavage site tetranucleotide was 5’-CTCG (Supplementary Data 1 and Supplementary Data 2).

Bernales *et al*[15], proposed a role of Orf121 in insertion specificity and target site based on transposition assay of a transposable cassette IR_L-kan-IR_R into the plasmid R388 using Orf121 insertion mutants and a P_{lac}-expressed *tnpA*, and the analysis of a limited number of clones at the

*ori*IS end (6 for each of the 3 *orf121* mutants). Our results, derived from the characterization of at least 50 insertion events, clearly indicate that *orf121* expression does not affect target site selection, and revealed that cleavage at the insertion site consistently occurs at 5'-CTTG or 5'-GTTC with no detectable preference, regardless of *orf121* expression, and a conserved T is retained at position 2 in all motifs analyzed. However, Bernales *et al*[15] used experimental assays that differ from ours (different transposition reporter plasmid and different transposable cassette), likely probing distinct mechanistic layers (site gating vs transposase availability/termination). However, our *in vivo* results diverge from our *in silico* analysis, in which 5′-GTTC emerged as the predominant target-site tetranucleotide (Table 2; Supplementary Table S1 and Supplementary Table S2; Supplementary Data 2). *In vivo* profiles can diverge from *in silico* because public plasmid repositories are not random samples: they reflect targeted sequencing campaigns (e.g., clinical/veterinary surveys), often contain clusters of near-identical plasmids, and may include redundant accessions for the same isoform factors, which can over-represent the target site. This caveat is evident in resources and audits such as PLSDB[38], COMPASS[39], and an INSDC[40] analysis documenting duplicates and redundancies.

The *in silico* analysis also revealed that *orf121* is a highly conserved feature in IS*91* isoforms, of which 80.6 % contain the overlap between the *orf121* stop codon and the *tnpA* start codon. This evolutionary conservation of the *orf121-tnpA* overlap of IS*91* isoforms suggests a selective pressure to maintain the translational coupling regulatory mechanism, ensuring that IS*91* remains active enough for mobilization while preventing uncontrolled transposition. Indeed, an overactive transposition can be detrimental to bacterial fitness by disrupting essential genes or imposing a metabolic burden due to excessive IS mobilization[41,42]. The presence of *orf121* in most IS*91* elements may have provided a selective advantage by fine-tuning IS activity in response to environmental conditions or the host genome context.

Our study highlights a pleiotropic role of Orf121 in IS*91* transposition, underscoring its dual function: i) inhibiting transposition frequency, and (ii) enhancing precise cleavage at *ter*IS, thereby limiting the spread of adjacent genes (Fig. S4C).

These findings not only provide a better understanding of IS*91* biology but also of the regulation of transposable elements associated with the dissemination of resistance genes.

Methods

Bacterial strains, media and growth conditions.

Bacterial strains and plasmids used in this study are listed in Table 1. All *E. coli* strains were grown in Lysogeny Broth (LB) at 37°C under agitation at 300 rpm. Media were supplemented, when necessary, with the appropriate antibiotics used at the following concentrations: rifampicin 50 µg/mL, spectinomycin 50 µg/mL, kanamycin 25 µg/mL, chloramphenicol 25 µg/mL, nalidixic acid 25 µg/mL, and ampicillin 100 µg/mL. The expression of IS*91 tnpA* alone or the co-expression with *orf121* from the P_{lac} promoter was induced by adding 0.5 mM IPTG to the media. The expression of *orf121* from the $P_{LTet\text{-}O1}$ promoter was induced by adding 50 nM anhydrotetracycline (aTc) to the media. Transformation of *E. coli* with plasmid DNA was performed as previously described[43]. All plasmid constructions were verified by DNA sequencing.

DNA manipulations.

PCR reactions were carried out using Phusion DNA Polymerase (ThermoScientific) to amplify the fragments used for cloning and Quick-Load® Taq 2X Master Mix (NEB) for all other applications. All PCR reactions were purified using MACHEREY-NAGEL's NucleoSpin Gel and PCR Clean-up kits according to the manufacturer's recommendations.

IS*91* derivate insertion sites in the pOX38Km reporter plasmid or the *E. coli DH5α* genome, were mapped by arbitrary priming PCR (AP-PCR). The first PCR round was performed in a final volume of 50 µl containing 0.8 µM of primers (arbitrary primer ARB1 and chloramphenicol resistance gene-specific primer, Cm-1005) and 15 µl of template (a single Cm^R colony from mating-out experiments lysed in 20 µl). PCR was performed as follows: 2 min 95°C, 6 cycles of 30 sec 95°C, 30 sec 30°C, 1 min 30 sec 72°C; 30 cycles of 30 sec 95°C, 30 sec 45°C, 2 min 72°C; and finally, 72°C for 5 min. The second PCR cycle was performed in a final volume of 50 µl

containing 0.8 μM of primers (arbitrary primer ARB2 and Cm-1049) and 5 μl of the purified PCR product from the first cycle as template. PCR was performed as follows: 2 min 95°C, 30 cycles (30 sec 95°C, 30 sec 57°C, 2 min 72°C) and a final step 72°C for 5 min. The PCR products were directly sequenced using the Ori91-1104 primer.

The *ter*IS cleavage site was mapped using the same procedure with primer pairs ARB1/Cm-473 for the first cycle of PCR and ARB2/Ter91-337 for the second cycle, and the PCR products were directly sequenced with the Ter91-276 primer.

Oligonucleotides used for PCR amplification of DNA fragments required for plasmid construction, diagnostic PCR or sequencing are described in Supplementary Table 3.

***In silico* analysis of IS*91* DNA sequences from the public database.**

Using the amino acid sequence of the TnpA transposase isoform V1 encoded by IS*91* in *Escherichia coli* (NCBI accession number CAD30457.1) as the query sequence, a search was performed with NCBI's basic local alignment search tool (BLAST). The search was performed using BLASTP (NCBI) with default parameter arguments. Resulting matching sequences were filtered to discard matches with less than 95% identity to the TnpA query sequence. Recovered sequences in which TnpA were partial or truncated were discarded. For each TnpA hit with ≥95% amino-acid identity across the alignment, we manually retrieved the corresponding nucleotide accessions (chromosomal/plasmid sequence entries) using the Identical Protein Groups (IPG) record of that protein. When an IPG contained more than 100 nucleotide entries, we collected the first 100 records. Each nucleotide sequence was then aligned to the IS*91* V1 reference (NCBI X17114.5) to delimit the full-length IS. Sequences that did not span the entire IS (partial/truncated IS) or showed less than 95% identity were excluded from analyses. We then defined isoforms as unique full-length IS sequences that differ by at least one nucleotide from IS*91* V1. Each isoform represents a group of one or more nucleotide accessions. For each accession number recovered,

we analyzed the target insertion and cleavage sites, as well as data related to the bacterial host, the host of the IS and the physical location of the IS (plasmid or chromosome). For each retained record, we annotated the physical location of the IS using the plasmid database (PLSDB)[38,44] by accession lookup when the nuccore accession was indexed in PLSDB. When PLSDB returned no match, we inspected the regions surrounding the IS for the presence of plasmid-specific markers (e.g., *repA*, *stbA*/*B*, *psiA*/*B*…) or chromosomal markers (e.g., *dnaA*, *seqA*, *dnaT*…) to assign location. The data harvest date was 2020-10-15; the nucleotide reference for identity comparisons and cloning was IS*91* V1 (X17114.5).

Estimation of *in vivo* transposition frequencies of IS*91* derivatives into *E. coli* by mating-out assay.

Mating-out assays were performed as previously described[45]. The donor strain was *E. coli* JS219[46] carrying three plasmids: i) *ter*IS-*cm*^{R} or *ter*IS::*orf121-cm*^{R} (used as IS*91*-Cm^{R} derivatives donor plasmids P_{IS}; Amp^{R}), ii) P_{lac}::*tnpA*, P_{lac}::*orf121-tnpA* or *orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA* (used as transposase +/- *orf121* expression plasmids; Sp^{R}) and iii) pOX38Km, the conjugative target plasmid (Km^{R}) (Table 1, Fig. 3A). The recipient strain was *E. coli* MC240 (Rif^{R}, Nal^{R}). In brief, after overnight growth, donor and recipient cultures were diluted 1:100 into 10 mL LB supplemented, where indicated, with the appropriate inducers (IPTG and/or aTc), and grown to $OD_{600} \approx 0.5$ for donors (to induce the transposition machinery) and $OD_{600} \approx 1$ for recipients. Cultures of donor and recipient cells were pooled (ratio 1:1) and incubated for conjugation for 75 minutes. Following mating of the donor with the recipient strain, cells were plated on LB plates supplemented with rifampicin and kanamycin (count of exconjugants i.e. all recipients carrying pOX38Km without and with IS*91*-Cm^{R} derivatives) or nalidixic acid, kanamycin and chloramphenicol (count of transposants i.e. recipients carrying pOX38Km with IS*91*-Cm^{R} derivatives). Transposition frequency was determined by dividing the CFU of transposants (Nal^{R},

Km^R and Cm^R) by the CFU of exconjugants (Nal^R, Km^R). Transposition event frequencies were estimated from more than 5 independent experiments. These experiments were used to calculate mean values and standard deviations.

Estimation of the proportion of one-ended transposition events in mating-out assay.

Several hundred clones resulting from transposition were subcultured onto both LB plates supplemented with kanamycin, ampicillin and nalidixic acid (numeration of recipient cells with pOX38Km carrying the whole derivatives donor plasmids P_{IS} resulting from OET), and LB plates supplemented with kanamycin, chloramphenicol and nalidixic acid (Total population of recipients carrying pOX38Km with IS*91*-Cm^R derivatives). The percentage of OET was determined by dividing the number of (Km^R, Amp^R and Nal^R) clones by the number of (Km^R, Cm^R and Nal^R) clones and multiplying by 100.

Estimation of *in vivo* insertion of IS*91* circular intermediates into *E. coli*.

To test the double-stranded circular intermediate (Fig. S3A), *E. coli DH5α* cells were rendered competent in the presence or absence of the inducer IPTG (0.5 mM) containing the plasmid P_{lac}::*tnpA* or P_{lac}::*orf121-tnpA*. These competent cells were transformed with 500 ng of suicide plasmids *ori*IS-*ter*IS or *ter*IS-*ori*IS-containing the junction in the pSW23T derivative (Table 1, Fig. 5A). Cells were heat-shocked at 42°C for 1 min, followed by 5 min on ice. The cells were then placed in 1 ml LB with or without inducer for 1 hour at 37°C with agitation, then plated. Determination of the total number of viable cells was performed on LB plates supplemented with spectinomycin (Sp^R), and insertion of junctions into the *E. coli DH5α* genome was selected on LB plates supplemented with spectinomycin and chloramphenicol (Sp^R, Cm^R).

To test the single-stranded circular intermediate (Fig. S3B), overnight cultures with Cm^R + DAP (0.3 mM) for donors (*E. coli β2163* (*pir*-) carrying the suicide plasmid (pSW23T (Empty), *ori*IS-

*ter*IS or *ter*IS-*ori*IS; Table 1) and Sp^R for recipients (*E. coli DH5α* target strain containing different *tnpA* and/or *orf121* expression plasmids (P_{lac}::*tnpA*, P_{lac}::*orf121-tnpA* or *orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA*; Table 1) were diluted 1:100, without antibiotics, and further grown to OD_{600} = 0.3. Expression of *tnpA* and/or *orf121* was then induced by adding respectively 0.5 mM IPTG and/or aTc at 50 ng/ml in the recipient cells cultures. The strains were grown for a further 1 h at 37°C with agitation. Filters containing the mixture of donor and recipient strains in a 1:8 ratio were incubated for 3 h at 37°C on LB plates supplemented with 0.3 mM DAP and 0.5 mM IPTG, or 0.3 mM DAP, 0.5 mM IPTG and 50 ng/ml aTc. Cells were then suspended in 2 mL LB by vortexing the filter, and appropriate dilutions were plated on selective LB plates (Sp^R) for recipients and LB plates (Sp^R + Cm^R) for transconjugants. The frequency of Cm^R transconjugants was calculated as the number of Cm^R-tagged recipient cells relative to the total number of recipient cells. The frequencies of the insertion events per viable cell from 4 independent experiments were used to calculate mean values and standard deviations.

Quantification of *orf121* and *tnpA* transcripts.

For mating-out transcripts, overnight cultures were diluted to OD_{600} = 0.2 into 10 mL LB supplemented with the indicated inducers (IPTG and/or aTc) and grown to the mean log phase (OD_{600} = 0.5). Total RNA was extracted from the donor culture immediately before mating with the recipient. Cells were pelleted and total RNA was extracted with the NucleoSpin® RNA Extraction Kit (Macherey-Nagel Inc.). Contaminating DNA was removed from RNA samples using the Turbo DNA-free Kit (Ambion) following manufacturer's recommendations. RNA integrity was check and concentration quantified. cDNAs were synthesized from 1 µg of DNase-treated total RNA by using PrimeScriptTM RT Reagent kit (TaKaRa Clontech) following manufacturer's instruction. cDNA of genes *tnpA*, *orf121* and *dxs* was quantified using the

PerfeCTa® SYBR® Green FastMix® Kit (Quanta BioSciencesTM) with appropriate oligonucleotides (Supplementary Table 3). Relative expression of *orf121* (Primers ORF121 F and ORF121 R) and *tnpA* (Primers TnpA F and TnpA R) genes was estimated by normalizing copy number of transcripts to that of the housekeeping gene *dxs* (Primer dxs-LC3 and dxs-LC4). The *tnpA*/*dxs* or *orf121*/*dxs* ratio was obtained from 3 independent experiments, which were used to calculate mean values and standard deviations.

β-galactosidases assays.

β-galactosidase assays were performed with the *E. coli* MG1656 strain carrying various plasmids (Table 1 ; [47]). Overnight cultures were diluted 1:100, and further grown to mid-log phase (OD_{600} = 0.4-0.6). The β-galactosidase assay was performed as previously described[48]. β-galactosidase activity was obtained from 3 independent experiments, which were used to calculate mean values and standard deviations.

Estimation of His-tagged Orf121 or $TnpA_{IS91}$ protein expression by western-immunoblotting.

Overnight cultures were diluted to OD_{600} = 0.2 and grown to mid-log phase (OD_{600} = 0.5). Gene expression was induced by adding 0.5 mM IPTG (*tnpA* alone or with *orf121*) and/or 50 nM aTc (*orf121*) in the media. Cultures were incubated for further 3 h at 37°C, and 1 mL of each culture was centrifuged and stored at -20°C. Cell extracts were prepared using the B-PER bacterial protein extraction kit according to the manufacturer's recommendations at a rate of 100 µL of reagent per OD unit (Thermo Scientific, Ref. 90078) and inclusion bodies proteins using the solubilization reagent used according to the manufacturer's instructions (Thermo Scientific, Ref. 78115). Total protein concentration was determined using the Bradford assay (Bio-Rad). Equal amounts of total protein (as determined by Bradford) were loaded per lane for SDS-PAGE analysis. Cell extract

samples were subjected to 15% SDS-PAGE electrophoresis, followed by a transfer to a PVDF membrane according to the manufacturer's instructions (Trans-Blot Turbo Transfer System, BioRad). The membrane was incubated in a 1:3000 dilution of mouse anti-His-tag antibody (Invitrogen, Ref. MA121315) overnight at 4°C, then washed 3 times for 10 minutes with PBS. Next, the membrane was incubated in a 1:3000 dilution of anti-Mouse IgG, HRP antibody (Invitrogen, Ref. A28177), then washed 6 times for 5 minutes with water. The membrane was revealed with SuperSignal West Pico PLUS (Thermo Scientific, Ref. 34577). The amount of protein was determined in arbitrary units (AU) by calculating the value of the area under the curve using ImageLab software.

Analysis of His-tagged Orf121 or TnpA$_{IS91}$ proteins by mass spectrometry.

Peptides from SDS-PAGE were analyzed by micro-LC-MS/MS using a nanoLC 425 system in micro-flow mode (Eksigent, Dublin, CA, USA) coupled with time-of-flight (TOF) (TripleTOF 5600+ Sciex, Framingham, MA, USA) operating in high-sensitivity mode. Reverse-phase LC was performed via a trap and elute configuration using a trap column (C18 Pepmap100 cartridge, 5 µm pore size; Thermo Fisher Scientific) and an analytical column (ChromXP C18 column, 12 nm, 3 µm pore size, Sciex) with the following mobile phases: loading solvent (water/ACN/TFA 98/2/0.05% (v/v)), solvent A (0.1% (v/v) TFA in water) and solvent B (water/ACN/TFA 5/95/0.1% (v/v)). All samples were loaded, trapped, and desalted using a flow rate of 10 µL/min with loading solvent for 5 minutes. The chromatographic separation was performed at a flow rate of 2 µL/min as follows: initial, 5% solvent B, increased to 25% for 90 minutes, then increased to 95% B for 10 minutes, maintained at 95% for 5 minutes and, finally, decreased to 5% B for re-equilibration. The peptides were then reprocessed via ProteinPilot with the Mascot module using *Escherichia coli* Uniprot 2022_03. The following parameters were entered: trypsin 1 maximum missed cleavage, cysteine carbamidomethylation (fixed), methionine oxidation (variable),

Precursor Tolerance 0.01Da; MS/MS fragment tolerance: 0.05 Da, a $p<0.05$ - peptide score > 25 - bold red required and a protein score > 100.

Statistical analyses.

All analyses were performed using R Statistical Software (v4.4.3) and RStudio (v 2024.12.1+563). A one-way ANOVA with a Tukey Kramer post-test was used on the $\log_{10}$-transformed values to compare the means in conjugation and in transposition. Target site usage was analyzed using a χ^2 test. Finally, the proportion of OET was compared across strains using a global χ^2 test. All statistical analysis results have been grouped together in Supplementary Data 4. Numerical data presented in graphs are available in Supplementary Data 3. Graphics were rendered via the ggplot2 R package (v3.5.1). All figures were prepared using Inkscape 1.4 (https://inkscape.org/).

Data Availability statement

All data generated or analyzed during this study are available within the paper and its Supplementary Information files.

References

1. von Wintersdorff, C. J. H. *et al.* Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Frontiers in Microbiology* **7**, (2016).
2. Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin Microbiol Rev* **31**, e00088-17 (2018).
3. Diaz-Aroca, E., de la Cruz, F., Zabala, J. C. & Ortiz, J. M. Characterization of the new insertion sequence IS91 from an alpha-hemolysin plasmid of Escherichia coli. *Mol Gen Genet* **193**, 493–499 (1984).
4. Lallement, C., Pasternak, C., Ploy, M.-C. & Jové, T. The Role of ISCR1-Borne POUT Promoters in the Expression of Antibiotic Resistance Genes. *Front Microbiol* **9**, 2579 (2018).
5. Romantschuk, M., Richter, G. Y., Mukhopadhyay, P. & Mills, D. IS801, an insertion sequence element isolated from Pseudomonas syringae pathovar phaseolicola. *Mol Microbiol* **5**, 617–622 (1991).
6. Tavakoli, N. *et al.* IS1294, a DNA element that transposes by RC transposition. *Plasmid* **44**, 66–84 (2000).
7. Haytham, Y. Etude de la séquence d'insertion IS1294b et de son implication dans la dissémination des gènes de résistance aux antibiotiques chez les entérobactéries. (Bordeaux, 2015).
8. Chandler, M. *et al.* Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol* **11**, 525–538 (2013).
9. Mendiola, M. V., Bernales, I. & de la Cruz, F. Differential roles of the transposon termini in IS91 transposition. *Proc Natl Acad Sci U S A* **91**, 1922–1926 (1994).
10. Garcillán-Barcia, M. P., Bernales, I., Mendiola, M. V. & de la Cruz, F. IS 91 Rolling-Circle Transposition. in *Mobile DNA II* 889–904 (John Wiley & Sons, Ltd, 2002). doi:10.1128/9781555817954.ch37.
11. Toleman, M. A., Bennett, P. M. & Walsh, T. R. ISCR elements: novel gene-capturing systems of the 21st century? *Microbiol Mol Biol Rev* **70**, 296–316 (2006).
12. Toleman, M. A. & Walsh, T. R. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol Rev* **35**, 912–935 (2011).
13. Mendiola, M. V. & de la Cruz, F. Specificity of insertion of IS91, an insertion sequence present in alpha-haemolysin plasmids of Escherichia coli. *Mol Microbiol* **3**, 979–984 (1989).
14. del Pilar Garcillán-Barcia, M., Bernales, I., Mendiola, M. V. & de la Cruz, F. Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Mol Microbiol* **39**, 494–501 (2001).
15. Bernales, I., Mendiola, M. V. & de la Cruz, F. Intramolecular transposition of insertion sequence IS91 results in second-site simple insertions. *Mol Microbiol* **33**, 223–234 (1999).
16. Mendiola, M. V., Jubete, Y. & de la Cruz, F. DNA sequence of IS91 and identification of the transposase gene. *J Bacteriol* **174**, 1345–1351 (1992).
17. Solovyev, V. & Salamov, A. Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies* 61–78 (2011).
18. Fortunato, G., Vaz-Moreira, I., Gajic, I. & Manaia, C. M. Insight into phylogenomic bias of blaVIM-2 or blaNDM-1 dissemination amongst carbapenem-resistant Pseudomonas aeruginosa. *Int J Antimicrob Agents* **61**, 106788 (2023).

19. Phuadraksa, T. *et al.* Emergence of plasmid-mediated colistin resistance mcr-3.5 gene in Citrobacter amalonaticus and Citrobacter sedlakii isolated from healthy individual in Thailand. *Front Cell Infect Microbiol* **12**, 1067572 (2022).
20. Li, L. *et al.* First report of two foodborne Salmonella enterica subsp. enterica serovar Bovismorbificans isolates carrying a novel mega-plasmid harboring blaDHA-1 and qnrB4 genes. *Int J Food Microbiol* **360**, 109439 (2021).
21. Berglund, F., Ebmeyer, S., Kristiansson, E. & Larsson, D. G. J. Evidence for wastewaters as environments where mobile antibiotic resistance genes emerge. *Commun Biol* **6**, 321 (2023).
22. Fan, Q., Zhang, J., Shi, H., Chang, S. & Hou, F. Metagenomic Profiles of Yak and Cattle Manure Resistomes in Different Feeding Patterns before and after Composting. *Appl Environ Microbiol* **89**, e0064523 (2023).
23. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429–3431 (2003).
24. Davis, M. A., Simons, R. W. & Kleckner, N. Tn10 protects itself at two levels from fortuitous activation by external promoters. *Cell* **43**, 379–387 (1985).
25. Sekine, Y., Izumi, K., Mizuno, T. & Ohtsubo, E. Inhibition of transpositional recombination by OrfA and OrfB proteins encoded by insertion sequence IS3. *Genes Cells* **2**, 547–557 (1997).
26. Huber, M. *et al.* Translational coupling via termination-reinitiation in archaea and bacteria. *Nat Commun* **10**, 4006 (2019).
27. Powell, M. L. Translational termination-reinitiation in RNA viruses. *Biochem Soc Trans* **38**, 1558–1564 (2010).
28. Powell, M. L., Brown, T. D. K. & Brierley, I. Translational termination-re-initiation in viral systems. *Biochem Soc Trans* **36**, 717–722 (2008).
29. Huber, M. *et al.* Unidirectional gene pairs in archaea and bacteria require overlaps or very short intergenic distances for translational coupling via termination-reinitiation and often encode subunits of heteromeric complexes. *Front Microbiol* **14**, 1291523 (2023).
30. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res* **22**, 4953–4957 (1994).
31. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).
32. Arutyunov, D. & Frost, L. S. F conjugation: back to the beginning. *Plasmid* **70**, 18–32 (2013).
33. Fernández-López, R. *et al.* Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol Rev* **30**, 942–966 (2006).
34. Zerbib, D., Polard, P., Escoubas, J. M., Galas, D. & Chandler, M. The regulatory role of the IS1-encoded InsA protein in transposition. *Mol Microbiol* **4**, 471–477 (1990).
35. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. The Leucine Zipper: A Hypothetical Structure Common to a New Class of DNA Binding Proteins. *Science* **240**, 1759–1764 (1988).
36. Matthews, J. M. *et al.* A class of zinc fingers involved in protein–protein interactions. *European Journal of Biochemistry* **267**, 1030–1038 (2000).
37. Bardaji, L., Añorga, M., Echeverría, M., Ramos, C. & Murillo, J. The toxic guardians - multiple toxin-antitoxin systems provide stability, avoid deletions and maintain virulence genes of Pseudomonas syringae virulence plasmids. *Mob DNA* **10**, 7 (2019).
38. Galata, V., Fehlmann, T., Backes, C. & Keller, A. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res* **47**, D195–D202 (2019).

39. Douarre, P.-E., Mallet, L., Radomski, N., Felten, A. & Mistou, M.-Y. Analysis of COMPASS, a New Comprehensive Plasmid Database Revealed Prevalence of Multireplicon and Extensive Diversity of IncF Plasmids. *Front Microbiol* **11**, 483 (2020).
40. Chen, Q., Zobel, J. & Verspoor, K. Duplicates, redundancies and inconsistencies in the primary nucleotide databases: a descriptive study. *Database (Oxford)* **2017**, baw163 (2017).
41. Miller, S. R. *et al.* Bacterial Adaptation by a Transposition Burst of an Invading IS Element. *Genome Biol Evol* **13**, evab245 (2021).
42. Wu, Y., Aandahl, R. Z. & Tanaka, M. M. Dynamics of bacterial insertion sequences: can transposition bursts help the elements persist? *BMC Evol Biol* **15**, 288 (2015).
43. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166**, 557–580 (1983).
44. Schmartz, G. P. *et al.* PLSDB: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res* **50**, D273–D278 (2022).
45. Polard, P., Prère, M. F., Fayet, O. & Chandler, M. Transposase-induced excision and circularization of the bacterial insertion sequence IS911. *EMBO J* **11**, 5079–5090 (1992).
46. Cam, K., Béjar, S., Gil, D. & Bouché, J. P. Identification and sequence of gene dicB: translation of the division inhibitor from an in-phase internal start. *Nucleic Acids Res* **16**, 6327–6338 (1988).
47. Espéli, O., Moulin, L. & Boccard, F. Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *J Mol Biol* **314**, 375–386 (2001).
48. Miller, J. H. *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia Coli and Related Bacteria*. (CSHL Press, 1992).
49. Demarre, G. *et al.* A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate Escherichia coli host strains. *Res Microbiol* **156**, 245–255 (2005).
50. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
51. Pasternak, C. *et al.* ISDra2 transposition in Deinococcus radiodurans is downregulated by TnpB. *Mol Microbiol* **88**, 443–455 (2013).
52. Jové, T., Da Re, S., Denis, F., Mazel, D. & Ploy, M.-C. Inverse correlation between promoter strength and excision activity in class 1 integrons. *PLoS Genet* **6**, e1000793 (2010).
53. Chandler, M. & Galas, D. J. IS1-mediated tandem duplication of plasmid pBR322. Dependence on recA and on DNA polymerase I. *J Mol Biol* **165**, 183–190 (1983).

Acknowledgements

A. Fauconnier acknowledges the French ministère de l'Enseignement Supérieur, de la Recherche (MESR) for his doctoral training grant. This work was supported by fundings from the French research institute Inserm and MESR.

The authors thank Emilie Pinault from BISCEm unit (Univ. Limoges, UAR 2015 CNRS, US 42 Inserm, CHU Limoges) for technical support regarding mass spectrometry.

The authors are grateful to Céline Loot (Institut Pasteur, Université Paris Cité, CNRS UMR3525, Unité Plasticité du Génome Bactérien, 75724 Paris, France) and Vincent Burrus (Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada) for critical review of the manuscript.

The authors dedicate this publication to our colleague, Cécile Pasternak (last author), who left us during the final phase of this work.

Author contributions

Aurélien Fauconnier [AF]: Conceptualization; Methodology; Investigation; Data curation; Formal analysis; Validation; Visualization; Writing – original draft; Writing – review & editing.

Sandra Da Re [SDR]: Conceptualization; Data curation; Validation; Visualization; Supervision; Writing – original draft; Writing – review & editing.

Margaux Gaschet [MG]: Investigation; Data curation; Formal analysis; Validation.

Thomas Jové [TJ]: Conceptualization; Methodology; Investigation; Data curation; Formal analysis; Validation; Supervision.

Marie-Cécile Ploy [MCP]: Conceptualization; Resources; Data curation; Formal analysis; Validation; Supervision; Project administration; Funding acquisition; Writing – original draft; Writing – review & editing.

Cécile Pasternak [CP]: Conceptualization; Methodology; Investigation; Resources; Data curation; Formal analysis; Validation; Visualization; Supervision; Project administration; Funding acquisition; Writing – original draft.

All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Figures legends

Figure 1. Genetic organization of members of the IS*91* family (IS*801*, IS*1294*, IS*1294b* and IS*91*).

*ter*IS and *ori*IS ends carrying the palindromic sequences are depicted as red and blue boxes, cleavage positions at *ter*IS and *ori*IS ends are shown by red and blue vertical arrows, respectively. The cleavage site at the *ter*IS end and the tetranucleotide target site adjacent to *ori*IS are indicated. *orf121* and *tnpA* encoding sequences are shown as orange and green arrows, respectively.

Figure 2. Estimation of the P_{orf121} and P_{tnpA} promoters' activity and the impact of the one base overlap

(A) Sequence from *ter*IS to the start of the *tnpA* gene of IS*91* accession number X17114.5. The *ter*IS region, *orf121* and *tnpA* genes are indicated by a red rectangle, an orange arrow and a green arrow above the nucleotidic sequence, respectively. Red arrows underline the terminal palindrome *ter*IS in bold red letters. The cleavage tetranucleotide is in black bold letter with a vertical red arrow indicating the cleavage site. The potential -35 and -10 boxes, SD region and start codon are shown in bold letters: orange for P_{orf121} and green for P_{tnpA}. Ribosome Binding Sites (RBS) are underlined.

(B) β-galactosidase activity of the transcriptional fusions between *lacZ* and the predicted promoters P_{orf121} and P_{tnpA} with *lacZ* RBS.

(C) β-galactosidase activity of the transcriptional fusions between *lacZ* and P_{tnpA} with its native RBS in the presence or absence of the overlap with *orf121*, and in the presence or absence of P_{orf121}.

(D) β-galactosidase activity of the transcriptional fusions between *lacZ* and native or mutated *tnpA* promoter (P_{tnpA} and P_{tnpA*}, respectively) in the presence of the overlap with *orf121* and the *orf121* native or mutated promoter (P_{orf121} and $P_{orf121*}$, respectively).

β-galactosidase activity is expressed as Miller units. Empty denotes pSU38ΔtotlacZ plasmid with the promoterless *lacZ* gene used as a negative control. Scheme of the P_{orf121} and/or P_{tnpA} *lacZ*-transcriptional fusion constructs are shown for panels B-D, with the different elements described on the figure. The results are the average of at least 3 independent experiments. Statistical significance was determined by one-way ANOVA followed by Tukey-Kramer post-test. Statistical significance is shown as a compact letter display for pairwise comparisons where means grouped under identical letters are not statistically different. Supplementary Data 3 contains the source data for all figures, and Supplementary Data 4 includes the detailed statistical analyses.

Figure 3. Expression of Orf121 decreases IS*91* transposition frequency.

(A) Schematic of the mating-out assays, the donor *E. coli* JS219 (black) carries three plasmids: one IS donor, one *tnpA*/*orf121* expression plasmid and the conjugative pOX38Km (Km^R) as the target for transposition. Upon TnpA expression, transposition events of the mini-IS occurs. Transposition events in the pOX38Km are isolated by conjugation in the *E. coli* MC240 recipient cell (dark red; Nal^R) to estimate transposition frequency.

(B) Transposition frequency of IS*91* derivatives (mini-IS *ter*IS-cm^R and *ter*IS::*orf121*-cm^R) estimated by mating-out assays when *tnpA* is expressed alone or with *orf121* from the same promoter *in trans*, or with *orf121* expressed *in cis* from *ter*IS::*orf121*-cm^R.

(C) Transposition frequency of IS*91* derivatives (mini-IS *ter*IS-*cm*^{R}) estimated by mating-out assays when *orf121* and *tnpA* are expressed *in trans* from independent promotors ($P_{LtetO-1}$ and P_{lac} respectively).

Scheme of constructs expressing *tnpA* and/or *orf121* from inducible promoters are shown on the figure on top of plasmid names and scheme of mini-IS on the left-hand side. The IS*91* TnpA transposase expression plasmids (expression *in trans* relative to the IS; selected by the *sp*^{R} resistance gene) feature an origin of replication p15A and the IPTG-inducible P_{lac} promoter (blue), which controls the *tnpA* gene (green) in a *tnpA* alone configuration (P_{lac}::*tnpA*) or in the native *orf121-tnpA* genes configuration (orange and green respectively; P_{lac}::*orf121-tnpA*). When the two genes are expressed under the control of different promoters, *tnpA* remains under the control of Plac, and *orf121* is under the control of the aTc-inducible $P_{LtetO-1}$ promoter (purple) (*orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA*) or without *tnpA* (*orf121*::$P_{LtetO-1}$-P_{lac}). IS*91* TnpA substrate plasmids (IS donor) are derivatives of pUC18 and confer ampicillin resistance (*amp*^{R} used for detecting the OET mechanism). Plasmid *ter*IS-*cm*^{R} carries a mini-IS composed of the *ter*IS (red) and *ori*IS (blue) flanking the *cm*^{R} mobility reporter gene (chloramphenicol resistance; yellow). Plasmid *ter*IS::*orf121-cm*^{R} is composed of the *ter*IS::*orf121* region (native IS*91* configuration, red and orange respectively) and *ori*IS (blue) flanking the *cm*^{R} gene. The - and + indicate the absence or presence of inducers, IPTG and aTc. Data are represented as box-plot. Experiments were performed at least 5 times. #, undetectable transposition (< 3.35 x 10^{-8}). Statistical significance was determined by one-way ANOVA followed by Tukey-Kramer post-test, and is shown as a compact letter display for pairwise comparisons where means, grouped under identical letters, are not statistically different. Supplementary Data 3 contains the source data for all figures, and Supplementary Data 4 includes the detailed statistical analyses.

Figure 4. Quantification of *tnpA* and *orf121* transcripts and protein levels in mating out experiments.

(A-B) Estimation by RT-qPCR of *tnpA* (A) and *orf121* (B) transcript levels in mating-out experiments (Fig. 3B-C). Transcript levels were normalized to the housekeeping gene *dxs*. Experiments were performed 6 times. Statistical significance was determined by one-way ANOVA followed by Tukey-Kramer post-test and constructs as previously detailed (Fig. 3). Statistical significance is shown as a compact letter display for pairwise comparisons where means, grouped under identical letters, are not statistically different. Supplementary Data 3 contains the source data for all figures, and Supplementary Data 4 includes the detailed statistical analyses.

(C-D) Detection of TnpA^{6his} (C) and ^{6his}Orf121 (D) protein by western immunoblotting. His-tagged TnpA (C-term; *tnpA*^{6his}; expected size 47 kDa) and His-tagged Orf121 (N-term; ^{6his}*orf121*; expected size 17 kDa) were extracted from the mating-out experiment cultures. An equal amount of total proteins was loaded for each culture. Densitometry values of the bands are reported as arbitrary units (ImageLab). AU: Arbitrary Unit of quantity of protein. The complete image of the PVDF membrane is available in Supplementary Figure 5.

Figure 5. Bottom strand ssDNA circular intermediates are successfully inserted into a new target sequence.

(A) Sequence of the *ter*IS-*ori*IS and *ori*IS-*ter*IS junctions present in the substrates used for insertion analyses. Sequence of *ter*IS and *ori*IS ends are shown in red and blue, respectively. Cleavage positions are shown by blue vertical arrows and the cleavage site at the *ter*IS is underlined (bold red letters). Scheme of the plasmids carrying the junctions are shown.

(B) Suicide transformation assays assessing *ter*IS-*ori*IS or *ori*IS-*ter*IS dsDNA-circular IS*91* intermediates insertion into *E. coli* chromosome in the absence of *tnpA* (P_{Lac} (empty)) and when

tnpA was expressed alone. The pUC18 plasmid was used as a positive control of transformation efficiency, and pSW23T as negative control for intermediates circular.

(C-D) Suicide conjugation assays assessing *ter*IS-*ori*IS or *ori*IS-*ter*IS ssDNA-circular IS*91* intermediates insertion into *E. coli* chromosome when *tnpA* was expressed (C) alone or with *orf121* from the same promoter (native configuration) or (D) with *orf121* from independent promoters. The - and + indicate the absence or presence of inducers, IPTG and aTc. Constructs are detailed in Fig. 3. Experiments were performed 4 times. #, non-detectable transformant or insertion (< 2.78 x 10^{-9}). Statistical significance was determined by one-way ANOVA followed by Tukey-Kramer post-test. Statistical significance is shown as a compact letter display for pairwise comparisons where means, grouped under identical letters, are not statistically different. Supplementary Data 3 contains the source data for all figures, and Supplementary Data 4 includes the detailed statistical analyses.

Tables

Table 1. Bacterial strains and plasmids

Strain or plasmid	Genotype or other relevant characteristics	Source or reference
	***E. coli* strains**	
DH5α	*supE44 ΔlacU(ϕ80lacZΔM15) hsdR17recA1endA1gyrA96 thi-1 relA1*	Laboratory collection
JS219	MC1061, *recA1*, *lacI*^{q}	[46]
MC240	XA103, F^{-}, ara, del(lac pro), gyrA (nal^{R}), metB, argE^{am}, rpoB, thi, supF, rif^{R}	Gift of B. Ton-Hoang
β2163	(F-) RP4-2-Tc::Mu dapA	[49]
MG1656	MG1655lac-	[47]
	Plasmids	
pSU2600	Source of IS*91* for the *orf121* gene, *tnpA* gene and the promoters P$_{orf121}$ and P$_{tnpA}$	[9]
pUC18	Amp^{R}; MCS from M13mp19 into *lacZ'*	[50]
pCP5	pUC18; P$_{cat}$::Cm^{R}	This work
pCP6	pUC18; *ter91* P$_{cat}$::Cm^{R}	This work
pAF1	pUC18; P$_{cat}$::Cm^{R} *ori91*	This work
pGY13556	pAPT110ΔKan^{r}; P$_{lac}$::*tnpA*$_{ISDra2}$-*his6*	[51]
pSW23T	Cm^{R}	[49]
pSU38ΔtotlacZ	vector carrying promoterless *lacZ*; p15A ori; Km^{R}	[52]
pOX38Km	Derivative of plasmid F, Km^{R}	[53]
	IS*91* substrate plasmids	
*ter*IS-*cm*^{R}	pUC18; *ter91* P$_{cat}$::Cm^{R} *ori91*	This work
*ter*IS::*orf121-cm*^{R}	pUC18; *ter91*::*orf121* P$_{cat}$::Cm^{R} *ori91*	This work
	TnpA expression plasmids	
P$_{Lac}$ (Empty) (pCP2)	pGY13556ΔtnpA$_{ISDra2}$Ωlinker79; Sp^{R}	This work
P$_{lac}$::*tnpA* (pCP4)	pCP2; P$_{lac}$::*tnpA*	This work
P$_{Lac}$::*tnpA*^{6his}	pCP4; P$_{lac}$::*tnpA*^{His6}	This work
P$_{lac}$::*orf121-tnpA* (pCP40)	pCP2; P$_{lac}$::*orf121tnpA*	This work
P$_{Lac}$::*orf121tnpA*^{6his}	pCP40; P$_{lac}$:: *orf121tnpA*^{His6}	This work
P$_{Lac}$:: ^{6his}*orf121tnpA*	pCP40; P$_{lac}$::^{His6}*orf121tnpA*	This work
orf121::P$_{LtetO-1}$-P$_{lac}$	pCP2; *orf121*::P$_{LtetO-1}$-P$_{lac}$	This work
orf121::P$_{LtetO-1}$-P$_{lac}$::*tnpA*	pCP2; *orf121*::P$_{LtetO-1}$-P$_{lac}$::*tnpA*	This work
orf121^{6his}::P$_{LtetO-1}$-P$_{Lac}$::*tnpA*^{6his}	pCP2; *orf121*^{His6}::P$_{LtetO-1}$-P$_{lac}$::*tnpA*^{His6}	This work
	Circular intermediates plasmids	
*ori*IS-*ter*IS	pSW23T; *ori*IS::*ter*IS	This work
*ter*IS-*ori*IS	pSW23T; *ter*IS::*ori*IS	This work
	Transcriptional fusion plasmids with *lacZ*	
P$_{tnpA}$::RBS$_{lacZ}$-*lacZ*	pSU38Δtot*lacZ*; P$_{tnpA}$::RBS$_{lacZ}$::*lacZ*	This work
P$_{orf121}$::RBS$_{lacZ}$-*lacZ*	pSU38Δtot*lacZ*; P$_{orf121}$::RBS$_{lacZ}$::*lacZ*	This work
P$_{tnpA}$::*lacZ*	pSU38Δtot*lacZ*; *orf121*::P$_{tnpA}$::RBS$_{tnpA}$::*lacZ* with overlap (-1)	This work
P$_{orf121}$-*orf121*$_{(PtnpA)}$::*lacZ*	pSU38Δtot*lacZ*; P$_{orf121}$::RBS$_{orf121}$::*orf121*::P$_{tnpA}$::RBS$_{tnpA}$::*lacZ* with overlap (-1)	This work
P$_{tnpA}$::*lacZ* Δ-1	pSU38Δtot*lacZ*; *orf121*::P$_{tnpA}$::RBS$_{tnpA}$::*lacZ* **without** overlap (-1)	This work
P$_{orf121}$-*orf121*$_{(PtnpA)}$::*lacZ* Δ-1	pSU38Δtot*lacZ*; P$_{orf121}$::RBS$_{orf121}$::*orf121*::P$_{tnpA}$::RBS$_{tnpA}$::*lacZ* **without** overlap (-1)	This work
P$_{orf121}$-*orf121*$_{(PtnpA*)}$::*lacZ*	pSU38Δtot*lacZ*; P$_{orf121}$::RBS$_{orf121}$::*orf121*::P$_{tnpA}$*::RBS$_{tnpA}$::*lacZ* with overlap (-1)	This work

P_{orf121}*-*orf121*$_{(PtnpA)}$::*lacZ*	pSU38Δtot*lacZ*; P_{orf121}*::RBS_{orf121}::*orf121*::P_{tnpA}::RBS_{tnpA}::*lacZ* with overlap (-1)	This work

Δ indicates a deletion; Ω indicates an insertion; :: indicates a novel joint; * indicates a mutation and ^{6his} indicates localization of His-tag.

Table 2. Tetranucleotide insertion target in the different mating-out assays

Substrate for transposition		*ter*IS-cm^R												*ter*IS::*orf121-*cm^R [d]	
Expression plasmids		P_{lac}::*tnpA* [a]		P_{lac}::*orf121-tnpA* [b]		*orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA* [c]								P_{lac}::*tnpA*	
IPTG [e]						-		-		+		+			
aTc [e]						-		+		-		+			
n [f]		277	%	292	%	50	%	49	%	50	%	50	%	240	%
Sequence of insertion site tetranucleotide	5'-CTTG	142	51.26	136	46.58	31	62.0	34	69.39	29	58.0	36	72.0	109	45.42
	5'-GTTC	113	40.79	134	45.89	19	38.0	15	30.61	21	42.0	12	24.0	113	47.08
	5'-GTCC	14	5.05	13	4.45									12	5.0
	5'-GTAC	3	1.08	1	0.34										
	5'-CTCG	3	1.08	5	1.71							2	4.0	2	0.83
	5'-GTCG	1	0.36												
	5'-CTTC	1	0.36											1	0.42
	5'CTGT			1	0.34										
	5'-CTGG			1	0.34										
	5'-CTGA			1	0.34										
	5'-GTGC													2	0.83
	5'-TTTG													1	0.42

[a] expression of TnpA from P_{lac} only, *in trans* relative to IS*91*.

[b] expression of Orf121 and TnpA in their native configuration from P_{lac}, *in trans* relative to IS*91*.

[c] expression of TnpA and Orf121 from independent promoters *in trans* relative to IS*91*.

[d] expression of Orf121 *in cis* relative to IS*91*.

[e] - and + indicate the absence or presence of inducers (IPTG and aTc).

[f] number of clones analyzed by AP-PCR.

Table 3. Estimation of the % of OET depending on the presence of Orf121

		P_{lac}::*tnpA* [d]	P_{lac}::*orf121-tnpA* [e]	*orf121*::$P_{LtetO-1}$-P_{lac}::*tnpA* [f]			
IPTG [a]		+	+	-	-	+	+
aTc [a]		/	/	-	+	-	+
*ter*IS-*cm*^{R}	n [b]	1235	674	1020	1020	1020	1020
	%[c]	53	18 [*]	29	11 [£]	48 [£]	32 [$]
*ter*IS::*orf121-cm*^{R} [g]	n [b]	1104					
	%[c]	29 [*]					

[a] - and + indicate the absence or presence of inducers.

[b] number of clones analyzed.

[c] *%* of OET in the different mating-out assays (*tnpA* expressed alone or with *orf121* expressed *in trans* (same and independent promoters) or *in cis*).

[d] expression of TnpA only *in trans* relative to the IS*91* derivative.

[e] expression of TnpA and Orf121 in their native configuration *in trans* relative to the IS*91* derivative.

[f] expression of TnpA and Orf121 from independent promoters *in trans* relative to the IS*91* derivative.

[g] expression of orf121 *in cis* relative to the IS*91* derivative.

[*] indicates that the difference between the percentages is significant ($p<0.05$) compared to the condition without *orf121* expression.

[£] indicates that the difference between the percentages is significant ($p<0.05$) compared to the condition without inducer.

[$] indicate that the percentage difference is significant ($p<0.05$) compared to the condition with IPTG-only.

Supplementary information

Supplementary Figure 1. His-tags does not impact transposition frequencies

Supplementary Figure 2. Insertion-site sequence logos around *ori*IS (positions −4 to −1)

Supplementary Figure 3. Principle of the suicide mating

Supplementary Figure 4. Global diagram of the role of Orf121 in IS*91* mobility.

Supplementary Figure 5. Quantification of TnpA and Orf121 protein levels in mating out experiments (Uncropped PVDF membranes)

Supplementary Table 1. Host of bacteria carrying IS*91* elements in GenBank®

Supplementary Table 2. Effect of Orf121 expression on target insertion sites in the *E. coli*

Supplementary Table 3. Oligonucleotides used in this study.

Supplementary Data 1. Sequences of the different IS*91* isoforms described in this study.

IS91-V1 corresponds to the first sequence of the IS*91* element identified (accession number X17114.5).

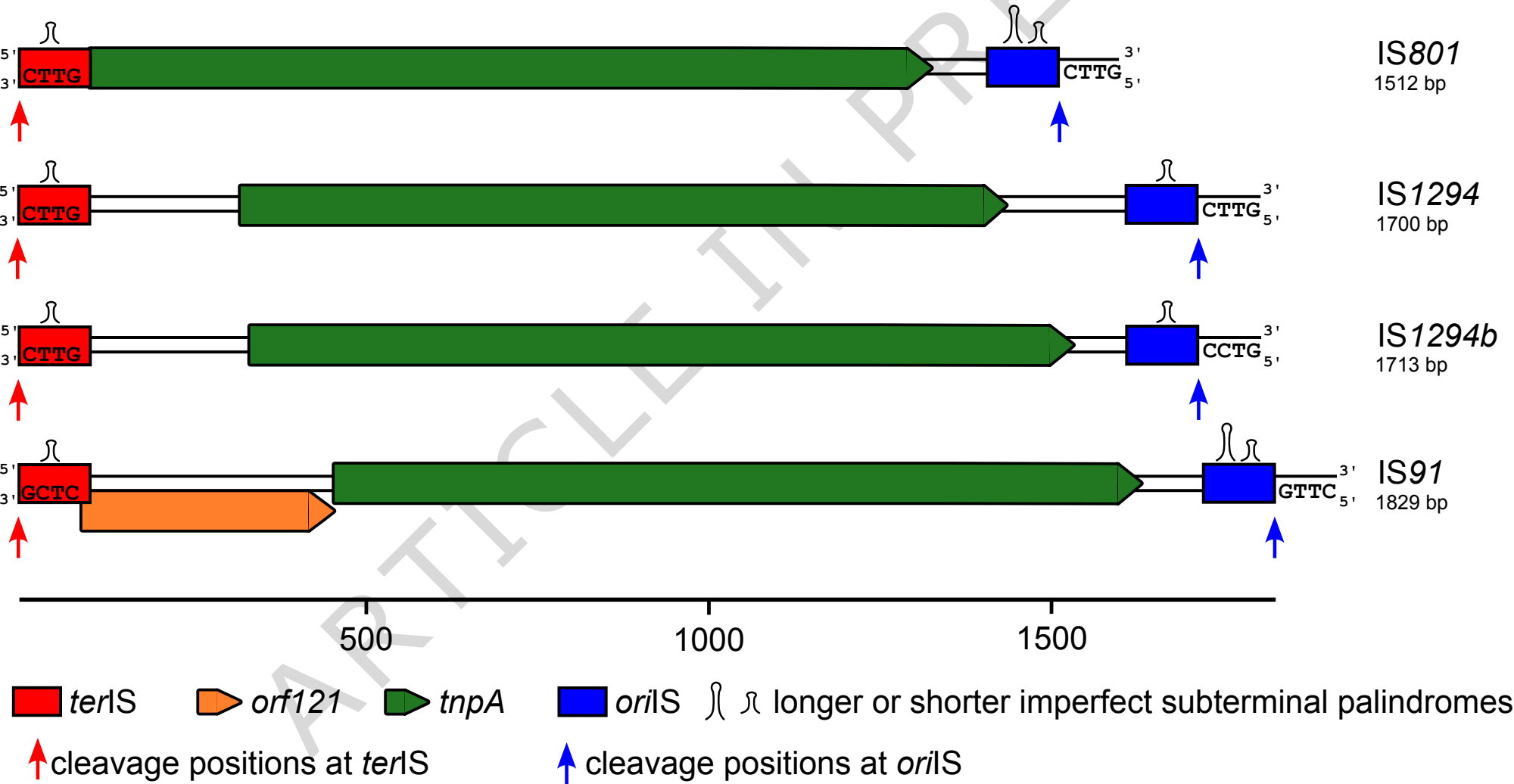
Supplementary Data 2. Description of the different IS*91* isoforms described in this study.

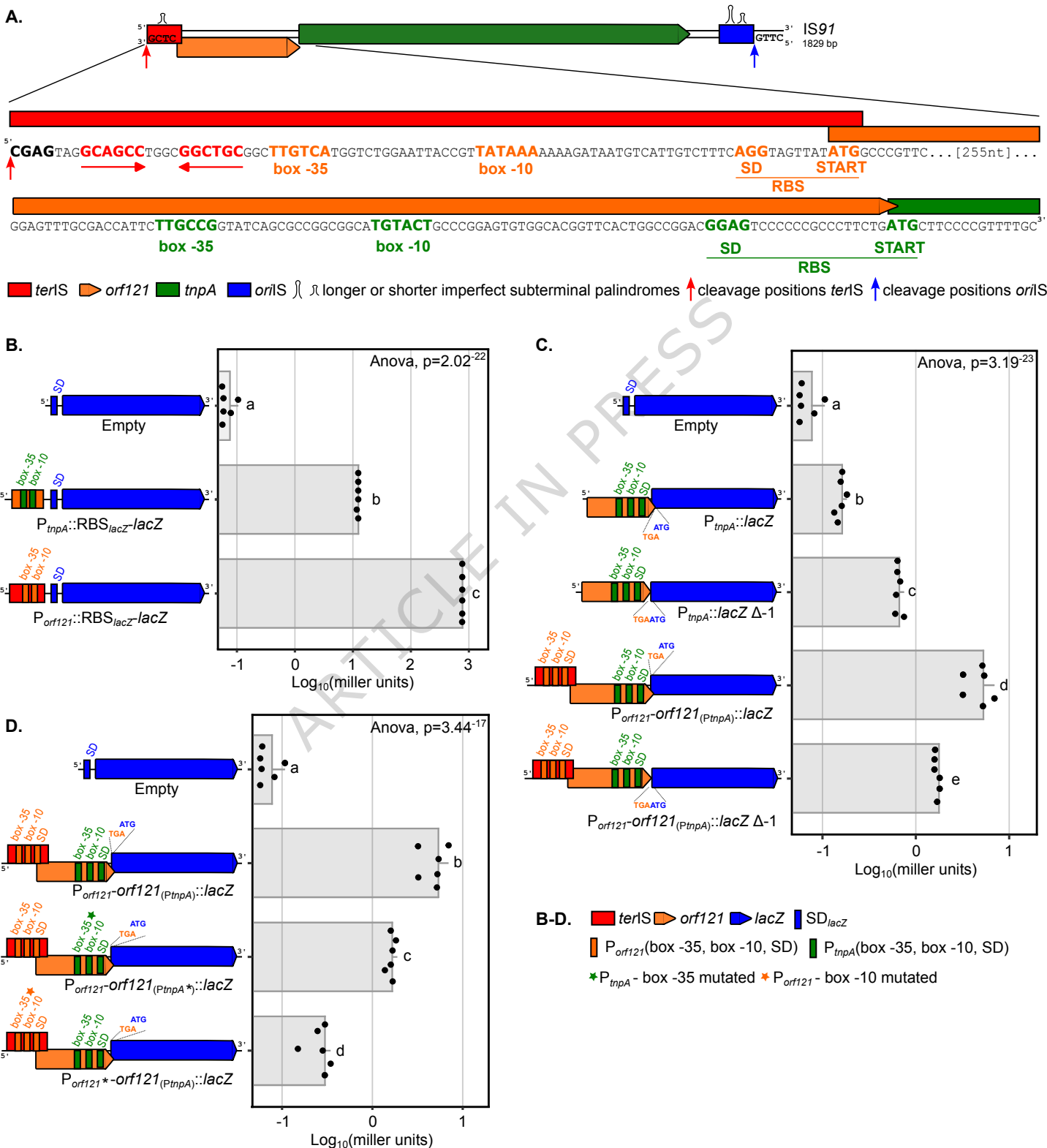
Supplementary Data 3. Complete data sets. (Excel workbook with multiple sheets).

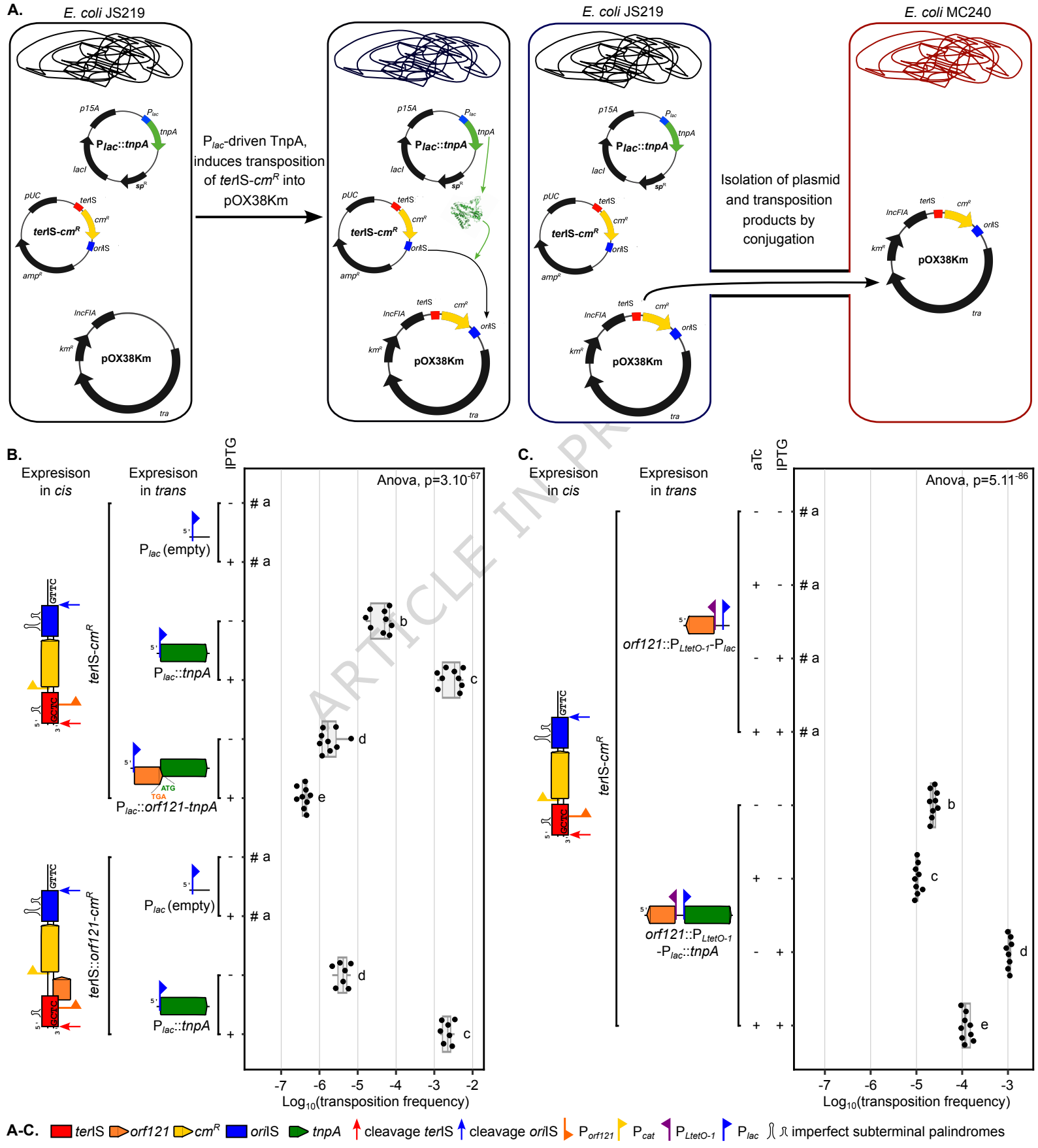
Supplementary Data 4. Statistical analyses. (Excel workbook with multiple sheets).

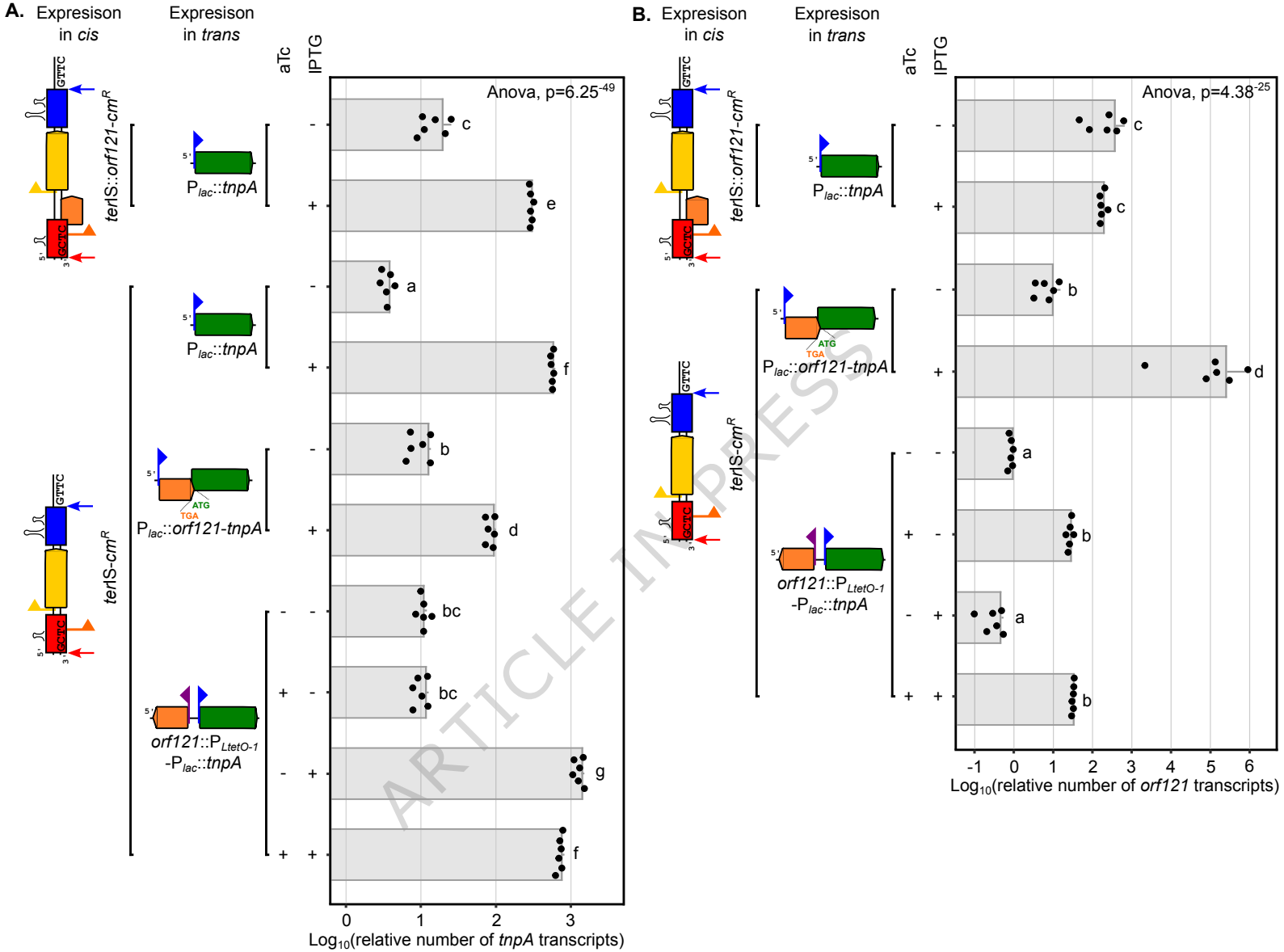
Editorial summary:
This study identifies Orf121 as a dual regulator that limits IS91 transposition while ensuring accurate excision of mobile DNA. The findings reveal a built-in mechanism that enables bacteria to balance genome plasticity with stability.

Peer review information:
Communications Biology thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Ranjana Pathania and Mengtan Xing. A peer review file is available.






A.
E. coli JS219
P lac -driven TnpA, induces transposition of terIS-cmR into pOX38Km
E. coli JS219
Isolation of plasmid and transposition products by conjugation
E. coli MC240
B.
Expresison in cis
Expresison in trans
IPTG
Anova, p=3.10-67
terIS-cmR
terIS::orf121-cmR
P lac (empty)
P lac::tnpA
P lac::orf121-tnpA
Log10(transposition frequency)
C.
Expresison in cis
Expresison in trans
aTc
IPTG
Anova, p=5.11-86
orf121::P LtetO-1-P lac
orf121::P LtetO-1 -P lac::tnpA
Log10(transposition frequency)
A-C.
terIS
orf121
cmR
oriS
tnpA
cleavage terIS
cleavage oriS
P orf121
P cat
P LtetO-1
P lac
imperfect subterminal palindromes



A-B. terIS orf121 cmR oriIS tnpA cleavage terIS cleavage oriIS Porf121 Pcat PLtetO-1 Plac imperfect subterminal palindromes