Toxin-Antitoxin Gene Pairs Found in Tn3 Family Transposons Appear To Be an Integral Part of the Transposition Module

Lima-Mendez, Gipsi; Alvarenga, Danillo Oliveira; Ross, Karen; Hallet, Bernard; Melderen, Laurence Van; Varani, Alessandro M.; Chandler, Michael

Published in:
mBio

DOI:
10.1128/mbio.00452-20

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Toxin-Antitoxin Gene Pairs Found in Tn3 Family Transposons Appear To Be an Integral Part of the Transposition Module

Gipsi Lima-Mendez, Danillo Oliveira Alvarenga, Karen Ross, Bernard Hallet, Laurence Van Melderen, Alessandro M. Varani, Michael Chandler

ABSTRACT Much of the diversity of prokaryotic genomes is contributed by the tightly controlled recombination activity of transposons (Tns). The Tn3 family is arguably one of the most widespread transposon families. Members carry a large range of passenger genes incorporated into their structures. Family members undergo replicative transposition using a DDE transposase to generate a cointegrate structure which is then resolved by site-specific recombination between specific DNA sequences (res) on each of the two Tn copies in the cointegrate. These sites also carry promoters controlling expression of the recombinase and transposase. We report here that a number of Tn3 members encode a type II toxin-antitoxin (TA) system, typically composed of a stable toxin and a labile antitoxin that binds the toxin and inhibits its lethal activity. This system serves to improve plasmid maintenance in a bacterial population and, until recently, was believed to be associated with bacterial persistence. At least six different TA gene pairs are associated with various Tn3 members. Our data suggest that several independent acquisition events have occurred. In contrast to most Tn3 family passenger genes, which are generally located away from the transposition module, the TA gene pairs abut the res site upstream of the resolvase genes. Although their role when part of Tn3 family transposons is unclear, this finding suggests a potential role for the embedded TA in stabilizing the associated transposon with the possibility that TA expression is coupled to expression of transposase and resolvase during the transposition process itself.

IMPORTANCE Transposable elements (TEs) are important in genetic diversification due to their recombination properties and their ability to promote horizontal gene transfer. Over the last decades, much effort has been made to understand TE transposition mechanisms and their impact on prokaryotic genomes. For example, the Tn3 family is ubiquitous in bacteria, molding their host genomes by the paste-and-copy mechanism. In addition to the transposition module, Tn3 members often carry additional passenger genes (e.g., conferring antibiotic or heavy metal resistance and virulence), and three were previously known to carry a toxin-antitoxin (TA) system often associated with plasmid maintenance; however, the role of TA systems within the Tn3 family is unknown. The genetic context of TA systems in Tn3 members suggests that they may play a regulatory role in ensuring stable invasion of these Tns during transposition.

KEYWORDS antitoxin, Tn3 family, toxin, transposition

Members of the Tn3 transposon (Tn) family form a tightly knit group having related transposase genes and related DNA sequences at their ends. However, they are highly diverse in the range of passenger genes that they carry (see reference 1) (Fig. 1).
The basic Tn3 family transposition module is composed of transposase and resolvase genes and two ends with related terminal inverted repeat DNA sequences, the IRs, of 38 to 40 bp or sometimes even longer (2). They encode a large (~1,000-amino-acid [aa]) DDE transposase, TnpA, significantly longer than the DDE transposases normally associated with insertion sequences (IS) (see reference 3). The TnpA transposase catalyzes DNA cleavage and strand transfer reactions necessary for formation of a cointegrate transposition intermediate during replicative transposition (4). The cointegrate is composed of donor (with the transposon) and target (without the transposon) circular DNA molecules fused into a single circular molecule and separated by two directly repeated transposon copies, one at each donor-target junction (4). Phylogenetic analysis based on TnpA sequence identified 7 clusters or subgroups named after representative transposons: Tn3, Tn21, Tn163, IS1071, IS3000, Tn4430, and Tn4651 (1, 5). A second feature of members of this transposon family is that they carry short internal (100- to 150-bp) DNA segments, at which site-specific recombination between each of the two Tn copies occurs to “resolve” the cointegrate into individual copies of the transposon donor and the target molecules each containing a single transposon copy (1).

This highly efficient recombination system is ensured by a transposon-specified site-specific recombinase: the resolvase. There are at present three known major resolvase types, TnpR, TnpI, and TnpT (Fig. 1), distinguished, among other
features, by the catalytic nucleophile involved in DNA phosphate bond cleavage and
rejoining during recombination. TnpR is a classic serine (S)-site-specific recombinase
(e.g., reference 6); TnpI is a tyrosine (Y) recombinase (7) (see reference 1); and
TnpS/H11001 TnpT is a heteromeric resolvase combining a tyrosine recombinase, TnpS, and a
divergently expressed helper protein, TnpT, with no apparent homology to other
proteins (8, 9). The tnpR gene can be in either the same orientation or opposite
orientation as tnpA. In the former case, the res site lies upstream of tnpR and, in the
latter case, between the divergent tnpR and tnpA genes. For relatives encoding TnpS
and TnpT, the corresponding genes are divergent and the res (rst) site lies between tnpS
and tnpT. Examples of these architectures are shown in Fig. 1. Each res includes a
number of short DNA subsequences which are recognized and bound by the cognate
resolvases. These are different for different resolvase systems and called
res (for resolution site), IRS (10) or irs (for internal recombination site [11]), or rst (for resolution site tnpS tnpT) (8) (see below). res sites that have been analyzed also include promoters
that drive both transposase and resolvase expression (see references 1, 10, and 12).
Indeed, TnpR from Tn3 itself was originally named for its ability to repress transposase
expression by binding to these sites (13, 14).

The diversity of these Tns resides in the variety of other mobile elements that have
been incorporated into their structures such as IS and integrons, as well as other Tn3
family members (see reference 1) and of their passenger genes. The most notorious of
these passenger genes are those for antibiotic and heavy metal resistance, although
other genes involved in virulence functions for both animals and plants (e.g., Fig. 1) or
in organic catabolite degradation also form part of the Tn3 family passenger gene
arsenal.

While studying Xanthomonas citri, a principal pathogen of citrus trees and an
important economic problem (e.g., reference 15), we had identified a number of Tn3
family structures in pXac64, a conjugative plasmid carrying a variety of pathogenicity
and virulence genes (Fig. 2) (2, 16). An interesting observation was that one of the
Tn3-related transposons, TnXc4, carries a toxin-antitoxin (TA) system belonging to the
type II TA class (17).

Type II TA systems are generally composed of 2 proteins: a stable toxin and a labile
antitoxin that binds the toxin and inhibits its lethal activity (see reference 18). The
antitoxin includes a DNA binding domain involved in promoter binding and negative
regulation of TA expression. They are involved in plasmid maintenance in growing
bacterial populations by a mechanism known as postsegregational killing. Upon plas-
mid loss, degradation of the labile antitoxin liberates the toxin from the inactive
complex, which in turn is free to interact with its target and cause cell death. Recently,
the Eva Top laboratory (19), while studying plasmid maintenance, observed that a
relatively unstable plasmid, pMS0506, could be stabilized by transposition of a 7.1-kb
Tn3-related transposon, Tn6231, from the non-self-transmissible plasmid pR28 (20)
indigenous to Pseudomonas moraviensis. Further analysis revealed that Tn6231 (which

FIG 2 An annotated map of plasmid pXAC64 (accession number CP004400) from Xanthomonas citri. The figure shows a section of the plasmid carrying four
Tn3 derivative transposons and one insertion sequence, ISXac2. Derivatives located to the left and right are minimal insertion cassettes (MICs) (56), which are
devoid of transposition genes. These include TALEs (transcription activator-like elements) (yellow) responsible for pathogenicity, which, in turn, include an array
of peptide repeats (DNA-binding domain in gray). Between the flanking MICs are two complete Tn3 family transposons. TnXax1 carries a tnpS/T resolvase with an
intervening res (rst) recombination site (green) and two genes (yellow) involved in plant-pathogen interaction. TnXc4 includes a toxin-antitoxin gene pair,
in orange, and divergent tnpA and tnpR genes with an intervening res site (green). Coordinates in base pairs are shown on the line at top.
is reported to be 99% identical to Tn4662 (19)) also carried a type II TA gene pair that presumably stabilized the target plasmid.

Although the presence of TA systems in Tn3 family transposons had been noted previously (21, 22) (see reference 1), neither the function of these systems within the Tn3 family nor their genetic context has been examined. These initial observations prompted us to investigate whether TA systems have been acquired by other Tn3 family members in a similar way and to examine their possible involvement in Tn behavior.

RESULTS

Identification of TA gene pairs in Tn3 family members. As a first step, we undertook a detailed annotation of available Tn3 family members in the ISfinder database (23) and also those listed in the work of Nicolas et al. (1). We also searched NCBI for previously annotated Tn3 family members (March 2018) and made use of an in-house script which searches for tnpA, tnpR, and TA genes located in proximity to each other (Tn3finder, https://ncbi.nlm.nih.gov/TnFinder; Tn3+TA_finder, https://github.com/danillo-alvarenga/tn3-ta_finder) to search complete bacterial genomes in the RefSeq database at NCBI. Of 190 Tn3 family transposons for which relatively complete sequence data (transposase, resolvase, and generally both IRS) were available, 39 carry TA systems (Fig. 3, colored squares, and Table 1; see also Table S1 in the supplemental material). A phylogenetic tree based on similarity between the tnpA gene products is shown in Fig. 3. Note that, with minor exceptions, the entire Tn3 library conforms closely to the previously defined Tn3 family subgroups (1).

The majority of Tn3 family members encode a TnpR resolvase (Fig. 3, purple circles), although several members of the Tn163 subgroup carry the TnpS/H11001 TnpT resolvase (Fig. 3, pink circles). Only three derivatives, Tn5401, TnBth4, and Tn4430, encode the TnpI resolvase (Fig. 3, salmon circles).

The Tn-carrying TA systems featured examples from all known combinations and orientations of transposase and resolvase genes (Fig. 4). While most cases occurred in Tn3 family members with tnpR resolvase genes, examples were also identified in transposons with tnpS+tnpT (TnPosp1_p and TnHdN1.1) and tnpI (Tn5401 and TnBth4) genes (Fig. 3). Illustrative examples are shown in Fig. 4.

A diversity of TA types. We examined the diversity of the TA modules associated with Tn3 transposons by comparison of the TA protein sequences with the Pfam database using hmmscan from the HMMER suite (24). Candidates with no Pfam match were searched against the PDB_mmCIF70 database (PDB filtered at 70% sequence identity) using HHsearch, a tool for protein remote homology detection based on profile-to-profile comparison (25). In total, 5 toxin families (RelE/ParE, Gp49, PIN_3, PIN, and HEPN) and 6 antitoxin families (ParD, HTH_37, RHH_6, Phd/YefM, AbrB/MazE, and MNT) were identified (Fig. 3, Table 1, and Table S1). All of these toxin families except ParE have been associated with RNase activity, either experimentally or by sequence similarity (26), while ParE inhibits gyrase activity by an unknown molecular mechanism (27). The majority of examples were found in two Tn3 subgroups, Tn3 (2 toxin families; 12 PIN_3 and 2 ParE) and Tn3000 (3 toxin families; 13 ParE, 5 Gp49, and 1 PIN), while 6 members of 5 different toxin families (ParE, Gp49, PIN_3, PIN, and HEPN) and 6 antitoxin families (ParD, HTH_37, RHH_6, Phd/YefM, AbrB/MazE, and MNT) were identified (Fig. 3, Table 1, and Table S1). All of these toxin families except ParE have been associated with RNase activity, either experimentally or by sequence similarity (26), while ParE inhibits gyrase activity by an unknown molecular mechanism (27).

In general, the TA genes are arranged with the antitoxin located upstream of the toxin gene. However, TA systems in reverse order in which the toxin gene precedes that of the antitoxin have been described in the literature (18). Among the 39 TA systems associated with the Tn3 family transposons, five members of the Tn3000 subgroup (Table 1 and Fig. 3) carried TA systems in which the toxin gene precedes that of the antitoxin. These systems are composed of a Gp49 (PF05973)-type toxin (T) of the
RelE/ParE superfamily and an HTH\_37 (PF13744)-type antitoxin (A) of the HigA superfamily. These all have the configuration /H11021 \( \text{tnpR} \) /H11022 \( \text{tnpA} \) (where the arrowheads point in the direction of transcription) (18, 28). A similar situation is found in the unrelated Tn4651 subgroup member TnPosp1\_p, which has the configuration T > A > \( \text{tnpR} \) > \( \text{tnpA} \).

**FIG 3** A phylogenetic tree of 190 Tn3 family members based on their TnpA sequences. We extracted Tn3 family members from the ISfinder database which served to generate the subgroups defined in the work of Nicolas et al. (1). Many others were drawn from the literature and have been given official names (Tn followed by digits, e.g., Tn1234; https://transposon.istmed.ac.uk/tn-registry), while others were identified using Tn3\_finder software (TnCentral, https://tncentral.proteininformationresource.org/TnFinder.html) and given temporary names. Each is associated with its GenBank accession number; the GenBank file contains either the extracted transposon or the DNA sequence from which it was extracted (e.g., DNA fragment, plasmid, or chromosome). Numbers above the lines of each clade indicate the maximum likelihood bootstrap values. The subgroups adhere closely to those defined by Nicolas et al. (1) with some minor variations resulting from the significantly larger Tn sample. The majority of members carry \( \text{tnpR} \), serine resolvases (purple circles). Those that include \( \text{tnpI} \) or \( \text{tnpT} / \text{tnpS} \) are indicated by salmon and pink circles, respectively. The TA gene pairs are indicated by colored squares. Note that Tn5501.5 carries a mutation which truncates its toxin gene, leaving the antitoxin intact. The outer squares represent the toxin, and the inner squares represent the antitoxin. The five toxin types are Gp49 (PF05973), purple; PIN\_3 (PF13470), dark green; PIN (PF01850), bright blue; ParE (PF05016), yellow; and HEPN, black. The antitoxins are HTH\_37 (PF13744), orange; RHH\_6 (PF16762), blue; PhdYeFM\_at (PF02604), magenta; RelB/ParD/CcdA/DinJ, dark gray; AbrB/MazE, light gray; and MNT, bright green. The corresponding Tn names and accession numbers are highlighted in bold for clarity. Note that the branches have been extended for clarity.
<table>
<thead>
<tr>
<th>TOXIN FAMILY</th>
<th>CROSS-REFERENCE</th>
<th>TOXIN SUPERFAMILY</th>
<th>ANTITOXIN</th>
<th>CROSS-REFERENCE</th>
<th>ANTITOXIN SUPERFAMILY</th>
<th>RESOLVASE</th>
<th>TNP A SUBGROUP</th>
<th>TNP A TRANSPOSITION MODULE CONFIGURATION</th>
<th>TRANSPON NAME (ACCESSION NO.)</th>
<th>Nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>ParE</td>
<td>PF05016</td>
<td>RelE/ParE</td>
<td>ParD</td>
<td>4Q2U_C</td>
<td>RelB/ParD</td>
<td>R</td>
<td>Tn3000</td>
<td>&lt;T &lt;A trpR&gt; tnpA&gt;</td>
<td>Tn5501 (JN648090.1), Tn5501.1 (CP016447.1), Tn5501.3 (GQ883559.1), Tn5501.4 (KY20693.21), Tn5501.5 (EF628291), Tn5501.6 (MF487840.1), Tn5501.7 (CP021651.1), Tn5501.8 (KC771559), Tn5501.9 (AJB63570.1), Tn5501.10 (MM053445), Tn5501.11 (CP002451)</td>
<td>11</td>
</tr>
<tr>
<td>ParE</td>
<td>PF05016</td>
<td>RelE/ParE</td>
<td>ParD</td>
<td>4Q2U_C</td>
<td>RelB/ParD</td>
<td>R</td>
<td>Tn4651</td>
<td>&lt;T &lt;A trpR&gt; tnpA&gt;</td>
<td>TnTsp1 (NC_014154)</td>
<td>1</td>
</tr>
<tr>
<td>ParE</td>
<td>PF05016</td>
<td>RelE/ParE</td>
<td>ParD-noDBD</td>
<td>4Q2U_C</td>
<td>ParD-noDBD</td>
<td>I</td>
<td>Tn3</td>
<td>&lt;T &lt;A resI trpL&gt; tnpA&gt;</td>
<td>TnPth4 (NZ_CP010092.1), TnS401 (U03554.1)</td>
<td>2</td>
</tr>
<tr>
<td>ParE</td>
<td>PF05016</td>
<td>RelE/ParE</td>
<td>Phd/YeFM</td>
<td>PF02604</td>
<td>Phd/yeFM</td>
<td>R</td>
<td>Tn3000</td>
<td>&lt;trpR A &gt; T&gt; tnpA&gt;</td>
<td>TnSpu1 (CP017991.1), TnAbapMCR4.3 (CP033872)</td>
<td>2</td>
</tr>
<tr>
<td>PIN_3</td>
<td>PF13470</td>
<td>PIN</td>
<td>RHH_6</td>
<td>PF16762</td>
<td>VapB</td>
<td>R</td>
<td>Tn3</td>
<td>&lt;trpR A &gt; T&gt; tnpA&gt;</td>
<td>Tn1412 (L36547), Tn5044 (Y17691.1), Tn5046 (Y18360.1), Tn5563a (K920395.1), Tn5563a.1 (CP028162), Tn5563a.2 (CP028567), TnThsp9 (FP475957.1), TncX4 (CP009309.1), TncX5 (Z73593), TncX4.1 (CP011958.1), TncX4.2 (CP023287), TncX4.3 (CP020887)</td>
<td>12</td>
</tr>
<tr>
<td>PIN_3</td>
<td>PF13470</td>
<td>PIN</td>
<td>RHH_6</td>
<td>PF16762</td>
<td>VapB</td>
<td>S/T</td>
<td>Tn4651</td>
<td>&lt;trpT trpS A &gt; T&gt; tnpA&gt;</td>
<td>TnHdn1 (FP929140)</td>
<td>1</td>
</tr>
<tr>
<td>Gp49</td>
<td>PF05973</td>
<td>RelE/ParE</td>
<td>HTH_37</td>
<td>PF13744</td>
<td>HigA</td>
<td>R</td>
<td>Tn3000</td>
<td>&lt;A &lt;T trpR&gt; tnpA&gt;</td>
<td>Tn4662a (NC_014124.1), Tn4662a 1 (AY631462.1), TnDsu1 (NC_016616.1), TnPuPupPGH1 (Y09450.1), Tn5501.12 (CP017294.1)</td>
<td>5</td>
</tr>
<tr>
<td>Gp49</td>
<td>PF05973</td>
<td>RelE/ParE</td>
<td>HTH</td>
<td>STN0_A</td>
<td>HigA</td>
<td>R</td>
<td>Tn4651</td>
<td>T&gt; A &lt;trpR&gt; tnpA&gt;</td>
<td>TnPsp1 (CP000316.1)</td>
<td>1</td>
</tr>
<tr>
<td>PIN</td>
<td>PF01850</td>
<td>PIN</td>
<td>Phd/yeFM</td>
<td>PF02604</td>
<td>Phd/yeFM</td>
<td>R</td>
<td>Tn163</td>
<td>T&gt; A &lt;trpR&gt; tnpA&gt;</td>
<td>TnSkul (CP003258.1)</td>
<td>1</td>
</tr>
<tr>
<td>PIN</td>
<td>PF01850</td>
<td>PIN</td>
<td>AbrB</td>
<td>2MRN_A</td>
<td>AbrB/Maze</td>
<td>R</td>
<td>Tn3000</td>
<td>&lt;A &lt;X &lt;trpR&gt; tnpA&gt;</td>
<td>TnAmu2 (NC_015188)</td>
<td>1</td>
</tr>
<tr>
<td>HEPN</td>
<td>SYEP_B</td>
<td>HEPN</td>
<td>Mnt</td>
<td>SYEP_A</td>
<td>Mnt</td>
<td>R</td>
<td>Tn21</td>
<td>&lt;trpR &lt;A &lt;A tnpA&gt;</td>
<td>TnPsy4 (KCX09060.1)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnSou9 (NC_004349)</td>
<td>1</td>
</tr>
</tbody>
</table>

4 Considered or PDB identifiers.
5 Associated resolvase type.
6 Associated Tn3 family subgroup.
7 TA module arrangement with respect to trpR and trpA (arrows represent the direction of gene expression). resI indicates the location of the res site.
8 Tn5501.5 carries a mutation which truncates its toxin gene, the antitoxin is intact.
9 Nb, number of transposons associated with the TA type.
Two additional members of the Tn3 subgroup, Tn5401 and TnBth4, both with the configuration \( \text{tnpI} \rightarrow \text{tnpA} \), carried a different TA system, a ParE toxin (Pfam: PF05016) and a ParD antitoxin, which appears to lack the DNA-binding domain. Among the TA gene pairs found in members of the Tn3000 transposon subgroup, a majority of toxins are of the ParE (Pfam: PF05016) type while the potential antitoxins have no Pfam match. Results from HHpred indicate that these are antitoxins have an RHH fold similar to that of the classical ParD antitoxin (PDB accession no. 4Q2U_C). It should be noted that this subgroup of Tns (Tn5501 and its derivatives) are highly related and differ mainly by the passenger genes they carry. With the exception of Tn5501.12 (discussed below), all Tn5501 derivatives have identical or nearly identical toxin and antitoxin protein sequences. They were all identified from the nonredundant NCBI nucleotide database using Tn5501 as query sequence. Interestingly, a single member of the Tn4651 subgroup, TnTsp1, also encodes a nearly identical ParE (identical protein sequence)-ParD (2 amino acid substitutions) TA pair.

**Acquisition and exchange of TA modules.** A relevant question is whether these TA modules (Table 1) were acquired once or multiple times during evolution. This question was addressed by phylogenetic analyses of Tn3-associated toxins assigned to the same Pfam group, along with the seed sequences used to build the Pfam hidden Markov model (HMM) (Fig. 5). If toxins share a recent common ancestor, there are two possible explanations. In cases where the TA module is found in related transposons (with
FIG 5 (Continued)
Toxin-Antitoxin Gene Pairs in Tn3 Family Transposons

C) Gp49

D) PIN

FIG 5 (Continued)
similar tnpA and/or resolvase genes), it is likely that it was first acquired by a transposon that subsequently diverged. Alternatively, for transposons which are generally not related (different tnpA family, different resolvase) but which harbor TA modules that are similar at the DNA level, it is likely that the TA module was acquired by recombination with another transposon. Tn3 toxins sharing their most common ancestor with non-Tn3 toxins are likely to have been acquired independently. The different TA modules identified and the Tn3 family members in which they are found are described below.

(i) ParE. There are 16 parE toxin genes in our collection; 14 are paired with a parD antitoxin gene and two are paired with a phd/yeFM antitoxin gene (Table 1; Fig. 5A). The 14 parE-parD modules are found in three transposon subfamilies, Tn3000 (11 examples), Tn3 (two examples), and Tn4651 (one example), suggesting that the parD-parE operon has been acquired three independent times in this collection (Fig. 5A).

The first acquisition event concerns the 11 parE-parD Tn3000 subgroup Tns, which are all Tn5501 relatives and have identical or nearly identical toxin protein sequences (Fig. 5A) and significant similarity at the DNA level within the TA modules.

One Tn5501 derivative, Tn5501.12, is an exception since it carries a gp49-HTH TA module and is described further below (Fig. S1B and D).

The parE-parD module located in TnTsp1 (Tn4651 subfamily) is identical at the protein level and nearly identical at DNA level (95%) with the 10 parE-parD modules of the set of Tn5501 derivatives (Fig. S1D), indicating that the parD-parE operon might have been acquired via recombination between a Tn5501 relative which contributes the DNA segment to the left and an unidentified transposon which contributed the DNA segment to the right to generate TnTsp1. The recombination point is likely to be at, or close to, the res site, where the homology between TnTsp1 res subsite I and Tn5501 res subsite I breaks down (Fig. 6A). Resolvase-mediated recombination probably occurs at the dAdT dinucleotide indicated in red in Fig. 6A (see references 29 and 30).

The second event is illustrated by TnAbapMCR4.3 (31) and TnSpu1 (Tn3000 subgroup) as indicated by the phylogenetic analysis (Fig. 5A). Moreover, they are coupled with a different antitoxin, from the Phd/YefM family (Table 1 and Fig. 3). Identity between the transposition modules of these two transposons is high in (including the TA genes), but they differ in their passenger genes.

The third event is represented by two examples, TnBth4 and Tn5401 (83% identity at the nucleotide level) in the Tn3 subgroup (Fig. 5A). Both carry a TnpI resolvase, and they do not have significant similarity with any of the other transposons carrying the parE-parD TA module (Fig. S1D).

(ii) PIN_3. The PIN_3 toxin domain is represented 13 times (8 unique sequences) in our collection. This domain is associated with an RH_6 antitoxin. Twelve Tns having this module are in the Tn3 subgroup, and one belongs to the Tn4651 subgroup (TnHdN1.1).

Interestingly, although TnHdN1.1 does not share significant sequence similarity at the nucleotide level with the Tn3 subgroup Tns featuring the same TA module, at the protein level both toxin and antitoxin from TnHdN1.1 share ~80% sequence identity with those in the Tn3 subgroup (Fig. 5B). Phylelogenetic analyses of all 13 PIN_3 toxins with the seed proteins used to build the corresponding PFAM family suggest that these

![Phylogenetic trees of toxin genes. The phylogenetic history of the transposon-associated toxins reconstructed along the corresponding relative seed proteins downloaded from the Pfam database. The position of transposon-associated toxins is indicated by an outlined colored background indicating the subgroup to which they belong, as in Fig. 3. Red dots indicate the tips where one toxin sequence was chosen as representative of a set of identical toxins (i.e., there are several Tn examples in the collection). (A) ParE. The phylogeny of the ParE toxins suggests that ParE has been recruited 3 times by Tn3s: twice by the Tn3000 subgroup and once by the Tn3 subgroup. Toxin sequences of Tn5501, Tn5501.1, Tn5501.2, Tn5501.3, Tn5501.4, Tn5501.6, Tn5501.7, Tn5501.8, Tn5501.9, Tn5501.10, and TnTsp1 are identical, which indicates that the last acquired the TA module by recombination with a Tn5501 ancestor. (B) PIN_3. The phylogeny of PIN_3 toxins suggests that this gene has been recruited once by the Tn3 subgroup and further recombined into an ancestor of TnHdN1.1. Toxin sequences of TnXc4, TnXc4.1, TnXc4.2, and TnXc4.3 are identical. Toxin sequences of Tn5563a, Tn5563a.1, and Tn5563a.2 are identical. (C) Gp49. The phylogeny of Gp49 toxins indicates that these have been recruited in 3 different events by the Tn3s: by the Tn3000 subgroup, the Tn4651 subgroup, and the Tn4651 subgroup. Toxin sequences of Tn4651.20 and Tn5501.12 are identical. (D) PIN. The phylogeny of the PIN toxin suggests that this toxin has been recruited in two separate events.](http://mbio.asm.org/)
proteins were recruited by a transposon in a single event and that an ancestor of TnHd1.1 acquired the TA module via recombination (Fig. 5B).

(iii) Gp49. Seven TA modules are composed of a Gp49 toxin and an HTH antitoxin. Phylogenetic analyses suggest that this TA module has been recruited on three occasions (Fig. 5C).

**FIG 6** Intertansposon recombination at the res site exchanges TA modules. The symbols are the same as defined in Fig. 1, 2, and 4. (A) Comparison of Tn5501 accession no. JH648090.1 and TnTsp1 accession no. NC_014154 showing a possible recombination point between the two Tns where exchange at the TA gene pair may have occurred. The bottom section shows the region of TnTsp1, including the TA gene module (orange), the res site (green), and tnpB and part of tnpA (purple). The top segment shows the equivalent map of Tn5501. Below is shown a DNA sequence alignment (magenta) with the equivalent region of Tn5501. The two transposons have similar DNA sequences to the left of res site I. The level of sequence identity is reduced in tnpR and is insignificant in tnpA. The res site I sequences (green) are shown between the two panels, and the AT dinucleotide at which recombination probably occurs is indicated in red. Sequence nonidentities are underlined. The two sequences are identical up to the probable recombination site and show some diversity to its right. (B) The region of Tn5501.12 (accession no. CP017294.1) showing the 5’ end of the tnpA gene, the tnpR gene, a res site typical of the tnpR res sites, and toxin-antitoxin gene pair (note that the toxin gene is upstream of the antitoxin gene [Table 1]). The horizontal magenta lines at the bottom show the alignment of Tn5501.12 with Tn5501 (accession no. JH648090.1) and Tn4662a (NC_014124.1). The right half of Tn5501 is clearly highly homologous to the right side of Tn5501.12 whereas the left side of Tn4662a is homologous to the left side of Tn5501.12. The DNA sequences at the top show the res subsite I (green) with the dinucleotide at which recombination should occur in red together with flanking sequences. Underlined bases indicate regions of nucleotide identity. This suggests a scenario in which Tn5501.12 was generated by recombination at res site I between transposons similar to Tn5501 and Tn4662a.
The first was acquisition by the Tn3000 subgroup transposons TnDsU1, TnPpupPGH1, Tn4662a, Tn4662a.1, and Tn5501.12. Tn5501.12 is the only relative of Tn5501 to have this TA module. Sequence comparisons suggest that Tn5501.12 resulted from exchange of the entire left end of transposon Tn5501 with a transposon very similar to Tn4662a (Fig. 6B) carrying a Gp49/HTH_37 TA module. The DNA sequence in this region indicates that recombination between the two transposons occurred at a sequence which resembles res site I containing the dinucleotide (in red) at which recombination takes place during the cointegrate resolution step of transposition (29, 30). This mechanism is similar to that proposed for acquisition of the parE-parD module by the Tn4651 subgroup Tn TnPsp1 as described above. The second acquisition concerns TnPosp1 from the Tn4651 subgroup. The TnPosp1 toxin is ~60% identical at the protein level to those of the Tn3000 subgroup, but they are not similar at the DNA level (Fig. S1B). The phylogenetic reconstruction (Fig. 5C) indicates that their most recent acquisition event is illustrated by TnSku1 (Tn163 subgroup), whose toxin does not share a recent ancestor with those of the Tn3000 group or with the TnPosp1 toxin (Fig. 5C).

(iv) PIN. There are two transposons in the set with a PIN toxin, which appear phylogenetically distant (Fig. 5D), suggesting these represent two independent acquisitions. Furthermore, TnAmu2 (Tn4430 subgroup) carries a Phd/YefM antitoxin and TnPsp42 has an AbrB/MazE antitoxin.

(v) HEPN. Finally, TnSod9 (Tn21 subgroup), located in the Shewanella oneidensis MR-1 megaplasmid, includes an HEPN (higher eukaryotes and prokaryote nucleotide-binding)-type toxin and an MNT (minimal nucleotidyltransferase) antitoxin as identified by their similarity to another toxin and antitoxin pair (HHpred hit PDB identifier 5YEP) encoded in the chromosome of the same strain (32).

Additional indications of independent TA acquisitions are evidenced by Tn3 derivatives in which the order of genes in the TA module is reversed, i.e., the toxin gene being located upstream of the antitoxin gene. This arrangement is found predominantly in members of the Tn3000 subgroup (Tn4662a, Tn5501.12, TnDsU1_p, TnPpupPGH1, and Tn4662a.1), although single examples are observed in the Tn4651 (TnPosp1_p), Tn21 (TnSod9), and Tn163 (Tn5393.1) subgroups.

The Tn3 family-associated TA passenger gene systems are located in a unique position. In most cases, the TA gene pairs are embedded within the transposition module comprising transposase and resolvase genes and the res site. They are positioned very close to the res sites (Fig. 7). This is in sharp contrast to all other Tn3 family passenger genes, which are generally located away from the resolution and transposon genes and, where known, have often been acquired as integron cassettes or by insertion of other transposons (1). Indeed, several TA-carrying transposons represent closely related derivatives with identical transposase, resolvase, and TA modules but contain different sets of passenger genes (e.g., Tn5501.1 and derivatives 5501.2, 5501.3, 5501.4, etc.).

In the majority of cases (33–37), the Tn3-associated TA gene pairs are located directly upstream of the resolvase genes (tnpR or tnpI) (Fig. S2). There are only three exceptions to this. The first is the single example of a derivative with the TnpS+TnpT resolvase, TnHdN1.1 (Fig. 3), where the TA genes are located between the resolvase tnpS and transposase genes (Fig. 7C). In the second, TnSku1 (not shown), the TA genes are located downstream of and transcribed toward tnpR, and in the third, a partial transposon copy, TnAmu2_p (not shown), there is a short open reading frame (ORF) of unknown function between the divergently transcribed antitoxin and tnpR genes.

Regulation of TA gene expression. Although it is possible that the TA genes are expressed from their own promoter if present, their position might permit expression from native Tn promoter elements. In Tn3, which has been examined in detail, transposase and resolvase gene expression is controlled by promoters found within the res site located between the two divergent genes (Fig. 7A, i). Resolvase binding to these
sites autoregulates both tnpR and tnpA expression (13, 14, 38, 39). The location of the TA genes in proximity to the res sites raises the possibility that their expression is also controlled by these promoters.

Few of the res sites in the collection of TA-associated Tn3 family members have been defined either experimentally or by sequence comparison. We therefore attempted to...
identify potential \textit{res} sites using as a guide the canonical \textit{res}-site organization schematised in the work of Nicolas et al. (1), a \textit{res} site library (kindly provided by Martin Boocock), and RSAT tools (Regulatory Sequence Analysis Tools; \url{http://rsat.sb-roscoff.fr/}) (see Materials and Methods). This analysis resulted in identification of 27 potential \textit{res} sites (Table S2). Their organization is shown in Fig. S2. For transposons with a TnpR resolvase, it is striking that in every single case, the TA genes are located just downstream from \textit{res} site I, whereas \textit{tnpR} is located next to \textit{res} site III. In transposons with divergent \textit{tnpA} and \textit{tnpR} such as TnXc5 and Tn5563a (Fig. S2, expanded in Fig. 7A, ii and iii), the \textit{tnpA} and \textit{tnpR} genes and \textit{res} sites are organized similarly to those of Tn3, which does not carry TA (Fig. 7A, i), except that \textit{tnpA} is separated from \textit{res} by the intervening TA genes. This organization is also similar in Tn3 members in which \textit{tnpA} is downstream of \textit{tnpR} and in the same orientation (e.g., Tn5501 and Tn5sp1 [Fig. S2]).

Promoters have been defined in the \textit{res} (\textit{irs}) site of the \textit{tnpl}-carrying Tn5401 (5, 12), and \textit{tnpl} and \textit{tnpA} expression is modulated by TnpI binding to the \textit{res} site (12) (Fig. 7B, i). The other \textit{tnpl}-carrying transposon with TA genes, TnBth4 (Fig. 7B, ii), has an identical \textit{res} site, and therefore expression is probably regulated in the same way. Again, the potential promoters are pertinently located for driving expression of the TA module.

Finally, transposon Tn\textit{HdN1.1} (Fig. 7C) is the only example in our collection of a \textit{tnps+tnpT} transposon carrying a TA module. The \textit{res} site and relevant promoter elements for the divergently expressed \textit{tnps+tnpT} have been identified between the \textit{tnps} and \textit{tnpT} genes in transposon Tn4651 (8, 9). In Tn\textit{HdN1.1}, the TA gene pair is located to the right of \textit{tnps}, between \textit{tnpS} and \textit{tnpA}, and all three genes are oriented in the same direction. Although the exact regulatory arrangement remains to be determined, it seems possible that the promoters in the \textit{res} site regulate expression of the TA gene pair.

These arrangements raise the possibility that some TA gene expression might occur from a \textit{res} promoter and be subject to control by resolvase binding. On the other hand, if the TA genes do carry their own promoters, then these might regulate downstream transposon genes such as \textit{tnpA}. Further experimental studies are necessary to examine the detailed regulation of toxin, antitoxin, and other transposon genes.

**DISCUSSION**

As part of our efforts to build a fully annotated transposon database (TnCentral, \url{https://tncentral.proteininformationresource.org/}), we identified and analyzed 190 Tn3 family transposons (Fig. 3; see also Table S1 in the supplemental material) and have observed that 39 of these include type II TA passenger genes from several distinct families (Fig. 3 and Table 1): 5 toxin families (\textit{ParE}, \textit{Gp49}, PIN\_3, PIN, and HEPN) and 6 antitoxin families (\textit{DinJ}, \textit{HTH\_37}, \textit{RHH\_6}, \textit{Phd/YefM}, \textit{AbrB/MazE}, and MNT). Several lines of evidence suggest that there have been multiple independent TA acquisition events, namely, (i) the transposons in our collection feature different families of toxin and antitoxin pairs (Table 1), (ii) in some cases the TA gene order is inverted, and (iii) we observed proteins with no significant sequence similarity within the same toxin/antitoxin family but predicted to share diverged TA gene pairs. Excluding those cases likely to have arisen from intermolecular Tn recombination (Fig. 6), and Tn\textit{HdN1.1}, which also appears to have acquired the TA via recombination, as indicated by the toxin tree (Fig. 5B), the most parsimonious interpretation of these observations is that the modules were acquired in 10 separate events. These include three for \textit{parE}, two for PIN3, three for \textit{gp49}, two for PIN, and one for HEPN (Table 1). At present, it is unclear how such multiple acquisitions have occurred at the molecular level.

In contrast to other passenger genes in Tn3 family transposons, the TA genes are located at an unusual position within the transposon. They are close to the DNA resolution site (\textit{res}, \textit{irs}, and \textit{rst}) (Fig. 4), and more precisely for those with TnpR resolvases, they consistently neighbor \textit{res} site I (Fig. S2), a DNA sequence which not only probably includes part of a promoter but is the point at which recombination occurs resulting in cointegrate resolution. For those transposons in which the \textit{tnpR} and \textit{tnpA} genes are divergently orientated (Fig. 4A, Fig. 6, and Fig. S2), the TA module is
located between the two genes and expressed in the same direction as \( tnpA \). For those in which \( tnpR \) precedes \( tnpA \) in the same orientation, the TA module lies upstream from \( tnpR \) and is oriented in the opposite orientation (e.g., Table 1, Fig. 7, and Fig. S2). A similar arrangement occurs for the two examples located on 4\( npl \)-carrying transposons. Only a single example of a 4\( nps+tnpT \)-carrying transposon with the TA module was identified, and here, the TA module is located between the resolvase gene pair and the transposase gene.

This location, close to the key enzymes involved in transposition, suggests that the role of the TA pair might not simply be to provide a general addiction system that stabilizes the host replicon, generally a plasmid, carrying the transposon. It seems possible that they play a more intimate role in stabilizing the associated transposon itself. We note, however, that there are two exceptions to this close association of TA genes with the \( Tn \) res site. For \( Tn5ku1 \), the TA genes are located downstream of and expressed toward \( tnpR \), while in the partial copy, \( TnAmu2_p \), there is a short ORF between the divergently transcribed antitoxin and \( tnpR \) genes. This does not appear to be related to the 3-component toxin-antitoxin-chaperone (TAC) systems (33).

Interestingly, type II TA expression, like that of \( tnpA \) and \( tnpR \), is tightly regulated at the transcriptional level. Where analyzed, the toxin-antitoxin complex binds to palindromic sequences located in the operon promoter via the antitoxin DNA-binding domain and acts as a negative transcriptional regulator. This regulation depends critically on the relative levels of toxin and antitoxin in a process known as conditional cooperativity, a common mechanism of transcriptional regulation of prokaryotic type II toxin-antitoxin operons in which, at low toxin/antitoxin ratios, the toxin acts as a corepressor together with the antitoxin. At higher ratios, the toxin behaves as a derepressor. It will be important to determine whether the \( Tn \)-associated TA genes include their indigenous promoters (18, 34).

In the case of \( Tn4631 \) (19), which is 99% identical to \( Tn4662 \) from plasmid \( pDK2 \) (35), the transposon clearly provides a level of stabilization of the host plasmid. This implies that TA expression occurs in the absence of transposition. There are a number of ways in which this could take place (Fig. 8). Expression could occur from a resident TA promoter (Fig. 8A) if present. However, TA expression might be expected to lead to expression of the downstream \( tnpA \) gene by readthrough transcription. Alternatively, in the absence of a TA promoter, TA expression could occur stochastically from the \( res \) promoter (Fig. 8B). However, this does not rule out the possibility that TA expression is regulated at two levels with a low-level “maintenance” expression, resulting in the plasmid stabilization properties described by Loftie-Eaton et al. (19) together with additional expression linked to derepression of the \( tnpA \) (and \( tnpR \)) promoters that must occur during the transposition process (Fig. 8C).

Indeed, regulation of \( tnpR \) and \( tnpA \) by TnpR is a mechanism allowing a burst of TnpA (and TnpR) synthesis, transitorily promoting transposition as the transposon invades a new host. Subsequent repression by newly synthesized TnpR would reduce transposition activity, reinstalling homeostasis once the transposon has been established, a process similar to zygotic induction (36) or plasmid transfer derepression as originally observed for \( ColI \) (37) and subsequently for \( R100 \) (40) and \( R1 \) (41). An alternative but nonexclusive explanation stems from the observation that the \( Tn6231 \) TnpR, in addition to the neighboring TA system, enhances plasmid stability (19). Resolvase systems are known to promote resolution of plasmid dimers (see reference 42), and it was suggested that integration of the TA system into \( Tn6231 \) “such that all the transposon genes shared a single promoter region” permits coordinated TA and TnpR expression and may facilitate temporary inhibition of cell division while resolving the multimers, promoting plasmid persistence. In this light, it is interesting that the \( ccd \) TA system of the \( Escherichia coli \) \( F \) plasmid is in an operon with a resolvase-encoding gene (43, 44).

Expression of the TA module from the \( tnpA/tnpR \) promoter at the time of the transposition burst could transiently increase invasion efficiency (“addiction”) over and above that provided by the endogenous TA regulation system. If the transposon is on a molecule (e.g., a conjugative plasmid) that is unable to replicate vegetatively in the
new host, expression of the TA module without transposition to a stable replicon would lead to loss of the transposon and consequent cell death, whereas cells in which transposition had occurred would survive and give rise to a new population in which all cells would contain the Tn. This might be seen as a “take me or die” mechanism, a notion which could be explored experimentally.

Clearly, there remain a number of important questions about the control of TA gene expression that arise from our in silico analyses and need to be addressed experimentally. These include whether the TA genes include their own promoters and whether expression is controlled by TA-associated promoter elements or by the resident promoters embedded in the res sites. Finally, it is an open question whether resolvase binding to res represses TA expression either from proximal TA promoters or from res-embedded promoters.

MATERIALS AND METHODS

Retrieval of prokaryotic genomes and database building. Nucleotide Fasta files from complete bacterial and archaeal genomes available in the RefSeq database (45, 46) were downloaded on 15 March 2018. Amino acid sequences of type II toxins and their corresponding antitoxins were retrieved from TADB (47, 48), while Tn3 transposases and resolvases were retrieved from the ISfinder database (23) and NCBI GenBank (49). These sequences were compiled into multifasta files to be used as databases in subsequent analyses.

Genomic screening for Tn3 transposons. The complete genomes were compared to the protein sequences from the TADB, ISfinder, and NCBI GenBank databases using tBLASTn 2.2.28 (50) and a custom Python script (Tn3finder available from https://tncentral.proteininformationresource.org/TnFinder.html; Tn3-TA_finder available from https://github.com/danillo-alvarenga/tn3-ta_finder). ORF prediction was performed with Prodigal 2.6.1 (51), and preannotated gbk files were produced with Biopython 1.66 (52). Genomic regions presenting translated protein similarity above 40% and alignment coverage of 60% against Tn3 transposases, resolvases, toxins, and antitoxins within maximum distances of 2,000 bp to each other were retrieved for manual curation.
Manual curation of transposons and accessory genes. Automatic annotations generated by the screening were manually verified in SnapGene Viewer 3.2.1 (GSL Biotech, San Diego, CA). TA gene pairs were evaluated in greater detail by comparisons with the Pfam 32.0 database (53) using hmmcan from the HMMER 3.1b2 suite (24). Remote homologs were searched with HHpred version 3.2.0 (25) against PDB_mmCIF70, a PDB entry filtered at 70% sequence identity. HHpred compares the query HMM to the database of HMMs based on PDB chains and generates query-template alignments. Toxin and antitoxin genes with matches against either Pfam or PDB were associated with the name and identifier (ID) of the Pfam or the PDB ID and known toxins or antitoxins featuring the identified fold (Table 1).

Phylogenetic reconstruction. TnpA protein sequences retrieved from our manually curated data set were aligned with MAFFT 7.309 (54), and their best-fit evolutionary models were predicted with ProTest 3.2.4 (55). A maximum likelihood tree was reconstructed with RaxML 8.2.9 (56) using a bootstrap value of 1,000. The final tree was visualized in FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree) and edited with Inkscape 0.92.4 (http://www.inkscape.org).

To reconstruct the phylogeny of the toxins, we built a nonredundant toxin set, by removing duplicated sequences. Following the classification of the toxin sequences by comparison with the Pfam database, we downloaded the seed protein sequences for each of the PFAM entries that matched the toxins. Protein sequence alignment of the toxins with the corresponding PFAM seed sequences and phylogenetic reconstruction followed the same procedure described above for TnpA proteins.

Sequence comparison between transposons. Transposons were compared all-against-all using blastn. For transposons having a toxin from the same family, all pairwise sequence similarities between the DNA segments comprising the transposase gene, the resolvase, and the TA were visualized as dot plots using flexidot version 1.06 (57) with 10 as wordsize (default).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 1.4 MB.

FIG S2, PDF file, 0.03 MB.

TABLE S1, DOCX file, 0.04 MB.

TABLE S2, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

This work was primarily funded by the Global Emerging Infections Surveillance (GEIS) and Response System (P0020_18_WR; awarded to Michael Chandler and Patrick McGann).

We thank Martin Boocock for providing information concerning the res sites and Jian Zhang (Protein Information Resource PIR, Georgetown University), Hongzhan Huang (Protein Information Resource PIR, University of Delaware), and Cathy Wu (Protein Information Resource PIR, Georgetown University) and Hongzhan Huang for providing essential expert assistance and both conceptual and practical support.

Invaluable contributions were made by Erik Snesrud and Patrick McGann (Walter Reed Army Institute of Research).

REFERENCES


