

Systematic Identification and Classification of β -Lactamases Based on Sequence Similarity Criteria: β -Lactamase Annotation

Evolutionary Bioinformatics
Volume 14: 1–11
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934318797351



Melise Chaves Silveira¹, Rangeline Azevedo da Silva¹,
Fábio Faria da Mota¹, Marcos Catanho², Rodrigo Jardim¹,
Ana Carolina R Guimarães² and Antonio B de Miranda¹

¹Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil. ²Laboratório de Genômica Funcional e Bioinformática, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil.

ABSTRACT: β -lactamases, the enzymes responsible for resistance to β -lactam antibiotics, are widespread among prokaryotic genera. However, current β -lactamase classification schemes do not represent their present diversity. Here, we propose a workflow to identify and classify β -lactamases. Initially, a set of curated sequences was used as a model for the construction of profiles Hidden Markov Models (HMM), specific for each β -lactamase class. An extensive, nonredundant set of β -lactamase sequences was constructed from 7 different resistance proteins databases to test the methodology. The profiles HMM were improved for their specificity and sensitivity and then applied to fully assembled genomes. Five hierarchical classification levels are described, and a new class of β -lactamases with fused domains is proposed. Our profiles HMM provide a better annotation of β -lactamases, with classes and subclasses defined by objective criteria such as sequence similarity. This classification offers a solid base to the elaboration of studies on the diversity, dispersion, prevalence, and evolution of the different classes and subclasses of this critical enzymatic activity.

KEYWORDS: β -lactamase, class, subclass, identification, sequence similarity

RECEIVED: March 15, 2018. **ACCEPTED:** August 8, 2018.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Melise Chaves Silveira, Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, Fiocruz, Av. Brasil 4365, Manguinhos, Rio de Janeiro RJ 21040-900, Brazil. Email: melisechaves@gmail.com

Introduction

The increasing amounts of genomic data produced by next-generation sequencing technologies made β -lactamases (BLs), the enzymes responsible for the irreversible inactivation of β -lactam antibiotics, one of the most numerous families of proteins studied to date.¹ First described by Abraham and Chain,² BLs can be found in pathogenic or commensal bacteria, isolated from humans or varied environments.³ Due to its great medical importance and clinical impact, several groups directed efforts in the development of a proper BL classification scheme, usually based on functional or structural criteria.^{4–6}

Function-based classification is achieved using experimental data to link the enzyme to its clinical role.^{6,7} The determination of these parameters for a large number of BLs, however, may be relatively costly and time-consuming. Also, they do not generate sequence data, essential for studies involving molecular evolution.⁸

Currently, the most widely used classification scheme for BLs is the Ambler structural classification, which is based on sequence similarity, and separates BLs into 4 classes: the classes A, C, and D of serine- β -lactamases (SBLs) and the class B of metallo- β -lactamases (MBLs).^{4,9,10} Class B is further divided into subclasses B1, B2, and B3, using sequence conservation data.¹¹

Despite SBLs and MBLs are able to break amide and ester bonds (EC 3.5.2.6), they belong to 2 distinct protein super-families that do not share a common ancestor.¹² Considering SBL's tertiary structures, they are similar enough among

themselves to be considered homologous,⁵ whereas the differences between their primary structures and catalytic mechanisms justify their division into classes A, C, and D.¹³

Experimental data indicate that the 3 subclasses of MBL should not be treated as equally separated groups. Subclasses B1 and B2 have detectable sequence similarity between them but not with B3,¹⁴ and structural evidence strongly suggest different Most Recent Common Ancestor between the group formed by subclasses B1 and B2 and the subclass B3.¹⁵

Based on this structural information, a reorganization of BL at 4 hierarchical levels was proposed by Hall and Barlow.⁵ In this scheme, the former subclasses B1 and B2 were merged and renamed as class MB, whereas subclass B3 was renamed as class ME. Thus, the 5 BL classes (third classification level) are SA, SC, SD, MB, and ME, and subclasses MB1 and MB2 represent the fourth and last hierarchical level.⁵

Moreover, recent studies employing phylogeny and amino acid-based sequence similarity networks showed that classes SA and SD could be further divided into 2 different groups, suggesting new BL subclasses.^{1,16} Indeed, the current molecular classification of BLs does not represent its actual sequence diversity, and there is yet no precise definition of various classes and subclasses.^{1,5,16}

Profiles Hidden Markov Models (HMM) and sequence clustering using similarity underlie the workflow proposed here. It is based on a previous hierarchical scheme which



reflects the distinct evolutionary origins of SBLs and MBLs.⁵ Application of this scheme to 2774 assembled bacterial genomes provided improvements in BL annotation and in the knowledge about BLs distribution among bacteria phyla, confirming previous studies suggesting new BL subclasses.^{1,16} Finally, our results propose the existence of a new BL class with fused domains and extended action spectrum.

Methodology

Data collection and preparation

On March 22, 2016, an online search for BL structures using the EC number (3.5.2.6) in the National Center for Biotechnology Information (NCBI) Structure Database¹⁷ returned 516 entries. Records with the word “mutant” in the description were excluded. The PDB (Protein Data Bank) IDs were retrieved and used for downloading PDB files and their corresponding FASTA files from RCSB PDB.¹⁸ The data were filtered using the following criteria: duplicate atoms positions were removed, the resolution of the structure should be less than 3 Å, and only monomers or the chain A from homomultimers were used.

A Non-Redundant Beta-Lactamase Dataset (NRBLD) was constructed with protein sequences retrieved from 7 different antibiotic resistance databases. Four databases are specific for BLs: (1) Comprehensive Beta-lactamase Molecular Annotation Resource (CBMAR, downloaded on August 2015),¹⁹ (2) The Institute Pasteur Database (downloaded on October 2015),²⁰ (3) DLact Antimicrobial Resistance Gene Database (downloaded on October 2015),²¹ and (4) Lactamase Engineering Database (LacED, downloaded on August 2015).²² The Metallo-Beta-Lactamase Engineering Database (MBLED, release 1.0) is the only specific for MBLs.²³ The remaining 2 databases, Comprehensive Antibiotic Resistance Database v1.0.0 (CARD)²⁴ and Resfams v1.2,³ have protein sequences related to antibiotic resistance in general. To recover only BL sequences, we search for those with the term “bla” and “beta” in the headers. Identical sequences have been removed using CD-HIT²⁵ at a 100% identity threshold. The following methodology steps are summarized in Figure 1.

Clustering curated BL structures and sequences

The clustering assays were made using BL structures and their corresponding amino acid sequences from PDB, in an attempt to reproduce the BL classes and subclasses proposed by previous works. The MaxCluster program²⁶ was used to cluster the structures based on root mean-squared deviation and hierarchical clustering, applying 3 tests: single, average, and maximum linkage. The BLASTClust v2.2.26 program²⁷ was used for the hierarchical clustering of the amino acid sequences, which performs a single linkage type clustering based on pairwise matches found by BLAST. Different threshold values of “BLAST score density” (BLAST score divided by the alignment length) and “minimum length coverage” were tested. An E value of 1E-05 was used.

Building a profile HMM for each BL class

The profiles HMM were constructed using the clusters of sequences corresponding to the 5 BL classes.⁵ Because these sequences came from proteins with a resolved structure, they were considered reliable to be used to construct the profiles. Sequences from each cluster were saved in separate multi-FASTA files. Identical sequences were removed using CD-HIT²⁵ at a 100% identity threshold. Sequences in each file were aligned using MUSCLE.²⁸ Profiles HMM were built from each alignment using the program *hmmbuild* from the HMMER package v3.1b2.²⁹

Calibration and validation of the profiles HMM

The profiles HMM generated were used against the protein sequences at NRBLD and superfamilies 3.40.710.10 (DD-peptidase/BL like) and 3.60.15.10 (Metallo-BL like) from Protein Structure Classification Database (CATH) (downloaded on March 17, 2016).³⁰ *Hmmsearch* from the HMMER package v3.1b2 with an E value of 1E-05 was employed.²⁹

The Class profile Specificity index (CpSp) evaluates whether each profile identifies a unique group of sequences. CpSp was calculated by dividing the number of NRBLD sequences identified exclusively with a given profile (Ne) by the total number of NRBLD sequences recovered by it (T), including intersections with others profiles results [$CpSp = (Ne/T) * 100$]. Profiles with CpSp below 100% were calibrated. Sequences from other classes that should not be identified were used as “negative training sequences,” following the HMM-ModE protocol.³¹ After this, the *hmmsearch* Gathering Threshold (GA) parameter was used, substituting the E value.

A total of 851 amino acid sequences of DD-peptidase/BL-like superfamily downloaded from CATH³⁰ were used to construct an unrooted Maximum Likelihood phylogenetic tree in MEGA-CC v7.0.18,³² using the Jones-Taylor-Thornton model, partial deletion for gaps/missing data treatment (95% site coverage cutoff) and 500 bootstrap replicates. Using in-house Perl scripts, the BL sequences were manually labeled with their respective class, and the clade node of each BL class was identified. The sequences in these clades were allocated in corresponding multi-FASTA files.

Graphics were created with all the sequences retrieved by each profile against the DD-peptidase/BL superfamily with their respective HMM bit score. These results were used to calibrate the profiles. A new *hmmsearch* parameter of HMM bit score threshold was established to separate true BL from other sequences in the superfamily. The Function Specificity (FSp) index verifies whether all sequences retrieved by a profile have the BL activity already described. For this, the number of superfamily sequences identified by a profile that has BL activity (Nbl) was divided by the total number of superfamily sequences identified by the profile (T) [$FSp = (Nbl/T) * 100$].

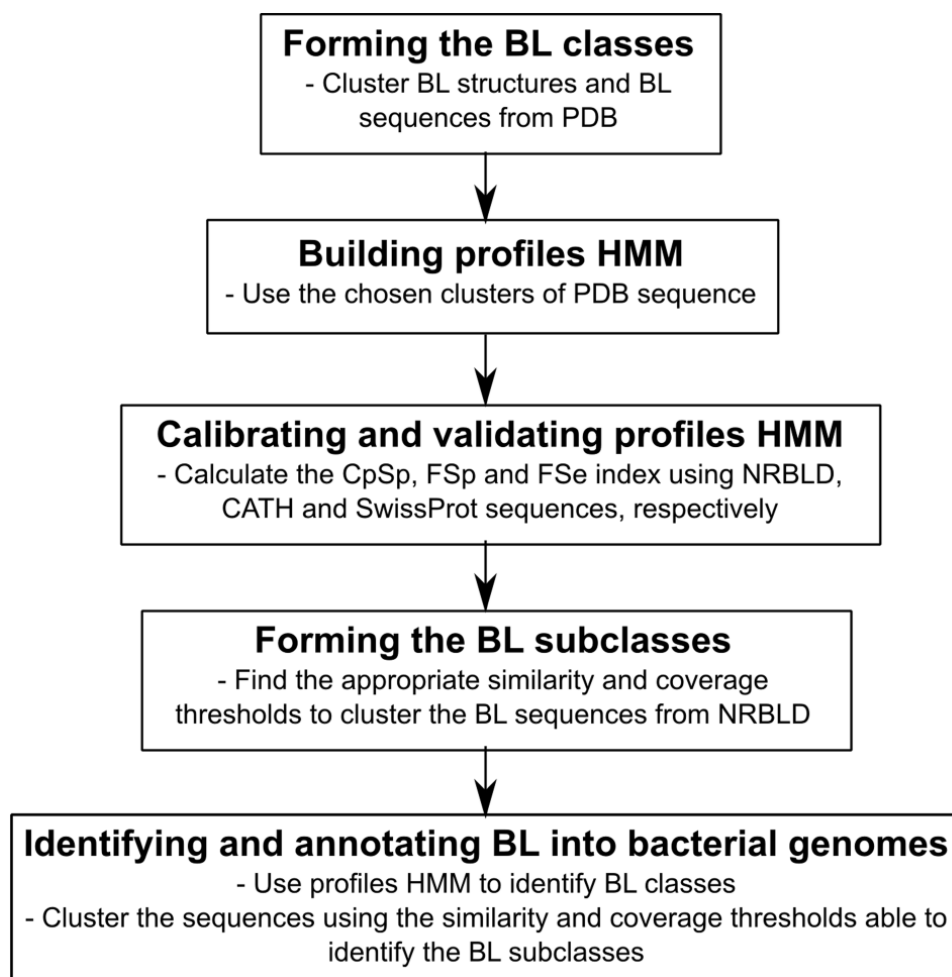


Figure 1. Main conceptual steps of the workflow.

BL indicates β -lactamase; CATH, Protein Structure Classification Database; PDB: Protein Data Bank; CpSp, Class profile Specificity; FSe, Function Sensibility; FSp, Function Specificity; NRBLD, Non-Redundant Beta-Lactamase Dataset.

A total of 144 sequences are attributed to the Gene Ontology (GO) molecular function “BL activity” (GO:0008800) in the SwissProt database (downloaded on March 2, 2016).³³ The validation index Function Sensitivity (FSe) evaluates the ability of all profiles together to identify the sequences annotated with the BL function in SwissProt. To calculate this index, the number of identified SwissProt sequences attributed to the BL GO term (N_{go}) was divided by 144 [$FSe = (N_{go}/144) * 100$].

Validation of thresholds used to form BL subclasses

The results obtained with the curated PDB data set were used as a basis to group the NRBLD sequences into subclasses. The BLs from PDB were added to NRBLD, and the profiles HMM were used against them.

In order to reproduce BL subclasses, different thresholds of “minimum length coverage” were tested to cluster the sequences in each class, together with the “BLAST score density” thresholds previously established using the PDB sequences. The sequences in each cluster were annotated using the BLASTP³⁴ best hit (v2.2.28) against the nonredundant NCBI protein database.¹⁷

The size of specific domains of BLs was used to stipulate a minimum size necessary for the enzyme to be functional.³⁵ The Pfam³⁶ models most common in MBL (PF00753), class SA (PF13354), and class SC (PF00144) have a length of 197, 202, and 330 residues, respectively. A total of 214 residues domain have been attributed to class SD enzymes.³⁷

The clusters formed were categorized as follows: (1) clusters corresponding to those obtained in the hierarchical classification of functionally characterized BLs (PDB sequences) and (2) clusters containing sequences that do not fit into the previous established similarity and coverage thresholds.

Comparison of the improved profiles HMM with Pfam profiles and BL motifs

To test the efficiency of our profiles HMM in identifying and classifying BL sequences, these were compared with profiles available in the Pfam database³⁶ and with motifs specific to BL classes from different sources.

The profiles HMM from Pfam³⁶ were used to search for BL sequences from PDB, and the CpSp indexes were calculated accordingly. Seven Pfam profiles (PF00144.22, PF13354.4,

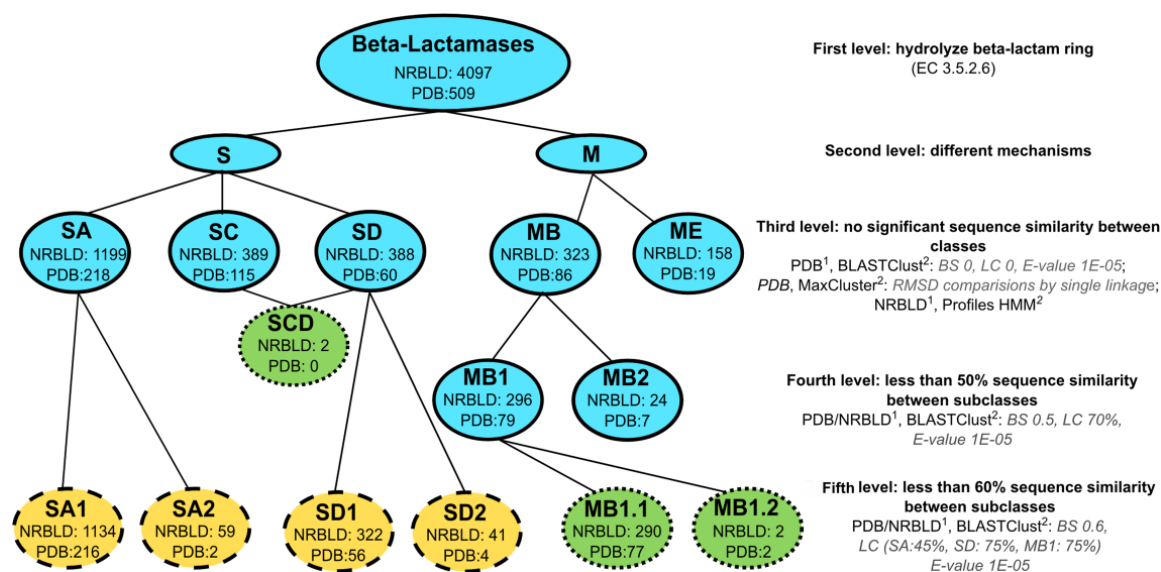


Figure 2. Hierarchical classification of β -lactamases.

The number of Protein Data Bank (PDB) and Non-Redundant Beta-Lactamase Dataset (NRBLD) sequences in each cluster after clustering is shown (PDB identifiers can be obtained in Table S2). 1 indicates the sequence dataset used for clustering. 2 indicates the program used for clustering, followed by the parameters used. Blue: Ambler's classification with real relationships as shown by Hall and Barlow⁵; green and dotted: new groups proposed for the first time in this study; yellow and dashed: groups described recently and confirmed in this work^{1,16}; BS: BLAST score density; LC: minimum length coverage; BL: Beta-lactamase; S: Serine-BL; M: Metallo-BL; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD; MB: Metallo-BL class MB (former subclasses B1 and B2); ME: Metallo-BL class ME (former subclass B3); SCD: Serine BL class SC-class SD; MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; MB1.1: Metallo-BL subclass MB1.1; MB1.2: Metallo-BL subclass MB1.2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2.

PF00753.25, PF12706.5, PF13483.4, PF14597.4, and PF16661.3) were downloaded on December 2016.

Motifs specific for BL classes obtained from the literature were evaluated for their ability to distinguish between classes and subclasses. An in-house Perl script was developed to identify the motifs in the BL sequences from PDB.

Identification and classification of BLs in fully assembled genomes

The workflow proposed here was applied to identify and classify BL sequences present in 2774 bacterial strains with fully assembled genomes that were deposited in NCBI.¹⁷ A genome was defined as the complete set of chromosome and plasmids of each strain. The protein sequences present in each genome were downloaded (June 2016). Using the profiles HMM, the BL classes were formed, which were then separated into their respective subclasses applying the BLASTClust program²⁷ with the "BLAST score density" and "minimum length coverage" thresholds set in the previous step. The taxonomic information of each sequence was determined using the Genome Online Database (GOLD)³⁸ and in-house Perl scripts.

The scripts, profiles HMM, and instructions required to apply the workflow presented here, in addition to the data used for the searches, are available at <https://github.com/melisesilveira/betaLactamase-classification.git>.

Results

Clustering curated BLs structures and sequences

In total, 509 PDB structures and their respective sequences were used in the clustering and in the construction of profiles

HMM. Of these, 208 were monomers and 301 were homomultimeric proteins.

Clustering of the curated set of structures using single linkage was consistent with the BL classification as proposed by Hall and Barlow, producing the same 5 classes (Figure 2).⁵ Average and maximum linkage resulted in 7 and 18 clusters, respectively, and were not further used.

The PDB sequences clustering applying "BLAST score density" thresholds of 0 and 0.5 were also consistent with the Hall and Barlow's classification scheme,⁵ producing the 5 BL classes and 2 MB subclasses, respectively. Using a "BLAST score density" threshold equal to 0.6, the subclasses proposed in previous works for classes SA and SD were reproduced,^{1,16} as well as the division of MB1 into 2 groups (Figure 2). The length coverage did not influence these results, and therefore a "minimum length coverage" of zero was chosen.

Separate clusters, one composed by 2 BL TEM-1 fused to a maltose-binding protein and another with 5 *Escherichia coli* penicillin-binding protein 5 (PBP5), were created in all clustering assays and were excluded because they are not true BLs, being wrongly associated with EC 3.5.2.6 in PDB. All sequence and structural clustering results are presented in Tables S1 to S3.

Building, calibration, and validation of the profiles HMM

Profiles HMM were built for each of the 5 clusters created using curated PDB sequences. These profiles were calibrated and validated using 3 sets of sequences: (1) NRBLD, our non-redundant protein set containing 4097 sequences; (2) the

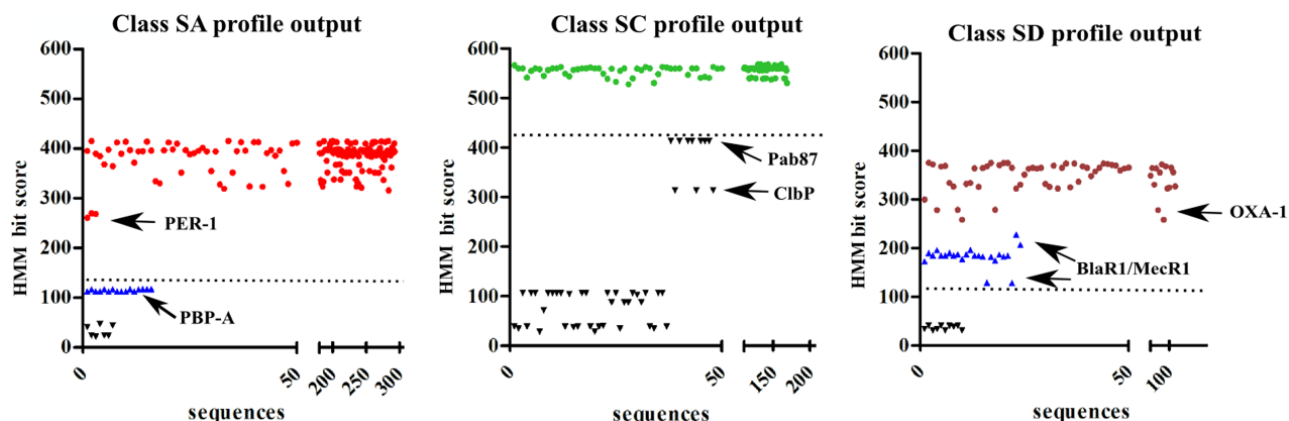


Figure 3. Plot of the 851 DD-peptidase/ β -lactamase-like superfamily (CATH 3.40.710.10) sequences.

Each dot represents a sequence identified by the profile with its HMM bit score result. The HMM bit score reflects whether the sequence is a better match to the profile model (positive score) or the null model of nonhomologous sequences (negative score). The dashed lines in the graphs indicate the established HMM bit score threshold for each profile. X: numbered sequence order, Y: HMM bit score. Red circle: 295 sequences from BL class SA, green circle: 177 sequences from BL class SC, brown circle: 103 sequences from BL class SD, blue triangle: proteins inserted within the BL clades that are not functionally characterized as BLs, and black inverted triangle: other non-BL sequences from DD-peptidase/ β -lactamase-like superfamily. Arrows indicate particular proteins quoted in the text; BL: Beta-lactamase.

DD-peptidase/BL- and MBL-like superfamilies, with 851 and 399 protein sequences, respectively; and (3) SwissProt/UniProt³³ database, with 550 552 sequences. The profiles for the SBL and MBL classes were analyzed separately as these 2 groups belong to different CATH superfamilies. The data and profiles HMM are available at <https://github.com/melisesilveira/betaLactamase-classification.git>.

The profiles for classes SA, SC, and SD were searched against NRBLD recovering 1292, 1121, and 396 sequences, respectively. The profile of the class SA was 100% specific for the class (CpSp), whereas the profiles of the classes SC and SD were 99% specific. Two sequences were identified by both profiles, BL LRA-13 (ACH58991.1) and an enzyme annotated as “class C BL” of *Janthinobacterium* sp. (WP_008451281.1). Both have domains of SC and SD classes, which led us to propose a new class, SCD, composed of BLs with fused domains (Figure 2).

The profiles for MBL classes recovered 527 (MB, CpSp=84%) and 612 (ME, CpSp=86%) NRBLD protein sequences, 84 of which were identified by both profiles. Therefore, the profiles were optimized, and after the calibration they recovered 323 (MB) and 159 (ME) sequences, without intersections, reaching 100% CpSp.

To compare the calibrated and noncalibrated MBL profiles, they were tested against the 598 MBL protein sequences from the MBLED database.²³ Initial profiles identified 440 (MB, 85% CpSp) and 222 (ME, 70% CpSp) sequences, which represents 99.8% of the database. However, calibrated profiles identified 424 (MB, 100% CpSp) and 164 (ME, 100% CpSp) sequences, representing 98.3% of the database. The remaining 1.7% was composed of protein fragments ranging from 75 to 131 amino acids, meaning that the calibrated profiles were able to identify all the complete BL sequences.

The phylogenetic relationships of the SBL superfamily proteins showed a few sequences with no BL activity inserted into BL clades (Figure S1). The penicillin-binding proteins A (PBP-A) are in the inner branches of class SA and have a

structure very similar to the BL PER-1 (subclass SA2) but do not have BL activity.³⁹ The regulatory proteins BlaR1 (and its cognate MecR1) are in inner branches of the class SD. Their extracellular domains are phosphorylated by β -lactams and, consequently, these proteins regulate resistance to these antibiotics in *Staphylococcus aureus*.⁴⁰

Initially, the profiles HMM for SBLs identified these sequences and others outside the BLs clades (all non-BLs). However, based on their HMM bit score values, a total of 75 non-BL sequences could be excluded (Figure 3). The separation of true SD2 BLs from BlaR1 proteins in this step was not possible because the HMM bit score of the single structure available of this subclass present in the CATH database³⁰ (OXA-1) was very close to some BlaR1 proteins (Figure 3). Additional tests have shown that when other variants of this subclass are included, their scores are smaller than those of some BlaR1 sequences. This can be explained by the structural homology of the extracellular sensor domain of BlaR1 to BLs from subclass SD2.⁴¹

The enzymes ClbP and Pab87 are associated with the BL EC number in the PDB and share a significant similarity between their active sites with the BLs from class SC.⁴² However, they can be separated from the class SC BLs by both phylogeny (Figure S1) and HMM bit score (Figure 3). Sequences in each BL clade and their respective HMM bit score are shown in Table S4.

HMM bit score thresholds of 120, 430, and 120 were defined for classes SA, SC, and SD, respectively (Figure 3). After the utilization of these bit scores, the FSp was equal to 100%, 100%, and 81%, for classes SA, SC, and SD, respectively. These thresholds retrieved 1199, 389, and 388 sequences from NRBD, respectively.

After the calibrations, all BL profiles together retrieved 132 proteins from SwissProt³³: 82, 10, 24, 15, and 1 sequences with the profiles for classes SA, SC, SD, MB, and ME, respectively. In all, 125 out of the 144 sequences annotated with the BL GO term in this database were identified by the profiles (87% FSe).

The remaining 19 sequences not identified as BLs presented one of the following annotations: BL fragments, BL-like protein, DacA carboxypeptidase, hydroxyacylglutathione hydrolases, and ribonuclease. Also unidentified were Hcp family proteins and a sequence described as a class SA BL PenA.⁴³ Hcp proteins, a family of cysteine-rich PBPs, do not have a significant sequence or structural similarity to known BLs.⁴⁴ A BLASTP³⁴ search in the NCBI protein database¹⁷ with the putative PenA returned as best-hit a class SC protein with only 13% coverage, meaning that it is probably not a BL and certainly not a PenA. Seven BlaR1/MecR1 proteins were also identified by the profiles but are regulatory proteins not associated with the GO term for BL.

Validation of BLASTClust thresholds to form BL subclasses

The “BLAST score density” values applied to the curated BL sequences to form the subclasses were validated in the PDB sequence set plus NRBLD. A significant length variation was observed between the BL sequences in PDB and those in NRBLD. The sequences in PDB range from 219 to 447 amino acids, whereas in NRBLD, they range from 96 to 619 amino acids. Therefore, in addition to the previous values of “BLAST score density,” different “minimum length coverage” thresholds were chosen to cluster NRBLD sequences. Clustering results are available in Tables S5 and S6 and the thresholds in Figure 2.

Application of the similarity and coverage thresholds stipulated for clustering resulted in the separation of true BLs from other sequences such as partial domains and the regulatory proteins BlaR1/MecR1. For instance, in the case of class SA, most non-BL sequences have a larger average size (345–637 amino acids) than BLs in subclasses SA1 and SA2 (285 and 300 amino acids, respectively). One cluster contains a functionally characterized BL (LRA-5, non-BL9, Table S5). No non-BL sequence was observed for the class SC. All clusters containing non-BL sequences from class SD have one sequence, which is similar in size to BlaR1/MecR1 proteins (~585 amino acids) or partial domains (<214 amino acids). Only one of them (YP_612206, non-BL13, Table S5) is similar in size to class SD BL (274 amino acids) and its best hit in the NCBI nonredundant protein database¹⁷ shows 43% identity with a sequence annotated as “class D BL” from *Oceanicaulis alexandrii* (E value of 1E-67). All non-BL sequences from class MB are partial domains (<197 amino acids). Among the 2 clusters containing non-BL sequences formed from MB1, one possesses partial domain sequences and the other has a 340-amino acid sequence from *Stigmatella aurantiaca*, considerably larger than BL sequences (~250 amino acids). The ME2 subclasses were not maintained after clustering of NRBLD sequences, even with higher minimum coverage thresholds (90%), not corroborating what was observed when only sequences from PDB were clustered.

Comparison of the improved profiles HMM with Pfam profiles and BL motifs

Seven Pfam³⁶ profiles were tested against the curated data set of BLs from PDB, showing low specificity for BL classes. The 2 Pfam profiles for SBL (PF00144.22 and PF13354.4) identified 339 and 227 enzymes out of a total of 399 SBLs, with an intersection of 220 sequences (35% and 3% CpSp, respectively). Three profiles for MBL (PF00753.25, PF12706.5, and PF16661.3) identified 98, 35, and 12 enzymes out of a total of 105 MBLs available (64%, 0%, and 0% CpSp, respectively). The other 2 MBL profiles did not identify any enzyme (PF13483.4 and PF14597.4).

Different sources were used to select 13 motifs related to the various BL classes, which were used to search among the 509 PDB sequences. About 11 specific motifs for the third classification level (SA, SC, and SD classes) and 2 motifs specific only for the second level (MBL) were used (Tables 1 and 2). The efficiency of these motifs was tested to separate BL sequences into subclasses (fourth and fifth levels).

No motifs developed for MBL nor motifs for class SA are present in all the sequences allocated to their respective groups. The KxxS motif⁴⁷ was found in all class SC sequences, whereas the motif SxV⁶ is present in all sequences allocated in the class SD. None of the motifs analyzed was specific to BL subclasses (MB1 or MB2, SA1 or SA2, SD1 or SD2; Tables 1 and 2).

Identification and classification of BLs in fully assembled genomes

A total of 1476 BL sequences were identified in 2774 prokaryotic genomes. SA, SC, SD, MB, and ME profiles recovered 616, 280, 366, 103, and 111 sequences, respectively. No SCD class members were found in the genomes surveyed.

After the clustering and annotation process, 123 (8.3%) sequences were considered non-BLs (Table S7). The remaining 1352 sequences (91.7%) were distributed among 12 phyla and classified according to the BL subclasses to which they belong (Table 3). All subclasses were found in the *Proteobacteria* phylum, excepted for MB1.2. SD1 is the most disseminated subclass among the phyla analyzed, whereas SC, SD2, and MB2 were mostly restricted to *Proteobacteria*. A clear difference can be observed between the phyla where BL sequences of the subclasses SA1 (*Proteobacteria*, *Firmicutes*, and *Actinobacteria*) and SA2 (*Bacteroidetes*, *Cyanobacteria*, and *Spirochaetes*) were identified. It should be noted that 49% of all analyzed strains belong to the phyla *Proteobacteria*. Regarding MBL subclasses, 51% (46) of MB1.1 sequences are in *Firmicutes*, whereas 82% (91) of ME are in *Proteobacteria* (Table 3). A single sequence of the MB1.2 class was identified, isolated from *Sediminispirochaeta smaragdinae* (WP_013255389.1), a bacteria belonging to the *Spirochaetes* phyla, with a specific domain of SPM-1 (cd16286).

Table 1. Efficiency of serine- β -lactamase motifs.

MOTIF	TARGET	CLASS	SUBCLASS 1*	SUBCLASS 2**
ExxLN ^a	SA	174/218 (80%)	174/216 (81%)	0/2 (0%)
SDN ^a	SA	183/218 (84%)	181/216 (83%)	2/2 (100%)
KTG ^a	SA	169/218 (78%)	167/216 (77%)	2/2 (100%)
S-[DG]-N-x(1,2)-A-[ACGNST]-x(2)-[ILMV]-x(4)-[AGSTV] ^b	SA	107/218 (49%)	105/216 (48%)	2/2 (100%)
[FY]-x-[LIVMFY]-{E}-S-[TV]-x-K-x(3)-(T)-[AGLM]-{D}-{KA}-[LC] ^c	SA	192/218 (88%)	190/216 (87%)	2/2 (100%)
YxN ^a	SC	116/119 (97%)	—	—
KxxS ^d	SC	119/119 (100%)	—	—
[FY]-E-[LIVM]-G-S-[LIVMG]-[SA]-K ^c	SC	118/119 (99%)	—	—
SxV ^a	SD	60/60 (100%)	56/56 (100%)	4/4 (100%)
SxxxxS ^d	SD	50/60 (83%)	46/56 (82%)	4/4 (100%)
[PA]-x-S-[ST]-F-K-[LIV]-[PALV]-x-[STA]-[LI] ^c	SD	43/60 (72%)	41/56 (73%)	2/4 (50%)

^aBush.⁶^bSingh et al.⁴⁵^cPROSITE.⁴⁶^dMACiE.⁴⁷ Target: BL class for which the motif was developed; BL: Beta-lactamase; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD.

*Subclasses SA1 and SD1; **subclasses SA2 and SD2; — represents classes that have no subclass.

Table 2. Efficiency of metallo- β -lactamase (MBL) motifs.

MOTIF	TARGET	MBL	ME	MB	MB1	MB2
[LI]-x-[STN]-[HN]-x-H-[GSTAD]-D-x(2)-G-[GP]-x(7,8)-[GS] ^a	MBL	55/105 (52%)	12/19 (63%)	43/86 (51%)	36/79 (46%)	7/7 (100%)
P-x(3)-[LIVM](2)-x-G-x-C-[LIVMF](2)-K ^a	MBL	46/105 (44%)	0	46/86 (54%)	39/79 (50%)	7/7 (100%)

^aPROSITE.⁴⁶ Target: BL class for which the motif was developed; BL: Beta-lactamase; MBL: Metallo-BL; MB: Metallo-BL class MB; ME: Metallo-BL class ME; MB1: Metallo-BL subclass 1; MB2: Metallo-BL subclass MB2.

A higher absolute number of BL sequences were observed in *Proteobacteria*. The distribution of BLs was also analyzed at the taxonomic class level for this phylum. *Gamma*proteobacteria class has 60% (545) of the BL sequences, divided among all BL subclasses. SD1 is the only subclass found in all classes, and the unique BL found in *Epsilon*proteobacteria (Table 4).

A total of 100% of the genomic sequences retrieved by the profiles (after the exclusion of non-BL sequences) were allocated to some subclass of BL. About 70% of their original annotations were designated only as BL (first level), 24% had information on the second or third level of classification or the gene name, and there were still 6% with erroneous or imprecise annotations, such as “hypothetical protein” (Figure 4).

Discussion

Classification schemes for BLs are of utmost importance due to the diversity of these enzymes and their importance in the scenario of bacterial resistance to antibiotics.^{1,4,5,16} In general, the identification of new sequences is most often done by sequence comparison methods,⁴⁸ such as the BLASTP

program.³⁴ Profiles HMM and other profile-sequence comparison methods led to a significant improvement in sensitivity over sequence comparison approaches and are already used in the identification of antibiotic resistance proteins.^{3,49}

The workflow developed here systematizes the annotation of BLs based mainly on 2 steps: searches using profiles HMM, followed by clustering the resulting sequences. The calibrated profiles HMM can assign a sequence to a specific class. They also discriminate functionally characterized BLs from proteins with other biochemical functions that belong to the same superfamily and therefore share fold signals that make this separation difficult.³¹ In addition, the calibrated profiles allowed the recognition of sequences erroneously attributed to the BL GO term (“BL activity”) in SwissProt³³ and also enable the identification of sequences imprecisely described as BL in different antibiotic resistance databases. In the clustering step, the established thresholds of similarity and coverage allowed the clearing of non-BL sequences, providing coherent BL subclasses.

Table 3. β -lactamase sequences identified in bacterial genomes.

PHYLA	GENOMES	SA		SC	SD		MB			ME	TOTAL
		SA1	SA2		SD1	SD2	MB1.1	MB1.2	MB2		
<i>Proteobacteria</i>	1176	302	5	276	132	68	24	—	5	91	903
<i>Firmicutes</i>	583	148	—	—	30	—	46	—	—	3	227
<i>Actinobacteria</i>	283	107	—	4	4	—	—	—	—	1	116
<i>Bacteroidetes</i>	88	—	16	—	14	—	18	—	—	3	51
<i>Cyanobacteria</i>	73	—	2	—	17	—	—	—	—	—	19
<i>Spirochaetes</i>	60	—	1	—	3	—	2	1	—	7	14
Other	135	2	4	—	9	1	—	—	—	6	22
Total	2398	559	28	280	209	69	90	1	5	111	1352

Genomes: number of strains analyzed by phylum. — represents phyla where no sequences were found. Other: Verrucomicrobia, Acidobacteria, Fusobacteria, Gemmatimonadetes, Chlorobi, and Chlamydiae; BL: Beta-lactamase; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD; MB: Metallo-BL class MB (former subclasses B1 and B2); ME: Metallo-BL class ME (former subclass B3); MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2; MB1.1: Metallo-BL subclass MB1.1; MB1.2: Metallo-BL subclass MB1.2.

Table 4. β -lactamase sequences identified in bacterial genomes from *Proteobacteria* phylum.

CLASS	GENOMES	SA		SC	SD		MB		ME	TOTAL
		SA1	SA2		SD1	SD2	MB1.1	MB2		
<i>Gammaproteobacteria</i>	546	148	3	216	77	20	13	4	64	545
<i>Alphaproteobacteria</i>	321	75	2	29	15	15	2	—	22	160
<i>Betaproteobacteria</i>	146	76	—	31	13	29	—	1	4	154
<i>Epsilonproteobacteria</i>	104	—	—	—	21	—	—	—	—	21
<i>Deltaproteobacteria</i>	59	3	—	—	6	4	9	—	1	23
Total	1176	302	5	276	132	68	24	5	91	903

Genomes: number of strains analyzed by phylum. — represents phyla where no sequences were found BL: Beta-lactamase; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD; MB: Metallo-BL class MB (former subclasses B1 and B2); ME: Metallo-BL class ME (former subclass B3); MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2; MB1.1: Metallo-BL subclass MB1.1.

Among the sequences from BL databases used to construct NRBLD, some non-BL sequences presented partial domains, suggesting assembly/sequencing artefacts or annotation errors (eg, non-BL3, Table S5).³⁵ Others had much larger sizes than BLs, such as the BlaR1 protein, homologous to the BLs of the class SD, but with different functions. However, 2 sequences classified as non-BL did not cluster within any BL subclass despite being similar in size to them. The LRA-5 protein, described and functionally characterized as a class SA BL, has low similarity and is a distant relative to functionally characterized BLs and their ancestors.⁵⁰ As the experimental and HMM profile data indicate that LRA-5 is a BL of class SA, we speculate that it may belong to a third subclass (SA3) considering its low similarity to other sequences. Availability of new sequences in the future will help confirm the existence of this new. The second exception (YP_612206) is 43% identical to a sequence annotated as

“class D BL” of the dimorphic rods *O alexandrii*, although the activity of this protein has not been demonstrated.⁵¹

It has been shown that the use of Pfam³⁶ profiles to identify sequences from the “Ser-BL-like superfamily” may capture unrelated sequences.¹⁶ In our comparisons, the improved profiles HMM displayed higher specificity when compared with Pfam profiles and BL motifs from literature. Recently, individualized subgroups of the class SA have been demonstrated, such as LSBL or TEM/SHV and CARB clusters. Characteristic residues have already been attributed to each of them, but in this work we have chosen to test motifs attributed to BL classes.¹ However, subclass-specific motifs, such for SA1 and SA2, should also be tested and compared in future studies.

Some BL subclasses that were previously described based on phylogenetic criteria were identified here using sequence similarity criteria. The sequences in subclasses MB1 and MB2 correspond to the subclasses “B1” and “B2” in the work of

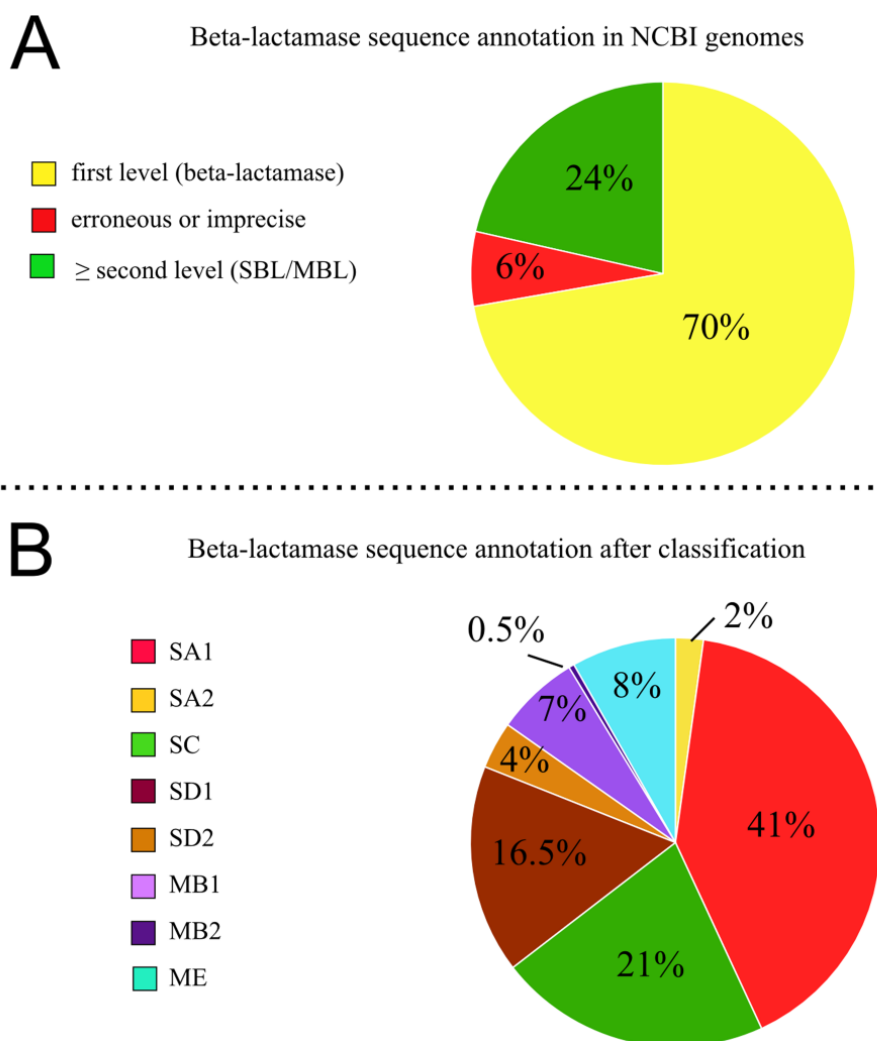


Figure 4. Original (A) annotation and (B) reannotation of the 1352 sequences according to the hierarchical levels of BL classification. BL indicates β -lactamase; MBL: metallo- β -lactamase; SBL: serine- β -lactamase; MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2; ME: Metallo-BL class ME.

Galleni et al.⁵² The sequences in the subclass SA2 correspond to BLs isolated mainly from the group Cytophagales-Flavobacteriales-Bacteroidales.^{1,16,53} The OXA alleles in subclass SD2 are the same as those found in the class called “D2” by Brandt et al.¹⁶

A new class of BLs with fused domains, the class SCD, is proposed. Two NRBLD proteins were captured by both SC and SD profiles without using the HMM bit score threshold. One of them, LRA-13, isolated from a noncultivated soil bacterium in Alaska, was confirmed experimentally as a BL, displaying a broad hydrolytic profile.⁵⁰

Subclass MB1.2 was first presented here, formed by members of the BL SPM family. It has been suggested that SPM-1 may be a structural hybrid between MB1 and MB2 subclasses.⁵⁴ SPM-1 genes were found only in isolates of *Pseudomonas aeruginosa*, a *Proteobacteria*, and its chromosomal location may have contributed to its isolation to other BL families of subclass MB1.⁵⁵ Curiously, we found that *S. smaragdinae*, a spirochaete, carries a protein with the SPM-1 domain

(WP_013255389.1). Further studies are needed to establish the evolutionary relationships between these proteins.

The distribution of BL sequences obtained from fully assembled genomes in different subclasses confirmed and complemented previous observations. For instance, the observed enrichment of BLs in *Actinobacteria* relative to other phyla³ is caused mainly by members of subclass SA1. The BLs of subclass SA2, related to the *Bacteroidetes* phylum,^{1,16} were also found in other phyla. The majority presence of class SC in the *Proteobacteria* phylum is a consensus, and the few sequences found in *Actinobacteria* confirm the occasional isolation of this class.¹⁶ The wide distribution of the subclass SD1 between the phyla analyzed confirms that class SD, more precisely the subclass SD1, has been underestimated.¹⁶ The recently described presence of this subclass in Gram-positive bacteria has also been confirmed here.⁵⁶ The association of BL sequences from subclass SD2 to *Proteobacteria* may be related to the fact that they are naturally occurring intrinsic genes, most likely chromosomally located.¹⁶ The occurrence of MB1 in *Bacteroidetes*

and *Firmicutes* phyla has been associated with chromosome, whereas the *Proteobacteria* MB1 enzymes are mostly mobile.⁵⁷ The association of class ME with soil bacteria may be related to its distribution among different phyla, which can share the same environment.³ New BLs of class ME have already been described in metagenomes, and the majority diverge deeply from other known enzymes of this class,⁵⁰ which reinforces the idea that class ME is widely distributed and diverse.

Acidobacteria are abundant mainly in the soil, but their cultivation is difficult.⁵⁸ Although few genomes of this phylum were available in 2016, the number of BLs identified was significant, suggesting that this is an important reservoir of BLs in nature.

The class *Gammaproteobacteria* includes common human pathogens such as *Enterobacteriaceae* and *Pseudomonadaceae*.¹⁶ These pathogens are exposed continuously to evolutionary pressure exerted by antibiotics, favoring the acquisition of resistance genes, which may be related to the presence of different BL subclasses in this group. *Epsilonproteobacteria*, which are widely distributed bacteria including genera with pathogenic species for humans such as *Helicobacter* and *Campylobacter*, presented BLs of only the SD1 subclass. Indeed, the production of 2 major BLs (OXA-61 and OXA-184) was detected in 85% of the *Campylobacter* strains.⁵⁹

Conclusions

The workflow developed in this study presents better specificity when compared with available BL motifs and Pfam profiles. Application of the improved profiles HMM and sequence similarity clustering parameters resulted in a 5-level hierarchical classification, consistent with previous BL classification scheme and recent proposed subclasses based on phylogeny. We also emphasize that the number of amino acids may be an important criterion for characterizing BLs, although there may be exceptions, such as the new class of fused domain enzymes (class SCD), proposed here. The workflow presented here, which can be further improved by the addition of functional and phylogenetic data, will be of great help in studies on the prevalence, distribution, and evolution of this critical enzymatic activity.

Acknowledgements

The authors would like to thank Dr Reema Singh and Dr Harpreet Singh for providing the DLact database sequences and Dr Alex Herbert for giving us the script to remove alternative positions of the atoms. MCS recognizes CAPES (Coordination for the Improvement of Higher Education Personnel, Brazil) for supporting her with a scholarship during her DSc program.

Author Contributions

MCS and ABM conceived of the manuscript outline; MCS, FFM, MC, RJ, ACRG, and ABM jointly developed the methodology; MCS, RAS, FFM, MC, and ABM wrote, edited and revised the manuscript.

REFERENCES

- Philippon A, Slama P, Dény P, Labia R. A structure-based classification of class A β -lactamases, a broadly diverse family of enzymes. *Clin Microbiol Rev.* 2016;29:29–57.
- Abraham EP, Chain E. An enzyme from bacteria able to destroy penicillin. *Nature.* 1940;146:837.
- Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 2014;9:207–216.
- Ambler RP. The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci.* 1980;289:321–331.
- Hall BG, Barlow M. Revised Ambler classification of β -lactamases. *J Antimicrob Chemother.* 2005;55:1050–1051.
- Bush K. The ABCD's of β -lactamase nomenclature. *J Infect Chemother.* 2013;19:549–559.
- Bush K, Jacoby GA, Medeiros AA. A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob Agents Chemother.* 1995;39:1211–1233.
- Ouzounis CA, Coulson RMR, Enright AJ, Kunin V, Pereira-Leal JB. Classification schemes for protein structure and function. *Nat Rev Genet.* 2003;4:508–519.
- Jaurin B, Grundström T. AmpC cephalosporinase of *Escherichia coli* K-12 has a different evolutionary origin from that of beta-lactamases of the penicillinase type. *Proc Natl Acad Sci U S A.* 1981;78:4897–4901.
- Ouellette M, Bissonnette L, Roy PH. Precise insertion of antibiotic resistance determinants into Tn21-like transposons: nucleotide sequence of the OXA-1 beta-lactamase gene. *Proc Natl Acad Sci U S A.* 1987;84:7378–7382.
- Rasmussen BA, Bush K. Carbapenem-hydrolyzing beta-lactamases. *Antimicrob Agents Chemother.* 1997;41:223–232.
- Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJE. Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol.* 2007;372:817–845.
- Frère JM, Galleni M, Bush K, Dideberg O. Is it necessary to change the classification of β -lactamases? *J Antimicrob Chemother.* 2005;55:1051–1053.
- Hall BG, Salipante SJ, Barlow M. The metallo- β -lactamases fall into two distinct phylogenetic groups. *J Mol Evol.* 2003;57:249–254.
- Alderson RG, Barker D, Mitchell JBO. One origin for metallo- β -lactamase activity, or two? An investigation assessing a diverse set of reconstructed ancestral sequences based on a sample of phylogenetic trees. *J Mol Evol.* 2014;79:117–129.
- Brandt C, Braun SD, Stein C, et al. In silico serine β -lactamases analysis reveals a huge potential resistome in environmental and pathogenic species. *Sci Rep.* 2017;7:43232. doi:10.1038/srep43232.
- NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2015;43:D6–D17.
- Burley SK, Berman HM, Christie C, et al. RCSB Protein Data Bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* 2017;27:316–330.
- Srivastava A, Singhal N, Goel M, Viridi JS, Kumar M. CBMAR: a comprehensive β -lactamase molecular annotation resource. *Database.* 2014;2014:bau111.
- Bialek-Davenet S, Criscuolo A, Ailloud F, et al. Genomic definition of hyper-virulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. *Emerg Infect Dis.* 2014;20:1812–1820.
- Singh R, Singh H. DLact: an antimicrobial resistance gene database. *J Comput Intell Bioinform.* 2008;1:93–108.
- Thai QK, Bös F, Pleiss J. The lactamase engineering database: a critical survey of TEM sequences in public databases. *BMC Genomics.* 2009;10:390.
- Widmann M, Pleiss J, Oelschlaeger P. Systematic analysis of metallo- β -lactamases using an automated database. *Antimicrob Agents Chemother.* 2012;56:3481–3491.
- McArthur AG, Waglechner N, Nizam F, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother.* 2013;57:3348–3357.
- Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;26:680–682.
- MaxCluster: a tool for protein structure comparison and clustering. <http://www.sbg.bio.ic.ac.uk/maxcluster/index.html>. Accessed March, 2016.
- Wei D, Jiang Q, Wei Y, Wang S. A novel hierarchical clustering algorithm for gene sequences. *BMC Bioinformatics.* 2012;13:174.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–1797.
- Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7:e1002195.
- Sillitoe I, Lewis TE, Cuff A, et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* 2015;43:D376–D381.
- Sinha S, Lynn AM. HMM-ModE: implementation, benchmarking and validation with HMMER3. *BMC Res Notes.* 2014;7:483.

32. Kumar S, Stecher G, Peterson D, Tamura K. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*. 2012;28:2685–2686.
33. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–D169.
34. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402.
35. Triant DA, Pearson WR. Most partial domains in proteins are alignment and annotation artifacts. *Genome Biol*. 2015;16:99.
36. Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44:D279–D285.
37. Pratap S, Katiki M, Gill P, Kumar P, Golemi-Kotra D. Active-site plasticity is essential to carbapenem hydrolysis by OXA-58 class D β -lactamase of *Acinetobacter baumannii*. *Antimicrob Agents Chemother*. 2016;60:75–86.
38. Mukherjee S, Stamatis D, Bertsch J, et al. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res*. 2017;45:D446–D456.
39. Urbach C, Evrard C, Pudzaitis V, Fastrez J, Soumillion P, Declercq JP. Structure of PBP-A from *Thermosynechococcus elongatus*, a penicillin-binding protein closely related to class A beta-lactamases. *J Mol Biol*. 2009;386:109–120.
40. Boudreau MA, Fishovitz J, Llarrull LI, Xiao Q, Mobashery S. Phosphorylation of BlaR1 in manifestation of antibiotic resistance in methicillin-resistant *Staphylococcus aureus* and its abrogation by small molecules. *ACS Infect Dis*. 2016;1:454–459.
41. Wilke MS, Hills TL, Zhang HZ, Chambers HF, Strymadka NCJ. Crystal structures of the Apo and penicillin-acylated forms of the BlaR1 beta-lactamase sensor of *Staphylococcus aureus*. *J Biol Chem*. 2004;279:47278–47287.
42. Dubois D, Baron O, Coughoux A, et al. ClbP is a prototype of a peptidase sub-group involved in biosynthesis of nonribosomal peptides. *J Biol Chem*. 2011;286:35562–35570.
43. Proenca R, Niu WW, Cacalano G, Prince A. The *Pseudomonas cepacia* 249 chromosomal penicillinase is a member of the AmpC family of chromosomal beta-lactamases. *Antimicrob Agents Chemother*. 1993;37:667–674.
44. Lüthy L, Grütter MG, Mittl PRE. The crystal structure of *Helicobacter pylori* cysteine-rich protein B reveals a novel fold for a penicillin-binding protein. *J Biol Chem*. 2002;277:10187–10193.
45. Singh R, Saxena A, Singh H. Identification of group specific motifs in beta-lactamase family of proteins. *J Biomed Sci*. 2009;16:109.
46. Sigrist CJA, De Castro E, Cerutti L, et al. New and continuing developments at PROSITE. *Nucleic Acids Res*. 2013;41:344–347.
47. Holliday GL, Andreini C, Fischer JD, et al. MACiE: exploring the diversity of biochemical reactions. *Nucleic Acids Res*. 2012;40:783–789.
48. Feuermann M, Gaudet P, Mi H, Lewis SE, Thomas PD. Large-scale inference of gene function through phylogenetic annotation of gene ontology terms: case study of the apoptosis and autophagy cellular processes. *Database*. 2016;2016:baw155.
49. Fróes AM, da Mota FF, Cuadrat RRC, D'Avila AMR. Distribution and classification of serine β -lactamases in Brazilian hospital sewage and other environmental metagenomes deposited in public databases. *Front Microbiol*. 2016;7:1790.
50. Allen HK, Moe LA, Rodbumer J, Gaarder A, Handelsman J. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J*. 2009;3:243–251.
51. Oh HM, Kang I, Vergin KL, Lee K, Giovannoni SJ, Cho JC. Genome sequence of *Oceanicaulis* sp. strain HTCC2633, isolated from the western Sargasso Sea. *J Bacteriol*. 2011;193:317–318.
52. Galleni M, Lamotte-brasseur J, Rossolini GM. Standard numbering scheme for class B beta-lactamases. *Society*. 2001;45:660–663.
53. Hall BG, Barlow M. Evolution of the serine beta-lactamases: past, present and future. *Drug Resist Updat*. 2004;7:111–123.
54. Bebrone C. Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem Pharmacol*. 2007;74:1686–1701.
55. Silveira MC, Albano RM, Asensi MD, Carvalho-Assef AP. Description of genomic islands associated to the multidrug-resistant *Pseudomonas aeruginosa* clone ST277. *Infect Genet Evol*. 2016;42:60–65.
56. Toth M, Antunes NT, Stewart NK, et al. Class D β -lactamases do exist in Gram-positive bacteria. *Nat Chem Biol*. 2016;12:9–14.
57. Berglund F, Marathe NP, Österlund T, et al. Identification of 76 novel B1 metallo- β -lactamases through large-scale screening of genomic and metagenomic data. *Microbiome*. 2017;5:134.
58. Kielak AM, Barreto CC, Kowalchuk GA, Van Veen JA, Karumac EE. The ecology of Acidobacteria: moving beyond genes and genomes. *Front Microbiol*. 2016;7:744.
59. Weis AM, Storey DB, Taff CC, et al. Genomic comparison of *Campylobacter* spp. and their potential for zoonotic transmission between birds, primates, and livestock. *Appl Environ Microbiol*. 2016;82:7165–7175.