Predatory myxobacteria lay at the cross-roads of metallo-β-lactamase evolution

Anastasia Lambropoulou[a], Anastasia Pavlidi[a], Martina Samiotaki[b], Mohamed Tammam[c], Efstathia Ioannou[c], Vassilios Roussis[c], Leonidas S. Tzouvelekis[a], Vivi Miriagou[a], Stathis D. Kotsakis[a]

a: Laboratory of Bacteriology, Hellenic Pasteur Institute, Athens, Greece
b: Mass-Spectrometry Core Facility, B.S.R.C., "Alexander Fleming", Athens, Greece
c: Section of Pharmacognosy and Chemistry of Natural Products, Department of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece

Supplemental material

1. Validation of beta-lactamase specific Hidden Markov Models

HMMs for class A, C, D, (serine reactive beta-lactamases) B1, B2 and B3 (metallo-beta-lactamases) were built using the reference sequences of the so far experimentally verified beta-lactamases contained in NCBI's Bacterial Antimicrobial Resistance Reference Gene Database (BioProject: PRJNA313047).

Protein sequences of the PRJNA313047 bioproject were downloaded from NCBI using eutils on 25-3-2019. Sequences belonging to each of the beta-lactamase class were extracted. The LRA-13 bifunctional beta-lactamase, containing distinct class C and D domains, was excluded (protein id: WP_063839877.). The VarG MBL-fold protein identified in a *V. cholera* isolate and suggested to have beta-lactamase activity (protein id: WP_000778180.1; 6) was also excluded as it doesn't contain the zinc binding motifs of the B1, B2 and B3 subclasses of metallo-beta-lactamases (8) and therefore it represents a separate protein family within the MBL superfamily.

Sequences of each beta-lactamase class were clustered using cd-hit with an identity threshold of 80% (-c 0.8) and a length cut-off of 100% (-s 1.0). The above yielded 6 data-sets (Table 1) that were used for training the respective HMMs.

Supplementary Table 1: Data sets used in model training

beta-lactamases	# sequences in PRJNA 313047	CD-HIT clusters [-c 0.8 –s 1.0]
class A	1002	124
class C	769	50
class D	724	93
class B1	245	50
class B2	15	3
class B3	68	38

CD-HIT extracted representative sequences for each cluster (except class B2 where the un-clustered data set was used due to the limited number of sequences) were aligned using clustal omega with default settings. The above alignments were used for the HMM training using the program hmmbuild of the HMMER v3.1b2 package. Cut-off scores for considering a positive HMMER hit against the above models as a likely beta-

lactamase were estimated using a one-out-cross-validation approach that was performed as follows. Each protein from each data-set was excluded and after alignment of the remaining sequences a HMM was build. Then the excluded protein was checked against the generated model using the program hmmsearch of HHMER with default inclusion criteria. The above process was repeated for each protein in the data set and the lowest observed HMMER score was considered as the cut-off.

Proteins subjected to one of the models generated with the whole data set using hmmsearch with default settings and having HMMER scores above the cut-off are consider positive for this specific model.

Supplementary Table 2: Estimation of cut-offs for each model

HMM	lowest HMMER bits score	Protein/cluster w lowest score
class_A_model	182	WP_004339683.1/CfxA family
class_C_model	415	WP_011339581.1/RHO family
class_D_model	188	WP_000722315.1/OXA-9
class_B1_model	157	WP_013255389.1/SPS-1
class_B2_model	316	WP_071766619.1/SFH-1
class_B3_model	180	WP_011094382.1/CAR-1

a. Serine reactive beta-lactamases

The class A, C and D models and the estimated cut-offs were evaluated regarding their ability to detect the respective proteins without yielding positive results against the other classes of serine reactive beta-lactamases and their homologous Penicillin Binding Proteins (PBPs). For these purpose each model was tested against data sets of functionally characterized class A, C and D beta-lactamases (BioProject: PRJNA313047) as well as Penicillin Binding Proteins as described by Massova and Mobashery (7). Moreover each model was searched against the respective UniProt families that contain manually (Swiss-Prot pre 1996) and computationally (trEMBL/unreviewed) annotated protein entries. The searches were performed using hhmsearch with default threshold criteria (E-value$\leq$10.0). The number of HMMER hits was estimated as well as the number of hits with HMMER scores above the respective cut-offs.

i) class A model

The class A model, by applying a cut-off of 182, was able to detect all the class A beta-lactamases without yielding positive results for the other serine reactive beta-lactamases and PBPs in the functionally characterized data sets (Table 3). Regarding the un-reviewed UniProt data sets the model yielded positive results for 89.38% of the entries in UniProt class-a beta-lactamases family. Around 10% of the above family, although they exhibited E-values within threshold criteria, have scores below the cut-off. Entries with scores between 100 and 182 in their majority concern fragments of class A beta-lactamases while at scores below 100 sequences without one or more of the typical class A motifs were identified and therefore they have been

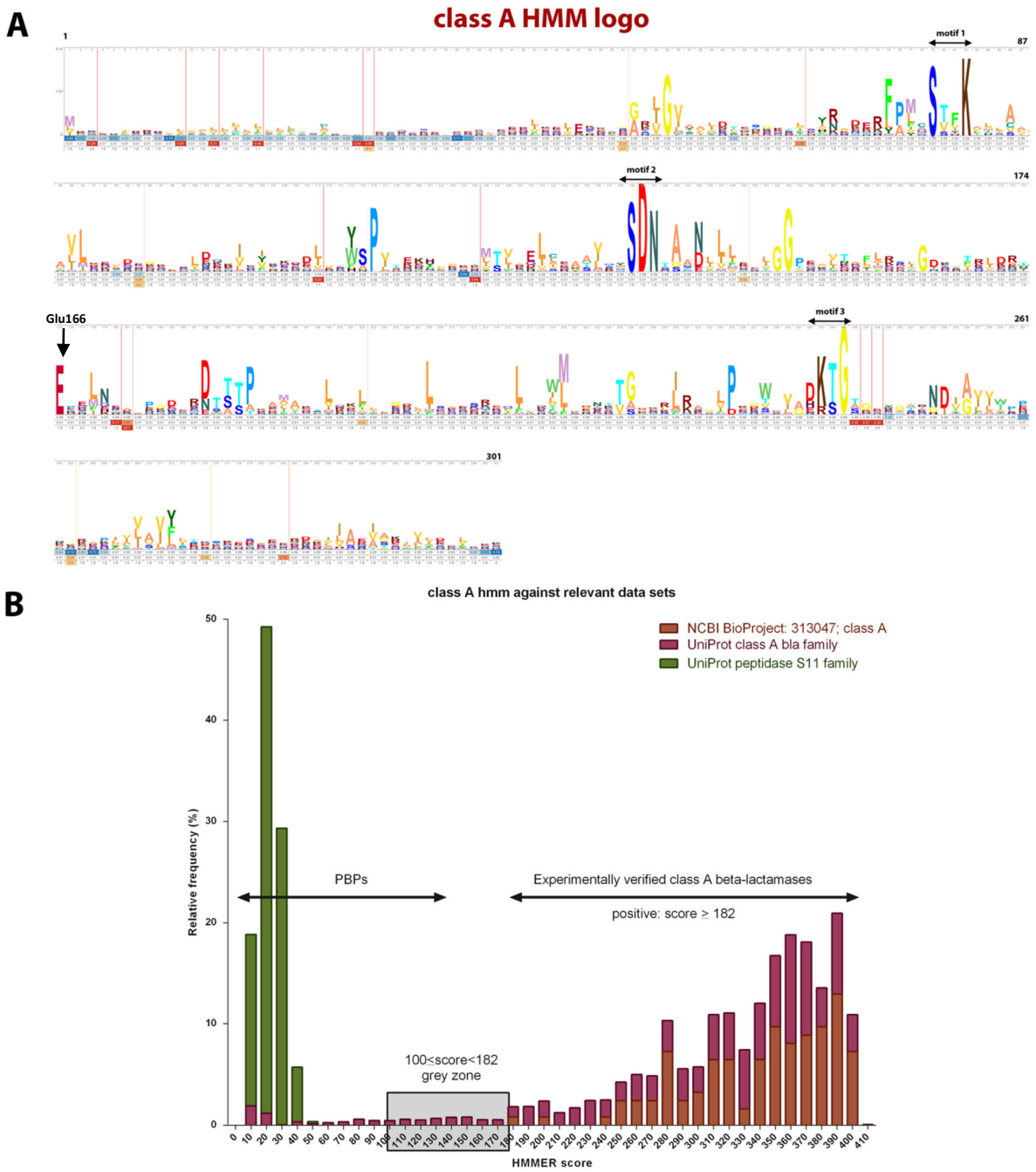
misclassified by trEMBL. Nonetheless a number of entries with scores within 100-182 range appear to be full length and having all the three class A catalytic motifs and therefore may represent yet uncharacterized beta-lactamases distantly related to those used for building the model giving ambiguous alignments (Example 1; Glu166). Hence establishment of a "grey-zone" is imperative in order to be able to detect both class A fragments (e.g. partially captured in WGS contigs) as well as highly divergent class A beta-lactamases (Figure 1). Every entry in the other UniProt families has been found negative regarding the validated model. The highest percentage of HMMER hits in non-class A families were identified in the peptidase S11 data set (Table 3). Although the majority of scores were below 100 (Figure 1), there were 22 sequences with scores within the above established grey-zone that may represent divergent class A (Example 2).

Example 1: Alignment of a Uniprot class-A family entry with score within the grey zone

```
>> tr|A0A0S2KGK2|A0A0S2KGK2_9GAMM  Beta-lactamase OS=Pseudohongiella spirulinae OX=1249552 GN=PS2015_2837 PE=3 SV=1
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  132.4   0.0   1.3e-38   1.3e-38      19     299 ..      10     292 ..       1     293 [] 0.80

  Alignments for each domain:
  == domain 1  score: 132.4 bits;  conditional E-value: 1.3e-38
  cdhit_clustered_class_A_aligned  19 lllaacaaa.aaaaasalaakleeklkel.ekksggrlGvaaldtes.gktlsyraderFpmcStfKallaaavLkev 93
                                      +llaac  a a+a  ++l+a+l e+l++   +  + ++Gv+al+ +s g+tl +     ++ +St+K+ll+a ++++v
   tr|A0A0S2KGK2|A0A0S2KGK2_9GAMM  10 VLLAACWIApASADFNSLQAELRETLQQAgDQGIRVSVGVSALEGGSaGATLLMGDALSYAPASTIKMLLVASLMQQV 87
                                      233333333233445566667777777662456 78********9987256667777789**********9********** PP

  cdhit_clustered_class_A_aligned  94 degklsldqkikikkddlvtysPvtekhvakt.mtvaeLceaavqySDNtAanlLleelgGpeavtaflrslgdkvtr 170
                                      d+g l+l+q+ +++++d+v     v ++ ++ + +t++ L+e +v+ SDNtA+n+L++ + G  a+  ++ +lg + ++
   tr|A0A0S2KGK2|A0A0S2KGK2_9GAMM  88 DAGLLTLQQTTTVSPADVVGGYGVLQEEAQPQdVTLGRLAELTVTISDNTATNVLVDVV-GYPAMAGLAAQLGLELMH 164
                                      *****************7766666666666 77*************9876.5667999*******9888 PP

  cdhit_clustered_class_A_aligned 171 ldrtEpeLneaipgderDtttPkamaktLqklllgdlLsaesreqLlqwmkdnttGdkriraglPkgwkvadKTGt.g 247
                                        r    +++ +p  + ++  + ++ ++L +++ g lLs++sr+q+++w++++t+  ++i ag+P+g+kva+KTG  g
   tr|A0A0S2KGK2|A0A0S2KGK2_9GAMM 165 FGRKM--PEAPEPPAKDNYINAPDALALLTAIYSGTLLSDDSRDQIMRWLSAQTV-RTKIAAGVPDGVKVAHKTGEnG 239
                                      88864..55667888888888889999*********************98755.799*************55 PP

  cdhit_clustered_class_A_aligned 248 eygarndiavvwppngrapivvavyltk...skesaee..rdaiiaeiakivldalk 299
                                          ++di++++p+  + ++va+++++   s+ +a++   + ++aei+++++ +l+
   tr|A0A0S2KGK2|A0A0S2KGK2_9GAMM 240 --PVSHDIGFIMLPD--STLAVAIFTENlkgSDFDASQalLNPLVAEITSTIYRHLG 292
                                      ..689********99..47*******9843333333334478899999999999875 PP
```

Example 2: Alignment of a Uniprot S11 peptidase family entry with score within the grey zone

```
>> tr|A0A2E7NKA6|A0A2E7NKA6_9PLAN  Uncharacterized protein OS=Planctomycetaceae bacterium OX=2026779 GN=CMJ68_04115 PE=3
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  138.2   0.0   5.3e-40   7.7e-40      12     298 ..      44     331 ..      33     333 .. 0.86
   2 !   23.1   0.0   6.2e-05     9e-05      57     144 ..     447     535 ..     445     544 .. 0.84

  Alignments for each domain:
  == domain 1  score: 138.2 bits;  conditional E-value: 5.3e-40
  cdhit_clustered_class_A_aligned  12 aallllllllaacaaaaaaaasalaakleeklkelekksggrlGvaaldtesgktlsyraderFpmcStfKallaaav 89
                                        +l+++++l+  c+       +a  a+l++++k l ++++g++ +++ +++sg+++s++ d  +p +S +K+ + + v
   tr|A0A0S2KGK2|A0A0S2KGK2_9PLAN  44 LLLVCAVMLA--CNGG---LWAADDARLAARIKPLVSRHQGQVAIVVRHLDSGREMSWQPDRPMPTASLIKLAVMVEV 116
                                      4455555444..4432...222334578*****************************************9********** PP

  cdhit_clustered_class_A_aligned  90 LkevdegklsldqkikikkddlvtysPvtekhvakt..mtvaeLceaavqySDNtAanlLleelgGpeavtaflrslg 165
                                        ++vd+++ +ld+++k ++dd v+ s v +k+ +++  +t+++ ++ ++ +SDNtA+nl+l+++ G e+ ++ + +lg
   tr|A0A0S2KGK2|A0A0S2KGK2_9PLAN 117 YRQVDDDRVDLDDRLKLTADDTVPGSGVLRKYFTTGsrLTLRDAVRLMIAVSDNTATNLVLDHI-GLESTNRTMSRLG 193
                                      *********************************889**********************9.69********** PP

  cdhit_clustered_class_A_aligned 166 dkvtrld....rtEpeLne.aipgderDtttPkamaktLqklllgdlLsaesreqLlqwmkdnttGdkriraglPkgw 238
                                        +tr+     r E+++     ++    +tt++++a++L+++  g + s++  +q+l+ +++++   +ri   lP+g+
   tr|A0A0S2KGK2|A0A0S2KGK2_9PLAN 194 HPNTRIHakvfRHETSIDPeRSQQFGLGSTTASETADLLEAIRSGRAVSKRASRQMLEHLRACEST-ERIPRYLPEGT 270
                                      ****998444467888887345556689*************************************965.9********** PP

  cdhit_clustered_class_A_aligned 239 kvadKTGtgeygarndiavvwppngrapivvavyltkskesaee....rdaiiaeiakivldal 298
                                      +va+KTG+ +   r d +++ +p+ + pivv v+++++++   +    ++++ia+i++ ++ ++
   tr|A0A0S2KGK2|A0A0S2KGK2_9PLAN 271 VVAHKTGSVS-RVRTDAGIIESPT-G-PIVVCVLTSENQDRRYSaenaASRLIADISREAWLHF 331
                                      ********84.7999*******88.6.9*******99776543333358899******999776 PP
```



Supplementary Figure 1: A) Sequence logo of the class A model generated with SkyAlign. The three conserved motifs of class A beta-lactamases as well as the general base for catalysis Glu166 are denoted. **B)** Frequency distribution of the HMMER scores observed in the relevant data sets. The grey zone is denoted by a grey box.

Supplementary Table 3: Class A model validation

Data set	Description	# Sequences	class A model				
			# HMMER hits[4]	% HMMER hits	max score	# score ≥ [cut-off=182]	% validated model
UNIPROT_blasA Family:"class-a beta-lactamase family"[1]	88: manually annotated (Swiss-Prot) 14388: Unreviewed (TrEMBL)	14476	14442	99.77	409.3	12938	89.38
UNIPROT_blasC Family: "class-c beta-lactamase family"	13: manually annotated (Swiss-Prot) 5358: Unreviewed (TrEMBL)	5371	20	0.37	18.3	0	0.00
UNIPROT_blasD Family: "class-d beta-lactamase family"	17: manually annotated (Swiss-Prot) 4080: Unreviewed (TrEMBL)	4097	286	6.98	55.9	0	0.00
UNIPROT_peptidase_M56 Family: "peptidase M56 family"	7: manually annotated (Swiss-Prot)	7	1	14.29	11.4	0	0.00
UNIPROT_peptidase_S11 Family: "peptidase S11 family"	16: manually annotated (Swiss-Prot) 51779: Unreviewed (TrEMBL)	51795	35706	68.94	138.2	0	0.00
UNIPROT_peptidase_S12 Family: "peptidase S12 family"	36: manually annotated (Swiss-Prot) 642: Unreviewed (TrEMBL)	678	29	4.28	14.8	0	0.00
UNIPROT_peptidase_S13 Family: "peptidase S13 family"	7: manually annotated (Swiss-Prot)	7	2	28.57	23.6	0	0.00
UNIPROT_transpeptidase Family: "transpeptidase family"	81: manually annotated (Swiss-Prot) 20682: Unreviewed (TrEMBL)	20763	83	0.40	18.4	0	0.00
Functionally characterized							
class_A	NCBI BioProject: 313047[2]; class A	1002	1002	100.00	405.1	1002	100.00
class_C	NCBI BioProject: 313047; class C[3]	769	0	0.00	NA	0	0.00
class_D	NCBI BioProject: 313047; class D[3]	724	30	4.14	21.3	0	0.00
HMWAPBP (transpeptidases)	High Molecular Weight class A PBPs	19	0	0.00	NA	0	0.00
HMWBPBP (transpeptidases)	High Molecular Weight class B PBPs	19	0	0.00	NA	0	0.00
HMWCPBP (M56, BlaR1)	High Molecular Weight class C PBPs	3	1	33.33	10.1	0	0.00
LMWAPBP (S11 peptidases)	Low Molecular Weight class A PBPs	13	10	76.92	35.6	0	0.00
LMWBPBP (S12 peptidases)	Low Molecular Weight class B PBPs	5	2	40.00	14.8	0	0.00
LMWCPBP (S13 peptidases)	Low Molecular Weight class C PBPs	5	2	40.00	17.5	0	0.00

1: search term used for sequence retrieval
2: NCBI BioProject containing sequences of functionally characterized antimicrobial resistance genes
3: The class-C/class-D multidomain LRA-13 beta-lactamase was excluded
4: Hits that justified HHMER default inclusion criteria E-value ≤ 10.0
NA: Not applicable

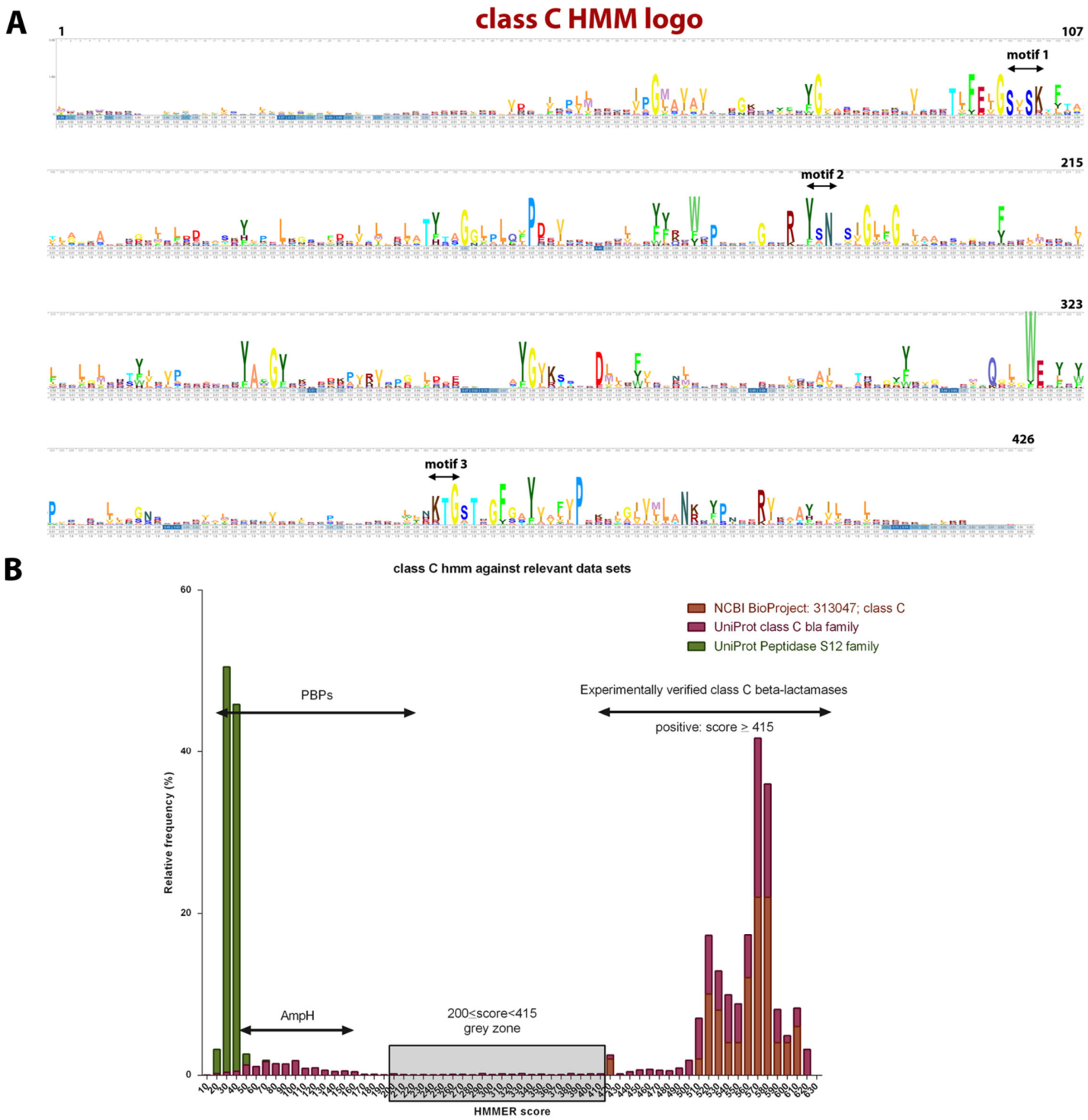
ii) class C model

The validated model (score ≥ 415) yielded positive results for the whole functionally characterized class C data set and negative for the rest serine reactive beta-lactamases and PBPs (Table 4). Regarding the Uniprot data sets the validated model identified 83.54% of the sequences as positive in the class C beta-lactamase families (Table 4). As it was the case for the class A model, sequences with scores below the cut-off corresponded to class C fragments as well as few whole proteins with the typical class C motifs. This was the case regarding scores as low as 200 (grey zone 200-415). Below that point PBPs were found and most importantly the AmpH peptidase and its homologues that exhibited scores centered around 100 (Figure 2). These proteins are highly homologous to class C beta-lactamases but as it has been experimentally demonstrated they cannot hydrolyze the beta-lactam ring (4). Therefore similarly as in class A beta-lactamase UniProt family, a number of entries in class C Uniprot family have been misclassified by the automatic annotation of EMBL.

In the class D Uniprot dat set 6 proteins were found positive with the validated model. These entries correspond to the bifunctional LRA-13 and its homologues that contain both class C and D beta-lactamase domains (and are included in both class C and class D Uniprot families). The class C model was correctly aligned with the class C domain (Example 3) yielding a score above the cut-off. The remaining UniProt data sets gave negative results with the maximum observed scores being below the grey-zone with peptidase S12 exhibiting the highest percentage of HMMER hits (Table 4).

Example 3: Alignment of the class C model with the respective domain of LRA-13.

```
>> tr|B5L5V6|B5L5V6_9BACT  Beta-lactamase OS=uncultured bacterium BLR13 OX=506515 GN=blaLRA-13 PE=3 SV=1
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 ?    0.8   0.1       0.3        32     330     352 ..     188     210 ..     171     237 .. 0.79
   2 !  519.9   0.9  6.1e-159  6.5e-157      32     386 ..     253     607 ..     226     608 .. 0.98

  Alignments for each domain:
  == domain 1  score: 0.8 bits;  conditional E-value: 0.3
  cdhit_clustered_class_C_aligned 330 qaeekakllnKtGstnGFgaYva 352
                                      q++   ++  KtG+ +G+g Yv
          tr|B5L5V6|B5L5V6_9BACT 188 QPPAGWRINGKTGAASGYGWYVG 210
                                      4556677888***********96 PP

  == domain 2  score: 519.9 bits;  conditional E-value: 6.1e-159
  cdhit_clustered_class_C_aligned  32 elkalvdaaikpllkeqdipGmavavivdgkkhyfnyGvasketkkpvteetlFElGSvSKtftatlagyakakgkls 109
                                       +k++vd+a++pl+k++dipGmavav+ +gk+++fnyG+as+et+++vt++tlFE+GS+SKt++atl++ya+++g+l
          tr|B5L5V6|B5L5V6_9BACT 253 MIKDMVDRAVQPLMKKYDIPGMAVAVTDNGKNYFFNYGLASRETGQAVTSHTLFEIGSLSKTMAATLTSYAQVNGQLA 330
                                      6899**********************************************************************9 PP

  cdhit_clustered_class_C_aligned 110 lddkaskylpeLkgsafdkitllelatytaGglpLqvPdevkdkaqllayyrqwqpkaapgtqRlYsNosiGLlGala 187
                                      l d++s+++p+L+g  fdki+ll+l+t+taG +p+qvPd++++ +ql++yy++w+p  a+g  R+YsN ++GLlG ++
          tr|B5L5V6|B5L5V6_9BACT 331 LTDTVSRHMPKLRGGGFDKISLLNLGTHTAGDFPMQVPDHIETYEQLMEYYKNWKPGVAAGGARTYSNLTVGLLGIIT 408
                                      ***********************************************************************9********** PP

  cdhit_clustered_class_C_aligned 188 akslgkpfeqllektllpaLglkhtyikvpeaamanYAfGyskedkpvrvspgvldaeaYGvkssaaDllrfveanle 265
                                      a+s+g+pf++++e++l+p+Lg++h+yi+vp+a+m+nYA+Gy++++ pvr++p vl+ eaYGvk+ aaDl+rfv+an+
          tr|B5L5V6|B5L5V6_9BACT 409 AQSMGMPFAEAMENRLFPQLGMHHSYINVPAAEMKNYAQGYNQANAPVRINPAVLATEAYGVKTDAADLIRFVDANMG 486
                                      ********************************************************************************** PP

  cdhit_clustered_class_C_aligned 266 peklekalqqaiaathkgyykvgemtqgLgWEsysypvsletllagnsekvalkanpvtaieapqaeekakllnKtGs 343
                                        kl+++lq+a++ th+ y+k+ge+tq+L+WE+y+   +l+++lag sek+++++np+t++++p ++++++l+nKtGs
          tr|B5L5V6|B5L5V6_9BACT 487 LVKLDEKLQRAVTGTHTAYFKTGELTQDLIWEQYPAASKLDRMLAGVSEKMVFESNPATRLAPPMPPQADVLINKTGS 564
                                      ********************************************************************************** PP

  cdhit_clustered_class_C_aligned 344 tnGFgaYvafvPekkigivmLaNknypneeRvkaayaileale 386
                                      t+GFgaY+ f P kk+givmLaNk+yp +eRv+aa++il++l+
          tr|B5L5V6|B5L5V6_9BACT 565 TGGFGAYALFNPGKKTGIVMLANKSYPGAERVTAAWHILDQLD 607
                                      ****************************************97 PP
```



Supplementary Figure 2: A) Class C model sequence logo with the three motifs participating in proton relay during catalysis being denoted. **B)** Frequency distributions of the HMMER bit scores during searches against the relevant data sets.

Supplementary Table 4: Class C model validation

Data set	Description	# Sequences	class C model				
			# HMMER hits[4]	% HMMER hits	max score	# score ≥ [cut-off=415]	% validated model
UNIPROT_blasA Family:"class-a beta-lactamase family"[1]	88: manually annotated (Swiss-Prot) 14388: Unreviewed (TrEMBL)	14476	469	3.24	170.10	0	0.00
UNIPROT_blasC Family: "class-c beta-lactamase family"	13: manually annotated (Swiss-Prot) 5358: Unreviewed (TrEMBL)	5371	5371	100.00	620.20	4487	83.54
UNIPROT_blasD Family: "class-d beta-lactamase family"	17: manually annotated (Swiss-Prot) 4080: Unreviewed (TrEMBL)	4097	39	0.95	519.9; 19.0[6]	6[5]	0.15
UNIPROT_peptidase_M56 Family: "peptidase M56 family"	7: manually annotated (Swiss-Prot)	7	0	0.00	NA	0	0.00
UNIPROT_peptidase_S11 Family: "peptidase S11 family"	16: manually annotated (Swiss-Prot) 51779: Unreviewed (TrEMBL)	51795	30	0.06	10.90	0	0.00
UNIPROT_peptidase_S12 Family: "peptidase S12 family"	36: manually annotated (Swiss-Prot) 642: Unreviewed (TrEMBL)	678	678	100.00	69.10	0	0.00
UNIPROT_peptidase_S13 Family: "peptidase S13 family"	7: manually annotated (Swiss-Prot)	7	0	0.00	NA	0	0.00
UNIPROT_transpeptidase Family: "transpeptidase family"	81: manually annotated (Swiss-Prot) 20682: Unreviewed (TrEMBL)	20763	0	0.00	NA	0	0.00
Functionally characterized							
class_A	NCBI BioProject: 313047[2]; class A	1002	0	0.00	NA	0	0.00
class_C	NCBI BioProject: 313047; class C[3]	769	769	100.00	620.30	769	100.00
class_D	NCBI BioProject: 313047; class D[3]	724	2	0.28	14.90	0	0.00
MOB_HMWAPBP	High Molecular Weight class A PBPs	19	0	0.00	NA	0	0.00
MOB_HMWBPBP	High Molecular Weight class B PBPs	19	0	0.00	NA	0	0.00
MOB_HMWCPBP	High Molecular Weight class C PBPs	3	0	0.00	NA	0	0.00
MOB_LMWAPBP_S11	Low Molecular Weight class A PBPs	13	0	0.00	NA	0	0.00
MOB_LMWBPBP_S12	Low Molecular Weight class B PBPs	5	5	100.00	69.10	0	0.00
MOB_LMWCPBP_S13	Low Molecular Weight class C PBPs	5	0	0.00	NA	0	0.00

1: search term used for sequence retrieval
2: NCBI BioProject containing sequences of functionally characterized antimicrobial resistance genes
3: The class-C/class-D multidomain LRA-13 beta-lactamase was excluded
4: Hits that justified HHMER inclusion criteria E-value ≤ 10.0
5: The 6 false positive hits correspond to LRA-13 class C/class D multidomain protein and its homologues. The model aligned with the class C domain
6: The first score is the max score for the whole data set and the second the max score when the LRA-13 homologues were excluded
NA: Not applicable

iii) class D model

The validated class D model identified all the experimentally confirmed class D beta-lactamases while it gave false positive results for the total of the high molecular weight class C PBPs corresponding to the BlaR1/MecR1 beta-lactam sensing membrane proteins found in some Gram positives (Table 5). The scores of these false positive results were well above the cut-off and the model aligned with the beta-lactam sensing domain. The latter is actually a class D beta-lactamase exhibiting very low de-acylation rates of the beta-lactam acyl-enzymes due to the replacement of the hydrophobic amino-acid at the third position of the catalytic motif SX(V/I/L) (Figure 3) with a hydrophilic one (usually Thr; Example 4) that most probably prohibits N-carboxylation of the side chain of lysine 67 (that is the catalytic base; 5, 9). Therefore a class D positive result should be reanalyzed with a BlaR1 specific model developed as above.

Regarding the Uniprot data sets the validated model gave positive results for 67.59% of entries in the class D beta-lactamase family with some of the sequences exhibiting scores between 100 and 188 possessing the typical class D motifs and this was established as the grey zone for the model (Figure 3). One protein in the class A Uniprot data set had score above the cut-off and the aligned model actually shows this to be a misclassified entry as it contains the typical class D motifs (Example 5). The 6 positive results observed during searches in Uniprot class C family correspond to LRA-13 and its homologues, with the model aligning to the class D domain (Example 6). The other families (except the M56 peptidase that corresponds to BlaR1/MecR1 family) gave negative results with the maximal scores below the grey zone of the model (Table 5).

Example 4: Alignment of the class D model with the beta-lactam sensing domain of BlaR1

```
>> sp|P12287|BLAR_BACLI  Regulatory protein BlaR1 OS=Bacillus licheniformis OX=1402 GN=blaR1 PE=1 SV=1
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  233.8   0.1     3e-73     3e-73      27     267 ..     355     595 ..     320     597 .. 0.91

  Alignments for each domain:
  == domain 1  score: 233.8 bits;  conditional E-value: 3e-73
  cdhit_clustered_class_D_aligned  27 eeeqekkelkklfkeaqtsgvivvadekkeklyendlkrakerfsPaSTFKilnaLigldsgvvkdekevfkwdgkkr 104
                                      ++ e ++ +++f++ ++sg +v+++++++k++ ++ k++  rf+PaST+K++ aL++l+sg+++++++ ++wdg+++
             sp|P12287|BLAR_BACLI 355 GTNVEYEDYSTFFDKFSASGGFVLFNSNRKKYTIYNRKESTSRFAPASTYKVFSALLALESGIITKNDSHMTWDGTQY 432
                                      4556778999*****************************************************************9 PP

  cdhit_clustered_class_D_aligned 105 klkawekdltlasamkaSvVpvyqelarriGeermqkyvkkldYGnadisgkldefWldgsLkIsaeeqvkflrkLak 182
                                       +k+w++d++l sam  S+ +++q+l r+iGe+ +++y+k+++YGn+d+s    ++WldgsL+Is+ eqv++l+k++
             sp|P12287|BLAR_BACLI 433 PYKEWNQDQDLFSAMSSTTWYFQKLDRQIGEDHLRHYLKSIHYGNEDFS-VPADYWLDGSLQISPLEQVNILKKFYD 509
                                      ********************A*****************************5.55669********************* PP

  cdhit_clustered_class_D_aligned 183 nkLpfseeaqeivkeilllleakedyklyaKtGvg.vdvepqiGWfvGwvekdgkvvvFalnidikkeadlkaReeilk 259
                                      n++ f+++++e+vk+ + le++++  l +KtG++ ++ e + GWf+G+ve+ +++ +Fa++i+ +k a  ++  ei+
             sp|P12287|BLAR_BACLI 510 NEFDFKQSNIETVKDSIRLEESNGRVLSGKTGTSvINGELHAGWFIGYVETADNTFFFAVHIQGEKRAAGSSAAEIAL 587
                                      *********************************55599*****************************************9 PP

  cdhit_clustered_class_D_aligned 260 kllkklkl 267
                                      ++l+k ++
             sp|P12287|BLAR_BACLI 588 SILDKKGI 595
                                      99999876 PP
```

Example 5: Alignment of the class D model with an entry in the Uniprot class A data set. The protein contains the class D typical motifs.

```
>> tr|A0A2E3QWK9|A0A2E3QWK9_9BACT  Beta-lactamase OS=Rhodothermaceae bacterium OX=2026787 GN=CMM85_10460 PE=3 SV=1
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  203.5   0.0   1.7e-61   2.5e-60      31     267 ..      55     296 ..      16     298 .. 0.94

  Alignments for each domain:
  == domain 1  score: 203.5 bits;  conditional E-value: 1.7e-61
  cdhit_clustered_class_D_aligned  31 ekkelkklfkeaqtsgvivvadekkeklyendlkrakerfsPaSTFKilnaLigldsgvvkdekevfkwdgkkrklka 108
                                      + ++++++f+++++ g++v+ d ++ ++  +d +ra+erfs aST K+n+L+ ld gv++d ++ f wdg +r  +
   tr|A0A2E3QWK9|A0A2E3QWK9_9BACT  55 SVADWSAAFEAEGAVGTMVLLDTQTGRTVRHDPARAAERFSLASTSKIYNSLVFLDRGVISDVDSLFAWDGVERWAEV 132
                                      5679**********************************************************************  PP

  cdhit_clustered_class_D_aligned 109 wekdltlasamkaSvVpvyqelarriGeermqkyvkkldYGnadisgkldefWldgsLkIsaeeqvkflrkLaknkLp 186
                                      w++d++l+s + +S+V+ +q+ a ++G+  + +  ++  +Gn+++s  l+ +Wldg+ ++sa eqv+fl++L +  L+
   tr|A0A2E3QWK9|A0A2E3QWK9_9BACT 133 WNRDHSLRSGLEVSAVWLFQRAALQVGRGGYDDVFAREPFGNSTMSDALEMSWLDGTWRVSADEQVAFLDRLRRGALA 210
                                      **************************************************************************  PP

  cdhit_clustered_class_D_aligned 187 fseeaqeivkeilll.eakedyklyaKtGwgvd.vepqiGWfvGwvekdgkvvvFalnidikk...eadlkaReeilk 259
                                      fs+e q +v++il +  +++++++KtGw v+  +p++GW+vGwve+ +  +vFa+n ++ +   ++ ++ R +i++
   tr|A0A2E3QWK9|A0A2E3QWK9_9BACT 211 FSAEDQATVRDILPVlAEAGEVRVKGKTGWYVRePDPELGWLVGWVERPDGDLVFAMNAEQASgasFDIMRGRLRIVR 288
                                      **************99999**********88799**************************99777777888999999 PP

  cdhit_clustered_class_D_aligned 260 kllkklkl 267
                                      ++l+ ++l
   tr|A0A2E3QWK9|A0A2E3QWK9_9BACT 289 AILEGEGL 296
                                      99987665 PP
```

Example 6: Alignment of the class D model with the respective domain of LRA-13

```
  >> tr|B5L5V6|B5L5V6_9BACT  Beta-lactamase OS=uncultured bacterium BLR13 OX=506515 GN=blaLRA-13 PE=3 SV=1
     #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
   ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
     1 !  200.5   0.0   8.7e-63   7.8e-60      42     262 ..      21     250 ..      14     255 .. 0.93

    Alignments for each domain:
    == domain 1  score: 200.5 bits;  conditional E-value: 8.7e-63
    cdhit_clustered_class_D_aligned  42 aqtsgvivvadekkeklyendlkrakerfsPaSTFKilnaLigldsgvvkdekevfkwdgkk...rklkawekdltla 116
                                        ++t+ +i +a++ +  +  +   +++++++PaSTFKi+ +L+g+dsg++kd + + k  +++       ++w++++++a
             tr|B5L5V6|B5L5V6_9BACT  21 HATEVCIAIAEAGTGAVLVQ-RGDCQRQVTPASTFKIAISLMGYDSGFLKDAHAP-KLPFRPgyvDWRPSWREPTDPA 96
                                        67888999999997777766.88*********************************.99998887644569******** PP

    cdhit_clustered_class_D_aligned 117 samkaSvVpvyqelarriGeermqkyvkkldYGnadisg.....kldefWldgsLkIsaeeqvkflrkLaknkLpfse 189
                                        ++m +SvV+++q++++ +G++r+++y ++++YGnad+sg     +l  +W+++sL+Is+ eq +fl+k+++++L +s+
             tr|B5L5V6|B5L5V6_9BACT  97 KWMSLSVVWYSQQVTKSLGMQRFADYTRNFKYGNADVSGdaendGLSMSWISSSLRISPLEQLAFLDKIVNRRLGVSA 174
                                        ***********************************99999999****************************** PP

    cdhit_clustered_class_D_aligned 190 eaqeivkeilll.eakedyklyaKtGwgvdvepqiGWfvGwvekdgkvvvFalnidikk......eadlkaReeilkk 260
                                        +a++++ +++ + +   ++++ +KtG ++     +GW+vGw++k +++ vFa+ ++ +       +a++ aR+e+lk+
             tr|B5L5V6|B5L5V6_9BACT 175 HAYDMTAQLTKFdQPPAGWRINGKTGAAS----GYGWYVGWASKGSRTFVFAHLMQRDAtqpqdvSAGVLARDEFLKE 248
                                        ************77789***********9....9*********************99988888999999999999998 PP

    cdhit_clustered_class_D_aligned 261 ll 262
                                        l+
             tr|B5L5V6|B5L5V6_9BACT 249 LP 250
                                        86 PP
```

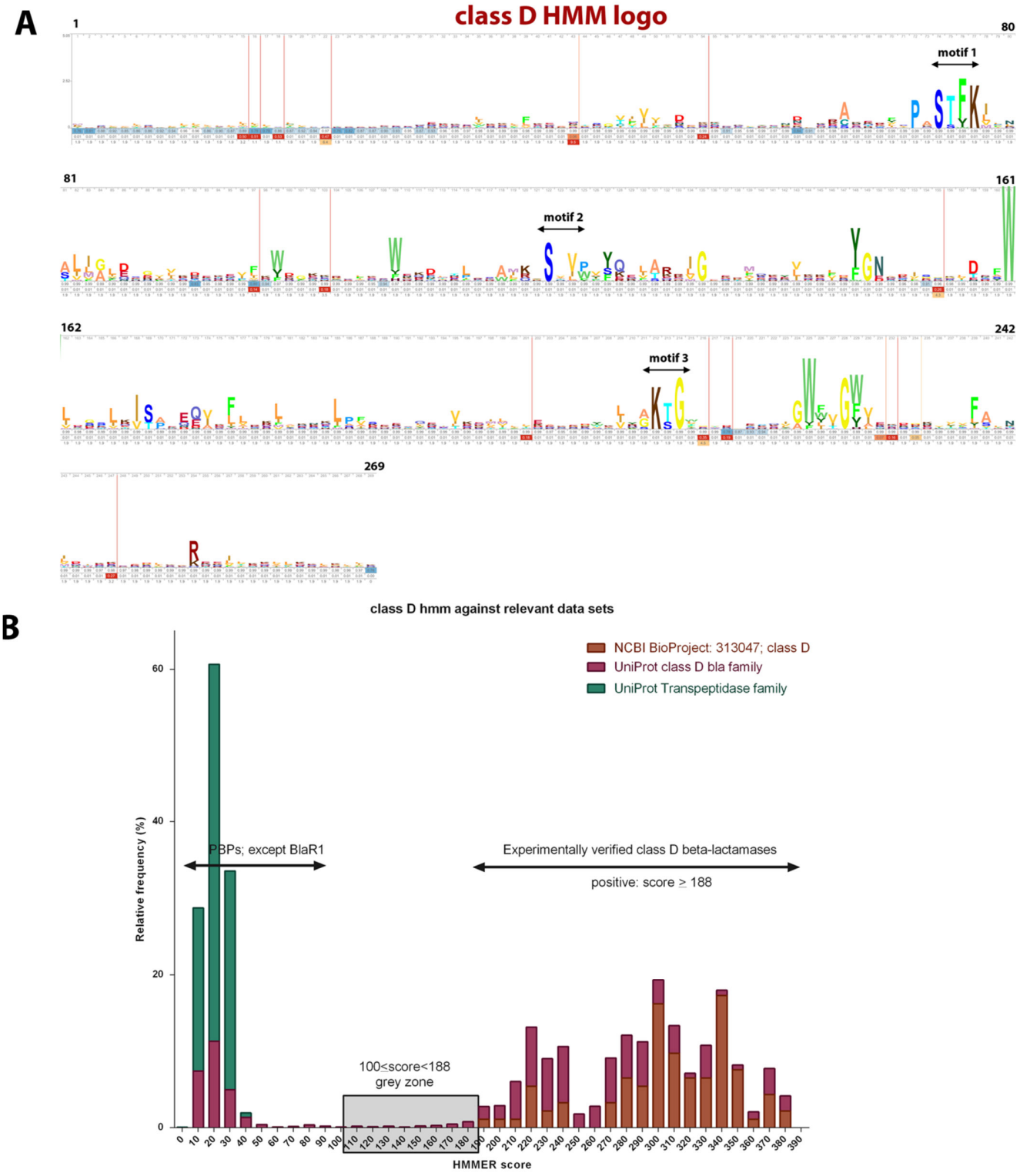


Figure 3: A) Class D model sequence logo with the three motifs participating in proton relay during catalysis being denoted. **B)** Frequency distributions of the HMMER bit scores during searches against the relevant data sets.

Supplementary Table 5: Class D model validation

Data set	Description	# Sequences	class D model				
			# HMMER hits[4]	% HMMER hits	max score	# score ≥ [cut-off=188]	% validated model hits
UNIPROT_blasA Family:"class-a beta-lactamase family"[1]	88: manually annotated (Swiss-Prot) 14388: Unreviewed (TrEMBL)	14476	957	6.61	203.5; 28.0[7]	1[5]	0.01
UNIPROT_blasC Family: "class-c beta-lactamase family"	13: manually annotated (Swiss-Prot) 5358: Unreviewed (TrEMBL)	5371	6	0.11	208.9; no hits[7]	6[6]	0.11
UNIPROT_blasD Family: "class-d beta-lactamase family"	17: manually annotated (Swiss-Prot) 4080: Unreviewed (TrEMBL)	4097	3891	94.97	381.80	2769	67.59
UNIPROT_peptidase_M56 Family: "peptidase M56 family"	7: manually annotated (Swiss-Prot)	7	7	100.00	234.50	7[8]	100.00
UNIPROT_peptidase_S11 Family: "peptidase S11 family"	16: manually annotated (Swiss-Prot) 51779: Unreviewed (TrEMBL)	51795	3573	6.90	27.70	0	0.00
UNIPROT_peptidase_S12 Family: "peptidase S12 family"	36: manually annotated (Swiss-Prot) 642: Unreviewed (TrEMBL)	678	0	0.00	NA	0	0.00
UNIPROT_peptidase_S13 Family: "peptidase S13 family"	7: manually annotated (Swiss-Prot)	7	0	0.00	NA	0	0.00
UNIPROT_transpeptidase Family: "transpeptidase family"	81: manually annotated (Swiss-Prot) 20682: Unreviewed (TrEMBL)	20763	7890	38.00	43.40	0	0.00
Functionally characterized							
class_A	NCBI BioProject: 313047[2]; class A	1002	66	6.59	21.40	0	0.00
class_C	NCBI BioProject: 313047; class C[3]	769	0	0.00	NA	0	0.00
class_D	NCBI BioProject: 313047; class D[3]	724	724	100.00	381.80	724	100.00
MOB_HMWAPBP	High Molecular Weight class A PBPs	19	1	5.26	6.80	0	0.00
MOB_HMWBPBP	High Molecular Weight class B PBPs	19	8	42.11	36.70	0	0.00
MOB_HMWCPBP	High Molecular Weight class C PBPs	3	3	100.00	233.80	3[8]	100.00
MOB_LMWAPBP_S11	Low Molecular Weight class A PBPs	13	0	0.00	NA	0	0.00
MOB_LMWBPBP_S12	Low Molecular Weight class B PBPs	5	0	0.00	NA	0	0.00
MOB_LMWCPBP_S13	Low Molecular Weight class C PBPs	5	1	20.00	7.00	0	0.00

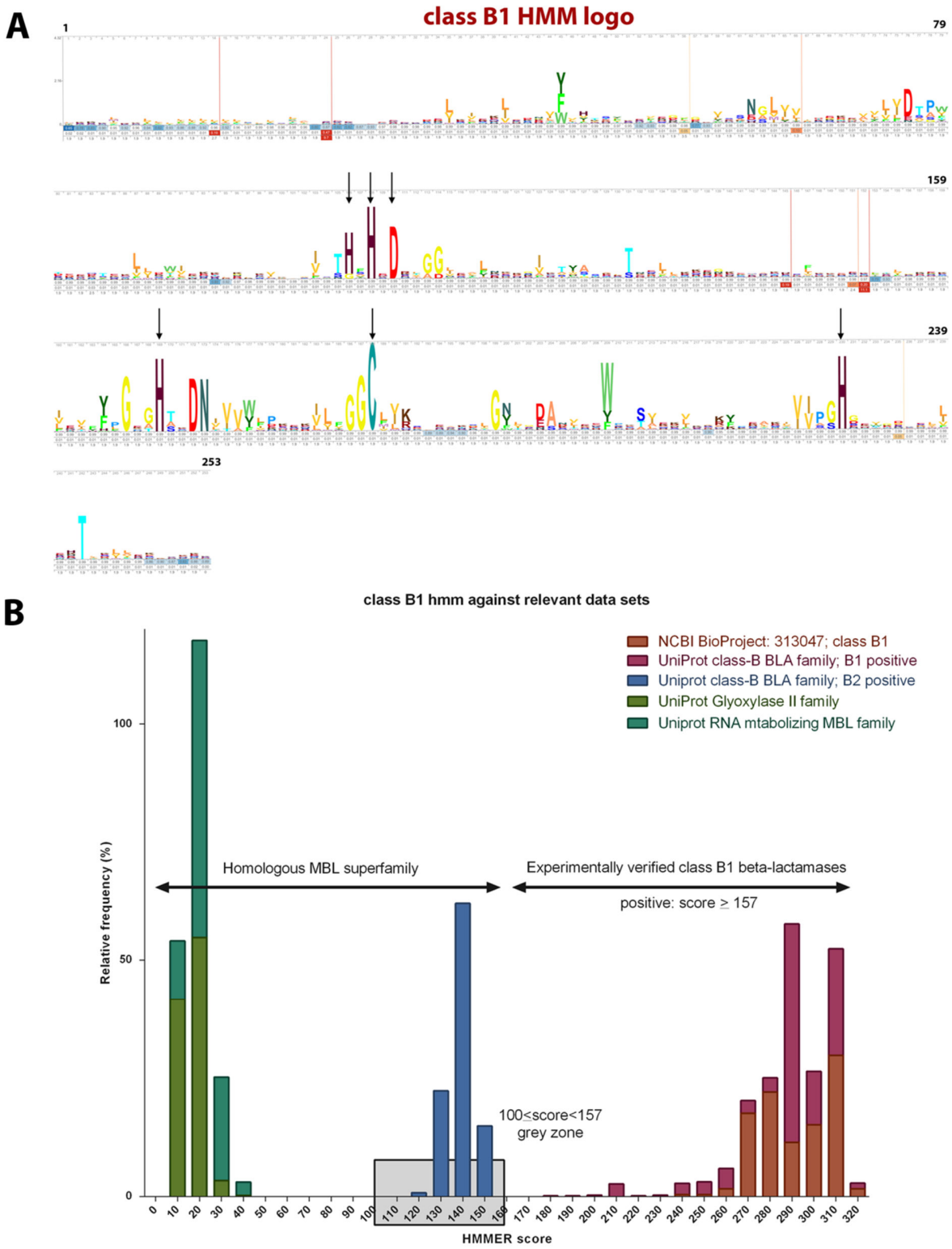
1: search term used for sequence retrieval
2: NCBI BioProject containing sequences of functionally characterized antimicrobial resistance genes
3: The class-C/class-D multidomain LRA-13 beta-lactamase was excluded
4: Hits that justified HHMER inclusion criteria E-value ≤10.0
5: Class D beta-lactamase erroneously added to class A beta-lactamase data set in trEMBL data base. See alignment
6: The 6 false positive hits correspond to LRA-13 class C/class D multidomain protein and its homologues. The model aligned with the class D domain
7: The first number is the max score for the whole data set and the second the max score with the misclassified sequence excluded
8: The model aligned with the beta-lactam sensing domain of BlaR1/MecR1 PBPs that is highly homologous with class D beta-lactamases. See discussion.
NA: Not applicable

b. Metallo-beta-lactamases

Metallo-beta-lactamases are members of a metalloenzyme superfamily sharing the same fold and exhibiting a degree of conservation in their metal binding motifs (MBL-fold superfamily; 1, 8). The accepted categorization of these proteins based on function and structure is the one proposed by Daiyasu et al. (3). B1, B2 and B3 subclasses of metallo-beta-lactamases found in Bacteria are forming the Group 1 family of MBLs while in the other 16 groups prokaryotic and eukaryotic enzymes performing a variety of functions are included (Table 6). The models of B1, B2 and B3 enzymes were subjected to hmmer searches against the data sets of the respective reference NCBI sequences as well as against the families of MBLs found in Uniprot and Interpro data bases. The Uniprot class-B beta-lactamase family data set contains entries of all the three metallo-beta-lactamase subclasses without being denoted the subclass they belong in their descriptions. Sequences found positive with each of the validated model were extracted yielding three separate data sets that were again subjected to each model in order to check for misclassification.

i) class B1 model

The validated (score $\geq$ 157) model classified correctly the reference metallo-beta-lactamase sequences (Table 6). Class B2 metallo-beta-lactamases exhibited scores that are near the cut-off (max sxore=145.5) as the two subclasses are phylogenetically related. In fact the SPS-1 enzyme that has been classified by others as B1 and yielded the lowest score for the B1 model is likely an intermediate evolutionary step between B1 and B2 beta-lactamases (2). Although there was no data pointing to the need of a grey zone establishment for this model we arbitrarily define one as in the serine reactive beta-lactamases in order to be able to detect B1 fragments and distantly related enzymes (grey zone 100 to 157). Therefore class B2 enzymes would fall inside the grey zone of B1 model (Figure 4). In order the annotation to be accurate the B1 model should be always run in parallel with that for B2 subclass. Regarding the Uniprot class-b-data set, every protein found positive for B1 (673) has been found negative using the B2 and B3 models and vice versa every protein found positive for the other two subclasses was negative for the B1 model. The model didn't yield a positive result for any other group of the MBL superfamily with the HMMER scores being below 100. For proteins exhibiting scores at the lower limit of the grey zone the diagnostic criterion for characterizing it as a likely B1 MBL is the presence of the amino-acids forming the two zinc binding sites and especially Cys221. If at this position an Asp or Glu is found then the enzyme catalyzes a different reaction and it is not a beta-lactamase (8).



Supplementary Figure 4: A) B1 model sequence logo with the conserved amino-acids participating in the binding of the two zinc atoms being indicated by arrows. **B)** Frequency distributions of the HMMER bit scores during searches against the relevant data sets.

Supplementary Table 6: Class B1 model validation

data set	Daiyasu et al classification	Description	# Sequences	class B1 model				
				# HMMER hits[3]	% HMMER hits	max score	# score ≥ [cut-off=157]	% validated model hits
class_B1	metallo beta-lactamase II	NCBI BioProject: 3130472[2]; class B1	245	245	100.00	316.30	245	100.00
class_B2	metallo beta-lactamase II	NCBI BioProject: 3130472; class B2	15	15	100.00	145.50	0	0.00
class_B3	metallo beta-lactamase II	NCBI BioProject: 3130472; class B3	68	38	55.88	25.00	0	0.00
Group_00_MBL	Pyrococcus furiosus	UP: P81414	1	0	0.00	NA	0	0.00
Group_01_MBL	metallo beta-lactamase II	UP: class-b beta-lactamase family[1]	795	795	100.00	315.50	673	84.65
UNIPROT_blasB class B1 positives	metallo beta-lactamase II	Found positive with validated HMM	673	673	100.00	315.50	673	100.00
UNIPROT_blasB class B2 positives	metallo beta-lactamase II	Found positive with validated HMM	121	121	100.00	146.70	0	0.00
UNIPROT_blasB class B3 positives	metallo beta-lactamase II	Found positive with validated HMM	1	1	100.00	15.30	0	0.00
Group_02_MBL	glyoxylase II	UP: glyoxylase II family	11621	10023	86.25	47.10	0	0.00
Group_03_MBL	flavoprotein	UP: zinc metallo-hydrolase group 3	9334	1921	20.58	36.00	0	0.00
Group_04_MBL	aryltransferase	UP: atsa family	7	0	0.00	NA	0	0.00
Group_05_MBL	cyclase	UP: Q9X466	1	1	100.00	82.40	0	0.00
Group_06_MBL	Cleavage and polyadenylation spec. factor	UP: RNA metabolizing MBL-like family	19812	17097	86.30	40.90	0	0.00
Group_07_MBL	DNA cross-link repair	UP: DNA repair MBL drmbl family	19	2	10.53	11.00	0	0.00
Group_08_MBL	comE	IP: ComA-like, MBL domain. IPR035681	22083	5702	25.82	40.60	0	0.00
Group_09_MBL	Choline binding protein E	UP: Q9KGZ1	1	0	0.00	NA	0	0.00
Group_10_MBL	PhnP protein	IP: Phosphonate metabolism protein, MBL domain. IPR035682	1524	7	0.46	14.20	0	0.00
Group_11_MBL	CMP-N-acetylneuraminate monoxygenase	UP: CMP-neu5ac hydroxylase family	12	0	0.00	NA	0	0.00
Group_12_MBL	Group 12: romA	UP: RomA family MBL fold metallo-hysrolase	208	0	0.00	NA	0	0.00
Group_13_MBL	alkyl sulfatase	UP: A0A1H1U5G4	1	1	100.00	22.60	0	0.00
Group_14_MBL	carbofuran hydrolase	UP: Q9KJV0	1	1	100.00	18.90	0	0.00
Group_15_MBL	methyl parathion hydrolase	UP: Q9ALW1	1	1	100.00	10.50	0	0.00
Group_16_MBL	3',5'-cyclic nucleotide phosphodiesterase	UP: cAMP phosphodiesterase class-ii family	735	6	0.82	13.90	0	0.00

1: search term used for sequence retrieval
2: NCBI BioProject containing sequences of functionally characterized antimicrobial resistance genes
3: Hits that justified HHMER inclusion criteria E-value ≤10.0
UP: UniProt; IP: InterPro; NA: not applicable

ii) class B2

The B2 hmm was able to annotate correctly the metallo-beta-lactamases of the reference data set using a cut-off score of 316 (Table 7). In the Uniprot class-b beta-lactamase family every protein found positive with this model (121 sequences) was found negative with those of B1 and B3 and those found positive with the latter models being negative when subjected to the B2 model. Therefore the validated model can detect with accuracy B2 MBLs. Every other family of MBL fold enzymes was found negative with scores below 100. The grey zone was established between 100 and 316 (Figure 5). Again for a protein with score near the lowest limit of the grey zone the presence of the zinc binding residues should be verified

ii) class B3

The 68 B3 metallo-beta-lactamase sequences of the reference data set were correctly annotated using a cut-off of 180 while the model gave negative results for the other subclasses (Table 8). Apparently the Uniprot class-b beta-lactamase family contains a solely B3 enzyme (L1 of *Stenotrophomonas maltophilia*) that was correctly annotated by the three MBL models. The rest MBL families were negative with scores below 100 (Table 8, Figure 6; grey zone: 100-180). The model gave positive scores for HARLDQ motif class B3 MβLs which do not possess meaningful hydrolytic activity and hence hits for this family should be reanalyzed with the specific model for this enzyme subset.

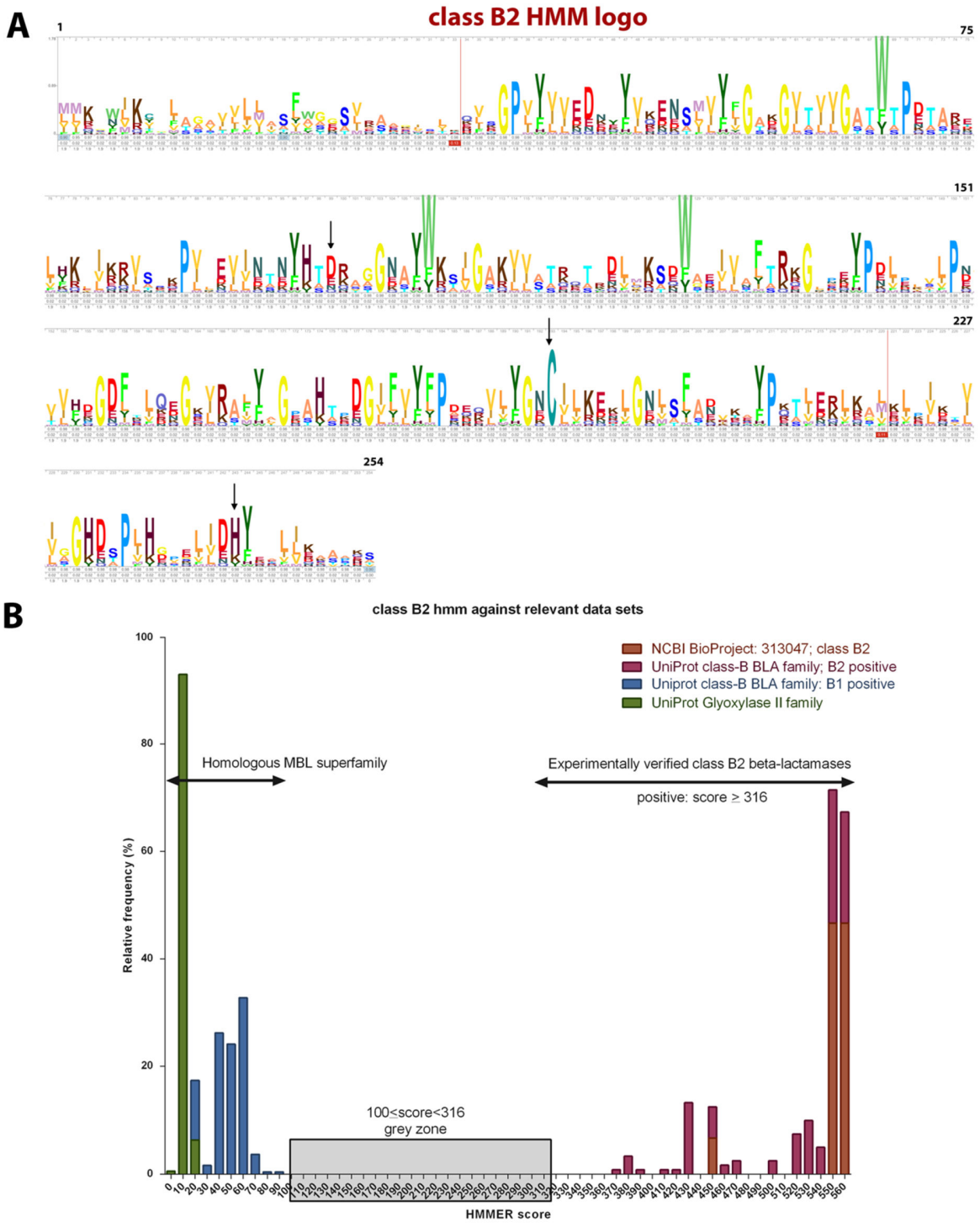
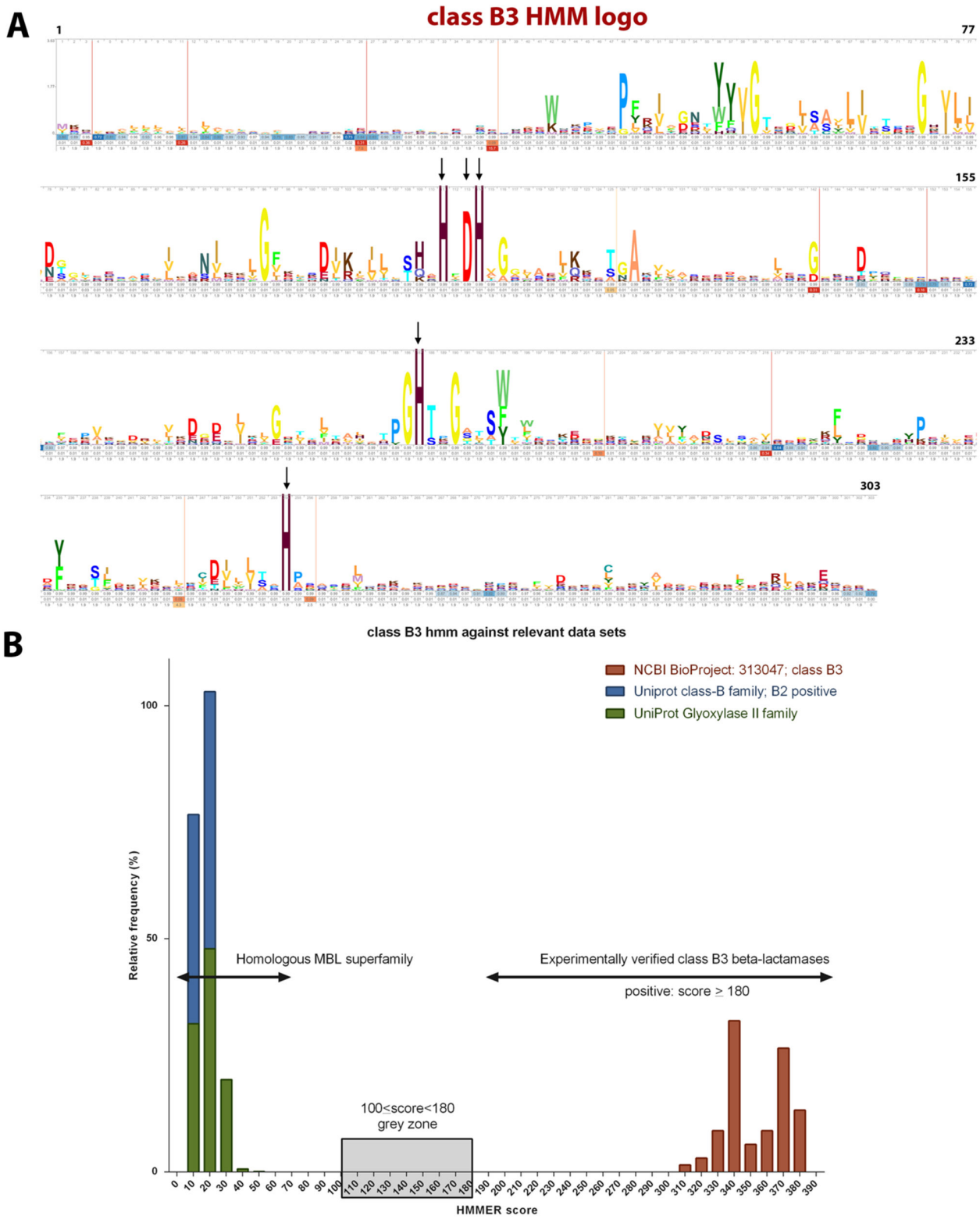


Figure 5: A) B2 model sequence logo with the conserved amino-acids participating in the binding of the zinc atom being indicated by arrows. **B)** Frequency distributions of the HMMER bit scores during searches against the relevant data sets.

Supplementary Table 7: Class B2 model validation

data set	Daiyasu et al classification	Description	# Sequences	class B2 model				
				# HMMER hits	% HMMER hits	max score	# score ≥ [cut-off=316]	% validated model hits
class_B1	metallo beta-lactamase II	NCBI BioProject: 3130472; class B1	245	245	100.00	85.40	0	0.00
class_B2	metallo beta-lactamase II	NCBI BioProject: 3130472; class B2	15	15	100.00	559.30	15	100.00
class_B3	metallo beta-lactamase II	NCBI BioProject: 3130472; class B3	68	1	1.47	13.70	0	0.00
Group_00_MBL	Pyrococcus furiosus	UP: P81414	1	0	0.00	NA	0	0.00
Group_01_MBL	metallo beta-lactamase II	UP: class-b beta-lactamase family	795	794	99.87	559.60	121	15.22
UNIPROT_blasB class B1 positives	metallo beta-lactamase II	Found positive with validated HMM	673	673	100.00	57.90	0	0.00
UNIPROT_blasB class B2 positives	metallo beta-lactamase II	Found positive with validated HMM	121	121	100.00	559.60	121	100.00
UNIPROT_blasB class B3 positives	metallo beta-lactamase II	Found positive with validated HMM	1	0	0.00	NA	0	0.00
Group_02_MBL	glyoxylase II	UP: glyoxylase II family	11621	365	3.14	24.60	0	0.00
Group_03_MBL	flavoprotein	UP: zinc metallo-hydrolase group 3	9334	34	0.36	18.90	0	0.00
Group_04_MBL	aryltransferase	UP: atsa family	7	0	0.00	NA	0	0.00
Group_05_MBL	cyclase	UP: Q9X466	1	1	100.00	40.70	0	0.00
Group_06_MBL	Cleavage and polyadenylation spec. factor	UP: RNA metabolizing MBL-like family	19812	1	0.01	13.30	0	0.00
Group_07_MBL	DNA cross-link repair	UP: DNA repair MBL drmbl family	19	0	0.00	NA	0	0.00
Group_08_MBL	comE	IP: ComA-like, MBL domain. IPR035681	22083	7	0.03	16.60	0	0.00
Group_09_MBL	Choline binding protein E	UP: Q9KGZ1	1	0	0.00	NA	0	0.00
Group_10_MBL	PhnP protein	IP: Phosphonate metabolism protein, MBL domain. IPR035682	1524	0	0.00	NA	0	0.00
Group_11_MBL	CMP-N-acetylneuraminate monoxygenase	UP: CMP-neu5ac hydroxylase family	12	0	0.00	NA	0	0.00
Group_12_MBL	Group 12: romA	UP: RomA family MBL fold metallo-hysrolase	208	0	0.00	NA	0	0.00
Group_13_MBL	alkyl sulfatase	UP: A0A1H1U5G4	1	0	0.00	NA	0	0.00
Group_14_MBL	carbofuran hydrolase	UP: Q9KJV0	1	0	0.00	NA	0	0.00
Group_15_MBL	methyl parathion hydrolase	UP: Q9ALW1	1	0	0.00	NA	0	0.00
Group_16_MBL	3',5'-cyclic nucleotide phosphodiesterase	UP: cAMP phosphodiesterase class-ii family	735	0	0.00	NA	0	0.00

1: search term used for sequence retrieval
2: NCBI BioProject containing sequences of functionally characterized antimicrobial resistance genes
3: Hits that justified HHMER inclusion criteria E-value ≤10.0
UP: UniProt; IP: InterPro; NA: not applicable



Supplementary Figure 6: A) B3 model sequence logo with the conserved amino-acids participating in the binding of the two zinc atoms being indicated by arrows. **B)** Frequency distributions of the HMMER bit scores during searches against the relevant data sets.

Supplementary Table 8: Class B3 model validation

data set	Daiyasu et al classification	Description	# Sequences	class B3 model				
				# HMMER hits[3]	% HMMER hits	max score	# score ≥ [cut-off=180]	% validated model hits
class_B1	metallo beta-lactamase II	NCBI BioProject: 3130472[2]; class B1	245	55	22.45	29.20	0	0.00
class_B2	metallo beta-lactamase II	NCBI BioProject: 3130472; class B2	15	11	73.33	17.70	0	0.00
class_B3	metallo beta-lactamase II	NCBI BioProject: 3130472; class B3	68	68	100.00	384.20	68	100.00
Group_00_MBL	Pyrococcus furiosus	UP: P81414	1	1	100.00	19.30	0	0.00
Group_01_MBL	metallo beta-lactamase II	UP: class-b beta-lactamase family[1]	795	504	63.40	341.80	1	0.13
UNIPROT_blasB class B1 positives	metallo beta-lactamase II	Found positive with validated HMM	673	387	57.50	27.10	0	0.00
UNIPROT_blasB class B2 positives	metallo beta-lactamase II	Found positive with validated HMM	121	116	95.87	18.10	0	0.00
UNIPROT_blasB class B3 positives	metallo beta-lactamase II	Found positive with validated HMM	1	1	100.00	341.80	1	100.00
Group_02_MBL	glyoxylase II	UP: glyoxylase II family	11621	11597	99.79	68.80	0	0.00
Group_03_MBL	flavoprotein	UP: zinc metallo-hydrolase group 3	9334	4800	51.42	42.40	0	0.00
Group_04_MBL	aryltransferase	UP: atsa family	7	1	14.29	24.90	0	0.00
Group_05_MBL	cyclase	UP: Q9X466	1	0	0.00	NA	0	0.00
Group_06_MBL	Cleavage and polyadenylation spec. factor	UP: RNA metabolizing MBL-like family	19812	13868	70.00	33.10	0	0.00
Group_07_MBL	DNA cross-link repair	UP: DNA repair MBL drmbl family	19	0	0.00	NA	0	0.00
Group_08_MBL	comE	IP: ComA-like, MBL domain. IPR035681	22083	9054	41.00	42.80	0	0.00
Group_09_MBL	Choline binding protein E	UP: Q9KGZ1	1	1	100.00	16.90	0	0.00
Group_10_MBL	PhnP protein	IP: Phosphonate metabolism protein, MBL domain. IPR035682	1524	215	14.11	19.20	0	0.00
Group_11_MBL	CMP-N-acetylneuraminate monoxygenase	UP: CMP-neu5ac hydroxylase family	12	0	0.00	NA	0	0.00
Group_12_MBL	Group 12: romA	UP: RomA family MBL fold metallo-hysrolase	208	0	0.00	NA	0	0.00
Group_13_MBL	alkyl sulfatase	UP: A0A1H1U5G4	1	1	100.00	17.80	0	0.00
Group_14_MBL	carbofuran hydrolase	UP: Q9KJV0	1	1	100.00	30.00	0	0.00
Group_15_MBL	methyl parathion hydrolase	UP: Q9ALW1	1	1	100.00	43.30	0	0.00
Group_16_MBL	3',5'-cyclic nucleotide phosphodiesterase	UP: cAMP phosphodiesterase class-ii family	735	112	15.24	16.80	0	0.00

1: search term used for sequence retrieval
2: NCBI BioProject containing sequences of functionally characterized antimicrobial resistance genes
3: Hits that justified HHMER inclusion criteria e≤10.0
UP: UniProt; IP: InterPro; NA: not applicable

2. Spreading of B1 MβLs in the *Prokaryotes*

Supplementary Table 9: Spreading of B1 metallo-β-lactamases genes in the *Prokaryotes*[a]

	Total hits		True hits		True hits secreted	
	% genomes	% species	% genomes	% species	% genomes	% species
Archaea (2312, 1192)[b]	0	0	0	0	0	0
Bacteria (378459, 54387)	6.16	6.26	5.97	5.98	5.88	5.82
Acidobacteriota (131, 100)	0.76	1.00	0.76	1.00	0.76	1.00
Actinomycetota (36654, 12333)	0.005	0.016	0	0	0	0
Aquificota (55, 40)	0	0	0	0	0	0
Bacillota (102102, 8980)	5.76	8.69	5.30	7.95	5.15	7.76
Bacteroidota (18764, 4305)	16.29	35.17	15.66	33.87	15.27	33.05
Balneolota (41, 32)	2.44	3.13	2.44	3.13	2.44	3.13
Bdellovibrionota (75, 43)	6.67	9.3	6.67	9.3	6.67	9.3
Campylobacterota (10201, 610)	0	0	0	0	0	0
Chlamydiota (544, 54)	0	0	0	0	0	0
Chlorobiota (48, 30)	0	0	0	0	0	0
Chloroflexota (211, 99)	0	0	0	0	0	0
Cyanobacteriota (1918, 1172)	0	0	0	0	0	0
Deferribacterota (22, 13)	0	0	0	0	0	0
Deinococcota (360, 165)	0	0	0	0	0	0
Fibrobacterota (90, 48)	0	0	0	0	0	0
Fusobacteriota (742, 118)	0	0	0	0	0	0
Gemmatimonadota (41, 14)	2.44	7.14	2.44	7.14	2.44	7.14
Lentisphaerota (25, 8)	0	0	0	0	0	0
Mycoplasmatota (1888, 364)	0	0	0	0	0	0
Myxococcota (223, 151)	65.92	66.23	65.92	66.23	59.64	62.91
Nitrospirota (43, 32)	0	0	0	0	0	0
Planctomycetota (247, 171)	0.81	1.17	0.40	0.58	0.40	0.58
Pseudomonadota (199899, 24417)	7.02	3.83	6.97	3.74	6.95	3.66
Rhodothermota (147, 19)	2.04	15.79	2.04	15.79	2.04	15.79
Spirochaetota (2201, 253)	5.86	19.76	5.54	18.18	3.68	13.44
Synergistota (135, 54)	0	0	0	0	0	0
Thermodesulfobacteriota (747, 344)	3.21	1.74	3.21	1.74	3.21	1.74
Thermotogota (170, 103)	0	0	0	0	0	0
Verrucomicrobiota (555, 195)	0.72	1.54	0.72	1.54	0.72	1.54

a: Taxa with ≥ 15 genomes sequenced are only listed. For a full list see Supplemental Data 5.

b: The first number in parenthesis indicate the number of genomes and the second the number of species for the respective taxon.

Supplementary Table 10: Bacterial phyla with significant carriage rates of resistance capable MβL genes[a] and spreading in their respective classes.[b]

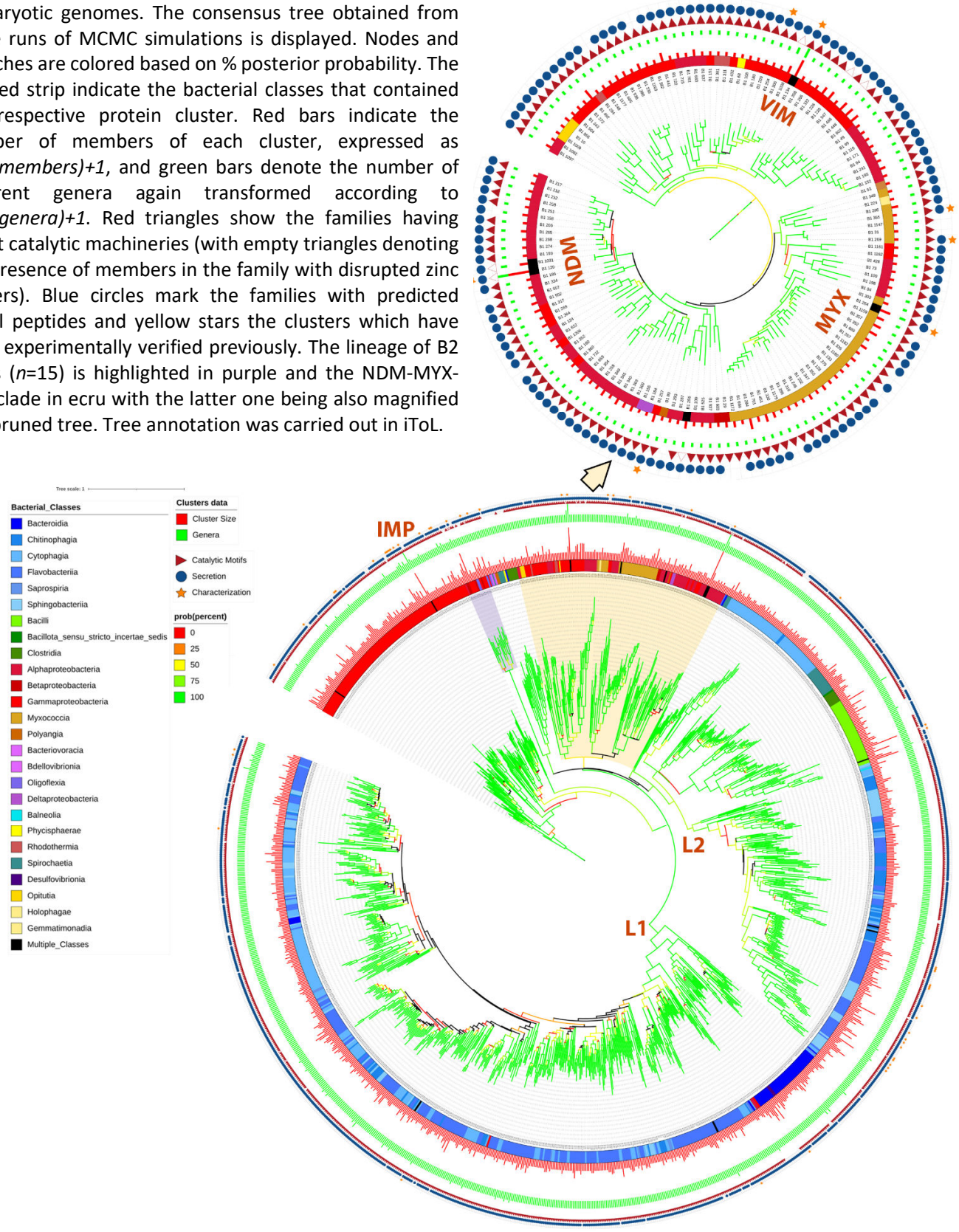
Taxon	% genomes	% species	Hits	Protein clusters	Richness
***Bacillota* (102102, 8980)[c]**	**5.15**	**7.76**	**5263**	**38**	**$7.2 \cdot 10^{-3}$**
Bacilli (85518, 6368)	6.12	10.69	5241	28	$5.3 \cdot 10^{-3}$
Bacillota incertae sedis (37, 14)	0	0	0	na	na
Clostridia (12641, 1997)	0.17	0.75	21	9	$4.3 \cdot 10^{-1}$
Erysipelotrichia (1651, 183)	0	0	0	na	na
Negativicutes (1721, 239)	0	0	0	na	na
Tissierellia (516, 165)	0	0	0	na	na
***Bacteroidota* (18764, 4305)**	**15.27**	**33.05**	**3156**	**774**	**$2.5 \cdot 10^{-1}$**
Bacteroidia (11728, 807)	2.96	8.67	347	40	$1.1 \cdot 10^{-1}$
Chitinophagia (295, 221)	44.07	43.44	143	79	$5.5 \cdot 10^{-1}$
Cytophagia (903, 694)	53.27	53.17	496	254	$5.1 \cdot 10^{-1}$
Flavobacteriia (5076, 2088)	31.67	34.00	1872	313	$1.7 \cdot 10^{-1}$
Saprospiria (47, 33)	34.04	36.36	16	9	$5.6 \cdot 10^{-1}$
Sphingobacteriia (712, 460)	39.61	36.09	282	83	$2.9 \cdot 10^{-1}$
***Myxococcota* (223, 151)**	**59.64**	**62.91**	**133**	**28**	**$1.5 \cdot 10^{-1}$**
Myxococcia (182, 120)	71.43	76.67	130	25	$1.9 \cdot 10^{-1}$
Polyangia (41, 31)	7.32	9.68	3	3	1.0
***Pseudomonadota* (199899, 24417)**	**6.95**	**3.66**	**14441**	**203**	**$1.4 \cdot 10^{-2}$**
Acidithiobacilla (139, 28)	0	0	0	na	na
Alphaproteobacteria (17474, 8269)	0.45	0.68	78	51	$6.5 \cdot 10^{-1}$
Betaproteobacteria (17197, 3131)	0.20	0.42	37	5	$1.4 \cdot 10^{-1}$
Gammaproteobacteria (165063, 12970)	8.34	6.36	14296	152	$1.0 \cdot 10^{-2}$
***Spirochaetota* (2201, 253)**	**3.68**	**13.44**	**86**	**18**	**$2.1 \cdot 10^{-1}$**
Spirochaetia (2201, 253)	3.68	13.44	86	18	$2.1 \cdot 10^{-1}$

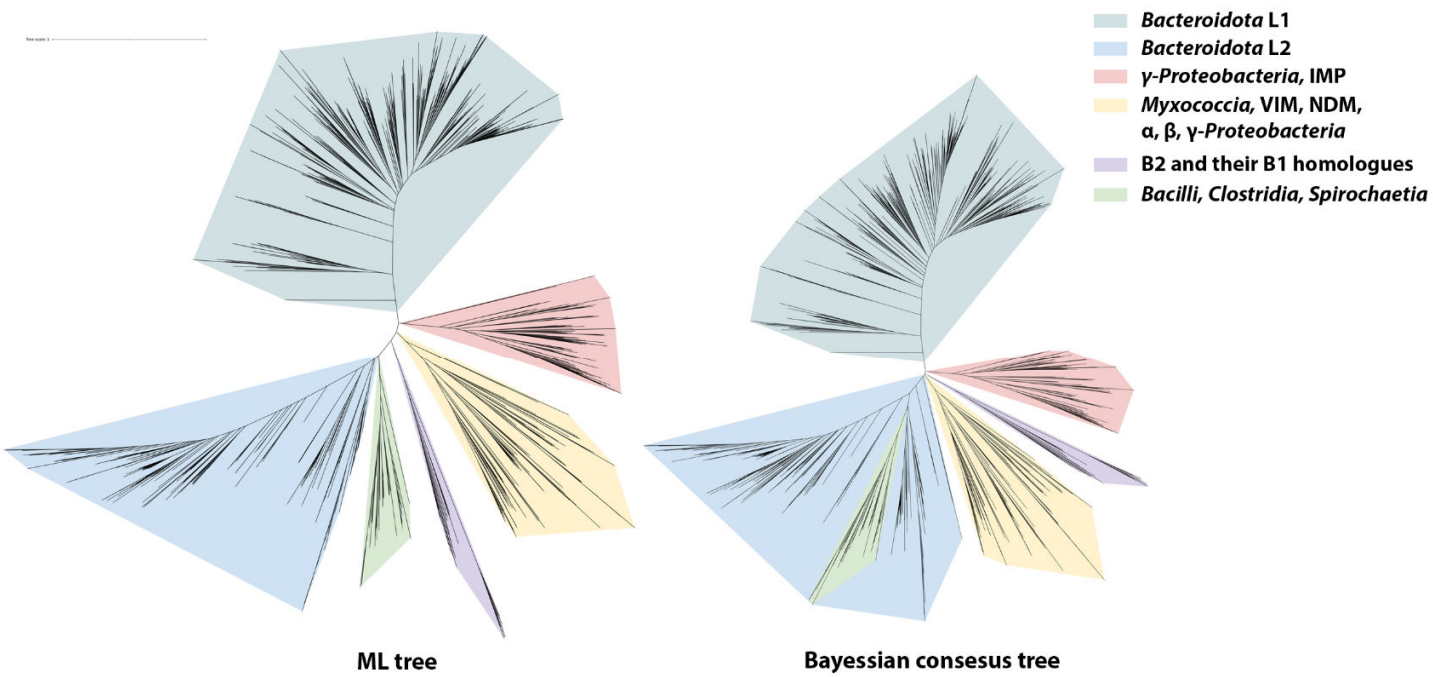
a: Intact open reading frames, expressing for secreted B1 MβL homologues with all catalytic motifs present.

b: Taxa with $\geq$ 15 genomes sequenced are only listed. For a full list see Supplemental Data 5.

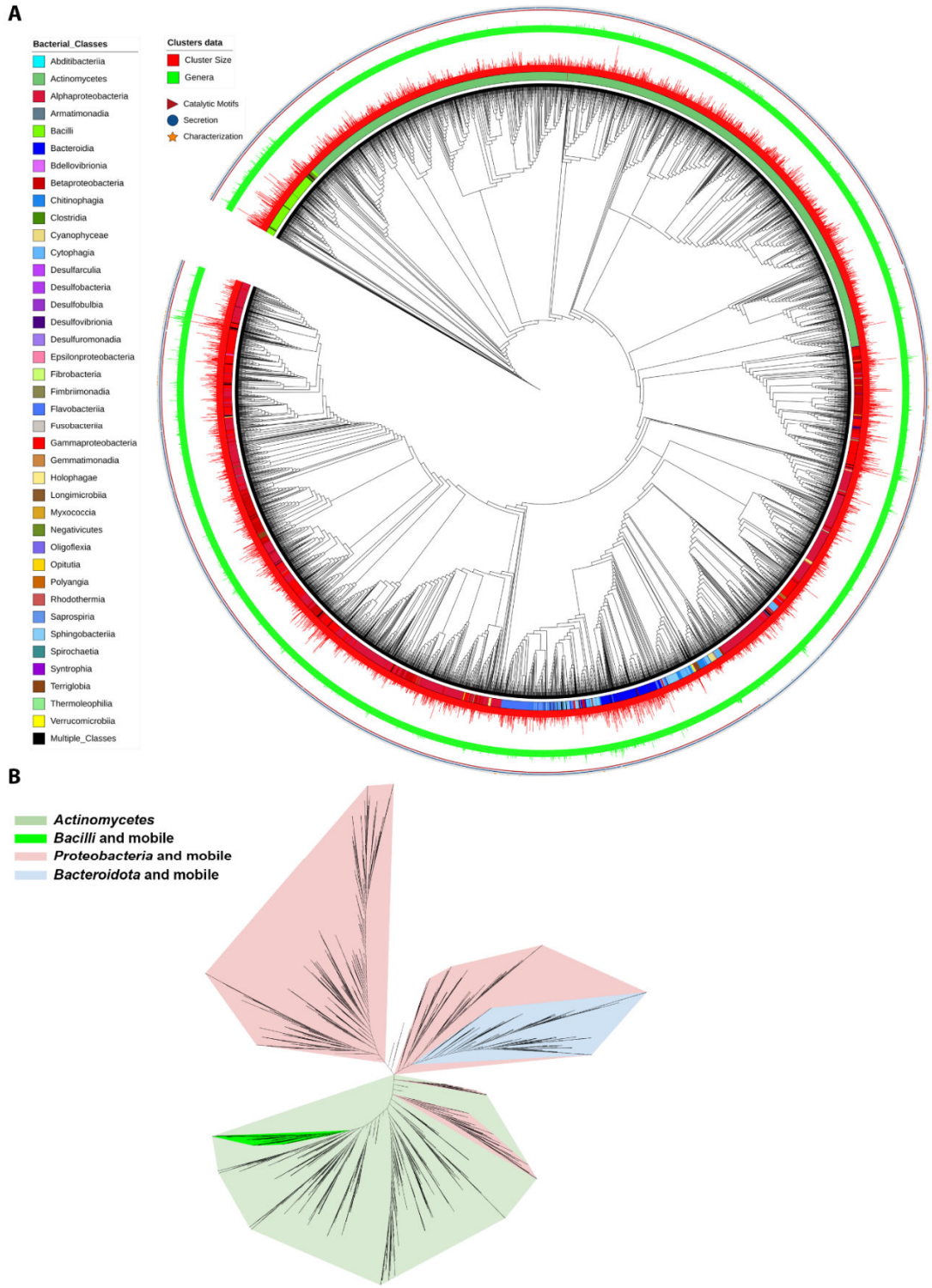
c: The first number in parenthesis indicate the number of genomes and the second the number of species for the respective taxon.

Supplementary Figure 7: Bayessian inferred phylogeny of the B1 MβL orthologous families (*n*=1196) detected in prokaryotic genomes. The consensus tree obtained from three runs of MCMC simulations is displayed. Nodes and branches are colored based on % posterior probability. The colored strip indicate the bacterial classes that contained the respective protein cluster. Red bars indicate the number of members of each cluster, expressed as *log(#members)+1*, and green bars denote the number of different genera again transformed according to *log(#genera)+1*. Red triangles show the families having intact catalytic machineries (with empty triangles denoting the presence of members in the family with disrupted zinc centers). Blue circles mark the families with predicted signal peptides and yellow stars the clusters which have been experimentally verified previously. The lineage of B2 MβLs (*n*=15) is highlighted in purple and the NDM-MYX-VIM clade in ecru with the latter one being also magnified in a pruned tree. Tree annotation was carried out in iToL.

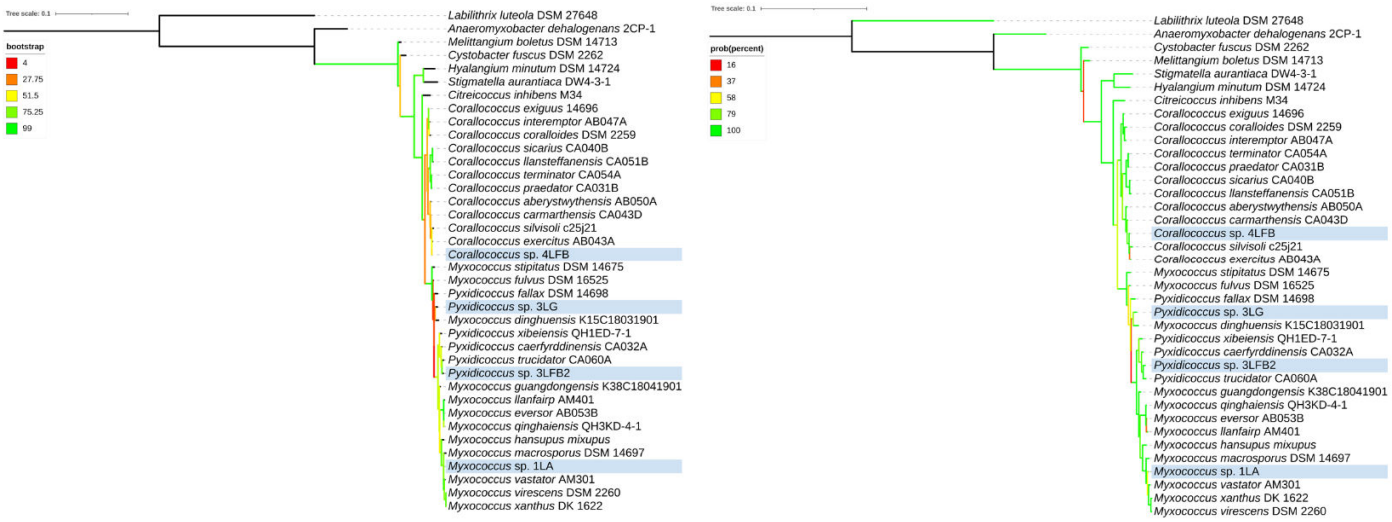




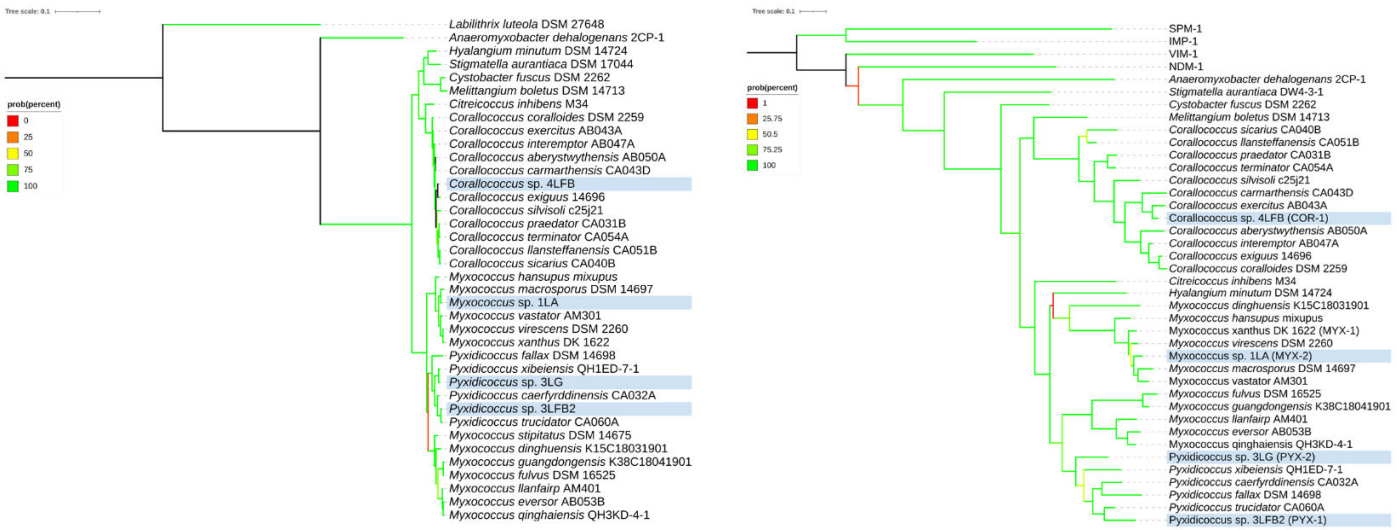
Supplementary Figure 8: Comparison of Maximum-Likehood (ML) and Bayessian inferred phylogenies of the B1 MβL orthologous families. The main lineages are highlighted in each tree. Based on our analyses each of the main lineage is mainly composed from representatives which either belonged to specific OTUs or to a specific type of enzymes (e.g. B2 MβLs and B1 enzymes lacking His116). The two methodologies yielded similar topologies. Note the different positioning of the *Bacilli*, *Clostridia*, *Spirochaetia* cluster between the two trees. ML phylogeny appeared more robust as the above branch was positioned closer to the B2 cluster.



Supplementary Figure 9: A) ML phylogeny of class A β-lactamases orthologous families (*n*=6120) identified in prokaryotic genomes. Tree annotation has been carried out as described in the legend of Figure 1. Genes identified in *Actinomycetes* exhibit a monophyletic radiation that gave rise to the families identified in other phyla as well as to mobile genes. B) Un-rooted representation of the ML tree highlighting the main lineages. Actinobacterial class A β-lactamases have been transferred into *Bacilli* and *Proteobacteria* during two independent mobilization events. Enzymes identified in *Bacteroidota* most likely originated in *Proteobacteria*.

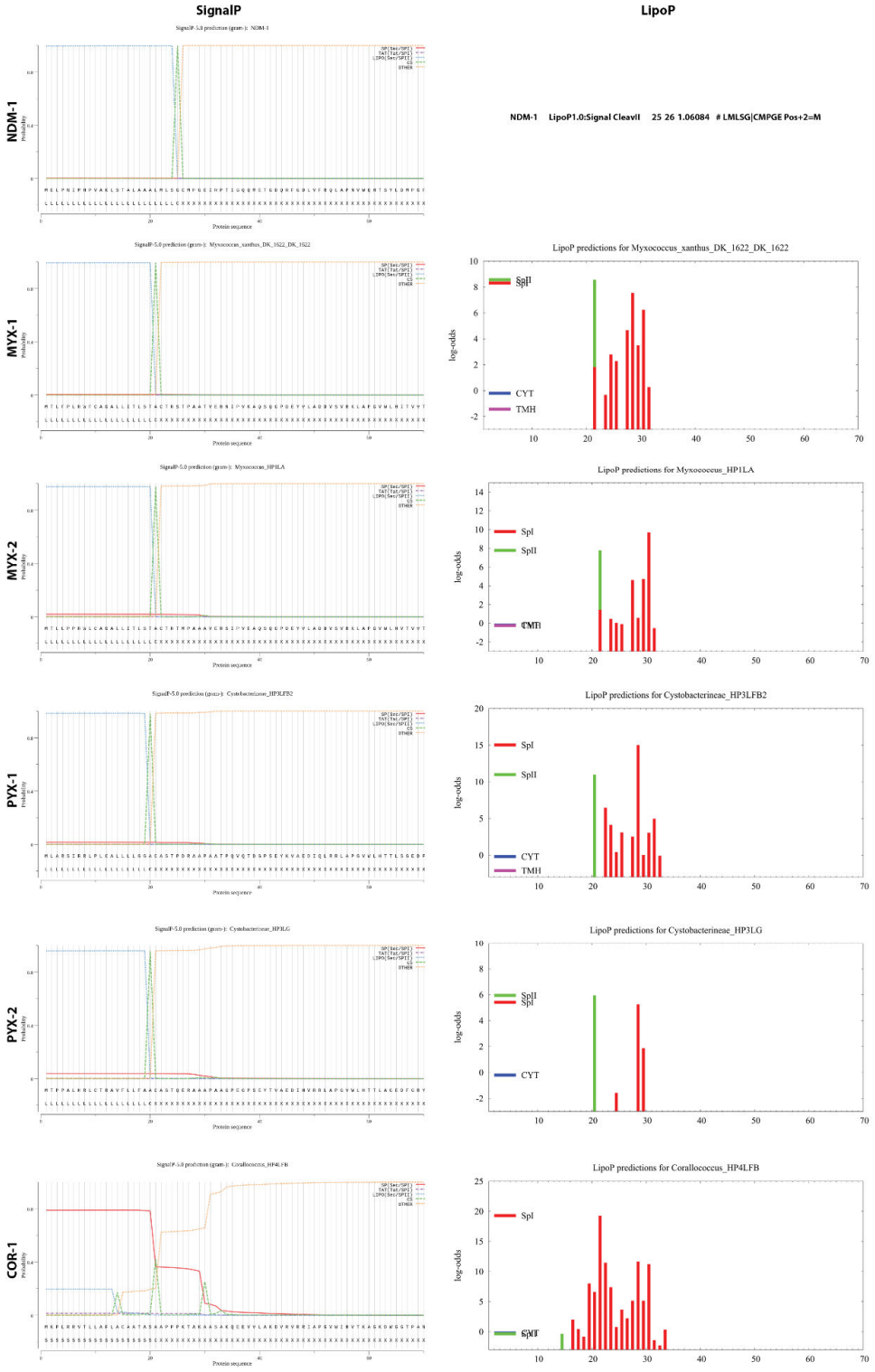


Supplementary Figure 10: Phylogenetic relationships of 16S rRNA genes from reference *Myxococcia* genomes as well as from novel strains identified in this study (highlighted in blue). The gene from *Labilithrix luteola* (*Polyangia*) was included as an outlier. ML (left) and Bayessian (right) phylogenies are displayed inferred as described in the Materials and Methods section of the main manuscript.

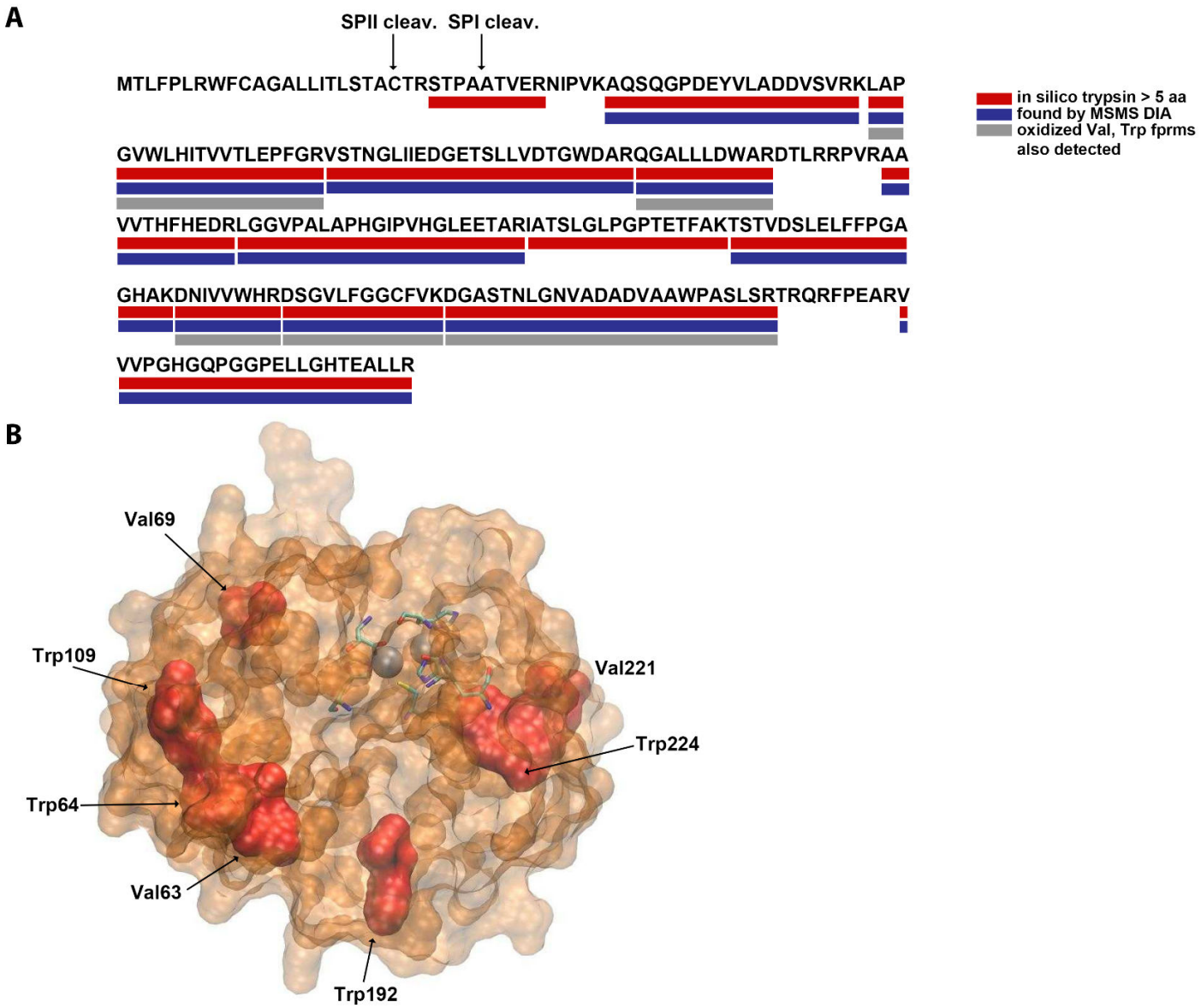


Supplementary Figure 11: Bayessian inferred phylogenies of RpoB and B1 MβLs from reference myxobacteria and strains isolated herein. The clinically relevant NDM-1, VIM-1, IMP-1 and SPM-1 are also included. Note that the B1 MβL from *Hyalangium minitum* DSM 14724 is within the myxococci-pyxidicocci lineage and it does not cluster with the enzymes from the other *Archangiaceae*.

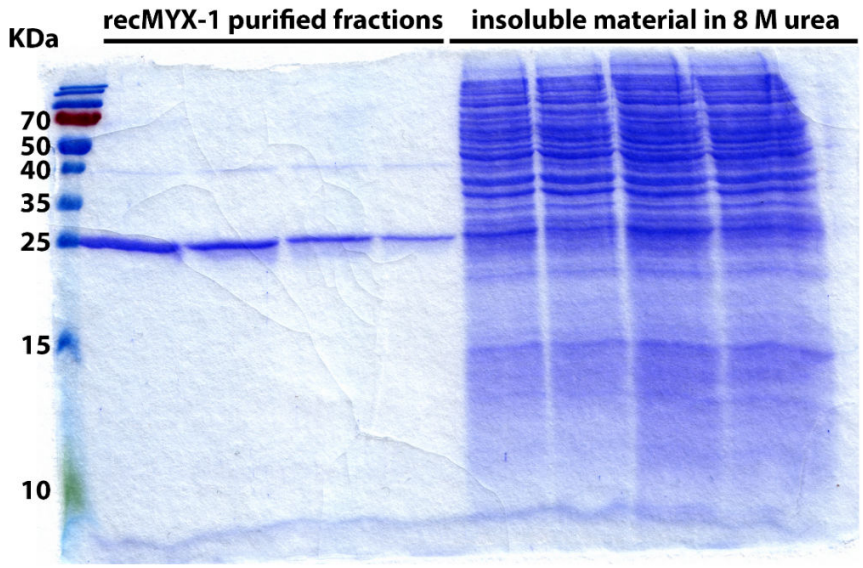
3. Expression of myxobacterial MβLs in *E. coli*



Supplementary Figure 12: Signal peptide predictions of NDM-1 and myxobacterial MβLs using SignalP and LipoP. Save for COR-1, SignalP predicted solely a type II signal peptide for myxobacterial enzymes while LipoP gave additional type I signal peptides having equivalent probabilities with the respective type II SP. NDM-1 is predicted to be a lipoprotein by both algorithms.



Supplementary Figure 13: A) Identified peptides in MSMS analysis of purified MYX-1. Red lines denote the *in silico* predicted peptides with size >5 amino acids after digestion with trypsin. Blue ranges indicate the unmodified detected peptides during our MSMS experiments. Apart from the native forms, oxidized species were also detected for several peptides marked with grey lines. Similar data were obtained for the purified MYX-2. B) Three dimensional model of MYX-1 with oxidized valine and tryptophan residues being represented as red surfaces. The above residues were located in the protein's interior.



Supplementary Figure 14: Purified fractions of recMYX-1 isolated from cell free extracts and protein content of the insoluble material left after membrane removal of the pellet obtained during cell disruption as analyzed with SDS-PAGE (15% polyacrylamide gel electrophoresed with Tris-glycine buffer and stained with Coomassie Briliant Blue R-250). The insoluble material has been dissolved with 8 M. A protein with the same apparent molecular mass as the purified enzyme was present in the insoluble material which precipitated again in the refolding buffer indicating that it is in a form not permitting the adaptation of the native fold *in vitro*. Due to the probable aggregation propensity of recMYX-1 we eventually purified it from the culture supernatant (see *Materials and Methods* in the main manuscript) obtained after induction at the presence of glycine as it yielded higher quantities of homogenous enzyme.

Supplementary Table 11: Kinetic and thermal stability of the studied metallo-β-lactamases

Enzyme	T_m (°C)			
	Kinetic stability [a] Periplasm	Kinetic stability [a] Pure enzymes (lysis buffer)	Kinetic stability [a] Pure enzymes	DSF [b]
MYX-1	65.06±1.17	69.02±1.51	63.13±1.02	61.05±0.22
MYX-2	73.20±0.81	75.46±1.40	71.73±1.52	68.78±0.33
COR-1	64.97±1.23	67.30±1.64	63.81±1.31	61.78±0.05
PYX-1	73.70±1.55	71.82±0.78	64.84±0.92	62.28±0.25
PYX-2	71.83±0.87	72.59±1.11	65.42±0.86	63.22±0.09
recMYX-1	66.18±1.34	69.02±1.08	63.42±1.63	61.05±0.08
recNDM-1	62.14±1.56	59.23±0.93	55.04±1.78	56.28±0.09

a: Data are the means of two independent experiments ± standard deviations.

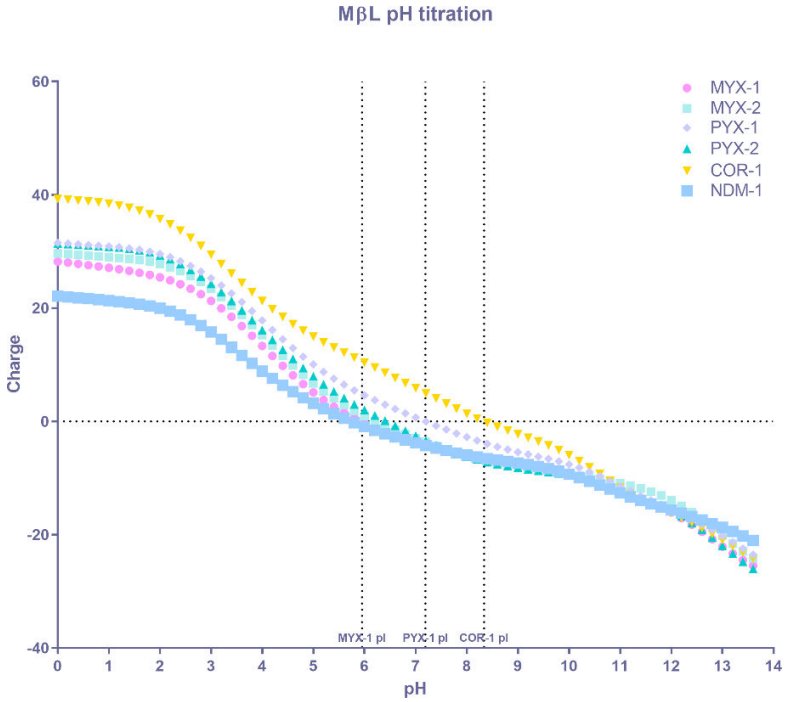
b: Data represent means of three measurements ± standard deviations.

Supplementary Table 12: Amino acid types of the residue 3 positions upstream the N-terminal cysteine of the mature lipoproteins predicted by SignalP v5 (probability >99%) in the genomes of *M. xanthus* and *E. coli*.

Genome	*n* lipoproteins			
	leucine	other hydrophobic	hydrophilic	total lipoproteins
M. xanthus	195 (49.7%)	160 (40.8%)	37[a] (9.4%)	392
E. coli	88 (82.2%)	17 (19.3%)	2[b] (0.02%)	107

a: 18 lipoproteins contained threonine, 18 serine and one glutamine

b: both hypothetical lipoproteins contained threonine



Supplementary Figure 15: *In silico* pH titration curves of the studied enzymes. Isoelectric points (pI) predicted by the method were used during purification using ion exchange chromatography.

References

1. **Bebrone C.** 2007. Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. Biochem Pharmacol 74: 1686-701
2. **Cheng Z, VanPelt J, Bergstrom A, Bethel C, Katko A, Miller C, Mason K, Cumming E, Zhang H, Kimble RL, Fullington S, Bretz SL, Nix JC, Bonomo RA, Tierney DL, Page RC, Crowder MW.** 2018. A Noncanonical Metal Center Drives the Activity of the Sediminispirochaeta smaragdinae Metallo-beta-lactamase SPS-1. Biochemistry 57: 5218-29
3. **Daiyasu H, Osaka K, Ishino Y, Toh H.** 2001. Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. FEBS Lett 503: 1-6
4. **Henderson TA, Young KD, Denome SA, Elf PK.** 1997. AmpC and AmpH, proteins related to the class C beta-lactamases, bind penicillin and contribute to the normal morphology of Escherichia coli. J Bacteriol 179: 6112-21
5. **Kumarasiri M, Llarrull LI, Borbulevych O, Fishovitz J, Lastochkin E, Baker BM, Mobashery S.** 2012. An amino acid position at crossroads of evolution of protein function: antibiotic sensor domain of BlaR1 protein from Staphylococcus aureus versus class D beta-lactamases. J Biol Chem 287: 8232-41

6. **Lin HV, Massam-Wu T, Lin CP, Wang YA, Shen YC, Lu WJ, Hsu PH, Chen YH, Borges-Walmsley MI, Walmsley AR.** 2017. The Vibrio cholerae var regulon encodes a metallo-beta-lactamase and an antibiotic efflux pump, which are regulated by VarR, a LysR-type transcription factor. PLoS One 12: e0184255
7. **Massova I, Mobashery S.** 1998. Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. Antimicrob Agents Chemother 42: 1-17
8. **Meini MR, Llarrull LI, Vila AJ.** 2015. Overcoming differences: The catalytic mechanism of metallo-beta-lactamases. FEBS Lett 589: 3419-32
9. **Vercheval L, Bauvois C, di Paolo A, Borel F, Ferrer JL, Sauvage E, Matagne A, Frere JM, Charlier P, Galleni M, Kerff F.** 2010. Three factors that modulate the activity of class D beta-lactamases and interfere with the post-translational carboxylation of Lys70. Biochem J 432: 495-504