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

Karen Ross, Alessandro M. Varani, Erik Snesrud, Hongzhan Huang, Danillo Oliveira Alvarenga, Jian Zhang, Cathy Wu, Patrick McGann, Mick Chandler

*School of Agricultural and Veterinary Sciences, Universidade Estadual Paulista, Jaboatão, São Paulo, Brazil
†Protein Information Resource, Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, Washington, DC, USA
‡Multidrug-Resistant Organism Repository and Surveillance Network, Walter Reed Army Institute of Research, Silver Spring, Maryland, USA
§Protein Information Resource, Center for Bioinformatics and Computational Biology, University of Delaware, Newark, Delaware, 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,
compound transposons, transposons from the Tn402, Tn554, and Tn7 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).

1. **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).
(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 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.
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 Tn3 Transposon Finder (Tn3_finder) performs the automatic prediction of transposable elements of the Tn3 family in bacteria and archaea. It compares user-provided bacterial and archael genome sequences to custom Tn3 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.
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 Tn554 family members.** The Tn554 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](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 Tn554 family members side by side, we searched for Tn554 in the TE family field of the transposon search interface (Fig. 5A). Fourteen Tn554 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 Tn554 transposons (Tn558.3 and Tn559) 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. Tn558.3 has a gene called fla, which contains a flavodoxin-like domain, and the ABR gene fexA, which confers resistance to phenicol antibiotics. Tn559 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 Tn3 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 Tn3-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, TnSku1 (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 TnXca1, which has two TA pairs, and TnS501.5, which has a parD antitoxin gene and no
toxin gene. The majority of Tn5501 derivatives in TnCentral have a parE toxin gene as well as the parD antitoxin, suggesting that Tn5501.5 may have undergone a deletion in the region containing parE (see Fig. S1 in the supplemental material).

(iii) Use case 3: Tn21 and its relatives. Tn21 is the canonical member of a subfamily of Tn3 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 (see, for example, references 40 and 41). Tn21 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, qacED1, 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 Tn21 (the hypothetical Tn21 backbone Tn21D in reference 40) was discovered (42). This transposon, Tn5060, was proposed to be the ancestor of Tn21 (42). Tn21 also has numerous relatives that carry different combinations of antibiotic resistance genes within and outside the integron. To
explore the Tn21 subfamily, we performed a TnCentral sequence search (BLAST) using the putative ancestral Tn5060 sequence (Fig. 7A). In addition to Tn5060 itself, we identified 10 transposons in the database (Tn20, Tn21, Tn21.1, Tn21.2, Tn5060, Tn2411, Tn2424, Tn4, Tn1935, and TnAs3; Fig. S2) that contain all (or nearly all) of the Tn5060 sequence. With the exception of Tn20, which is almost identical to Tn5060 (99.5%), these transposons have two or more discontinuous subregions that align to Tn5060. This suggests that these transposons arose from Tn5060 via the insertion of other sequences. For example, Tn21 has two subregions that align with the Tn5060 sequence: bases 1 to 4633 of Tn5060 align with bases 1 to 4633 of Tn21 (Fig. 7B, left red bar in the alignment column for Tn21) and bases 4629 to 8667 of Tn5060 align with bases 15634 to 19635 of Tn21 (Fig. 7B, right red bar in the alignment column for Tn21). The region of Tn21 from 4633 to 15633 does not align with Tn5060 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. Tn20, like Tn5060, carries no ABR genes. The other nine
FIG 7 Analysis of ABR in Tn21 Relatives. (A) TnCentral sequence search using the sequence of Tn5060, the proposed ancestor of Tn21, 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 (Tn\textsuperscript{1935}, Tn\textsubscript{21}, Tn\textsubscript{2411}, Tn\textsubscript{4}, Tn\textsubscript{As3}, Tn\textsubscript{2424}, Tn\textsubscript{S086}), carbapenams (Tn\textsubscript{1935} and Tn\textsubscript{4}), cephalosporins (Tn\textsubscript{1935} and Tn\textsubscript{4}), carbapenems (Tn\textsubscript{4}), phenicolis (Tn\textsubscript{As3}, Tn\textsubscript{2424}, Tn\textsubscript{21.1}, Tn\textsubscript{21.2}), diaminopyrimidines (Tn\textsubscript{S086}, Tn\textsubscript{21.1}, Tn\textsubscript{21.2}), and tetracyclines (Tn\textsubscript{21.2}). 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 \textit{CAT} in TnAs\textsubscript{3}, \textit{catB2} in Tn\textsubscript{2424}, and \textit{cmi6A} in Tn\textsubscript{21.1} and Tn\textsubscript{21.2}. Similarly, sulfonamide and sulfone resistance is conferred by \textit{sul1} in all of the antibiotic-resistant family members except for Tn\textsubscript{21.1} and Tn\textsubscript{21.2}, where resistances are conferred by \textit{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., Tn\textsubscript{3.1} in Tn\textsubscript{4}; 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 Tfnd 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 Tfnd, 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 \textit{ab initio} prediction (e.g., OASIS [48], ISseeker [49], and ISScan [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. Tfnd, 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 Tn\textsubscript{S060}. (C) ABR genes and targeted antibiotic classes in Tn\textsubscript{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.
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

FIG 8 TnCentral curation workflow (see the text for a description).
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 (S3) accession number is provided as a synonym. There is some ambiguity in the literature concerning class 1 integrons and members of the Tn4022 transposon family. Class 1 integrons appear to be derivatives of this transposon family and include members with a range of Tn4022 transposition genes with various degrees of completeness. We have therefore elected to include all class 1 integrons as members of the Tn4022 family (see Table S1). ISfinder classification is used for the individual IS and, in the case of compound transposons, the group to which they 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 (S4) and InterPro resources (S5) are used to define sequence family information.

**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.

**ACKNOWLEDGMENTS**

We thank John Dekker (NIAID, NIH, Bethesda, MD), Fred Dyda and Alison Hickman (NIDDK, NIH, Bethesda, MD), Patricia Sigueri (CNRS, Toulouse, France), Susu He (Nanjing University Medical School, Nanjing, China), Laurence van Melder (Université Libre de Bruxelles, Brussels, Belgium), and Gipsi Lima-Mendez and Bernard Hallet (Université de Louvain la Neuve, Louvain la Neuve, Belgium) for helpful discussions. We also thank 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 their contributions to the TnCentral website implementation. TE features and sequence information are extracted from the enhanced GenBank files. TE feature information is used for the search and 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.

**REFERENCES**


TnCentral: a prokaryotic TE resource


