TnCentral: A Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis

Karen Ross[1§], Alessandro M. Varani[2§], Erik Snesrud[3*], Hongzhan Huang[4], Danillo Oliveira Alvarenga[2], Jian Zhang[1], Cathy Wu[4], Patrick McGann[3], Mick Chandler[5]

[1]Protein Information Resource, Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, Washington, DC, USA.

[2] School of Agricultural and Veterinary Sciences, Universidade Estadual Paulista, Jaboticabal, Sao Paulo, Brazil.

[3]Multidrug-Resistant Organism Repository and Surveillance Network, Walter Reed Army Institute of Research, Silver Spring, Maryland, USA.

[4] Protein Information Resource, Center for Bioinformatics and Computational Biology, University of Delaware, USA

[5]Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, Washington, DC, USA

[§] These authors contributed equally to this work.

*We would like to dedicate this article to the memory of Erik Snesrud who was instrumental in initiating this work but was sadly unable to see it completed.

Keywords: Mobile Genetic Elements, genome evolution, antibiotic resistance, virulence

Importance

The ability of bacteria to undergo rapid evolution and adapt to changing environmental circumstances drives the public health crisis of multiple antibiotic resistance as well as outbreaks of disease in economically important agricultural crops and animal husbandry. Prokaryotic transposable elements (TE) play a critical role in this. Many carry "passenger genes" (not required for the transposition process) conferring resistance to antibiotics or heavy metals or causing disease in plants and animals. Passenger genes are spread by the normal TE transposition activities, by insertion into plasmids which then spread via conjugation within and across bacterial populations. Thus, an understanding of TE composition and transposition mechanisms is key to developing strategies to combat bacterial pathogenesis. Toward this end, we have developed TnCentral, a bioinformatics resource dedicated to describing and exploring the structural and functional features of prokaryotic TE and whose use is intuitive and accessible to users with or without bioinformatics expertise.

Abstract

We describe here the structure and organization of TnCentral (https://tncentral.proteininformationresource.org/), a web resource for prokaryotic transposable elements (TE). TnCentral currently contains ~400 carefully annotated TE, including transposons from the Tn*3,* Tn*7,* Tn*402* and Tn*554* families, compound transposons, integrons and associated insertion sequences (IS). These TE carry passenger genes, including genes conferring resistance to over 25 classes of antibiotics and nine types of heavy metal as well as genes responsible for pathogenesis in plants, toxin/antitoxin gene pairs, transcription factors and genes involved in metabolism. Each TE has its own entry page providing details about its transposition genes, passenger genes, and other sequence features required for transposition as well as a graphical map of all features. TnCentral content can be browsed and queried through text and sequence-based searches with a graphic output. We describe three use cases, which illustrate how the search interface, results tables, and entry pages can be used to explore and compare TEs.

TnCentral also includes downloadable software to facilitate user-driven identification, with manual annotation, of certain types of TE in genomic sequences.

Through the TnCentral homepage, users can also access TnPedia which provides comprehensive reviews of the major TE families including an extensive general section, and specialised sections with descriptions of insertion sequence and transposon families. TnCentral and TnPedia are intuitive resources that can be used by clinicians and scientists to assess TE diversity in clinical, veterinary and environmental samples.

Introduction

Transposable elements (TE) are key facilitators of bacterial evolution and adaptation and central players in the emergence of antibiotic and heavy metal resistance and to the transmission of virulence and pathogenic traits. Some TE can capture "passenger genes" (genes not involved in the transposition process) encoding these traits and transmit them to plasmids, where they accumulate and are then transferred within and between bacterial populations by conjugation. TE also contribute significantly to the on-going reorganization of bacterial genomes giving rise to new strains that are more adept at proliferating in clinical and agricultural environments, as well as in natural ecosystems.

Understanding TE nature, distribution, and activity is therefore an indispensable part of the struggle to cope with the public health crisis of multiple antibiotic resistance (ABR) [1,2]. To understand the impact of TE on bacterial populations, it is essential to provide a detailed description and catalog of TE structures and diversity. The simplest TE, known as Insertion Sequences (IS), have a profound impact on genome organization and function (see [3–7]) but do not themselves generally carry integrated passenger genes. There are a large number of significantly more complex TE (Figure 1), arguably even more important in the global emergence of ABR and other virulence and pathogenicity traits. These are generically called transposons and may carry multiple passenger genes, including some of the most clinically important antibiotic resistance genes. Like IS, these TE are grouped into a number of distinct families with characteristic organizations [3]. Their transposition activities facilitate the rapid spread of groups of antibiotic resistance genes and promote their horizontal transfer. Yet another important aspect of their impact is their ability to assemble passenger genes into resistance clusters [8,9]. While there appears to be a wide-spread appreciation that mobile plasmids are responsible for the spread of antibiotic resistance, it

is less well-known that IS and transposons are the conduits that transfer this information between chromosomes and plasmids.

There are a number of other bioinformatics resources that cover aspects of prokaryotic TE biology. These include databases for TE passenger genes such as antibiotic resistance (CARD [10], ARDB [11]) or toxin/antitoxin gene pairs (TADB [12], TASmania [13]) as well as the various classes of TE themselves such as insertion sequences (ISfinder [14]), integrons (INTEGRALL [15]), integrative conjugative elements, ICE (ICEberg [16,17]), plasmids (PlasmidFinder [18]) or more general databases which include a variety of these genome components (ACLAME [19–21]). However, there is a need for a resource that collects, compares and collates detailed information on the various different classes of TE that are responsible for the transmission of medically and economically important passenger genes in an intuitive and accessible way.

Here, we describe TnCentral (https://tncentral.proteininformationresource.org/), a database of detailed structural and functional information on bacterial TE. Additionally, TnCentral provides access to TnPedia (https://tnpedia.fcav.unesp.br/), a comprehensive encyclopedia describing the current state of our knowledge of the biology of IS and transposons. Together, TnCentral and TnPedia provide a detailed description of TE diversity with easy-to-understand graphics outputs that are accessible to users without significant bioinformatic knowledge. They allow users to rapidly analyse the landscape of TE in genomes (chromosomes and plasmids) isolated from clinical, veterinary and environmental samples.

Results

TnCentral Website Content

As of May 2021, TnCentral contains information on ~400 TE. About half of these TE are *Tn3*-family transposons. The remainder are integrons, compound transposons, transposons from the *Tn402*, *Tn554* and *Tn7* families, and IS that are associated with TE or are part of compound transposons (Supplementary Table 1). They include TE with resistance to over 25 different classes of antibiotics and nine different heavy metals. The collection also contains TE that carry a toxin/antitoxin system for bacterial plasmid maintenance [22–24] and TE from xanthomonads carrying genes for plant pathogenicity.

TnCental Web Portal

The TnCentral home page is designed to give the user easy access to the contents of TnCentral with a number of options (Figure 2A), including:

TnCentral Search *(search of the TnCentral database),*

Sequence Search *(BLAST-like search for sequence similarities in the database),*

Browse Tn list *(view all TE in TnCentral),*

Tnfinder Software *(access to downloadable scripts for identifying potential TE in sequence databases),*

Documentation *(downloadable documentation for TnCentral),*

For Curators *(detailed curation guidelines),*

TnPedia *(TE Encyclopedia),*

Related links, *and*

Feedback.

TnCentral Search. The interface provides a variety of search functions divided into two search types: ***Transposon search*** and ***Gene search*** (Figure 2B).

Transposon search. The transposon collection can be searched using the transposon **name**, **synonyms** which may have been used in the literature, the **type** of mobile genetic element (e.g., insertion sequence, transposon or integron), the **family** and **subgroup** to which it belongs, the **host organism**, **country** of identification and **date** of identification. The latter three search terms are intended for use in epidemiological tracking. These search terms result in a table that can be sorted, customized and downloaded (See Use Case #1, below).

Gene search. It is also possible to search for TE-associated genes by name, by class (Transposase, Accessory Gene or Passenger Gene) or by function (Antibiotic Resistance, Heavy Metal Resistance) and to retrieve information on the transposons in which they are found (see Use Case #2, below).

Sequence Search. Sequence Search allows users to perform sequence similarity

157 searches using BLAST [26,27] (see Use Case #3, below). By default, the search database is
158 the TnCentral database, but the page also provides links to BLAST against the ISfinder
159 (https://isfinder.biotoul.fr/blast.php), NCBI https://blast.ncbi.nlm.nih.gov/Blast.cgi),
160 Comprehensive Antibiotic Resistance Database (CARD;
161 https://card.mcmaster.ca/analyze/blast) and the Toxin-Anitoxin (TADB;
162 https://bioinfo-mml.sjtu.edu.cn/TADB2/) databases. The BLAST tool automatically
163 distinguishes between DNA and protein query sequences.
164
165 **Browse Tn list.** This option allows the user to browse the entire TnCentral database.
166
167 **The Transposon entry page.** All of the search and browse options provide links to entry
168 pages for each TE (Figure 3), which provide detailed information about TE features and
169 origins. The page includes: 1) host information: host species, strain, and
170 plasmid/chromosome in which the transposon was found as well as the date and geographic
171 location of the isolate; 2) a graphic representation of the annotated sequence with color-
172 coded features; 3) Terminal Inverted Repeats (IR); 4) DNA sequence; 5) internal
173 recombination sites (e.g. *res* sites) including their coordinates, length and DNA sequence; 6)
174 ORF summary, which includes all protein coding genes in the order in which they appear, 5'-
175 3', in the TE sequence, the element with which they are associated (important for nested TE
176 in which one TE is inserted into another), their coordinates, their class (e.g., Transposase,
177 Accessory Gene, Passenger Gene) and subclass (e.g., Antibiotic Resistance, Heavy Metal
178 Resistance) and their relative orientation within the TE; 7) a detailed ORF description
179 including the amino acid sequence; 8) if applicable, a table of Internal Transposable
180 Elements (TE inserted in the main element) including the name, type location and length; 9)
181 if applicable, a table of Internal Repeats (repeat elements, other than the terminal inverted
182 repeats, that are found within the TE), including the associated TE, coordinates and DNA
183 sequence; 10) Bibliographic references with direct links to PubMed [25]. Each section can be
184 collapsed using a button on the right-hand side of the section heading. Sections can be
185 viewed either by scrolling down on the page or by clicking on the section name in the menu
186 located on the left side of the page. Sequence files in FASTA and GenBank format can be
187 downloaded using the links on the left side of the page under the menu.

Tnfinder Software. This section provides three user-downloadable scripts written in-house for identifying transposons. They provide users with local control over analyses and help them screen datasets containing large numbers of genomic sequences using their own servers for identifying potential candidates and which are then manually curated.

Tn3 Transposon Finder (Tn3_finder) performs the automatic prediction of transposable elements of the Tn*3* family in bacteria and archaea. It compares user-provided bacterial and archaeal genome sequences to custom Tn*3* transposase and resolvase databases by BLAST alignments. The criteria for identifying potential transposon regions according to similarity, coverage and distance values can be adjusted by the user. Additional ORFs that might be related to passenger genes are also predicted, and flanking regions can also be retrieved and analyzed. The automatic prediction results are written in report files and pre-annotated GenBank files to help in subsequent manual curation. Tn*3*_finder allows for the concurrent analysis of multiple genomes by multithreading.

Composite Transposon Finder (TnComp_finder) predicts the putative composite transposons in bacterial and archaeal genomes based on insertion sequence replicas in a relatively short span. It works by comparing nucleotide sequences from bacterial and archaeal genomes to a custom transposon database and identifying duplicated transposons in user-defined genomic regions from BLAST alignments. Similar to Tn3_finder, multithreaded analyses of multiple genomes are available and parameters for similarity, coverage, distance and flanking regions can be adjusted by the user. Results are written in report files and pre-annotated GenBank files to help in subsequent manual curation.

Antibiotic Resistance Gene-associated IS Finder (ISAbR_finder) is an experimental program for the automatic prediction of antibiotic resistance genes associated with known IS elements derived from the ISfinder database and has yet to be tested extensively. It works by comparing IS nucleotide sequences from bacterial and archaeal genomes to a custom antibiotic resistance database based on the parsing of BLAST alignment results, using a number of parameters that can be customized by the user for stricter or more relaxed criteria and allowing multithreaded alignments of multiple genomes. ISAbR_finder

219 also produces report files and pre-annotated GenBank files on which the recommended
220 manual curation should be performed.
221
222 **Documentation.** This section, which can be downloaded as a pdf file, provides a short
223 background description of transposons and TnCentral together with a short description of
224 the **curation workflow** and of planned future developments.
225
226 **For Curators.** This section provides a detailed description of the curation workflow used in
227 generating the annotated TnCentral data.

228 **TnPedia.** TnCentral provides access from the homepage to TnPedia, an online knowledge
229 base which contains information concerning transposition in prokaryotes. TnPedia is
230 developed using MediaWiki (https://www.mediawiki.org) and can also be accessed directly
231 (https://tnpedia.fcav.unesp.br/). It is structured into three main sections: **General**
232 **Information, IS Families and Transposon Families** (Figure 4).
233 The **General Information** section provides a series of clickable sections with an
234 extensive bibliography and direct links to the articles in PubMed. It includes a historical
235 perspective, definitions and descriptions of a variety of prokaryotic TE, the basic
236 mechanisms involved in their movement and the enzymes involved in these processes. It
237 also contains information describing their impact on their host genomes and how their
238 activities are controlled.
239 The **IS Families** section consists of individual chapters describing each of the ~25 IS
240 families in detail and covers, where possible, the identification of the founding members,
241 their organisation, distribution, variability and phylogenetic relationships, regulation of their
242 transposition, impact on their host genomes, and their transposition mechanisms including
243 genetic, biochemical and structural studies.
244 The **Transposon Families** section describes each transposon family with similar
245 information to that included in the IS family descriptions but, in addition, including a
246 detailed description of their structures and the passenger genes which they may carry.
247
248 **Examples of TnCentral Use**

Use Case #1: Comparing Protein Coding Genes in Tn*554* Family Members

The Tn*554* family is a small family restricted to the Firmicutes. Members encode three genes, *tnpA*, *tnpB* and *tnpC*, involved in transposition [28,29] (https://tnpedia.fcav.unesp.br/index.php/Transposons_families/Tn554_family). TnpA and TnpB both exhibit a C-terminal motif which shares all the important catalytic residues of a typical tyrosine site-specific recombinase [28,29]. They insert in a sequence-specific way into the DNA repair gene *radC* [30,31] and can also be found in a circular form [32–36]. To compare the protein coding genes in Tn*554* family members side by side, we searched for Tn*554* in the TE family field of the Transposon Search interface (Figure 5A). Fourteen Tn*554* family members were found (of which only 10 are shown in Figure 5B). In order to perform a side-by-side comparison of the protein-coding genes in these TE, we used the Customize Display option on the search results page, to add the "All Gene Fields" columns, which provide information about the protein coding genes, to the display and to remove several columns (e.g., Host Organism, Country) (Figure 5B). Results for two of the Tn*554* transposons (Tn*558.3* and Tn*559*) are shown in Figure 5C. Both transposons have the three-part transposition module (*tnpA*, *tnpB*, *tnpC*) characteristic of the family. However, the two transposons are quite diverse in their passenger genes. Tn*558.3* has gene called *fla*, which contains a flavodoxin-like domain, and the ABR gene *fexA*, which confers resistance to phenicol antibiotics. Tn*559* has just a single passenger gene, the ABR gene, *dfrK*, which confers resistance to diaminopyrimidine antibiotics. As shown by this example, the flexible search results page makes it easy to compare features across multiple transposons.

Use Case #2: Type II Toxin/Antitoxin Systems in Tn*3* Transposons

Toxin/Antitoxin (TA) systems are implicated in plasmid maintenance in bacterial populations [37]. These systems are characterized by a stable toxin and an unstable antitoxin that binds to the toxin and inhibits its lethal effect. Loss of a plasmid carrying a TA system will lead to rapid depletion of the antitoxin, allowing the persistent toxin to kill the cell. Thus, only members of a population that retain the plasmid will survive. Recently, a set of Tn*3*-family transposons carrying TA systems were characterized and included in the TnCentral database [22]. To explore these Tn, we used the TnCentral Gene Search function, selecting "Passenger Gene" from the Gene Class pull-down menu and "Toxin" from the Gene Sub-Class pull-down

menu (Figure 6A, red box). The search results included eight different toxin genes (Gp49, HEPN, PIN, PIN_3, *abiEii*, *higB*, *parE*, and zeta) found in 43 different transposons. Similarly, transposons carrying antitoxin genes were identified using the Gene Search function with the Gene Sub-Class menu set to "Antitoxin" (Figure 6B, red box). There were 44 transposons carrying 11 different antitoxin genes. Combinations of toxin and antitoxin genes in individual transposons were examined by going to the ORF Summary section of the entry pages for the TA transposons. For example, Tn*Sku1* (Figure 6B, yellow box; Figure 6C) has a Gp49 toxin gene and an antitoxin gene containing an HTH domain (referred to as HTH). Most transposons have a single toxin/antitoxin gene pair except for Tn*Xca1*, which has two TA pairs, and Tn*5501.5*, which has a *parD* antitoxin gene and no toxin gene. The majority of Tn*5501* derivatives in TnCentral have a *parE* toxin gene as well as the *parD* antitoxin, suggesting that Tn*5501.5* may have undergone a deletion in the region containing *parE* (Supplementary Figure 1).

Use Case #3: Tn*21* and its Relatives

Tn*21* is the canonical member of a subfamily of Tn*3* transposons that confers a variety of antibiotic resistances [38–40] and several analyses have proposed mechanisms to explain how *Tn21* arose from simpler ancestor transposons (e.g., [40,41]). Tn*21* has a mercury resistance operon at the 5'- (left) end, a *tnpA*/*tnpR* transposition module at the 3'-(right) end, and a transposition-deficient integron (In2) carrying several ABR genes (a GCN5-related N-acetyltransferase (GNAT_fam), *sul1*, *qacEdelta1*, and *aadA*) in the middle (Supplementary Figure 2). These ABR genes confer resistance to aminoglycosides, sulfones, sulfonamides, quaternary ammonium salts, and acridine dye. More recently, a transposon that lacks the integron insertion but is otherwise identical to Tn*21* (the hypothetical Tn*21* backbone Tn21Δ in [40]) was discovered [42]. This transposon, Tn*5060*, was proposed to be the ancestor of Tn*21* [42]. Tn*21* also has numerous relatives that carry different combinations of antibiotic resistance genes within and outside the integron. To explore the Tn*21* subfamily, we performed a TnCentral Sequence Search (BLAST) using the putative ancestral Tn*5060* sequence (Figure 7A). In addition to Tn*5060* itself, we identified ten transposons in the database (Tn*20*, Tn*21*, Tn*21.1*, Tn*21.2*, Tn*5086*, Tn*2411*, Tn*2424*, Tn*4*, Tn*1935*, and Tn*As3*; Supplementary Figure 2) that contain all (or nearly all) of the Tn*5060* sequence. With the

exception of Tn*20*, which is almost identical to Tn*5060* (99.5%), these transposons have two or more discontinuous sub-regions that align to Tn*5060*. For example, Tn*21* has two sub-regions, one of which is a close match to the left half of Tn*5060* and the other of which is a close match to the right half of Tn*5060* (red bars in Figure 7B). This suggests that these transposons arose from Tn*5060* via the insertion of other sequences.

We compared the antibiotic resistance profiles of the ten transposons by inspecting their TnCentral entry pages. Tn*20*, like Tn*5060*, carries no ABR genes. The other nine transposons carry ABR genes targeting aminoglycosides, sulfones, sulfonamides, and quaternary ammonium salts (Figure 7C). Other resistances found in a subset of the six include acridine dye (Tn*1935*, Tn*21*, Tn*2411*, Tn*4*, TnA*s3*, Tn*2424*, Tn*5086*), carbapenams (Tn*1935* and Tn*4*), cephalosporins (Tn*1935* and Tn*4*), carbapenems (Tn*4*), monobactams (Tn*4*), phenicols (Tn*As3*, Tn*2424*, Tn*21.1*, Tn*21.2*), diaminopyrimidines (Tn*5086*, Tn*21.1*, *Tn21.2*), and tetracyclines (Tn*21.2*). Interestingly, in some cases where the transposons have resistances in common, they are conferred by different genes (Figure 7C). For example, phenicol resistance is conferred by *CAT* in Tn*As3, catB2 in* Tn*2424*, and *cmlA6* in Tn*21.1* and Tn*21.2*. Similarly, sulfonamide and sulfone resistance is conferred by *sul1* in all of the antibiotic-resistant family members except for Tn*21.1* and Tn*21.2*, where those resistances are conferred by *sul3*. Thus, even this closely related subfamily of transposons shows diversity in its antibiotic resistance genes. This is partially due to the flexibility of the integron to incorporate new antibiotic resistance gene cassettes but also to insertion of ABR-gene containing elements outside of the integron region (e.g., Tn*3.1* in Tn*4*, Supplementary Figure 2).

Discussion

Here, we have described TnCentral, a user-friendly resource for exploration of prokaryotic TE. TnCentral provides a flexible search interface, TE-specific entry pages with intuitive graphics and detailed information about TE features, and a BLAST interface that allows users to identify TE that carry features of interest or to identify TE that are present in sequences of interest (e.g., plasmids). As shown in the use cases, the flexible search results page makes it easy to compare features across multiple transposons, the detailed entry pages allow

exploration of TE passenger genes, such as ABR genes, and the Sequence Search enables retrieval of TE with related sequences that could be used as a starting point for evolutionary analyses. Moreover, TnCentral provides access to Tnfinder software for locating candidate TE in sequence data and to TnPedia, a comprehensive review of the biology of selected TE families.

As discussed in the Introduction, a variety of resources dedicated to aspects of prokaryotic TE biology currently exist. TnCentral's unique contribution to this universe of resources lies in its coverage of a variety of TE (e.g. different transposon families and compound transposons with their associated IS and integrons) and its detailed focus on both core transposition genes and passenger genes of clinical, environmental, and economic importance. It has the additional feature of providing a clear graphic output for visualizing the often complex structures of TE.

The next step beyond annotation of individual TE is to annotate and visualize the TE content of prokaryotic chromosomes and plasmids. These studies are critical for understanding the propagation of high impact passenger genes, such as those that confer antibiotic resistance. Several tools that address this problem are available. For example, ISsaga [43], which is integrated into ISfinder, annotates IS present in user-provided sequences. Other software suites have been designed specifically to annotate IS in short read raw data (e.g. ISQuest [44], Transposon Insertion Finder [45], ISMapper [46] and panISa [47]) using preassembled libraries of TE and their components, while yet other approaches are based on *ab initio* prediction (e.g., OASIS [48], ISseeker [49] ISEscan [50], or provide a comparative view of IS mobilisation events (e.g. ISCompare [51]). These annotation tools are only as good as their underlying TE databases. ISfinder, which includes nearly 6000 individual examples of IS classified in distinct families and subfamilies according to their transposition mechanism and structural organization, provides such a rigorous framework for IS and has been incorporated into a number of annotation pipelines (e.g., ISsaga [43], MobileElementFinder [52]). However, IS represent only a fraction of prokaryotic TE, and unlike transposons and integrons, they rarely carry passenger genes. We hope that TnCentral will become a benchmark for more complex TE as ISfinder is for IS.

TnCentral is an ongoing project, and we will continue to expand and update the content. In addition to the exporting annotated TE in GenBank format, we plan to make all files available in a SnapGene file format which will allow users to use SnapGene (https://www.snapgene.com/), a commercial software tool (with a free viewer version) for visualizing and documenting nucleotide sequences and their features, to analyze and explore them. We also intend to enhance the visualization of TnCentral Sequence Search (i.e., BLAST) results to better support the analysis of plasmid sequences that may carry a complex complement of TE although it should be noted that the Sequence Search tool can already accommodate analysis of large plasmids. Ultimately, we envision that TnCentral could be used to analyze the TE content of a collection of sequences, such as patient, veterinary and environmental samples from an antibiotic resistance outbreak, to understand TE-driven evolution of the prokaryotic mobilome.

Methods

Curation Workflow

The TnCentral curation workflow is depicted in Figure 8. Curation is performed by members of the TnCentral development team as well as by graduate students in bioinformatics courses at Georgetown University Medical Center. TnFinder scripts are run against RefSeq and other sequence databases and GenBank files potentially containing TE are retrieved. TE sequences are isolated and annotated using SnapGene (https://www.snapgene.com). Features of interest (i.e., protein coding genes, TE, repeat elements, and recombination sites) are annotated according to detailed curation guidelines (provided in the "For Curators" of TnCentral). Fully annotated features are saved in a SnapGene Custom Library. New transposon sequences can be searched against this library, enabling detection of features previously identified in other TEs. All annotated TE files are checked by a second curator. An enhanced GenBank file containing all annotations is exported from SnapGene and checked for common curation formatting errors using a custom Perl script. Detected errors are manually corrected in the SnapGene file, which is then exported as a revised enhanced GenBank file. Information from this GenBank file is used to populate the TnCentral database, which, in turn, serves as the backend for the TnCentral web portal. An

image file showing a color-coded map of TE features is also exported from SnapGene and displayed on the TE entry page.

Although we have adhered to the standard nomenclature for transposons extracted from the literature, for the many transposons newly identified during TnCentral database-building, we have temporarily used names indicating their source. In all cases, the Transposon Registry [53] accession number is provided as a synonym. There is some ambiguity in the literature concerning class 1 integrons and members of the Tn*402* transposon family. Class 1 integrons appear to be derivatives of this transposon family and include members with a range of Tn*402* transposition genes with varying degrees of completeness. We have therefore elected to include all Class 1 integrons as members of the Tn*402* family (Supplementary Table S1). ISfinder classification is used for the individual IS and in the case of compound transposons, the group to which they are belong is defined by the flanking IS.

Properties of protein coding genes are annotated with cross-references to database or ontology identifiers whenever possible. Antibiotic resistance gene properties, including gene name, sequence family, antibiotic resistance mechanism, and target drug classes are annotated according to the Antibiotic Resistance Ontology (ARO) as presented in Comprehensive Antibiotic Resistance Database (CARD) [10]. The Pfam [54] and InterPro resources [55] are used to define sequence family information.

TnCentral Website implementation

TE features and sequence information are extracted from the enhanced GenBank files. TE feature information is used for the search and the entry pages, and the TE DNA and protein sequence information are used for the Sequence Search and display. The extracted data is loaded into the TnCentral database, implemented using MySQL. The website is built on a Linux server with Apache, and the web application is built on Perl CGI. Apache Lucene is used to index the data for flexible and fast search and retrieval. JavaScript is used for the interactive web-interface and display. BLAST is used for similarity search.

Acknowledgements

We would like to thank John Dekker (NIAID, NIH Bethesda, Md, USA), Fred Dyda and Alison Hickman (NIDDK, NIH Bethesda, Md, USA), Patricia Siguier (CNRS Toulouse, France), Susu He (Nanjing University Medical School, Nanjing, China), Laurence van Melderen (Université Libre de Bruxelles, Belgium), and Gipsi Lima-Mendez and Bernard Hallet (Université de Louvain la Neuve, Belgium) for helpful discussions; Ben Glick (University of Chicago, SnapGene) for his help with the SnapGene software; the student curators at Georgetown University Medical Center for their contributions to the annotation process; and the Protein Information Resource (University of Delaware, Georgetown University Medical Center) for informatics support. This project was supported the U.S. Department of Defense (DoD) Global Emerging Infections Surveillance (GEIS) Branch (P0020_18_WR) and by institutional resources at the Center for Bioinformatics and Computational Biology, University of Delaware.

Figure Legends

Figure 1. Structural arrangement of prokaryotic Transposable Elements. The TE is indicated by a pale-yellow horizontal bar at the top of each section. Open reading frames are shown as horizontal arrows with the arrowheads indicating the direction of expression: purple, transposition-associated genes; red, antibiotic resistance genes; green, other passenger genes. The inverted terminal repeats found at the ends of the majority of TE are shown as grey arrows and the direct target repeats generally produced by insertion are indicated by small black arrows. **A.** Insertion Sequence (IS), a short DNA segment encoding only the mobilization protein (Transposase, TnpA), flanked by two imperfect Inverted Repeats (IRs), and generally containing a short flanking directly repeated duplication (DR) on the target of insertion. **B.** tIS (transporter IS) are structurally similar to an IS, but contain passenger genes. They are presently restricted to the IS*1595* and IS*66* families. **C.** Compound transposons are formed by two IS in either direct or inverted orientation, flanking a variety of passenger genes including those for antibiotic resistance. **D.** Transposons are more heterogeneous structures and include different sets of transposition-related genes which are specific to each Tn family and multiple antibiotic resistances, virulence and other passenger genes. This is an example of a Tn*3* family transposon with transposon, *tnpA*, and resolvase genes, *tnpR*.

Figure 2. A) TnCentral homepage showing clickable links to various TnCentral sections in the box on the left. B) TnCentral search interface showing search choices for TE on the left and for transposition-related and passenger genes on the left.

Figure 3. TnCentral TE Entry Page. #1-10: Sections of the entry page (see text for details).

Figure 4. The main sections of TnPedia, a TnCentral-related wiki compiling information on prokaryotic transposable elements. Only three of the four sections (General Information, IS families and Transposon families) are illustrated. The fourth section is a Transposition Glossary, which is under construction.

Figure 5. Comparing Protein Coding Genes in Tn*554* Family Members. A) TnCentral Transposon Search interface, showing a search for Tn*554* in the TE family field. B) Interface for customizing the columns in the search results display. Clicking on "Customize Display" (red box) opens the interface. C) Partial Tn*554* family search results after customization to show information on protein coding genes (All Gene Fields).

Figure 6. Exploring Toxin/Antitoxin Genes in TnCentral. A) Partial results of searching TnCentral for toxin genes. The settings used to obtain these results are shown in the red box. Links to entry pages for the TE carrying the indicated genes are provided in the MGE Accession column (e.g., Tn*Sku1*-CP002358.1, yellow box). B) Partial search results for antitoxin genes in TnCentral. Settings are shown in the red box. C) ORF Summary section of the entry page for Tn*Sku1*-CP002358.1, showing the presence of a toxin/antitoxin gene pair (Gp49 toxin/HTH antitoxin).

Figure 7. Analysis of ABR in Tn*21* Relatives. A) TnCentral Sequence Search using the sequence of Tn*5060*, the proposed ancestor of Tn*21*, as a query. B) Sequence Search results. The query sequence is represented by the width of the Alignment column. The red bars represent regions of the matched transposons that are highly similar to regions of Tn*5060*. C) ABR genes and targeted antibiotic classes in Tn*21* relatives. Red shading in the table cells indicates that the transposon carries at least one gene targeting the antibiotic class; blue shading indicates that it does not. The ABR genes found in each transposon are indicated in the table cells.

Figure 8. TnCentral Curation Workflow (see text for description).

Supplementary Figures

Supplementary Figure 1. Maps of Tn*5501* and Tn*5501.5* showing the loss of *parE* toxin gene in Tn*5501.5*. Feature color code: yellow--TE; purple--transposition genes; dark orange--toxin/anti-toxin genes; light orange--other open reading frames; grey--repeat elements; green--recombination sites. Maps were created with SnapGene.

Supplementary Figure 2. Maps of Tn*21* and its relatives. The feature color code is the same as in Supplementary Figure 1. Maps were created with SnapGene. Note that the different transposon derivatives are not to scale but their individual lengths are included.

Supplementary Table S1. The table displays the entire collection of TE at present in the database (May 2021) with columns indicating their **TnCentral accession numbers**, their **names**, **synonyms** from the literature and/or the Transposon Registry [53], TE **Type, Family Group**), **Host Organism** and **Molecular Source** (e.g., plasmid or chromosome). If no information is provided in the Molecular Source column, the source is chromosomal or unknown.

Bibliography

1. Spellberg B, Guidos R, Gilbert D, Bradley J, Boucher HW, Scheld WM, Bartlett JG, Edwards J, Infectious Diseases Society of America. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clin Infect Dis. 2008 Jan 15;46(2):155–164. PMID: 18171244

2. O'Neill J. Tackling Drug-Resistant Infections Globally:' ' Final Report and Recommendations. London?: Welcome Foundation HM Government UK; 2016 May.

3. Craig, Chandler, Gellert, Lambowitz, Rice, Sandmeyer, editors. Mobile DNA III. American Society of Microbiology; 2015.

4. Mahillon J, Chandler M. Insertion sequences. Microbiol Mol Biol Rev. 1998 Sep;62(3):725–774. PMCID: PMC98933

5. Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M. Everyman's guide to bacterial insertion sequences. Microbiol Spectr. 2015 Apr;3(2):MDNA3–0030. PMID: 26104715

6. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014 Sep;38(5):865–891. PMCID: PMC7190074

7. Vandecraen J, Chandler M, Aertsen A, Van Houdt R. The impact of insertion sequences on bacterial genome plasticity and adaptability. Crit Rev Microbiol. 2017 Nov;43(6):709–730. PMID: 28407717

8. He S, Chandler M, Varani AM, Hickman AB, Dekker JP, Dyda F. Mechanisms of Evolution in High-Consequence Drug Resistance Plasmids. MBio. 2016 Dec 6;7(6). PMCID: PMC5142620

9. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, Dyda F. Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition. MBio. 2015 Jun 9;6(3):e00762. PMCID: PMC4471558

10. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020 Jan 8;48(D1):D517–D525. PMCID: PMC7145624

552 11. Liu B, Pop M. ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res. 2009
553 Jan;37(Database issue):D443–7. PMCID: PMC2686595

554 12. Xie Y, Wei Y, Shen Y, Li X, Zhou H, Tai C, Deng Z, Ou H-Y. TADB 2.0: an updated
555 database of bacterial type II toxin-antitoxin loci. Nucleic Acids Res. 2018 Jan
556 4;46(D1):D749–D753. PMCID: PMC5753263

557 13. Akarsu H, Bordes P, Mansour M, Bigot D-J, Genevaux P, Falquet L. TASmania: A
558 bacterial Toxin-Antitoxin Systems database. PLoS Comput Biol. 2019 Apr
559 25;15(4):e1006946. PMCID: PMC6504116

560 14. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference
561 centre for bacterial insertion sequences. Nucleic Acids Res. 2006 Jan 1;34(Database
562 issue):D32–6. PMCID: PMC1347377

563 15. Moura A, Soares M, Pereira C, Leitão N, Henriques I, Correia A. INTEGRALL: a database
564 and search engine for integrons, integrases and gene cassettes. Bioinformatics. 2009
565 Apr 15;25(8):1096–1098. PMID: 19228805

566 16. Bi D, Xu Z, Harrison EM, Tai C, Wei Y, He X, Jia S, Deng Z, Rajakumar K, Ou H-Y. ICEberg:
567 a web-based resource for integrative and conjugative elements found in Bacteria.
568 Nucleic Acids Res. 2012 Jan;40(Database issue):D621–6. PMCID: PMC3244999

569 17. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, Tai C, Deng Z, Ou H-Y. ICEberg 2.0: an updated
570 database of bacterial integrative and conjugative elements. Nucleic Acids Res. 2019
571 Jan 8;47(D1):D660–D665. PMCID: PMC6323972

572 18. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller
573 Aarestrup F, Hasman H. In silico detection and typing of plasmids using PlasmidFinder
574 and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014
575 Jul;58(7):3895–3903. PMCID: PMC4068535

576 19. Leplae R, Lima-Mendez G, Toussaint A. ACLAME: a CLAssification of Mobile genetic
577 Elements, update 2010. Nucleic Acids Res. 2010 Jan;38(Database issue):D57–61.
578 PMCID: PMC2808911

579 20. Leplae R, Hebrant A, Wodak SJ, Toussaint A. ACLAME: a CLAssification of Mobile
580 genetic Elements. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D45–9. PMCID:
581 PMC308818

582 21. Leplae R, Lima-Mendez G, Toussaint A. A first global analysis of plasmid encoded
583 proteins in the ACLAME database. FEMS Microbiol Rev. 2006 Nov;30(6):980–994.
584 PMID: 17064288

585 22. Lima-Mendez G, Oliveira Alvarenga D, Ross K, Hallet B, Van Melderen L, Varani AM,
586 Chandler M. Toxin-Antitoxin Gene Pairs Found in Tn3 Family Transposons Appear To
587 Be an Integral Part of the Transposition Module. MBio. 2020 Mar 31;11(2). PMCID:
588 PMC7157771

589 23. Szuplewska M, Czarnecki J, Bartosik D. Autonomous and non-autonomous Tn3-family
590 transposons and their role in the evolution of mobile genetic elements. Mob Genet
591 Elements. 2014 Dec;4(6):1–4. PMCID: PMC4588545

592 24. Loftie-Eaton W, Yano H, Burleigh S, Simmons RS, Hughes JM, Rogers LM, Hunter SS,
593 Settles ML, Forney LJ, Ponciano JM, Top EM. Evolutionary Paths That Expand Plasmid
594 Host-Range: Implications for Spread of Antibiotic Resistance. Mol Biol Evol. 2016
595 Apr;33(4):885–897. PMCID: PMC4840908

596 25. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N,
597 Funk K, Hefferon T, Holmes JB, Kim S, Kimchi A, Kitts PA, Lathrop S, Lu Z, Madden TL,
598 Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD, Ostell J. Database
599 resources of the National Center for Biotechnology Information. Nucleic Acids Res.
600 2019 Jan 8;47(D1):D23–D28. PMCID: PMC6323993

601 26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool.
602 J Mol Biol. 1990 Oct 5;215(3):403–410. PMID: 2231712

603 27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
604 BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.
605 PMCID: PMC2803857

606 28. Murphy E, Huwyler L, de Freire Bastos M do C. Transposon Tn554: complete
607 nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive
608 mutants. EMBO J. 1985 Dec 1;4(12):3357–3365. PMCID: PMC554666

609 29. Bastos MC, Murphy E. Transposon Tn554 encodes three products required for
610 transposition. EMBO J. 1988 Sep;7(9):2935–2941. PMCID: PMC457088

611 30. Krolewski JJ, Murphy E, Novick RP, Rush MG. Site-specificity of the chromosomal
612 insertion of Staphylococcus aureus transposon Tn554. J Mol Biol. 1981 Oct
613 15;152(1):19–33. PMID: 6279864

614 31. Murphy E, Reinheimer E, Huwyler L. Mutational analysis of att554, the target of the
615 site-specific transposon Tn554. Plasmid. 1991 Jul;26(1):20–29. PMID: 1658834

616 32. Haroche J, Allignet J, El Solh N. Tn5406, a new staphylococcal transposon conferring
617 resistance to streptogramin a and related compounds including dalfopristin.
618 Antimicrob Agents Chemother. 2002 Aug;46(8):2337–2343. PMCID: PMC127375

619 33. Kehrenberg C, Schwarz S. Florfenicol-chloramphenicol exporter gene fexA is part of
620 the novel transposon Tn558. Antimicrob Agents Chemother. 2005 Feb;49(2):813–815.
621 PMCID: PMC547298

622 34. Kadlec K, Schwarz S. Identification of the novel dfrK-carrying transposon Tn559 in a
623 porcine methicillin-susceptible Staphylococcus aureus ST398 strain. Antimicrob Agents
624 Chemother. 2010 Aug;54(8):3475–3477. PMCID: PMC2916295

625 35. Li D, Li X-Y, Schwarz S, Yang M, Zhang S-M, Hao W, Du X-D. Tn6674 Is a Novel
626 Enterococcal optrA-Carrying Multiresistance Transposon of the Tn554 Family.
627 Antimicrob Agents Chemother. 2019 Aug 23;63(9). PMCID: PMC6709469

628 36. Schwendener S, Perreten V. New transposon Tn6133 in methicillin-resistant
629 Staphylococcus aureus ST398 contains vga(E), a novel streptogramin A, pleuromutilin,
630 and lincosamide resistance gene. Antimicrob Agents Chemother. 2011
631 Oct;55(10):4900–4904. PMCID: PMC3187003

632 37. Hayes F, Van Melderen L. Toxins-antitoxins: diversity, evolution and function. Crit Rev
633 Biochem Mol Biol. 2011 Oct;46(5):386–408. PMID: 21819231

634 38. De La Cruz F, Grinsted J. Genetic and Molecular Characterization of Tn21, a Multiple
635 Resistance Transposon from R100.1. J Bacteriol. 1982;151:222–228. PMCID: 6282806

636 39. Grinsted J, de la Cruz F, Schmitt R. The Tn21 subgroup of bacterial transposable
637 elements. Plasmid. 1990 Nov;24(3):163–189. PMID: 1963947

638 40. Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome.
639 Microbiol Mol Biol Rev. 1999 Sep;63(3):507–522. PMCID: PMC103744

640 41. Tanaka M, Yamamoto T, Sawai T. Evolution of complex resistance transposons from an
641 ancestral mercury transposon. J Bacteriol. 1983 Mar;153(3):1432–1438. PMCID:
642 PMC221794

643 42. Kholodii G, Mindlin S, Petrova M, Minakhina S. Tn5060 from the Siberian permafrost is
644 most closely related to the ancestor of Tn21 prior to integron acquisition. FEMS
645 Microbiol Lett. 2003 Sep 26;226(2):251–255. PMID: 14553919

646 43. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. ISsaga is an ensemble of
647 web-based methods for high throughput identification and semi-automatic annotation
648 of insertion sequences in prokaryotic genomes. Genome Biol. 2011 Mar 28;12(3):R30.
649 PMCID: PMC3129680

650 44. Biswas A, Gauthier DT, Ranjan D, Zubair M. ISQuest: finding insertion sequences in
651 prokaryotic sequence fragment data. Bioinformatics. 2015 Nov 1;31(21):3406–3412.
652 PMID: 26116929

653 45. Nakagome M, Solovieva E, Takahashi A, Yasue H, Hirochika H, Miyao A. Transposon
654 Insertion Finder (TIF): a novel program for detection of de novo transpositions of
655 transposable elements. BMC Bioinformatics. 2014 Mar 14;15:71. PMCID: PMC4004357

656 46. Hawkey J, Hamidian M, Wick RR, Edwards DJ, Billman-Jacobe H, Hall RM, Holt KE.
657 ISMapper: identifying transposase insertion sites in bacterial genomes from short read
658 sequence data. BMC Genomics. 2015 Sep 3;16:667. PMCID: PMC4558774

659 47. Treepong P, Guyeux C, Meunier A, Couchoud C, Hocquet D, Valot B. panISa: ab initio
660 detection of insertion sequences in bacterial genomes from short read sequence data.
661 Bioinformatics. 2018 Nov 15;34(22):3795–3800. PMID: 29931098

662 48. Robinson DG, Lee M-C, Marx CJ. OASIS: an automated program for global investigation
663 of bacterial and archaeal insertion sequences. Nucleic Acids Res. 2012
664 Dec;40(22):e174. PMCID: PMC3526298

665 49. Adams MD, Bishop B, Wright MS. Quantitative assessment of insertion sequence
666 impact on bacterial genome architecture. Microb Genom. 2016 Jul 18;2(7):e000062.
667 PMCID: PMC5343135

668 50. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in
669 prokaryotic genomes. Bioinformatics. 2017 Nov 1;33(21):3340–3347. PMID: 29077810

670 51. Mogro EG, Ambrosis N, Lozano MJ. Easy identification of insertion sequence
671 mobilization events in related bacterial strains with ISCompare. BioRxiv. 2020 Oct 16;

672 52. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN.
673 Detection of mobile genetic elements associated with antibiotic resistance in
674 Salmonella enterica using a newly developed web tool: MobileElementFinder. J
675 Antimicrob Chemother. 2021 Jan 1;76(1):101–109. PMCID: PMC7729385

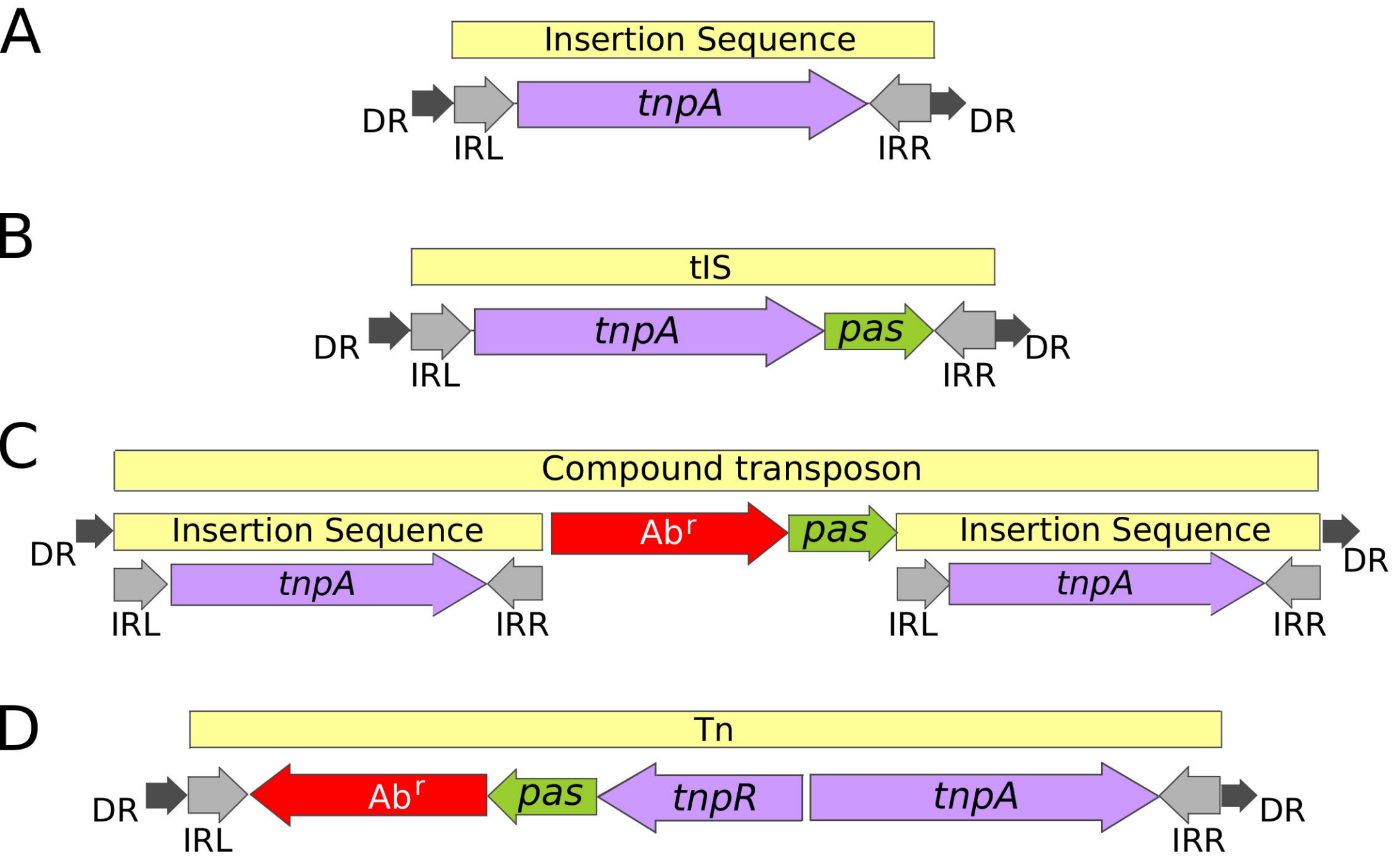
676 53. Tansirichaiya S, Rahman MA, Roberts AP. The Transposon Registry. Mob DNA. 2019
677 Oct 9;10:40. PMCID: PMC6785933

678 54. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson
679 LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE,
680 Finn RD. The Pfam protein families database in 2019. Nucleic Acids Res. 2019 Jan
681 8;47(D1):D427–D432. PMCID: PMC6324024

682 55. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y,
683 El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A,
684 Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C,
685 Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND,
686 Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe

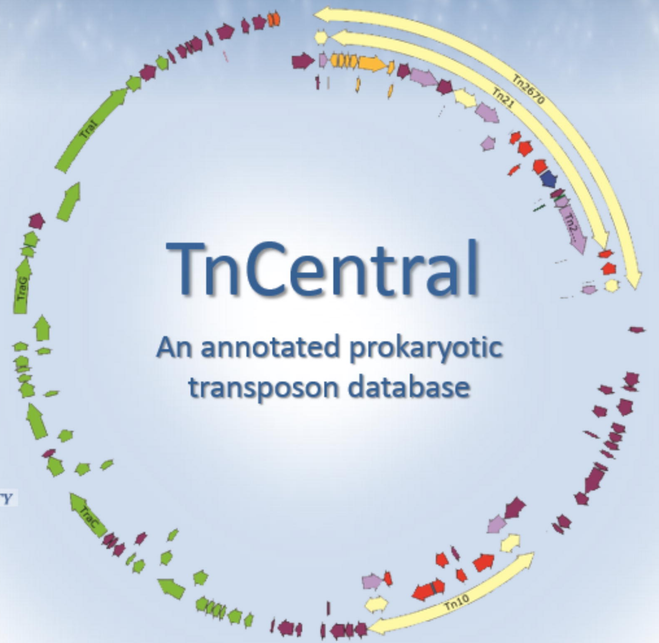
687 I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong S-Y, Finn RD. InterPro in 2019:
688 improving coverage, classification and access to protein sequence annotations. Nucleic
689 Acids Res. 2019 Jan 8;47(D1):D351–D360. PMCID: PMC6323941

690


A
Insertion Sequence
DR
IRL
tnpA
IRR
DR
B
tIS
DR
IRL
tnpA
pas
IRR
DR
C
Compound transposon
DR
Insertion Sequence
Ab^r
pas
Insertion Sequence
DR
IRL
tnpA
IRR
IRL
tnpA
IRR
D
Tn
DR
IRL
Ab^r
pas
tnpR
tnpA
IRR
DR

A

- About
- TnCentral Search
- Sequence Search
- Browse Tn List
- TnFinder Software
- Documentation
- For Curators
- TnPedia (TE Encyclopedia)
- Related Links
- Feedback



TnCentral

An annotated prokaryotic transposon database

B

TnCentral Search

Transposon Search

Any field:

TE Name contains
TE Synonym: contains
TE type: All
TE family: contains
TE group: contains

TnCentral accession: contains
Host organism: contains

Country:
Date of Isolation:

Submit Reset

Gene Search

By Gene Name:

Submit

By Gene Class:

-- Select one ----------

Submit

By Gene Function:

-- Select one ----------

Submit

1

Transposon

Name: Tn5501

Family: Tn3 **Group:** Tn3000

Evidence of Transposition: Yes

Host

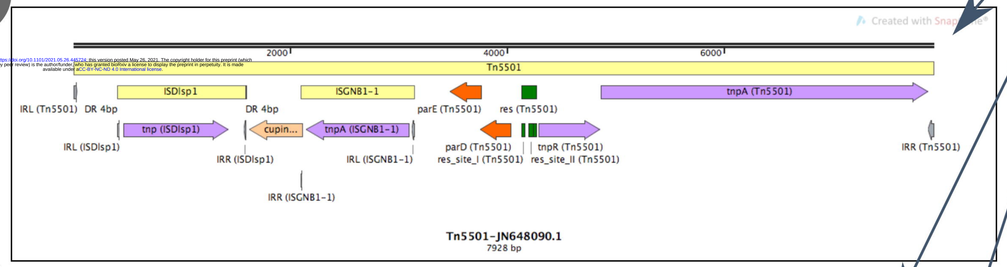
Host Organism: Delftia sp. KV29

Molecular Source: plasmid pKV29

Date of Isolation: 1998

2

Map



3

Terminal Inverted Repeats (IR)

IRL (Length: 38 bp) → GGGGTTCTAAGCCGGAACCGCCGAAAATTCCGTCAGCC

IRR (Length: 38 bp) → GGGGTTCTAAGCCAGAACCGCCGAAATTTCCGTCATCC

4

Sequence

DNA Sequence Length 7928

```
---------10 ---------20 ---------30 ---------40 ---------50 ---------60 ---------70 ---------80 ---------90 ---------100
GGGGTTCTAA GCCGGAACCG CCGAAAATTC CGTCAGCCGA TCAACGTGGC TTGTCCCGCG CCCGGTCGAT GGGGTAGACC CAACAGTCGT GTCACTAGCC 100
GCCATTTCGA TCACGGCAAT GCCAGCCGGA CGTCACGTCC AGATTGTTCC GGTCTGGATG AGGCCGACTG ACGTCTCGGA TGACGGGTGG CATACAACTG 200
CTGTGAGTCC TGCAGGGGGG CAGCTGCCTG ACCGGACGGC GAGCATCAGC CCATCTCATG TATTAGTCAT GTCAGCTTTG ACACTGCGCA CGCGACGGCA 300
CCCGACCCGT TGCAGACCCC CAGACATATG GAAAGCTGAC GCTCAACGTG GAGTTAGCCG GCGCGGCGCG GCTTCATCGC GCAGCGTCCG TGTTGATGGA 400
```

Jump to:

- Top
- Host
- Map
- Terminal Inverted Repeats
- Sequence
- Recombination Sites
- ORFs
  - ORF Summary
  - ORF Details
- Internal Transposable Elements
- Internal Repeats
- References

Download:

Source File (GenBank format)

Sequence File (FASTA format)

5

Recombination Sites

Name	Coordinates	Length	Sequence
res	4139-4269	131	GCCTGTCGGA AAACATTTGT TTTTCGACAG GCCTTCAACG GTCCTCTGCA CCAACCTCCG AGTGGCCGCA AAATTGTGCG GAAAACTCTG TCGCCAGACG CTACCATACG GAAACCTCGT CTTAATGGTT T

6

ORFs

ORF Summary

Gene Name	Associated TE	Coordinates	Class	Sub Class	Orientation
tnp	ISDlsp1	462-1436	Transposase		+
cupin2	Tn5501	1626-2111	Passenger Gene	Other	-
tnpA	ISGNB1-1	2150-3097	Transposase		-

ORF Details

Gene Name	Protein Name	Associated TE	Length	Coordinates	Strand
tnp	Tnp	ISDlsp1	975	462-1436	+

Class: Transposase

Transposase Chemistry: DDE

Protein Sequence:
MLTGMKQSSL ELNLSTRKTR KQELLAQMDR VVPWAALVEL IAPYYPEGKN GRPPFALEAM LRVHCMQQWF TLSDLAMEEA FFDTPIYREF AGLDAHGRMP
DESTILRFRH RLEKHRLAEQ ILATVNDLLA ARGLLLKAGT AVDATLIAAP SSTKNKDRKR DPEMHSSQKG NEWHFGMKAH IGVDADSGLV HTVIGTSGNV
ADVTEGNSLL HGEETDAFGD AGYQGAHKRP DARKDVTWHV AMRPGKRKEL DKENNPVDAL IDQVEKIKAS IRAKVEHPFR VIKRQFGYTK VRYRGLKKNT
LQLKTLFALS NLWMVRHQLL GAQG

8

Internal Transposable Elements (TE)

TnCentral Accession	TE Name	Type	Coordinates	Length
ISDlsp1-JN648090.1	ISDlsp1	Insertion Sequence	409-1592	1184
ISGNB1-1-EF628291	ISGNB1	Insertion Sequence	2097-3143	1047

9

Internal Repeat Elements

Name	Associated Mobile Element	Coordinates	Sequence (Top Strand)
IRL	ISDlsp1	409-428	GGAAATCCTG CAAAACCTCG
IRR	ISDlsp1	1574-1592	GCTCAACAAG TCCTGTAGG

10

References

Stolze Y, Eikmeyer F, Wibberg D, Brandis G, Karsten C, Krahn I, Schneiker-Bekel S, Viehover P, Barsch A, Keck M, Top EM, Niehaus K, Schluter A. *IncP-1beta plasmids of Comamonas sp. and Delftia sp. strains isolated from a wastewater treatment plant mediate resistance to and decolorization of the triphenylmethane dye crystal violet.* Microbiology. 2012 Aug;158(Pt 8):2060-2072. doi: 10.1099/mic.0.059220-0. Epub 2012 May 31. PubMed ID: 22653947

General Information

1. **Overview**
2. **Insertion Sequence History and Early Transposition Models**
3. **What Is an IS?**
4. **ISfinder and the Growing Number of IS**
5. **IS Identification, nomenclature and naming attribution**
6. **IS Distribution**
7. **Major Groups are Defined by the Type of Transposase They Use**
8. **Fuzzy Borders**
9. **tIS - IS and relatives with passenger genes**
10. **IS derivatives of Tn*3* family transposons**
11. **IS related to Integrative Conjugative Elements (ICEs)**
12. **IS*91* and IS*CR* families**
13. **Non-autonomous IS derivatives**
14. **Relationship Between IS and Eukaryotic TE**
15. **Impact of IS on Genome Evolution - The Importance of Time Scale**
16. **Target Choice**
17. **Influence of transposition mechanisms on genome impact**
18. **IS and Gene Expression**
19. **IS Organization**
20. **Control of transposition activity**
21. **Transposase expression and activity**
22. **Reaction mechanisms**
23. **The Casposases**

Main Page

Welcome to TnPedia, the TnCentral Wiki

This Tnpedia has been written in an attempt to assemble a body of information (including many of the historical articles) generally dispersed in the literature as an aid to understanding how knowledge has been built up to our present view of the key role played by transposable elements (TE) in prokaryotes – both in influencing gene expression, in driving genome evolution and in facilitating horizontal gene transfer (HGT). It is divided into four sections:

General Information on Prokaryotic Elements

A section providing general information about Transposition and **T**ransposable **E**lements (TE) with emphasis on prokaryotic elements. It was originally written for ISfinder (https://www-is.biotoul.fr/index.php) with contributions from P. Siguier and E. Gourbeyre. It contains historical, mechanistic and genetic information.

Insertion Sequence (IS) Families

A second section describing each **I**nsertion **S**equence (IS) family in some detail. This section has been entirely updated compared to that which was included in a previous ISfinder version. It contains information from a number of reviews and from the primary literature together with analyses undertaken in the framework of ISfinder and TnCentral. P. Siguier and E. Gourbeyre provided a large proportion of these analyses.

Transposon families [In progress]

A third section presenting detailed descriptions of transposon and transposon families written in the framework of TnCentral.

IS Families

Prokaryotic Insertion Sequences (IS)
1. IS*1* family
2. IS*1595* family
3. IS*3* family
4. IS*481* family
5. IS*4* and related families
5a. IS*701* family
5b. IS*H3* family
5c. IS*1634* family
6. IS*5* and related IS*1182* families
7. IS*6* family
8. IS*21* family
9. IS*30* family
10. IS*66* family
11. IS*110* and IS*1111* families
12. IS*256* family
13. IS*630* family
14. IS*982* family
15. IS*1380* family
16. ISA*s1* family
17. ISL*3* family
18. ISA*zo13* family
19. IS*607* family [In progress]
20. IS*91* and related IS*CR* families
21. IS*200*/IS*605* family

Transposon families

Prokaryotic Transposon Families
1. Composite or compound transposons
2. Tn*3* family transposons
3. Tn*7* family transposons
4. Tn*402* family transposons
5. Tn*554* family transposons

A

Transposon Search

Any field:

TE Name: contains

TE Synonym: contains

TE type: All

TE family: contains Tn554

TE group: contains

TnCentral accession: contains

Host organism: contains

Country:

Date of Isolation:

Submit Reset

B

Transposon Search

Transposon Search Output Page

search | MGE Family | Tn554 | AND | Any Field | + add input box | − del input box

Page Size: 10 items/page

Fields Not in Display	Fields In Display
Country	All Gene Fields
Date of Isolation	MGE Family
Evidence of Transposition	MGE Name
First Isolate	

apply | cancel

14 entries | 2 pages | 10 / page | 1 | 2

Select | Customize Display | Save Result As: TABLE | FASTA

TnCentral Accession	MGE Name	MGE Synonyms	MGE Type	MGE Family	MGE Group	Host Organism	Country	Date of Isolation
Tn5406-AF186237	Tn5406		transposon	Tn554		Staphylococcus aureus MRSA	Spain	January 2011
Tn554-X03216.1.1	Tn554		transposon	Tn554		Staphylococcus aureus	USA	1979
Tn554Cad-AB037671.1	Tn554Cad		transposon	Tn554		Staphylococcus aureus 85_2082	New Zealand	2001
Tn558-AJ715531.1	Tn558		transposon	Tn554		Staphylococcus lentus	USA	2004
Tn558.1-AM408573.1	Tn558.1		transposon	Tn554		Staphylococcus warneri	Denmark	2006
Tn558.2-MH018573	Tn558.2		transposon	Tn554		Enterococcus Enterococcus avium strain C674	China	2018
Tn558.3-CP025122	Tn558.3		transposon	Tn554		Bacillus sp. HBCD-sjtu	China	2017
Tn559-FN677369	Tn559		transposon	Tn554		Staphylococcus aureus ST398	Germany	2010
Tn6133-FR772051.1	Tn6133		transposon	Tn554		Staphylococcus aureus subsp. aureus ST398	Switzerland	2011
Tn6188-HF565366	Tn6188		transposon	Tn554		Listeria monocytogenes 6179	Austria	2013

C

TnCentral Accession	MGE Name	MGE Family	All Gene Fields						
			Gene Name	Gene Class	Gene Function	ORF Target	Gene Length (bp)	Protein Length (aa)	Protein Molecular Function	Sequence Family
Tn558.3-CP025122	Tn558.3	Tn554	tnpA	Transposase			1086	361		
			tnpB	Transposase			1920	639		
			tnpC	Accessory Gene	Helper		366	121		Tn554_family
			fla	Passenger Gene	Other		417	138		flavodoxin
			fexA (ARO:3002704)	Passenger Gene	Antibiotic Resistance	phenicol antibiotic (ARO:3000387)	1428	475	antibiotic efflux (ARO:0010000)	major facilitator superfamily (MFS) antibiotic efflux pump (ARO:0010002)
Tn559-FN677369	Tn559	Tn554	tnpA	Transposase			1086	361		
			tnpB	Transposase			1893	630		
			tnpC	Accessory Gene	Helper		378	125		Tn554_family
			dfrK (ARO:3002869)	Passenger Gene	Antibiotic Resistance	diaminopyrimidine antibiotic (ARO:3000171)	492	163	antibiotic target replacement (ARO:0001002)	trimethoprim resistant dihydrofolate reductase dfr (ARO:3001218)

A

Gene Search

Gene Search Output Page

Gene Class: Passenger Gene | Gene Sub Class: Toxin | Gene Target: -- Select one ----------

8 items | 1 page | 10 / page |

Select | Customize Display | Save As TABLE

Gene Name	Gene Class	Gene Function	ORF Target	MGE Accession	MGE Name	Host Organism	Country
				TnpPGH1-Y09450.1	TnpPGH1	Pseudomonas putida	
				TnSku1-CP002358.1	TnSku1	Sulfuricurvum kujiense DSM 16994	Japan
Gp49	Passenger Gene	Toxin	ribosome associated mRNA	Tn4662a.1-AY831462.1	Tn4662a.1	Pseudomonas putida GJ31	
				Tn4662a-NC_014124.1	Tn4662a	Pseudomonas putida HS	U.S.A
				Tn5501.12-CP017294.1	Tn5501.12	Pseudomonas aeruginosa PA83	Germany
HEPN	Passenger Gene	Toxin	RNA	TnSod9-NC_004349	TnSod9	Shewanella oneidensis MR-1	
				TnXca1-NC_007507	TnXca1	Xanthomonas campestris pv. vesicatoria	
PIN	Passenger Gene	Toxin	single stranded RNA	TnPsy42-KX009060.1	TnPsy42	Pseudomonas syringae pv. actinidiae RT594	Japan
				TnXax1.1-NC_016053	TnXax1.1	Xanthomonas arboricola pv. pruni CFBP 55306	

Gene Class: Passenger Gene | Gene Sub Class: Antitoxin | Gene Target: -- Final --

11 items | 2 pages | 10 / page | 1 | 2

Select | Customize Display | Save As TABLE

B

Gene Name	Gene Class	Gene Function	ORF Target	MGE Accession	MGE Name	Host Organism	Country
HTH	Passenger Gene	Antitoxin		TnSku1-CP002358.1	TnSku1	Sulfuricurvum kujiense DSM 16994	Japan
				Tn4662a.1-AY831462.1	Tn4662a.1	Pseudomonas putida GJ31	
				Tn4662a-NC_014124.1	Tn4662a	Pseudomonas putida HS	U.S.A
HTH_37	Passenger Gene	Antitoxin		Tn5501.12-CP017294.1	Tn5501.12	Pseudomonas aeruginosa PA83	Germany
				TnpPGH1-Y09450.1	TnpPGH1	Pseudomonas putida	
PIN_12	Passenger Gene	Antitoxin		TnBth3-CP003766	TnBth3	Bacillus thuringiensis HD-789	

ORF Summary

C

Gene Name	Associated TE	Coordinates	Class	Sub Class	Orientation
HTH	TnSku1	126-416	Passenger Gene	Antitoxin	+
Gp49	TnSku1	419-712	Passenger Gene	Toxin	+
tnpR	TnSku1	713-1336	Accessory Gene	Resolvase	-
tnpA	TnSku1	1492-4410	Transposase		+

A

Sequence Search (BLAST)

Job Name:

Database

TnCentral

Other Sequence Search Services:

iSfinder: ○ CARD: ○ Toxin/Antitoxin: ○ NCBI: ○

Query Sequence

Enter a sequence here:

```
>Tn5060
GGGGGCACCTCAGAAAACGGAAAATAAAGCACGCTAAGGCATAGCTGACCTTGCCAGGCCTGCTTCGCCCTGTAGTGACGCGATCAACGGGCAG
GAAACATTCCCCTTTCGTGCATGGCAGGCGCACACGAGTTCAGACAGCACGGTTTCCATGCGCGCCAAGTCGGCCATCTTCTCGCGCACGTCCT
TGAGCTTGTGTTCGGCCAGGCTGCTGGCCTCCTCGCAGTGGGTGCCATCGTCGAGCCGCAACAGCTCGGCAATCTCGTCCAGACTGAACCCCAG
CCGCTGTGCCGATTTCACGAATTTCACCCGAACCACGTCCGCCTCCCCATAGCGGCGGATGCTGCCGTAAGGCTTGTCCGGTTCCGGCAACAGG
CCCTTGCGCTGATAGAAGCGGATTGTCTCCACGCTGACCCCGGCCGCCTTGGCAAAAACGCCAATGGTCAGGTTTTCCAAATTATTTTCCATAT
CGCTTGACTCCGTACATGAGTACGGAAGTAAGGTTACGCTATCCAATCCAAATTCAAAAGGGCCAACGTATGTCTGAACCACAAAACGGGCGCG
GTGCGCTCTTCGCCGGCGGGCTGGCCGCCATTCTTGCATCGACCTGCTGCCTGGGGCCGCTAGTACTGGTCGCCCTGGGCTTCTCCGGTGCTTG
GATCGGCAACCTGACGGTGCTGGAACCCTATCGACCGTTGTTCATCGGCGCGGCGCTAGTGGCGCTGTTCTTCGCCTGGAAGCGGATTTACCGG
CCCGTGCAGGCATGCAAGCCAGGTGAGGTCTGCGCGATTCCGCAGGTGCGCGCCACCTACAAGCTGATTTTCTGGATCGTGGCCGTGCTGGTCC
```

Or, upload file (Fasta format) : Choose File No file chosen

B

	TnCentral Accession	TE Name	TE Length	Host Organism	TE Family	Alignment (by score)
☐	Tn5060-AJ551280.1	Tn5060	8667	Pseudomonas sp. A19-1	Tn3	
☐	Tn20-AF457211.1	Tn20	8644	Escherichia coli	Tn3	
☐	Tn21-AF071413	Tn21	19672	Shigella flexneri	Tn3	
☐	TnAs3-CP000645.1	TnAs3	18735		Tn3	
☐	Tn21.2-MH626558	Tn21.2	35400	Salmonella enterica subsp. enterica serovar Typhimurium	Tn3	
☐	Tn2424-UGCJ01000005	Tn2424	26008	Escherichia coli NCTC11186	Tn3	
☐	Tn5086-CP054343	Tn5086	15341	Escherichia coli SCU-164	Tn3	
☐	Tn1935-MK797990	Tn1935	23364	Salmonella enterica subsp. enterica serovar Wien ZM3	Tn3	
☐	Tn21.1-MH257753	Tn21.1	21668	Salmonella enterica subsp. enterica serovar Typhimurium	Tn3	
☐	Tn2411-FN554766	Tn2411	18055	Escherichia coli 042	Tn3	
☐	Tn4-KY749247.1	Tn4	23009	Salmonella enterica subsp. enterica serovar Paratyphi B	Tn3	

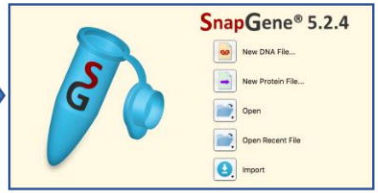
C

	Tn1935	Tn21	Tn2411	Tn4	TnAs3	Tn2424	Tn5086	Tn21.1	Tn21.2
aminoglycosides	GNAT_fam, aadA, aph3’-Ia	GNAT_fam, aadA				GNAT_fam, AAC(6')-Ia, aadA3	GNAT_fam	aadA, aadA2	
sulfonamides	sul1							sul3	
sulfones									
quaternary ammonium salts	qacEdelta1							qacL	
acridine dye									
phenicols					CAT	catB2		cmlA6	
diaminopyrimidines							dfrA7	dfrA12	
tetracyclines									tetR, tet(B), tetC_p
penams	bla-OXA-1			bla-TEM-1					
cephalosporins									
penems									
monobactams									

TnFinder Scripts, Literature

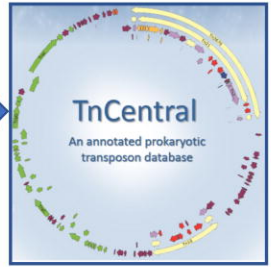
Transposon Feature Map (Image file)

GenBank File



Enhanced GenBank File

TnCentral Database



TnCentral Website

Curation Guidelines

QC: Validation by Second Curator Automated Format Checking (Perl Script)