

TADB 3.0: an updated database of bacterial toxin–antitoxin loci and associated mobile genetic elements

Jiahao Guan¹, Yongkui Chen¹, Ying-Xian Goh¹, Meng Wang¹, Cui Tai¹, Zixin Deng¹, Jiangning Song^{2,3} and Hong-Yu Ou^{1,*}

¹State Key Laboratory of Microbial Metabolism, Joint International Laboratory on Metabolic & Developmental Sciences, School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

²Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC 3800, Australia

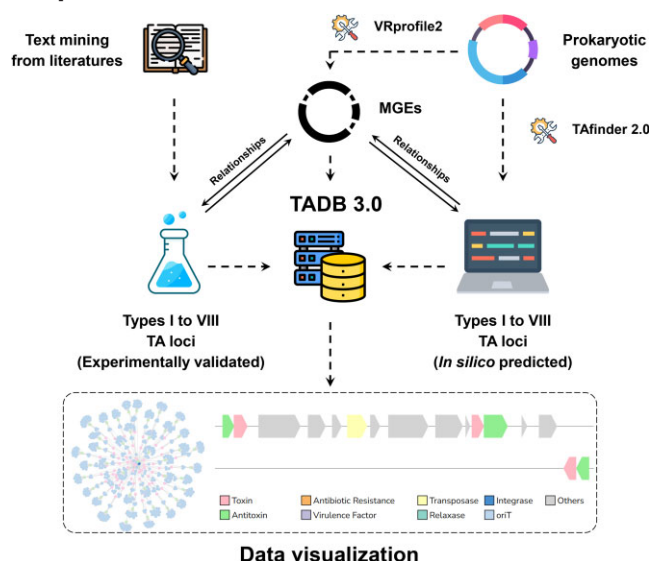
³Monash Data Futures Institute, Monash University, Melbourne, VIC 3800, Australia

*To whom correspondence should be addressed. Tel: +86 2162932943; Fax: +86 2162932418; Email: hyou@sjtu.edu.cn

Abstract

TADB 3.0 (<https://bioinfo-mml.sjtu.edu.cn/TADB3/>) is an updated database that provides comprehensive information on bacterial types I to VIII toxin–antitoxin (TA) loci. Compared with the previous version, three major improvements are introduced: First, with the aid of text mining and manual curation, it records the details of 536 TA loci with experimental support, including 102, 403, 8, 14, 1, 1, 3 and 4 TA loci of types I to VIII, respectively; Second, by leveraging the upgraded TA prediction tool TAFinder 2.0 with a stringent strategy, TADB 3.0 collects 211 697 putative types I to VIII TA loci predicted in 34 789 completely sequenced prokaryotic genomes, providing researchers with a large-scale dataset for further follow-up analysis and characterization; Third, based on their genomic locations, relationships of 69 019 TA loci and 60 898 mobile genetic elements (MGEs) are visualized by interactive networks accessible through the user-friendly web page. With the recent updates, TADB 3.0 may provide improved *in silico* support for comprehending the biological roles of TA pairs in prokaryotes and their functional associations with MGEs.

Graphical abstract



Introduction

Toxin–antitoxin (TA) loci are small genetic modules commonly found in prokaryote genomes, whose prevalence and diversity have been extensively documented in numerous chromosomes and plasmids (1). Typically, a TA locus consists of a pair of genes: one encodes a stable toxin that inhibits essential cellular functions, and another encodes a labile antitoxin that neutralizes the activity of its cognate toxin.

TA systems have been found to participate in various biological processes essential for bacterial survival and adaptation, such as stress response (2), bacteriophage defense (3), bacterial virulence (4), antimicrobial resistance (5) and plasmid stability (6). The known TA loci can be classified into eight types based on the nature and mode of action of the antitoxins, namely types I to VIII (7,8). The majority of described toxins are proteins (types I to VII), while the toxins are RNAs

Received: September 15, 2023. Revised: October 11, 2023. Editorial Decision: October 13, 2023. Accepted: October 14, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

for type VIII TA pairs. Antitoxins are proteins for types II and IV to VII TA pairs but are RNAs in types I, III and VIII TA pairs.

Bacterial mobile genetic elements (MGEs), such as prophages, genomic islands (GIs), integrative and conjugative elements (ICEs), and plasmids, often harbor genes conferring selective advantages, including antibiotic resistance determinants and virulence factors, thereby shaping bacterial evolution and dissemination of these traits (9,10). Notably, growing evidence indicates the significant associations between TA loci and MGEs, highlighting their functional interplay (11–13). For example, the type II TA pairs have been well-documented to be an addiction module to ensure the maintenance of plasmids during bacterial replication (6,14). Recently, a plasmid-carrying *prpT/prpA* locus of *Pseudoalteromonas rubra* was found to control the plasmid copy number (15). TA loci have also been identified in relation to the genetic stability of other MGEs such as genomic islands (13), prophages (16), integrative and conjugative elements (17), and integrons (18). Interestingly, a *creTA* type VIII TA locus was proposed to maintain CRISPR immunity by making cells addicted to CRISPR-Cas (19). Therefore, understanding the relationships between TA loci and MGEs might shed light on the mechanisms underlying horizontal gene transfer, bacterial survival under environmental stress and bacterial adaptation to changing environments.

The field of TA systems has witnessed remarkable progress over the years, with various databases providing valuable resources for characterizing these systems. TASmania (20) is a database archiving *in silico* discovered types I to IV TA loci using curated HMM models. The T1Tadb database (21) specifically collected information on type I TA loci, including annotations of toxin peptides and antitoxin RNA molecules. The RASTA online tool (22) searches the conserved functional domains of individual toxins or antitoxin proteins, but its website maintenance was discontinued in 2011. In the year of 2011, we proposed the open-access database TADB1.1 (23) archiving both the experimentally validated and *in silico* predicted type II TA pairs. In 2018, we released the updated database TADB 2.0 (24), which systematically archives both the updated experimentally validated and *in silico* predicted type II TA loci, along with an online tool TAffinder for the identification of type II TA loci in bacterial genomes. Since then, a rapidly increasing number of new type II and other types of TA pairs have been characterized experimentally. Furthermore, although the interplay between TA loci and MGEs plays a crucial role in bacterial survival and evolution, existing TA-related databases often lack comprehensive information on their associations. Thus, the demand for a major database update became urgent.

Here, we report the release of TADB version 3.0, which provides an up-to-date comprehensive deposition of experimentally supported types I to VIII TA pairs. Meanwhile, we have upgraded TAffinder to version 2.0, considerably expanding its capabilities to predict a wide range of types I to VIII TA loci in completely sequenced bacterial and archaeal genomes. Furthermore, TADB 3.0 provides a user-friendly webpage to interrogate and display the relationships of TA loci and MGEs based on their genomic locations. We expect that TADB 3.0 will provide better support for exploring the diversity of TA pairs and offering crucial insights into the intricate relationships between TA loci and MGEs.

Materials and methods

Data updated by text mining and manual curation

Following a meticulous manual curation process of literatures from PubMed using the keyword ‘toxin–antitoxin system’, TADB 3.0 has incorporated 335 TA system-related publications published since 2017, bringing the total papers in the database to 921. After text mining and manual curations of these papers, the number of experimentally validated type II TA loci has significantly expanded from 105 to 403 in TADB 3.0. Moreover, this updated database includes 102, 8, 14, 1, 1, 3 and 4 experimentally validated TA loci of types I, III, IV, V, VI, VII and VIII, respectively (Supplementary Table S1). Notably, the families and genomic locations of 36 type I TA loci were collected with reference to T1Tadb (21), and hyperlinks to T1Tadb entries were built for these type I TA loci. In addition, using a rigorous prediction strategy, TADB 3.0 also records 16 009, 168 794, 55, 16 685, 2 786, 2, 388 and 6 978 *in silico* predicted TA loci of types I to VIII, respectively, as elaborated in the following section. Consequently, the TADB 3.0 database now archives a total of 16 111, 169 197, 63, 16 699, 2 787, 3, 391 and 6 982 TA loci, all accessible within the database (Supplementary Table S1).

Similar to the previous version, TADB 3.0 continues to utilize the PostgreSQL relational database, PHP data pipeline and HTML web interfaces for organizing the TA loci data. In addition, the Bootstrap library (<https://getbootstrap.com/>) and data visualization library powered by JavaScript, such as ECharts (25) and SVGene (<https://github.com/kblin/svgene>), are added to various web interfaces for more user-friendly browsing. Illustrations of the TADB 3.0 data presentation interfaces are provided in Figure 1 and Supplementary Figure S1. The domains of toxin and antitoxin proteins were predicted by InterProScan (26). For each toxin or antitoxin, similar proteins were identified using NCBI BLASTp (27). Multiple sequence alignments of these similar proteins were performed by Clustal Omega (28) and subsequently visualized by the R package msa (29). For protein toxins and antitoxins, the experimentally determined and predicted three-dimensional structures were obtained from the RCSB Protein Data Bank (PDB) (<https://www.rcsb.org/>) (30) and AlphaFold Protein Structure Database (31), respectively. These structures were then visualized using PDBe Mol* (32). In the case of RNA toxins and antitoxins, the RNA secondary structures were predicted by RNAfold (33) and visualized by VARNA (34).

Prediction of types I to VIII TA loci in completely sequenced prokaryotic genomes using TAffinder 2.0

The TAffinder tool, initially introduced in the TADB 2.0 database (24), is designed to predict type II TA loci through the search of homologs for toxin or antitoxin proteins by utilizing NCBI BLASTp (27) and HMMER3 (35). Types II to VII TA loci typically consist of two tandem genes located on the same DNA strand, coding for cognate protein partners. In contrast, the type I TA system consists of a toxin gene and an antitoxin RNA that are located on the opposite DNA strands. The type VIII TA system consists of a toxin RNA and an antitoxin RNA, which are located either on the same strand or the opposite strands. In this study, we updated the TAffinder prediction pipeline to version 2.0, enabling the prediction of all types of TA loci in addition to type II (Supplementary Figure S2).

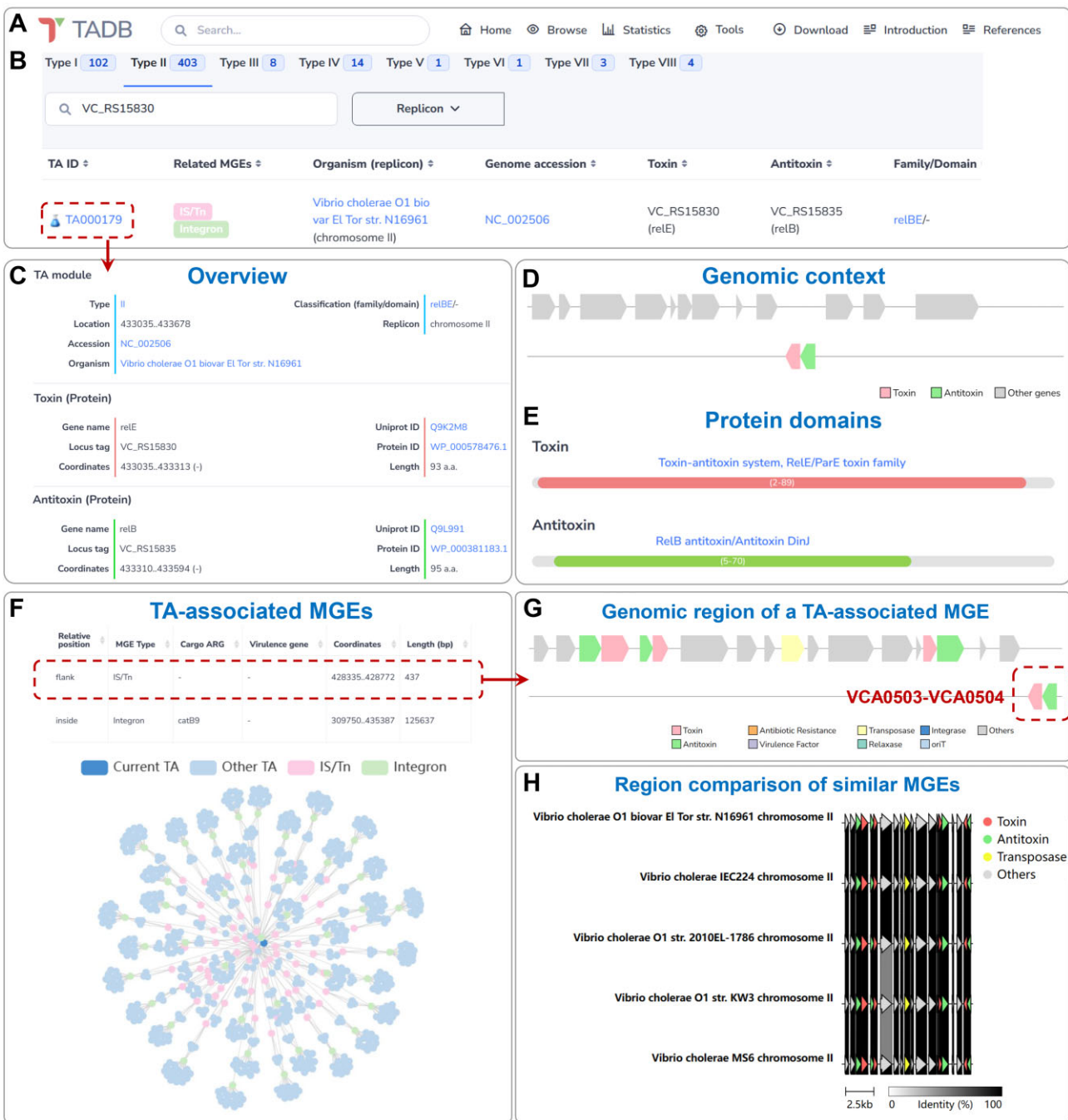


Figure 1. Updated web interface of TADB 3.0. **(A)** The functional modules of TADB 3.0 on the home page. **(B)** The 'Browse' web page displaying information on the *relBE* family TA locus (*VCA0503/VC_RS15830-VCA0504/VC_RS15835*) in *Vibrio cholerae* O1 bio var El Tor str. N16961. **(C)** An overview of the features of TA locus *VCA0503-VCA0504* on the detailed information page. **(D)** Visualization of the genomic context of TA locus by SVGene. **(E)** Protein domains of the toxin *VCA0503* and the antitoxin *VCA0504* predicted by InterProScan. **(F)** Interactive networks highlighting the MGEs associated with the *VCA0503-VCA0504* TA locus identified by VRprofile2. Similar MGEs were defined as MGEs with pairwise Mash distances lower than 0.01. **(G)** Visualization of the gene structures around the TA-associated MGEs by SVGene using an IS element flanked by TA locus *VCA0503-VCA0504* as an example. **(H)** Partial view of the comparison of similar MGEs by clinker and clustermap.js.

In brief, the prediction of RNA homologs by BLASTn was added to the pipeline. The protein homolog search module and operon structure detection module remain the same as the original version for types II to VII TA loci prediction. Notably, for the identification of type I TA loci, the toxin gene and antitoxin RNA must be located on opposite DNA strands, rather than forming an operon structure on the same strand. In addition, for the identification of type VIII TA loci, we took

into consideration the characteristics of the two experimental validated type VIII TA loci: *creTA*, where the toxin and antitoxin RNAs are located either on the same strand or the opposite strands (19,36,37), and *SdsR-RyeA*, where the two RNAs are located on opposite strands (38). Accordingly, we made predictions for these two distinct type VIII TA loci based on their respective characteristics. The BLAST reference dataset comprises all the TADB 3.0-archived experimentally validated

proteins and RNAs of TA loci archived in TADB 3.0. In addition, in comparison with the previous version of TAFinder, we have expanded the number of HMM profiles for the toxin domains from 108 to 207, and for the antitoxin domains, the number has increased from 201 to 228.

To perform TA loci prediction on completely sequenced bacterial and archaeal genomes using TAFinder 2.0, we first downloaded 34 257 bacterial and 532 archaeal complete genomes from the NCBI RefSeq database (39) on 10 July, 2023, including 36 751 chromosomes and 40 457 plasmids. Subsequently, we used these genome sequences and annotated protein-coding sequences as inputs to TAFinder 2.0 for TA loci prediction (Supplementary Figure S3). To ensure the high specificity of TA loci prediction results, we only kept the proteins and nucleotides with BLAST-based *H*-value > 0.36 (40) as candidate toxins and antitoxins. In addition, the criteria for the length of toxin or antitoxin proteins were set to range between 30 and 500 amino acid residues. Except for the type I TA loci, the intergenic distance between the putative toxin and antitoxin genes on the same DNA strand was set to -20 to 150 bp. For type I TA loci prediction, the maximum distance between the toxin gene and the antisense antitoxin sRNA is set to 200 bp, without any limitation of overlap length. For the prediction of type VIII TA loci, we applied distinct criteria: the maximum distance between toxin RNA and antitoxin RNA was set to 200 bp for *creTA* while this distance was set to <0 bp for *SdsR-RyeA*.

Identification of MGEs associated with TA loci

In this study, TA-associated MGEs are defined as the MGEs harboring TA loci or the ISs/transposons flanked by TA loci with an interval < 5 kbp (Supplementary Figure S3). To identify MGEs associated with both experimentally validated and *in silico* predicted TA loci, the MGE identification pipeline VRprofile2 (41) was first utilized to identify MGEs in bacterial genomes, including prophages, GIs, ICEs, integrons, ISs/transposons and IS clusters. The genomic locations of TA loci and predicted MGEs were then analyzed to identify TA-associated MGEs. Pairwise Mash distances (42) of these MGEs were also calculated, and MGEs with Mash distances < 0.01 were considered similar MGEs.

TA-associated MGEs were displayed on the 'Browse' page, 'Statistics' page and the detailed information page of TA loci. In the 'Statistics' page and detailed information pages of TA loci, TA-associated MGEs, the corresponding similar MGEs, and all TA loci harbored by these MGEs are dynamically presented as interactive networks utilizing ECharts (25). Users can access comprehensive details about each MGE by simply clicking on MGE nodes in the network graphs or via MGE buttons available in tables on the 'Browse' page. Besides, within the detailed information page for MGEs, cargo genes including TA pairs, antibiotic resistance genes and virulence genes are visually displayed through gene structure plots by SVGene (<https://github.com/kblin/svgene>). In addition, comparative analysis and interactive visualization of a group of similar MGEs were performed by clinker and clustermap.js (43).

Results and discussion

Compared to TADB 2.0, the updated TADB 3.0 database offers three major improvements: (i) manual curation of >500

types I to VIII TA loci with experimental evidence and support; (ii) collection of the TAFinder 2.0-predicted TA loci in > 34000 completely sequenced prokaryotic genomes; (iii) graphical representation of the relationships of the TA loci and MGEs.

TADB 3.0 browse module for types I to VIII TA loci

TADB 3.0 provides a flexible and user-friendly web interface with several new features incorporated (Figure 1). Specifically, the 'Statistics' module was developed for visualization of the relationships between TA loci and MGEs, the distribution of TA loci across taxonomies and the homology networks of toxin and antitoxin proteins (Supplementary Figure S1). On the 'Browse' page and in the detailed information pages dedicated to various types of TA loci, users can access comprehensive details about the strains, TA families, genomic contexts, protein domains and associated MGEs (Figure 1B–F). For instance, the *relBE* family TA locus (*VCA0503/VC_RS15830-VCA0504/VC_RS15835*) is one of the TA loci previously characterized in the chromosome II of *Vibrio cholerae* O1 biovar El Tor str. N16961 (18). This TA locus is located within a superintegron and plays a crucial role in stabilizing this MGE. Meanwhile, as identified by VRprofile2 (41), the TA locus is also flanked by a putative IS5 family transposase within a 5 kbp distance, supporting the idea that the TA locus might contribute to the stability of its adjacent region (Figure 1F). Visualizing the 5 kbp flanking region of this IS5 family transposase reveals its connections with three additional TA loci, apart from *VCA0503-VCA0504* (Figure 1G). Moreover, by analyzing other predicted TA loci and MGEs in *V. cholerae*, we found that superintegrons and transposases are widespread in *V. cholerae* genomes, which also encode or are flanked by several putative TA loci, as previously described by Iqbal *et al.* (18) (Figure 1F). Comparison of the flanking genomic regions of these similar transposases indicated that these regions are conserved among a certain amount of *V. cholerae* strains being analyzed (Figure 1H).

Rigorous prediction of types I to VIII TA loci in prokaryotic genome sequences

In recent years, there has been a substantial increase in both the number of experimentally validated TA pairs and the availability of prokaryotic genome sequencing data. This growing wealth of information necessitates the need to obtain relatively high confidence predicted TA loci dataset to support further experimental research. With the experimentally validated types I to VIII TA loci available in the TADB 3.0 database (Supplementary Table S1), we have upgraded our type II TA loci prediction tool, TAFinder, to version 2.0 (Supplementary Figure S2). This new version is capable of predicting types I to VIII TA loci by utilizing the newly collected experimentally validated data from the TADB 3.0 database. For the prediction of TA loci in completely sequenced prokaryotic genomes, a total of 34 257 bacterial and 532 archaeal completely sequenced genomes were downloaded from the NCBI RefSeq database (39). Subsequently, TAFinder 2.0 was employed to predict TA loci in these genome sequences. To enhance the specificity and reliability of our predictions, we adopted a rigorous methodology that exclusively used experimentally validated TA pairs as the reference dataset and implemented a stringent strategy (Supplementary Figure S3). Briefly, only toxin and antitoxin BLAST hits with *H*-value >0.36 (40)

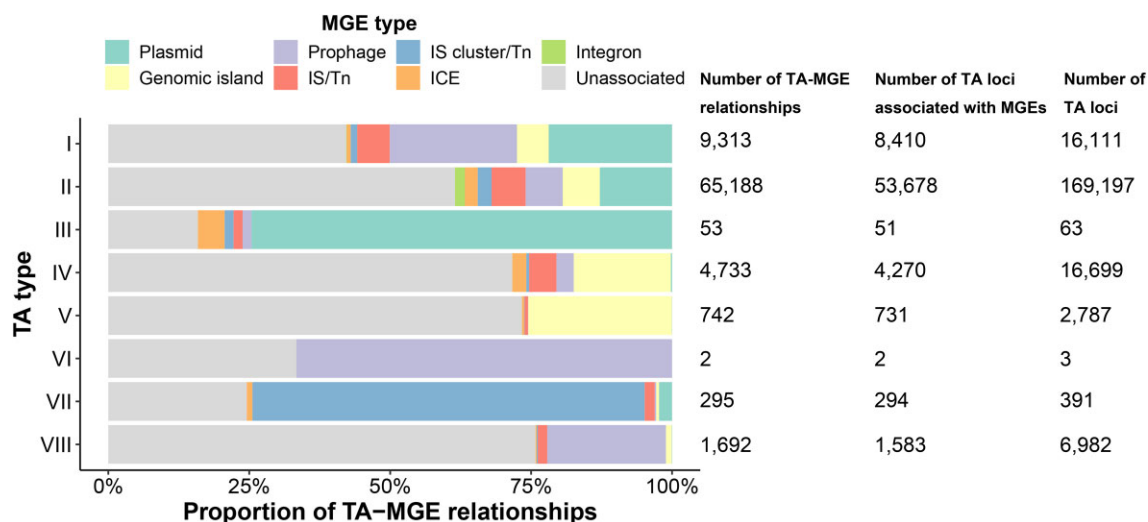


Figure 2. Relationships of the TADB-archived TA loci and MGEs of the prokaryotic organisms. TA-associated MGEs are defined as the MGEs harboring TA loci or the ISs/transposons flanked by TA loci with an interval <5 kb in this study. The stacked bar plots display the proportion of TA-MGE relationships of each TA type. If one TA locus was associated with multiple MGEs or one MGE was associated with multiple TA loci, the relationships were calculated separately. The number of TA-MGE relationships, the number of TA loci associated with MGEs and the total number of TA loci were listed on the right side of the bar plot.

were retained. As a result, our predictions yielded a total of 211 697 TA loci homologs in 34 789 completely sequenced prokaryotic genomes, including 16 009, 168 794, 55, 16 685, 2 786, 2, 388 and 6 978 TA homologs of types I to VIII TA loci, respectively (Supplementary Table S1).

The species with the highest number of *in silico* predicted TA loci for each type and the distribution of TA families within each TA type for each species were displayed in Supplementary Figure S4. As the most extensively studied type of TA systems, type II TA loci homologs are the most abundant among all types of TA loci homologs, partially due to the large number of experimentally validated type II loci used for prediction. Among the species harboring type II TA loci homologs, 99.1% (2769/2795) of the analyzed *Escherichia coli* genomes harbored 39 324 putative type II TA loci in both chromosomes and plasmids, with around 14 type II TA loci per strain (Supplementary Figure S4B). The pronounced representation of type II TA loci among *E. coli* can be attributed, in part, to the extensive availability of completely sequenced *E. coli* genomes. Apart from that, *Mycobacterium tuberculosis* strains possess an average of >50 TA loci, significantly outnumbering other species. Meanwhile, the average number of putative type II TA loci per strain is high across several clinical pathogens, such as *V. cholerae* (17.3) and *Klebsiella pneumoniae* (11.3). Among the predicted type II TA loci, the TA families *vapBC*, *higBA*, *relBE* and *mazEF* are widespread in the top 20 species with the highest number of predicted type II TA loci, while some other families such as *hipBA* appear to be abundant in certain species.

The predicted TA loci have been cataloged in the TADB 3.0 database available for browsing and download. However, it is noteworthy that the predicted outcomes might not accurately represent the distribution of TA loci and TA families across prokaryotes, but rather serve as high-confidence predictions for reference purposes, highlighting potential TA loci worthy of further experimental validation and investigation. Furthermore, for users specifically interested in particular bacterial strains, it is highly recommended to utilize TAFinder 2.0 with less strict thresholds for TA loci prediction.

Investigation of the TA-MGE relationships based on genomic location

TA pairs are proposed to play a role in maintaining MGEs that contain these genetic modules (13,16–18). After obtaining experimentally validated TA loci together with the *in silico* predicted TA loci, we then utilized VRprofile2 (41) to identify various types of MGEs in TA-harboring prokaryotic strains (Supplementary Figure S3). In this study, the TA-associated MGEs are defined as the MGEs harboring TA loci or the ISs/transposons flanked by TA loci with an interval <5 kb (Supplementary Figure S3). As a result, out of the 211 697 TA loci archived in TADB 3.0, 69 019 (32.6%) TA loci were predicted to be associated with 60 898 MGEs, resulting in a total of 82 018 TA-MGE relationships (Figure 2, Supplementary Table S2 and Supplementary Figure S5). Among various types of MGEs, prophages and plasmids exhibited the highest frequency of relationships with TA loci. Out of 40 457 plasmids being analyzed, 14 163 (35.0%) were found to encode at least one TA locus. Besides, genomic islands and ISs/transposons also showed significant associations with TA loci. For type II TA loci, it was observed that 31.7% (53 678/169 197) of them were associated with MGEs, especially plasmids, genomic islands and prophages (Figure 2). Interestingly, a majority of identified type III TA loci (51/63, 81.0%) were found to be associated with MGEs, particularly plasmids. This observation might be attributed to the relatively low number of experimentally validated type III TA loci currently available, coupled with our stringent prediction methodology, which might have resulted in fewer predictions for type III TA loci. Moreover, we also analyzed the distribution of TA families among different types of MGEs (Supplementary Figure S6). Among the TA families, the type II TA family *relBE* has the most connections with MGEs, followed by the type II TA family *higBA* and the type I TA family *hok-sok*. Meanwhile, the type IV TA family *yeeUV* (also known as *cbtA-cbeA*) is also frequently associated with MGEs. Notably, the type II TA family *relBE* is associated with all types of MGEs, whereas a large number of TA loci of type IV TA family *yeeUV* are located within genomic islands but were seldom found on plasmids. How-

ever, it is worth noting that, since our MGE prediction focused solely on strains that harbor TA loci, further investigations are needed to determine the proportion of MGEs in each MGE type associated with TA loci.

Conclusion

Here, we reported a major upgrade of TADB with a primary focus on TA loci and their relationships with MGEs. TADB 3.0 collected and integrated the systematic information of types I to VIII TA loci. Overall, our rigorous analysis, using a robust methodology that incorporates experimentally validated TA loci and stringent threshold settings, has yielded highly specific prediction results of TA loci in completely sequenced prokaryotic genomes. Furthermore, TADB 3.0 provides insightful visualizations of the relationships of the putative TA loci and MGEs. Altogether, these results available in TADB 3.0 provide researchers a solid foundation for further investigations while recognizing the need for future experimental studies to comprehensively explore the distribution and functional significance of TA pairs in prokaryotes.

Data availability

TADB 3.0 is freely available at <https://bioinfo-mml.sjtu.edu.cn/TADB3/>.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

National Natural Science Foundation of China [32070572]; Science and Technology Commission of Shanghai Municipality [19JC1413000]. Funding for open access charge: National Natural Science Foundation of China [32070572] and Science and Technology Commission of Shanghai Municipality [19JC1413000].

Conflict of interest statement

None declared.

References

- Yamaguchi,Y., Park,J.H. and Inouye,M. (2011) Toxin-antitoxin systems in bacteria and archaea. *Annu. Rev. Genet.*, **45**, 61–79.
- Gerdes,K. and Maisonneuve,E. (2012) Bacterial persistence and toxin-antitoxin loci. *Annu. Rev. Microbiol.*, **66**, 103–123.
- Fineran,P.C., Blower,T.R., Foulds,I.J., Humphreys,D.P., Lilley,K.S. and Salmond,G.P. (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 894–899.
- Lobato-Marquez,D., Diaz-Orejas,R. and Portillo,G.-D., F. (2016) Toxin-antitoxins and bacterial virulence. *FEMS Microbiol. Rev.*, **40**, 592–609.
- Yang,Q.E. and Walsh,T.R. (2017) Toxin-antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. *FEMS Microbiol. Rev.*, **41**, 343–353.
- Gerdes,K., Christensen,S.K. and Lobner-Olesen,A. (2005) Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Micro.*, **3**, 371–382.
- Jurenas,D., Fraikin,N., Goormaghtigh,F. and Van Melderren,L. (2022) Biology and evolution of bacterial toxin-antitoxin systems. *Nat. Rev. Micro.*, **20**, 335–350.
- Qiu,J., Zhai,Y., Wei,M., Zheng,C. and Jiao,X. (2022) Toxin-antitoxin systems: classification, biological roles, and applications. *Microbiol. Res.*, **264**, 127159.
- van Hoek,A.H., Mevius,D., Guerra,B., Mullany,P., Roberts,A.P. and Aarts,H.J. (2011) Acquired antibiotic resistance genes: an overview. *Front. Microbiol.*, **2**, 203.
- Smillie,C., Garcillan-Barcia,M.P., Francia,M.V., Rocha,E.P. and de la Cruz,F. (2010) Mobility of plasmids. *Microbiol. Mol. Biol. Rev.*, **74**, 434–452.
- Yao,X., Chen,T., Shen,X., Zhao,Y., Wang,M., Rao,X., Yin,S., Wang,J., Gong,Y., Lu,S., et al. (2015) The chromosomal SezAT toxin-antitoxin system promotes the maintenance of the SsPI-1 pathogenicity island in epidemic *Streptococcus suis*. *Mol. Microbiol.*, **98**, 243–257.
- Dziewit,L., Jazurek,M., Drewniak,L., Baj,J. and Bartosik,D. (2007) The SXT conjugative element and linear prophage N15 encode toxin-antitoxin-stabilizing systems homologous to the tad-ata module of the *Paracoccus aminophilus* plasmid pAM12. *J. Bacteriol.*, **189**, 1983–1997.
- Huguet,K.T., Gonnet,M., Doublet,B. and Cloeckaert,A. (2016) A toxin antitoxin system promotes the maintenance of the IncA/C-mobilizable *Salmonella* Genomic Island 1. *Sci. Rep.*, **6**, 32285.
- Harms,A., Brodersen,D.E., Mitarai,N. and Gerdes,K. (2018) Toxins, Targets, and Triggers: an Overview of Toxin-Antitoxin Biology. *Mol. Cell*, **70**, 768–784.
- Ni,S., Li,B., Tang,K., Yao,J., Wood,T.K., Wang,P. and Wang,X. (2021) Conjugative plasmid-encoded toxin-antitoxin system PrpT/PrpA directly controls plasmid copy number. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2011577118.
- Hallez,R., Geeraerts,D., Sterckx,Y., Mine,N., Loris,R. and Van Melderren,L. (2010) New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in *Escherichia coli* O157:H7. *Mol. Microbiol.*, **76**, 719–732.
- Wozniak,R.A. and Waldor,M.K. (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.*, **5**, e1000439.
- Iqbal,N., Guerout,A.M., Krin,E., Le Roux,F. and Mazel,D. (2015) Comprehensive functional analysis of the 18 *Vibrio cholerae* N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. *J. Bacteriol.*, **197**, 2150–2159.
- Li,M., Gong,L., Cheng,F., Yu,H., Zhao,D., Wang,R., Wang,T., Zhang,S., Zhou,J., Shmakov,S.A., et al. (2021) Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. *Science*, **372**, eabe5601.
- Akarsu,H., Bordes,P., Mansour,M., Bigot,D.J., Genevaux,P. and Falquet,L. (2019) TASmania: a bacterial Toxin-Antitoxin Systems database. *PLoS Comput. Biol.*, **15**, e1006946.
- Tourasse,N.J. and Darfeuille,E. (2021) T1Tadb: the database of type I toxin-antitoxin systems. *RNA*, **27**, 1471–1481.
- Sevin,E.W. and Barloy-Hubler,F. (2007) RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biol.*, **8**, R155.
- Shao,Y., Harrison,E.M., Bi,D., Tai,C., He,X., Ou,H.Y., Rajakumar,K. and Deng,Z. (2011) TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Res.*, **39**, D606–D611.
- Xie,Y., Wei,Y., Shen,Y., Li,X., Zhou,H., Tai,C., Deng,Z. and Ou,H.Y. (2018) TADB 2.0: an updated database of bacterial type II toxin-antitoxin loci. *Nucleic Acids Res.*, **46**, D749–D753.
- Li,D.Q., Mei,H.H., Shen,Y., Su,S., Zhang,W.L., Wang,J.T., Zu,M. and Chen,W. (2021) ECharts: a declarative framework for rapid construction of web-based visualization (vol 2, pg 136, 2018). *Vis. Inform.*, **5**, 43.
- Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A., Nuka,G., et al. (2014)

- InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240.
27. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf.*, 10, 421.
 28. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, 7, 539.
 29. Bodenhofer,U., Bonatesta,E., Horejs-Kainrath,C. and Hochreiter,S. (2015) msa: an R package for multiple sequence alignment. *Bioinformatics*, 31, 3997–3999.
 30. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, 28, 235–242.
 31. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A., *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, 50, D439–D444.
 32. Sehnal,D., Bittrich,S., Deshpande,M., Svobodova,R., Berka,K., Bazgier,V., Velankar,S., Burley,S.K., Koca,J. and Rose,A.S. (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.*, 49, W431–W437.
 33. Gruber,A.R., Lorenz,R., Bernhart,S.H., Neubock,R. and Hofacker,I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, 36, W70–W74.
 34. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25, 1974–1975.
 35. Potter,S.C., Luciani,A., Eddy,S.R., Park,Y., Lopez,R. and Finn,R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.*, 46, W200–W204.
 36. Cheng,F., Wang,R., Yu,H., Liu,C., Yang,J., Xiang,H. and Li,M. (2021) Divergent degeneration of creA antitoxin genes from minimal CRISPRs and the convergent strategy of tRNA-sequestering CreT toxins. *Nucleic Acids Res.*, 49, 10677–10688.
 37. Wang,R., Shu,X., Zhao,H., Xue,Q., Liu,C., Wu,A., Cheng,F., Wang,L., Zhang,Y., Feng,J., *et al.* (2023) Associate toxin–antitoxin with CRISPR-Cas to kill multidrug-resistant pathogens. *Nat. Commun.*, 14, 2078.
 38. Choi,J.S., Kim,W., Suk,S., Park,H., Bak,G., Yoon,J. and Lee,Y. (2018) The small RNA, SdsR, acts as a novel type of toxin in *Escherichia coli*. *RNA Biol.*, 15, 1319–1335.
 39. O’Leary,N.A., Wright,M.W., Brister,J.R., Ciufu,S., Haddad,D., McVeigh,R., Rajput,B., Robbertse,B., Smith-White,B., Ako-Adjei,D., *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, 44, D733–D745.
 40. Li,J., Tai,C., Deng,Z., Zhong,W., He,Y. and Ou,H.Y. (2018) VRprofile: gene-cluster-detection-based profiling of virulence and antibiotic resistance traits encoded within genome sequences of pathogenic bacteria. *Brief Bioinform.*, 19, 566–574.
 41. Wang,M., Goh,Y.X., Tai,C., Wang,H., Deng,Z. and Ou,H.Y. (2022) VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens. *Nucleic Acids Res.*, 50, W768–W773.
 42. Ondov,B.D., Treangen,T.J., Melsted,P., Mallonee,A.B., Bergman,N.H., Koren,S. and Phillippy,A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.*, 17, 132.
 43. Gilchrist,C.L.M. and Chooi,Y.H. (2021) clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics*, 37, 2473–2475.