Activation of TnSmu1, an integrative and conjugative element, is lethal to *Streptococcus mutans*

Shawn King[1], Allison Quick[1], Kalee King[1], Alejandro R. Walker[2], and Robert C. Shields[1*]

[1] Department of Biological Sciences, Arkansas State University, Jonesboro, Arkansas, USA

[2] Department of Oral Biology, University of Florida, Gainesville, Florida, USA

ORCiDs:

Shields: http://orcid.org/0000-0002-8823-9504

Walker: https://orcid.org/0000-0002-5839-6584

Keywords: *Streptococcus mutans*, essential gene, integrative and conjugative element, TnSmu1, transcriptional regulator

*Corresponding Author: rshields@astate.edu

Abstract

Integrative and conjugative elements (ICEs) are chromosomally encoded mobile genetic elements that can transfer DNA between bacterial strains. They are important for generating microbial diversity and can encode for antibiotic, virulence, and metabolism determinants. Recently, as part of efforts to determine hypothetical essential gene functions, we have discovered an important regulatory module encoded on an ICE known as TnSmu1 on the *Streptococcus mutans* chromosome. The regulatory module consists of a transcriptional regulator we are naming essential regulatory factor (*erfR*) and a metalloprotease (*erfP*). The *efrR* gene cannot be deleted and repression by CRISPR interference (CRISPRi) causes a severe growth defect. We used a bypass of essentiality (BoE) screen, with transposon mutagenesis allied to genome sequencing, to discover genes that allow deletion of the regulatory module. This revealed that conjugation genes, located within TnSmu1, can partially restore the viability of an *erfR* mutant. Loss of *erfR* also leads to production of a circular intermediate form of TnSmu1, that is also inducible by the genotoxic agent mitomycin C. To gain further insights into potential regulation of TnSmu1 by ErfR and broader effects on *S. mutans* UA159 physiology we used a combination of CRISPRi and RNA-seq. The RNA-seq analysis identified 134 genes whose $\log_2$ fold change was ≥ 2 after repression of *erfR*. Strongly induced genes included all the TnSmu1 mobile element, genes involved in amino acid metabolism, transport systems, and a Type I-C CRISPR-Cas system. Lastly, bioinformatic analysis shows that the TnSmu1 mobile element and its associated genes are well distributed across *S. mutans* isolates. Taken together, our results show that activation of a newly characterized ICE is strongly deleterious to *S. mutans*, highlighting the complex interplay between mobile elements and their host. We are continuing to investigate TnSmu1 as a model to understand ICE distribution, and effects on host fitness and cell biology in oral streptococci.

INTRODUCTION

Streptococcus mutans is a Gram-positive bacterium that colonizes the human oral cavity (Loesche, 1986). Like many streptococci, *S. mutans* can cause disease when environmental conditions are favorable. Poor oral hygiene combined with frequent ingestion of simple sugars creates an environment in which *S. mutans* can cause dental caries (tooth decay) (Sheiham and James, 2015; Pitts *et al.*, 2017). If this disease is allowed to progress it causes the breakdown of teeth, with symptoms that include pain, difficulty eating and tooth loss (Pitts *et al.*, 2017). Due to its widespread presence in the oral cavity and resulting disease burden it is necessary to identify targets for improved therapeutics and understand processes essential for the pathogen. With this goal in mind, we have recently applied transposon sequencing (Tn-seq) and CRISPR interference (CRISPRi) and have identified >200 *Streptococcus mutans* essential genes that are required for the viability of the organism (Shields *et al.*, 2018; Shields *et al.*, 2020). For *S. mutans*, most essential genes can be broadly sorted into three categories: processing of genetic information, energy production, and maintenance of the cell envelope (Shields *et al.*, 2018).

It was during CRISPRi experiments that we discovered an essential gene, *erfR* (essential regulator factor), which is annotated as a transcriptional regulator, and resides in a two-gene operon with *erfP*. This prompted us to explore the genomic context of *erfRP* and to evaluate the effects of interfering with production of ErfR or EfrP. Transposon cassette insertions into *erfR* are not tolerated by *Sm*, and repression of this gene leads to significant growth defects (Figure 1). Probing ErfRP with SMART (Letunic and Bork, 2018), we found that the N-terminus of ErfR contains a cI-like repressor of phage λ DNA-binding domain, often found in *Streptococcus* phages. ErfP contains a putative ImmA/IrrE family metallo-endopeptidase domain, based on possession of a conserved metalloprotease zinc-binding motif, HEXXH (Jongeneel *et al.*, 1989). In addition, the toxin-antitoxin database RASTA (Sevin and Barloy-Hubler, 2007) annotates ErfRP as a putative toxin-antitoxin module. We found this interesting as the antitoxin component (*e.g. erfR*) of toxin-antitoxin modules are often essential, because they prevent the accumulation of the toxin protein. While we cannot exclude the hypothesis that ErfRP is a toxin-antitoxin module, exhaustive literature searches provide evidence for other functions for ErfRP. First, model λ cI repressors have an N-terminal domain that binds DNA and a 3' domain that functions in cI autoproteolysis, in conjunction with RecA (Galkin *et al.*, 2009). In *S. mutans,* it is possible that these functions are separated into two proteins, ErfR and ErfP. Next, cI-like repressors regulate prophage induction in bacteria but they have also been shown to regulate integrative

conjugative element (ICE) expression (Bellanger *et al.*, 2007; Bose *et al.*, 2008). Notably, *erfRP* reside in the flanking region of the putative ICE designated as TnSmu1 (Fig. 1C). Very little is known about the activity or the mobility of this element, but it shares similarities with ICE*St1*, ICE*St3*, Tn*916* and ICE*Bs1*. Although the regulatory regions of these different ICEs are substantially re-arranged compared to TnSmu1, ICE*St1* and ICE*Bs1* contain cI-like repressors that repress ICE expression (Bellanger *et al.*, 2007; Bose *et al.*, 2008; Carraro *et al.*, 2011). Therefore, we have formulated two major hypothesis explaining the essential nature of *erfR*: 1) ErfR is an antitoxin in a toxin-antitoxin module or 2) ErfR is involved in the activation of TnSmu1. CRISPR-mediated knockdown of *erfR* would therefore lead either to the over-accumulation of ErfP, which acts as a toxin, or cause over-expression of TnSmu1, with lethality arising from an unknown mechanism.

Here we investigate both of our hypotheses for explaining the essentiality of *erfR*. We use a combinational molecular approach employing transposon mutagenesis, genome sequencing, RNA-seq, bioinformatics and standard molecular microbiology techniques. We find that *erfR* is very likely to be a repressor of the ICE TnSmu1. Loss of repression leads to excision and circularization of TnSmu1, upregulation of the TnSmu1 element, and broader effects on *S. mutans* physiology. The observation that activation of an ICE is lethal extends similar findings made in other microorganisms and emphasizes the complex interaction of mobile elements with their host.

RESULTS AND DISCUSSION

Evidence that SMU_218/*erfR* is not an antitoxin.

Following our initial observations, we began by investigating the possibility that *erfRP* is a toxin-antitoxin module. We reasoned that we would be able to mutagenize both *erfR* and *erfP* at the same time, as this would not lead to the accumulation of the putative toxin. For other toxin-antitoxin modules (e.g. *mazEF* or *relBE*) double knockout of both genes is possible, whereas single knockout of the antitoxin module is not permitted. Under the conditions tested, we have not been able to obtain a double mutant of *erfRP* (Figure S1). However, deletion of the putative metalloprotease, *erfP*, is permitted by *S. mutans* cells (Figure S1). A lack of viability for a double knockout provides some evidence that *erfRP* is not a toxin-antitoxin module. To confirm, we also created over-expression strains in *E. coli* to test the toxicity of each protein. The *erfR* and

erfP genes were cloned into a pBAD protein expression vector, where the genes are induced by the addition of the monosaccharide arabinose. When the expression of *erfP* was induced, there was only a minor impact on the growth rate of *E. coli*, suggesting that ErfP accumulation in *E. coli* is not toxic (Figure S2). Over-expression of the EfrR protein did cause a moderate growth phenotype in *E. coli*. The growth phenotype was absent when *erfR* expression was not induced by arabinose. To summarize, the data do not support the hypothesis that ErfRP constitute a toxin-antitoxin module.

A bypass of essentiality screen reveals genes that when inactivated allow the deletion of *erfR.*

Under the conditions we have tested, *erfRP* is non-mutable (*i.e.* essential) (Figure 1; Figure 2A). However, we hypothesized that transformation of the *erfRP* deletion construct may be possible if introduced into a transposon library previously generated by our laboratory (Shields *et al.*, 2018). This is because cells that contain a transposon that interrupt a gene responsible for *erfRP* essentiality would tolerate *erfRP* deletion. This is known as bypass of essentiality (BoE), with systematic studies of essential gene bypass having been conducted in yeast (Li *et al.*, 2019; van Leeuwen *et al.*, 2020). BoE has been shown for several essential genes in bacteria including *zipA*, c-di-AMP null mutants, RNase E, *bamD*, and *ftsH* (Anupama *et al.*, 2011; Pichoff *et al.*, 2015; Krüger *et al.*, 2021; Wang *et al.*, 2021; Hart and Silhavy, n.d.). Utilizing a BoE screen we were able to isolate eight double mutants capable of tolerating *erfRP* deletion (Figure 2A). These mutants were then subjected to complete genome sequencing, to identify the location of the transposon insertion and identify suppressor mutations in the event these were also generated. Genome sequencing revealed the transposon insertions to be within the following regions: SMU_197c, SMU_201c (isolated twice), SMU_141c, *recN*, *perR*, the *murE* and SMU_1678 intergenic region, and the intergenic region between 16S rDNA and SMU_1750c (Table S1). The growth phenotypes of these strains were variable (Figure 2B). The intergenic transposon insertion into *murE* caused a considerable growth phenotype, as did an insertion in SMU_197c. Certain double mutant isolates grew almost the same as wild-type *S. mutans*, including *peR*::Tn, *recN*::Tn and SMU_141c::Tn.

We hypothesized that because of the strong selective pressure placed on cells during transformation, mutants could be created that are experiencing a duplication event; this is the loss of the essential module *erfRP* but a copy of *erfRP* is still present. To account for this, we

used PCR to visualize the size of the *erfRP* region. Gene duplication events occurred in the strains *perR*::Tn, *recN*::Tn, and the strain with a transposon insertion between 16S rDNA and SMU_1750c (Figure S3; Table S1). Double mutants with an *erfR* gene duplication grew better than strains that fully lacked *erfR* (Figure 2B). For the other five mutants, only one band for *erfR* was amplified and the size was as expected for a deletion mutant. As an additional control, PCR was also conducted to confirm that the Magellan6 transposon was inserted in the regions identified by whole genome sequencing. Transposon insertions were apparent for all strains except for SMU_141c::Tn where the results were inconclusive (Figure S4; Table S1).

The observation that inactivation of nearby conjugation-related genes (SMU_197c and SMU_201c) leads to increased viability of *erfRP* deletion is a prominent finding. This finding has synergy with our hypothesis that *erfR* is a regulator involved in the expression of the putative ICE TnSmu1 (Figure 1C). This suggests that loss of *erfR* leads to increased expression of TnSmu1, which is deleterious to *S. mutans*, and that a putative conjugation apparatus is involved in this phenotype. ICEs have complex relationships with their hosts. In many cases they carry beneficial genes, such as virulence factors or antibiotic resistance genes. However, expression of ICEs has been found to have a major impact on certain hosts. For example, *setR* cannot be deleted, a CI-like repressor found on a *Vibrio cholerae* mobile element known as SXT (Beaber *et al.*, 2002). It was also recently observed that activation of Tn*916* causes severe cell growth defects (Bean *et al.*, 2022:916). Tn*916* only activates in a small percentage (0.1-3%) of cells and the lethality of Tn*916* was discovered by activating it in a much larger proportion of cells. The deleterious effects of Tn*916* were less impactful in cells lacking conjugation genes that reside in Tn*916*, and a gene, *yqaR* found within a defective phage-like element *skin*. Intriguingly, Tn*916* can also cause cell death in *E. faecalis* and *B. subtilis*, species that do not encode for a *yqaR*-like gene.

Loss of *erfR* leads to increased genome sequencing reads from the putative ICE TnSmu1.

We have evidence that *erfR* may be involved in the activation of the putative ICE TnSmu1. Next, we noticed that in several double mutant backgrounds, read coverage was significantly elevated for a region between SMU_191c and SMU_220c (Figure 3; Figure S5). Read coverage depth in this region was ~300 for *S. mutans* UA159 but was elevated to as much as ~5000 reads in double mutant backgrounds. This increased read depth corresponds exactly to a region of the

genome that is predicted to contain the putative ICE, TnSmu1. ICEs are mobile genetic elements that exist on a host chromosome that are able to excise and transfer to a new host by conjugation (Johnson and Grossman, 2015). This region is predicted to be an ICE and exists in the ICE database ICEberg 2.0 (Liu *et al.*, 2019) but to our knowledge has not been experimentally verified. To be considered an ICE, these mobile elements must contain several important features: a recombinase (also known as an integrase), a specific site of attachment (typically a tRNA gene), conjugation machinery, a relaxase, and an *oriT* site. Several of these features exist within the TnSmu1 region (Figure 3C). SMU_191c encodes for a putative tyrosine recombinase that may function to catalyze DNA breakage and rejoining. Immediately adjacent to the 5' end of SMU_191c is tRNA-leu, which is likely the integration site. The attachment (*att*) sites *attL* and *attR* have the sequence CTATACCGGCGGCCG. A putative MOB_T family relaxase is encoded by the gene SMU_207c and may function to convert a circular form of TnSmu1 into ssDNA prior to conjugation. Several genes within TnSmu1 are predicted to be members of a type IV secretion system (SMU_196c, SMU_197c, SMU_198c, SMU_199c, SMU_201c, and SMU_208c). A putative *oriT* site is located between SMU_207c and SMU_208c, with the sequence ACCCCCCTATTAGTATCGGGGGG. This shares similarities with the *oriT* site in the ICEs ICE*Bs1* (ACCCCCCCACGCTAACAGGGGGGT) and ICE*St3* (ACCCCCGATTTCTAATAGGGGGGT). The relaxase encoded by SMU_207c may recognize this *oriT* sequence, where it would nick the ICE DNA to form transfer DNA (T-DNA). Altogether, TnSmu1 contains many of the genes that are required for an ICE, and may allow excision, replication, transfer, and integration of the element. In addition, when *erfR* is deleted from the chromosome there is a significant increase in read coverage depth across the TnSmu1 region. This observation is consistent with ErfR being involved in the activation of the mobile element.

Evidence of a circular intermediate form of TnSmu1.

When excised from the host chromosome ICEs exist as circular DNA molecules (Johnson and Grossman, 2015). For some ICEs these circular intermediate forms are capable of rolling-circle replication (Wright *et al.*, 2015). Increased read coverage depth suggests that loss of *erfR* is leading to excision and circularization of TnSmu1. In addition to the observation that sequencing reads increase when *erfR* is deleted, there was also evidence of unexpected genomic junctions in strains of *S. mutans* missing the *erfR* gene. There was evidence from the genome sequencing that *attL* and attR are combining, which is leading to an unexpected junction forming between SMU_191c and SMU_220c. This is additional evidence that TnSmu1 is excising from

the host chromosome and forming a circular dsDNA molecule. To confirm these observations, we designed primers to test excision and circularization. For excision, these primers will only amplify a PCR product if TnSmu1 is no longer present on the host chromosome. For circularization, a PCR product will form if there is a junction between the ends of the ICE (*i.e.* SMU_191c and SMU_220c). Excision and circularization of TnSmu1 in wild-type *S. mutans*, grown to mid-log, occurs close to the limit of detection for the assay (Figure 4A and B). However, excision and circularization of TnSmu1 is apparent in *erfR* double mutant strains. For Δ*erfR* strains carrying transposon insertions within putative TnSmu1 conjugation genes (SMU_197c and SMU_201c) there are multiple copies of TnSmu1 per cell. In these strains excision is occurring 1000-fold more compared to wild-type *S. mutans*. We also observed circularization/excision in two strains experiencing duplication of *erfR*, *perR*::Tn and *recN*::Tn. We chose to examine these strains because they are much healthier than the strains with disrupted conjugation genes (Figure 2B). For these two strains, circularization of TnSmu1 was occurring but excision of the element was measured close to the limit of detection for the assay. Taken together, our observations of excision and circularization are consistent with TnSmu1 having this important life-cycle feature of ICEs.

Induction of TnSmu1 by the genotoxic agent mitomycin C.

The *erfRP* module shares similarities to CI-like phage repressors that are sensitive to DNA damage (via induction of the SOS response) (Casjens and Hendrix, 2015). Other ICEs contain similar regulatory modules, ICE*Bs1* being the most extensively characterized but also ICE*St3* and SXT (Beaber *et al.*, 2004; Auchtung *et al.*, 2005; Bellanger *et al.*, 2007). ICE*Bs1* contains a repressor, ImmR, that is cleaved by a protease ImmA, in a two gene organization like *erfRP* but in a different location within the ICE (next to the integrase gene). To determine if DNA damage was an inducer of TnSmu1, possibly via cleavage of ErfR, we incubated *S. mutans* UA159 in the presence of genotoxic agent mitomycin C (MMC). When treated with MMC there was a clear dose-dependent response, with clear growth hindrance at 25 ng/mL (Figure 4C). Induction of TnSmu1 circularization was investigated at MMC concentrations of 3.125 ng/mL, 6.25 ng/mL and 12.5 ng/mL. With increasing concentrations of MMC, the relative abundance of circular TnSmu1 increased in a dose-dependent manner (Figure 4D). At a concentration of 12.5 ng/mL not all cells have a circular copy of TnSmu1 but circularization is 1000-fold above levels seen in untreated *S. mutans* cells. Induction of TnSmu1 circularization via MMC is consistent with similar phenotypes displayed by other ICEs, including those with similar regulatory modules. It is

hypothesized that ICE induction in response to DNA damage acts to maintain the element by forcing horizontal transmission if vertical transmission is no longer guaranteed (because of cell death) (Ryan *et al.*, 2016). Further experiments are required to fully understand the nature of the induction, if it is via an SOS response and if the putative protease ErfP cleaves ErfR to cause de-repression of the element.

Repression of *erfR* with CRISPR inference strongly up-regulates TnSmu1.

Next, we wanted to gain further insights into potential regulation of TnSmu1 by ErfR and broader effects on *S. mutans* UA159 physiology. To do this we performed RNA-seq profiling of an *S. mutans* strain experiencing knockdown of the *erfR* gene using CRISPRi. As already shown, when *erfR* is repressed by CRISPRi there is a strong growth defect (Figure 1). RNA was collected from uninduced and induced (with 0.1% xylose) cultures grown to OD_{600} ~0.4, and this RNA was then processed for RNA-seq. The RNA-seq analysis identified 134 genes whose $\log_2$ fold change was ≥2 after repression of *erfR* (FDR < 0.05; Table S2). Of these genes, 34 were repressed and 100 were induced (Figure 5A).

There are several notable trends identified by transcriptome analysis of the *erfR* depleted strain. Firstly, every gene within the TnSmu1 element is strongly induced (Figure 5A). Most TnSmu1 genes were up-regulated by greater than 20-fold, with SMU_204c (hypothetical protein) induced by over 100-fold ($\log_2$ fold change = 10.03). Of the twenty most strongly induced genes, all belonged to the TnSmu1 region. Visual inspection of the TnSmu1 region, with reads shown and comparable to a non-induced strain, highlights the strong induction across the element (Figure 6). Worth highlighting is that two intergenic regions within TnSmu1 are strongly induced. The first region is between SMU_217c and SMU_218, and the second region is between SMU_219 and SMU_220c. For the first region, we do not currently know what is being produced (*e.g.* small RNA, small peptide, antitermination system *etc*). The second region contains a chromosomally encoded type I toxin-antitoxin system (Koyanagi and Lévesque, 2013). Expression of a toxic peptide known as Fst-Sm, encoded by this system, is toxic when over-expressed in *S. mutans* (in the absence of the anti-toxin) (Koyanagi and Lévesque, 2013). The mechanism of action of the toxic peptide is unknown but plasmid encoded Fst peptides affect the integrity of the membrane, cause defects in cell division, and sensitize cells to nisin (probably by altering the cell membrane) (Weaver *et al.*, 2003; Patel and Weaver, 2006). More broadly, type I toxins have been shown to induce membrane perturbations, ROS production,

ATP drop, induction of membrane proteins and sugar transport systems, and induction of macromolecular synthesis (e.g. fatty acid biosynthesis and purine metabolism) (Nonin-Lecomte *et al.*, 2021).

Several other genes are up or downregulated after *erfR* depletion. A broad overview of the effect on the *S. mutans* transcriptome is shown in Figure 5B with categorization by Cluster of Orthologous Groups (COGs). Several genes involved in amino acid transport or metabolism are up-regulated including a cysteine transport operon *tcyDEFGH*, *argD*, *cysK*, *opcD*, *pepQ*, *ilvE*, *gatA*, *hipO*, SMU_1216c (putative cystine transporter), SMU_1486c (putative histidinol-phosphatase), and SMU_1938c (putative methionine transporter). Figure 6 shows reads mapped to both *cysK* and *tcyDEFGH* in uninduced and induced conditions. There is a sulfur amino acid regulatory pathway in *S. mutans*, with regulation controlled by MetR, CysR and HomR (Sperandio *et al.*, 2010). Up-regulation of *cysK* and *tcyDEFGH* may suggest that cells are experiencing aberrant cysteine metabolism. Cysteine is an important sulfur containing amino acid with many physiological roles, including disulfide bond formation, biogenesis of [Fe-S] clusters and protection against oxidative stress. Other transport systems that are also up regulated include *treB* (trehalose PTS), *fruA* (fructose PTS), ammonium transporter (*nrgA*), bacteriocin/competence transporters (*comB*, SMU_1889c, SMU_1897, *comG*, *mutE2* and *mutG*), *tauC* and *msmK*. The expression of several purine metabolism genes is altered with *purDEK* up-regulated and *purC* down-regulated. All the genes in the Type I-C CRISPR-Cas system are up-regulated when *erfR* is repressed (Figure 6). This system is presumed to not function in providing immunity against bacteriophage because of truncation in CRISPR spacer acquisition proteins (Cas1) (Serbanescu *et al.*, 2015). CRISPR-Cas systems have been found to confer protection against DNA damage (UV and MMC) but this protective role was not seen with the *S. mutans* Type I-C system (Serbanescu *et al.*, 2015). This may rule out a DNA repair hypothesis for explaining why this system is up regulated. Several genes are downregulated after CRISPRi repression of *erfR*, many of which are hypothetical proteins. Six translation genes are down-regulated, predominately 50S and 30S ribosomal proteins. Some, but not all, genes located in the genomic island TnSmu2 are downregulated (*mubD*, *mubC*, *mubH*, *mubG*, and *mubE*). This region encodes for a hybrid nonribosomal peptide (NRP) and polyketide (PK) system that produces mutanobactin (Wu *et al.*, 2010). A putative acyl carrier protein (SMU_27, *acpP*) is also down-regulated that may participate in PK synthesis (Crosby and Crump, 2012).

The main question that motivated this CRISPRi RNA-seq study was whether ErfR represses TnSmu1. When *erfR* is repressed there is a clear and strong up-regulation of TnSmu1, which indicates that this is the regulon of ErfR. In addition to this finding was the observation of two intergenic transcripts being expressed within TnSmu1, including a Type I toxin-antitoxin system. Besides strongly induced TnSmu1 expression other effects on *S. mutans* physiology were observed. These impacts were on amino acid transport/metabolism, transporter expression, CRISPR-Cas, purine metabolism, and translation. These broader effects may indicate processes that are impacted by the lethal/severe growth defect caused by TnSmu1 expression, but it is unclear what specific process is impacted and which are cascaded effects downstream. In previous transcriptome studies, acquisition of ICEs (i.e. ICE*clc* and ICE*Pmi*ChnXH1653) has been shown to have a small but measurable effect on gene expression (Miyazaki *et al.*, 2018; He *et al.*, n.d.). Transcriptome alterations would be expected to be small in those studies if the ICEs are only induced in a small sub-population of cells. However, as with induction of TnSmu1, global pathways such as amino acid metabolic pathways are altered when hosts acquire ICEs. Overall, RNAseq has provided additional insight into the host-mobile element partnership and associated costs when TnSmu1 is expressed.

Distribution and conjugative transfer of TnSmu1 genes between *S. mutans* strains.

To begin to understand if TnSmu1 is capable of conjugative transfer, we wanted to determine if closely and distantly related strains of *S. mutans* carry this ICE. Notably, we have recently discovered that *S. mutans* strains carry CRISPR spacers against this element (Walker and Shields, 2022). This finding suggests that *S. mutans* attempts to prevent transfer/acquisition of TnSmu1 via CRISPR-Cas mediated immunity, and it also provides evidence of potential transfer between strains. Figure 7 shows strains that carry TnSmu1 genes and is organized by strain relatedness. Distantly related strains, such as smu342 and UA159, carry full length TnSmu1 elements, which might be indicative of conjugal transfer. Although many *S. mutans* strains carry TnSmu1 genes, only nine strains carry the full element as organized on *S. mutans* UA159. ICEs have plasticity and lose/gain genes, often interacting with other mobile elements such as transposons, lysogenic phages and genomic islands. It is therefore not surprising that the structure of TnSmu1 changes as it transfers within the *S. mutans* species, particularly if its expression is lethal in other *S. mutans* hosts. Under normal conditions we anticipate that a very small sub-fraction of *S. mutans* cells would express TnSmu1 because of ErfR repression and induction being strongly deleterious. It is therefore likely that transfer of TnSmu1 from a donor

strain to a recipient strain occurs at very low efficiency. Similar ICEs in other Firmicutes bacteria, like ICE*Bs1* and Tn*916*, exhibit considerable differences in conjugation rates. In the future we plan to explore conjugative transfer of this element between strains in greater detail.

CONCLUSION

In this study, we sought to reveal the role of *erfR*, an essential gene annotated as a transcriptional regular or an antitoxin. We showed that *erfR* is likely not an antitoxin and is a repressor that controls the activation of the ICE TnSmu1. Activation of the ICE is lethal to *S. mutans*, hence why removal or repression of *erfR* is not tolerated. We explored several steps of the ICE life cycle, showing that TnSmu1 can be induced, that it produces a circular intermediate, and the element, or parts of it, are distributed among *S. mutans* strains. Several details of how TnSmu1 functions remain to be elucidated, including showing conjugation into recipient strains, a detailed understanding of the ErfRP regulatory module and how it controls cell fate, and what is causing activation of the ICE to be lethal. In addition, there are several hypothetical genes and an overproduced RNA within an intergenic region that need to be explored. Taken together, our results show that activation of a newly characterized ICE is strongly deleterious to *S. mutans*, highlighting the complex interplay between mobile elements and their host. We are continuing to investigate TnSmu1 as a model to understand ICE distribution among streptococci, and their effects on host fitness and cell biology.

MATERIALS AND METHODS

Bacterial strains and culture conditions.

Streptococcus mutans strains were cultured from single colonies in Brain Heart Infusion (BHI, Difco) broth. Unless otherwise stated *S. mutans* was routinely cultured at 37°C in a 5% CO_2, microaerophilic atmosphere. *Escherichia coli* strains were routinely cultured in LB broth (Lennox formula; 10 g/L tyrptone, 5 g/L yeast extract and 5 g/L NaCl) at 37°C with aeration. Antibiotics were added to growth media at the following concentrations: kanamycin (1.0 mg/mL for *S. mutans*, 50 µg/mL for *E. coli*), spectinomycin (1.0 mg/mL for *S. mutans*), ampicillin (100 µg/mL for *E. coli*). A list of strains and plasmids (Table S3) and oligonucleotide primers (Table S4) can be found in the supplementary material.

Gene mutagenesis and plasmid cloning.

Standard DNA manipulation techniques were used to engineer deletion mutant strains (Sambrook and Russell, 2001). A PCR ligation mutagenesis method was used to replace genes with non-polar kanamycin markers (Lau *et al.*, 2002). For each gene deletion, primers A and B were designed to amplify 500-600 bp upstream of the coding sequence (with ca. 50-bp overlapping the coding sequence of the gene). Primers C and D were designed to amplify 500 to 600 bp downstream of the coding sequence (with ca. 50-bp overlapping the coding sequence of the gene). Primers B and C contained *Bam*HI restriction enzyme sites for ligation of the AB and CD fragments to a non-polar kanamycin cassette digested from plasmid pALH124 (Ahn and Burne, 2006). Transformants were selected on BHI agar containing kanamycin. Double-crossover recombination, without introduction of nearby secondary mutations, was confirmed by PCR and Sanger sequencing using primers E and F, away from the site of recombination.

Plasmid cloning was conducted using the protein expression plasmid pBAD/His/A that contains an arabinose inducible promoter for tightly regulated protein production. PCR products for SMU_218, SMU_219 and SMU_218-219 were amplified and digested, before being ligated into digested pBAD/His/A plasmid. Correct in-frame insertion was verified with Sanger sequencing.

Transposon interaction screen

To look for genetic interactions of *erfR* with other genes, a previously created Tn-seq library (Shields *et al.*, 2018) was transformed with a *erfR* deletion product, described above. Transformation of the library was completed in triplicate and each transformation reaction was screened on at least ten agar plates containing spectinomycin (transposon cassette) and kanamycin (*erfR* deletion construct). Colonies that grew and were potentially viable *erfR* deletions (plus a transposon cassette) were picked and plated onto fresh antibiotic-containing media. After this, colonies were cultured overnight and frozen in glycerol at -80°C. In addition, control experiments where conducted. Controls included attempting to delete *erfR* with the deletion construct and selection of the Tn-seq library on kanamycin (to observe background spontaneous resistance).

Interaction screen genome sequencing analysis

Genomic DNA was isolated from strains using a MasterPure Gram Positive DNA (Epicentre) purification kit with modifications as previously described (Shields *et al.*, 2013). After DNA purification, total DNA concentration and purity was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific). Purified genomic DNA was sent to the Microbial

Genome Sequencing Center, and samples were sequenced according to their protocol. Sequencing reads were compared to a *S. mutans* UA159 GenBank file (NC_004350.2) using Breseq. Following genomic sequencing, transposon insertion locations were verified with PCR and Sanger sequencing (primers available in Table S4). Correct deletion of *erfRP* was also verified using the primer pair 218_219E and 219F, followed by gel electrophoresis and comparison to expected product sizes.

qPCR measurements of TnSmu1 excision and circularization

Quantitative PCR (qPCR) was used to calculate the excision and circularization frequency of TnSmu1. Specific primers, which only create products during excision or circularization, were used in the qPCR assay (Table S4). Primers which amplified a chromosomal single copy gene, *sloR*, close to TnSmu1 were also used to standardize results. Primers were designed using the qPCR settings in the Primer3plus online application (Untergasser *et al.*, 2007). Cell samples were grown to mid-log in rich media and DNA was collected using a MasterPure Gram Positive DNA (Epicentre) purification kit with modifications as previously described (Shields *et al.*, 2013). Standard curves for each primer pair were generated using eight 10-fold dilutions of PCR products, starting with 10^8 copies/µl. Triplicates of standard curve DNAs, samples and cDNA controls were added to wells containing iQ SYBR Green Supermix (Bio-Rad) with primers (0.4 µM). Thermocycling was carried out using a CFX Touch Real-Time PCR Detection System (Bio-Rad) set to the following protocol: 40 cycles of 95°C for 10 s and 60°C for 45 s, with a starting cycle of 95°C for 30 s.

CRISPRi RNAseq

RNA was extracted from OD_{600} = 0.5 bacterial cultures using the RNeasy Mini Kit. RNA was extracted from biological triplicates of CRISPRi strains carrying a short guide RNA (sgRNA) targeting SMU_218 in the presence or absence of the dCas9 inducing molecule xylose. Next, RNA was sent to the Microbial Genome Sequencing Center who generated RNAseq libraries and sequenced samples according to their protocol. After sequencing, read counts were aligned to the *S. mutans* UA159 genome using bioinformatics tools hosted on the Galaxy server maintained by the Research Computing Center at the University of Florida. Gene expression changes between samples were quantified with Degust (http://degust.erc.monash.edu/) using the edgeR methodology.

TnSmu1 distribution bioinformatics

Upon acknowledging the TnSmu1 operon presence in *S. mutans* UA159 we isolated the gene sequences of all the individual genes by constructing a BED type file, which was parsed to the getfasta command from the Bedtools suite (Quinlan and Hall, 2010). The resulting individual FASTA files with each gene sequence were blasted against a collection of over 600 clinical isolate genomes in our possession. The blast output was then filtered with the objective of keeping only those strains showing both a percent identity greater than 75% and a ratio between the query length and the sequence length greater than 0.9. From this filtered blast output we isolated the PID column for each isolate across all genes and generated a matrix as input for the heatmap. In parallel we conducted a Multi-locus Sequence Typing (MLST) analysis with all resulting strains in the curated blast output. The MLST sequence was then parsed to PHYML (Guindon and Gascuel, 2003) to construct a Maximum-likelihood phylogeny tree. Finally the phylogenies and the heatmap were plot together in R Statistical Language (R: a language and environment for statistical computing, n.d.) with packages ggplot2 (Valero-Mora, 2010) and GGTREE (Yu *et al.*, 2017).

ACKNOWLEDGMENTS

This work was supported by the National Institute for Dental and Craniofacial Research (NIDCR) grant DE029882 awarded to R.C.S.

CONFLICT OF INTEREST

The authors declare no competing financial interest.

AUTHOR CONTRIBUTIONS

S.K., A.Q., K.K., A.R.W., and R.C.S. were involved in the investigation, methodology, data curation, formal analysis, and revised the manuscript; S.K. and R.C.S. were involved in project administration; A.R.W. performed software development (implementation of code and supporting algorithms); R.C.S. was involved with funding acquisition, supervision, conceptualization of the study and drafted the manuscript.

References

Ahn, S.-J., and Burne, R.A. (2006) The *atlA* operon of *Streptococcus mutans*: role in autolysin maturation and cell surface biogenesis. *J Bacteriol* **188**: 6877–88.

Anupama, K., Leela, J.K., and Gowrishankar, J. (2011) Two pathways for RNase E action in *Escherichia coli in vivo* and bypass of its essentiality in mutants defective for Rho-dependent transcription termination. *Mol Microbiol* **82**: 1330–1348.

Auchtung, J.M., Lee, C.A., Monson, R.E., Lehman, A.P., and Grossman, A.D. (2005) Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc Natl Acad Sci* **102**: 12554–12559.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2002) Genomic and Functional Analyses of SXT, an Integrating Antibiotic Resistance Gene Transfer Element Derived from Vibrio cholerae. *J Bacteriol* **184**: 4259–4269.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**: 72–74.

Bean, E.L., McLellan, L.K., and Grossman, A.D. (2022) Activation of the integrative and conjugative element Tn *916* causes growth arrest and death of host bacteria. http://biorxiv.org/lookup/doi/10.1101/2022.04.01.486793.

Bellanger, X., Morel, C., Decaris, B., and Guédon, G. (2007) Derepression of Excision of Integrative and Potentially Conjugative Elements from *Streptococcus thermophilus* by DNA Damage Response: Implication of a cI-Related Repressor. *J Bacteriol* **189**: 1478–1481.

Bose, B., Auchtung, J.M., Lee, C.A., and Grossman, A.D. (2008) A conserved anti-repressor controls horizontal gene transfer by proteolysis. *Mol Microbiol* **70**: 570–582.

Carraro, N., Libante, V., Morel, C., Decaris, B., Charron-Bourgoin, F., Leblond, P., and Guédon, G. (2011) Differential regulation of two closely related integrative and conjugative elements from *Streptococcus thermophilus*. *BMC Microbiol* **11**: 238.

Casjens, S.R., and Hendrix, R.W. (2015) Bacteriophage lambda: Early pioneer and still relevant. *Virology* **479–480**: 310–330.

Crosby, J., and Crump, M.P. (2012) The structural role of the carrier protein – active controller or passive carrier. *Nat Prod Rep* **29**: 1111–1137.

Galkin, V.E., Yu, X., Bielnicki, J., Ndjonka, D., Bell, C.E., and Egelman, E.H. (2009) Cleavage of bacteriophage lambda cI repressor involves the RecA C-terminal domain. *J Mol Biol* **385**: 779–787.

Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.

Hart, E.M., and Silhavy, T.J. Functions of the BamBCDE Lipoproteins Revealed by Bypass Mutations in BamA. *J Bacteriol* **202**: e00401-20.

He, J., Sun, L., Zhang, L., Leptihn, S., Yu, Y., and Hua, X. A Novel SXT/R391 Integrative and Conjugative Element Carries Two Copies of the blaNDM-1 Gene in *Proteus mirabilis*. *mSphere* **6**: e00588-21.

Johnson, C.M., and Grossman, A.D. (2015) Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu Rev Genet* **49**: 577–601.

Jongeneel, C.V., Bouvier, J., and Bairoch, A. (1989) A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Lett* **242**: 211–214.

Koyanagi, S., and Lévesque, C.M. (2013) Characterization of a *Streptococcus mutans* intergenic region containing a small toxic peptide and its cis-encoded antisense small RNA antitoxin. *PloS One* **8**: e54291.

Krüger, L., Herzberg, C., Rath, H., Pedreira, T., Ischebeck, T., Poehlein, A., *et al.* (2021) Essentiality of c-di-AMP in *Bacillus subtilis*: Bypassing mutations converge in potassium and glutamate homeostasis. *PLOS Genet* **17**: e1009092.

Lau, P.C.Y., Sung, C.K., Lee, J.H., Morrison, D.A., and Cvitkovitch, D.G. (2002) PCR ligation mutagenesis in transformable streptococci: application and efficiency. *J Microbiol Methods* **49**: 193–205.

Leeuwen, J. van, Pons, C., Tan, G., Wang, J.Z., Hou, J., Weile, J., *et al.* (2020) Systematic analysis of bypass suppression of essential genes. *Mol Syst Biol* **16**: e9828.

Letunic, I., and Bork, P. (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* **46**: D493–D496.

Li, J., Wang, H.-T., Wang, W.-T., Zhang, X.-R., Suo, F., Ren, J.-Y., *et al.* (2019) Systematic analysis reveals the prevalence and principles of bypassable gene essentiality. *Nat Commun* **10**: 1002.

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., *et al.* (2019) ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* **47**: D660–D665.

Loesche, W.J. (1986) Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* **50**: 353–80.

Miyazaki, R., Yano, H., Sentchilo, V., and Meer, J.R. van der (2018) Physiological and transcriptome changes induced by *Pseudomonas putida* acquisition of an integrative and conjugative element. *Sci Rep* **8**.

Nonin-Lecomte, S., Fermon, L., Felden, B., and Pinel-Marie, M.-L. (2021) Bacterial Type I Toxins: Folding and Membrane Interactions. *Toxins* **13**: 490.

Patel, S., and Weaver, K.E. (2006) Addiction Toxin Fst Has Unique Effects on Chromosome Segregation and Cell Division in *Enterococcus faecalis* and *Bacillus subtilis*. *J Bacteriol* **188**: 5374–5384.

Pichoff, S., Du, S., and Lutkenhaus, J. (2015) The bypass of ZipA by overexpression of FtsN requires a previously unknown conserved FtsN motif essential for FtsA–FtsN interaction supporting a model in which FtsA monomers recruit late cell division proteins to the Z ring. *Mol Microbiol* **95**: 971–987.

Pitts, N.B., Zero, D.T., Marsh, P.D., Ekstrand, K., Weintraub, J.A., Ramos-Gomez, F., *et al.* (2017) Dental caries. *Nat Rev Dis Primer* **3**: 17030.

Quinlan, A.R., and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Oxf Engl* **26**: 841–842.

R: a language and environment for statistical computing https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing.

Ryan, M.P., Armshaw, P., and Pembroke, J.T. (2016) SXT/R391 Integrative and Conjugative Elements (ICEs) Encode a Novel 'Trap-Door' Strategy for Mobile Element Escape. *Front Microbiol* **7** 2022.

Sambrook, J., and Russell, D. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Serbanescu, M.A., Cordova, M., Krastel, K., Flick, R., Beloglazova, N., Latos, A., *et al.* (2015) Role of the *Streptococcus mutans* CRISPR-Cas Systems in Immunity and Cell Physiology. *J Bacteriol* **197**: 749–761.

Sevin, E.W., and Barloy-Hubler, F. (2007) RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biol* **8**: R155.

Sheiham, A., and James, W.P.T. (2015) Diet and Dental Caries: The Pivotal Role of Free Sugars Reemphasized. *J Dent Res* **94**: 1341–1347.

Shields, R.C., Mokhtar, N., Ford, M., Hall, M.J., Burgess, J.G., Elbadawey, M.R., and Jakubovics, N.S. (2013) Efficacy of a Marine Bacterial Nuclease against Biofilm Forming Microorganisms Isolated from Chronic Rhinosinusitis. *PloS ONE* **8**: e55339–e55339.

Shields, R.C., Walker, A.R., Maricic, N., Chakraborty, B., Underhill, S.A.M., and Burne, R.A. (2020) Repurposing the *Streptococcus mutans* CRISPR-Cas9 System to Understand Essential Gene Function. *PLoS Pathog* **16**: e1008344.

Shields, R.C., Zeng, L., Culp, D.J., and Burne, R.A. (2018) Genomewide Identification of Essential Genes and Fitness Determinants of *Streptococcus mutans* UA159. *mSphere* **3**: e00031-18.

Sperandio, B., Gautier, C., Pons, N., Ehrlich, D.S., Renault, P., and Guédon, E. (2010) Three Paralogous LysR-Type Transcriptional Regulators Control Sulfur Amino Acid Supply in *Streptococcus mutans*. *J Bacteriol* **192**: 3464–3473.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A.M. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**: W71-4.

Valero-Mora, P.M. (2010) ggplot2: Elegant Graphics for Data Analysis. *J Stat Softw* **35**: 1–3.

Walker, A.R., and Shields, R.C. (2022) Investigating CRISPR spacer targets and their impact on genomic diversification of *Streptococcus mutans*. 2022.04.14.488371 https://www.biorxiv.org/content/10.1101/2022.04.14.488371v1

Wang, Y., Cao, W., Merritt, J., Xie, Z., and Liu, H. (2021) Characterization of FtsH Essentiality in *Streptococcus mutans* via Genetic Suppression. *Front Genet* **12**: 659220.

Weaver, K.E., Weaver, D.M., Wells, C.L., Waters, C.M., Gardner, M.E., and Ehli, E.A. (2003) *Enterococcus faecalis* Plasmid pAD1-Encoded Fst Toxin Affects Membrane Permeability and Alters Cellular Responses to Lantibiotics. *J Bacteriol* **185**: 2169–2177.

Wright, L.D., Johnson, C.M., and Grossman, A.D. (2015) Identification of a Single Strand Origin of Replication in the Integrative and Conjugative Element ICEBs1 of *Bacillus subtilis*. *PLOS Genet* **11**: e1005556.

Wu, C., Cichewicz, R., Li, Y., Liu, J., Roe, B., Ferretti, J., *et al.* (2010) Genomic Island TnSmu2 of *Streptococcus mutans* Harbors a Nonribosomal Peptide Synthetase-Polyketide Synthase Gene Cluster Responsible for the Biosynthesis of Pigments Involved in Oxygen and H2O2 Tolerance. *Appl Environ Microbiol* **76**: 5815–5826.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**: 28–36.

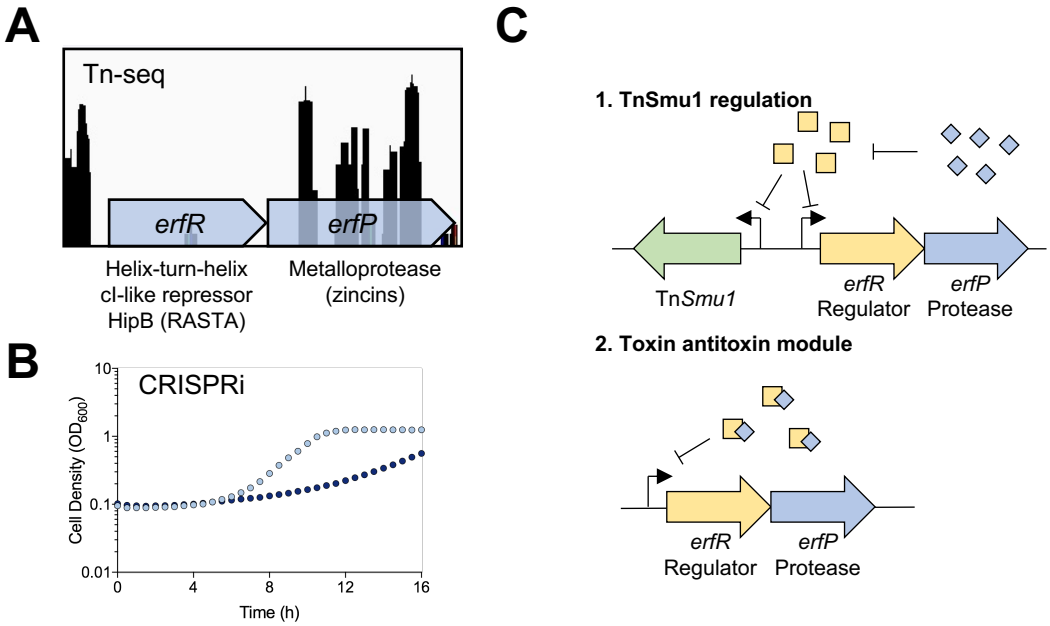

Figure 1. Discovery of an essential gene with an undefined function. There are no transposon insertions (black bars) found within *erfR* during Tn-seq analysis, which indicates that the gene is essential (A). Annotations of *erfR* and *erfP* are shown below each gene. Repression of *erfR* using a CRISPRi system leads to a significant growth defect (dark blue dots) compared to a control strain (light blue dots) (B). (C) ErfR is annotated as either a transcriptional regulator or a HipB-like antitoxin. Considering both annotation and the genomic context of *erfR*, there are two potential roles of this gene: 1) it is part of a regulatory module contained within the ICE TnSmu1, or 2) it is an antitoxin gene in a toxin-antitoxin modue.

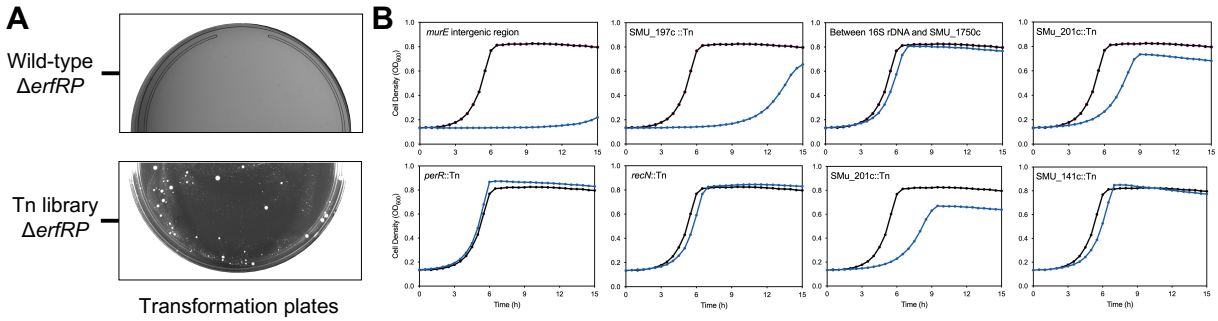

Figure 2. A transposon interaction screen identifies gene mutations that bypass the essentiality of *erfRP*. (A) An *erfRP*::*aphA3* mutagenesis fragment was transformed into a *S. mutans* Tn-seq library with selection on antibiotic-containing agar. Deletion of *erfRP* is not possible in a wild-type strain of *S. mutans* (top panel). White *S. mutans* colonies were obtained when *erfRP* was transformed into the Tn-seq library (bottom panel). (B) After genome sequencing analysis, several strains were identified that allow *erfRP* deletion. Each strain, with the corresponding transposon insertion site, is shown in a single growth analysis chart (dark line, wild-type strain; light blue line, double mutant strain).

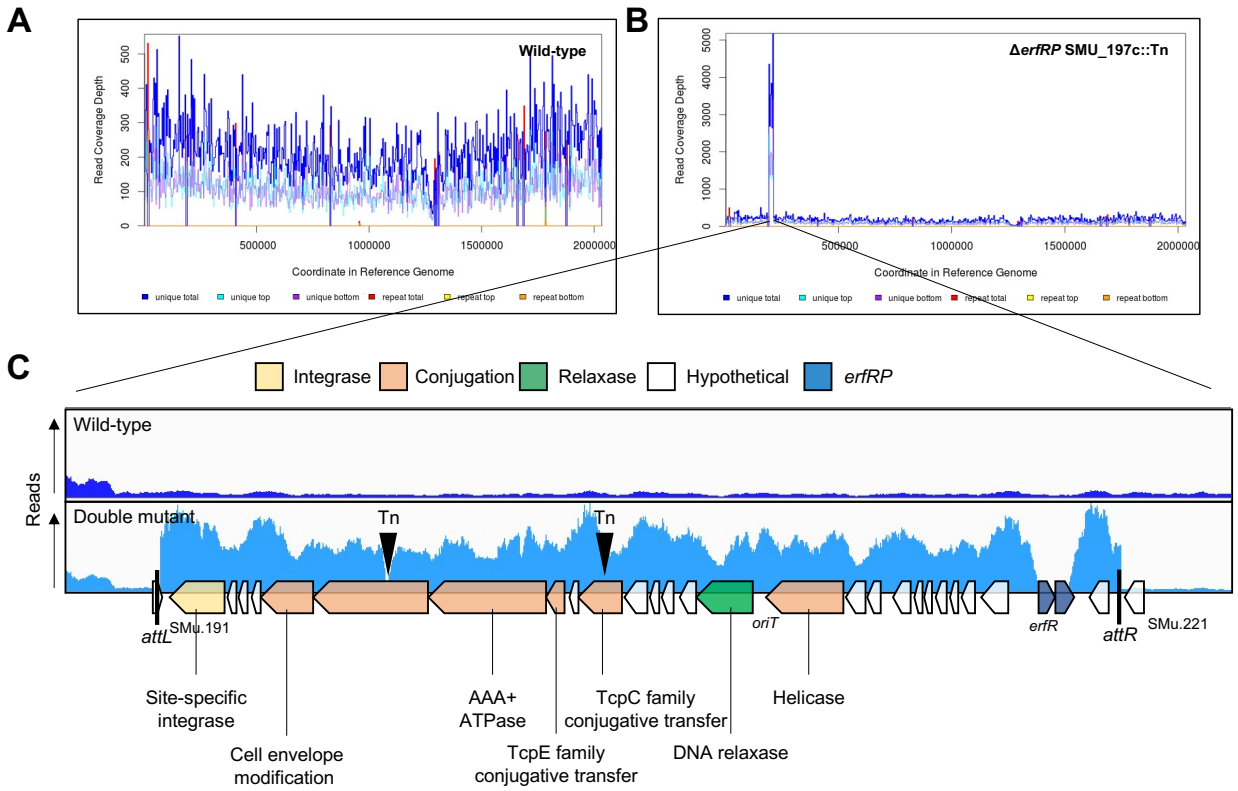

Figure 3. Mutations that bypass the essentiality of *erfRP* cause an increase in genome sequencing reads from the ICE TnSmu1. Read coverage depth for both a wild-type (A) and a double mutant strain Δ*erfRP*::*aphA3* SMU_197c::Tn (B) is shown. Genome sequencing read depth was substantially higher across the TnSmu1 region in the double mutant background. (C) Sequencing reads were visualized with the Integrated Genome Viewer for only the TnSmu1 region (and immediate up- and downstream genes). Read coverage is ca. 10-fold higher across the entire TnSmu1 region, except for the deleted *erfRP* genes. The location of each transposon insertion, within putative conjugation genes, which allowed *erfRP* deletion is shown as upside-down triangles. The TnSmu1 region is annotated with putative functions related to excision/recombination (integrase), conjugation, T-DNA formation (relaxase), strand nicking (*oriT*) and attachment sites (*attL* and *attR*).

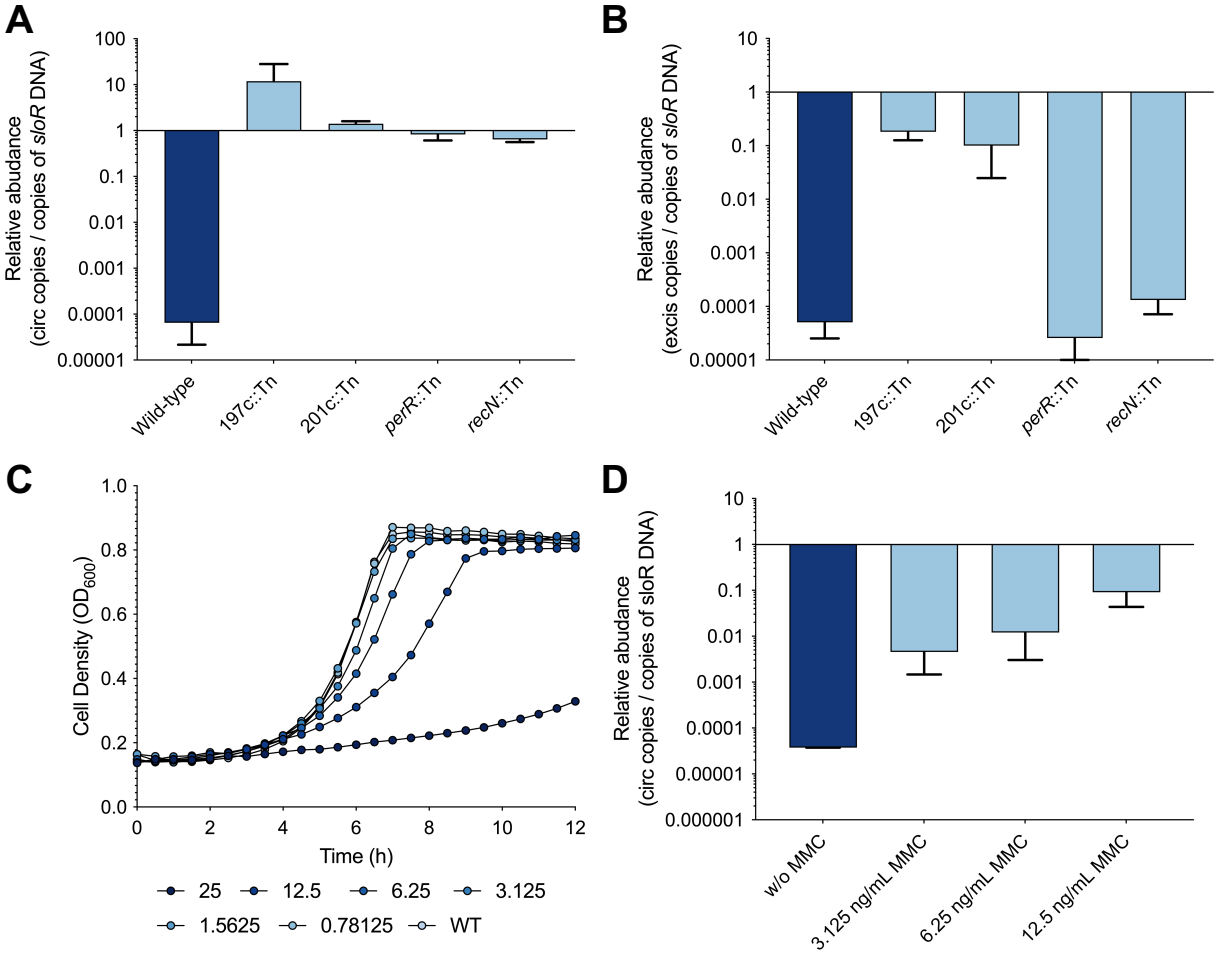

Figure 4. TnSmu1 is capable of circularization and excision. Quantitative PCR was employed to measure the relative abundance of circular (A) and excised (B) copies of TnSmu1. The number of copies was compared to a nearby single-copy chromosomal gene, *sloR*. Induction of TnSmu1 circularization was also measured by inducing TnSmu1 with the genotoxic agent mitomycin C (MMC). The impact of different concentrations of MMC was quantified using a growth assay (C), and then different doses of MMC were used to induce TnSmu1 circularization (D).

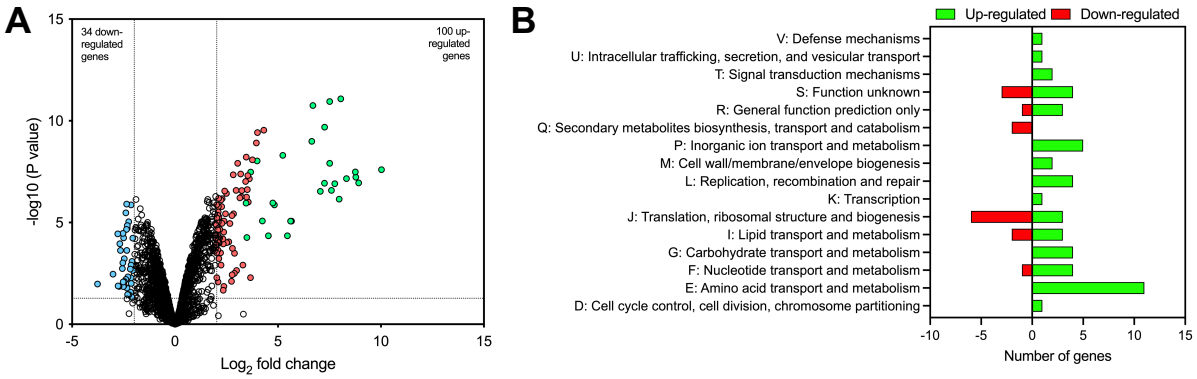

Figure 5. Measuring the *S. mutans* transcriptome response to *erfR* repression with CRISPRi-RNAseq. (A) Volcano plot of differential gene expression of *S. mutans* in a CRISPRi strain targeting *erfR* (with and without xylose *dcas9* induction). The expression difference was considered significant for a log2 fold change > 2 (vertical dotted line) and p value < 0.05 (horizontal dotted line). TnSmu1 genes are shown as light green dots. (B) The number of genes up- or downregulated with a Cluster of Orthologous Group (COG) assignment is indicated in a bar chart.

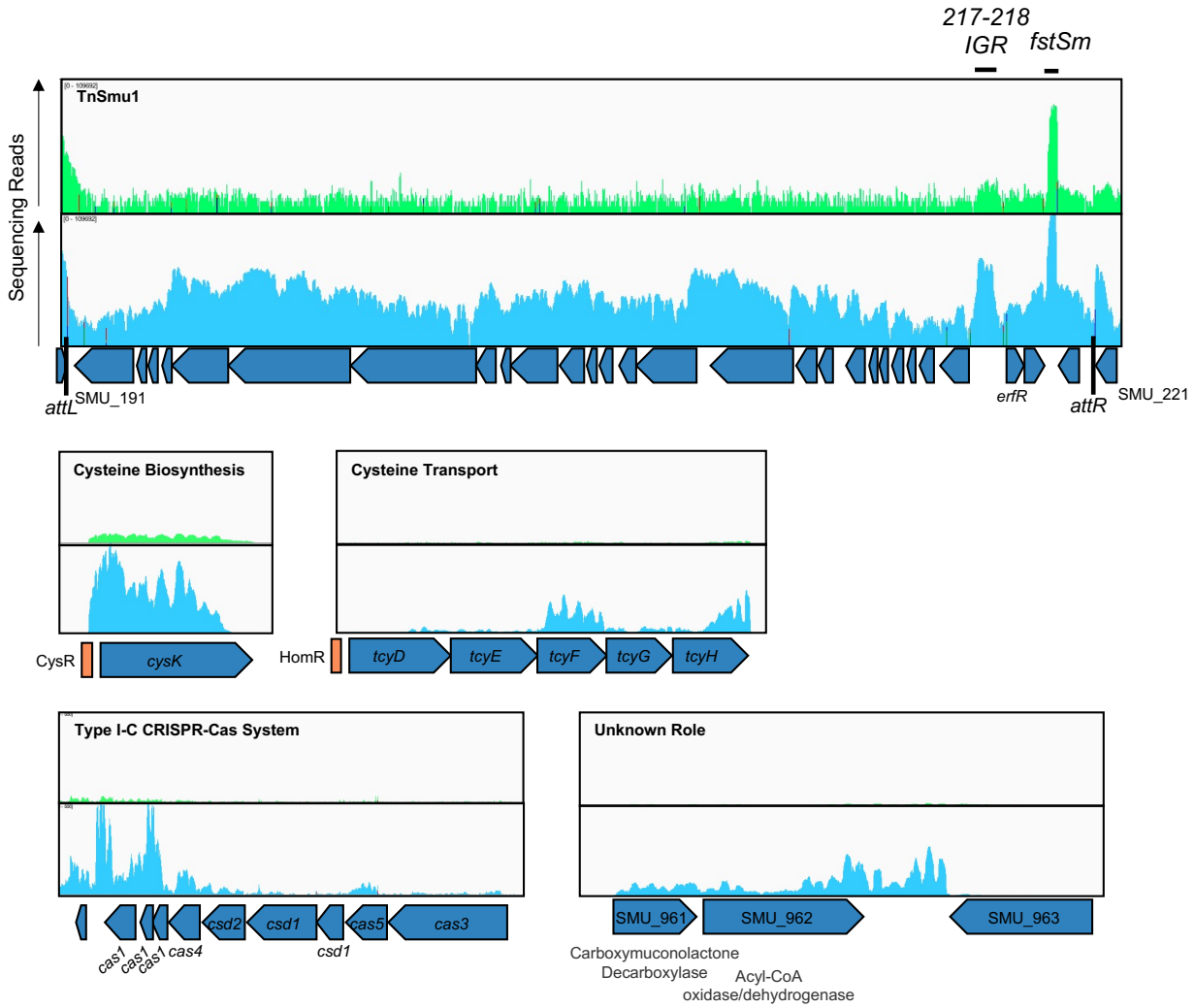

Figure 6. High-resolution transcriptome maps of selected differentially regulated regions. RNAseq reads were visualized using the Integrated Genome Viewer, with the control sample shown in green, and the *erfR*-targeting sample shown in blue. Several regions are highlighted, with the top panel depicting read coverage across the TnSmu1 mobile element. The *fstSm* and 217-218 intergenic region (IGR) small RNAs are highlighted. In addition to TnSmu1 read visualization, reads are also shown for *cysK*, a cysteine transport operon (*tcyDEFGH*), the Type I-C CRISPR-Cas system and an operon with an unknown role. For each region, genes are annotated, and orange boxes are shown for potential promoter regulator sites.

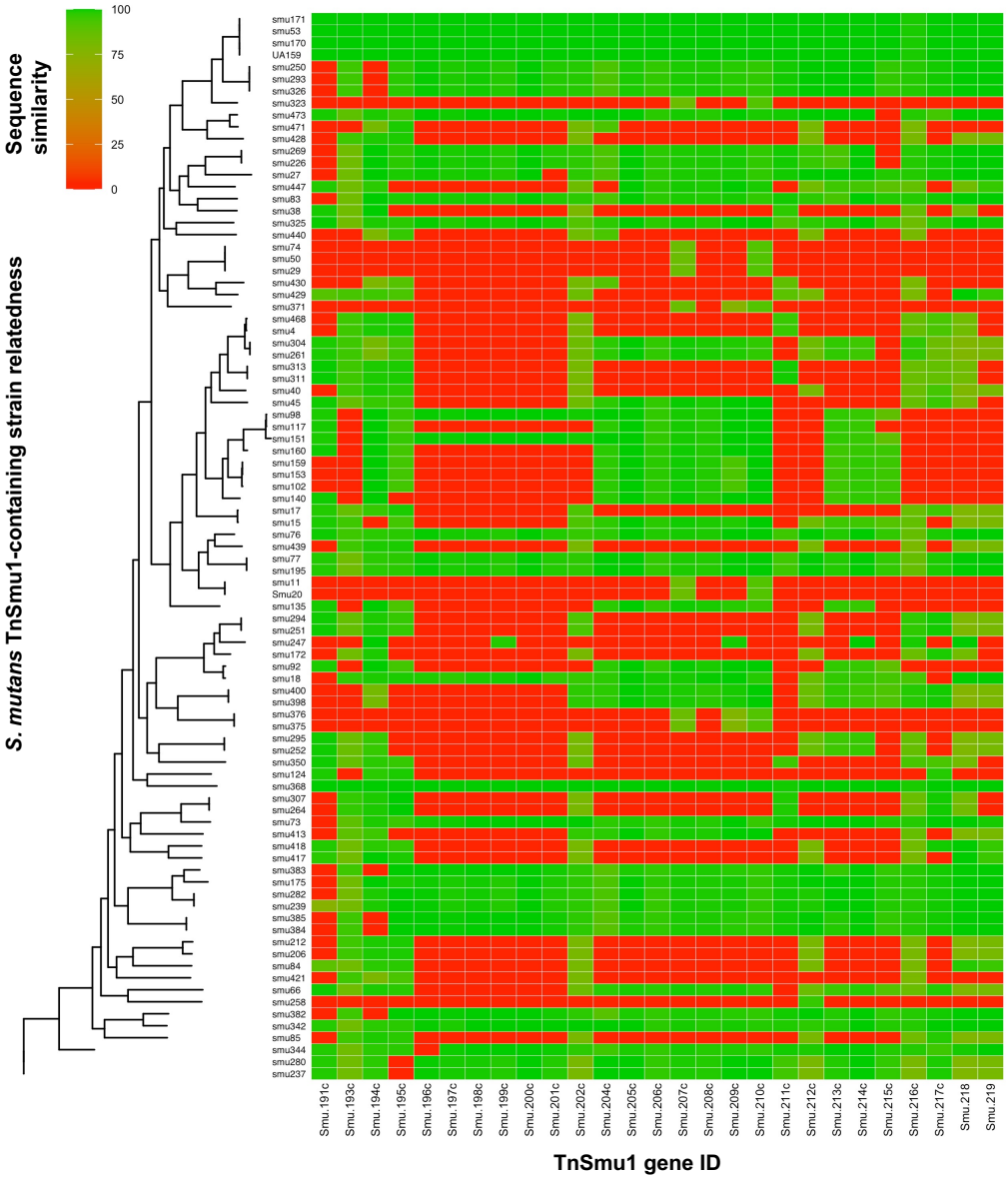

Figure 7. TnSmu1 gene distribution across sequenced *S. mutans* genomes. *S. mutans* strains were searched for TnSmu1 genes, comparing to those carried on *S. mutans* UA159. Strains are organized by strain relatedness (*y*-axis) and gene ID is shown on the *x*-axis (SMU_191c to SMU_219). The percent sequence similarity for each gene is shown (compared to the gene on *S. mutans* UA159 TnSmu1). Red areas depict a lack of a TnSmu1 gene on a strain genome.