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

Karen Ross,[a] Alessandro M. Varani,[b] Erik Snesrud,[c] Hongzhan Huang,[d] Danillo Oliveira Alvarenga,[b] Jian Zhang,[a] Cathy Wu,[d] Patrick McGann,[c] Mick Chandler[e]

aProtein Information Resource, Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, Washington, DC, USA
bSchool of Agricultural and Veterinary Sciences, Universidade Estadual Paulista, Jaboticabal, Sao Paulo, Brazil
cMultidrug-Resistant Organism Repository and Surveillance Network, Walter Reed Army Institute of Research, Silver Spring, Maryland, USA
dProtein Information Resource, Center for Bioinformatics and Computational Biology, University of Delaware, Newark, Delaware, USA
eDepartment of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, Washington, DC, USA

Karen Ross and Alessandro M. Varani contributed equally to this article. Author order was determined alphabetically.

ABSTRACT We describe here the structure and organization of TnCentral (https://tncentral.proteininformationresource.org/ [or the mirror link at https://tncentral.ncc.unesp.br/]), a web resource for prokaryotic transposable elements (TE). TnCentral currently contains ~400 carefully annotated TE, including transposons from the Tn3, Tn7, Tn402, and Tn554 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 TE. 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 specialized 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.

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 normal TE transposition activities and 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 whose use is intuitive and accessible to users with or without bioinformatics expertise.

Transposable elements (TE) are key facilitators of bacterial evolution and adaptation. They are central players in the emergence of antibiotic and heavy metal resistance and contribute 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 ongoing 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 references 3, 4, 5, 6, and 7) but do not themselves generally carry integrated passenger genes. There are a large number of significantly more complex TE (Fig. 1), which are arguably even more important in the global emergence of antibiotic resistance (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. Another important aspect of their impact is their ability to assemble passenger genes into resistance clusters (8, 9). While there appears to be a widespread 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] and ARDB [11]) or toxin/antitoxin gene pairs (TADB [12] and 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/ [or the mirror link at https://tncentral.ncc.unesp.br/]), a database of detailed structural and functional information on bacterial TE. In addition, 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. These databases allow users to rapidly analyze the landscape of TE in genomes (chromosomes and plasmids) isolated from clinical, veterinary, and environmental samples.

RESULTS

TnCentral website content. As of August 2021, TnCentral contains information on ~400 TE. About half of these TE are Tn3-family transposons. The remainder are integrons,
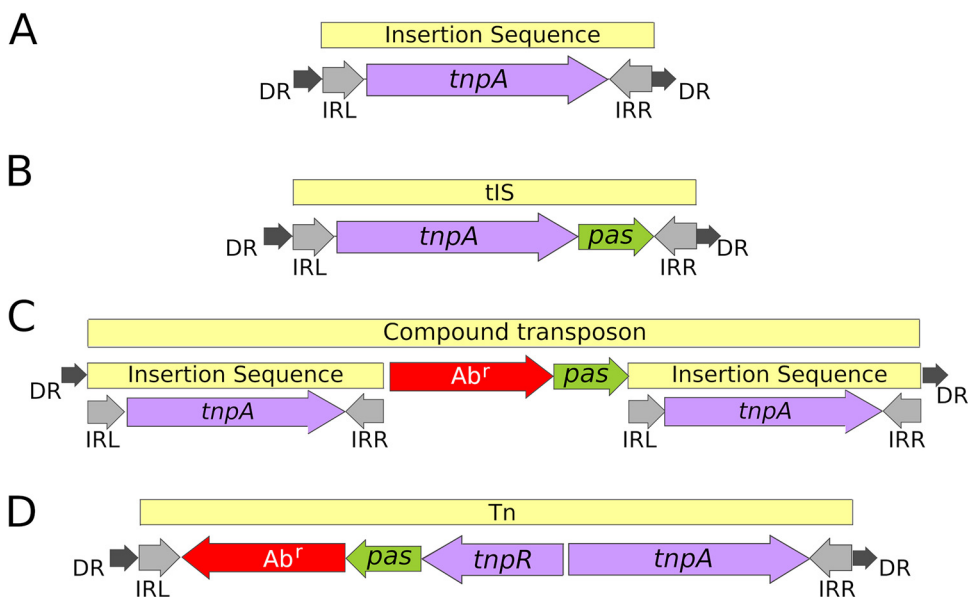
FIG 1 Structural arrangement of prokaryotic transposable elements. The TE is indicated by a pale-yellow horizontal bar at the top of each section. ORFs are shown as horizontal arrows with arrowheads indicating the direction of expression: purple, transposition-associated genes; red, antibiotic resistance genes; and green, other passenger genes. The inverted terminal repeats found at the ends of the majority of TE are shown as gray 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) The tIS (transporter IS) is structurally similar to an IS but contains passenger genes. These are presently restricted to the IS*1595* and IS*66* families. (C) Compound transposons are formed by two IS in either a direct or an 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 that are specific to each transposon 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*.

compound transposons, transposons from the Tn*402*, Tn*554*, and Tn*7* families, and IS that are associated with TE or are part of compound transposons (see Table S1 in the supplemental material). These 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. Although not considered *per se* as transposons, we have included the mobile integrons systems because of their important impact in shaping many transposons and their importance in the acquisition and dissemination of ABR.

TnCentral web portal. The TnCentral home page is designed to give the user easy access to the contents of TnCentral with a number of options (Fig. 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 (Fig. 2B).

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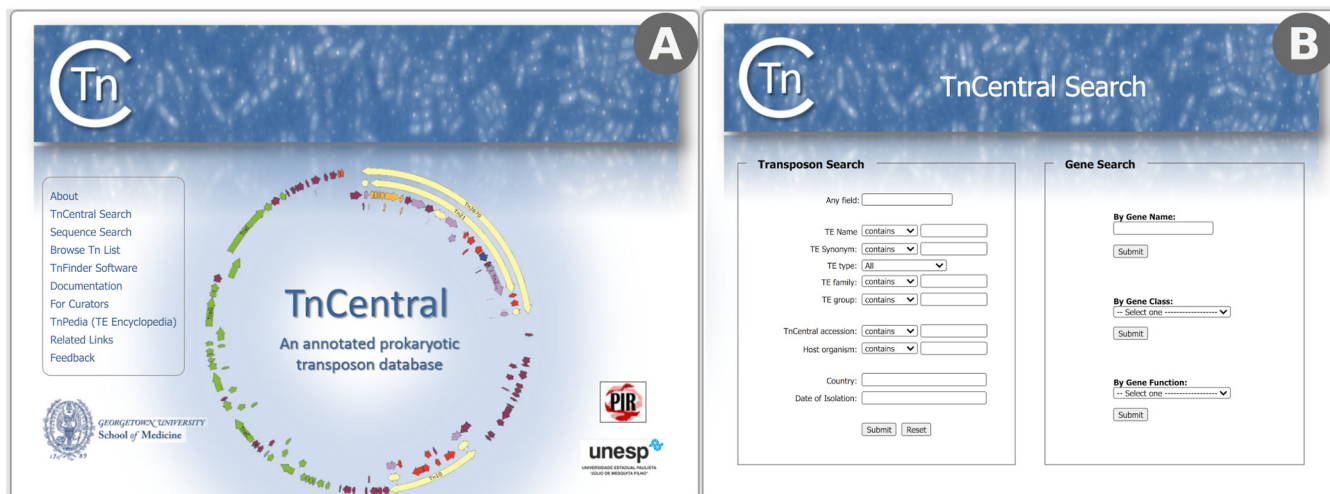
FIG 2 (A) TnCentral homepage showing clickable links to various TnCentral sections in the box on the left. (B) TnCentral search interface showing search choices for TE on the left and for transposition-related and passenger genes on the left.

(ii) 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 or 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 searches against the TnCentral database using BLAST (25, 26) (see use case 3, below). A text box for entering query sequences is provided. The BLAST tool automatically distinguishes between DNA and protein query sequences. BLAST parameters (e.g., maximum expect value, maximum number of results displayed, and scoring matrix for protein BLAST) can be customized using the menus in the options box, which is located below the query entry box. The page also provides links to several other BLAST interfaces where searches can be initiated. These include the ISfinder (https://isfinder.biotoul.fr/blast.php), NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi), Comprehensive Antibiotic Resistance (CARD [https://card.mcmaster.ca/analyze/blast]), and Toxin-Antitoxin (TADB [https://bioinfo-mml.sjtu.edu.cn/TADB2/]) databases.

The sequence search results display is currently quite basic, so it is best suited for simple searches such as querying with a transposon sequence to find related transposons in the database or querying with a protein sequence to find transposons in the database that encode the protein. In the future, we plan to enhance the display to facilitate more complex searches such as analyzing the transposon content of a plasmid or genome (see the Discussion). The results table, which is sorted by score, provides the TnCentral Accession for each significant hit hyperlinked to the corresponding entry page, information such as the transposon or protein name, and an alignment column. In the alignment column, the query sequence is represented by the width of the column, and the hits are shown as colored bars positioned according to the portion of the query sequence to which they align. The color of the bar indicates the strength of the match (red indicates strongest; black indicates weakest). Clicking on the alignment column brings up a page that shows the alignment in detail, as well as statistics such as the score and E value.

Browse Tn List. The browse Tn list option allows the user to browse the entire TnCentral database.

Transposon entry page. All of the search and browse options provide links to entry pages for each TE (Fig. 3), which provide detailed information about TE features and origins. The page includes various sections: (i) host information (the host species, strain, and plasmid/chromosome in which the transposon was found, as well as the date and geographic location of the isolate) (Fig. 3, section 1); (ii) a
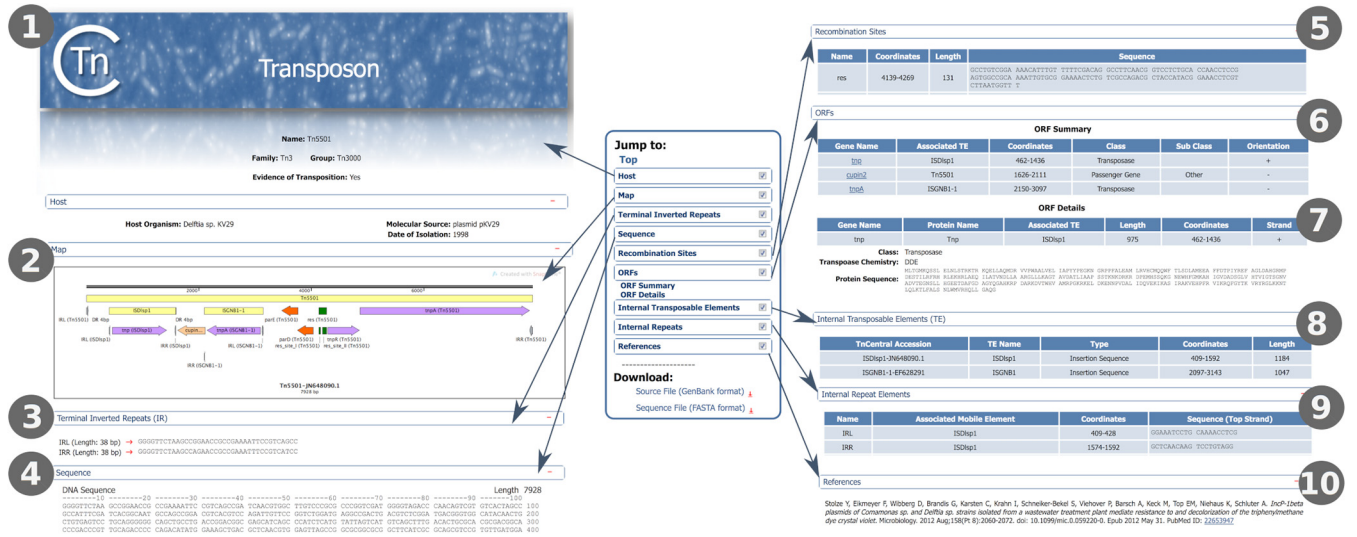
FIG 3 TnCentral TE entry page. Numbers 1 to 10 indicate various sections of the entry page (see the text for details).

graphic representation of the annotated sequence with color-coded features (Fig. 3, section 2); (iii) terminal inverted repeats (Fig. 3, section 3); (iv) the DNA sequence (Fig. 3, section 4); (v) internal recombination sites (e.g., *res* sites), including their coordinates, length, and DNA sequence (Fig. 3, section 5); (vi) open reading frame (ORF) summary, which includes all protein coding genes in the order in which they appear, 5'–3', in the TE sequence, the element with which they are associated (important for nested TE in which one TE is inserted into another), their coordinates, their class (e.g., transposase, accessory gene, passenger gene) and subclass (e.g., antibiotic resistance or heavy metal resistance), and their relative orientation within the TE (Fig. 3, section 6); (vii) a detailed ORF description, including the amino acid sequence (Fig. 3, section 7); (viii) if applicable, a table of internal transposable elements (TE inserted in the main element), including the name, type, location, and length (Fig. 3, section 8); (ix) if applicable, a table of internal repeats (i.e., repeat elements, other than the terminal inverted repeats, that are found within the TE), including the associated TE, coordinates, and DNA sequence (Fig. 3, section 9); and (x) bibliographic references with direct links to PubMed (Fig. 3, section 10) (27). Each section can be collapsed using a button on the right-hand side of the section heading. Sections can be viewed either by scrolling down on the page or by clicking on the section name in the menu located on the left side of the page. Sequence files in FASTA and GenBank format can be 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 to identify transposons. These scripts help users to screen data sets containing large numbers of genomic sequences using their own servers to identify potential candidate transposons, which can then be manually curated.

The Tn*3* 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. Tn3_finder allows for the concurrent analysis of multiple genomes by multithreading.
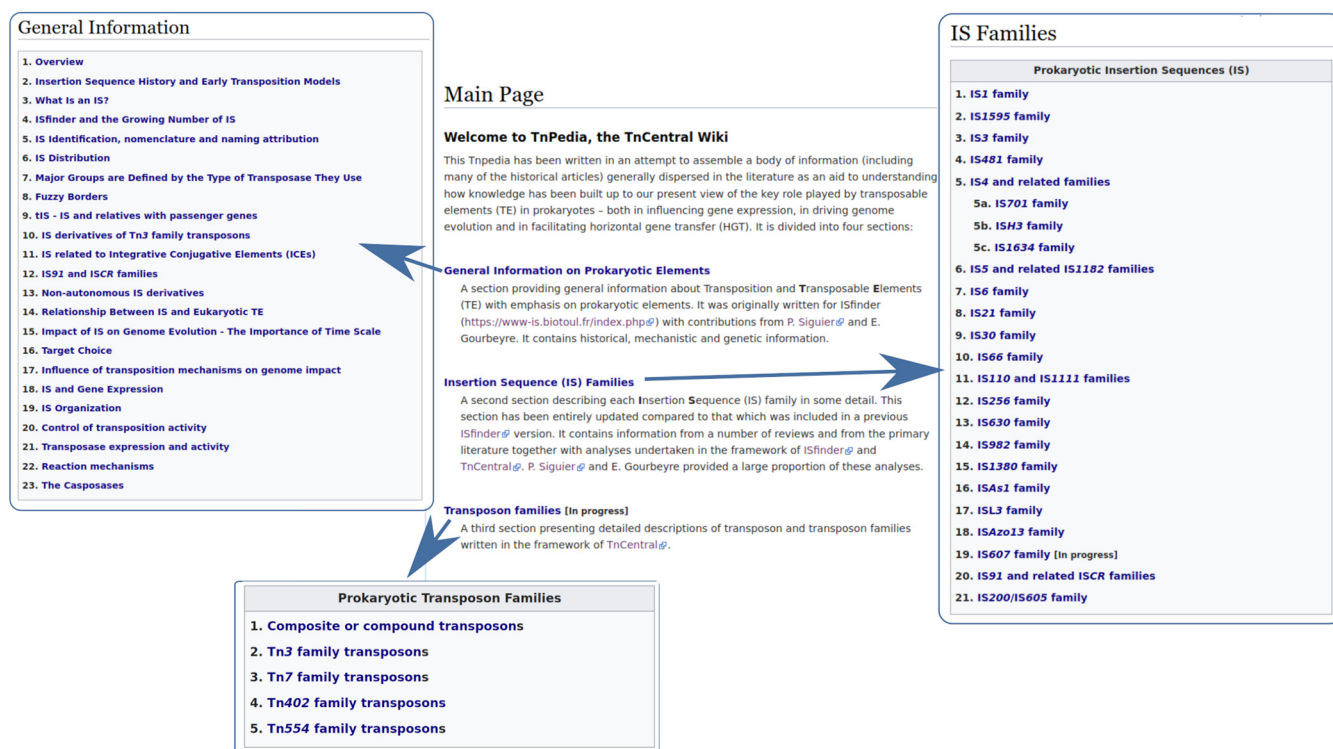
General Information

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 Tn3 family transposons
11. IS related to Integrative Conjugative Elements (ICEs)
12. IS91 and ISCR 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 Transposable Elements (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 Insertion Sequence (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. IS1 family
2. IS1595 family
3. IS3 family
4. IS481 family
5. IS4 and related families
  5a. IS701 family
  5b. ISH3 family
  5c. IS1634 family
6. IS5 and related IS1182 families
7. IS6 family
8. IS21 family
9. IS30 family
10. IS66 family
11. IS110 and IS1111 families
12. IS256 family
13. IS630 family
14. IS982 family
15. IS1380 family
16. ISAs1 family
17. ISL3 family
18. ISAzo13 family
19. IS607 family [In progress]
20. IS91 and related ISCR families
21. IS200/IS605 family

Prokaryotic Transposon Families

1. Composite or compound transposons
2. Tn3 family transposons
3. Tn7 family transposons
4. Tn402 family transposons
5. Tn554 family transposons

FIG 4 Main sections of TnPedia, a TnCentral-related wiki compiling information on prokaryotic transposable elements. Only three of four sections (general information, IS families, and transposon families) are illustrated. A fourth section, the transposition glossary, is under construction.

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 the parameters for similarity, coverage, distance, and flanking regions can be adjusted by the user. The 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 also produces report files and pre-annotated GenBank files on which the recommended manual curation should be performed.

Documentation. This section, which can be downloaded as a .pdf file, provides a short background description of transposons and TnCentral, together with a short description of the curation workflow and planned future developments.

For Curators. This section provides a detailed description of the curation workflow used to generate the annotated TnCentral data.

TnPedia. TnCentral provides access from the homepage to TnPedia, an online knowledge base that contains information concerning transposition in prokaryotes. TnPedia was developed using MediaWiki (https://www.mediawiki.org) and can also be accessed directly (https://tnpedia.fcav.unesp.br/). It is structured into three main sections: general information, IS families, and transposon families (Fig. 4).

The general information section provides a series of clickable sections with an extensive bibliography and direct links to the articles in PubMed. It includes a historical perspective, definitions, and descriptions of a variety of prokaryotic TE, the basic mechanisms involved in their movement, and the enzymes involved in these processes. It also contains information describing their impact on their host genomes and how their activities are controlled.

The IS families section consists of individual chapters describing each of the ~25 IS families in detail and covers, where possible, the identification of the founding members, their organization, distribution, variability, and phylogenetic relationships; the regulation of their transposition; the impact on their host genomes; and their transposition mechanisms, including genetic, biochemical, and structural studies.

The transposon families section describes each transposon family with information similar to that included in the IS family descriptions but, in addition, includes a detailed description of their structures and the passenger genes that they may carry.

Examples of TnCentral use. (i) 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 that shares all of the important catalytic residues of a typical tyrosine site-specific recombinase (28, 29). *Tn554*-family transposons 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 (Fig. 5A). Fourteen Tn*554* family members were found (only 10 of which are shown in Fig. 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 and country) (Fig. 5B). The results for two of the Tn*554* transposons (Tn*558.3* and Tn*559*) are shown in Fig. 5C. Both transposons have the three-part transposition module (*tnpA*, *tnpB*, and *tnpC*) characteristic of the family. However, the two transposons are quite diverse in their passenger genes. Tn*558.3* has a 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 in this example, the flexible search results page makes it easy to compare features across multiple transposons.

(ii) 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 transposons, we used the TnCentral gene search function, selecting "Passenger Gene" from the gene class pulldown menu and "Toxin" from the gene subclass pulldown menu (Fig. 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 subclass menu set to "Antitoxin" (Fig. 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* (Fig. 6B, yellow box; Fig. 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
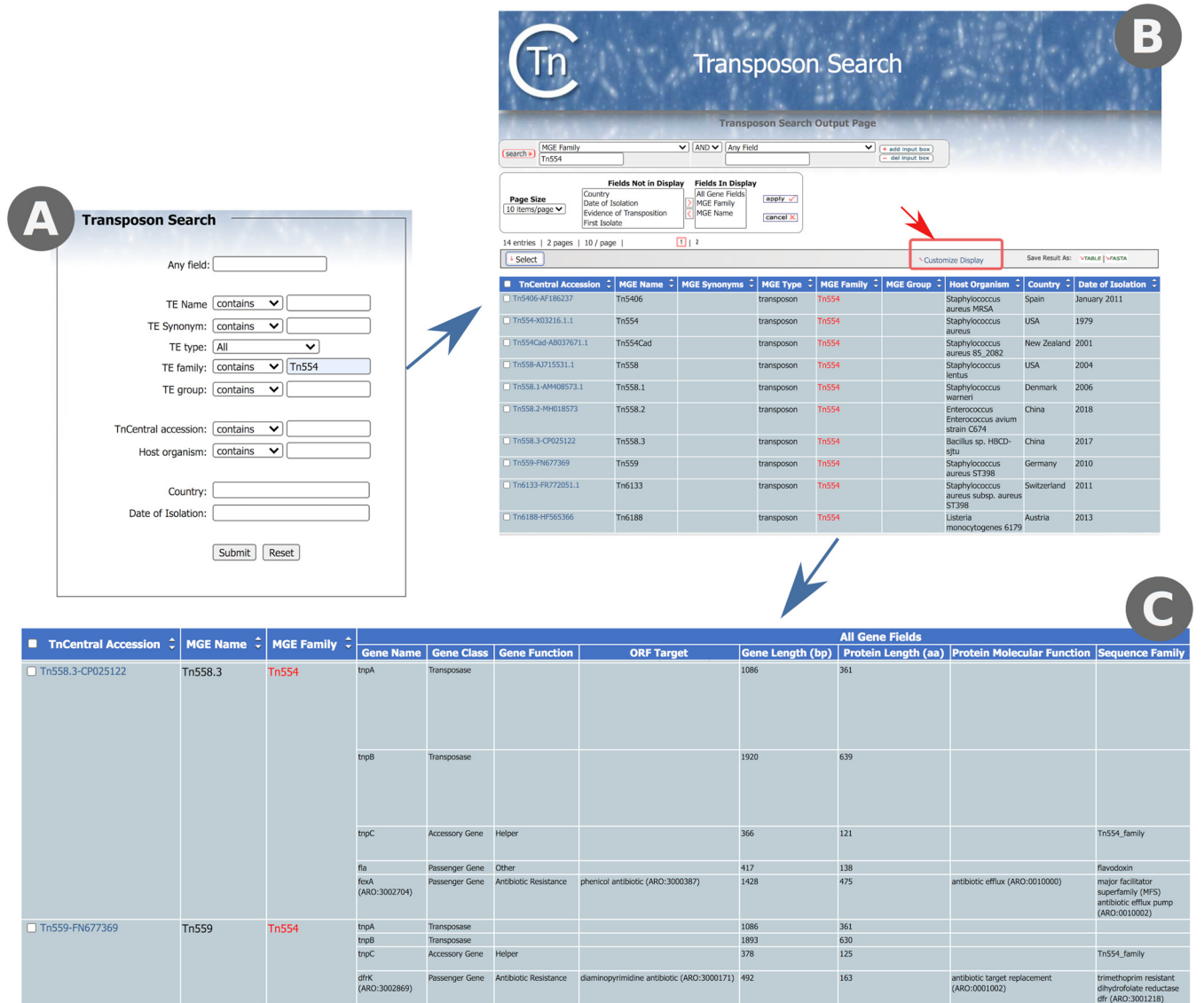
FIG 5 Comparing protein coding genes in Tn*554* family members. (A) TnCentral transposon search interface, showing a search for Tn*554* in the TE family field. (B) Interface for customizing the columns in the search results display. Clicking on "customize display" (red box) opens the interface. (C) Partial Tn*554* family search results after customization to show information on protein coding genes (All Gene Fields).

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* (see Fig. S1 in the supplemental material).

(iii) 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 Tn*21* arose from simpler ancestor transposons (see, for example, references 40 and 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*, *qacE*Δ*1*, and *aadA*) in the middle (see Fig. S2). 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 Tn*21*Δ in reference 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

FIG 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).

explore the Tn*21* subfamily, we performed a TnCentral sequence search (BLAST) using the putative ancestral Tn*5060* sequence (Fig. 7A). In addition to Tn*5060* itself, we identified 10 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*; Fig. S2) 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 subregions that align to Tn*5060*. This suggests that these transposons arose from Tn*5060* via the insertion of other sequences. For example, Tn*21* has two subregions that align with the Tn*5060* sequence: bases 1 to 4633 of Tn*5060* align with bases 1 to 4633 of Tn*21* (Fig. 7B, left red bar in the alignment column for Tn*21*) and bases 4629 to 8667 of Tn*5060* align with bases 15634 to 19635 of Tn*21* (Fig. 7B, right red bar in the alignment column for Tn*21*). The region of Tn*21* from 4633 to 15633 does not align with Tn*5060* because it contains an insertion of the In2 integron in the *urfM* gene (see Fig. S2A and E).

We compared the antibiotic resistance profiles of the 10 transposons by inspecting their TnCentral entry pages. Tn*20*, like Tn*5060*, carries no ABR genes. The other nine

A Sequence Search (BLAST)

Job Name: [   ]

Database
TnCentral

Other Sequence Search Services:
iSfinder: ○   CARD: ○   Toxin/Antitoxin: ○   NCBI: ○

Query Sequence

Enter a sequence here:
```
>Tn5060
GGGGGCACCTCAGAAAACGGAAAATAAAGCACGCTAAGGCATAGCTGACCTTGCCAGGCCTGCTTCGCCCTGTAGTGACGCGATCAACGGG
CAGGAAACATTCCCCTTTCGTGCATGGCAGGCGCACACGAGTTCAGACAGCACGGTTTCCATGCGGCGCCAAGTCGGCCATCTTCTCGCGCA
CGTCCTTGAGCTTGTGTTCGGCCAGGCTGCTGGCCTCCTCGCAGTGGGTGCCATCGTCGAGCCGCAACAGCTCGGCAATCTCGTCCAGACT
GAACCCCAGCCGCTGTGCCGATTTCACGAATTTCACCCGAACCACGTCCGCCTCCCCATAGCGGCGGATGCTGCCGTAAGGCTTGTCCGGT
TCCGGCAACAGGCCCTTGCGCTGATAGAAGCGGATTGTCTCCACGCTGACCCCGGCCGCCTTGGCAAAAACGCCAATGGTCAGGTTTTCCA
AATTATTTTCCATATCGCTTGACTCCGTACATGAGTACGGAAGTAAGGTTACGCTATCCAATCCAAATTCAAAAGGGCCAACGTATGTCTG
AACCACAAAACGGGCGCGGTGCGCTCTTCGCCGGCGGGCTGGCCGCCATTCTTGCATCGACCTGCTGCCTGGGGCCGCTAGTACTGGTCGC
CCTGGGCTTCTCCGGTGCTTGGATCGGCAACCTGACGGTGCTGGAACCCTATCGACCGTTGTTCATCGGCGCGGCGCTAGTGGCGCTGTTC
TTCGCCTGGAAGCGGATTTACCGGCCCGTGCAGGCATGCAAGCCAGGTGAGGTCTGCGCGATTCCGCAGGTGCGCGCCACCTACAAGCTGA
```

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

Options

Alignment view: [pairwise   ▾]

≫

Composition-based ☑
Choose filter   ☐ Low complexity   ☐ Mask for lookup table only   ☐ Mask lower case
Expect   [10 ▾]
Word size   [11 ▾]   Adjust Gap Cost to
Matrix   [BLOSUM62 (Gap Costs: 11,1) ▾]   open [0 ▾] extend [-2 ▾]
No.of hits to display [250 ▾]

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									

FIG 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. The query entry box and the customizable BLAST

(Continued on next page)

transposons carry ABR genes targeting aminoglycosides, sulfones, sulfonamides, and quaternary ammonium salts (Fig. 7C). Other resistances found in a subset of the six include acridine dye (Tn1935, Tn21, Tn2411, Tn4, TnAs3, Tn2424, Tn5086), carbapenams (Tn1935 and Tn4), cephalosporins (Tn1935 and Tn4), carbapenems (Tn4), monobactams (Tn4), phenicols (TnAs3, Tn2424, Tn21.1, Tn21.2), diaminopyrimidines (Tn5086, Tn21.1, Tn21.2), and tetracyclines (Tn21.2). Interestingly, in some cases where the transposons have resistances in common, they are conferred by different genes (Fig. 7C). For example, phenicol resistance is conferred by *CAT* in TnAs3, *catB2 in* Tn2424, and *cmlA6* in Tn21.1 and Tn21.2. Similarly, sulfonamide and sulfone resistance is conferred by *sul1* in all of the antibiotic-resistant family members except for Tn21.1 and Tn21.2, where 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 the integron region (e.g., Tn3.1 in Tn4; Fig. S2).

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. 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], and ISEscan [50]) or they provide a comparative view of IS mobilization events (e.g., ISCompare [51]). These annotation tools are only as good as their underlying TE databases. ISfinder, which includes nearly 6,000 individual examples of IS classified in distinct families and subfamilies according to their transposition mechanism and struc-

FIG 7 Legend (Continued)

parameters are shown. The display of the parameters can be hidden/shown by clicking on the red double arrowhead. (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 Tn5060. (C) ABR genes and targeted antibiotic classes in Tn21 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.
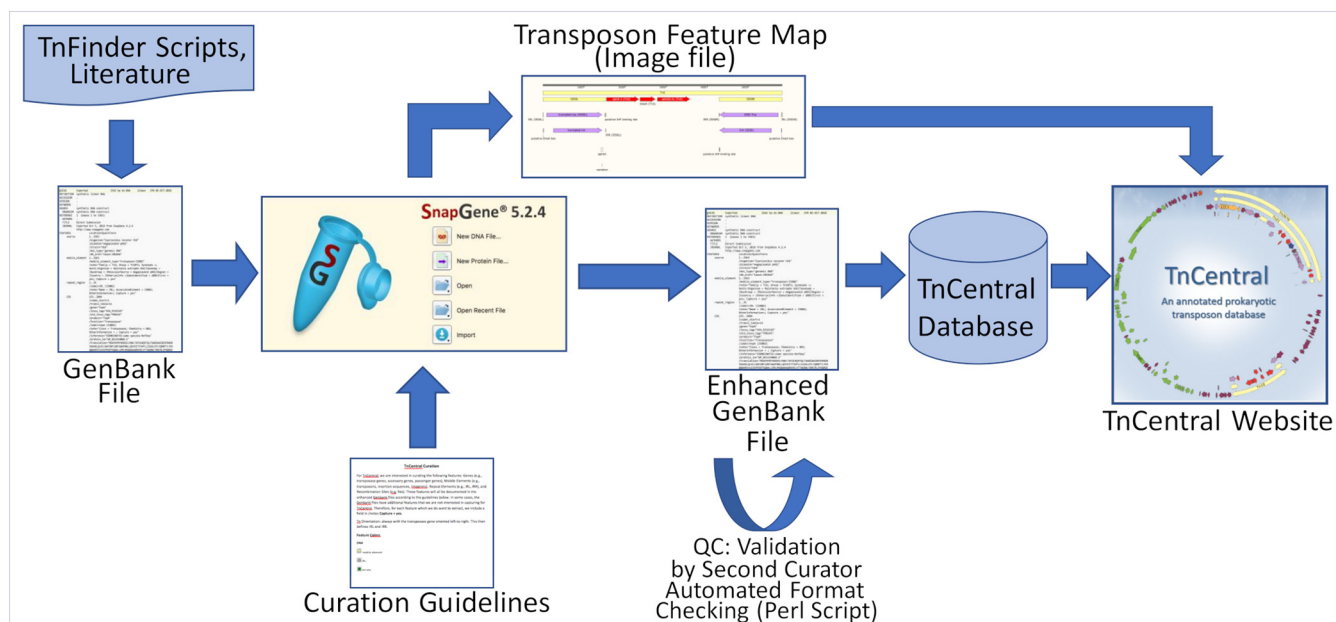
FIG 8 TnCentral curation workflow (see the text for a description).

tural organization, provides such a rigorous framework for IS, and has been incorporated into a number of annotation pipelines (e.g., ISsaga [43] and 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 that will allow users to use SnapGene, 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. For example, we will improve the graphics to show the alignment of multiple hits (i.e., multiple transposons) along the query sequence, enable tooltips that will display the coordinates of the alignment when hovering over a hit in the graphical display, and display the features (e.g., passenger genes or repeat elements) that are included in each hit. 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.

MATERIALS AND METHODS

Curation workflow. The TnCentral curation workflow is depicted in Fig. 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. 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" section 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 various degrees of completeness. We have therefore elected to include all class 1 integrons as members of the Tn*402* family (see 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 the 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 are 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.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
FIG S1, TIF file, 0.3 MB.
FIG S2, TIF file, 1.4 MB.
TABLE S1, XLSX file, 0.04 MB.

REFERENCES

1. Spellberg B, Guidos R, Gilbert D, Bradley J, Boucher HW, Scheld WM, Bartlett JG, Edwards J, Infectious Diseases Society of America. 2008. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clin Infect Dis 46:155–164. https://doi.org/10.1086/524891.
2. O'Neill J. 2016. Tackling drug-resistant infections globally: final report and recommendations. Welcome Foundation HM Government UK, London, United Kingdom.
3. Craig NL (ed). 2015. Mobile DNA III. American Society of Microbiology, Washington, DC.
4. Mahillon J, Chandler M. 1998. Insertion sequences. Microbiol Mol Biol Rev 62:725–774. https://doi.org/10.1128/MMBR.62.3.725-774.1998.
5. Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M. 2015. Everyman's guide to bacterial insertion sequences. Microbiol Spectr 3:MDNA3. https://doi.org/10.1128/microbiolspec.MDNA3-0030-2014.
6. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev 38:865–891. https://doi.org/10.1111/1574-6976.12067.
7. Vandecraen J, Chandler M, Aertsen A, Van Houdt R. 2017. The impact of insertion sequences on bacterial genome plasticity and adaptability. Crit Rev Microbiol 43:709–730. https://doi.org/10.1080/1040841X.2017.1303661.
8. He S, Chandler M, Varani AM, Hickman AB, Dekker JP, Dyda F. 2016. Mechanisms of evolution in high-consequence drug resistance plasmids. mBio 7:e01987-16. https://doi.org/10.1128/mBio.01987-16.

9. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, Dyda F. 2015. Insertion sequence IS26 reorganizes plasmids in clinically isolated multidrug-resistant bacteria by replicative transposition. mBio 6:e00762. https://doi.org/10.1128/mBio.00762-15.

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. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935.

11. Liu B, Pop M. 2009. ARDB: Antibiotic Resistance Genes Database. Nucleic Acids Res 37:D443–D447. https://doi.org/10.1093/nar/gkn656.

12. Xie Y, Wei Y, Shen Y, Li X, Zhou H, Tai C, Deng Z, Ou H-Y. 2018. TADB 2.0: an updated database of bacterial type II toxin-antitoxin loci. Nucleic Acids Res 46:D749–D753. https://doi.org/10.1093/nar/gkx1033.

13. Akarsu H, Bordes P, Mansour M, Bigot D-J, Genevaux P, Falquet L. 2019. TASmania: a bacterial toxin-antitoxin systems database. PLoS Comput Biol 15:e1006946. https://doi.org/10.1371/journal.pcbi.1006946.

14. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36. https://doi.org/10.1093/nar/gkj014.

15. Moura A, Soares M, Pereira C, Leitão N, Henriques I, Correia A. 2009. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics 25:1096–1098. https://doi.org/10.1093/bioinformatics/btp105.

16. Bi D, Xu Z, Harrison EM, Tai C, Wei Y, He X, Jia S, Deng Z, Rajakumar K, Ou H-Y. 2012. ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. Nucleic Acids Res 40:D621–D626. https://doi.org/10.1093/nar/gkr846.

17. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, Tai C, Deng Z, Ou H-Y. 2019. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. Nucleic Acids Res 47:D660–D665. https://doi.org/10.1093/nar/gky1123.

18. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895–3903. https://doi.org/10.1128/AAC.02412-14.

19. Leplae R, Lima-Mendez G, Toussaint A. 2010. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res 38:D57–D61. https://doi.org/10.1093/nar/gkp938.

20. Leplae R, Hebrant A, Wodak SJ, Toussaint A. 2004. ACLAME: a CLAssification of Mobile genetic Elements. Nucleic Acids Res 32:D45–D49. https://doi.org/10.1093/nar/gkh084.

21. Leplae R, Lima-Mendez G, Toussaint A. 2006. A first global analysis of plasmid encoded proteins in the ACLAME database. FEMS Microbiol Rev 30:980–994. https://doi.org/10.1111/j.1574-6976.2006.00044.x.

22. Lima-Mendez G, Oliveira Alvarenga D, Ross K, Hallet B, Van Melderen L, Varani AM, Chandler M. 2020. Toxin-antitoxin gene pairs found in Tn3 family transposons appear to be an integral part of the transposition module. mBio 11:e00452-20. https://doi.org/10.1128/mBio.00452-20.

23. Szuplewska M, Czarnecki J, Bartosik D. 2014. Autonomous and non-autonomous Tn3-family transposons and their role in the evolution of mobile genetic elements. Mob Genet Elements 4:1–4. https://doi.org/10.1080/2159256X.2014.998537.

24. Loftie-Eaton W, Yano H, Burleigh S, Simmons RS, Hughes JM, Rogers LM, Hunter SS, Settles ML, Forney LJ, Ponciano JM, Top EM. 2016. Evolutionary paths that expand plasmid host-range: implications for spread of antibiotic resistance. Mol Biol Evol 33:885–897. https://doi.org/10.1093/molbev/msv339.

25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

26. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

27. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, Holmes JB, Kim S, Kimchi A, Kitts PA, Lathrop S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD, Ostell J. 2019. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 47:D23–D28. https://doi.org/10.1093/nar/gky1069.

28. Murphy E, Huwyler L, de Freire Bastos MC. 1985. Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. EMBO J 4:3357–3365. https://doi.org/10.1002/j.1460-2075.1985.tb04089.x.

29. Bastos MC, Murphy E. 1988. Transposon Tn554 encodes three products required for transposition. EMBO J 7:2935–2941. https://doi.org/10.1002/j.1460-2075.1988.tb03152.x.

30. Krolewski JJ, Murphy E, Novick RP, Rush MG. 1981. Site specificity of the chromosomal insertion of Staphylococcus aureus transposon Tn554. J Mol Biol 152:19–33. https://doi.org/10.1016/0022-2836(81)90093-0.

31. Murphy E, Reinheimer E, Huwyler L. 1991. Mutational analysis of att554, the target of the site-specific transposon Tn554. Plasmid 26:20–29. https://doi.org/10.1016/0147-619x(91)90033-s.

32. Haroche J, Allignet J, El Solh N. 2002. Tn5406, a new staphylococcal transposon conferring resistance to streptogramin a and related compounds including dalfopristin. Antimicrob Agents Chemother 46:2337–2343. https://doi.org/10.1128/AAC.46.8.2337-2343.2002.

33. Kehrenberg C, Schwarz S. 2005. Florfenicol-chloramphenicol exporter gene fexA is part of the novel transposon Tn558. Antimicrob Agents Chemother 49:813–815. https://doi.org/10.1128/AAC.49.2.813-815.2005.

34. Kadlec K, Schwarz S. 2010. Identification of the novel dfrK-carrying transposon Tn559 in a porcine methicillin-susceptible Staphylococcus aureus ST398 strain. Antimicrob Agents Chemother 54:3475–3477. https://doi.org/10.1128/AAC.00464-10.

35. Li D, Li X-Y, Schwarz S, Yang M, Zhang S-M, Hao W, Du X-D. 2019. Tn6674 is a novel enterococcal optrA-carrying multiresistance transposon of the Tn554 family. Antimicrob Agents Chemother 63:e00809-19. https://doi.org/10.1128/AAC.00809-19.

36. Schwendener S, Perreten V. 2011. New transposon Tn6133 in methicillin-resistant Staphylococcus aureus ST398 contains vga(E), a novel streptogramin A, pleuromutilin, and lincosamide resistance gene. Antimicrob Agents Chemother 55:4900–4904. https://doi.org/10.1128/AAC.00528-11.

37. Hayes F, Van Melderen L. 2011. Toxins-antitoxins: diversity, evolution, and function. Crit Rev Biochem Mol Biol 46:386–408. https://doi.org/10.3109/10409238.2011.600437.

38. De La Cruz F, Grinsted J. 1982. Genetic and molecular characterization of Tn21, a multiple resistance transposon from R100.1. J Bacteriol 151:222–228. https://doi.org/10.1128/jb.151.1.222-228.1982.

39. Grinsted J, de la Cruz F, Schmitt R. 1990. The Tn21 subgroup of bacterial transposable elements. Plasmid 24:163–189. https://doi.org/10.1016/0147-619x(90)90001-s.

40. Liebert CA, Hall RM, Summers AO. 1999. Transposon Tn21, flagship of the floating genome. Microbiol Mol Biol Rev 63:507–522. https://doi.org/10.1128/MMBR.63.3.507-522.1999.

41. Tanaka M, Yamamoto T, Sawai T. 1983. Evolution of complex resistance transposons from an ancestral mercury transposon. J Bacteriol 153:1432–1438. https://doi.org/10.1128/jb.153.3.1432-1438.1983.

42. Kholodii G, Mindlin S, Petrova M, Minakhina S. 2003. Tn5060 from the Siberian permafrost is most closely related to the ancestor of Tn21 prior to integron acquisition. FEMS Microbiol Lett 226:251–255. https://doi.org/10.1016/S0378-1097(03)00559-7.

43. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol 12:R30. https://doi.org/10.1186/gb-2011-12-3-r30.

44. Biswas A, Gauthier DT, Ranjan D, Zubair M. 2015. ISQuest: finding insertion sequences in prokaryotic sequence fragment data. Bioinformatics 31:3406–3412. https://doi.org/10.1093/bioinformatics/btv388.

45. Nakagome M, Solovieva E, Takahashi A, Yasue H, Hirochika H, Miyao A. 2014. Transposon Insertion Finder (TIF): a novel program for detection of de novo transpositions of transposable elements. BMC Bioinformatics 15:71. https://doi.org/10.1186/1471-2105-15-71.

46. Hawkey J, Hamidian M, Wick RR, Edwards DJ, Billman-Jacobe H, Hall RM, Holt KE. 2015. ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. BMC Genomics 16:667. https://doi.org/10.1186/s12864-015-1860-2.

47. Treepong P, Guyeux C, Meunier A, Couchoud C, Hocquet D, Valot B. 2018. panISa: ab initio detection of insertion sequences in bacterial genomes from short read sequence data. Bioinformatics 34:3795–3800. https://doi.org/10.1093/bioinformatics/bty479.

48. Robinson DG, Lee M-C, Marx CJ. 2012. OASIS: an automated program for global investigation of bacterial and archaeal insertion sequences. Nucleic Acids Res 40:e174. https://doi.org/10.1093/nar/gks778.

49. Adams MD, Bishop B, Wright MS. 2016. Quantitative assessment of insertion sequence impact on bacterial genome architecture. Microb Genom 2:e000062. https://doi.org/10.1099/mgen.0.000062.

50. Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics 33:3340–3347. https://doi.org/10.1093/bioinformatics/btx433.

51. Mogro EG, Ambrosis N, Lozano MJ. 2020. Easy identification of insertion sequence mobilization events in related bacterial strains with ISCompare. BioRxiv https://www.biorxiv.org/content/10.1101/2020.10.16.342287v4.

52. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN. 2021. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. J Antimicrob Chemother 76:101–109. https://doi.org/10.1093/jac/dkaa390.

53. Tansirichaiya S, Rahman MA, Roberts AP. 2019. The transposon registry. Mob DNA 10:40. https://doi.org/10.1186/s13100-019-0182-3.

54. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432. https://doi.org/10.1093/nar/gky995.

55. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong S-Y, Finn RD. 2019. InterPro in 2019: improving coverage, classification, and access to protein sequence annotations. Nucleic Acids Res 47:D351–D360. https://doi.org/10.1093/nar/gky1100.