AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

EPIDEMIOLOGY AND SURVEILLANCE

An Integrative Database of β-Lactamase Enzymes: Sequences, Structures, Functions, and Phylogenetic Trees

Vivek Keshri,[a] Seydina M. Diene,[a] Adrien Estienne,[b] Justine Dardaillon,[c] Olivier Chabrol,[b] Laurent Tichit,[d] Jean-Marc Rolain,[a] Didier Raoult,[a] Pierre Pontarotti[a,e]

[a]Aix Marseille Université, IRD, APHM, MEPHI, IHU-Méditerranée Infection, Marseille, France
[b]Aix-Marseille Université, I2M, UMR-CNRS 7373, Evolution Biologique et Modélisation, Marseille, France
[c]Centre de Recherche en Biologie cellulaire de Montpellier (CRBM), UMR5237, CNRS-Université de Montpellier, Montpellier, France
[d]Institut de Mathématiques de Marseille, Aix Marseille Université, I2M Centrale Marseille, CNRS UMR 7373, Marseille, France
[e]CNRS, Marseille, France

ABSTRACT β-Lactamase enzymes have attracted substential medical attention from researchers and clinicians because of their clinical, ecological, and evolutionary interest. Here, we present a comprehensive online database of β-lactamase enzymes. The current database is manually curated and incorporates the primary amino acid sequences, closest structural information in an external structure database (the Protein Data Bank [PDB]) and the functional profiles and phylogenetic trees of the four molecular classes (A, B, C, and D) of β-lactamases. The functional profiles are presented according to the MICs and kinetic parameters that make them more useful for the investigators. Here, a total of 1,147 β-lactam resistance genes are analyzed and described in the database. The database is implemented in MySQL and the related website is developed with Zend Framework 2 on an Apache server, supporting all major web browsers. Users can easily retrieve and visualize biologically important information using a set of efficient queries from a graphical interface. This database is freely accessible at http://ifr48.timone.univ-mrs.fr/beta-lactamase/public/.

KEYWORDS β-lactamase genes, gene characterization, online database

Countless lives are saved by antibiotics, including β-lactams, which continue to be a leading resource for treatment against bacterial infections. β-Lactamase enzymes are the primary cause of the β-lactam resistance phenomenon because they are able to inactivate β-lactams by cleaving their four-membered β-lactam ring (1). Nowadays, more than 1,300 natural β-lactamase enzymes have been documented (2), and their classification into families helps investigators illustrate how genes are related to each other. All genes of the same family can be closely packed together to form a cluster that can be used to predict the function of novel genes. The molecular classification of β-lactamases is based on the amino acid sequences and divides sequences into four molecular classes, A, B, C, and D (2–4).

The existing publicly available β-lactamase databases offer partial information and have limited functions, such as sequence search by BLAST analysis. Therefore, we designed a new database that contains the primary amino acid sequences, three-dimensional protein structures in an external structure database (Protein Data Bank), phylogenetic trees, and functional profiles. The functional profiles are available in the form of MIC (5) and kinetic parameters (k_{cat}, K_M, and k_{cat}/K_M) (6). A few β-lactamase databases are available in the public domain, such as a Comprehensive β-lactamase Molecular Annotation Resource (CBMAR) (7), the Lahey Clinic Database (www.lahey.org/Studies/), the Bacterial Antimicrobial Resistance Reference Gene Database (BioProject accession number PRJNA313047), the Antibiotic Resistance Genes Database

Citation Keshri V, Diene SM, Estienne A, Dardaillon J, Chabrol O, Tichit L, Rolain J-M, Raoult D, Pontarotti P. 2019. An integrative database of β-lactamase enzymes: sequences, structures, functions, and phylogenetic trees. Antimicrob Agents Chemother 63:e02319-18. https://doi.org/10.1128/AAC.02319-18.

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Pierre Pontarotti, pierre.pontarotti@univ-amu.fr.

Received 5 November 2018
Returned for modification 12 December 2018
Accepted 10 February 2019
Accepted manuscript posted online 19 February 2019
Published 25 April 2019

(ARDB) (8), Lactamase Engineering Database (LacED) (9), the Comprehensive Antibiotic Resistance Database (CARD) (10), a comprehensive database of widely circulated β-lactamases (BLAD) (11), and the Beta-Lactamase DataBase (BLDB; http://bldb.eu/). The contents of these databases are very useful, but none of the aforementioned databases individually provides a complete set of information containing the phylogenetic trees of each molecular class of enzymes, the closest three-dimensional protein structures, and functional profiles. Therefore, a centralized database with updated resources on β-lactamases is required.

This article presents the first release of an integrative β-lactamase database that contains sequences, structures, functions, and phylogenetic information on naturally occurring β-lactamase enzymes. Our motivation in creating this database is to provide a centralized collection of information based on antibiotic resistance genes. The current database offers a user-friendly interface and query results for future studies.

RESULTS

How to access the data. The home page of the current database offers different sections for the data search, such as a particular β-lactamase enzyme, Ambler molecular classes, phylogenetic trees, and literature references of individual proteins. Two types of search options are available in the database, text and advanced search. On the home page of the database, text search can be accomplished by entering simple text or a phrase in the search box; for instance, the enzyme name "TEM." This type of text search option is associated with all data structures. On the other hand, the advanced search option is more computationally productive, since the user can search β-lactamase enzymes according to their Ambler class and get complete information.

Selecting a group of β-lactamases. An easy way to obtain information on enzymes is to browse by name, for instance, by typing "TEM" in the given box and clicking on the green button, "Find β-lactamase." If the user does not want to use an enzyme name but rather other criteria, such as the Ambler class, this can be selected on the home page (Fig. 1). Once the user has selected a particular data set, all available information in the database can be accessed with more details, such as β-lactamase class, the accession numbers of gene sequences, and coding sequence (CDS) regions or location on the genetic DNA support (Fig. 2). It is also possible to download a comma-separated values (CSV) file containing the name, description, accession number, and sequence of the β-lactamase of interest.

Data links, downloads, and interface. Basic information, such as amino acid sequences, closest 3D structure in the Protein Data Bank (PDB) structural database, kinetic parameters, and MIC values, are accessible via the download links provided at the top of the page (Fig. 2). The primary sequence of each gene is available in CSV format. Users can also access the closest three-dimensional structure present in the external database. The structure-related page shows the PDB identifiers (IDs) of the structure, along with their sequence similarity to known β-lactamase sequences. The percent identity, query coverage, query sequences, alignment length, total number of mismatches, gaps, E value, and the bit scores are listed in separate columns. The external database link provides easy access to the complete three-dimensional structure (Fig. 3).

The functional profiles of each β-lactamase are accessible in the form of two data sets; MIC and kinetic parameters. These data were gathered from published literature and manually curated. In the present database, users can access all of these data, such as a list of enzymes, with and without β-lactamase MIC and fold change between them (Fig. 4A). Regarding kinetic parameters, K_M, k_{cat}, and k_{cat}/K_M values are obtainable (Fig. 4B). These parameters represent the catalytic contact for the conversion of substrate to product, substrate concentration (the initial rate of which is one-half of the maximum velocity), and specificity constant or catalytic efficiency, respectively. The kinetic parameters are tabulated and are readily available to users (Fig. 4B).

The phylogenetic trees are available in different file formats, including Newick, Nexus, Portable Network Graphics (PNG), phyloXML, PDB, and Encapsulated PostScript

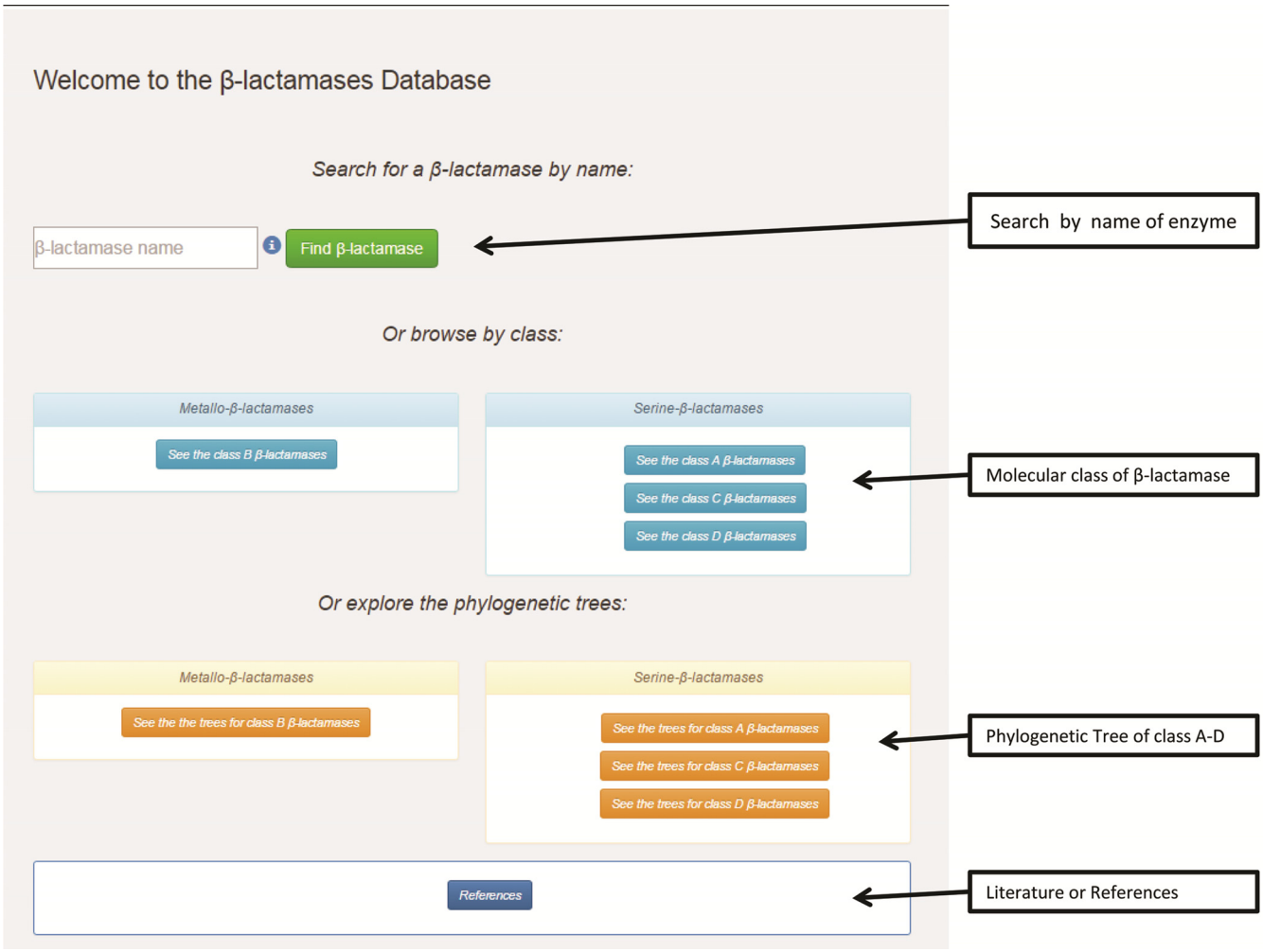

FIG 1 Screenshot depicting the database interface. Database home page and β-lactamase search engine for information collection. Queries using one β-lactamase enzyme name, for example, "TEM" or "Tem." The home page allows a search for all β-lactamase enzymes according to their Ambler molecular classes, such as serine β-lactamases (classes A, C, and D) and metallo-β-lactamases (class B). The bottom section of the page offers the phylogenetic tree of all enzymes according to their respective molecular classes. The references for the individual enzyme (if published in scientific journals) are also available in the reference section.

(EPS) (Fig. 5). The user can download all database resources from the downloadable links. These include (i) primary amino acid sequences of all β-lactamases, (ii) phylogenetic trees of Ambler molecular classes A, B, C, and D, (iii) homologous protein structures from external PDB databases, (iv) MIC and kinetic parameters, and (v) published scientific literature or references.

Data volume summary. This section describes the total data volume available for users. The current release of the database presents four phylogenetic trees of Ambler classes. The phylogenetic trees of classes A, B, C, and D contains 620, 174, 151, and 202 sequences, respectively. For the designing of targeted therapy against any protein target, three-dimensional structural information is essential; we therefore brought to light the three-dimensional crystal structure of β-lactamases. The BLAST search mentioned above to select sequences against a protein structure database (Protein Data Bank) revealed multiple significant hits. Currently, this database provides the external link to 1,516 unique three-dimensional structures in PDB. The extracted functional information from published journal articles in the form of MICs and kinetic parameters is accessible. Users can access MICs and kinetic data for 218 and 137 β-lactamase enzymes.

Results for "TEM%"

Number of proteins: 151

- *Get the csv* of all these β-lactamases, their sequence and their accession number
- See the *structural information* for these β-lactamases
- See the *kinetic information* for these β-lactamases
- See the *MIC data* (Minimum inhibitory concentration) for these β-lactamases

Download Links

#	Protein Name	Protein Description	Accession number	Localisation
1	TEM-1	Betalactamase class A	JQ735917	1-861:861
2	TEM-10	Betalactamase class A	U09188	198-1005:858
3	TEM-101	Betalactamase class A	AF495873	1-861:861
4	TEM-102	Betalactamase class A	AY029354	3-833:831
5	TEM-104	Betalactamase class A	AF516719	215-1075:861
6	TEM-105	Betalactamase class A	AF516720	215-1075:861
7	TEM-106	Betalactamase class A	AY101578	215-1075:861
8	TEM-107	Betalactamase class A	AY101764	207-1067:861
9	TEM-108	Betalactamase class A	AF506748	39-899:861

FIG 2 Graphical user interface of the sequential result. The database contains the primary amino acid sequences. Therefore, this interface provides sequence information, such as the name, molecular class, accession number, and CDS region or localization.

DISCUSSION

Here, we describe a database of *β*-lactamases that brings together most of the available *β*-lactam resistance genes, their phylogenetic trees, structural information, and functional profiles. We have provided a complete set of readily accessible infor-

Results for "TEM%"

Number of results: 15251

- *Get the csv* of all the structural information about those β-lactamases

#	Protein Name	Subject id	Percent Identity	Alignment length	Mismatches	Gap opens	q.start	q.end	s.start	s.end	E-value	Bit-score
1	TEM-1D	gi|67464382|pdb|1ZG4|A	99.3	286	2	0	1	286	1	286	0	588.0
2	TEM-1D	gi|67464385|pdb|1ZG6|A		286	3	0	1	286	1	286	0	587.0
3	TEM-1D	gi|822599600|pdb|4ZJ1|A	98.95	286	3	0	1	286	1	286	0	586.0
4	TEM-1D	gi|822599601|pdb|4ZJ2|A	98.6		4	0	1	286	1	286	0	585.0
5	TEM-1D	gi|325533970|pdb|3P98|A	98.6	286	4	0	1	286	1	286	0	583.0
6	TEM-1D	gi|822599602|pdb|4ZJ3|A	98.25	286	5	0	1	286	1	286	0	582.0
7	TEM-1D	gi|8569630|pdb|1ERO|A	100	263	0	0	24	286	1	263	0	541.0
8	TEM-1D	db|1ESU|A	99.62	263	1	0	24	286	1	263	0	540.0
9	TEM-1D	gi|15988093|pdb|1JTD|A	99.62	263	1	0	24	286	1	263	0	539.0
10	TEM-1D	gi|21465759|pdb|1JWV|A	99.62	263	1	0	24	286	1	263	0	539.0

PDB ID

Identity (%)

Enzymes

FIG 3 Graphical user interface of the structural information. The database incorporates primary sequences of *β*-lactamases. These sequences have greater similarities with the available nonredundant three-dimensional structures in the structure database. This snapshot presents the closest structure of TEM-1D as an example in the external PDB database.

A

MIC data for "TEM%"

Number of results: 402

- Get the csv of all the physiological data linked to those β-lactamases

#	Protein Name	Antibiotics	Without Bla MIC (μg/ml)	With Bla MIC (μg/ml)	Fold
1	TEM-116	Cefotaxime	0.0625	32	512
2	TEM-116	Imipenem	0.0625	0.125	2
3	TEM-116	Cefoxitin	2	2	1
4	TEM-116	Ampicillin	4	256	64
5	TEM-116	Aztreonam	0.0625	32	512

B

Kinetic data for "TEM%"

Number of results: 191

- Get the csv of all those kinetic data

#	Protein Name	Antibiotic	k_{cat} (s^{-1})	Km (μM)	k_{cat}/K_M
1	TEM-109	Amoxicillin	66±3	15±1	4
2	TEM-109	Benzylpenicillin	171±2	20±2	8
3	TEM-109	Aztreonam	38±5	130±15	0.3
4	TEM-109	Cefuroxime	22±3	93±7	0.2
5	TEM-109	Ticarcillin	42±2	14±2	3

FIG 4 Graphical user interface of functional profile. (A) MIC (without and with Bla, along with fold change between them). (B) Enzymatic activity of β-lactamases with their corresponding β-lactam antibiotic, in the form of k_{cat}, K_M, and k_{cat}/K_M values.

mation on β-lactamases to complement and to assist in the growth of its study. The current database incorporates compiled data from existing resources, such as published literature and public domain databases, and analyzes the data for meaningful information. For the collection and processing of biological information, a set of 1,147 amino acid sequences was selected. A total of 54.05% sequences belonged to class A, 15.17% to class B, 13.17% to class C, and 17.61% to class D. Because of unpublished articles/data and a lack of information in the literature, we were unable to get experimental data on all enzymes. Therefore, the present version of the database offers a lesser number of functional profiles (MICs and kinetic parameters) compared to the 1,147 sets of β-lactamases.

The β-lactamase enzyme has a crucial role in β-lactam antibiotic resistance. There is an urgent need to understand the roles and mechanisms of enzymes in antibiotic resistance, especially in β-lactam resistance. The β-lactamase databases developed previously compiled information which was insufficient to understand their evolutionary relations and their functional or structural information. The Antibiotic Resistance Genes Online (ARGO) (12), ARDB (8), CBMAR (7), CARD (10), BLAD (11), and Antibiotic Resistance Gene-ANNOTation (ARG-ANNOT) (13) databases are popular antibiotic resistance databases. These databases cover many aspects of β-lactamase enzymes, such as sequence information, subfamily phylogenetic trees, protein sequences, nucleotide sequences, and mutational information. However, these databases lack functional

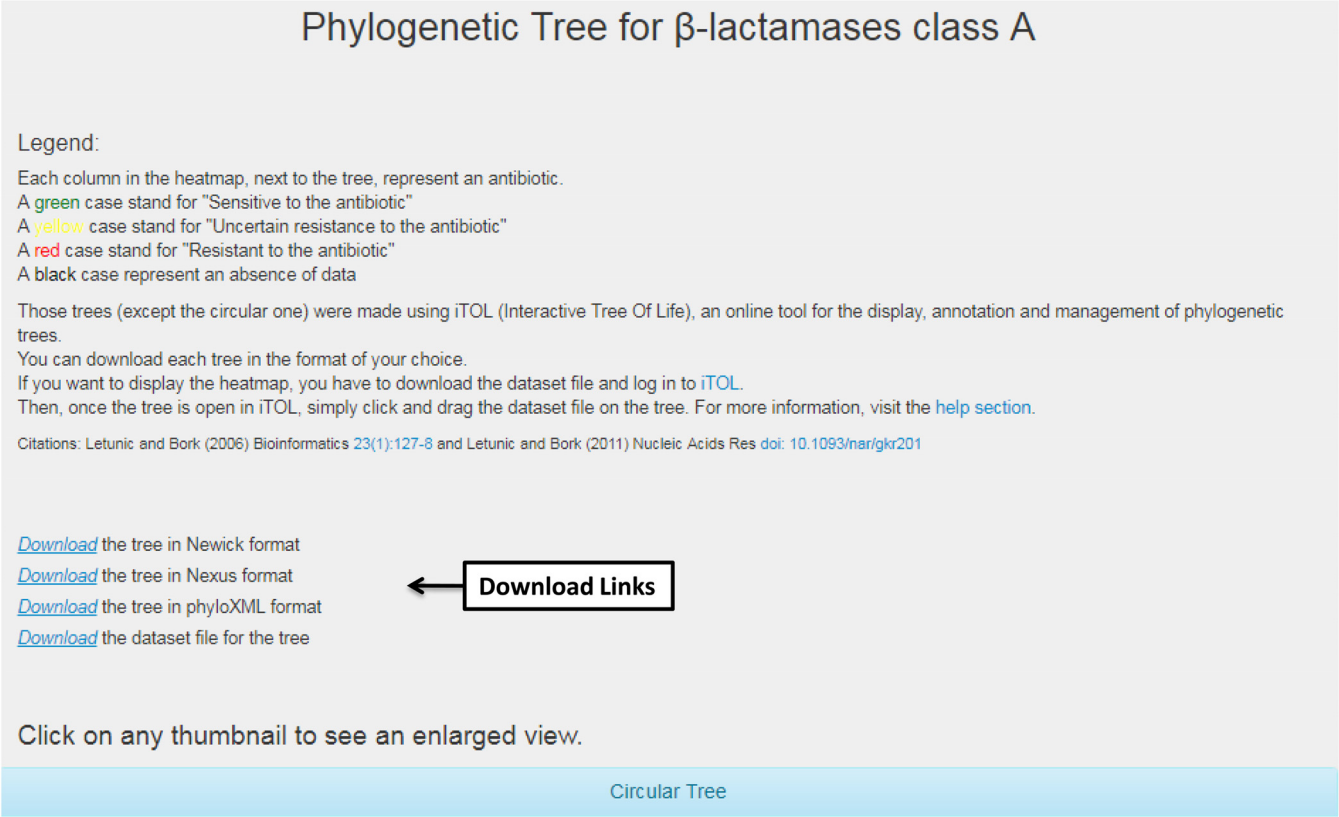

FIG 5 Graphical user interface of a phylogenetic tree. The tree describes the evolutionary relation among β-lactamases at an intraclass scale.

information (MICs and kinetic parameters), large-scale phylogenetic trees of molecular classes, and all possible homologous enzymes or protein 3D structures. The phylogenetic trees provided in the database described here will help researchers in the identification of the closest enzymes and functional profiles providing hydrolyzing properties.

Because of the huge diversity of β-lactamase enzymes and rapid identification of novel enzymes from different environments, this database aims at becoming the first index of all available information on the topic and will continue to be updated over time. Future efforts will also strive to include all the diversity of β-lactamase enzymes present in the environmental and human microbiome, as well as newly reported β-lactamases. Currently, this database contains only protein 3D structure IDs and their similarities with the core input set of query sequences. Therefore, we will incorporate three-dimensional coordinate values in the future version.

In conclusion, the current database is an online database for studying β-lactam resistance enzymes. This database contains comprehensive information, and researchers can use and/or share the contents of the database without any restriction. The current database is the first substantial effort to integrate sequences, structures, functions, and phylogenetic trees of β-lactamases. Because of the primary amino acid sequences, structures, phylogenetic trees, and functional information, the present database is different from other existing databases, making it a useful resource for researchers. With the vast amount of sequences and three-dimensional structural data becoming available, future improvements of the current database will focus on expanding available information. With this release, we are committed to being proactive in understanding the needs of all our users and will provide a continuously updated version of this database.

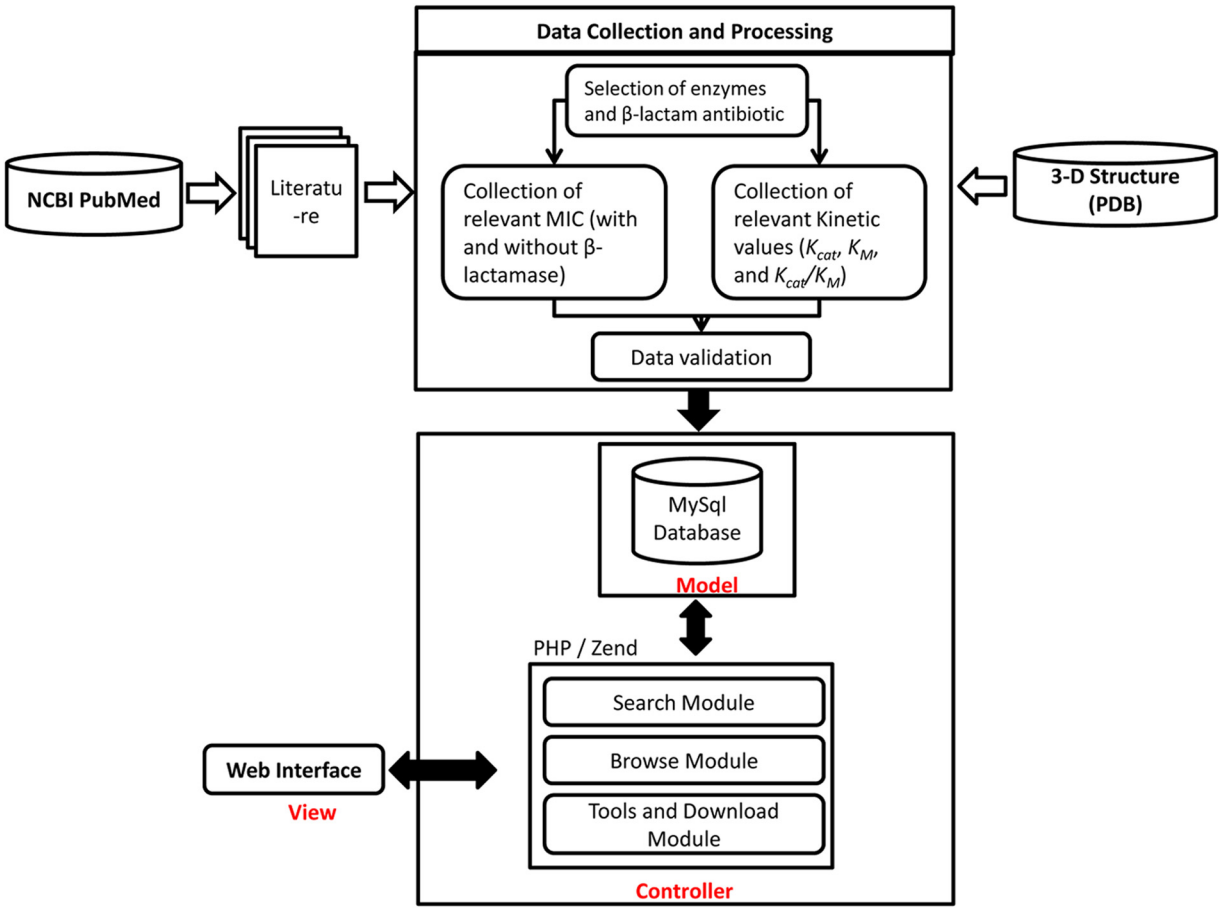

FIG 6 Database scheme. This figure explains the data collection steps and literature review, followed by the methods used to retrieve information.

MATERIALS AND METHODS

Data collection and analysis. (i) Input sequences and classification. A total of 1,147 amino acid sequences of β-lactamase were downloaded from the ARG-ANNOT database (13). These sequences were then classified into the four Ambler molecular classes. Classes A, B, C, and D include totals of 620, 174, 151, and 202 sequences, respectively.

(ii) Phylogenetic tree construction. All amino acid sequences of each class were aligned through the MUSCLE algorithm (14), followed by truncation of poor alignment regions by the trimAl program (15). Subsequently, the phylogenetic tree was constructed in FastTree (16) with the approximate maximum-likelihood method. FigTree-v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) was used for visualization of the phylogenetic tree.

(iii) Three-dimensional structure available in an external database, the Protein Data Bank (PDB). A total of 81,117 nonredundant three-dimensional protein structures were downloaded from the PDB database. The total of 1,147 β-lactamase sequences was used as queries and subjected to BLAST search analysis (17) against the nonredundant structures from the PDB database (18). The protein BLAST search was accomplished with default settings, and results are available in tabular format in the current database with significant BLAST hits, structure IDs, identity, alignment regions, expected thresholds, and bit score.

(iv) Collection of MIC and kinetic data from published literature. The MIC is the lowest concentration of the antibiotic that inhibits the visible growth of a microorganism after overnight incubation (5). The available MIC value (with and without β-lactamase) of each enzyme with respective β-lactam antibiotic composition was collected from the most relevant published literature and followed by the calculated fold change between them. Here, the meaning of with and without β-lactamases simply means that composition contains or does not contain β-lactamase along with β-lactam antibiotics. Also, the kinetic parameters (k_{cat}, K_M, and k_{cat}/K_M) were obtained for available antibiotics.

(v) Database design and implementation. The current database is implemented in MySQL 5.6 and aims at being readily accessible according to different search criteria. The online release of the database works in all key browsers and is compatible with all operating systems. The web interface was developed with Zend Framework 2.5 using PHP 5.5 and an Apache 2.4 web server, which runs on Linux (Ubuntu 15.10). Figure 6 depicts the database workflow, which explains the data collection process from published journal articles and retrieval of information from the web interface.

ACKNOWLEDGMENTS

We gratefully acknowledge TradOnline for English correction of manuscript and Olivier Chabrol and Sylvain Buffet for uploading the database and maintenance of the website.

This work was supported by the French Government under the "Investments for the Future" program managed by the National Agency for Research (ANR), Méditerranée Infection grant 10-IAHU-03.

We have no conflicts of interest to declare.

REFERENCES

1. Holten KB, Onusko EM. 2000. Appropriate prescribing of oral beta-lactam antibiotics. Am Fam Physician 62:611–620.
2. Bush K. 2013. The ABCD's of β-lactamase nomenclature. J Infect Chemother 19:549–559. https://doi.org/10.1007/s10156-013-0640-7.
3. Ambler RP, Coulson AF, Frère JM, Ghuysen JM, Joris B, Forsman M, Levesque RC, Tiraby G, Waley SG. 1991. A standard numbering scheme for the class A beta-lactamases. Biochem J 276:269–270. https://doi.org/10.1042/bj2760269.
4. Ambler R. 1980. The structure of β-lactamases. Philos Trans R Soc Lond B Biol Sci 289:321–331. https://doi.org/10.1098/rstb.1980.0049.
5. Andrews JM. 2001. Determination of minimum inhibitory concentrations. J Antimicrob Chemother 48:5–16. https://doi.org/10.1093/jac/48.suppl_1.5.
6. Koshland DE. 2002. The application and usefulness of the ratio k_{cat}/K_M. Bioorg Chem 30:211–213. https://doi.org/10.1006/bioo.2002.1246.
7. Srivastava A, Singhal N, Goel M, Virdi JS, Kumar M. 2014. CBMAR: a comprehensive β-lactamase molecular annotation resource. Database (Oxford) 2014:bau111. https://doi.org/10.1093/database/bau111.
8. Liu B, Pop M. 2009. ARDB—Antibiotic Resistance Genes Database. Nucleic Acids Res 37:D443–D447. https://doi.org/10.1093/nar/gkn656.
9. Thai QK, Bös F, Pleiss J. 2009. The Lactamase Engineering Database: a critical survey of TEM sequences in public databases. BMC Genomics 10:390. https://doi.org/10.1186/1471-2164-10-390.
10. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348–3357. https://doi.org/10.1128/AAC.00419-13.
11. Danishuddin M, Baig MH, Kaushal L, Khan AU. 2013. BLAD: A comprehensive database of widely circulated beta-lactamases. Bioinformatics 29:2515–2516. https://doi.org/10.1093/bioinformatics/btt417.
12. Scaria J, Chandramouli U, Verma SK. 2005. Antibiotic Resistance Genes Online (ARGO): a database on vancomycin and beta-lactam resistance genes. Bioinformation 1:5–7. https://doi.org/10.6026/97320630001005.
13. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain J-M. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220. https://doi.org/10.1128/AAC.01310-13.
14. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
15. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
16. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
17. 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.
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235.