Received: 4 December 2025 | Revised: 18 June 2026 | Accepted: 19 June 2026

DOI: 10.1002/imo2.70118

RESEARCH ARTICLE

iMetaOmics WILEY

Antarctic soil microbiomes encode structurally conserved and phylogenetically diverse beta-lactamases

José Coche-Miranda[1] | **Patricio Arros**[1] | **Nicolás Canales**[1] | **Camilo Berríos-Pastén**[1] | **Gastón Azziz**[2,3] | **Rosalba Lagos**[1] | **Julieta Orlando**[4,5] | **Francisco P. Chávez**[6] | **Andrés E. Marcoleta**[1]

[1]Grupo de Microbiología Integrativa, Laboratorio de Biología Estructural y Molecular BEM, Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

[2]Microbiology Laboratory, Faculty of Agronomy, Universidad de la República, Montevideo, Uruguay

[3]Molecular Microbiology Unit, Clemente Estable Biological Research Institute, Montevideo, Uruguay

[4]Laboratorio de Ecología Microbiana, Departamento de Ecología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

[5]Instituto Milenio Biodiversidad de Ecosistemas Antárticos y Subantárticos, Santiago, Chile

[6]Laboratorio de Microbiología de Sistemas, Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

Correspondence
Andrés E. Marcoleta, Grupo de Microbiología Integrativa, Laboratorio de Biología Estructural y Molecular BEM, Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, 7800003, Chile.
Email: amarcoleta@uchile.cl

Funding information
Agencia Nacional de Investigación y Desarrollo, Grant/Award Number: FONDECYT 1221193

Abstract

Antimicrobial resistance is a major global health threat shaped not only by clinical settings but also by environmental reservoirs, in line with the One Health framework. Antarctic and Subantarctic soils represent a largely unexplored component of the environmental resistome, particularly beta-lactamases, whose diversity, structural features, and biogeographic patterns remain poorly resolved. Here, we applied an integrative metagenomic approach combining sequence-, structure-, and function-based similarity to systematically identify putative beta-lactamase orthologs across an extensive collection of Antarctic and Subantarctic soil metagenomes. We detected multiple beta-lactamase classes and families, with a predominance of class A serine beta-lactamases and subclass B3 metallo-beta-lactamases. Several previously uncharacterized families and divergent variants were identified, many of which retain conserved catalytic architectures. Beta-lactamase diversity and relative abundance were strongly patterned across three major biogeographic clusters: Subantarctic islands, the Antarctic Peninsula and surrounding islands, and continental cold deserts, mirroring shifts in microbial community composition. The Antarctic Peninsula and maritime Antarctica exhibited the highest overall beta-lactamase abundance. Analysis of more than 3500 metagenome-assembled genomes indicated that most beta-lactamases were encoded by members of *p_Pseudomonadota* and *p_Bacteroidota*, with additional representatives distributed across other bacterial phyla. A subset was linked to mobile genetic elements, suggesting a latent but limited potential for horizontal gene transfer. Together, these findings reveal a geographically organized and compositionally diverse repertoire of beta-lactamases in (sub) Antarctic soils, underscoring the evolutionary persistence of conserved

José Coche-Miranda and Patricio Arros contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). *iMetaOmics* published by John Wiley & Sons Australia, Ltd on behalf of iMeta Science.

catalytic frameworks in cold environments and highlighting the contribution of remote ecosystems to global resistome diversity.

KEYWORDS
antibiotic inactivation, beta-lactamase, cryosphere, natural resistome, structure-guided metagenomics

Highlights

- Antarctic soils harbor a diverse and geographically structured reservoir of beta-lactamases.
- Class A and subclass B3 beta-lactamases dominate across (sub)Antarctic metagenomes.
- Integrative sequence-structure-function analysis reveals novel and highly divergent beta-lactamases.
- Beta-lactamases are primarily encoded by *p_Pseudomonadota* and *p_Bacteroidota* across Antarctic soils.

1 | INTRODUCTION

Antimicrobial resistance (AMR) is a major global health challenge, responsible for an estimated 1.3 million deaths annually and projected to cause up to 10 million deaths per year by 2050 if effective countermeasures are not implemented [1]. Addressing AMR requires understanding not only clinical dynamics but also its environmental origins, in line with the One Health framework, which integrates the interconnected roles of humans, animals, and ecosystems [2]. Within this context, it has been widely demonstrated that the antibiotic resistance genes (ARGs) are ancient and ubiquitous, and that environmental microbiota, even in remote and extreme environments, constitute reservoirs from which resistance determinants may emerge [3–5].

Among ARGs of particular concern are those encoding beta-lactamases (BLAs), the most widespread and efficient resistance mechanism against beta-lactam antibiotics in Gram-negative bacteria [6–8]. Beta-lactams remain the most extensively used antimicrobial agent in clinical practice [9], underscoring the importance of understanding the diversity and evolution of BLAs. Based on sequence similarity and structural features, BLAs are classified into two major groups: serine beta-lactamases (SBLs; classes A, C, and D) and metallo-beta-lactamases (MBLs; subclasses B1, B2, and B3) [10, 11]. SBLs belong to the PBP-like superfamily and share conserved active-site motifs (SXXK, SXN, and KTG (T/S)), whereas MBLs belong to the MBL-fold superfamily and coordinate zinc ions through conserved histidine and aspartate residues [12, 13]. MBLs are characterized by an alpha-beta-beta-alpha-fold, and two conserved active-site metal-binding motifs involving H116/H118/H196 and D120/H121/H263, along with D221 that coordinates both metals. Both groups of beta-lactamases are evolutionarily ancient, with origins estimated at approximately 2–2.5 billion years ago [12, 14], suggesting that their primary biological roles extend beyond clinical antibiotic resistance.

BLAs have been identified across diverse environments, including ecosystems with minimal human impact [15, 16]. Maritime Antarctic soils, in particular, have been reported to harbor potential ARGs, including BLAs, and to host bacteria resistant to multiple beta-lactam antibiotics [5, 17, 18]. Nevertheless, the distribution, taxonomic origin, and structural characteristics of beta-lactamases across Antarctic soil environments remain incompletely understood.

To date, only a limited number of studies have explored BLAs in Antarctic metagenomes. A survey of 11 microbial mat metagenomes along a latitudinal gradient identified putative BLAs from classes A–D, with class C predominating [19]. However, soils, expected to exhibit higher microbial and genetic diversity than microbial mats [20, 21], have not been comprehensively surveyed across distinct Antarctic regions. Moreover, previous studies did not systematically assess the taxonomic distribution, structural conservation, or evolutionary placement of the identified BLA sequences, limiting insight into their ecological roles and evolutionary trajectories.

Identifying beta-lactamase orthologs based solely on sequence identity presents important limitations, particularly for deeply divergent environmental variants. Remote homologs may exhibit low similarity to clinically characterized enzymes, increasing the likelihood of both false positives and false negatives in conventional searches. We therefore hypothesized that Antarctic soils encode a broad and largely unexplored repertoire of beta-lactamases that can be more reliably detected by integrating multiple

complementary layers of information derived from metagenomic data.

To test this hypothesis, we applied an integrative metagenomic workflow combining sequence similarity, structural prediction, and multidimensional functional inference to an extensive collection of soil metagenomes from Subantarctic islands, maritime Antarctica, and continental cold desert environments. This strategy enabled systematic identification of putative beta-lactamase orthologs while reducing misclassification within large protein superfamilies, thereby allowing a comprehensive assessment of their diversity, structural organization, and distribution across Antarctic ecosystems.

2 | RESULTS

Biogeographic patterns of microbial community structure across Antarctic soils

We constructed a set of 81 soil metagenomes, including six from sub-Antarctic Islands (SI): South Georgia, Marion, Kerguelen, Bartolome, Possession, and Bouvet Islands. Additionally, 37 metagenomes from the Antarctic Peninsula and surrounding islands (AP), mainly the South Shetland Archipelago; and 38 from high-latitude cold desert areas (CD), primarily the McMurdo Dry Valleys, the Shackleton Glacier, and the Union Glacier (Figure 1A). The complete list of metagenomes included in this study, along with their accession numbers and relevant metadata, is provided in Table S1.

First, we aimed to explore the diversity and taxonomic composition of the microbial communities inhabiting the different Antarctic soils studied. We used the SingleM pipeline, which is specially designed for microbial diversity analyses from shotgun metagenomic data based on a collection of 59 essential single-copy genes (ESCGs) [22]. However, we only kept the results for 43 ESCGs because 16 were not observed in at least one sample. Shannon Index values differed significantly among soil types, with the lowest diversity in CD (median = 5.12), followed by AP (median = 5.55), and the highest SI (median = 6.40) (Figure 1B). This pattern likely reflects similar richness in CD and AP (median = 393 and 381, respectively), both markedly lower than in SI (median = 882) (Figure 1C).

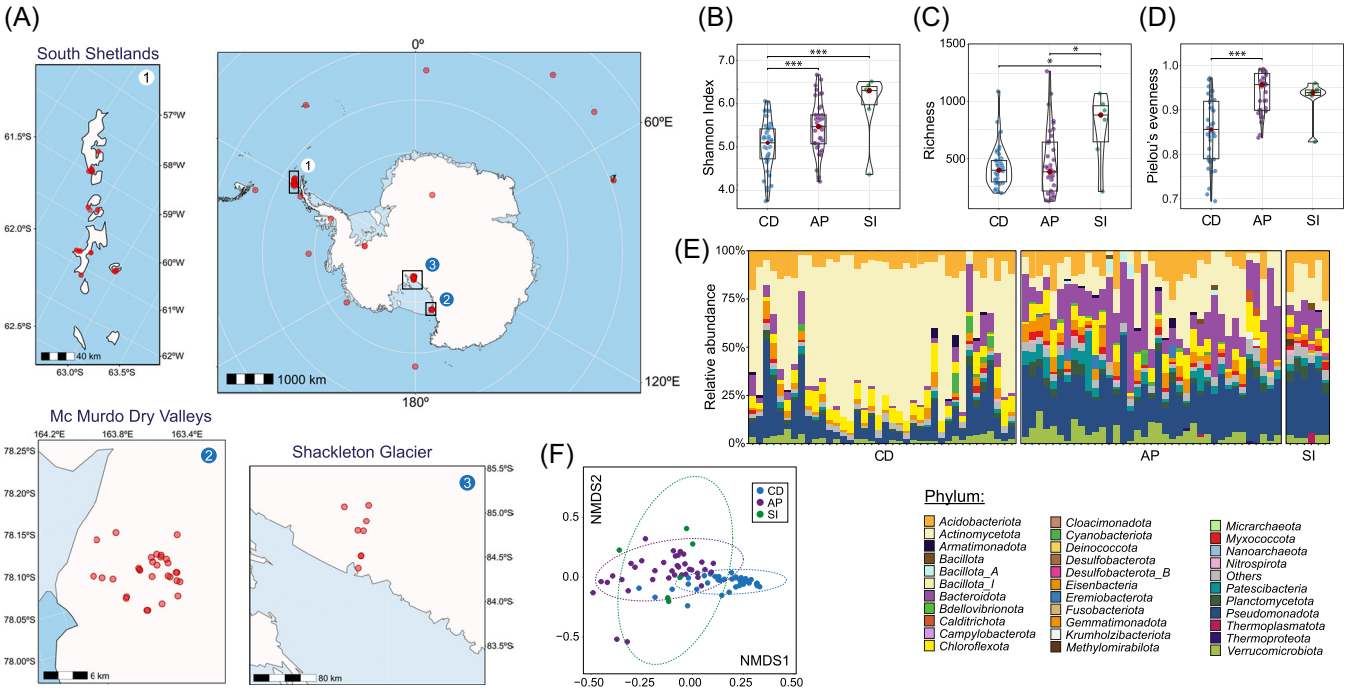


FIGURE 1 Geographic distribution and microbial diversity of Antarctic and Subantarctic soil metagenomes. (A) Map showing the geographic location of the 81 soil metagenomes included in this study. Insets 1–3 show enlarged views of the regions with higher sample density. (B–D) distribution of diversity indices across cold desert (CD), Antarctic Peninsula and surrounding islands (AP), and Subantarctic Islands (SI) soils, based on 43 essential single-copy marker genes. *$p_{\text{FDR-adjusted}} < 0.05$ and ***$p_{\text{FDR-adjusted}} < 0.001$. (E) Relative abundance of microbial phyla across the analyzed metagenomes. (F) Non-metric multidimensional scaling (NMDS) ordination based on Bray–Curtis dissimilarities calculated from family-level taxonomic profiles. PERMANOVA indicated significant differences in microbial community structure between AP and CD and between CD and SI soils. Complementary ordination quality and beta-dispersion analyses are shown in Figure S1.

Additionally, evenness was high in AP and SI (medians: 0.95 and 0.94) but significantly reduced in CD (median: 0.86) (Figure 1D).

Regarding microbial community structure, the SI and AP soils showed a more similar phyla composition, predominating *p_Pseudomonadota* (mainly *o_Burkholderiales*), *p_Bacteroidota* (mainly *o_Chitinophagales*), *p_Acidobacteriota* (mainly *o_Pyrinomonadales*), and, to a lesser extent, *p_Actinomycetota* (mainly *o_Acidimicrobiales*) (Figure 1E). Conversely, in most CD soils, *p_Actinomycetota* (mainly *o_Solirubrobacterales*) predominated, whereas *p_Bacteroidota*, *p_Pseudomonadota*, *p_Gemmatimonadota*, and *p_Patescibacteria* were markedly less abundant. In agreement, beta-diversity dissimilarity analysis at the family level revealed a more similar taxonomic composition between SI and AP soils, whereas the CD soils clustered apart, due to their distinct community structure (Figure 1F, Figure S1).

An integrative strategy for Antarctic beta-lactamase identification

To identify beta-lactamase orthologs, we applied an integrative workflow that combines complementary protein assembly and clustering strategies with sequence-, structure-, and function-based similarity embeddings to generate confident protein ortholog clusters (Figure 2). This approach enables discrimination between true beta-lactamase orthologs and phylogenetically related proteins with different functions.

Metagenomic reads from the 81 soil datasets were assembled at the contig level using metaSPAdes, followed by open reading frame (ORF) prediction, yielding approximately 9 million protein sequences. Additionally, direct protein assembly from metagenomic reads was performed using PLASS [23]. PLASS enables ultra-fast, scalable protein-level assembly, overcoming biases in contig reconstruction to uncover low-abundance proteins, protein variants, and novel sequences that are often missed by traditional workflows. This complementary strategy increased the total number of predicted proteins to approximately 98 million.

To reduce redundancy, all predicted proteins were clustered at 99% sequence identity and 99% bidirectional coverage, resulting in approximately 62 million non-redundant sequences. These were combined with curated reference beta-lactamases from the Beta-Lactamase DataBase (BLDB; 5,458 non-redundant representatives) and a set of 24 phylogenetically related non-beta-lactamase

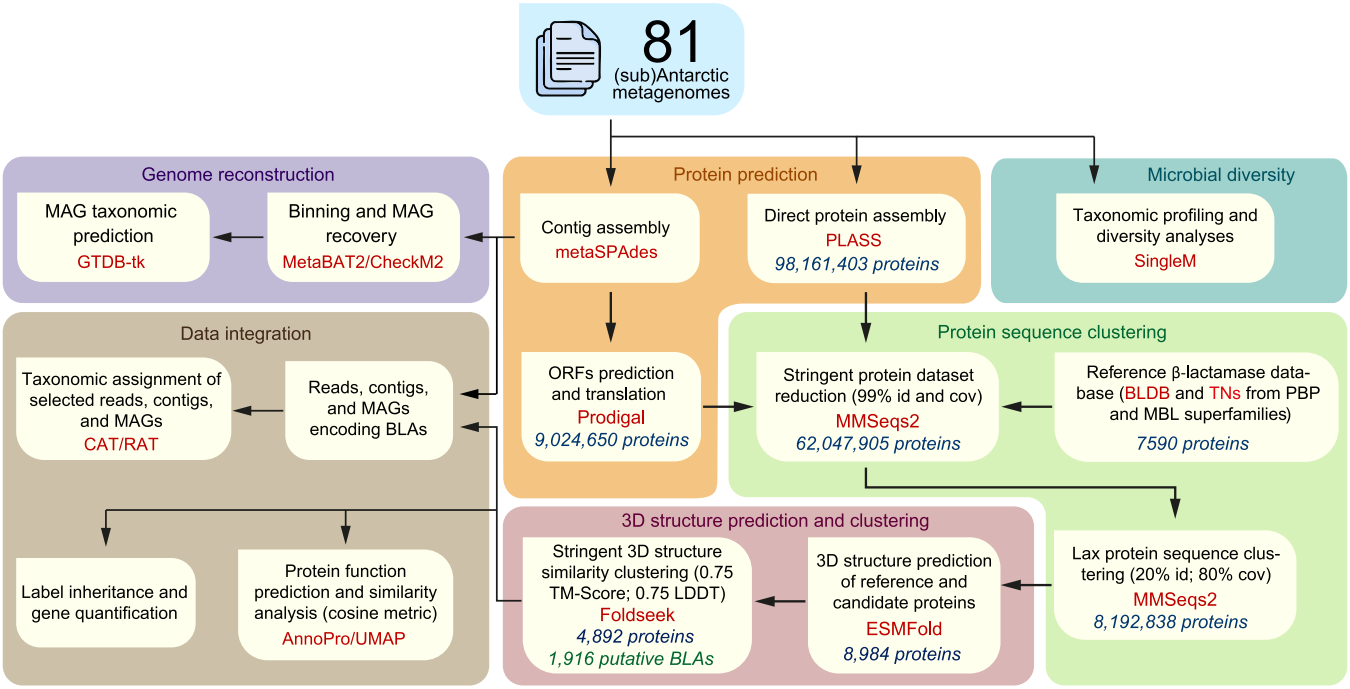


FIGURE 2 Integrative bioinformatic workflow for Antarctic beta-lactamase identification and microbial diversity analysis. The workflow summarizes the main analytical stages used in this study, including metagenome assembly, direct protein assembly, protein prediction, protein sequence clustering, structure prediction and structural clustering, functional similarity analysis, genome reconstruction, taxonomic assignment, gene quantification, and microbial diversity profiling. Main workflow modules are highlighted by color. Bioinformatic tools used in each step are indicated in red, intermediate protein counts are shown in blue, and the final number of putative Antarctic beta-lactamases is shown in green.

proteins from the PBP-like and MBL-fold superfamilies used as negative controls (true negatives; TNs) (Table S2).

An initial clustering step at 20% sequence identity and 80% bidirectional coverage identified 3,502 Antarctic protein sequences grouping either with reference beta-lactamases (2181 sequences) or with negative controls (1321 sequences). To further refine candidate selection, three-dimensional structures were predicted for these Antarctic proteins, as well as for all reference beta-lactamases and negative controls. Structure-based clustering was then performed using strict similarity criteria (LDDT $\geq$ 0.75; TM-score $\geq$ 0.75), enabling discrimination of beta-lactamase orthologs from related proteins with distinct functions.

This integrative filtering yielded a final set of 1916 non-redundant Antarctic beta-lactamase candidates forming 25 clusters with reference beta-lactamases (Figure 2; Table S3), and 692 proteins clustering with TNs. No mixed clusters containing both reference beta-lactamases and negative controls were observed.

Benchmarking and validation of beta-lactamase identification

To evaluate the performance of our integrative approach, we compared the resulting candidate set with two widely used sequence-based methods for beta-lactamase detection: MMSeqs2, based on local sequence alignments, and HMMER, using hidden Markov models (HMMs) derived from beta-lactamase profiles. Both methods were applied to the complete set of predicted proteins from the 81 Antarctic soil metagenomes (see Methods for details).

Using thresholds optimized through Receiver Operating Characteristic (ROC) analyses (Figure S2), MMSeqs2 and HMMER each detected the full set of 1916 structurally supported Antarctic beta-lactamase candidates. At their respective optimal discrimination thresholds (defined by maximum Matthews Correlation Coefficient), MMSeqs2 identified 76,837 additional potential beta-lactamase hits, whereas HMMER identified 80,186.

The union of both sequence-based candidate sets comprised 92,283 non-redundant proteins. When aligned against our structurally curated reference set, including confirmed beta-lactamases and phylogenetically related non-beta-lactamase proteins, 52,348 (56.73%) were classified as putative beta-lactamases, whereas 39,935 (43.27%) grouped with non-beta-lactamase homologs. This proportion indicates that reliance on sequence similarity alone would lead to substantial misclassification of deeply divergent homologs within large protein superfamilies.

Notably, 90.55% of these additional candidates exhibited amino acid identity values below 40% relative to reference beta-lactamases (Figure S2). Within this low-identity range, extensive overlap was observed among proteins from the PBP-like superfamily, including carboxylesterase EstU1-like proteins and class C beta-lactamases. This convergence likely reflects shared evolutionary ancestry and highlights the reduced discriminatory resolution of sequence similarity-based methods when evaluating deeply divergent putative orthologs.

In contrast, the integrative workflow incorporating structural similarity constraints yielded a more conservative set of 1916 non-redundant beta-lactamase candidates, all supported by coherent structural clustering. These findings indicate that combining structural and sequence information refines candidate selection, particularly within evolutionary distance ranges where sequence similarity alone provides limited discrimination.

Furthermore, most additional sequence-based candidates belonged to the PBP-like superfamily (79.79%), including carboxylesterase EstU1-like proteins (29.77%), class C beta-lactamases (26.82%), and class A beta-lactamases (15.83%). Within the <40% identity range, these protein groups occupy overlapping regions of sequence similarity space, consistent with their shared evolutionary origin. This overlap may contribute to the preferential detection of class C-like sequences in previous Antarctic metagenomic surveys [19].

Structural diversity and conserved catalytic features of Antarctic beta-lactamases

The 1,916 putative beta-lactamases (BLAs), defined according to combined sequence and structural criteria, were classified into 986 SBLs and 930 MBLs. Among SBLs, 512 sequences corresponded to Class A, 77 to Class C, and 397 to Class D, with amino acid identities ranging from 42.1% to 100% relative to reference BLDB proteins. Six Class A sequences shared 97.9%–100% identity with known BLDB entries (five TEM variants and CARB-16), and one Class D sequence exhibited 99.6% identity to OXA-455. These high-identity sequences likely represent widely distributed BLA variants that have been reported in both clinical and environmental contexts, including plasmid-associated forms such as TEM-116 [24, 25].

Within MBLs, 263 sequences were assigned to subclass B1, two to subclass B2, and 665 to subclass B3, with sequence identities ranging from 41.3% to 89.5% relative to BLDB entries (Table S3). In parallel, 127 non-beta-lactamase orthologs from the PBP-like superfamily were identified (33 EstU1-like, 86 PBP-5-like, and 8 PBP-A-like),

together with 565 orthologs from the MBL-fold superfamily, including type II glyoxylases, methyl-parathion hydrolases, PQQ-biosynthesis protein B, pyridoxolactonases, and ribonuclease Z-like proteins (Table S4). These distributions further support the capacity of the workflow to discriminate BLAs from closely related proteins within the same structural superfamilies.

To assess whether predicted BLAs retained canonical catalytic features, we examined conserved motifs described for each class across sequences within the 40%–60% identity range. Forty-six out of 51 sequences of class A (40%–50% identity) possessed the S(70)XXK, S(130)XN, and K(234)TG active site motifs, together with E166 (Figure S3). The differences were observed in the catalytic motif K(234)TG, where three sequences displayed the motif RIG, while two sequences exhibited the motif NTG. Also, the six class C sequences (50%–60% identity) comprised the S(64)XXK, Y(150)XN, and K(315)T/SG active site motifs (Figure S4). Furthermore, 49 out of 55 sequences of class D (44%–50% amino acid identity) exhibited the S(70)TFK, S(118)XV, Y/F(144)GN, and K(216)T/SG active site motifs (Figure S5). Differences were mainly identified at valine 120 within the S(118)XV motif, with five sequences showing leucine at position 120 and one sequence showing isoleucine.

For MBLs, 22 of 23 subclass B1 sequences (47%–50% identity) comprised the H(116)-H(118)-H(196) and D(120)-C(221)-H(263) zinc-binding residues (Figure S6), with one sequence having a Tyr instead of a His residue at position 196. Additionally, 26 out of 30 sequences of subclass B3 (40%–50% amino acid identity range) contained the H(116)-H(118)-H(196) and D(120)-H(121)-H(263) zinc-binding residues (Figure S7), where four sequences exhibit a non-conserved substitution at serine 221 (Arg, Met, Gly, or Cys). Finally, in subclass B2, only two sequences were identified in this study, both of which contained the D(120)-C(221)-H(263) zinc-binding residues. Therefore, most identified BLAs, including those with low sequence identity to reference proteins, exhibit the key motifs typical for the respective structural class. At the highest sequence identity thresholds defined for each class, at least 91% conservation was observed across all catalytic residues.

To further evaluate functional coherence, we performed multidimensional function profiling using the AnnoPRO framework [26]. Each protein was represented as a 6109-dimensional probability vector corresponding to Gene Ontology terms (molecular function, cellular compartment, and biological process), and pairwise similarities were calculated using cosine distance, followed by UMAP dimensionality reduction (Figure S8). SBLs formed well-defined clusters distinct from both PBP-like and MBL-fold-like non-beta-lactamase proteins, with clear separation among Classes A, C, and D. Carboxylesterases, PBP-5-like, and PBP-A-like also formed distinct functional clusters, indicating that functional embedding distinguishes SBLs from related PBP-like proteins independently of the sequence and structural features used for initial identification.

In contrast, most members of the broader MBL-fold-like superfamily displayed overlapping functional profiles, except for ribonuclease Z-like proteins, which formed a distinct cluster. This observation underscores the greater functional diversity within the MBL-fold superfamily and highlights the importance of integrating sequence, structural, and functional information to resolve orthologous groups within this protein family accurately.

Together, these analyses demonstrate that Antarctic beta-lactamases encompass substantial class diversity while maintaining conserved catalytic architectures and coherent functional profiles, even at relatively low levels of sequence identity.

Phylogenetic relationships and structural clustering of Antarctic beta-lactamases

We next examined the phylogenetic relationships and structural diversity of Antarctic BLAs relative to reference beta-lactamases and to non-beta-lactamase proteins from the PBP-like and MBL-fold-like superfamilies (TNs). Antarctic BLAs and their closest reference sequences within each structural cluster were clustered at 90% sequence identity and bidirectional coverage to define representative sets. An analogous approach was applied to Antarctic proteins clustering with true negatives.

This strategy yielded representative Antarctic BLAs distributed across classes A (128), B (233; including 53 B1, 1 B2, and 179 B3), C (11), and D (72), together with 23 Antarctic PBP-like proteins and 116 MBL-fold-like non-beta-lactamase proteins. These were analyzed alongside 587 BLDB reference sequences and 14 true-negative representatives. Phylogenetic trees were constructed starting from multiple sequence alignments (MSAs), incorporating structural cluster information derived from the Foldseek-based workflow (Figures 3 and 4).

Regarding the PBP-like superfamily, as expected, reference beta-lactamase sequences from classes A, C, and D formed deeply separated monophyletic clades, including Antarctic orthologs (Figure 3). The highest sequence diversity was observed in class A, followed by class D and then class C. Accordingly, a markedly higher

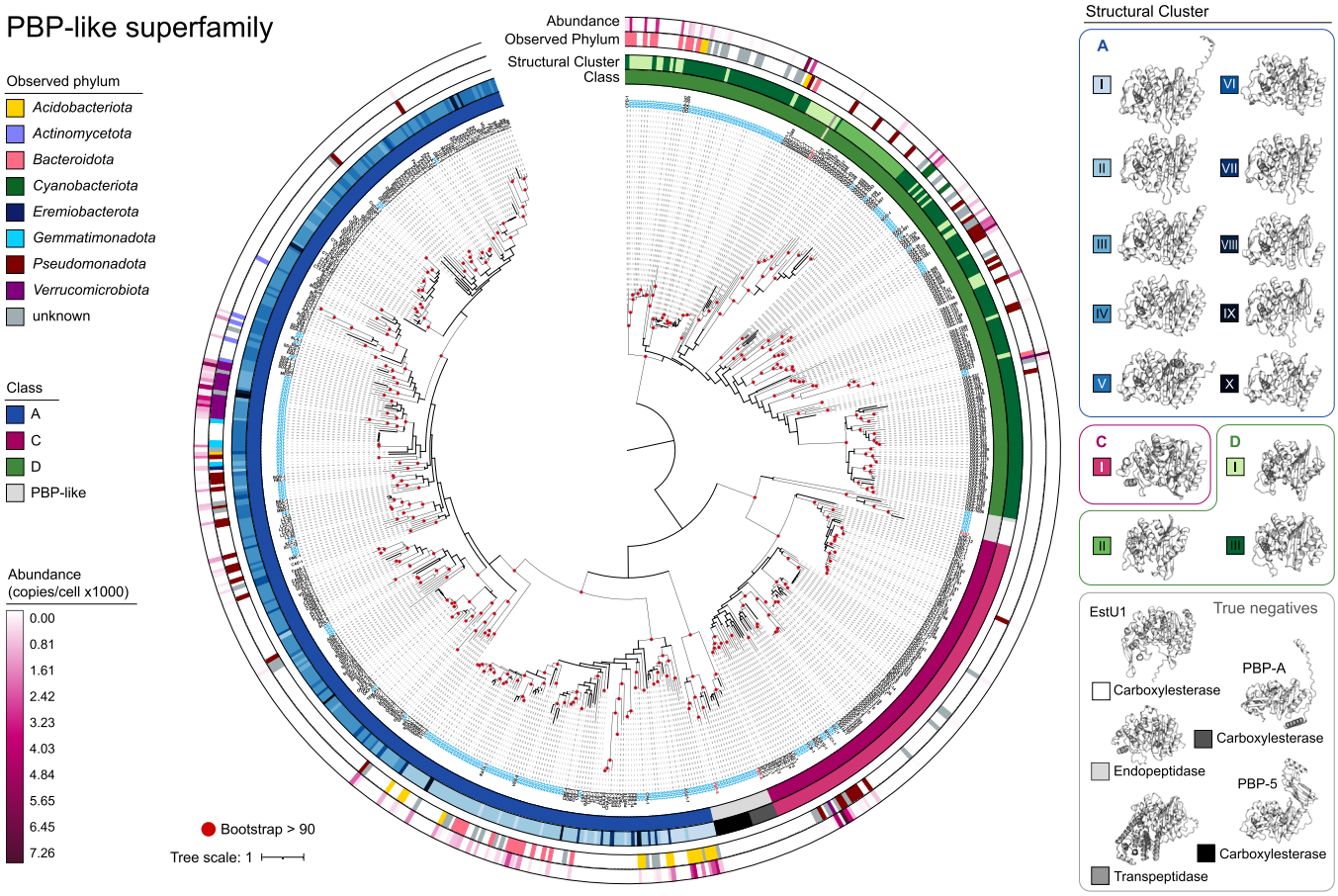


FIGURE 3 Phylogenetic relationships, structural clusters, and abundance patterns of Antarctic PBP-like beta-lactamases. Maximum-likelihood tree showing representative Antarctic class A, C, and D beta-lactamases and related PBP-like superfamily proteins. Red, black, and blue tip labels indicate reference non-beta-lactamase proteins, reference beta-lactamases, and Antarctic beta-lactamase orthologs, respectively. Structural clusters were defined using LDDT ≥ 0.75 and TM-score ≥ 0.75 thresholds, and representative structures are shown for each cluster. Outer rings indicate beta-lactamase class, structural cluster, observed phylum, and cumulative abundance across all sites. Beta-lactamase gene abundance and taxonomic affiliation were inferred from metagenomic reads mapped to each beta-lactamase. Red circles indicate bootstrap support >90.

number of structural clusters was found for class A (10), compared with class D (3) and class C (1). Conversely, all the non-beta-lactamase PBP-like proteins, except for BlaR1 (class D), formed separate clades from beta-lactamases. Regarding BlaR1, this result is consistent with previous phylogenetic analyses within the PBP-like superfamily, indicating that this protein may represent a paralog that retains structural similarity to class D beta-lactamases but has lost hydrolytic activity, evolving to act as a beta-lactam sensor and signal transducer [12].

Inside class A, two major clades were observed. A more numerous one, including most of the references plus some Antarctic sequences, and a smaller one, including a higher proportion of Antarctic sequences. This separation is supported by structural clustering: the minor clade primarily included sequences from structural clusters I, II, and III, whereas the major clade was composed mainly of clusters IV, V, and VI. In class C, only one structural cluster was observed, indicating greater homogeneity within this group. In class D, three structural clusters were observed, with I and III intermingled in the tree and II more separated, all of which included Antarctic orthologs.

For all the classes, some Antarctic sequences were grouped into clades distinct from BLDB references, which could represent novel beta-lactamase families (Figure 3). For class A beta-lactamases, most monophyletic clades consist exclusively of Antarctic sequences, sharing 41.1% to 59.8% sequence identity with their closest references. In contrast, clades that share a common ancestor with known references mostly exhibit <70% identity, except for one clade of four sequences related to RATA-1, which shows 80.92%–86.51% identity. RATA-1 is a class A beta-lactamase identified from *s_Riemerella anatipestifer* isolates, with most of its known variants reported in marine

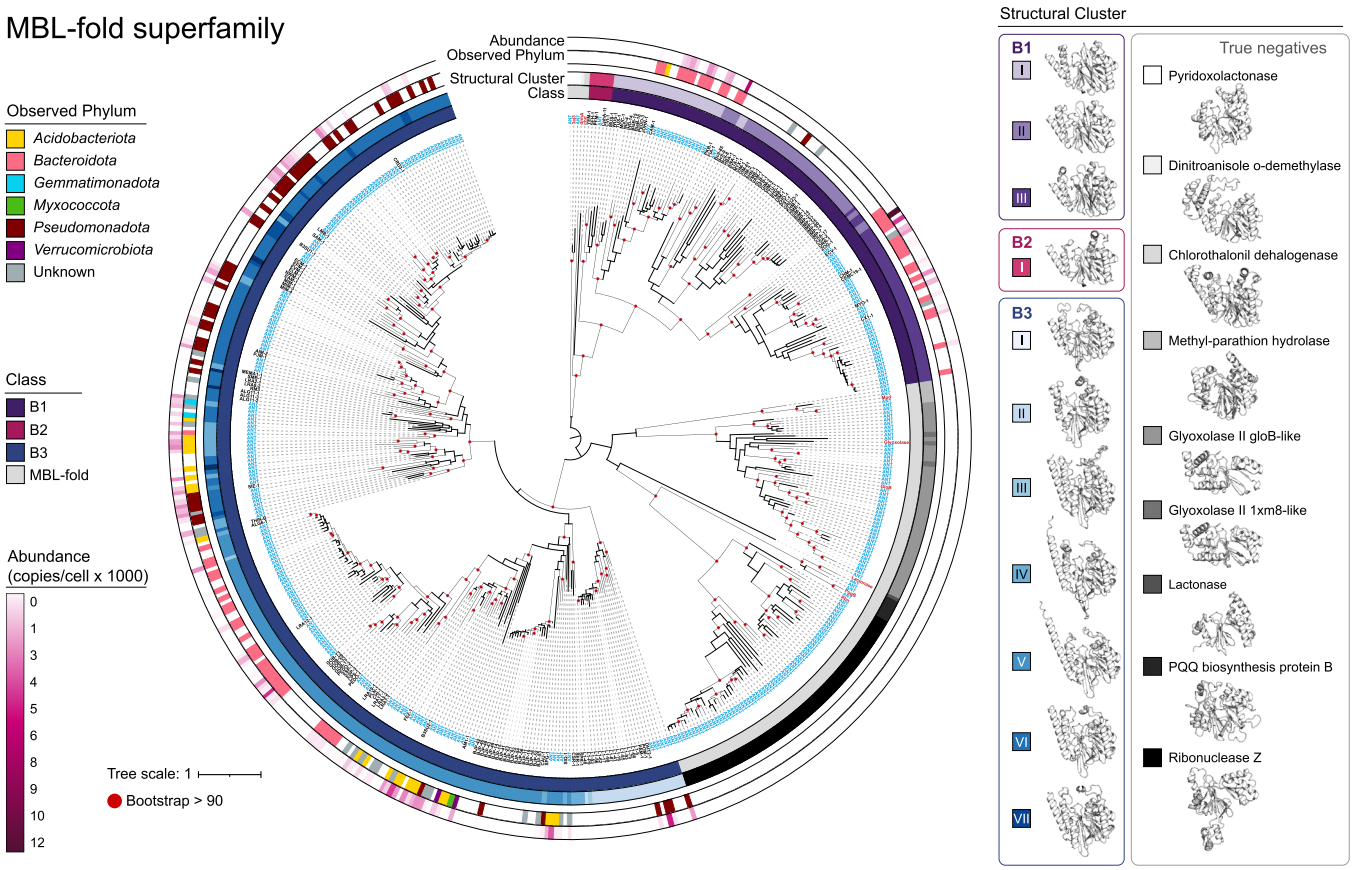


FIGURE 4 Phylogenetic relationships, structural clusters, and abundance patterns of Antarctic MBL-fold beta-lactamases. Maximum-likelihood tree showing representative Antarctic class B beta-lactamases and related MBL-fold superfamily proteins. Red, black, and blue tip labels indicate reference non-beta-lactamase proteins, reference beta-lactamases, and Antarctic beta-lactamase orthologs, respectively. Structural clusters were defined using LDDT ≥ 0.75 and TM-score ≥ 0.75 thresholds, and representative structures are shown for each cluster. Outer rings indicate beta-lactamase subclass, structural cluster, observed phylum, and cumulative abundance across all sites. Beta-lactamase gene abundance and taxonomic affiliation were inferred from metagenomic reads mapped to each beta-lactamase. Red circles indicate bootstrap support >90.

o_Flavobacteriaceae (*p_Bacteroidota*) [27]. This finding indicates its presence in soil-associated environmental bacteria of the same phylum.

For Class C BLAs, the monophyletic clades primarily group with known references, including LRA18-1, IDC-1, MYCC1-1, LRA10-1, and UCB-1. Notably, two Antarctic sequences share at least 83% identity with IDC-1, an integron-derived cephalosporinase previously detected by functional metagenomics in river sediments contaminated with municipal and hospital sewage [28]. Within class D, the largest monophyletic clade, composed solely of Antarctic sequences, comprises 30 BLAs with 48.3%–55.1% identity to the closest references, potentially constituting a novel family. Another clade of four Antarctic sequences clustered with OXA-1091 (70%–74% sequence identity), a beta-lactamase detected in the non-clinical bacterium *s_Bradyrhizobium* sp. BTAi, which exhibits activity against oxacillin and imipenem [29].

Concerning the MBL-fold-like superfamily, all Antarctic sequences and BLDB reference sequences from subclasses B1, B2, and B3 clustered into separate clades compared to the non-beta-lactamase proteins within the superfamily (Figure 4). Subclass B1 beta-lactamases included a large monophyletic clade of 19 Antarctic sequences sharing 48.5%–58% identity with their closest references, corresponding to a putative novel family. A distinct clade of eight sequences is closely related to MYO-1 (70%–88% identity), a beta-lactamase identified in *s_Myroides odoratimimus* able to hydrolyze imipenem [30]. Additionally, we identified novel Antarctic variants of CHM-1 and CEMC19-1 BLAs, and a highly divergent Antarctic BLA sharing a common ancestor with the CAM-1 carbapenemase found in *s_Pseudomonas aeruginosa* [31].

For subclass B3, the largest monophyletic clade contains 28 Antarctic BLAs showing 63.4%–69.4% identity

with their nearest reference (LRA12-1) and belonging to structural cluster IV. Remarkably, LRA12-1 was identified through functional metagenomics in a remote Alaskan soil sample [32]. In addition, we found a clade of 18 Antarctic BLAs related to CRD3-1 (77%–84.19% identity), a BLA previously found in *g_Erythrobacter* species that displays high levels of imipenem-hydrolyzing activity [33]. Additionally, we observed a putative novel family forming a monophyletic clade comprising six sequences that share a common ancestor with SIE-1, first identified in *s_Sphingobium indicum*, which hydrolyzes carbapenems and cephalosporins and exhibits the peculiarity of having a Glu residue in the active site (EHH/DHH active-site motif) [34]. We also observed an additional clade comprising mostly Antarctic BLAs related to FEZ-1 and B3SU2-1, initially identified in environmental *g_Flavobacterium* species [35].

The single Antarctic subclass B2 beta-lactamase was phylogenetically distant from its closest reference (CphA-11), consistent with a divergent lineage.

Overall, phylogenetic and structural analyses reveal extensive diversification of beta-lactamases across Antarctic soils, including multiple monophyletic clades composed exclusively of Antarctic sequences and several lineages sharing moderate similarity with previously described environmental enzymes. These results highlight the broad evolutionary breadth of Antarctic BLAs while reinforcing the structural coherence that underlies their classification.

Biogeographic patterns and taxonomic affiliations of Antarctic beta-lactamases

We analyzed the abundance of BLA genes across the 81 soil metagenomes to determine their geographic distribution. Overall, MBLs accounted for 53% of the total abundance across all the sites, whereas serine beta-lactamases (SBLs) comprised 47%. Subclass B3 was the most abundant group (39.42%), followed by class A (28.09%), class D (14.40%), and subclass B1 (13.41%) (Figure 5A).

Soils from AP exhibited a higher total BLA abundance than CD soils (Figure 5B). At the class and subclass levels, A, C, and B1 followed this same pattern, whereas no significant differences were observed among regions for class D, subclass B2, or subclass B3 (Figure 5C–H).

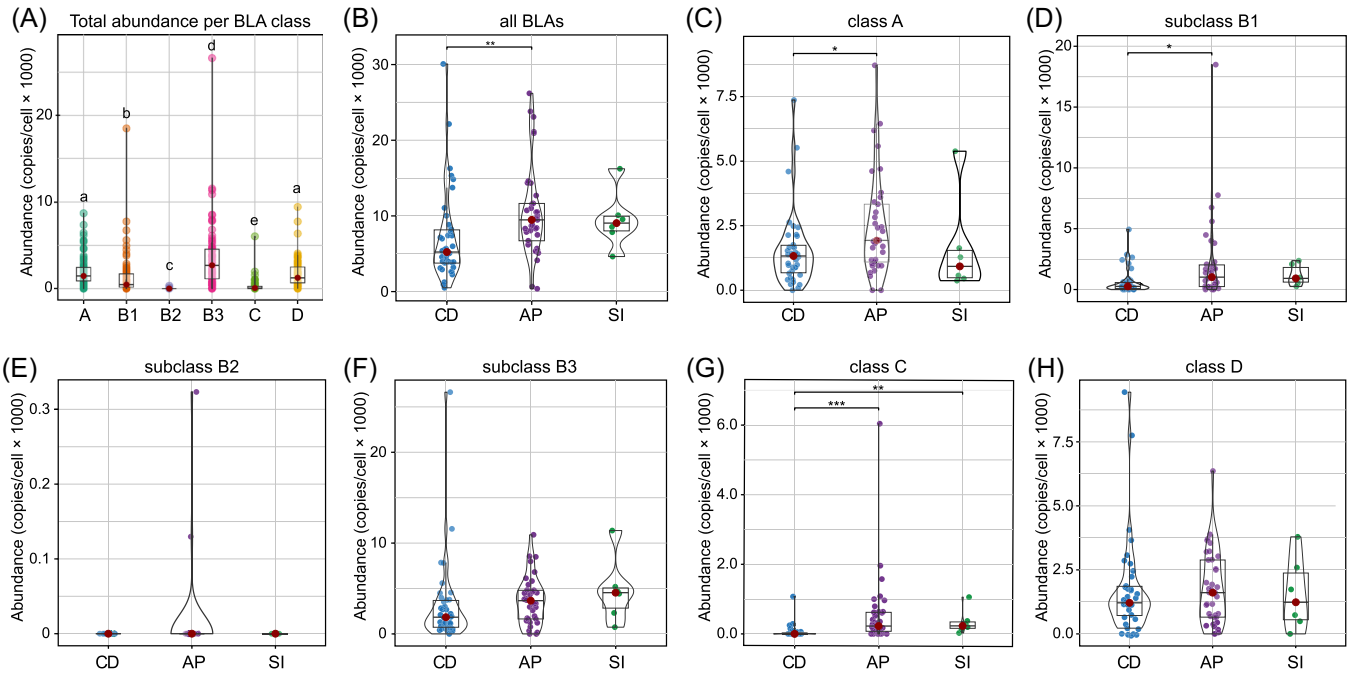


FIGURE 5 Relative abundance of Antarctic beta-lactamases across soil ecotypes. Distribution of beta-lactamase abundance values, expressed as copies per cell × 1000, across cold desert (CD), Antarctic Peninsula and surrounding islands (AP), and Subantarctic Islands (SI) soils. (A) Total abundance by beta-lactamase class and subclass across all Antarctic and Subantarctic sites. Statistical differences were assessed using the non-parametric Friedman test, followed by Durbin–Conover pairwise comparisons; different letters indicate significant differences among groups (pFDR-adjusted < 0.001). (B) Total beta-lactamase abundance by soil ecotype. (C–H) Beta-lactamase abundance by class or subclass across soil ecotypes. For panels (B)–(H), statistical differences were assessed using the Kruskal–Wallis test followed by Dunn's pairwise test. Significance is indicated as *pFDR-adjusted <0.05, **pFDR-adjusted <0.01, and ***pFDR-adjusted <0.001.

To investigate the taxonomic affiliations of Antarctic BLAs, we applied two complementary approaches. First, metagenomic reads mapping to BLA genes were extracted and taxonomically classified using RAT. Then, the BLA gene received the taxonomic classification of the mapping reads using an LCA criterion (Figure 6A).

Among SBLs, class A beta-lactamases were primarily associated with *p_Acidobacteriota* (*c_Blastocatellia*), *p_Verrucomicrobiota* (*c_Verrucomicrobiae*), and *p_Bacteroidota* (*c_Bacteroidia*). Class C sequences were mainly linked with *p_Pseudomonadota* (predominantly *c_Gammaproteobacteria*), and *p_Verrucomicrobiota* (*c_Verrucomicrobiae*), whereas class D BLAs were related to *p_Bacteroidota* (*c_Bacteroidia*), *p_Pseudomonadota* (*c_Alphaproteobacteria*), and *a_Acidobacteriota* (*c_Blastocatellia*) (Figure 6A).

Within MBLs, subclass B1 was predominantly assigned to *p_Bacteroidota* (*c_Bacteroidia*), subclass B2 to sequences of undetermined phyla, and subclass B3 to *p_Pseudomonadota* (*c_Alphaproteobacteria* and *c_Gammaproteobacteria*), *p_Acidobacteriota* (*c_Terriglobia* and *c_Vicinamibacteria*), and *p_Bacteroidota* (*c_Bacteroidia*) (Figure 6A). Together, these results provide a broad overview of the potential taxonomic distribution of beta-lactamases across Antarctic soils.

In a second approach, we examined the predicted taxonomic origins of reconstructed Antarctic BLAs within the phylogenetic trees (Figures 3 and 4), integrating relative abundance information.

For class A, the highest abundance corresponded to sequences from *p_Acidobacteriota* (*f_Pyrinomonadaceae*), forming clades related to ASU1-1, a beta-lactamase previously identified in environmental functional metagenomic libraries [36]. Additional Antarctic BLAs formed clades related to CME-3 and CME-4 within *p_Bacteroidota* (*f_Weeksellaceae*), and to RATA-1 and VEB-9 among *f_Flavobacteriaceae* and related lineages. Notably, a substantial fraction of class A BLAs belonged to monophyletic clades composed exclusively of Antarctic sequences, particularly within *p_Verrucomicrobiota* (*o_Chthoniobacterales*) and diverse *p_Pseudomonadota* genera, consistent with lineage-specific diversification.

For class C, the most abundant Antarctic BLAs were affiliated with *p_Pseudomonadota* (*g_Pseudomonas_E* and *f_Rhodanobacteraceae*) and *p_Verrucomicrobiota* (*f_UBA10450*, *o_Chthoniobacterales*), sharing ancestry with UCB-1, identified in a functional metagenomic library of agricultural and plant-associated soil [37], and LYL-1 from *s_Lysobacter lactamgenus*, a bacterium known for cephalosporin biosynthesis [38].

Within class D, abundant OXA-like proteins were associated mainly with *p_Pseudomonadota* (including *g_Bradyrhizobium* and members of *f_Burkholderiaceae* and *f_Rhodanobacteraceae*), as well as with NOD-1-like enzymes from *p_Bacteroidota* and *p_Cyanobacteriota* lineages. Other abundant class D beta-lactamases were related to CPD-1 from *s_Chitinophaga* sp. (*p_Bacteroidota*, *f_Chitinophagaceae*) and OXA-347 and OXA-209 found in *s_Bacteroides* spp. [29].

Among MBLs, subclass B3 displayed substantial taxonomic breadth. A prominent clade corresponded to highly abundant sequences from *p_Acidobacteriota*, particularly within *o_Vicinamibacterales*, *o_Bryobacterales*, and *o_Acidoferrales*, consistent with previously noted lineage-specific diversification.

We also observed a high abundance of B1 enzymes, mainly from *p_Bacteroidota* (*g_Ferruginibacter* and *f_Cyclobacteriaceae*), sharing ancestry with CAM-1, a

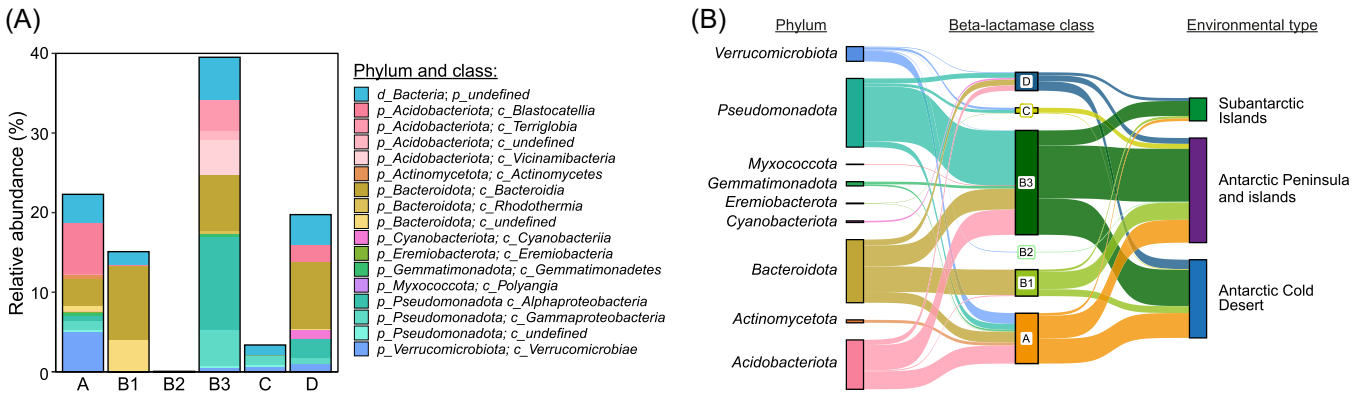


FIGURE 6 Taxonomic distribution of Antarctic beta-lactamases. (A) Taxonomic assignment of beta-lactamase-coding genes based on metagenomic reads mapped to beta-lactamase sequences and resolved using a last common ancestor approach. Bars show total relative abundance across all sites, grouped by beta-lactamase class or subclass and predicted taxonomic affiliation. (B) Sankey plot showing the association between predicted host phyla, beta-lactamase classes or subclasses, and environmental types for beta-lactamases detected in assembled contigs.

carbapenemase found in *s_P. aeruginosa* but proposed to have originated in *p_Bacteroidota* [31, 39]. Also, we identified another BLA from Bacteroidota (*f_Spirosomataceae*; *g_Spirosoma*) related to FIA-1, identified in the genome from *s_Fibrella aestuarina* BUZ 2 and capable of hydrolyzing imipenem [30]. Additionally, we observed a high abundance of Antarctic BLAs (most from *f_Flavobacteriaceae*) sharing ancestry with CHM-1, CEMC19-1, and MYO-1.

The B2 beta-lactamase, distantly related to CphA, was found in very low abundance. In contrast, B3 beta-lactamases were more diverse and abundant, with the representation among *p_Pseudomonadota* (*g_Rhizomicrobium*), *p_Verrucomicrobiota* (*o_Limisphaerales*), and *p_Acidobacteriota* (*o_Terriglobales*, *o_Bryobacterales*, and *o_Acidoferrales*), showing evolutionary relationships with FEZ-1, B3SU2-1, and AM1-1. Enzymes from *p_Bacteroidota* (mostly from *f_Chitinophagaceae*) were also identified in high abundance, displaying similarity to LRA12-1. Furthermore, we observed in high abundance enzymes from *p_Acidobacteriota* (*c_Vicinamibacterales*, *f_Fen-336*, *f_UBA2999*, and *f_SCN-69-37*) and *p_Pseudomonadota* (*f_Woeseiaceae*, *f_Xanthomonadaceae*, and *f_Sphingomonadaceae*) clustered closely to SIE-1 and THIN-B, while others from *p_Gemmatimonadota* (*c_Gemmatimonadetes*) were related to ALG11. Finally, several *p_Pseudomonadota* BLA sequences affiliated with *g_Sphingomicrobium*, distantly related to CRD3-1, were detected in high abundance.

Overall, these analyses reveal that Antarctic beta-lactamases exhibit marked regional variation in abundance and broad taxonomic distribution, encompassing diverse bacterial phyla and including several lineage-specific clades that appear enriched in Antarctic soils. To minimize confounding by unequal sample numbers across ecotypes, we also calculated per-region normalized abundance averages for each beta-lactamase, which revealed ecotype-specific abundance patterns within both the PBP-like and MBL-fold superfamilies (Figures S9 and S10).

Genomic context and host distribution of Antarctic beta-lactamases

To investigate the genomic context of Antarctic BLAs, we analyzed assembled contigs and metagenome-assembled genomes (MAGs) recovered from the 81 metagenomes. In total, 36.8 million contigs were assembled (median N50 = 1143 bp), and binning resulted in 3576 MAGs. MAG taxonomic assignments were largely consistent with taxa identified from unassembled reads and included several poorly described lineages (Table S5).

A total of 470 BLAs were identified in contigs and 175 in MAGs (Figure S11). Consistent with read-based abundance estimates, subclass B3 and class A beta-lactamases were the most frequently detected in both contigs and MAGs.

At the contig level (Figure 6B), class A BLAs were associated with the broadest range of phyla, primarily *p_Acidobacteriota*, *p_Bacteroidota*, and *p_Verrucomicrobiota*. Class C sequences were mainly affiliated with *p_Pseudomonadota* and *p_Verrucomicrobiota*, whereas class D BLAs were associated with *p_Bacteroidota*, *p_Acidobacteriota*, and *p_Pseudomonadota*. Some phylum-specific associations were observed: class A BLAs were uniquely detected in *p_Actinomycetota*, and class D BLAs were uniquely identified in *p_Cyanobacteriota*. Among MBLs, subclass B3 exhibited the broadest phylogenetic distribution, predominantly within *p_Pseudomonadota*, *p_Acidobacteriota*, and *p_Bacteroidota*. Subclass B1 was largely restricted to *p_Bacteroidota*, with occasional representation in *p_Acidobacteriota*, whereas subclass B2 was detected only in *p_Verrucomicrobiota*. Notably, Verrucomicrobiota harbored five of the six BLA classes and subclasses, representing the highest class diversity within a single phylum.

Geographically, contigs encoding class A, class D, and subclasses B1 and B3 BLAs were detected across all three regions. In contrast, class C BLAs were restricted to Antarctic Peninsula and continental desert soils, while subclass B2 was detected only in AP soils, which uniquely contained representatives of all six classes and subclasses.

Within MAGs (Figure 7; Table S5), class A enzymes were distributed across the broadest range of phyla, predominantly *p_Verrucomicrobiota*, *p_Acidobacteriota*, and *p_Bacteroidota*. Class C was restricted to *p_Pseudomonadota* and *p_Verrucomicrobiota*, whereas class D spanned five phyla, with *p_Acidobacteriota*, *p_Pseudomonadota*, and *p_Bacteroidota* most prevalent. Among MBLs, subclass B3 again showed the broadest host range, mainly within *p_Pseudomonadota*, *p_Acidobacteriota*, and *p_Bacteroidota*. Subclass B1 was confined to *p_Bacteroidota*, and subclass B2 was detected only in MAGs not meeting the ≥50% completeness and <10% contamination criteria, and was, therefore, excluded from further analysis. Notably, two *p_Planctomycetota* MAGs encoded class A BLAs, and one *p_Myxococcota* MAG encoded a subclass B3 BLA, although these phyla were not associated with BLAs at the contig level.

Consistent with contig-based analyses, MAGs affiliated with *p_Pseudomonadota*, *p_Bacteroidota*, and *p_Verrucomicrobiota* harbored the highest diversity of BLA classes (four out of six classes each). In two cases, individual MAGs encoded multiple BLAs from different

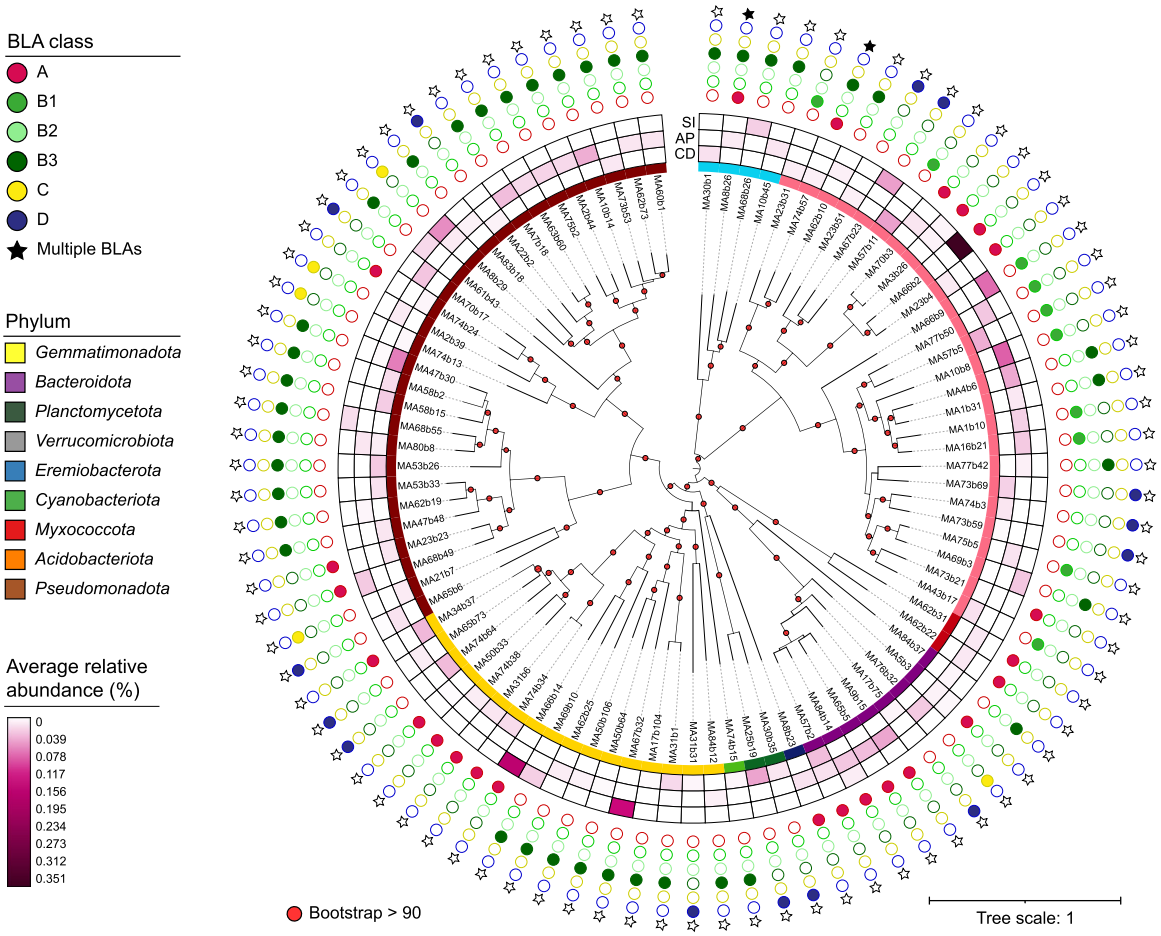


FIGURE 7 Phylogenetic distribution and regional abundance of beta-lactamase-containing metagenome-assembled genomes. Midpoint-rooted maximum-likelihood phylogenetic tree inferred from the multiple sequence alignment of Genome Taxonomy Database (GTDB) bacterial marker genes for MAGs containing beta-lactamase genes. The inner colored ring indicates the MAG phylum. Heatmap rings show the mean relative abundance of each MAG across Subantarctic Islands (SI), Antarctic Peninsula and surrounding islands (AP), and cold desert (CD) soils. Outer symbols indicate the presence of beta-lactamase genes from each class or subclass. Black stars indicate MAGs encoding multiple beta-lactamases. Red circles indicate bootstrap support >90.

structural classes: MA8b26 (*p_Gemmatimonadota*) and MA74b57 (*p_Bacteroidota*), each of which contained both a class A and a subclass B3 enzyme. More broadly, MAGs belonging to the same phylogenetic lineages tended to encode BLAs from the same structural classes, suggesting lineage-associated distribution patterns.

Overall, these results demonstrate that Antarctic beta-lactamases are broadly distributed across assembled genomes and contigs, spanning multiple bacterial phyla and exhibiting consistent host associations across structural classes. Beyond host distribution, understanding the genomic context of these beta-lactamase genes is essential to evaluate their potential mobility. We therefore examined the association of Antarctic BLAs with mobile genetic elements (MGEs) to assess their possible capacity for horizontal transfer.

Antarctic beta-lactamases found associated with MGEs are limited, although likely underestimated due to assembly constraints

Detecting MGEs in metagenomic assemblies remains challenging, particularly when relying on short-read data, due to limited contiguity and fragmentation of assembled sequences. These constraints reduce the ability to resolve full genetic contexts and may limit detection sensitivity for MGEs in environmental metagenomes [40]. Despite these limitations, we used state-of-the-art bioinformatic tools to identify potential MGEs associated with BLA genes, including integrons, prophages, and transposable elements. Putative plasmid sequences were not considered due to the high false-positive rates

reported for plasmid prediction pipelines in metagenomic datasets [41].

We identified one class D OXA-205-like beta-lactamase located within a complete integron in a *p_Pseudomonadota* contig (*f_Burkholderiaceae*), as well as four beta-lactamase genes embedded in CALIN-type integrons (Figure S12). Among these, one corresponded to a class D OXA-1172-like enzyme, whereas the others represented putative novel variants, including a class A enzyme (62.5% identity to GES-41), a class C enzyme (83.5% identity to IDC-1), and a class D enzyme (52% identity to OXA-1089), associated with *p_Pseudomonadota* and *p_Acidobacteriota* lineages.

Analysis of insertion sequence proximity revealed additional associations. A class D BLA gene (47.4% identity against NOD-1) was observed in the vicinity of an IS982 in a *p_Acidobacteriota* contig from *o_Pyrinomonadales*. Additionally, a class C BLA gene (61.2% identity to LRA10-1) was identified near an IS21 in a *p_Pseudomonadota* contig from *c_Alphaproteobacteria*, and a subclass B2 BLA gene (61.9% identity to CVI-1) was observed near an IS200/IS605 element in a *p_Verrucomicrobiota* contig from *o_Chthoniobacterales*.

No associations between beta-lactamase genes and predicted prophages were detected. Overall, only a limited number of BLAs showed proximity to identifiable MGEs in the assembled contigs.

Given the inherent fragmentation of short-read assemblies, these observations likely represent a conservative estimate of MGE associations in the sampled soils. Nevertheless, the detection of integron- and insertion sequence-linked beta-lactamases indicates that at least a subset of Antarctic BLAs reside within genomic contexts compatible with horizontal gene transfer.

3 | DISCUSSION

The emergence of novel beta-lactamases and their rapid dissemination via plasmids and other MGEs represents a critical escalation in the global AMR crisis [42, 43]. While much attention has focused on clinical settings, increasing evidence indicates that environmental microbiomes constitute a vast reservoir of beta-lactamase diversity. In this study, we demonstrate that Antarctic and Subantarctic soil microbiota harbor a broad repertoire of beta-lactamases spanning multiple structural classes, with a predominance of class A serine beta-lactamases and subclass B3 metallo-beta-lactamases.

Structural class distribution analysis revealed that subclass B3 and class A beta-lactamases exhibited both the highest relative abundance (39.43% and 22.3%, respectively) and the largest number of unique sequences ($n = 665$ and $n = 512$). This pattern contrasts with previous analyses of Antarctic microbial mats, where class C beta-lactamases predominated [19], and with studies of cryosol soils reporting higher proportions of class C enzymes [44]. Such discrepancies likely reflect differences in habitat type and microbial community composition, as microbial mats and cryosols exhibit distinct taxonomic structures compared with soil environments [20, 21].

Broader environmental surveys have frequently reported a predominance of class B beta-lactamases outside human-associated microbiomes [45, 46], consistent with the high abundance of subclass B3 enzymes observed here. Together, these comparisons suggest that Antarctic soils align with global environmental trends in beta-lactamase class distribution, while also harboring distinct lineage-specific diversification.

Nonetheless, the abundance of class B beta-lactamases in human-derived samples may be on the rise, as suggested by recent studies [47], in response to the increased clinical use of beta-lactam antibiotics that are not efficiently hydrolyzed by class A enzymes. This is particularly relevant, since it has been reported that class B beta-lactamases that originate from environmental bacteria, especially those of the subclass B3, are capable of hydrolyzing all beta-lactam antibiotics [48]. These findings highlight Antarctic environmental bacteria as a potential rich reservoir of previously uncharacterized beta-lactamase variants, whose hydrolytic activity against clinically relevant beta-lactam antibiotics needs to be studied.

Beyond the genomic and structural characterization, the observed geographical distribution of beta-lactamase classes across Subantarctic islands (SI), Antarctic Peninsula and surrounding islands (AP), and cold desert (CD) soils likely reflects ecological filtering along strong environmental gradients [49]. Subantarctic and maritime Antarctic soils, which generally exhibit higher organic matter and microbial productivity, may foster more intense microbial interactions and secondary metabolite production, thereby maintaining a richer repertoire of natural resistance determinants [5]. In contrast, the nutrient-poor and extremely oligotrophic cold deserts of continental Antarctica host simpler communities shaped by extreme abiotic stress, where beta-lactamases could persist as multifunctional enzymes under relaxed selection [50, 51].

In terms of taxonomic distribution, consistent with previous studies, beta-lactamases were predominantly associated with the phyla *p_Pseudomonadota* and *p_Bacteroidota*, regardless of the enzyme class or soil origin [30, 46]. Notably, we also unveiled an unprecedented diversity of BLAs from members of *p_Acidobacteriota*, being the first large-scale BLA

screening to report such an association, particularly relevant for class A beta-lactamases. Although *p_Acidobacteriota* primarily comprises uncultivable representatives, recent studies have confirmed members of this phylum as ubiquitous and abundant in diverse environments [52], which highlights its potential as a gene dispersal vector. A subclass B2 beta-lactamase (CAM-2) has previously been reported within the chromosome of a MAG affiliated with *f_Pyrinomonadaceae*, recovered from sediments in Shenzhen Bay, China [53]. This enzyme was proposed as a potential ancestor of the clinically disseminated CAM-1 variant found in *s_P. aeruginosa*. In the present study, 57 sequences showing 52%–83% sequence identity to CAM-1 were identified, suggesting that Antarctic environmental *p_Acidobacteriota* may also serve as reservoirs for novel beta-lactamase variants of clinical interest.

Although all the beta-lactamases identified in this study are chromosomal, five were associated with complete or partial integrons, and three were near insertion sequences. Notably, we identified a gene encoding an OXA-205-like BLA in an integron, in agreement with previous reports of this enzyme associated with class 1 integrons in a clinical *s_P. aeruginosa* isolate [54], as well as in Antarctic microbial mats [55], microplastic-associated samples along watersheds, and activated sludge from wastewater treatment plants in China [56, 57]. Additionally, two BLA genes previously associated with clinical or human-impacted environments were identified here for the first time in a pristine, non-human-associated setting: the class D variant OXA-1171/OXA-544, initially described in clinical *s_P. aeruginosa* isolates, and a class C beta-lactamase sharing 83.1% identity with IDC-1, an integron-derived cephalosporinase previously detected only in wastewater samples from Asia [28, 58].

The remaining beta-lactamases exhibited less than 65% sequence identity to reference sequences in the Beta-Lactamase DataBase (BDLB), indicating that they may represent previously uncharacterized families. This highlights the need for further functional characterization and assessment of their potential for horizontal gene transfer, particularly given that many beta-lactamases acquired by clinical pathogens have originated from environmental *p_Pseudomonadota* [30, 59, 60], the same phylum to which the potentially mobile beta-lactamases identified in this study belong.

Studies of the Antarctic resistome consistently indicate that resistance genes in remote, low-human-impact environments are shaped primarily by long-standing ecological interactions rather than recent anthropogenic selection [5, 17, 61]. It has been proposed that Antarctic resistance determinants evolved in response to microbial competition and natural antibiotic production within these ecosystems. In particular, beta-lactamases may function as defense mechanisms against beta-lactam-producing fungi or bacteria present in soil environments [62, 63], supporting the view that environmental beta-lactamases reflect a long coevolutionary history between antibiotic-producing and antibiotic-degrading microorganisms, potentially spanning billions of years [64]. Because metadata regarding the degree of human impact were inconsistently reported across the analyzed datasets, we could not systematically evaluate the effect of local anthropogenic activity on beta-lactamase profiles, and therefore our interpretations primarily reflect broad ecological patterns rather than site-specific disturbance gradients.

The predominance of class A and subclass B3 enzymes observed here is consistent with their deep evolutionary origins among serine beta-lactamases and metallo-beta-lactamases, respectively. Their long-term persistence and structural conservation may contribute to their widespread distribution in low-human-impact environments such as Antarctica, where selective pressures are primarily ecological rather than clinical. Moreover, several class A and subclass B3 variants have been shown to possess relatively broad substrate spectra [11, 48], which could confer adaptive flexibility in heterogeneous soil environments.

Alternatively, the persistence of these enzymes in cold, nutrient-limited ecosystems may reflect ancestral multifunctionality. Promiscuous catalytic activities within the PBP-like and MBL-fold superfamilies can be co-opted for diverse physiological roles under fluctuating selective pressures [65]. For example, serine beta-lactamases share structural features with proteins involved in cell-wall remodeling [66, 67], while metallo-beta-lactamase-fold proteins have been implicated in quorum sensing and other regulatory processes [68, 69]. This perspective is particularly relevant in Antarctic soils, where diverse homoserine lactone-based quorum-sensing systems have been documented as adaptive strategies in extreme environments [70]. Elucidating the native physiological roles of Antarctic beta-lactamases will therefore be essential to understanding their ecological significance.

Advances in next-generation sequencing and structure-informed bioinformatics have greatly expanded the detection of resistance gene homologs in environmental microbiomes [71]. For beta-lactamases in particular, it has been estimated that tens of thousands of variants remain undiscovered [72]. The integrative workflow applied here, combining sequence, structural, and functional similarity, enables the identification of deeply divergent orthologs while minimizing misclassification within large superfamilies. In this context,

notably, metagenomic-metatranscriptomic studies of gut microbiomes have shown that a substantial proportion of detected β-lactam ARGs are not transcriptionally active, indicating that gene presence alone does not necessarily reflect gene expression [73]. Antarctic soils provide a valuable system for exploring the evolution of beta-lactamase architectures under predominantly natural selective regimes.

Finally, several enzymes identified in this study originate from bacteria inhabiting permanently cold environments, where winter temperatures may reach −60°C [74]. As such, they may represent models for studying structural stability and catalytic adaptation under low-temperature conditions. Investigating their biochemical properties could yield insights into the evolutionary constraints and flexibility of beta-lactamase scaffolds in extreme ecosystems.

4 | CONCLUSION

Together, our findings demonstrate that Antarctic and Subantarctic soils constitute geographically structured and phylogenetically diverse reservoirs of beta-lactamases, dominated by class A and subclass B3 enzymes and largely encoded by members of *p_Pseudomonadota* and *p_Bacteroidota*. The integrative sequence–structure–function framework employed here revealed a wide repertoire of previously uncharacterized and deeply divergent beta-lactamases that nonetheless retain conserved catalytic architectures, underscoring the evolutionary persistence of these enzyme families in cold environments. These results support the view that remote ecosystems contribute significantly to global resistome diversity and reflect long-standing ecological and evolutionary processes rather than recent anthropogenic selection. Moreover, the prevalence of novel variants highlights the need for future functional characterization to determine their biochemical properties, substrate spectra, and potential for mobilization. In this context, Antarctic soils provide a valuable natural system for understanding the evolution, diversification, and functional potential of beta-lactamases within the broader One Health framework.

5 | METHODS

Antarctic soil metagenome database construction

This study included 69 metagenomes downloaded from the NCBI Sequence Read Archive (SRA) database as of September 26, 2023, and 12 metagenomes sequenced by our research group. As inclusion criteria, we used all the metagenomes that matched the query "Antarctic soil metagenomes," which included their geographical localization, effectively came from soil samples, and corresponded to shotgun metagenomic sequencing. For replicate sequence files from the same soil sample, we selected the one with the highest sequencing effort. All analyzed metagenomes correspond to soil samples; differences in the level of environmental detail reported in Table S1 reflect variation in the original metadata submitted to NCBI rather than differences in soil type. Because technical and environmental covariates were not consistently available across datasets, they could not be incorporated into the inferential models; therefore, unmodeled covariates may contribute to residual variation in community structure and beta-lactamase abundance estimates. The metagenomes' FASTQ files were downloaded using their SRA run accession and a combination of prefetch and fasterq-dump tools from SRA-toolkit v2.11.3.

Metagenomic reads quality assessment and trimming

Raw Illumina reads were quality-checked and processed using fastp v0.23.4 [75] with default parameters to filter low-quality reads and possible adapters. We also assessed human contamination through mapping all the metagenomes to the human reference genome GRCh38.p14 (accession: GCF_000001405.40) using bwa-mem2 v2.2.1 [76] with default parameters and samtools v1.13 [77] to discard all the mapped reads.

Sequence and microbial diversity analysis

Sequence reads for each metagenome were processed using SingleM pipe v0.18.3 [22] with default parameters, and then summarized using the summarize subcommand to determine operational taxonomic units (OTUs) at the species level, by clustering similar sequences at a 96.66% identity threshold for 43 essential single-copy universal bacterial and archaeal genes (ESCGs). Next, for each ESCG, α-diversity metrics including richness, Shannon indices, Pielou evenness and others related metrics were calculated using the custom Perl script get_indices_from_unifrac_tables.pl, where the input corresponds to the summarized by OTUs output of SingleM (SingleM summarize –unifrac-by-otu). When the above command is executed, 59 tables are generated, one for each of the archaeal/bacterial ESCGs.

These 59 tables are used as input for the custom Perl Script, recovering all α-diversity metrics for each ESCG, which can then be used to generate Boxplots and related statistical tests presented in Figure 1. Only 43 out of 59 ESCGs were kept for further analyses, discarding those genes where the estimated richness value equaled zero or one, and genes whose α-diversity metric estimations fell outside the 95% confidence interval for the observed α-diversity metrics distribution.

The Shannon Index H was calculated standardly according to the following formula:

$$H = -\sum_{x \epsilon X} p(x) \quad \ln p(x),$$

where x corresponds to each observed OTU for the ESCG X, $p(x)$ corresponds to the relative abundance of OTU x, and the log base is e.

The Pielou Index J is calculated standardly according to the following formula:

$$J = \frac{H}{\ln(R)},$$

where H corresponds to the Shannon Index, and R corresponds to the Richness for ESCG X, which corresponds to the number of distinct OTUs × observed for ESCG X.

Next, for each α-diversity metric (richness, Pielou index, and Shannon index), we implemented a multivariate non-parametric rank-based analysis using the mv.Csample.test function from the MNM R package [78], to compare α-diversity metric distributions between environmental types. This approach considers the α-diversity metrics of all 43 ESCGs simultaneously as a single multivariate vector for each metagenome. We employed outer standardization, spatial rank scores, and a permutation-based p-value estimation (1000 permutations) (score = "r", stand = "o", method = "p") to rigorously account for the non-normal distribution and potential outliers inherent in microbial diversity data. This was followed by post-hoc pairwise comparisons with FDR p-value adjustment. Then, for visualization purposes, we implemented the ggstatsplot package [79] in R to plot α-diversity metrics. In these plots, each data point corresponds to the mean value across all 43 ESCGs α-diversity metric estimations for each metagenome ($n = 1$ per site), ensuring the visualization is consistent with the multivariate statistical design and free from pseudoreplication.

Additionally, SingleM results were used to determine the taxonomic profiles of the observed communities in each soil sample based on the GTDB r220 taxonomy database [80]. These profiles were used to estimate beta-diversity (Bray–Curtis dissimilarity) at the family level among all the metagenomes and assess possible associations with the environmental type using the vegan [81] and pairwiseAdonis (https://github.com/pmartinezarbizu/pairwiseAdonis) packages in a custom R script. The family level was chosen to maximize the proportion of reads assigned to any taxa (70.80%–94.82%).

Next, we complemented the NMDS 2D projection and corresponding PERMANOVA, with a permutation-based multivariate homogeneity of group dispersions test, to estimate if the average distances of group members to the spatial median of the group, used as group centroid, in multivariate space are different between groups. This complementary analysis was performed using the betadisper function from the vegan R package [81], considering the bias correction for unequal number of samples (type = "median," bias.adjust = TRUE), followed by a permutation test (permutations = 999) to test the null hypothesis of homogeneity of multivariate dispersions, and post-hoc pairwise comparisons with FDR-adjusted p-values. Distances to centroids distributions for each ecotype were plotted as a boxplot, using the ggstatsplot [79] package in R, presented in Figure S1.

NonPareil v3.401 [82] was used in kmer mode to estimate metagenomic coverage and sequencing effort for each sample. MASH v2.3 [83] was used to estimate k-mer distance between metagenomic samples, using the options "sketch," "dist," and default parameters. Figures were plotted using the ggplot2 and ggstatsplot [79] packages in R.

Metagenome assembly and protein prediction

Metagenomic trimmed reads were assembled into contigs using metaSPAdes v3.15.5 [84] with default parameters. These contigs were processed using Prodigal v2.6.3 [85] in meta mode to translate the nucleotide sequences from all metagenomic samples. Additionally, protein sequences were directly assembled from trimmed reads using PLASS [23] in assemble mode with default parameters, adjusting the –min-length to 30 or 40 depending on the read length of the samples (100 bp or <150 bp, respectively). Only complete protein sequences identified by Prodigal and PLASS were retained for subsequent analyses. Partial protein sequences assembled by PLASS were excluded using the custom Perl script plass_orf_filter.pl.

Beta-lactamase protein sequence database construction and identification of (sub) Antarctic beta-lactamases

On October 5th, 2023, 7566 BLA protein sequences were downloaded from the Beta-Lactamase database BLDB [86],

preserving the protein family and enzyme class associated with each. Additionally, six PBP-fold superfamily and 18 MBL-fold-like superfamily non-beta-lactamase protein sequences were added as true negatives (TN DB, Table S3).

BLDB protein amino acid lengths were determined, and standard statistical metrics were calculated for the obtained length distribution (Q0 = 95, Q1 = 284, Q2 = 313, Q3 = 377, Q4 = 428, arithmetic mean = 329.336, standard deviation = 52.4, mode = 377). Based on the length distribution of BLDB proteins, PLASS-assembled and Prodigal-predicted metagenomic proteins were filtered by amino acid length (minimum length = 100 aa, maximum length = 500 aa) using the custom Perl script filter_by_length.pl. After applying this size filter, 98,161,403 PLASS proteins (PLASS DB) and 9,024,650 Prodigal (Meta DB) proteins were kept. Proteins from each database (BLDB, PLASS DB, and Meta DB) were redundancy filtered separately with MMSeqs2 v15.6f452 [87] in Linclust mode [88] using sequence identity ≥0.99 and bidirectional coverage ≥0.99 thresholds (--alignment-mode 3 --cov-mode 0 --cluster-mode 2 --realign), only keeping the representative protein sequences for further steps.

All 65,532,405 redundancy-filtered protein sequences (5458 from BLDB, 58,058,386 from PLASS DB, 7,468,537 from Meta DB, and all 24 from TN DB) were redundancy-filtered as a single set following the same procedure described previously. Then, the resulting 62,047,905 protein set was clustered using low stringency thresholds (sequence identity ≥ 0.2 and bidirectional coverage ≥ 0.8) with MMSeqs2, as described previously. All 267 low-stringency clusters containing BLDB or TN DB proteins were kept for further analysis, using the custom Perl script get_relevant_clusters_list_v2.pl. Eighty-one clusters also contained 3502 Antarctic proteins (PLASS DB or Meta DB).

All 3502 relevant Antarctic proteins selected in the previous step and 5482 redundancy-filtered reference sequences (5458 BLDB proteins and 24 TN DB proteins) were used as input for the protein structure predictor ESMFold v1.0.3 (ESM-2 model) [89]. Next, Foldseek v8.ef4e960 [90] was used to cluster all 8984 predicted protein structure models with highly stringent global and local structural similarity thresholds (TM-Score ≥ 0.75 and LDDT ≥ 0.75) while keeping the lax sequence identity thresholds of the previous MMSeqs2 clustering (sequence identity ≥ 0.2 and bidirectional coverage ≥ 0.8) (--cov-mode 0 -e 1 --cluster-mode 2 --cluster-reassign -s 7.5 --tmalign-hit-order 3 --tmalign-fast 0 --sort-by-structure-bits 1). Then, from the 70 structural clusters obtained, 34 contained both reference and Antarctic proteins (25 clusters contained BLDB reference proteins, while 9 contained TN reference proteins).

All 3025 Antarctic protein sequences in 34 relevant structural similarity clusters were considered putative beta-lactamases (2174) or TN non-beta-lactamase proteins (851). Next, the enzyme class from reference proteins in each structural cluster was inherited to putative Antarctic BLAs and TNs (no cluster contained more than one enzyme class). Similarly, redundant protein sequences from previous steps (≥0.99 sequence identity and ≥0.99 bidirectional coverage clusters) were assigned to structural clusters by reverse inheritance. Finally, the 3025 selected Antarctic proteins and all 7590 reference proteins were clustered iteratively using DIAMOND BLASTP v2.1.9 [91] in ultra-sensitive mode until each protein belonged to a unique single-linkage cluster (99% identity and 99% bidirectional coverage thresholds). The 4892 representatives of these clusters were selected as the final Antarctic BLA set (1916 proteins), Antarctic TN set (692 proteins), reference BLA set (2260 proteins), and reference true negatives set (24 proteins).

Multiple amino acid sequence alignment and phylogenetic analyses

For each of the 34 previously mentioned relevant structural clusters and five more structural clusters containing only reference TNs, the final protein dataset corresponding to each of those clusters was redundancy filtered with MMSeqs2 as previously described using sequence identity ≥0.9 and bidirectional coverage ≥0.9 thresholds. Then, the obtained 1184 non-redundant protein dataset was divided into two groups: the corresponding PBP-like (677 proteins) and MBL-like (507 proteins). Next, the protein sequences in each group were aligned using MAFFT v7.505 [92] in globalpair mode with 1000 iterations, followed by trimming of constant and uninformative sites based on the alninfo analysis performed by IQ-TREE v2.2.5 [93], using the custom Perl script make_informative_msa.pl. Trimmed alignments were then used as input for maximum likelihood tree inference using IQ-TREE (--seed 139808 --bnni --allnni) with 1000 ultrafast bootstraps [94] and the ModelFinderPlus algorithm [95] to select the best amino acid substitution model for each tree according to the BIC criterion, Q.pfam +R10 for the PBP-like superfamily tree and WAG + F + R9 for the MBL-like superfamily tree. Phylogenetic trees visualization was performed using iTOL v7.0 [96].

Beta-lactamases structural alignment and visualization

All the predicted structures for the 1184 non-redundant proteins, which were used as input for phylogenetic trees

inference, were structurally aligned using US-align v20240319 [97] (-mol prot -mm 4 -ter 2). Structure PDB files were imported into PyMOL v3.1.3 (Schrodinger, LLC), then signal peptides detected by SignalP v6.0 [98] were removed. Finally, structure visualization cartoons were made in ray_trace_mode 3.

Contig binning and reconstruction of metagenome-assembled genomes (MAGs)

The trimmed reads were mapped to their respective contigs in all the samples using bwa-mem2 for metagenomic contig binning. The SAM files were then filtered and sorted with samtools and processed with the jgi_summarize_bam_contig_depths command to obtain per-sample contig coverage depth. These files were required as input for binning the contigs using MetaBAT2 v2.15 [99]. Using the default parameters, 3576 bins were obtained from these metagenomic samples. All the bins were quality-checked using CheckM2 v1.0.1 [100]. Only MAGs with ≥50% completeness and <10% contamination (MiMAG criteria) [101] were retained for phylogenetic and taxonomic analyses. Under this parameter, we obtained 1523 MAGs.

Taxonomic assignment of MAGs and metagenomic contigs

Previous BLA representatives were used to determine which contigs and MAGs contained these enzymes using DIAMOND BLASTX with default parameters. The results were filtered using a custom R script that required a minimum 70% identity and 80% coverage match, yielding 547 contigs and 90 MAGs containing BLAs. Taxonomic assignment of contigs containing BLAs was performed using the Contig Annotator Tool CAT v6.0 [102] in sensitive mode supplemented with the GTDB r220 database [80]. The taxonomy and BLA class information for all the contigs were summarized in a Sankey diagram plotted in R using the ggsankey package (https://github.com/davidsjoberg/ggsankey). High- and medium-quality MAGs were also taxonomically identified using GTDB-tk v2.3.0 in classify_wf mode with the GTDB r220 database. The MSA output of this workflow was used as input to IQ-TREE in autodetection mode (--seed 314716, best model: LG + F + R6, bootstrap: 1000) for maximum likelihood tree inference of the MAGs comprising BLAs. Tree visualization and editing were performed with ITOL.

Additionally, we mapped metagenomic reads back to BLA-containing MAGs to calculate their average regional relative abundance (percentage of mapped reads), only considering alignments where at least 75% of the read was mapped to the reference sequence with 95% identity or more, following the recommendations of Poff et al. [103], using CoverM v0.7.0 [104] (-p bwa-mem2 --min-read-aligned-percent 75 --min-read-percent-identity 95 --proper-pairs-only --exclude-supplementary -m relative_abundance).

Beta-lactamase gene taxonomic assignment and relative abundance calculation

The final set of 4176 BLA sequences was divided into six groups according to structural class (A, B1, B2, B3, C, and D). For each group, a MSA was generated using MAFFT as described previously, and profile hidden Markov models (HMMs) were constructed using HMMER v3.2.1 [105]. All BLAs of the respective class identified in taxonomically assigned contigs were subsequently realigned into their corresponding HMM to generate refined reference MSAs.

Then, each class-specific MSA and HMM was used as input for SingleM seqs v0.18.3 [22] to find the most conserved 20 amino acid region within each class profile. The resulting HMMs were then used to build GraftM packages using GraftM v0.15.1 [106], and these were incorporated into class-specific SingleM packages.

The six SingleM packages (one per BLA class) were used to identify BLA-containing sequencing reads in each metagenome using SingleM in pipe mode. This class-specific conserved-region strategy enabled reliable detection of BLA-associated reads independently of contig length, thereby minimizing biases introduced by short or fragmented assemblies.

All identified BLA-containing reads for each metagenome were taxonomically assigned using the Read Annotator Tool RAT v6.0 in sensitive mode [107], supplemented with the GTDB r220 database and previously assigned contig-level taxonomy obtained using CAT. For each read, the specific observed BLA sequence was determined using MMSeqs2 v15.6f452 in easy-search mode (-s 7 --alignment-mode 3) against the curated BLA database (4,176 proteins) and the TN database (716 proteins). Stringent parameters previously recommended for ARG detection [108] were applied (e-value ≤ 1e-7, identity ≥ 80%, query coverage ≥ 75%, and alignment length of at least 25 amino acids). Reads failing to meet these criteria were discarded.

For each BLA gene, all taxonomic assignments derived from mapped reads were integrated using a hierarchical Last Common Ancestor (LCA) resolution procedure implemented in the custom script

get_abundance_and_taxonomy.pl. Taxonomic reconciliation proceeded sequentially from the domain to the species level. At each rank (domain, phylum, class, order, family, genus, species), assignments were compared across reads; the process stopped at the highest rank showing no conflict among observed classifications. Missing taxonomic ranks were not treated as conflicts. This multi-step consensus approach ensured conservative and robust taxonomic assignment even when BLAs were encoded on short or fragmented contigs.

The relative abundance of each BLA (copies/cell) in all metagenomes was determined according to the following equation [109]:

$$\text{Abundance} = \sum_{i=1}^{n}\left(K \times \frac{\frac{Ni\ BLAi\ read \times Li\ read}{Li\ BLAi\ reference\ sequence}}{Ncell\ number}\right)$$

where "*Ni BLAi read*" is the number of reads coding for a specific BLA, "*Li read*" is the read length, "*Li BLAireference sequence*" is the length of the nucleotide sequence of the BLA, *Ncell number* corresponds to the estimated number of cells in the metagenome, and K is a normalization parameter for multi-gene antibiotic resistance mechanisms (for BLAs, $K = 1$). The number of cells in each metagenome was estimated using the SingleM microbial_fraction subcommand [110] as the quotient between the number of prokaryotic nucleotide bases and the average prokaryotic genome size. These abundance values should therefore be interpreted as standardized metagenome-level estimates rather than direct measurements with explicitly propagated uncertainty.

Next, we calculated per-region normalized abundance averages (expressed as copies per cell ×1000) for each individual beta-lactamase.

Functional prediction and similarity clustering of beta-lactamases and related proteins

For each of the 4982 proteins in the database, functional prediction annotation based on Gene Ontology (GO) terms was performed using AnnoPRO v0.2-alpha+2.g547770e [26]. Then, functional similarity between proteins was determined by calculating the cosine distance between GO-term probability vectors [111] using the custom Perl script make_annopro_matrices.pl. Additionally, Uniform Manifold Approximation and Projection (UMAP) [112] was performed using all 4982 GO-term vectors to determine functional similarity groups between proteins and plot results (n_components = 2, random_state = 42, n_neighbors = 100, min_dist = 0.1, metric ='cosine').

Performance evaluation of the bioinformatic workflow for structure-guided identification of BLA from metagenomic data

We considered using two standard processes to search for BLA orthologs: MMSeqs2, using local alignments, and HMMER to perform searches using hidden Markov Models (HMMs) of BLA profiles. First, for MMSeqs2 v15.6f452, we used all complete PLASS proteins and all complete translated ORFs observed in the contigs assembled by metaSPAdes as queries to search against all BLDB targets (easy-search mode –alignment-mode 3) and then retrieved all 413,632 best hits according to bitscore without any filtering.

Next, for the HMMER v3.2.1 implementation, we first split BLDB into 6 fasta files, one for each BLA class (A, B1, B2, B3, C, and D). We inferred a MSA for each of them with MAFFT v7.505 (--globalpair –maxiterate 1000), then constructed HMMs using hmmbuild and hmmpress, and searched for BLA orthologs in all queries described above using hmmscan against all built HMMs. All outputs were concatenated, and for each query, only the best hit according to the full-sequence HMMER score was retained. A total of 422,398 hits were observed using this method.

Then, assuming our final database of 1916 candidate Antarctic BLAs and 692 Antarctic True Negatives candidates (not-BLA proteins from PBP-like and MBL-like superfamilies) obtained by the structural alignment-based workflow as a high-confidence reference set to test the boundaries of search algorithms' sensitivity, we evaluated the performance of both methods, MMSeqs2 and HMMER, using successive thresholds of transformed e-values, according to the following formula, where T corresponds to the transformed score and E to the *e*-value: $T = -\ln(E)$.

For each threshold, performance metrics including sensitivity, specificity, precision, false discovery rate, and Matthews Correlation Coefficient (MCC) were calculated, and the corresponding Receiver Operating Characteristic (ROC) curves were plotted using the custom Perl scripts get_ROC_table_hmm.pl and get_ROC_table_mmseqs.pl. For all thresholds, all 1916 candidate Antarctic BLAs were always observed by both methods. Next, optimal thresholds for each method were determined according to MCC scores.

For MMSeqs2, the optimal e-value corresponded to 4.96e-15 (MCC = 1, FDR = 0), allowing the detection of 76,837 new potential BLAs not included in our set; and for HMMER, the optimal e-value corresponded to 2.99e-22 (MCC = 1, FDR = 0), allowing the detection of 80,186 potential Antarctic BLAs.

Next, we combined both lists of potential new Antarctic BLAs into a non-redundant set of 92,283 proteins (64,740 detected by both methods, 15,446 by HMMER alone, and 12,097 by MMSeqs2 alone). This set was locally aligned against our high-confidence reference set, supplemented with all BLDB reference BLAs and the 24 True Negative reference proteins using MMSeqs2 (easy-search mode –alignment-mode 3). Best hits were selected according to bitscore and quantified.

MGE analysis

All contigs coding for BLAs were used as input for predicting complete integrons, clusters of attC sites lacking integron-integrases (CALIN), and integron integrases using IntegronFinder v2.0.5 [113]. Integrated prophages were predicted using the PHASTEST [114] web server (accessed November 26th, 2024) and Phigaro v2.4.0 [115]. Complete and partial insertion sequences were predicted using ISEScan v 1.7.2.3 [116]. Predicted ISs located ≤5000 bp upstream or downstream from BLA genes were considered relevant hits. MGEs' genomic features were identified and annotated using Bakta v1.10.1 [117] in meta mode using the full Bakta database v5.1 (https://zenodo.org/records/10522951).

AUTHOR CONTRIBUTIONS

José Coche-Miranda: Conceptualization; methodology; software; formal analysis; investigation; data curation; visualization; writing—original draft; writing—review and editing. **Patricio Arros**: Methodology; software; formal analysis; investigation; data curation; visualization; writing—review and editing. **Nicolás Canales**: Investigation; data curation; writing—review and editing. **Camilo Berríos-Pastén**: Investigation; data curation; writing—review and editing. **Gastón Azziz**: Resources; data curation; writing— review and editing. **Rosalba Lagos**: Supervision; writing—review and editing. **Julieta Orlando**: Resources; supervision; writing—review and editing. **Francisco P. Chávez**: Conceptualization; supervision; funding acquisition; writing—review and editing. **Andrés E. Marcoleta**: Conceptualization; methodology; supervision; project administration, funding acquisition; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

We would like to sincerely acknowledge Dr. Silvia Batista, Dr. Matías Giménez, and Josefina Vera (Montevideo, Uruguay) for providing us with access to three Antarctic soil metagenomes and for their very insightful discussions regarding Antarctic beta-lactamases. Computational resources were provided by the supercomputing infrastructure of Soroban (SATREPS MACH – JPM/JSA1705) at Centro de Modelación y Computación Científica at Universidad de La Frontera, Chile. JO acknowledges the support of the ICM-ANID project ICN2021_002. A previous version of this work was posted as a preprint at the bioRxiv server (https://doi.org/10.1101/2025.09.30.679653). This work was funded by Agencia Nacional de Investigación y Desarrollo ANID (Chile), grants FONDECYT 1221193, and Anillo mBioClim ACT210044. P. Arros was supported by the María Ghilardi Venegas Foundation (Chile). J. Coche-Miranda was supported by Agencia Nacional de Investigación y Desarrollo ANID (Chile), Beca Doctorado Nacional 2024 21241736.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in NCBI at https://www.ncbi.nlm.nih.gov/, reference number PRJNA1425240. The metagenomic data used in this study come from two sources. 69 of them were downloaded from the SRA archive (NCBI database). The rest of the metagenomes were generated from Antarctic soil samples acquired by the research groups involved in this study, and the reads were deposited in the NCBI database under the BioProject number PRJNA1425240 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1425240). The accession numbers and relevant metadata are included in Table S1. The Antarctic beta-lactamase and True Negatives collection generated by this research is available at https://github.com/mudamaid62/Coche_2026_BLA_database. All custom Perl, Python, and R scripts, and raw data used for this research are available in the following Github repository: https://github.com/mudamaid62/Coche2026imetaomics. Supplementary materials (figures, tables, graphical abstract, slides, videos, Chinese translated version, and updated materials) may be found in the online DOI or iMetaOmics http://www.imeta.science/imetaomics/.

ETHICS STATEMENT

No animals or humans were involved in this study.

ORCID

Francisco P. Chávez https://orcid.org/0000-0002-6838-208X

Andrés E. Marcoleta https://orcid.org/0000-0003-2950-7510

REFERENCES

1. Naghavi, Mohsen, Stein Emil Vollset, Kevin S. Ikuta, Lucien R. Swetschinski, Authia P. Gray, Eve E. Wool,

Gisela Robles Aguilar, et al. 2024. "Global Burden of Bacterial Antimicrobial Resistance 1990–2021: A Systematic Analysis With Forecasts to 2050." *The Lancet* 404: 1199–226. https://doi.org/10.1016/S0140-6736(24)01867-1

2. Larsson, D. G. J., W. H. Gaze, R. Laxminarayan, and EAMR Topp. 2023. "AMR, One Health and the Environment." *Nature Microbiology* 8: 754–5. https://doi.org/10.1038/s41564-023-01351-9
3. D'Costa, Vanessa M., Christine E. King, Lindsay Kalan, Mariya Morar, Wilson W. L. Sung, Carsten Schwarz, Duane Froese, et al. 2011. "Antibiotic Resistance Is Ancient." *Nature* 477: 457–61. https://doi.org/10.1038/nature10388
4. Agudo, Rubén, and M Paloma Reche. 2024. "Revealing Antibiotic Resistance's Ancient Roots: Insights From Pristine Ecosystems." *Frontiers in Microbiology* 15: 1445155. https://doi.org/10.3389/fmicb.2024.1445155
5. Van, Goethem, Marc W Rian Pierneef, Oliver KIB. ezuidt, Yves Van De Peer, Don A. Cowan, and Thulani P. Makhalanyane. 2018. "A Reservoir of 'Historical' Antibiotic Resistance Genes in Remote Pristine Antarctic Soils." *Microbiome* 6: 40. https://doi.org/10.1186/s40168-018-0424-5
6. de Souza, Joyce, Alexandre Zanatta Vieira, Hellen Geremias Dos Santos, and Helisson Faoro. 2024. "Potential Involvement of Beta-Lactamase Homologous Proteins in Resistance to Beta-Lactam Antibiotics in Gram-Negative Bacteria of the ESKAPEE Group." *BMC Genomics* 25: 508. https://doi.org/10.1186/s12864-024-10410-2
7. Zeng, Ximin, and Jun Lin. 2013. "Beta-Lactamase Induction and Cell Wall Metabolism in Gram-Negative Bacteria." *Frontiers in Microbiology* 4: 128. https://doi.org/10.3389/fmicb.2013.00128
8. Mohammad, Shahid, Anuradha Singh, and Hiba Sami. 2022. *Beta-Lactam Resistance in Gram-negative Bacteria: Threats and Challenges*. Springer.
9. Mora-Ochomogo, Montserrat, and Christopher T. Lohans. 2021. "β-Lactam Antibiotic Targets and Resistance Mechanisms: From Covalent Inhibitors to Substrates." *RSC Medicinal Chemistry* 12: 1623–39. https://doi.org/10.1039/d1md00200g
10. Akhtar, Anees, Nazish Fatima, and Haris M. Khan. 2022. "Beta-Lactamases and Their Classification: An Overview." In *Beta-Lactam Resistance in Gram-Negative Bacteria: Threats and Challenges*, 25–33. https://doi.org/10.1007/978-981-16-9097-6_3
11. Bush, Karen, and George A. Jacoby. 2010. "Updated Functional Classification of β-Lactamases." *Antimicrobial Agents and Chemotherapy* 54: 969–76. https://doi.org/10.1128/AAC.01009-09
12. Fröhlich, Christopher, John Z. Chen, Sevan Gholipour, Ayse N. Erdogan, and Nobuhiko Tokuriki. 2021. "Evolution of β-Lactamases and Enzyme Promiscuity." *Protein Engineering, Design and Selection* 34: gzab013. https://doi.org/10.1093/protein/gzab013
13. Zapun, André, Carlos Contreras-Martel, and Thierry Vernet. 2008. "Penicillin-Binding Proteins and β-Lactam Resistance." *FEMS Microbiology Reviews* 32: 361–85. https://doi.org/10.1111/j.1574-6976.2007.00095.x
14. Bush, Karen. 2018. "Past and Present Perspectives on β-lactamases." *Antimicrobial Agents and Chemotherapy* 62. https://doi.org/10.1128/AAC.01076-18
15. Allen, Heather K., Luke A. Moe, Jitsupang Rodbumrer, Andra Gaarder, and Jo Handelsman. 2009. "Functional Metagenomics Reveals Diverse β-Lactamases in a Remote Alaskan Soil." *The ISME Journal* 3: 243–51. https://doi.org/10.1038/ismej.2008.86
16. Bhullar, Kirandeep, Nicholas Waglechner, Andrew Pawlowski, Kalinka Koteva, Eric D. Banks, Michael D. Johnston, Hazel A. Barton, and Gerard D. Wright. 2012. "Antibiotic Resistance Is Prevalent in an Isolated Cave Microbiome." *PLoS One* 7: e34953. https://doi.org/10.1371/journal.pone.0034953
17. Marcoleta, Andrés E, Patricio Arros, Macarena A. Varas, José Costa, Johanna Rojas-Salgado, Camilo Berríos-Pastén, Sofía Tapia-Fuentes, et al. 2022. "The Highly Diverse Antarctic Peninsula Soil Microbiota As a Source of Novel Resistance Genes." *Science of The Total Environment* 810: 152003. https://doi.org/10.1016/j.scitotenv.2021.152003
18. Santos, Andrés, Burgos Felipe, Jaime Martinez-Urtaza, and Leticia Barrientos. 2022. "Metagenomic Characterization of Resistance Genes in Deception Island and Their Association with Mobile Genetic Elements." *Microorganisms* 10: 1432. https://doi.org/10.3390/microorganisms10071432
19. Azziz, Gastón, Matías Giménez, Héctor Romero, Patricia M. Valdespino-Castillo, Luisa I. Falcón, Lucas A. M. Ruberto, Walter P. Mac Cormack, and Silvia Batista. 2019. "Detection of Presumed Genes Encoding Beta-Lactamases by Sequence Based Screening of Metagenomes Derived From Antarctic Microbial Mats." *Frontiers of Environmental Science & Engineering* 13: 44. https://doi.org/10.1007/s11783-019-1128-1
20. Finke, Nico, R. L. Simister, A. H. O'Neil, Sulung Nomosatryo, C. Henny, L. C., MacLean, et al. 2019. "Mesophilic Microorganisms Build Terrestrial Mats Analogous to Precambrian Microbial Jungles." *Nature Communications* 10: 4323. https://doi.org/10.1038/s41467-019-11541-x
21. Lavoie, Kathleen H., Ara S. Winter, Kaitlyn J. H. Read, Evan M. Hughes, Michael N. Spilde, and Diana E. Northup. 2017. "Comparison of Bacterial Communities From Lava Cave Microbial Mats to Overlying Surface Soils From Lava Beds National Monument, USA." *PLoS One* 12: e0169339. https://doi.org/10.1371/journal.pone.0169339
22. Woodcroft, Ben J., Samuel T. N. Aroney, Rossen Zhao, Mitchell Cunningham, Joshua A. M. Mitchell, Rizky Nurdiansyah, Linda Blackall, and Gene W. Tyson. 2025. "Comprehensive Taxonomic Identification of Microbial Species in Metagenomic Data Using SingleM and Sandpiper." *Nature Biotechnology* 44: 1–6. https://doi.org/10.1038/s41587-025-02738-1
23. Steinegger, Martin, Milot Mirdita, and Johannes Söding. 2019. "Protein-Level Assembly Increases Protein Sequence Recovery from Metagenomic Samples Manyfold." *Nature Methods* 16: 603–6. https://doi.org/10.1038/s41592-019-0437-4
24. Jeong, Seok Hoon, Il Kwon Bae, Jung Hun Lee, Seung Ghyu Sohn, Geun Ho Kang, Ghil Ja Jeon, Young Ho Kim, Byeong Chul Jeong, and Sang Hee Lee. 2004. "Molecular Characterization of Extended-Spectrum Beta-Lactamases Procoli from a Korean Nationwide Survey." *Journal of Clinical Microbiology* 42: 2902–6. https://doi.org/10.1128/JCM.42.7.2902-2906.2004

25. Maravić, Ana, Mirjana Skočibušić, Željana Fredotović, Ivica Šamanić, Svjetlana Cvjetan, Mia Knezović, and Jasna Puizina. 2016. "Urban Riverine Environment Is a Source of Multidrug-Resistant and ESBL-Producing Clinically Important *Acinetobacter* spp." *Environmental Science and Pollution Research* 23: 3525–35. https://doi.org/10.1007/s11356-015-5586-0
26. Zheng, Lingyan, Shuiyang Shi, Mingkun Lu, Pan Fang, Ziqi Pan, Hongning Zhang, Zhimeng Zhou, et al. 2024. "AnnoPRO: A Strategy for Protein Function Annotation Based on Multi-Scale Protein Representation and a Hybrid Deep Learning of Dual-Path Encoding." *Genome biology* 25: 41. https://doi.org/10.1186/s13059-024-03166-1
27. Luo, Hongyan, Zhishuang Yang, Ting Lei, Caixia Li, Zuoyong Zhou, Mingshu Wang, Dekang Zhu, Pei Li, and Anchun Cheng. 2024. "RATA: A Novel Class A Carbapenemase With Broad Geographic Distribution and Potential for Global Spread." *Science of The Total Environment* 931: 172873. https://doi.org/10.1016/j.scitotenv.2024.172873
28. Böhm, Maria-Elisabeth, Mohammad Razavi, Carl-Fredrik Flach, and DG Joakim Larsson. 2020. "A Novel, Integron-Regulated, Class C β-lactamase." *Antibiotics* 9: 123. https://doi.org/10.3390/antibiotics9030123
29. Lupo, Valérian, Paola Sandra Mercuri, Jean-Marie Frère, Bernard Joris, Moreno Galleni, Denis Baurain, and Frédéric Kerff. 2022. "An Extended Reservoir of Class-D Beta-Lactamases in Non-Clinical Bacterial Strains." *Microbiology Spectrum* 10: e00315–22. https://doi.org/10.1128/spectrum.00315-22
30. Berglund, Fanny, Nachiket P. Marathe, Tobias Österlund, Johan Bengtsson-Palme, Stathis Kotsakis, Carl-Fredrik Flach, DG Joakim Larsson, and Erik Kristiansson. 2017. "Identification of 76 Novel B1 Metallo-β-Lactamases Through Large-Scale Screening of Genomic and Metagenomic Data." *Microbiome* 5: 134. https://doi.org/10.1186/s40168-017-0353-8
31. Boyd, David A., Luiz F. Lisboa, Robert Rennie, George G. Zhanel, Tanis C. Dingle, and Michael R. Mulvey. 2019. "Identification of a Novel Metallo-β-Lactamase, CAM-1, in Clinical *Pseudomonas aeruginosa* Isolates From Canada." *Journal of Antimicrobial Chemotherapy* 74: 1563–7. https://doi.org/10.1093/jac/dkz066
32. Rodríguez, María Margarita, Raphaël Herman, Barbara Ghiglione, Frédéric Kerff, Gabriela D'Amico González, Fabrice Bouillenne, Moreno Galleni, et al. 2017. "Crystal Structure and Kinetic Analysis of the Class B3 Di-Zinc Metallo-β-Lactamase LRA-12 From an Alaskan Soil Metagenome." *PloS one* 12: e0182043. https://doi.org/10.1371/journal.pone.0182043
33. Gudeta, Dereje D., Valeria Bortolaia, Simona Pollini, Jean-Denis, Docquier, Gian M. Rossolini, Gregory C. A. Amos, Elizabeth M. H. Wellington, and Luca Guardabassi. 2016. "Expanding the Repertoire of Carbapenem-Hydrolyzing Metallo-ß-Lactamases by Functional Metagenomic Analysis of Soil Microbiota." *Frontiers in Microbiology* 7: 1985. https://doi.org/10.3389/fmicb.2016.01985
34. Wilson, Liam A., Esmée G Knaven, Marc T. Morris, Marcelo Monteiro Pedroso, Christopher J. Schofield, Thomas B. Brück, and Mikael Boden, et al. 2021. "Kinetic and Structural Characterization of the First B3 Metallo-β-Lactamase With an Active-Site Glutamic Acid." *Antimicrobial Agents and Chemotherapy* 65(10): 00936–21. https://doi.org/10.1128/AAC.00936-21
35. Boschi, Letizia, Paola Sandra Mercuri, Maria Letizia Riccio, Gianfranco Amicosante, Moreno Galleni, Jean-Marie Frère, and Gian Maria Rossolini. 2000. "The *Legionella* (*Fluoribacter*) *gormanii* Metallo-β-Lactamase: A New Member of the Highly Divergent Lineage of Molecular-Subclass B3 β-Lactamases." *Antimicrobial agents and chemotherapy* 44: 1538–43. https://doi.org/10.1128/AAC.44.6.1538-1543.2000
36. Ren, Hao, Zhaoxiang Lu, Ruanyang Sun, Xiran Wang, Jiahao Zhong, Tiantian Su, Qian He, et al. 2023. "Functional Metagenomics Reveals Wildlife as Natural Reservoirs of Novel β-Lactamases." *Science of The Total Environment* 868: 161505. https://doi.org/10.1016/j.scitotenv.2023.161505
37. Torres-Cortés, Gloria, Vicenta Millán, Hugo C. Ramírez-Saad, Rafael, Nisa-Martínez, Nicolás Toro, and Francisco Martínez-Abarca. 2011. "Characterization of Novel Antibiotic Resistance Genes Identified by Functional Metagenomics on Soil Samples." *Environmental Microbiology* 13: 1101–14. https://doi.org/10.1111/j.1462-2920.2010.02422.x
38. Kimura, H., M. Izawa, and Y. Sumino. 1996. "Molecular Analysis of the Gene Cluster Involved in Cephalosporin Biosynthesis From *Lysobacter lactamgenus* YK90." *Applied Microbiology and Biotechnology* 44: 589–96. https://doi.org/10.1007/BF00172490
39. Berglund, F., A. Johnning, D. G. J. Larsson, and E. Kristiansson. 2021. "An Updated Phylogeny of the Metallo-v-Lactamases." *Journal of Antimicrobial Chemotherapy: JAC*. https://doi.org/10.1093/JAC/DKAA392
40. Kerkvliet, Jesse J., Alex Bossers, Jannigje G. Kers, Rodrigo Meneses, Rob Willems, and Anita C. Schürch. 2024. "Metagenomic Assembly Is the Main Bottleneck in the Identification of Mobile Genetic Elements." *PeerJ* 12: e16695. https://doi.org/10.7717/peerj.16695
41. Pradier, Léa, Tazzio Tissot, Anna-Sophie Fiston-Lavier, and Stéphanie Bedhomme. 2021. "PlasForest: A Homology-Based Random Forest Classifier for Plasmid Detection in Genomic Datasets." *BMC Bioinformatics* 22: 349. https://doi.org/10.1186/s12859-021-04270-w
42. Yahav, Dafna, Christian G. Giske, Alise Grāmatniece, Henrietta Abodakpi, Vincent H. Tam, and Leonard Leibovici. 2020. "New β-Lactam–β-Lactamase Inhibitor Combinations." *Clinical Microbiology Reviews* 34: 00115–20. https://doi.org/10.1128/CMR.00115-20
43. Sun, Shengwei. 2025. "Emerging Antibiotic Resistance by Various Novel Proteins/Enzymes." *European Journal of Clinical Microbiology & Infectious Diseases* 44: 1551–66. https://doi.org/10.1007/s10096-025-05126-4
44. Rigou, Sofia, Eugène Christo-Foroux, Sébastien Santini, Artemiy Goncharov, Jens Strauss, Guido Grosse, Alexander N. Fedorov, et al. 2022. "Metagenomic Survey of the Microbiome of Ancient Siberian Permafrost and Modern Kamchatkan Cryosols." *Microlife* 3: uqac003. https://doi.org/10.1093/femsml/uqac003
45. Gatica, Joao, Edouard Jurkevitch, and Eddie Cytryn. 2019. "Comparative Metagenomics and Network Analyses Provide Novel Insights into the Scope and Distribution of

β-Lactamase Homologs in the Environment." *Frontiers in Microbiology* 10: 146. https://doi.org/10.3389/fmicb.2019.00146

46. Inda-Díaz, Juan Salvador, David Lund, Marcos Parras-Moltó, Anna Johnning, Johan Bengtsson-Palme, and Erik Kristiansson. 2023. "Latent Antibiotic Resistance Genes Are Abundant, Diverse, and Mobile in Human, Animal, and Environmental Microbiomes." *Microbiome* 11: 44. https://doi.org/10.1186/s40168-023-01479-0
47. Jia, Baolei, Ju Hye Baek, Jae Kyeong Lee, Ying Sun, Kyung Hyun Kim, Ji Young Jung, and Che Ok Jeon. 2024. "Expanding The-Lactamase Family in the Human Microbiome." *Advanced Science* 11: 2403563. https://doi.org/10.1002/advs.202403563
48. Krco, Stefan, Samuel J. Davis, Pallav Joshi, Liam A. Wilson, Marcelo Monteiro Pedroso, Andrew Douw, Christopher J. Schofield, et al. 2023. "Structure, Function, and Evolution of metallo-β-Lactamases From the B3 Subgroup—Emerging Targets to Combat Antibiotic Resistance." *Frontiers in Chemistry* 11: 1196073. https://doi.org/10.3389/fchem.2023.1196073
49. Varliero, Gilda, Pedro H. Lebre, Byron Adams, Steven L., Chown, Peter Convey, Paul G. Dennis, Dandan Fan, et al. 2024. "Biogeographic Survey of Soil Bacterial Communities Across Antarctica." *Microbiome* 12: 9. https://doi.org/10.1186/s40168-023-01719-3
50. Cary, S Craig, Ian R. McDonald, John E. Barrett, and Don A. Cowan. 2010. "On the Rocks: The Microbiology of Antarctic Dry Valley Soils." *Nature Reviews Microbiology* 8: 129–38. https://doi.org/10.1038/nrmicro2281
51. Yergeau, Etienne, and George A. Kowalchuk. 2008. "Responses of Antarctic Soil Microbial Communities and Associated Functions to Temperature and Freeze–Thaw Cycle Frequency." *Environmental Microbiology* 10: 2223–35. https://doi.org/10.1111/j.1462-2920.2008.01644.x
52. McReynolds, Ella, Mostafa S. Elshahed, and Noha H. Youssef. 2025. "An Ecological-Evolutionary Perspective on the Genomic Diversity and Habitat Preferences of the *Acidobacteriota*." *Microbial Genomics* 11: 001344. https://doi.org/10.1099/mgen.0.001344
53. Xiang, Rong, and Meng Li. 2022. "Detection of a Novel Metallo-β-Lactamase, CAM-2, in a Metagenome-Assembled Genome From China." *Microbiology Spectrum* 10: e00261–00222. https://doi.org/10.1128/spectrum.00261-22
54. Krasauskas, Renatas, Danutė Labeikytė, Arvydas Markuckas, Justas Povilonis, Julija Armalytė, Rita Plančiūnienė, Povilas Kavaliauskas, and Edita Sužiedėlienė. 2015. "Purification and Characterization of a New β-lactamase OXA-205 from *Pseudomonas aeruginosa*." *Annals of Clinical Microbiology and Antimicrobials* 14: 52. https://doi.org/10.1186/s12941-015-0113-1
55. Antelo, Verónica, Matías Giménez, Gastón Azziz, Patricia Valdespino-Castillo, Luisa I. Falcón, Lucas A. M. Ruberto, Walter P. Mac Cormack, Didier Mazel, and Silvia Batista. 2021. "Metagenomic Strategies Identify Diverse Integron-Integrase and Antibiotic Resistance Genes in the Antarctic Environment." *MicrobiologyOpen* 10: e1219. https://doi.org/10.1002/mbo3.1219
56. Zhang, Houpu, Zihan Zhang, Jiajin Song, Lin Cai, Yunlong Yu, and Hua Fang. 2021. "Foam Shares Antibiotic Resistomes and Bacterial Pathogens With Activated Sludge in Wastewater Treatment Plants." *Journal of hazardous materials* 408: 124855. https://doi.org/10.1016/j.jhazmat.2020.124855
57. Li, Ruilong, Longji Zhu, Yijin Wang, and Yong-Guan Zhu. 2022. "Metagenomic Insights into Environmental Risk of Field Microplastics in an Urban River." *Water Research* 223: 119018. https://doi.org/10.1016/j.watres.2022.119018
58. Agnieszka, E. Laudy, Patrycja Róg, Katarzyna Smolińska-Król, Milena Ćmiel, Alicja Słoczyńska, and Jan Patzer, et al. 2017. "Prevalence of ESBL-Producing Pseudomonas Aeruginosa Isolates in Warsaw, Poland, Detected by Various Phenotypic and Genotypic Methods." *PLoS One* 12: e0180121. https://doi.org/10.1371/journal.pone.0180121
59. Gholipour, Sevan, John Chen, Dongkyu Lee, and Nobuhiko Tokuriki. 2024. "The global β-lactam resistome revealed by comprehensive sequence analysis." *bioRxiv*. https://doi.org/10.1101/2024.03.01.583042
60. Ebmeyer, Stefan, Erik Kristiansson, and DG Joakim Larsson. 2021. "A Framework for Identifying the Recent Origins of Mobile Antibiotic Resistance Genes." *Communications Biology* 4: 8. https://doi.org/10.1038/s42003-020-01545-5
61. Arros, Patricio, Daniel Palma, Matías Gálvez-Silva, Alexis Gaete, Hugo Gonzalez, Gabriela Carrasco, José Coche, et al. 2024. "Life on the Edge: Microbial Diversity, Resistome, and Virulome in Soils From the Union Glacier Cold Desert." *Science of The Total Environment* 957: 177594. https://doi.org/10.1016/j.scitotenv.2024.177594
62. Majidzadeh, Mohadeseh, Parvin Heidarieh, Mehdi Fatahi-Bafghi, and Mahmood Vakili. 2021. "Antimicrobial Activity of Actinobacteria Isolated From Dry Land Soil in Yazd, Iran." *Molecular Biology Reports* 48: 1717–23. https://doi.org/10.1007/s11033-021-06218-y
63. Pantůček, Roman, Adéla Ivo Sedláček, Veronika Indráková, Ivana Vrbovská, Vojtěch Mašlaňová, Pavel Kovařovic, Stanislava Švec, et al. 2018. "Staphylococcus Edaphicus Sp. Nov., Isolated in Antarctica, Harbors the Mecc Gene and Genomic Islands with a Suspected Role in Adaptation to Extreme Environments." *Applied and Environmental Microbiology* 84: e01746–01717. https://doi.org/10.1128/AEM.01746-17
64. Gholipour, Sevan, Dongkyu Lee, and Nobuhiko Tokuriki. 2025. "Molecular Evolution and Origins of Antibiotic Resistance Genes." *Biochemistry* 64: 2516–28. https://doi.org/10.1021/acs.biochem.5c00044
65. Tawfik, Olga, and Dan S. Khersonsky. 2010. "Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective." *Annual Review of Biochemistry* 79: 471–505. https://doi.org/10.1146/annurev-biochem-030409-143718
66. Jacobs, Christine, Li Jun Huang, Eveline Bartowsky, Staffan Normark. JT395403 Park 1994. ""Bacterial Cell Wall Recycling Provides Cytosolic Muropeptides as Effectors for Beta-Lactamase Induction." *The EMBO Journal* 13: 4684–94. https://doi.org/10.1002/j.1460-2075.1994.tb06792.x
67. Massova, Irina, and Shahriar Mobashery. 1998. "Kinship and Diversification of Bacterial Penicillin-Binding Proteins and β-Lactamases." *Antimicrobial Agents and Chemotherapy* 42: 1–17. https://doi.org/10.1128/aac.42.1.1

68. Fisher, Jed F., and Shahriar Mobashery. 2023. "β-Lactams From the Ocean." *Marine Drugs* 21: 86. https://doi.org/10.3390/md21020086
69. Selleck, Christopher, Marcelo Monteiro Pedroso, Liam Wilson, Stefan Krco, Esmée Gianna Knaven, Manfredi Miraula, Nataša Mitić, et al. 2020. "Structure and Mechanism of Potent Bifunctional β-Lactam-and Homoserine Lactone-Degrading Enzymes From Marine Microorganisms." *Scientific Reports* 10: 12882. https://doi.org/10.1038/s41598-020-68612-z
70. Wong, Sin Yin, James C. Charlesworth, Nicole Benaud, Brendan P. Burns, and Belinda C. Ferrari. 2019. "Communication Within East Antarctic Soil Bacteria." *Applied and Environmental Microbiology* 86: e01968–01919. https://doi.org/10.1128/AEM.01968-19
71. Joakim Larsson, D. G., and Carl-Fredrik Flach. 2022. "Antibiotic Resistance in the Environment." *Nature Reviews Microbiology* 20: 257–69. https://doi.org/10.1038/s41579-021-00649-x
72. Bush, Karen. 2023. "Classification for β-lactamases: Historical Perspectives." *Expert Review of Anti-infective Therapy* 21: 513–22. https://doi.org/10.1080/14787210.2023.2194633
73. Wang, Yanan, Yongfei Hu, Fei Liu, Jian Cao, Na Lv, Baoli Zhu, Gaiping Zhang, and George Fu Gao. 2020. "Integrated Metagenomic and Metatranscriptomic Profiling Reveals Differentially Expressed Resistomes in Human, Chicken, and Pig Gut Microbiomes." *Environment International* 138: 105649. https://doi.org/10.1016/j.envint.2020.105649
74. Turner, John, Gareth J. Marshall, Kyle Clem, Steve Colwell, Tony Phillips, and Hua Lu. 2020. "Antarctic Temperature Variability and Change From Station Data." *International Journal of Climatology* 40: 2986–3007. https://doi.org/10.1002/joc.6378
75. Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. "fastp: An Ultra-Fast All-In-One FASTQ Preprocessor." *Bioinformatics* 34: i884–90. https://doi.org/10.1093/bioinformatics/bty560
76. Vasimuddin, Md, Sanchit Misra, Heng Li, and Srinivas Aluru. 2019. "Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems." 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 314–324.
77. Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. "Twelve Years of SAMtools and BCFtools." *Gigascience* 10: giab008. https://doi.org/10.1093/gigascience/giab008
78. Oja, Hannu. 2010. *Multivariate Nonparametric Methods With R: An Approach Based on Spatial Signs and Ranks*. Springer Science & Business Media
79. Patil, Indrajeet. 2021. "Visualizations With Statistical Details: The 'Ggstatsplot' Approach." *Journal of Open Source Software* 6: 3167. https://doi.org/10.21105/joss.03167
80. Chaumeil, Pierre-Alain, Aaron J. Mussig, Philip Hugenholtz, and Donovan H. Parks. 2022. "GTDB-Tk v2: Memory Friendly Classification With the Genome Taxonomy Database." *Bioinformatics* 38: 5315–6. https://doi.org/10.1093/bioinformatics/btac672
81. Oksanen, Jari, Gavin L. Simpson, F Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, and R. B. O'hara, et al. 2026. "Vegan: Community Ecology Package." R package version 2.8-0. https://vegandevs.github.io/vegan/
82. Rodriguez-R, Luis M., Santosh Gunturu, James M. Tiedje, James R. Cole, and Konstantinos T. Konstantinidis. 2018. "Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity." *MSystems* 3: 00039–18. https://doi.org/10.1128/mSystems.00039-18
83. Ondov, Brian D., Todd J. Treangen, Páll Melsted, Adam B. Mallonee, Nicholas H. Bergman, Sergey Koren, and Adam M. Phillippy. 2016. "Mash: Fast Genome and Metagenome Distance Estimation Using MinHash." *Genome biology* 17: 132. https://doi.org/10.1186/s13059-016-0997-x
84. Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. 2017. "metaSPAdes: A New Versatile Metagenomic Assembler." *Genome Research* 27: 824. https://doi.org/10.1101/gr.213959.116
85. Hyatt, Doug, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11: 119. https://doi.org/10.1186/1471-2105-11-119
86. Naas, Thierry, Saoussen Oueslati, Rémy A Bonnin, Maria Laura, Dabos, Agustin Zavala, Laurent Dortet, Pascal Retailleau, and Bogdan I. Iorga. 2017. "Beta-Lactamase Database (BLDB)–structure and Function." *Journal of Enzyme Inhibition and Medicinal Chemistry* 32: 917–9. https://doi.org/10.1080/14756366.2017.1344235
87. Steinegger, Martin, and Johannes Söding. 2017. "MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets." *Nature Biotechnology* 35: 1026–8. https://doi.org/10.1038/nbt.3988
88. Steinegger, Martin, and Johannes Söding. 2018. "Clustering Huge Protein Sequence Sets in Linear Time." *Nature Communications* 9: 2542. https://doi.org/10.1038/s41467-018-04964-5
89. Lin, Zeming, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, et al. 2023. "Evolutionary-Scale Prediction of Atomic-Level Protein Structure With a Language Model." *Science* 379: 1123–30. https://doi.org/10.1126/science.ade2574
90. Van Kempen, Michel, Stephanie S. Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L. M. Gilchrist, Johannes Söding, and Martin Steinegger. 2024. "Fast and Accurate Protein Structure Search with Foldseek." *Nature Biotechnology* 42: 243–6. https://doi.org/10.1038/s41587-023-01773-0
91. Buchfink, Benjamin, Klaus Reuter, and Hajk-Georg Drost. 2021. "Sensitive Protein Alignments at Tree-Of-Life Scale Using DIAMOND." *Nature Methods* 18: 366–8. https://doi.org/10.1038/s41592-021-01101-x
92. Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30: 772–80. https://doi.org/10.1093/molbev/mst010
93. Minh, Bui Quang, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams,

Arndt Von Haeseler, and Robert Lanfear. 2020. "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era." *Molecular Biology and Evolution* 37: 1530–4. https://doi.org/10.1093/molbev/msaa015

94. Hoang, Diep Thi, Olga Chernomor, Arndt, Von Haeseler, Bui Quang Minh, Le Sy Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35: 518–22. https://doi.org/10.1093/molbev/msx281
95. Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K. F. Wong, Arndt Von Haeseler, and Lars S. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14: 587–9. https://doi.org/10.1038/nmeth.4285
96. Letunic, Ivica, and Peer Bork. 2021. "Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation." *Nucleic acids research* 49: W293–6. https://doi.org/10.1093/nar/gkab301
97. Zhang, Chengxin, Morgan Shine, Anna Marie Pyle, and Yang Zhang. 2022. "US-align: Universal Structure Alignments of Proteins, Nucleic Acids, and Macromolecular Complexes." *Nature Methods* 19: 1109–15. https://doi.org/10.1038/s41592-022-01585-1
98. Teufel, Felix, José Juan Almagro Armenteros, Alexander Rosenberg Johansen, Magnús Halldór Gíslason, Silas Irby Pihl, Konstantinos D. Tsirigos, Ole Winther, et al. 2022. "SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models." *Nature Biotechnology* 40: 1023–5. https://doi.org/10.1038/s41587-021-01156-3
99. Kang, Dongwan D., Feng Li, Edward, Ashleigh Kirton, Rob Thomas, Hong Egan, Zhong Wang. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies." *PeerJ* 7: e7359. https://doi.org/10.7717/peerj.7359
100. Chklovski, Alex, Donovan H. Parks, Ben J. Woodcroft, and Gene W. Tyson. 2023. "CheckM2: A Rapid, Scalable and Accurate Tool for Assessing Microbial Genome Quality Using Machine Learning." *Nature Methods* 20: 1203–12. https://doi.org/10.1038/s41592-023-01940-w
101. Bowers, Robert M., Nikos C. Kyrpides, Ramunas Stepanauskas, Miranda Harmon-Smith, Devin Doud, T. B. K. Reddy, Frederik Schulz, et al. 2017. "Minimum Information About a Single Amplified Genome (MISAG) and a Metagenome-Assembled Genome (MIMAG) of Bacteria and Archaea." *Nature Biotechnology* 35: 725–31. https://doi.org/10.1038/nbt.3893
102. Meijenfeldt, Von, F. A. Bastiaan, Ksenia Arkhipova, Diego D. Cambuy, Felipe H. Coutinho, and Bas E. Dutilh. 2019. "Robust Taxonomic Classification of Uncharted Microbial Sequences and Bins With CAT and BAT." *Genome Biology* 20: 217. https://doi.org/10.1186/s13059-019-1817-x
103. Poff, Kirsten E., Andy O. Leu, John M. Eppley, David M. Karl, and Edward F. DeLong. 2021. "Microbial Dynamics of Elevated Carbon Flux in the Open Ocean's Abyss." *Proceedings of the National Academy of Sciences* 118: e2018269118. https://doi.org/10.1073/pnas.2018269118
104. Aroney, Samuel T. N., Rhys J. P. Newell, Jakob N. Nissen, Antonio Pedro Camargo, Gene W. Tyson, and Ben J. Woodcroft. 2025. "CoverM: Read Alignment Statistics for Metagenomics." *Bioinformatics* 41: btaf147. https://doi.org/10.1093/bioinformatics/btaf147
105. Eddy, Sean R. 2011. "Accelerated Profile HMM Searches." *PLoS Computational Biology* 7: e1002195. https://doi.org/10.1371/journal.pcbi.1002195
106. Boyd, Joel A., Ben J. Woodcroft, and Gene W. Tyson. 2018. "GraftM: a Tool for Scalable, Phylogenetically Informed Classification of Genes Within Metagenomes." *Nucleic Acids Research* 46: e59. https://doi.org/10.1093/NAR/GKY174
107. Hauptfeld, Ernestina, Nikolaos Pappas, Sandra van Iwaarden, Basten L. Snoek, Andrea Aldas-Vargas, Bas E. Dutilh, and FA Bastiaan von Meijenfeldt. 2024. "Integrating Taxonomic Signals From MAGs and Contigs Improves Read Annotation and Taxonomic Profiling of Metagenomes." *Nature Communications* 15: 3373. https://doi.org/10.1038/s41467-024-47155-1
108. Yin, Xiaole, Xiawan Zheng, Liguan Li, An-Ni Zhang, Xiao-Tao Jiang, and Tong Zhang. 2023. "ARGs-OAP v3. 0: Antibiotic-Resistance Gene Database Curation and Analysis Pipeline Optimization." *Engineering* 27: 234–41. https://doi.org/10.1016/j.eng.2022.10.011
109. Yin, Xiaole, Xi Chen, Xiao-Tao Jiang, Ying Yang, Bing Li, Marcus Ho-Hin Shum, Tommy T. Y. Lam, et al. 2023. "Toward a Universal Unit for Quantification of Antibiotic Resistance Genes in Environmental Samples." *Environmental Science & Technology* 57: 9713–21. https://doi.org/10.1021/acs.est.3c00159
110. Eisenhofer, Raphael, Antton Alberdi, and Ben J. Woodcroft. 2026. "Large-Scale Estimation of Bacterial and Archaeal DNA Prevalence in Metagenomes Reveals Biome-Specific Patterns." *MSystems* 11: 1062–01025. https://doi.org/10.1101/2024.05.16.594470
111. Koehler, Leman Julia, Pawel Szczerbiak, P Douglas Renfrew, Vladimir Gligorijevic, Daniel Berenberg, Tommi Vatanen, Bryn C. Taylor, et al. 2023. "Sequence-Structure-Function Relationships in the Microbial Protein Universe." *Nature Communications* 14: 2351. https://doi.org/10.1038/s41467-023-37896-w
112. McInnes, Leland, John Healy, and James Melville. 2018. "Umap: Uniform Manifold Approximation and Projection For Dimension Reduction." *arXiv:1802.03426* [preprint]. https://doi.org/10.48550/arXiv.1802.03426
113. Néron, Bertrand, Eloi Littner, Matthieu Haudiquet, Amandine Perrin, Jean Cury, and Eduardo P. C. Rocha. 2022. "IntegronFinder 2.0: Identification and Analysis of Integrons Across Bacteria, With a Focus on Antibiotic Resistance in Klebsiella." *Microorganisms* 10: 700. https://doi.org/10.3390/microorganisms10040700
114. Wishart, David S., Scott Han, Sukanta Saha, Eponine Oler, Harrison, Peters, Jason R. Grant, Paul Stothard, and Vasuk Gautam. 2023. "PHASTEST: Faster Than PHASTER, Better Than PHAST." *Nucleic Acids Research* 51: W443–50. https://doi.org/10.1093/nar/gkad382
115. Starikova, Elizaveta V., Polina O. Tikhonova, Nikita A. Prianichnikov, Chris M. Rands, Evgeny M. Zdobnov, Elena N. Ilina, and Vadim M. Govorun. 2020. "Phigaro: High-Throughput Prophage Sequence Annotation." *Bioinformatics* 36: 3882–4. https://doi.org/10.1093/bioinformatics/btaa250
116. Xie, Zhiqun, and Haixu Tang. 2017. "ISEScan: Automated Identification of Insertion Sequence Elements in Prokaryotic Genomes." *Bioinformatics* 33: 3340–7. https://doi.org/10.1093/bioinformatics/btx433

117. Schwengers, Oliver, Lukas Jelonek, Marius Alfred Dieckmann, Sebastian Beyvers, Jochen Blom, and Alexander Goesmann. 2021. "Bakta: Rapid and Standardized Annotation of Bacterial Genomes Via Alignment-Free Sequence Identification." *Microbial Genomics* 7: 000685. https://doi.org/10.1099/MGEN.0.000685

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1: NMDS ordination quality and beta-dispersion analysis of Antarctic soil microbial communities.

Figure S2: Benchmarking of sequence-based beta-lactamase searches against the structurally curated candidate set.

Figure S3: Conservation of catalytic and other relevant motifs in divergent Antarctic class A beta-lactamases.

Figure S4: Conservation of catalytic and other relevant motifs in divergent Antarctic class C beta-lactamases.

Figure S5: Conservation of catalytic and other relevant motifs in divergent Antarctic class D beta-lactamases.

Figure S6: Conservation of catalytic and other relevant motifs in divergent Antarctic subclass B1 metallo-beta-lactamases.

Figure S7: Conservation of catalytic and other relevant motifs in divergent Antarctic subclass B3 metallo-beta-lactamases.

Figure S8: Functional similarity clustering of Antarctic candidate beta-lactamases and related true-negative proteins.

Figure S9: Distribution of beta-lactamase classes across assembled metagenomic contigs and metagenome-assembled genomes.

Figure S10: Regional normalized abundance of PBP-like beta-lactamases across Antarctic soil ecotypes.

Figure S11: Regional normalized abundance of MBL-fold beta-lactamases across Antarctic soil ecotypes.

Figure S12: Genomic contexts of Antarctic beta-lactamases associated with putative mobile genetic elements.

Table S1: Antarctic and Subantarctic soil metagenomes analyzed in this study.

Table S2: PBP-like and MBL-like proteins lacking beta-lactamase activity used as negative controls for the search of Antarctic beta-lactamase orthologs.

Table S3: Predicted beta-lactamase orthologs in the Antarctic and Subantarctic soil metagenomes.

Table S4: Putative orthologs of true negatives from the PBP-like and MBL-fold-like superfamilies by protein class identified in 81 Antarctic metagenomes using the workflow proposed in this study.

Table S5: Assembled contigs and MAGs from Antarctic and Subantarctic soil metagenomes.