

Letter to the Editor

T4SP: A Novel Tool and Database for Type IV Secretion Systems in Bacterial Genomes*

ZHANG Wen¹, YU Wei Wen¹, LIU Di², LI Ming³, DU Peng Cheng¹, WU Yi Lei⁴,
George F. GAO^{5,6,7,#}, and CHEN Chen^{1,#}

Secretion systems, which can mediate the macromolecules to pass across cellular membranes, are essential for virulent and genetic material exchange among bacterial species^[1]. Type IV secretion system (T4SS) is one of the secretion systems and it usually consists of 12 genes: VirB1, VirB2 ...VirB11, and VirD4^[2]. The structure and molecular mechanisms of these genes have been well analyzed in Gram-negative strains^[3] and Gram-positive strains were once believed to be lack of T4SS. However, some recent studies revealed that one or more *virB/D* genes also exist in some kinds of Gram-positive bacteria and play similar role, and form a T4SS-like system^[3]. The VirB1-like, VirB4, VirB6, and VirD4 genes were identified in the chromosome of Gram-positive bacterium *Streptococcus suis* in our previous studies and their role as important mobile elements for horizontal transfer to recipients in an 89 K pathogenicity island (PAI) was demonstrated^[4-5]. However, their structure and molecular mechanisms in other strains, especially in Gram-positive strains, are remained unclear.

T4SS plays variety of roles in the process of transferring a wide range of components, such as single proteins, or protein-protein and protein-DNA complexes^[1-2,6] and its most important role is to mediate the conjugative plasmid DNA or transposons transfer, which leads to the spread of antibiotic-resistant and pathogenic genes among a wide range of bacterial species^[2,6]. In addition, T4SS can directly mediated the secretion of effect protein into host cells during infection^[7]. Because all T4SS are evolutionally related^[2], they can be identified based on the structures of its genes. However, the versatile genes of T4SS limit the identification ability of the system. Also, kinds of pathogens activity modes in different bacteria are remained largely unsolved to

date. Therefore, reliable predictive method for uncovering the homologs of T4SS can be very helpful to elucidate the candidate drug-resistant genes and pathogenic mechanisms of bacterial species.

In the present study, a database with an embedded bioinformatics tool, named the Type IV Secretion Project (T4SP), was developed. Experimentally verified *virB/D* genes, putative functional *virB/D* genes which were predicted based on genomic location and homology, and homologs of *virB/D* genes which were predicted based on homology in the complete bacterial gene pool, as well as the integrative bioinformatics software package for predicting T4SS are publicly available online (<http://T4SS.bioinfo-icdc.org>, mirror website <http://T4SS.genotyping.org>). T4SS genes (*virB/D* genes) can be identified by T4SP in the complete bacterial gene pool by either using the sequence alignment method, or the Hidden Markov Model algorithm, or gene annotation tools, which also presents the localization and distribution of all candidate T4SS genes. The T4SP database is useful for predicting and testing the possible gene exchange mechanisms and effect or molecule delivery ways in various microbiological species. It will also help to elucidate the function of T4SS homologous genes.

The T4SP database is an integrative system for finding all *virB/D* genes related with T4SS in bacterial gene pool. The latest version of T4SP contains a total of 717 experimentally verified *virB/D* genes, 3852 *virB/D* genes predicted based on genomic location and homology, and 21 785 homologs of *virB/D* genes. The T4SP database also covers distribution information (e.g., gene number, gene name, gene type, position, sequence, related articles, quick links to other websites, and so on) of these *virB/D* genes. Users may access the T4SP database (<http://T4SS>).

doi: 10.3967/0895-3988.2013.07.015

1. National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention/State Key Laboratory for Infectious Disease Prevention and Control, Beijing 102206, China; 2. Network Information Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; 3. Department of Microbiology, Third Military Medical University, Chongqing 400038, China; 4. Beijing Computing Center, Beijing, China; 5. CAS Key laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Science, Beijing 100101, China; 6. Beijing Institute of Life Science, Chinese Academy of Sciences; 7. Chinese Center for Disease Control and Prevention, Beijing, China

bioinfo-icdc.org) using an internet browser. A mirror website (<http://T4SS.genotyping.org>) was also created to balance the main website load, which can be tested by online stability testing tools (<http://www.websitepulse.com>).

Most of biologists interested in T4SS usually focus on two issues. The first issue is whether a certain species has already been proven to have the T4SS. We have checked the gene annotations, and all related references for all reported *virB/D* genes, which are listed in our database. We hope to annotate and mark all of the genes in the future. The second issue is whether new species or strain contains T4SS. Users usually know little about a new target bacterial strain and a reliable tool is therefore needed for identifying whether the strain has T4SS

pathogenicity which caused via the molecular mechanisms. The T4SP quick identification system then helps us elucidate these. Two lists “Experimentally verified *virB/D* genes” with their cited references and “Putative Functional *virB/D* genes predicted based on genomic location and homology” were exhibited in the T4SP database, which are identified by the bioinformatics tools on the search page. With the help of the T4SP database, researchers can easily address the two issues mentioned above. Furthermore, it is very convenient for users to make use of the database because of the additional support functions, such as a search engine, taxonomy tree, sequence download, result comparison, reference lists, and quick links to other websites.

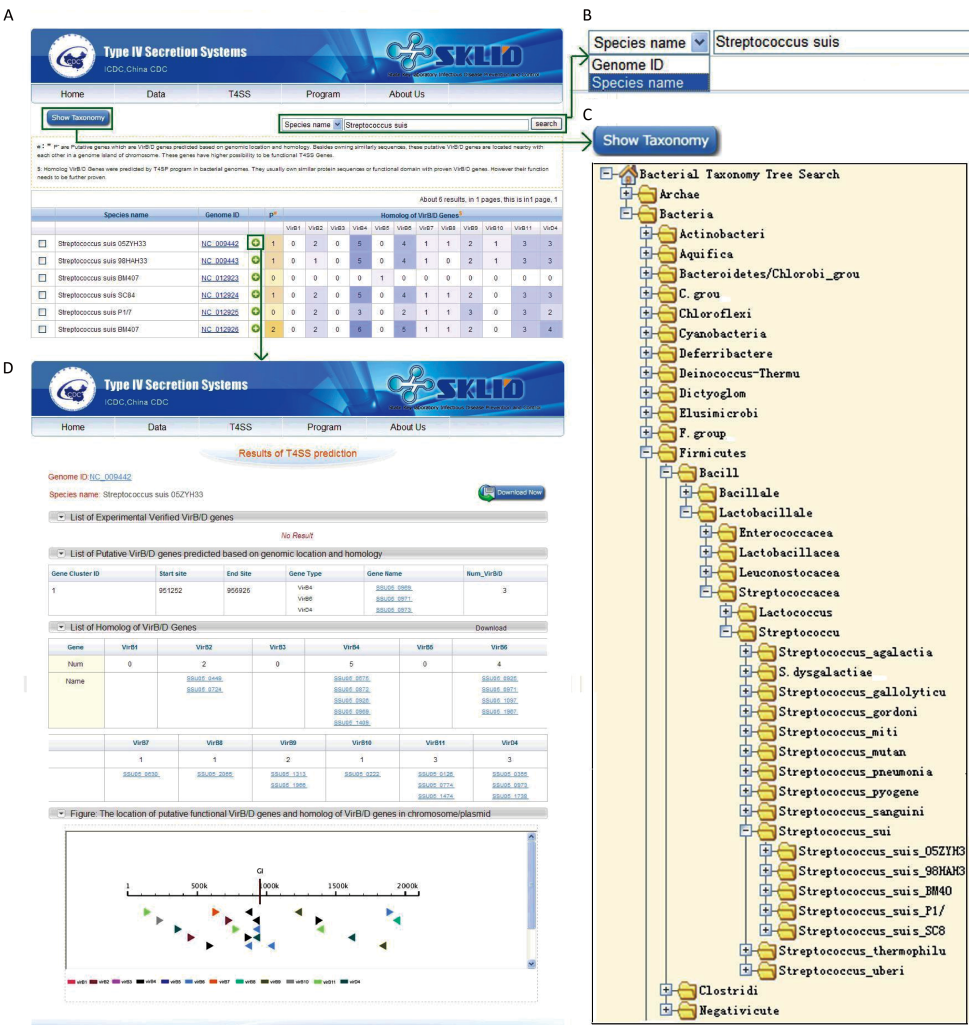


Figure 1. User interface of the T4SP database (A)sample of the search page showing tabular results. The database can be accessed by two ways, as shown in Figures B and C. (B) A search box using key words. (C) Part of the bacterial taxonomy tree. (D) A sample of the results page.

Data example from *Streptococcus suis* (*S. suis*) are illustrated here to display the functions of the T4SP database. *S. suis* is a species of *Streptococcus* found in pigs and it can cause human infections^[8]. In our previous work, the candidate T4SS genes in *S. suis* were reported for the first time and were proved to have the lethal and virulent as *S. suis* strain^[4-5]. When users search for experimentally verified *virB/D* genes in *S. suis* by using the search engine (Figure 1B) or taxonomy tree (Figure 1C), all the T4SS genes previously identified displayed in the database, including the predicted *virB/D* genes. The 5 strains of *S. suis* are shown in a tabular list (Figure 1A), which contains the number of homologs and putative *virB/D* genes in each genome. Until now, there was no reliable data that can be used to prove the *S. suis* hazing T4SS, except for the homologs of *virB/D* genes mentioned by Li, 2011^[5]. Therefore, there is no experimentally verified *virB/D* gene in the *S. suis* category in our database. Yet, all five *Streptococcus* strains contained the homologs of *virB/D* genes, but only in 4 strains contained the putative functional *virB/D* genes. A simple genome comparison function is available in our database. Users can click the button under the corresponding genome for comparison and result will then be displayed at the bottom of the search page, with a figure representing the genome selected. In the figure, mark Δ represents the location of *virB/D* homologs in the chromosomes, whereas different colors represent related types of *virB/D* genes. More detailed information can be obtained by directly clicking the plus button at each line. A sample page containing detailed information about *virB/D* genes under a genome is shown in Figure 1D. The list of experimentally verified *virB/D* genes, putative functional *virB/D* genes and homologs of *virB/D* genes are then displayed in a tabular format. All information and sequences of identified *virB/D* genes can be freely downloaded. The T4SP database also have shortcuts linking to related literature and other bioinformatics resources (NCBI, Uniprot and KEGG) for each listed gene, which also help the user identify their function. There are 3 putative functional *virB/D* genes (SSU05_0969, SSU05_0971, and SSU05_0973) listed for *S. suis* 05ZYH33 (NC_009442). These 3 genes were identified in a new PAI with a length of 89 kb by using PCR and other molecular technologies^[4]. These three genes are involved in T4SS, mediating the horizontal transfer of PAI between strains^[5], and the lethality effect of the virulent *S. suis* strain was affected^[9], which may

possibly listed under experimentally verified *virB/D* genes in the T4SP database if the functions of these three *virB/D* genes in T4SS are proved further.

As we all known, new disease outbreak is usually caused by rare strain(s). In our opinion, the T4SP database can help quickly access T4SS distribution in newly identified strain(s). Furthermore, some of homolog genes might have similar function. For users who want to further study the function of these genes by gene knocking outs or other ways, our database can provide all homolog sequences and help users to avoid the false-positive or false-negative results caused by homologs in experiments. Our database also helps users to better understand the distribution of T4SS genes in their nearest species/strains by using comparative genome analysis. Actually, our previous researches covered more than 50 species, including *V. cholera*, *S. suis* and *E. Coli*, which were successfully identified by the database, such as *E. Coli* O157^[10] and *Streptococcus*^[11]. Therefore, we hope the database also be useful to other microbiologists and clinical scientists.

The T4SP database is also an open platform that allows users to submit their own data. Data on *virB/D* genes proven by experiments or predicted by our software package are both permitted to be logged into the T4SP database and can be shared with other users upon approval of the administrator.

AVAILABILITY

The T4SP database is available free at website <http://T4SS.bioinfo-icdc.org/>. The mirror website is <http://T4SS.genotyping.org/>.

*This work was supported by the National Natural Science Foundation of China (No. 81201322); the Priority Project on Infectious Disease Control and Prevention 2011ZX10004-001 and 2013ZX10003006-002 by the Chinese Ministry of Science and Technology and the Chinese Ministry of Health; and the Foundation of State Key Laboratory for Infectious Disease Prevention and Control (Grand No. 2011SKLID303).

#Correspondence should be addressed to CHEN Chen, Tel: 8610-58900724; E-mail: chenchen@icdc.cn; Geroge F.GAO, E-mail: gaof@im.ac.cn

Biographical note of first Author: Dr. ZHANG Wen, female, born in 1982, assistant researcher, majoring in microbiology genome.

Received: October 25, 2012;

Accepted: February 28, 2013

REFERENCES

1. Waksman G, Fronzes R. Molecular architecture of bacterial type IV secretion systems. *Trends in biochemical sciences*, 2010; 35, 691-8.
2. Fronzes R, Christie PJ, Waksman G. The structural biology of type IV secretion systems. *Nat Rev Micro*, 2009; 7, 703-14.
3. Abajy MY, Kopec J, Schiwon K, et al. A Type IV-Secretion-Like System Is Required for Conjugative DNA Transport of Broad-Host-Range Plasmid pIP501 in Gram-Positive Bacteria. *J Bacteriol*, 2007; 189, 2487-96.
4. Chen C, Tang J, Dong W, et al. A Glimpse of Streptococcal Toxic Shock Syndrome from Comparative Genomics of *S. suis* 2 Chinese Isolates. *PLoS ONE*, 2007; 2, e315.
5. Li M, Shen X, Yan J, et al. GI-type T4SS-mediated horizontal transfer of the 89K pathogenicity island in epidemic *Streptococcus suis* serotype 2. *Molecular Microbiology*, 2011; 79, 1670-83.
6. Cascales E, Christie PJ. The versatile bacterial type IV secretion systems. *Nat Rev Micro*, 2003; 1, 137-49.
7. Christie PJ, Vogel JP. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends in Microbiology*, 2000; 8, 354-60.
8. Feng Y, Zhang H, Ma Y, et al. Uncovering newly emerging variants of *Streptococcus suis*, an important zoonotic agent. *Trends in Microbiology*, 2010; 18, 124-31.
9. Zhao Y, Liu G, Li S, et al. Role of a Type IV-Like Secretion System of *Streptococcus suis* 2 in the Development of Streptococcal Toxic Shock Syndrome. *Journal of Infectious Diseases*, 2011; 204, 274-81.
10. Juhas M, Crook DW, Hood DW. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cellular Microbiology*, 2008; 10, 2377-86.
11. Zhang W, Rong C, Chen C, et al. Type-IVC Secretion System: A Novel Subclass of Type IV Secretion System (T4SS) Common Existing in Gram-Positive Genus *Streptococcus*. *PLoS ONE*, 2012; 7, e46390.