Technical Report

efsa SUPPORTING PUBLICATIONS

APPROVED: 3 July 2024
doi: 10.2903/sp.efsa.2024.EN-8931

# Catalogue of Antimicrobial Resistance Genes in Species of *Bacillus* used to Produce Food Enzymes and Feed Additives

European Food Safety Authority (EFSA),
Silvia Peluso, Margarita Aguilera-Gómez, Valeria Bortolaia, Francesco Catania, Pier Sandro Cocconcelli, Lieve Herman, Simon Moxon, Laurence Vernis, Giovanni Iacono, Simone Lunardi, Elisa Pettenati, Arianna Gallo, Jaime Aguilera

## Abstract

A key step in the characterisation of bacterial strains used in the food and feed chain is the identification of antimicrobial resistance (AMR) genes in their genomes. The presence of acquired AMR genes influences important aspects of the risk assessment, such as the applicability of the qualified presumption of safety (QPS) approach, which can have a direct impact on the data requirements. Aiming to implement the EFSA approach to discriminate between 'intrinsic' and 'acquired' AMR genes, a bioinformatics pipeline was developed and applied to the species of the genus *Bacillus* that are most frequently subjects of applications for regulated products submitted to EFSA. The results are presented as a catalogue of genes indicating their abundance and distribution among complete and confirmed genomes publicly available for each species. The results of this work are aimed to support the evaluation of AMR genes in a consistent and harmonised way.



**Keywords:** antimicrobial resistance, QPS, whole genome sequence, *Bacillus*, genomic screening, intrinsic AMR genes, acquired AMR genes

**Requestor:** European Commission

**Question number:** EFSA-Q-2023-00741

**Correspondence:** fip@efsa.europa.eu

**Acknowledgements**: EFSA wishes to thank the following for the support provided to this scientific output: Beatriz Guerra Roman, Estefania Noriega Fernandez and Yi Liu.

**Suggested citation:** EFSA (European Food Safety Authority), Peluso S, Aguilera-Gómez M, Bortolaia V, Catania F, Cocconcelli PS, Herman L, Moxon S, Vernis L, Iacono G, Lunardi S, Pettenati E, Gallo A, Aguilera J, 2024. Catalogue of antimicrobial resistance genes in species of *Bacillus* used to produce food enzymes and feed additives. EFSA supporting publication 2024:EN-8931. 33 pp. doi:10.2903/sp.efsa.2024.EN-8931

**ISSN:** 2397-8325

# Table of contents

# 1 Introduction

A key step in the characterisation of microbial active agents and production strains of fermentation products is the assessment of antimicrobial resistance (AMR) genes as intrinsic or acquired. Definitions of intrinsic and acquired AMR has been addressed by EFSA for the purposes of EFSA risk assessments (EFSA BIOHAZ Panel, 2023):

- 'Intrinsic' AMR genes are genes inherent to strains of a bacterial species that limit the action of antimicrobial agents, thereby allowing bacteria to survive and multiply in the presence of the antimicrobial agents. An AMR gene is considered 'intrinsic' if it is shared by the vast majority of wild type strains of the same species (or subspecies) and restricted to those located on the chromosome.

- 'Acquired' AMR genes are interpreted as novel resistance genes for the strain under assessment, acquired through horizontal transfer, enabling a bacterial strain to survive or multiply in the presence of concentrations of an antimicrobial agent higher than those that inhibit the growth of the majority of wild-type strains of the same species without this AMR gene (EFSA FEEDAP Panel, 2018). 'Acquired' AMR genes could be integrated in the bacterial chromosome or harboured on a separate genetic element.

The presence of acquired AMR genes influences important aspects of the risk evaluation, such as the applicability of the qualified presumption of safety (QPS) approach, which can have a direct impact on the data requirement for the risk assessment. For production strains, the presence of acquired AMR genes in their genome triggers the need to assess the possible presence of DNA from the strain in the product under assessment.

EFSA has proposed a methodology starting from a bioinformatics analysis to evaluate AMR genes in bacterial strains under assessment, aiming to classify such genes as intrinsic or acquired (EFSA BIOHAZ Panel, 2023).

# 2 Terms of reference

EFSA will develop a bioinformatic pipeline that implements the methodology to analyse AMR genes distribution among genomes in line with the EFSA BIOHAZ Panel Statement on how to interpret the QPS qualification on 'acquired antimicrobial resistance genes'. By applying the workflow to the main species of the genus *Bacillus* and associated genera whose strains are used as active agents or production strains[1] in applications submitted to EFSA, EFSA will produce a catalogue of AMR genes in those species, aimed to support the safety evaluation of bacterial strains in a consistent and harmonised way.

# 3 Data and methodologies

## 3.1 Data

Whole genome sequence (WGS) data was obtained from the RefSeq database of the National Centre for Biotechnology Information (NCBI, USA)[2]. RefSeq includes non-redundant, well-

[1] As defined in the Guidance on the characterisation of microorganisms used as feed additives or as production organisms (EFSA FEEDAP Panel, 2022).

[2] https://www.ncbi.nlm.nih.gov/refseq/about/

annotated sequence data. Genomes annotated by NCBI RefSeq with an assembly level of complete genomes as updated on the 30th of November 2023 were retained for the analysis. Accession numbers to sequences of AMR genes or AMR gene products were collected from data provided to EFSA in the context of applications for food enzymes and feed additives. Those accession numbers were used to retrieve the original sequence (nucleotide or amino acid) from the NCBI database.

## 3.2 Methodologies

This work follows the general principles for the genome-based identification of intrinsic AMR genes as provided in the *Statement on how to interpret the QPS qualification on 'acquired antimicrobial resistance genes'* (EFSA BIOHAZ Panel, 2023). A summary workflow of the methodology used is depicted in Figure 1.

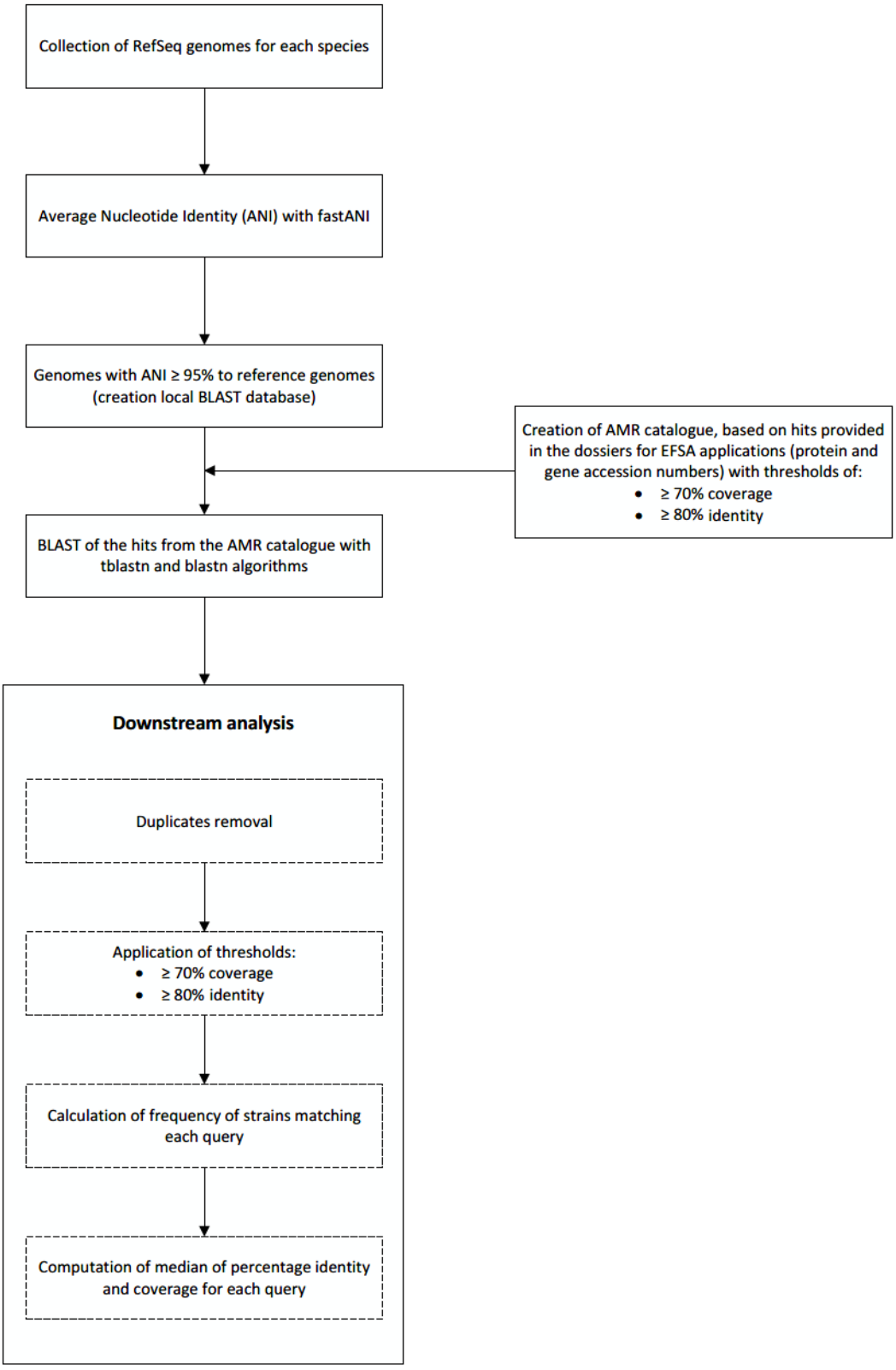


Figure 1: Schematic workflow of the methodology used in this work. Details are provided in Sections 3.2.2. and 3.2.3.

### 3.2.1 Species included in the study

A list of species from the genus *Bacillus* and other associated genera (*e.g.*, *Geobacillus, Priestia*, etc) most of them formerly considered also as *Bacillus*) were obtained from the updated list of QPS-recommended microorganisms for safety risk assessments carried out by EFSA.[3] The species were identified in the RefSeq database and those which contained at least 30 genomes were selected for the analysis, with a few exceptions described further in this document.

### 3.2.2 AMR genes analysed

For the purpose of this guidance, any sequence showing identity and coverage above the established thresholds (EFSA, 2024 and future updates) with an AMR gene included in a curated database is defined as a 'hit'.

For each of the species included in the study, hits to AMR genes found by the applicants in WGS data of production strains or strains used as active agents were extracted from each dossier submitted to EFSA until December 2023. Hits originated from genetically modified strains were also included. Only the results including AMR genes with percentage of coverage equal or over 70 and percentage of identity equal or over 80, as recommended by EFSA (EFSA SC, 2024; EFSA BIOHAZ Panel, 2023) were processed further. EFSA requires the use of two regularly updated databases to investigate the possible presence of AMR genes in genomes (EFSA SC, 2024). Depending on the database used, the data were provided as different molecular identifiers, such as NCBI protein ID (*e.g.*, WP_123456.1, NP_123456.1), as NCBI nucleotide accession number (*e.g.*, gene(A)_1_X1234, geneB_1_XX123456), as CARD ID[4] (*e.g.*, ARO:1234567) or as UniProt[5] accession number (*e.g.*, P12345).

In the case of data were provided as:

- Protein ID: the protein ID was retained.
- Nucleotide accession number: the coordinates of the coding sequence were used to obtain the sequence from NCBI.
- CARD ID: the ARO code was entered into the CARD database to obtain the protein ID from NCBI.
- UniProt accession code: the UniProt accession code was entered into the UniProt database to obtain the protein ID from NCBI.

Once all the data was retrieved, there were assembled into two CSV files per species, one containing the protein IDs and one containing the gene accession numbers.

If the information provided in the dossier did not correspond to any sequence record in databases (such as no identification number or only the gene name provided), the corresponding hit was excluded from this study (Annex A). The accession number of each hit was considered as its unique identifier. From each unique identifier, the original (nucleotide or amino acid) sequence was retrieved from NCBI and compiled into two FASTA files, one for nucleotide sequences and one for amino acid sequences, to be used for further analysis. The sequences of the genes *rpoB*,

[3] Available at https://zenodo.org/records/8124409

[4] https://card.mcmaster.ca

5 https://www.uniprot.org/

*gyrA*, *gyrB* and 16S rRNA from the reference genome of each species were included as internal positive controls (housekeeping genes), whereas the FimA (fimbrial assembly)-encoding gene from *Salmonella* was used as negative control.

### 3.2.3 Bioinformatic analyses

Genomic data were collected from the NCBI. Only RefSeq genomes maintained by NCBI were used for the scope of this analysis. A reference list of all genomes intended for analysis was compiled using a bash script. Average Nucleotide Identity (ANI) was estimated using the 'fastANI' tool V 1.32 (Jain et al., 2018), on modality 'One to Many' with default settings and with the command: -q [QUERY_GENOME] -rl [REFERENCE_LIST] -o [OUTPUT_FILE]. QUERY GENOME indicates the reference genome (type strain when available) of the species analysed. The heatmaps and the files in Newick format clustering dendrogram used to generate the phylogenetic trees were obtained using the ANIclustermap tool (Gould et al., 2023). Genomes with an ANI equal to or greater than 95% (Chun et al., 2018) were retained for downstream analysis. In each ANI analysis reference genomes from other species of *Bacillus* were included as controls. However, it is recognised that the threshold of 95% ANI is not sufficient to distinguish between *B. velezensis* and *B. amyloliquefaciens* in line with published work (Chun et al., 2019). Consequently, the reference genome of *B. amyloquefaciens* was not used as control in the analysis of *B. velezensis* and *vice versa*, other species were used as controls. The discrimination between *B. velezensis* and *B. amyloliquefaciens* only relies on the annotation by RefSeq, which constitutes a limitation in the analysis.

A local BLAST database was created using the filtered genomes. Plasmid sequences were identified and separated from the chromosomes using a custom Python script. BLAST searches of the AMR hits collected from EFSA's dossiers were conducted against the newly generated BLAST database using both protein and gene sequences with the 'tblastn' and 'blastn' commands, respectively. For the BLAST searches, an E-value threshold of 0.05 was used. The raw catalogue obtained underwent downstream analysis as follows: (i) Where more than one hit for an AMR gene occurred in the same genome, all duplicates were removed, retaining only the hit with the highest percentage of identity. (ii) A raw file containing all the matches was retained for parallel analysis (visual outputs). (iii) Only the results including AMR genes with percentage of coverage equal or over 70 and percentage of identity equal or over 80 (EFSA SC, 2024) were processed further (iv). The frequency of strains matching each query (AMR gene) was calculated. (v) Median percentage identity and coverage values were computed for each query. The results were saved as CSV files for further analysis and visualization.

Most of the scripts were run in python3[6] and some of the graphical outputs were produced in R[7] Statistical Software (v4.3.3; R Core Team 2024).

# 4 Results

## 4.1 Species and verified complete genomes

Nineteen species of the genus *Bacillus* or related genus (including those formerly classified as *Bacillus*) were found in the QPS list (version of June 2023). Of these, five species (*B. subtilis*, *B. velezensis*, *B. amyloliquefaciens*, *B. lichenifomis* and *Priestia megaterium*) had 30 or more

[6] Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

[7] R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

complete genomes deposited in the RefSeq database (Table 1). No dossier containing genomic data on *P. megaterium* has been received by EFSA, so this species was excluded from the analysis. On the other hand, it was decided to include *B. paralicheniformis,* even though the number of genomes (=26) was slightly below the threshold of 30 (EFSA BIOHAZ Panel, 2023), because of the relevant number of dossiers containing genomic data. Therefore, the species included in this study were *B. subtilis*, *B. velezensis*, *B. amyloliquefaciens*, *B. licheniformis* and *B. paralicheniformis.*

Table 1: *Bacillus* and related species and confirmed genomes

| Species[a] | Screened genomes[b] | Confirmed genomes (% of screened) |
|---|---|---|
| ***B. velezensis*** | 329 | 328 (99.7%) |
| ***B. subtilis*** | 299 | 299 (100%) |
| ***B. amyloliquefaciens*** | 76 | 57 (75%) |
| ***B. licheniformis*** | 40 | 40 (100%) |
| *Priestia megaterium (B. megaterium)* | 40 | NA |
| ***B. paralicheniformis*** | 26 | 26 (100%) |
| *Weizmannia coagulans (B. coagulans)* | 21 | NA |
| *B. pumilus* | 19 | NA |
| *Parageobacillus thermoglucosidasius* | 6 | NA |
| *Geobacillus stearothermophilus* | 5 | NA |
| *Lysinibacillus fusiformis (B. fusiformis)* | 5 | NA |
| *Niallia circulans (B. circulans)* | 5 | NA |
| *Priestia flexa (B. flexus)* | 5 | NA |
| *Geobacillus thermodenitrificans* | 4 | NA |
| *Alkalihalobacillus clausii (B. clausii)* | 3 | NA |
| *B. vallismortis* | 2 | NA |
| *B. smithii* | 1 | NA |
| *Lederbergia lenta (B. lentus)* | 1 | NA |
| *Paenibacillus illinoisensis* | 0 | NA |

(a): Species in bold letters were included in this study.
(b): Species with ≥ 30 genomes (as from 30 November 2023) were considered for inclusion in this study, with the exceptions explained in the text.

## 4.2 AMR genes and their presence in verified genomes

### *4.2.1 Bacillus velezensis*

All but one analysed genomes of *B. velezensis* were confirmed using the threshold of 95% ANI.

Table 2: Hits to AMR genes reported in dossiers in the publicly available genomes of *B. velezensis*. Grey rows correspond to amino acid sequences, while yellow rows to nucleic acid sequences. Internal positive and negative control sequences are depicted in green and red text, respectively.

| Accession number * | Gene symbol* | Description as indicated in NCBI | MIA[(a)] | Frequency of genomes with at least one hit for gene (%) | Median identity (%) | Median coverage (%) |
|---|---|---|---|---|---|---|
| KM359438.1 | *cfrB* | *Peptoclostridium difficile* strain 11140508 23S rRNA methyltransferase gene | Yes | 81.7 | 88.0 | 100 |
| KM359439.1 | *cfrB* | *Peptoclostridium difficile* strain 11107643 23S rRNA methyltransferase gene | Yes | 81.7 | 88.2 | 100 |
| KR610408.1 | *cfrB* | *Enterococcus faecium* strain 448-18961R 23S rRNA methyltransferase gene | Yes | 81.7 | 88.3 | 100 |
| WP_012116915.1/ AGZ55247.1 | *clbA* | *Bacillus amyloliquefaciens* 23S rRNA (adenine(2503)-C(8))-methyltransferase | Yes | 80.8 | 99.4 | 100 |
| KIX81495.1 | *lmrB* | Bacillales lincomycin resistance protein | Yes | 99.7 | 88.4 | 99.4 |
| WP_087347987.1 | *rphC* | *Brevibacillus brevis* rifamycin-inactivating phosphotransferase | Yes | 15.2 | 80.1 | 100 |
| D12567.1 | *tetL* | *Bacillus subtilis* tetBSR gene | Yes | 83.5 | 86.7 | 99.0 |
| HM235948.1 | *tetL* | *Bacillus* sp. (plasmid pBHS24) *tet* gene | Yes | 82.6 | 80.8 | 98.2 |

| | | | | | | |
|---|---|---|---|---|---|---|
| X08034.1/ NG_048204.1 | *tetL* | *Bacillus subtilis* 168) gene for tetracycline efflux MFS transporter | Yes | 83.5 | 86.8 | 100 |
| CAB13167.1 | *ykkD* | *Bacillus subtilis* 168 guanidinium efflux transporter subunit | | 99.4 | 81.0 | 100 |
| NZ_CP009679.1 9759-11308 | 16S rRNA | *B. velezensis* strain JS25R 16S ribosomal RNA | NA[(b)] | 100 | 99.9 | 100 |
| ATY38581.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | NA | NA |
| KY367272.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | NA | NA |
| NZ_CP009679.1 7003-9462 | *gyrA* | *B. velezensis* strain JS25R DNA gyrase subunit A | NA | 100 | 98.6 | 100 |
| WP_014416706.1 | *gyrA* | *B. velezensis* strain JS25R DNA gyrase subunit A | NA | 99.7 | 99.8 | 98.4 |
| NZ_CP009679.1 4871-6787 | *gyrB* | *B. velezensis* strain JS25R DNA gyrase subunit B | NA | 100 | 99.0 | 100 |
| WP_014720600.1 | *gyrB* | *B. velezensis* strain JS25R DNA gyrase subunit B | NA | 99.7 | 99.9 | 100 |
| NZ_CP009679.1 117191-120772 | *rpoB* | *B. velezensis* strain JS25R RNA polymerase subunit beta | NA | 100 | 99.4 | 100 |
| WP_003156440.1 | *rpoB* | *B. velezensis* strain JS25R RNA polymerase subunit beta | NA | 100 | 99.9 | 100 |

* As provided in dossiers.
MIA: medically important antimicrobial (WHO, 2024).
NA: non applicable

Table 2 shows the results of the BLAST searches for the hits to AMR genes identified in strains from dossiers against the publicly available, confirmed genomes of *B. velezensis.* All hits were found to match above thresholds in all strains (Figure 2) with the exception of *rphC*, that was widespread among the genomes with a median identity of 79.8% (Figure 1A).

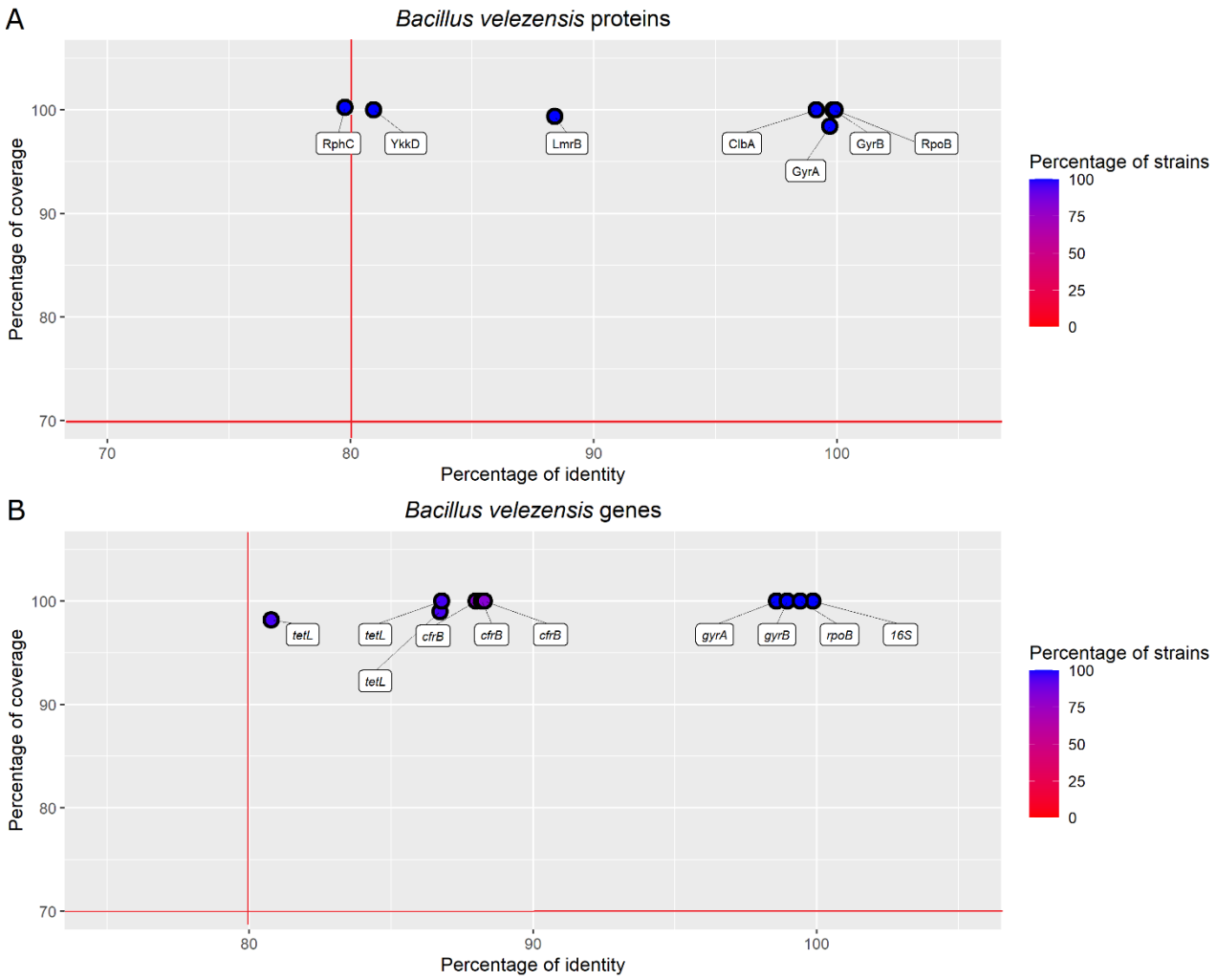


Figure 2: Median identity and coverage of matches to AMR genes found in *B. velezensis* publicly available genomes. Searches based on NCBI accession codes to amino acid (A) and nucleotide (B) sequences. The percentage of analysed genomes containing matches is depicted with colours. Thresholds for identity (80%) and coverage (70%) as indicated in EFSA BIOHAZ Panel, 2023 are indicated with red lines. Dots with the same gene symbol correspond to different allelic variants of the same gene (see table 2).

### *4.2.2 Bacillus subtilis*

All analysed genomes of *B. subtilis* were confirmed using the threshold of 95% ANI. Twenty five percent of the screened genomes (76) included plasmids. Plasmid sequences were analysed independently, and hits found in plasmids are reported separately (Table 3).

Table 3: Hits to AMR genes reported in dossiers in the publicly available genomes of *B. subtilis*. Grey rows correspond to amino acid sequences, while yellow rows to nucleic acid sequences. Frequencies of hits found in plasmids are reported in brackets and expressed as percentages with respect to all genomes containing plasmids (76). Internal positive and negative control sequences are depicted in green and red text, respectively. Two accession numbers leading to the same sequence are reported separated by a slash (/).

| | | | | Chromosome (plasmid) | | |
|---|---|---|---|---|---|---|
| **Accession number*** | **Gene symbol*** | **Description as indicated in NCBI** | **MIA [a]** | **Frequency of genomes with at least one hit (%)** | **Median identity (%)** | **Median coverage (%)** |
| WP_003229862.1/ CAB14620.1 | *aadK* | Bacillales aminoglycoside 6-adenylyltransferase | Yes | 97.0 | 98.6 | 100 |
| M26879.1/ AL009126.3 2735682-2736536 (-) | *aadK* | *B. subtilis* 168 aminoglycoside 6-adenylyltransferase gene | Yes | 97.0 | 98.8 | 100 |
| WP_001014230.1 | ANT(4')-Ib (protein symbol) | *Staphylococcus aureus* pUB110 aminoglycoside nucleotidyltransferase | Yes | 3.3 | 99.2 | 100 |
| WP_003229880.1/ AAC36944.1 | *blt* | Bacillales 168 multidrug efflux transporter | | 98.3 | 99.5 | 100 |
| L32599.1 | *blt* | *B. subtilis* 168 multidrug efflux transporter | | 98.3 (1.3) | 99.3 (98.1) | 100 (100) |
| WP_003229881.1 | *bltR* | Bacillales 168 multidrug efflux transcriptional regulator | | 98.0 | 99.6 | 100 |
| M33768.1 | *bmr* | *B. subtilis* Efflux-mediated multidrug resistance protein | | 99.7 | 98.9 | 100 |
| WP_003230328.1/ AAA22277.1[b] | *bmr* | Bacillales Efflux-mediated multidrug resistance protein | | 99.0 | 100 | 100 |
| WP_003230325.1 | *bmrR* | *B. subtilis* multidrug efflux transcriptional regulator | | 99.7 | 100 | 100 |
| ANM47690.1 | *cat* | *Streptococcus* phage chloramphenicol acetyltransferase | Yes | 3.7 (1.3) | 100 (100) | 92.7 (92.7) |
| AAB53259.1 | *cat-TC* | *Limosilactobacillus reuteri* chloramphenicol acetyltransferase (plasmidic) | Yes | 3.7 (1.3) | 92.5 (92.5) | 100 (100) |

| | | | | | | |
|---|---|---|---|---|---|---|
| WP_002320869.1[c] | EF-Tu (protein symbol) | *Enterococcus faecium* elongation factor Tu (Sequence superseeded) | No | 99.3 | 87.9 | 93.6 |
| AAF86219.1 | *ermB* | *Enterococcus faecium* streptogramin resistance | Yes | 1 | 98.8 | 98.8 |
| AJ605334.1 | *fosB1* | *Bacillus cereus* MIC231V/D mobile insertion cassette containing a fosfomycin resistance gene and D-stereospecific endopeptidase gene, strain As4-12 | Yes | 0 | NA[d] | NA |
| CP001903.1 1933754-1934170 (+) | *fosB1* | *Bacillus thuringiensis* BMB171 fosfomycin resistance protein | Yes | 0 | NA | NA |
| WP_003246449.1/ CAB12062.1 | *lmrA* | Bacillales 168 transcriptional repressor of lmrAB and yxaGH operons | Yes | 98.3 | 99.5 | 100 |
| WP_010886396.1 | *lmrB* | Bacillales lincomycin efflux MFS transporter | Yes | 98.3 | 99.8 | 100 |
| KIX81495.1[e] | *lmrB* | Bacillales lincomycin efflux MFS transporter | Yes | 98.3 | 99.6 | 100 |
| JYFL01000006.1 112070-113503 (-) | *lmrB* | *B. subtilis* lincomycin efflux MFS transporter | Yes | 98.3 | 98.6 | 100 |
| WP_003246254.1/ AJE92936.1 | *mphK* | Bacillales macrolide 2'-phosphotransferase | Yes | 94.0 | 98.7 | 100 |
| NC_000964.3 275838-276758 (+) | *mphK* | *B. subtilis* 168 putative macrolide 2'-phosphotransferase | Yes | 98.3 | 98.1 | 100 |
| WP_003242495.1/ CAX52582.1 | *mprF* | Bacillales Bifunctional lysylphosphatidylglycerol flippase/synthetase | No | 99.7 | 99.8 | 100 |
| AL009126.3 916778-919348 (+) | *mprF* | *B. subtilis* 168 Bifunctional lysylphosphatidylglycerol flippase/synthetase | No | 99.7 | 98.8 | 100 |
| WP_003243306.1/ BAA85265.1 | *pgsA / capA* | Bacillales capsular polyglutamate synthetase | No | 99.7 | 100 | 100 |

| | | | | | | |
|---|---|---|---|---|---|---|
| APB03222.1 | *rphB* | *Paenibacillus sp*. rifampin phosphotransferase | Yes | 1.3 | 80.1 | 99.9 |
| KX531052.1 | *rphB* | *Paenibacillus sp*. rifampin phosphotransferase | Yes | 85.3 | 80.4 | 99.8 |
| WP_000918664.1 | *rpoB* | *Staphylococcus aureus* DNA-directed RNA polymerase mutant | Yes | 99.3 | 80.9 | 97.6 |
| WP_003242546.1 | *satA* | Bacillales streptothricin N-acetyltransferase | No | 98.3 | 91.6 | 100 |
| NP_052129.1[f] | TEM-116 (protein symbol) | Broad-spectrum class A beta-lactamase TEM-116 (Bacteria, Archaea) | Yes | 1.0 | 98.8 | 100 |
| GU584222.2 | *tet(45)* | *Escherichia coli* DH5α tetracycline resistance gene, partial cds | Yes | 0 | NA | NA |
| AAA22851.1[g] | *tetL* | *B. stearothermophilus* plasmid pTHT15 tetracycline resistance gene | Yes | 63.2 (1.3) | 81.2 (99.6) | 99.8 (100) |
| M11036.1 | *tetL* | *B. stearothermophilus* plasmid pTHT15 tetracycline resistance gene | Yes | 55.9 (1.3) | 80.5 (99.8) | 98.9 (100) |
| X08034.1/ NG_048204.1 | *tetL* | *Bacillus subtilis* 168 gene for tetracycline efflux MFS transporter | Yes | 63.2 (1.3) | 100 (80.5) | 100 (98.8) |
| WP_003242953.1 | *tetL* | Bacillales tetracycline efflux MFS transporter | Yes | 63.2 (1.3) | 99.8 (81.6) | 100 (99.6) |
| WP_003246258.1 / CAB12108.2 | *tmrB* | Bacillales Tunicamycin resistance ATP-binding | Yes | 98.3 | 98.5 | 100 |
| AL009126.3 339156-339749 (-) | *tmrB* | Bacillales Tunicamycin resistance ATP-binding | No | 98.7 | 99.0 | 100 |
| WP_003234144.1/ NP_388442.1 | *vmlR* | Bacillales ABC-F type ribosomal protection protein | Yes | 97.9 | 98.5 | 100 |
| WP_044456553.1 | VOC[h](protein symbol) | *B. subtilis* VOC (vicinal oxygen chelate) family protein[e] | Yes | 0.7 | 98.6 | 90.8 |
| WP_003232583.1 / CAB13166.1 | *ykkC* | Bacillales multidrug efflux SMR transporter subunit | | 99.7 | 100 | 100 |

| | | | | | | |
|---|---|---|---|---|---|---|
| AL009126.3 1376517-1376855 (+) | *ykkC* | *B. subtilis* 168 multidrug efflux SMR transporter subunit | | 99.7 | 99.7 | 100 |
| WP_003245695.1 / CAB13167.1 | *ykkD* | Bacillales multidrug efflux SMR transporter subunit | | 99.3 | 99 | 100 |
| AL009126.3 1376855-1377172 (+) | *ykkD* | *B. subtilis* 168 multidrug efflux SMR transporter subunit | | 99.7 | 99.7 | 100 |
| NC_000964.3 | 16S rRNA | *B. subtilis* 168 Ribosomal RNA | NA | 99.7 | 100 | 100 |
| ATY38581.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | NA | NA |
| KY367272.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | NA | NA |
| NP_387888.1 | *gyrA* | *B. subtilis* 168 DNA gyrase subunit A | NA | 99.7 | 100 | 98.2 |
| NC_000964.3 6994-9459 (+) | *gyrA* | *B. subtilis* 168 DNA gyrase subunit A | NA | 99.7 | 99 | 100 |
| NP_387887.1 | *gyrB* | *B. subtilis* 168 DNA gyrase subunit B | NA | 99.7 | 100 | 100 |
| NC_000964.3 | *gyrB* | *B. subtilis* 168 DNA gyrase subunit B | NA | 99.7 | 99. | 100 |
| NP_387988.2 | *rpoB* | *B. subtilis* 168 RNA polymerase subunit beta | NA | 99.3 | 99.9 | 100 |
| NC_000964.3 | *rpoB* | *B. subtilis* 168 RNA polymerase subunit beta | NA | 99.3 | 99.6 | 100 |

* As provided in dossiers.
MIA: medically important antimicrobial (WHO, 2024).
The proteins with the two codes in column 1 differ in 1 amino acid, but both entries refer to the same publication and original submission.
This record is currently superseded in the NCBI database at the time of publication of this report. The same sequence is also deposited with accession number EFR69584.1.
NA: non-applicable.
Variant described in a genetically modified vitamin B2-overproducing strain.
This record is currently superseded in the NCBI database at the time of publication of this report. The same sequence is also deposited as broad-spectrum class A beta-lactamase (accession number WP_000027050.1).
Gene described in pTHT15, a plasmid found in thermophilic *Bacillus*.
This record is currently superseded in the NCBI database. The same sequence is also deposited as a bleomycin resistance protein from an EU-unauthorized genetically modified *Bacillus subtilis* (accession number KIX80085.1).

Most of the hits correspond to genes that were described in *B. subtilis* strain 168 and were in fact found above thresholds in the majority of the genomes (Figure 3). However, some others are underrepresented, such as *tetL* (NG_048204.1/WP_003242953.1), which is present in ca. 63.2% of the chromosomes. Another *tetL* (AAA22851.1/M11036.1) gene described in plasmid

pTHT15 from *B. stearothermophilus*, was also found with similarly low frequency in the chromosomes (63.2/55.9%).

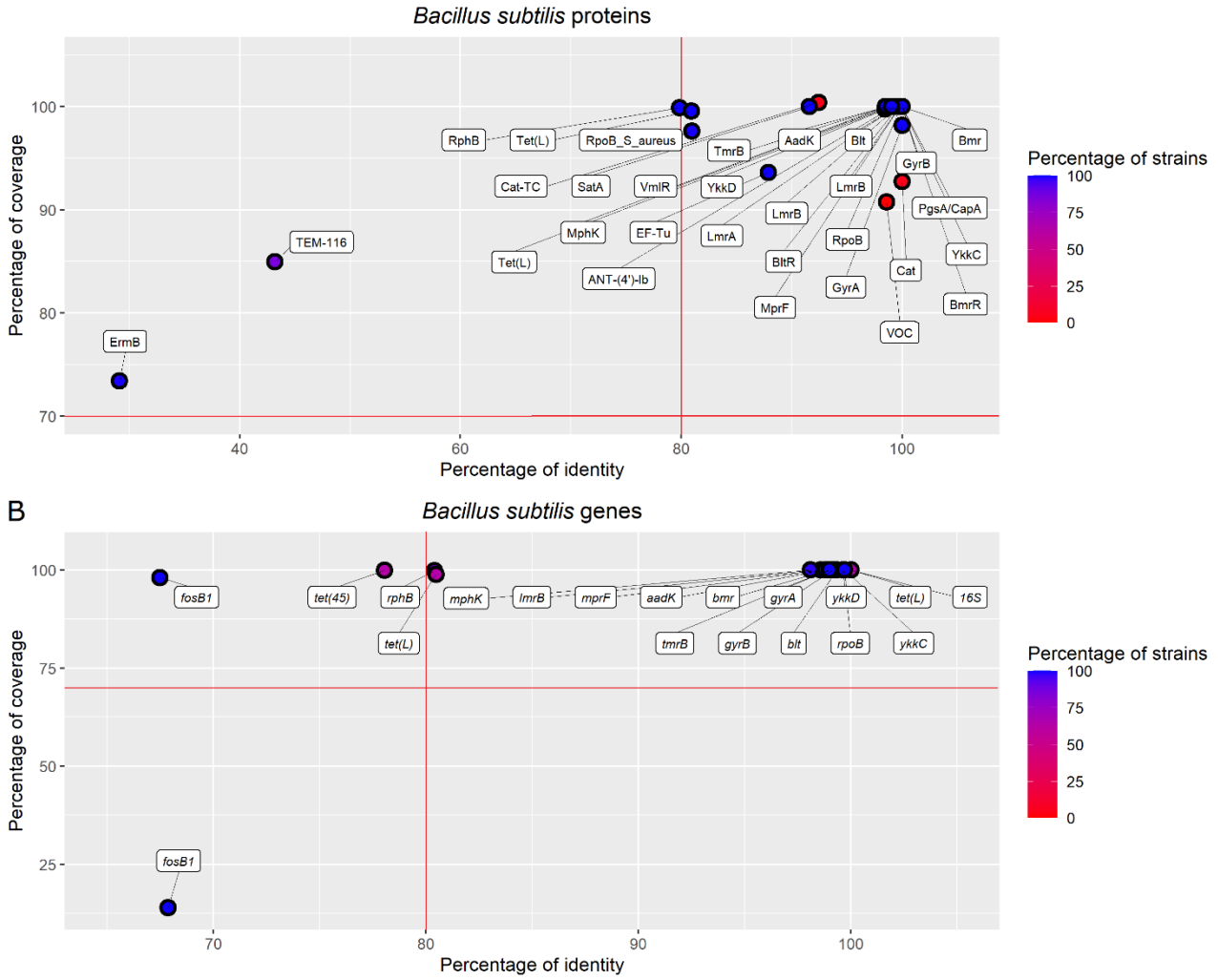


Figure 3: Median identity and coverage of matches to AMR genes found in *B. subtilis* publicly available genomes (chromosome). Searches based on NCBI accession codes to amino acid (A) and nucleotide (B) sequences. The percentage of analysed genomes containing matches is depicted with colours. Thresholds for identity (80%) and coverage (70%) as indicated in EFSA BIOHAZ Panel, 2023 are indicated with red lines. Dots for *tetL* correspond to different allelic variants (see Table 3). Dots for *fosB1* correspond to *B. thuringiensis* CP001903.1 (upper) and *B. cereus* AJ605334.1 (lower).

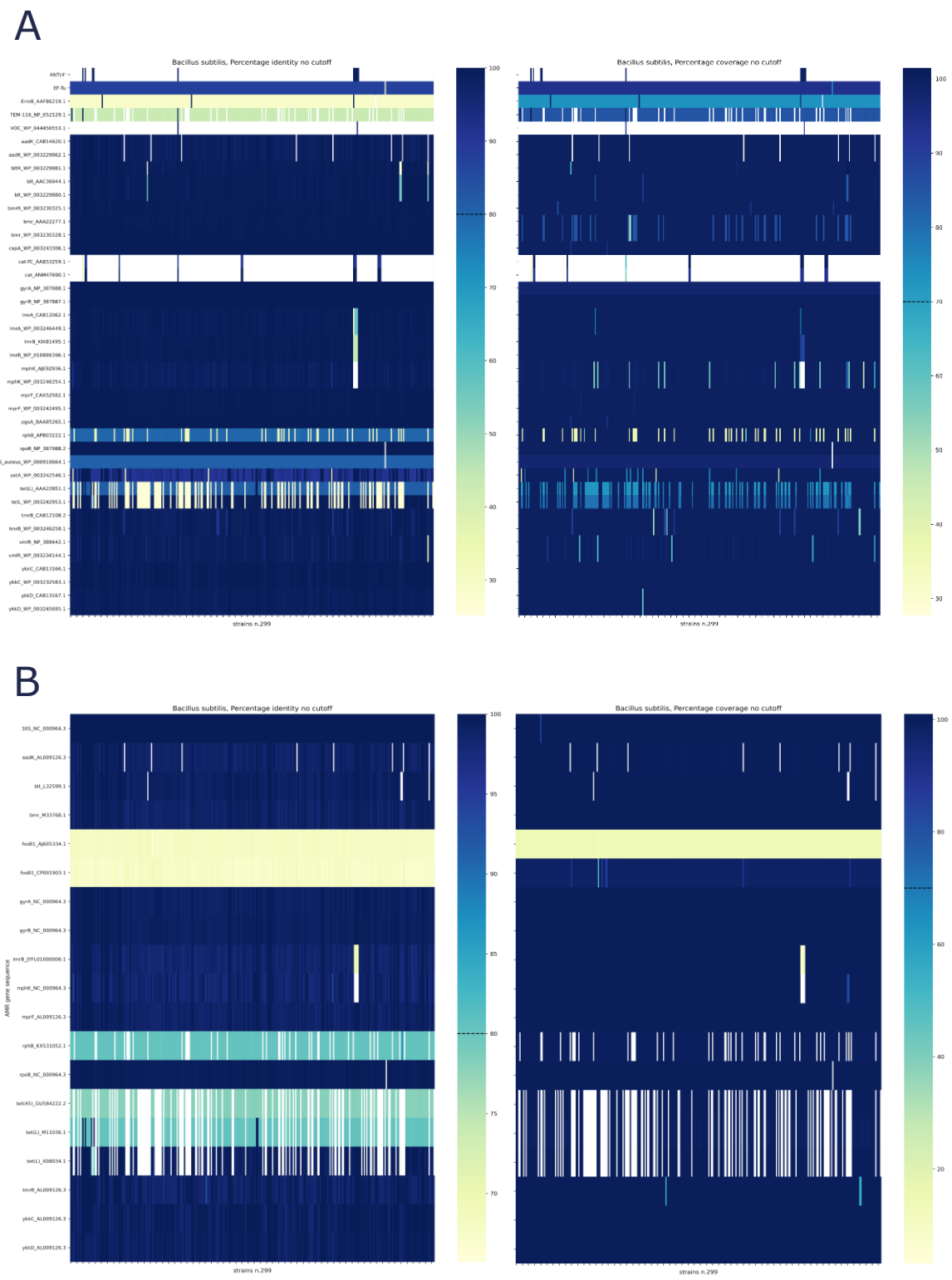


Figure 4: Distribution of identity (left) and coverage (right) of matches to AMR genes among publicly available genomes of *B. subtilis*. Searches based on accession numbers to (A) amino acid or (B) nucleotide sequences. A high-resolution plot can be found in Annex C, Figure 2.

The hit to *rphB* (KX531052.1) was found above thresholds in ca. 85% of the screened genomes when the nucleotide sequence was blasted, but only 1.3 % when the query corresponded to the amino acid sequence (APB03222.1) due to the 80% threshold of identity applied (Figure 3A). Looking at the distribution of the hit among the genomes (Figure 4A) shows that the hit is present in most of the genomes (84%) with a percentage of identity between 79 % and 79.9%.

A few hits to genes described in species different from *B. subtilis* were found in very few genomes or not found. This is the case of ANT(4')-Ib (WP_001014230.1), which is an aminoglycoside O-nucleotidyltransferase encoded by a gene present in pUB110 (McKenzie et al., 1986). Likewise, two hits to chloramphenicol acetyltransferases (ANM47690.1 from *Streptococcus phage* and AAB53259.1 from *Limosilactobacillus reuteri*), matched to a small proportion of genomes. The gene *ermB* from *Enterococcus faecium* (AAF86219.1), conferring macrolide resistance, was found to be present in only three genomes. The fosfomycin resistance gene *fosB1*, described in *B. cereus* (AJ605334.1) and also annotated in *B. thuringensis* (CP001903.1 [1933754-1934170(+)]) was not found in any of the analysed genomes.

### *4.2.3 Bacillus amyloliquefaciens*

Twenty five percent of the deposited genomes showed an ANI of < 95% with respect to the reference strain (Table 1, Annex B Figure 3), and therefore they were excluded from the analysis.

Table 4: Hits to AMR genes reported in dossiers in the publicly available genomes of *B. amyloliquefaciens*. Grey rows correspond to amino acid sequences, while yellow rows to nucleic acid sequences. Internal positive and negative control sequences are depicted in green and red text, respectively.

| Accession number* | Gene symbol * | Description as indicated in NCBI | MIA[(a)] | Frequency of genomes with at least one hit (%) | Median identity (%) | Median coverage (%) |
|---|---|---|---|---|---|---|
| WP_087343787.1 | *aadK* | *Brevibacillus brevis* aminoglycoside nucleotidyltransferase ANT(6)-Ic | Yes | 0 | NA[(b)] | NA |
| ADA62098.1 | *aadD2*/ ANT(4')-Ib (protein symbol) | *Staphylococcus aureus* pSAP079A aminoglycoside nucleotidyltransferase | Yes | 0 | NA | NA |
| M19465.1<br>2202-2963 (-) | ANT(4')-Ib (protein symbol) | *Staphylococcus aureus* pUB110 aminoglycoside nucleotidyltransferase | Yes | 0 | NA | NA |
| AF181950.1<br>3176-3946 (+) | ANT(4')-Ia (protein symbol) | *Staphylococcus aureus* pUB110 aminoglycoside nucleotidyltransferase | Yes | 0 | NA | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| KM359438.1 | *cfrB* | *Peptoclostridium difficile* strain 11140508 23S rRNA methyltransferase gene | Yes | 84.2 | 88 | 100 |
| KM359439.1 | *cfrB* | *Peptoclostridium difficile* strain 11107643 23S rRNA methyltransferase gene | Yes | 84.2 | 88.2 | 100 |
| KR610408.1 | *cfrB* | *Enterococcus faecium* strain 448-18961R 23S rRNA methyltransferase gene | Yes | 84.2 | 88.3 | 100 |
| CDF47262.1 | *cfrB* | *Clostridioides difficile* chloramphenicol/fluorfenicol resistance protein | Yes | 84.2 | 86.8 | 100 |
| CP006845.1 (539695 - 540745) | *clbA* | *Bacillus amyloliquefaciens* CC178 23S rRNA methyltransferase gene | Yes | 84.2 | 98.9 | 100 |
| WP_012116915.1/ AGZ55247.1 | *clbA* | *Bacillus amyloliquefaciens* 23S rRNA (adenine(2503)-C(8))-methyltransferase | Yes | 84.2 | 99.4 | 100 |
| AIX48090.1 | *clcD* | *Clostridioides difficile* 23S rRNAmethyltransferase | Yes | 84.2 | 86.2 | 100 |
| WP_000938987.1 | *fosB* | *B. anthracis* strain Ames fosfomycin resistance bacillithiol transferase | Yes | 0 | NA | NA |
| KIX81495.1 | *lmrB* | *Bacillus subtilis* Lincomycin resistance protein | Yes | 100 | 88.2 | 99.4 |
| WP_063842248.1 | *bla1/ penP* | *Bacillus cereus* class A beta-lactamase | Yes | 0 | NA | NA |
| KX531052.1 | *rphB* | *Paenibacillus sp.* rifampin phosphotransferase | Yes | 0 | NA | NA |
| WP_087347987.1 | *rphC* | *Brevibacillus brevis* rifamycin-inactivating phosphotransferase | Yes | 0 | NA | NA |
| NG_064662.1 | *satA* | *Bacillus subtilis* 168 gene for streptothricin N-acetyltransferase | No | 91.2 | 84.8 | 100 |
| WP_003242546.1 | *satA* | *Bacillales* streptothricin N-acetyltransferase | No | 21.1 | 80.3 | 100 |
| WP_003242953.1 | *tetL* | *Bacillales* tetracycline efflux MFS transporter | Yes | 80.7 | 86.4 | 99.8 |

| | | | | | | |
|---|---|---|---|---|---|---|
| D12567.1 | *tetL* | *Bacillus subtilis* tetBSR gene | Yes | 82.5 | 86.7 | 99.0 |
| HM235948.1 | *tetL* | *Bacillus* sp. (plasmid pBHS24) *tet* gene | Yes | 82.5 | 80.9 | 98.2 |
| X08034.1/ NG_048204.1 | *tetL* | *Bacillus subtilis* 168 gene for tetracycline efflux MFS transporter | Yes | 82.5 | 87.1 | 97.0 |
| CP003583.1 (60337 - 61609) | *tufA* | *Enterococcus faecium* translation elongation factor EF-Tu 1 mutant | No | 100 | 81.6 | 93.6 |
| WP_003234144.1/ NP_388442.1 | *vmlR* | *B. subtilis* ABC-F type ribosomal protection protein | Yes | 0 | NA | NA |
| WP_003232583.1 / CAB13166.1 | *ykkC* | Bacillales multidrug efflux SMR transporter subunit | | 0 | NA | NA |
| CAB13167.1 | *ykkD* | *Bacillus subtilis* 168 guanidinium efflux transporter subunit | | 100 | 81.0 | 100 |
| CP072120.1 (818121 - 819670) | 16S rRNA | *Bacillus amyloliquefaciens* strain GKT04 Ribosomal RNA | NA | 100 | 99.9 | 100 |
| ATY38581.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | NA | NA |
| KY367272.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | NA | NA |
| WP_015239009.1 | *gyrA* | *Bacillus amyloliquefaciens* strain GKT04 DNA gyrase | | 100 | 99.5 | 98.4 |
| CP072120.1 (1819414 - 1821873) | *gyrA* | *Bacillus amyloliquefaciens* strain GKT04 DNA gyrase | NA | 100 | 98.5 | 100 |
| WP_014720600.1 | *gyrB* | *Bacillus amyloliquefaciens* strain GKT04 DNA gyrase | | 100 | 99.8 | 100 |
| CP072120.1 (1817282 - 1819198) | *gyrB* | *Bacillus amyloliquefaciens* strain GKT04 DNA gyrase | NA | 100 | 98.9 | 100 |
| WP_003156440.1 | *rpoB* | *Bacillus amyloliquefaciens* strain GKT04 RNA polymerase | | 100 | 99.9 | 100 |
| CP072120.1 (1935392 - 1938973) | *rpoB* | *Bacillus amyloliquefaciens* strain GKT04 RNA polymerase | NA | 100 | 99.4 | 100 |

* As provided in dossiers.

MIA: medically important antimicrobial (WHO, 2024).
NA: non-applicable.

Several hits were found above thresholds in the majority of the genomes. None of the hits provided in dossiers to aminoglycoside nucleotidyltransferases described to be present in different plasmids of *S. aureus* were found in the RefSeq genomes of *B. amlyloliquefaciens*. No matches above thresholds were found to *fosB*, encoding a fosfomycin bacillithiol transferase described in *B. anthracis* (WP_000938987.1), *penP,* a beta-lactamase resistance gene described in *B. cereus* (WP_063842248.1), *vmlR*, encoding a ribosomal protection protein described in *B. subtilis* and *ykkC*, a multidrug efflux transporter described in some *Bacillus* species (Figure 5A).

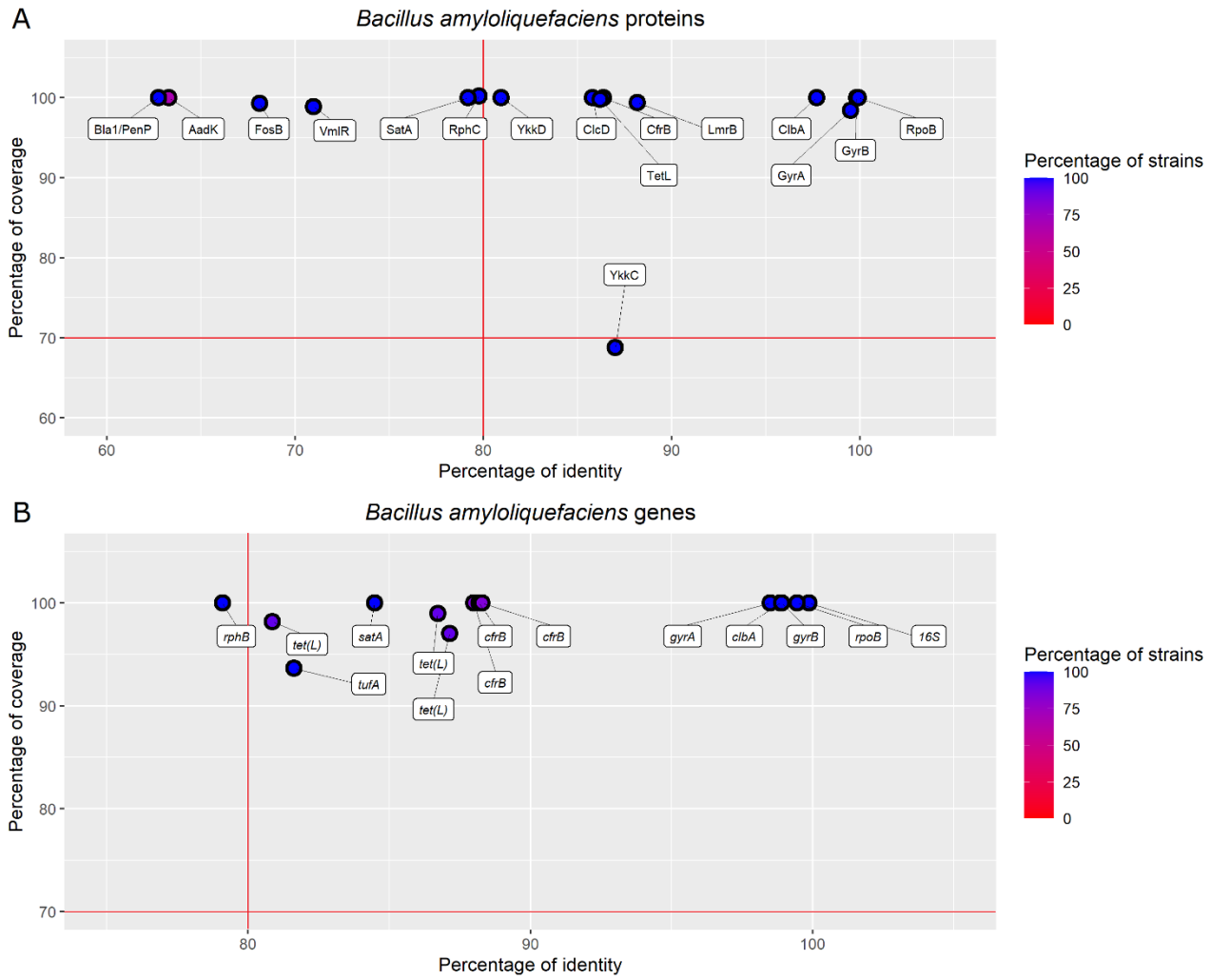


Figure 5: Median identity and coverage of matches to AMR genes found among the publicly available genomes of *B. amyloliquefaciens* included in this analysis. Searches based on NCBI accession codes to amino acid (A) and nucleotide (B) sequences. The percentage of analysed genomes containing matches is depicted with colours. Thresholds for identity (80%) and coverage (70%) as indicated in EFSA BIOHAZ Panel, 2023 are indicated with red lines. Dots with the same gene symbol correspond to different allelic variants (see Table 4).

As observed in *B. subtilis, rphB* did not match above thresholds to any screened genome but was found in most of them with a median identity of 79.1% (Figure 5B). A similar situation was found for *satA*, a streptothricin N-acetyltransferase: whereas the hit to the nucleotide sequence (NG_064662.1) was found above the thresholds in nearly all genomes (Figure 5B), the hit to the

amino acid sequence (WP_003242546.1) only matched above thresholds in ca. 12% of the genomes but it matched to the majority of genomes very close to the threshold of identity (median identity 79.2% Figure 5A). Similarly, another hit corresponding to a rifamycin phosphotransferase (*rphC*, from *Brevibacillus brevis*), which did not match to any *B. amyloliquefaciens* genome above 80% identity, matched to almost all genomes at 79.8% median identity (Figure 4A).

### *4.2.4 Bacillus licheniformis*

All analysed genomes of *B. licheniformis* were confirmed using the threshold of 95% ANI.

All hits were found with identity and coverage above thresholds in all or nearly all the analysed genomes (Table 5, Figure 6) No hits were found in plasmids.

Table 5: Hits to AMR genes reported in dossiers in the publicly available genomes of *B. licheniformis*. Grey rows correspond to amino acid sequences, while yellow rows to nucleic acid sequences. Internal positive and negative control sequences are depicted in green and red text, respectively.

| Accession number* | Gene symbol* | Description as indicated in NCBI | MIA[a] | Frequency of genomes with at least one hit (%) | Median identity (%) | Median coverage (%) |
|---|---|---|---|---|---|---|
| WP_142782184.1 | *apH* | *B. licheniformis* phosphotransferase | Yes | 100 | 99.4 | 100 |
| WP_003178584.1 | *blaP/ penP* | *B. licheniformis* class A beta-lactamase | Yes | 100 | 100 | 100 |
| WP_009327810.1 | *catH* | *Bacillus* type A chloramphenicol O-acetyltransferase | Yes | 100 | 97.3 | 100 |
| WP_003180775.1 | *lmrA* | *Bacillus* TetR/AcrR family transcriptional regulator | Yes | 100 | 100 | 100 |
| WP_011197798.1 | *lmrB* | *Bacillus* DHA2 family efflux MFS transporter permease subunit | Yes | 100 | 100 | 100 |
| WP_087347987.1 | *rphC* | *Brevibacillus brevis* rifamycin-inactivating phosphotransferase | Yes | 97.5 | 80.1 | 100 |
| ATY38581.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA[b] | 0 | 0 | 0 |
| WP_003178095.1 | *gyrA* | *B. licheniformis* strain SCDB 14 DNA gyrase subunit A | NA | 100 | 100 | 98.1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| WP_009330043.1 | *gyrB* | *B. licheniformis* strain SCDB 14 DNA gyrase subunit B | NA | 100 | 100 | 100 |
| WP_003178309.1 | *rpoB* | *B. licheniformis* strain SCDB 14 RNA polymerase subunit beta | NA | 100 | 99.8 | 100 |

* As provided in dossiers.
MIA: medically important antimicrobial (WHO, 2024).
NA: non-applicable.

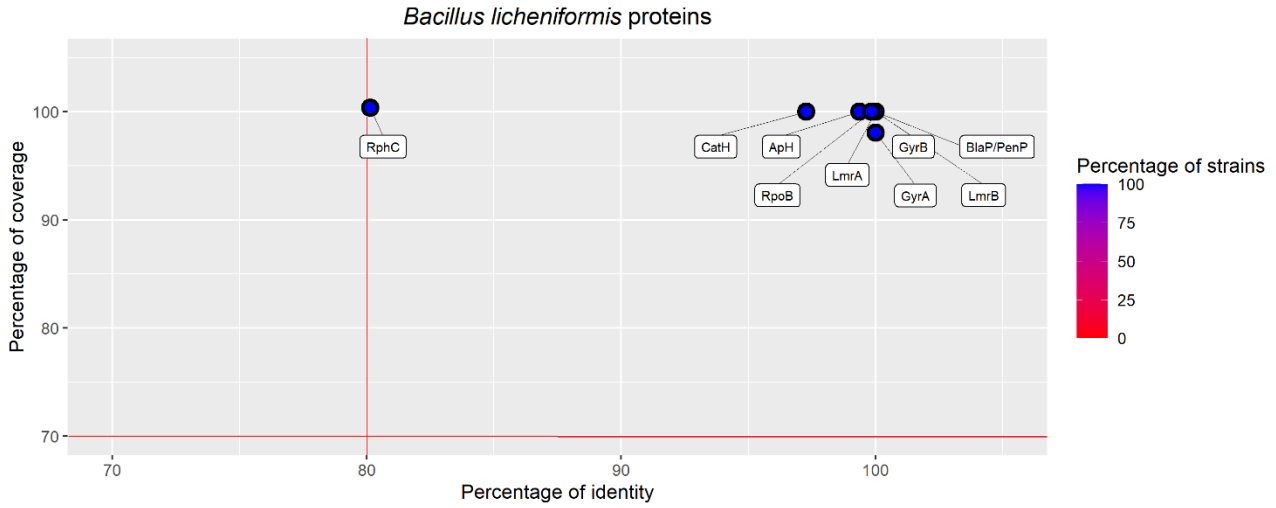


Figure 6: Median identity and coverage of matches to AMR genes found in *B. licheniformis* publicly available genomes. Searches based on NCBI accession codes to amino acid sequences. The percentage of analysed genomes containing matches is depicted with colours. The threshold for identity (80%) and coverage (70%) as indicated in EFSA BIOHAZ Panel, 2023 is shown with red lines.

### *4.2.5 Bacillus paralicheniformis*

All analysed genomes of *B. paralicheniformis* were confirmed using the threshold of 95% ANI.

Table 6 shows the results of the BLAST searches for the hits to AMR genes identified in strains from dossiers against the publicly available genomes of *B. paralicheniformis*.

Table 6: Hits to AMR genes reported in dossiers in the publicly available genomes of *B. paralicheniformis*. Grey rows correspond to amino acid sequences, while yellow rows to nucleic acid sequences. Internal positive and negative control sequences are depicted in green and red text, respectively.

| Accession number* | Gene symbol* | Description as indicated in NCBI | MIA[(a)] | Frequency of genomes with at least one hit (%) | Median identity (%) | Median coverage (%) |
|---|---|---|---|---|---|---|
| WP_023855962.1 | *bcrA* | *B. paralicheniformis* strain Bac84 bacitracin resistance | Yes | 100 | 99.7 | 100 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | ABC transporter subunit | | | | |
| WP_041817242.1 | *bcrA* | *B. paralicheniformis* strain ATCC 9945a bacitracin resistance ABC transporter subunit | Yes | 100 | 98.7 | 100 |
| AAA99505.1 | *bcrB* | *B. paralicheniformis* strain ATCC 9945a bacitracin resistance ABC transporter subunit | Yes | 100 | 98.6 | 100 |
| WP_023855961.1 | *bcrB* | *B. paralicheniformis* strain Bac84 bacitracin resistance ABC transporter subunit | Yes | 100 | 99.6 | 100 |
| AGN36968.1 | *bcrC* | *B. paralicheniformis* strain ATCC 9945a bacitracin resistance ABC transporter subunit | Yes | 100 | 98.4 | 94.1 |
| WP_023855960.1 | *bcrC* | *B. paralicheniformis* strain Bac84 bacitracin resistance ABC transporter subunit | Yes | 100 | 100 | 94.1 |
| WP_023857755.1 | *blaP* | *B. paralicheniformis* class A beta-lactamase | Yes | 100 | 100 | 100 |
| WP_025811090.1 | *ermD* | *Bacillus* 23S rRNA (adenine(2058)-N(6))-methyltransferase | Yes | 96.2 | 100 | 100 |
| M29832.1 | *ermD* | *Bacillus* 23S rRNA (adenine(2058)-N(6))-methyltransferase gene | Yes | 96.2 | 99.5 | 100 |
| WP_063844769.1 | *ermD* | *B. anthracis* 23S rRNA (adenine(2058)-N(6))-methyltransferase | Yes | 96.2 | 99.7 | 100 |
| WP_087347987.1 | *rphC* | *Brevibacillus brevis* rifamycin-inactivating phosphotransferase | Yes | 30.8 | 80.4 | 100 |

| | | | | | | |
|---|---|---|---|---|---|---|
| NZ_CP014842.1<br>3306836-3308385 | 16S rRNA | *B. paralicheniformis* strain Bac84 16S ribosomal RNA | NA[(b)] | 100 | 99.6 | 100 |
| ATY38581.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | 0 | 0 |
| KY367272.1 | *fimA* | *Salmonella enterica* serovar Dublin strain 67 fimbrin | NA | 0 | NA | NA |
| WP_023856793.1 | *gyrA* | *B. paralicheniformis* strain Bac84 DNA gyrase subunit A | NA | 96.0 | 100 | 98.1 |
| NZ_CP014842.1<br>33726-36194 | *gyrA* | *B. paralicheniformis* strain Bac84 DNA gyrase subunit A | NA | 100 | 96.0 | 100 |
| WP_031305110.1 | *gyrB* | *B. paralicheniformis* strain Bac84 DNA gyrase subunit B | NA | 96.0 | 100 | 100 |
| NZ_CP014842.1 | *gyrB* | *B. paralicheniformis* strain Bac84 DNA gyrase subunit B | NA | 100 | 96.1 | 100 |
| WP_020449852.1 | *rpoB* | RNA polymerase subunit beta | NA | 100 | 99.9 | 100 |
| NZ_CP014842.1 | *rpoB* | *B. paralicheniformis* strain Bac84 DNA gyrase subunit B | NA | 100 | 96.9 | 100 |

* As provided in dossiers.
MIA: medically important antimicrobial (WHO, 2024).
NA: non-applicable

Similarly to *B. licheniformis* and *B. velezensis*, all reported hits were found above thresholds, with the only exception of *rphC*, that matched to all genomes with 79.9% median identity (Figure 7). The other hits reported were the genes involved in bacitracin resistance, which are typical in this species, and *ermD*, a macrolide-lincosamide-streptogramin resistance gene, described in *B. anthracis* and *B. licheniformis*. However, respect to the latter, identical proteins are found mainly in strains reported as *B. paralicheniformis* in the NCBI database (data not shown).

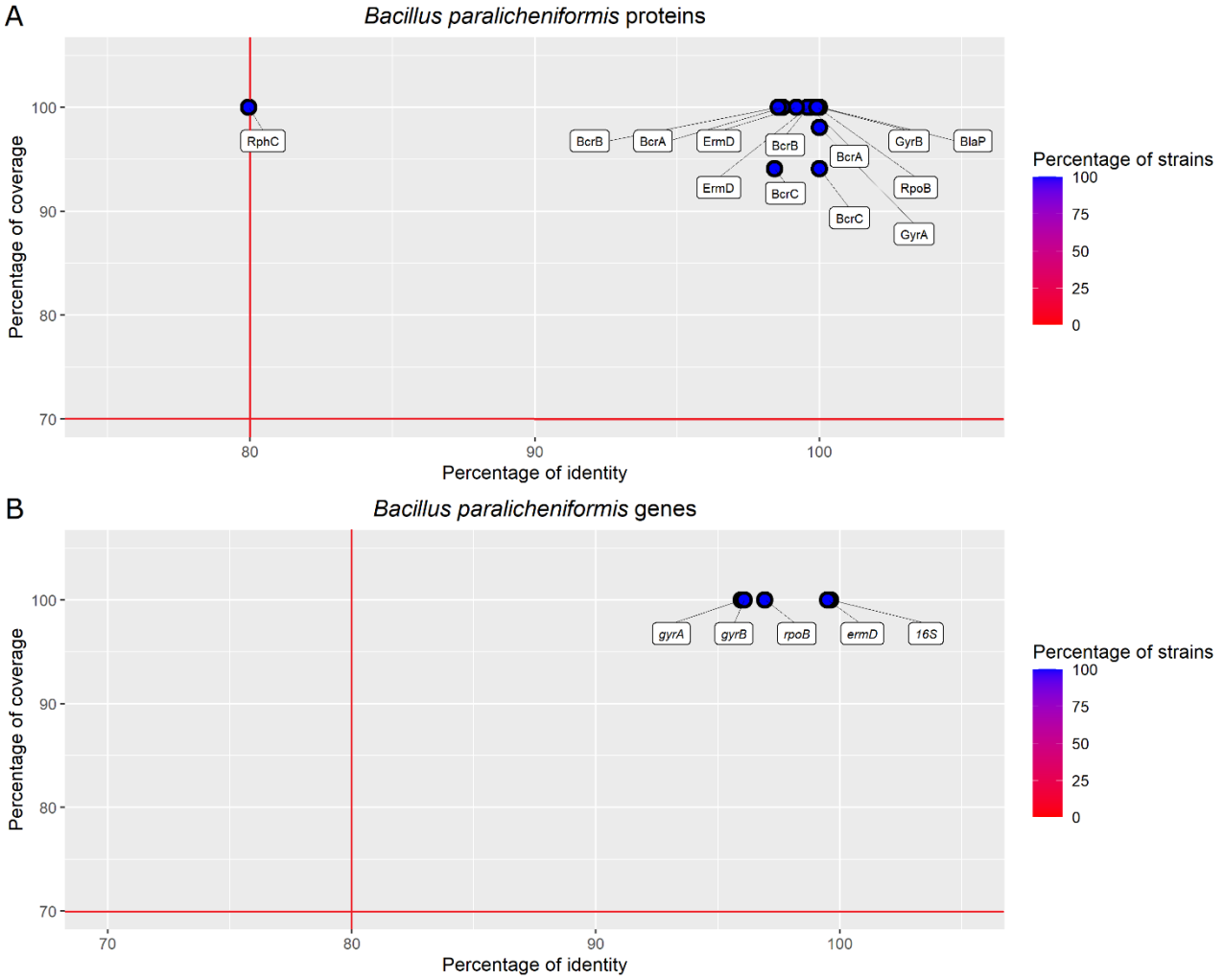


Figure 7: Median identity and coverage of matches to AMR genes found in *B. paralicheniformis* publicly available genomes. Searches based on NCBI accession codes to amino acid (A) and nucleotide (B) sequences. The percentage of analysed genomes containing matches is depicted with colours. Thresholds for identity (80%) and coverage (70%) as indicated in EFSA BIOHAZ Panel, 2023 are indicated with red lines. Dots with the same name correspond to different allelic variants (Table 6).

# 5 Conclusion

The presence of an AMR gene in the majority of genomes of strains belonging to a given species is one of the criteria used by EFSA to determine if that gene can be regarded as intrinsic and therefore of no safety concern. A bioinformatic pipeline was designed and applied to produce a catalogue of hits to AMR genes reported in *Bacillus* strains from dossiers for regulated products submitted to EFSA. The catalogue indicates the prevalence of each gene in publicly available, complete and verified genomes of selected species. This catalogue aims to strengthen and harmonise the assessment of such genes. The bioinformatics pipeline designed to produce this catalogue, which is made available, can be used for the assessment of any other AMR gene in any species for which a sufficient number of genomes is available.

# References

Chun BH, Kim KH, Jeong SE, Jeon CO, 2019. Genomic and metabolic features of the Bacillus amyloliquefaciens group– B. amyloliquefaciens, B. velezensis, and B. siamensis– revealed by pan-genome analysis, Food Microbiology, 77, 146–157. doi: 10.1016/j.fm.2018.09.001.

Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S and Trujillo ME, 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. International Journal of Systematic and Evolutionary Microbiology, 68, 61–466.

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D, Bover-Cid S, Chemaly M, De Cesare A, Hilbert F, Lindqvist R, Nauta, M, Nonno R, Peixe L, Ru G, Simmons M, Skandamis P, Suffredini E, Cocconcelli PS, Suarez JE, …Herman L, 2023. Statement on how to interpret the QPS qualification on 'acquired antimicrobial resistance genes'. EFSA Journal, 21, 1–13. doi: 10.2903/j.efsa.2023.8323

Gould AL, Henderson JB, 2023. Comparative genomics of symbiotic Photobacterium using highly contiguous genome assemblies from long read sequences. Microbial Genomics, 9, 001161. doi: 10.1099/mgen.0.001161.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis K, Aluru S, 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications, 9, 5114. doi: 10.1038/s41467-018-07641-9

McKenzie T, Hoshino T, Tanaka T, Sueoka N, 1986. The nucleotide sequence of pUB110: some salient features in relation to replication and its regulation. Plasmid, 15, 93–103. doi: 10.1016/0147-619x(86)90046-6

# Abbreviations

| | |
|---|---|
| AMR | Antimicrobial resistance |
| NCBI | National Center of Biotechnology Information |
| QPS | Qualified Presumption of Safety |
| WGS | Whole Genome Sequence |
| ANI | Average Nucleotide Identity |
| BLAST | Basic Local Alignment Search Tool |
| MIA | Medically important antimicrobial |

# Annex A Hits included in the analysis

**Table 1:** Hits to AMR genes extracted from dossiers provided by the applicants to EFSA. Invalid matches are those not corresponding to any sequence in the NCBI database.

| Species | Total matches | Invalid matches | Valid matches without duplications | Hits (matches above threshold) included in the analysis |
|---|---|---|---|---|
| *B. velezensis* | 28 | 0 | 28 | 10 |
| *B. subtilis* | 151 | 2 | 54 | 42 |
| *B. amyloliquefaciens* | 124 | 1 | 56 | 26 |
| *B. licheniformis* | 39 | 1 | 38 | 6 |
| *B. paralicheniformis* | 16 | 0 | 16 | 11 |

# Annex B Cluster analysis and phylogenetic trees of the analysed species

The figures are available in the supplementary information.

Figure 1. FastANI-based cluster map (A) and phylogenetic tree (B) of the *B. velezensis* complete genomes available at the RefSeq database (NCBI). Genomes are indicated with their respective NCBI accession numbers, except reference genomes that are indicated in A with a red line (*B. velezensis*) and green lines (*B. subtilis*, *B. licheniformis* and *B. paralicheniformis*), and in B with the strain name.

Figure 2. FastANI-based cluster map (A) and phylogenetic tree (B) of the *B. subtilis* complete genomes available at the RefSeq database (NCBI). Genomes are indicated with their respective NCBI accession numbers, except reference genomes that are indicated in A with a red line (*B. subtilis*) and green lines (*B. paralicheniformis*, *B. licheniformis* and *B. amyloliquefaciens*), and in B with the strain name.

Figure 3. FastANI-based cluster map (A) and phylogenetic tree (B) of the *B. amyloliquefaciens* complete genomes available at the RefSeq database (NCBI). Genomes are indicated with their respective NCBI accession numbers, except reference genomes that are indicated with the strain name. Nineteen genomes showed an ANI of <95% with respect to the reference genome GKT04 (orange sectors). These sequences were excluded from the analysis following the criteria for species confirmation adopted for this study.

Figure 4. FastANI-based cluster map (A) and phylogenetic tree (B) of the *B. licheniformis* complete genomes available at the RefSeq database (NCBI). Genomes are indicated with their respective NCBI accession numbers, except reference genomes that are indicated with the strain name.

Figure 5. FastANI-based cluster map (A) and phylogenetic tree (B) of the *B. paralicheniformis* complete genomes available at the RefSeq database (NCBI). Genomes are indicated with their respective NCBI accession numbers, except reference genomes that are indicated with the strain name.

# Annex C Distribution of matches to AMR genes in publicly available genomes

Figure 1. Distribution of identity (left) and coverage (right) of matches to AMR genes among publicly available genomes of *B. velezensis*. Searches based on accession numbers to (A) amino acid or (B) nucleotide sequences.

Figure 2. Distribution of identity (left) and coverage (right) of matches to AMR genes among publicly available genomes of *B. subtilis*. Searches based on accession numbers to (A) amino acid or (B) nucleotide sequences.

Figure 3. Distribution of identity (left) and coverage (right) of matches to AMR genes among publicly available genomes of *B. amyloliquefaciens*. Searches based on accession numbers to (A) amino acid or (B) nucleotide sequences.

Figure 4. Distribution of identity (left) and coverage (right) of matches to AMR genes among publicly available genomes of *B. licheniformis*. Searches based on accession numbers to amino acid sequences.

Figure 5. Distribution of identity (left) and coverage (right) of matches to AMR genes among publicly available genomes of *B. paralicheniformis*. Searches based on accession numbers to (A) amino acid or (B) nucleotide sequences.

# Annex D Pipeline for the automated analysis of gene distribution in microbial species

The pipeline is available at the EFSA Knowledge Junction (https://zenodo.org/records/12608405).