Supplementary material

Supplementary text

Culture conditions

Strains were isolated from clinical material following established decontamination procedures. Bacteria were cultured in BD MGIT 960 tubes and Middlebrook 7H10 agar media. Strains collected as a part of the IeDEA study were cultured on Löwenstein-Jensen slants and sent to the National Center for Mycobacteria (University of Zürich, Zürich, Switzerland) for phenotypic drug susceptibility testing.

Phenotypic drug susceptibility testing – BD MGIT 960

DST for first- and second-line anti-TB drugs was performed using the Becton Dickinson (BD) MGIT 960 and EpiCenter devices equipped with the TB eXiST module (BD, Franklin Lakes, NJ, USA) (1). The MGIT 960 system was used for primary bacterial isolation and DST for first-line drugs (rifampin, isoniazid, ethambutol, pyrazinamide) as recommended by the manufacturer, using currently established epidemiological cut-offs (ECOFF – epidemiological cut-off, i.e the highest observed wild-type MIC (2)) summarized in Table 1 in the main text. MGIT tubes enriched with 0.8 ml of supplement (MGIT 960 OADC supplement; BD) were inoculated with 0.1 ml of the drug solution (supplementary Table 2) and 0.5 ml of the test strain suspension. For the drug-free growth control tube, the *M. tuberculosis* suspension was diluted 1:100 in sterile saline and 0.5 ml was inoculated into the tube (proportion testing).

Phenotypic drug susceptibility testing - Proportion method by 7H10 agar dilution

M. tuberculosis strains were grown in Middlebrook 7H9 liquid broth with 10% OADC supplement (BD) until McFarland 0.5 was reached. Using a replicator micropipettor robot VIAFLOW96 (Integra Biosciences, Zizers, Switzerland), 1 µl of the culture suspension was inoculated onto Middlebrook 7H10 agar plates with twofold serially diluted drug concentrations to determine the MIC (supplementary Table 2), whereby the MIC is defined as the lowest drug concentration that inhibits growth of more than 99% the bacteria during 21 days of incubation at 37°C (3). Agar plates were read automatically using an AID Microplate reader and automated software developed by AID (AID Diagnostika, Strassberg, Germany). 7H10 agar dilution MICs of the reference strains differed between the different

experimental runs by more than one dilution step for capreomycin, kanamycin A, amikacin and isoniazid. For these antibiotics, the MICs for all strains from the last run were rescaled by the differences observed for the reference strains. Phenotypic DST for all strains was performed at the Swiss National Center for Mycobacteria (University of Zürich, Zürich, Switzerland).

Whole genome sequencing

Libraries were prepared using the Illumina Nextera XT kit and sequenced on an Illumina HiSeq 2500, generating 125 bp paired-end reads. Sequencing was performed at the genomics facility of the ETHZ/University of Basel in Basel, Switzerland and at the Broad Institute, Cambridge, Massachusetts, United States.

The raw data was processed with an in-house python pipeline as follows: reads were adaptor clipped and quality trimmed with Trimmomatic (v.0.33), whereby resulting reads < 20 bp were discarded. Overlapping paired-end reads were merged using SeqPrep (https://github.com/jstjohn/SeqPrep). The processed reads were subsequentltly mapped to a reconstructed hypothetical MTBC ancestor (4) with BWA (v.0.7.12) (5). Duplicated reads were marked with Picard (v.2.1.1) (https://github.com/broadinstitute/picard) using the MarkDuplicates module. Local realignments of reads around indels were performed with the GATK (v.3.4.0) modules RealignerTargetCreator and IndelRealigner (6). Pileups were generated with Samtools (v.1.2) (7) and SNPs were subsequently called with VarScan (v.2.4.1) (8) using the following thresholds: minimum mapping & minimum base quality of 20, minimum read depth of 7X at a given position. For a SNP to be called, the alternative base call needed to be supported by at lest 5 reads without strand bias. Furthermore, SNPs were considered as fixed when they reached a frequency of ≥ 90%. The position was called as ancestral when the frequency was found to be < 10%. SNPs were annotated with SnpEff (v.4.11) (9) corresponding to the *M. tuberculosis* H37Rv reference annotation (NC_000962.3). SNPs in regions that share a minimum of 50 bp of sequence identity with other regions in the genome were excluded (10).

Variable SNP alignment and phylogenetic analysis

A variable SNP alignment was generated by concatenating all filtered SNPs in the dataset, whereby the IUPAC nucleotide ambiguity codes were used for unfixed positions (10% ≤ variant frequency < 90%). Positions were considered variable if at least one genome had

62 SNP-call at the position in question. If the SNP fell into an excluded region (see above) or
63 was covered by less than 7 reads it was encoded as X in the alignment. If there was no
64 sequence information at all available for a position, it was encoded as a gap. Furthermore,
65 positions known to be involved in drug resistance were not considered in the alignment. The
66 variable SNP alignment was used to infer a maximum likelihood phylogeny with RAxML
67 (v.8.2.8) (11), using the general time-reversible model of sequence evolution and branch
68 support values were inferred by bootstrapping the highest scoring maximum likelyhood tree
69 (1000 pseudoreplicates). *Mycobacterium canettii* (SRR011186) was used to root the
70 phylogeny. The genomes were classified into main and sub lineages based on the presence of
71 previously established markers (12). Average genetic distances between strains based on
72 fixed single nucleotide polymorphisms (SNP) were calculated with the ape package (v.4.1)
73 (13) for R (v.3.3.3) (14). Two strains were defined as clustered if their average genetic
74 distance was ≤ 12 SNPs (15) and one strain was subsequently omitted from the analysis at
75 random. Phylogenetic trees with associated metadata were visualized with the R package
76 ggtree (v.1.6.11) (16).

77 **WGS-based resistance profile inferrence**

78 The WGS data was screened for non-synonymous mutations and indels in genes known to be
79 involved in drug resistance for the drugs assayed in *M. tuberculosis* (17, 18). Phylogenetic
80 markers in drug resistance-associated genes (Table 2 – main text) shared by all strains
81 belonging to a main- or sublineage (12, 19) (supplementary Table 1), as well as previously
82 described phylogenetic markers (20), including main lineage markers derived from an
83 unpublished collection of 400 phylogenetically diverse strains, were removed. After filtering
84 for phylogenetic markers, all fixed non-synonymous polymorphisms in resistance-related
85 genes were treated as linked to drug resistance. Where necessary, WGS results were
86 complemented by molecular DST approaches as described previously (21).

87 **Data availability**

88 The datasets generated and analysed in the study were deposited at the National Center
89 National Center for Biotechnology Information (NCBI) under the BioProject IDs
90 PRJNA454477 - http://www.ncbi.nlm.nih.gov/bioproject/454477 and PRJNA300846 -
91 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA300846. Supplementary Table S3 provides a
92 summary of sample identifyers and accession numbers and drug resistance information. The

93 quantitative readout of the 7H10 agar dilution-based DST and the results of the MGIT 960-
94 based testing at the WHO-defined critical concentration are available upon request.

95 **Supplementary Tables**

96 Supplementary Table S1 List of major phylogenetic markers in drug-resistance related genes

Gene	Rv number	Position	Ref NT	Alt NT	Codon change	Lineage	Coll (12) sublineage	Comments
embC	Rv3793	4242182	G	T	Ala774Ser	L4	L4.3.3	Sublineage of L4.3.3
gid	Rv3919c	4407967	T	C	Leu79Ser	L4	L4.1	Sublineage of L4.1

97
98
99
100

Supplementary Table S2 List of drug concentrations tested for MGIT 960 and 7H10 agar dilution for the training and test set. ND = not determined

Antibiotic	**MGIT 960 concentrations (mg/L) Training set (n = 56)**	**MGIT 960 concentrations (mg/L) Test set (n = 120)**	**7H10 agar dilution concentration ranges (mg/L), $\log_2$ steps, Training set (n = 56)**	**7H10 agar dilution concentration, ranges (mg/L), $\log_2$ steps, Test set (n = 120)**
Ethionamide	1.25, 2.5, 5, 10, 25	ND	0.25-128	0.25-256
Ethambutol	1.25, 2.5, 5, 12.5, 50	5	0.5-64	0.5-64
Capreomycin	2.5, 5, 25	ND	0.5-64	0.5-64
Streptomycin	1, 4, 20	ND	0.06-128	0.06-128
Kanamycin A	1, 2, 4, 20	ND	0.06-128	0.06-128
Amikacin	1, 4, 20	1	0.06 - 128	0.06-128
Moxifloxacin	0.25, 0.5, 2.5, 7.5	0.25	0.016-32	0.016-32
Isoniazid	0.1, 1, 3, 10	0.1, 1	0.004-32	0.004-32
Rifampin	1, 4, 20	1	0.016-256	0.016-256
Rifabutin	0.1, 0.4, 2	ND	0.004-32	0.004-32
Pyrazinamide	100	100	ND	ND

101 ## Supplementary References

102 1. Springer B, Lucke K, Calligaris-Maibach R, Ritter C, Bottger EC. 2009. Quantitative
103 Drug Susceptibility Testing of Mycobacterium tuberculosis by Use of MGIT 960 and
104 EpiCenter Instrumentation. J Clin Microbiol 47:1773–1780.

105 2. Ängeby K, Juréen P, Kahlmeter G, Hoffner S, Schön T. 2012. Challenging a dogma:
106 antimicrobial susceptibility testing breakpoints for Mycobacterium tuberculosis. Bull
107 World Health Organ 90:693–698.

108 3. Sirgel FA, Wiid IJF, van Helden PD. 2009. Measuring Minimum Inhibitory
109 Concentrations in Mycobacteria, p. 173–186. *In* Parish T., BA (ed.), Mycobacteria
110 Protocols. Methods in Molecular Biology (Methods and Protocols). Humana Press,
111 Totowa, NJ.

112 4. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD,
113 Gagneux S. 2010. Human T cell epitopes of Mycobacterium tuberculosis are
114 evolutionarily hyperconserved. Nat Genet 42:498–503.

115 5. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
116 transform. Bioinformatics 25:1754–1760.

117 6. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella
118 K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis
119 Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing
120 data. Genome Res 20:1297–1303.

121 7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
122 Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics
123 25:2078–2079.

124 8. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller C a, Mardis
125 ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number
126 alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.

127 9. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden
128 DM. 2012. A program for annotating and predicting the effects of single nucleotide
129 polymorphisms, SnpEff. Fly (Austin) 6:80–92.

130 10. Stucki D. 2015. Transmission and evolution of Mycobacterium tuberculosis studied by
131 whole genome sequencing and single nucleotide polymorphism-typing.

132 11. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
133 analysis of large phylogenies. Bioinformatics 30:1312–1313.

134 12. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M,
135 Portugal I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing
136 Mycobacterium tuberculosis complex strains. Nat Commun 5:4812.

137 13. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution
138 in R language. Bioinformatics 20:289–290.

139 14. R Core Team. 2013. R: A Language and Environment for Statistical Computing. R
140 Foundation for Statistical Computing, Vienna, Austria.

141 15. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson
142 DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P,
143 Smith EG, Peto TE. 2013. Whole-genome sequencing to delineate Mycobacterium
144 tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13:137–
145 146.

146 16. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an R package for
147 visualization and annotation of phylogenetic trees with their covariates and other
148 associated data. Methods Ecol Evol 8:28–36.

149 17. Gygli SM, Borrell S, Trauner A, Gagneux S. 2017. Antimicrobial resistance in
150 Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. FEMS
151 Microbiol Rev 41:354–373.

152 18. Safi H, Lingaraju S, Amin A, Kim S, Jones M, Holmes M, McNeil M, Peterson SN,
153 Chatterjee D, Fleischmann R, Alland D. 2013. Evolution of high-level ethambutol-
154 resistant tuberculosis through interacting mutations in decaprenylphosphoryl-β-D-
155 arabinose biosynthetic and utilization pathway genes. Nat Genet 45:1190–1197.

156 19. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihwa L,
157 Borrell S, Luo T, Gao Q, Kato-Maeda M, Ballif M, Egger M, Macedo R, Mardassi H,
158 Moreno M, Tudo Vilanova G, Fyfe J, Globan M, Thomas J, Jamieson F, Guthrie JL,

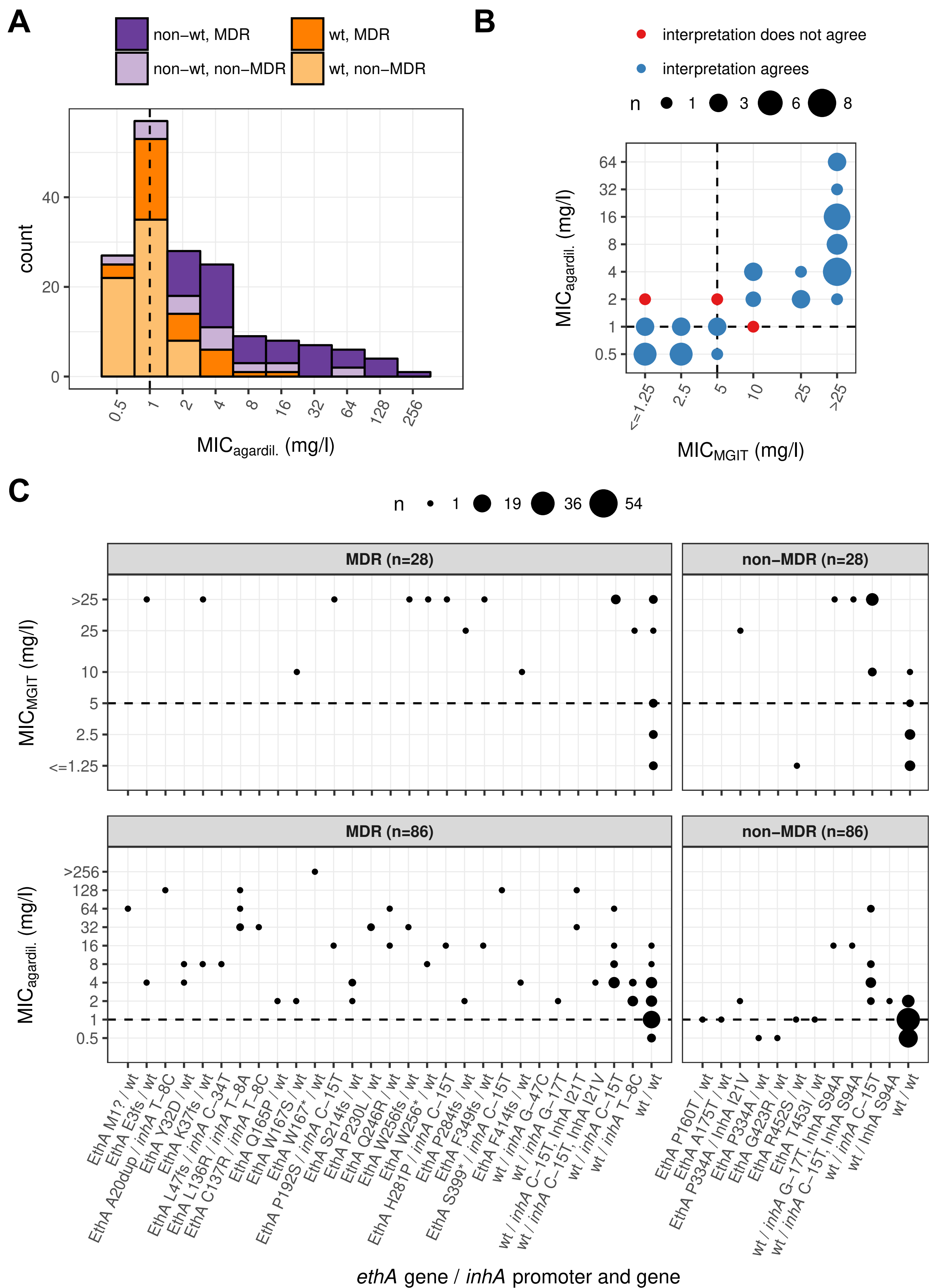
159 Asante-Poku A, Yeboah-Manu D, Wampande E, Ssengooba W, Joloba M, Henry
160 Boom W, Basu I, Bower J, Saraiva M, Vaconcellos SEG, Suffys P, Koch A,
161 Wilkinson R, Gail-Bekker L, Malla B, Ley SD, Beck H-P, de Jong BC, Toit K,
162 Sanchez-Padilla E, Bonnet M, Gil-Brusola A, Frank M, Penlap Beng VN, Eisenach K,
163 Alani I, Wangui Ndung'u P, Revathi G, Gehre F, Akter S, Ntoumi F, Stewart-
164 Isherwood L, Ntinginya NE, Rachow A, Hoelscher M, Cirillo DM, Skenders G,
165 Hoffner S, Bakonyte D, Stakenas P, Diel R, Crudu V, Moldovan O, Al-Hajoj S, Otero
166 L, Barletta F, Jane Carter E, Diero L, Supply P, Comas I, Niemann S, Gagneux S.
167 2016. Mycobacterium tuberculosis lineage 4 comprises globally distributed and
168 geographically restricted sublineages. Nat Genet 48:1535–1543.

169 20. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G,
170 Mallard K, Nair M, Miranda A, Alves A, Perdigão J, Viveiros M, Portugal I, Hasan Z,
171 Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-
172 tuberculosis drug resistance from whole-genome sequences. Genome Med 7:51.

173 21. Ritter C, Lucke K, Sirgel FA, Warren RW, van Helden PD, Bottger EC, Bloemberg G
174 V. 2014. Evaluation of the AID TB Resistance Line Probe Assay for Rapid Detection
175 of Genetic Alterations Associated with Drug Resistance in Mycobacterium
176 tuberculosis Strains. J Clin Microbiol 52:940–946.

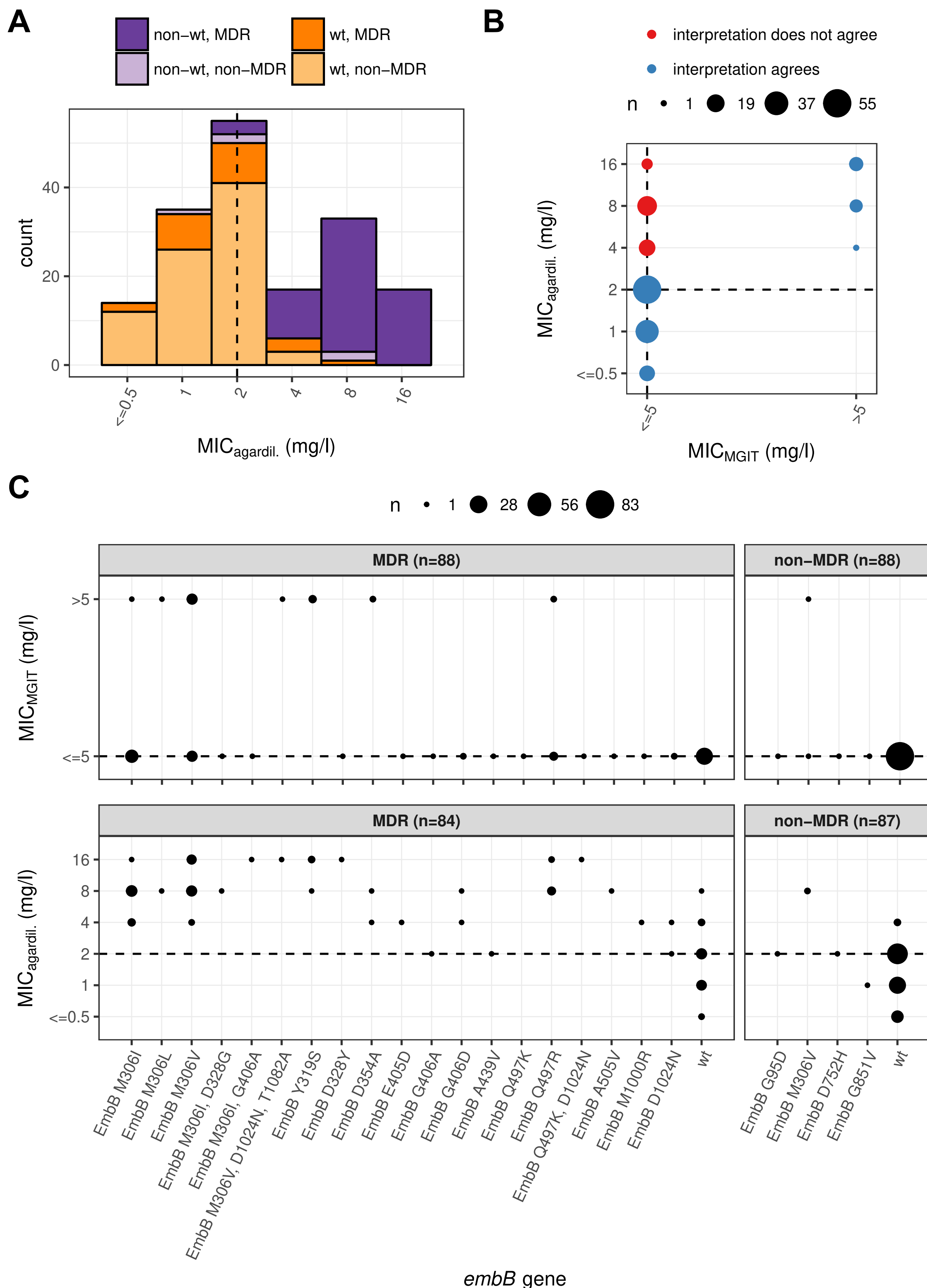
177

Ethionamide (ETH)



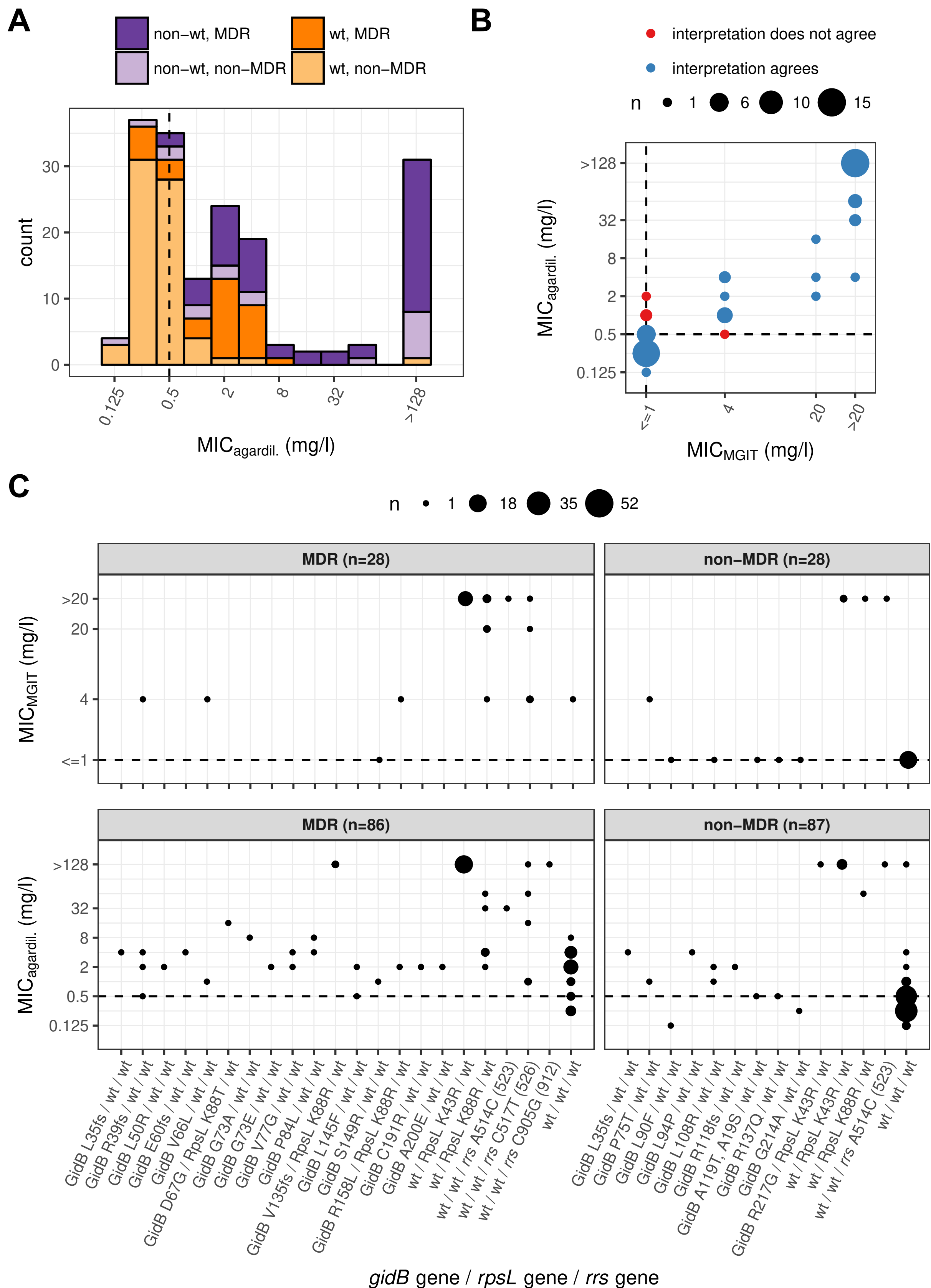
Supplementary Figure S1 - **A** Histogram of 7H10 agar dilution MICs for **ethionamide** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Top panel MGIT 960, bottom panel 7H10 agar dilution. "M1?" indicates an initiator codon variant, "fs" indicates a frame shift variant, "*" indicates a stop-gain variant. In all panels dashed lines indicate the ECOFF (see Table 1 in the main text).

Ethambutol (EMB)



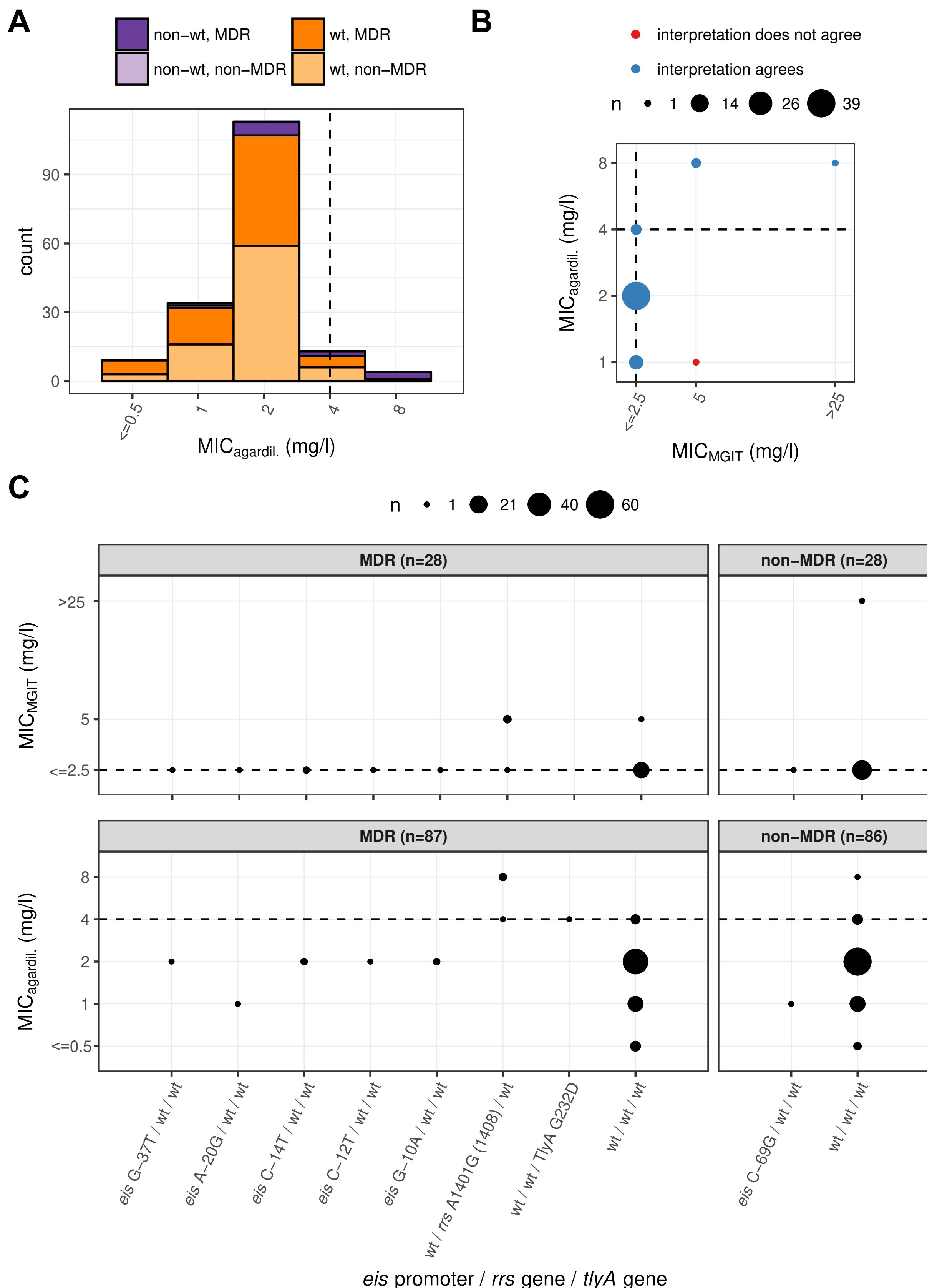
Supplementary Figure S2 - **A** Histogram of 7H10 agar dilution MICs for **ethambutol** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Top panel MGIT 960, bottom panel 7H10 agar dilution. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Streptomycin (STR)



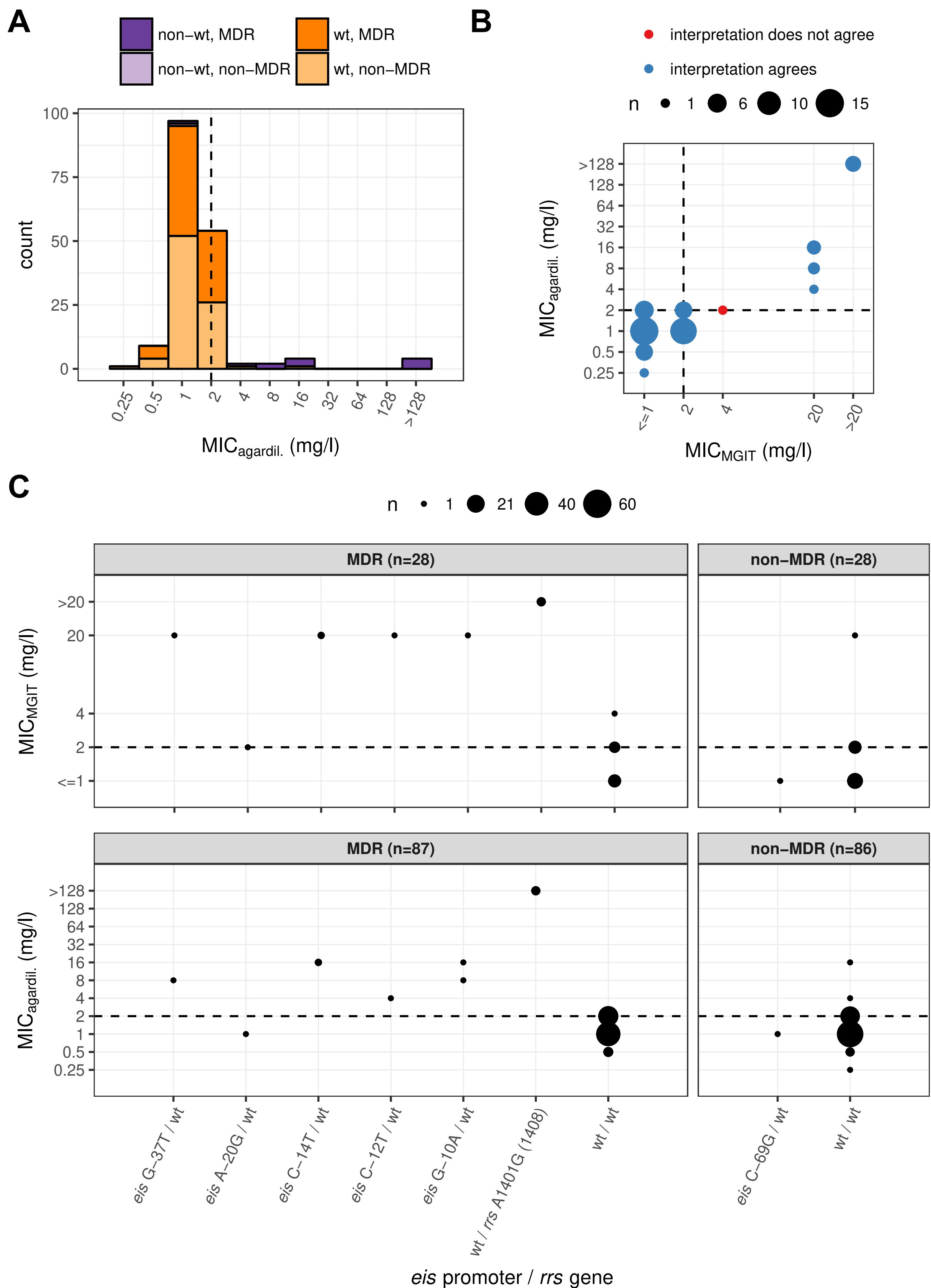
Supplementary Fgure S3 - **A** Histogram of 7H10 agar dilution MICs for **streptomycin** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Numbers in parentheses indicate the corresponding *Escherichia coli* nucleotide numbering. Top panel MGIT 960, bottom panel 7H10 agar dilution. "fs" indicates a frame shift variant. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Capreomycin (CAP)



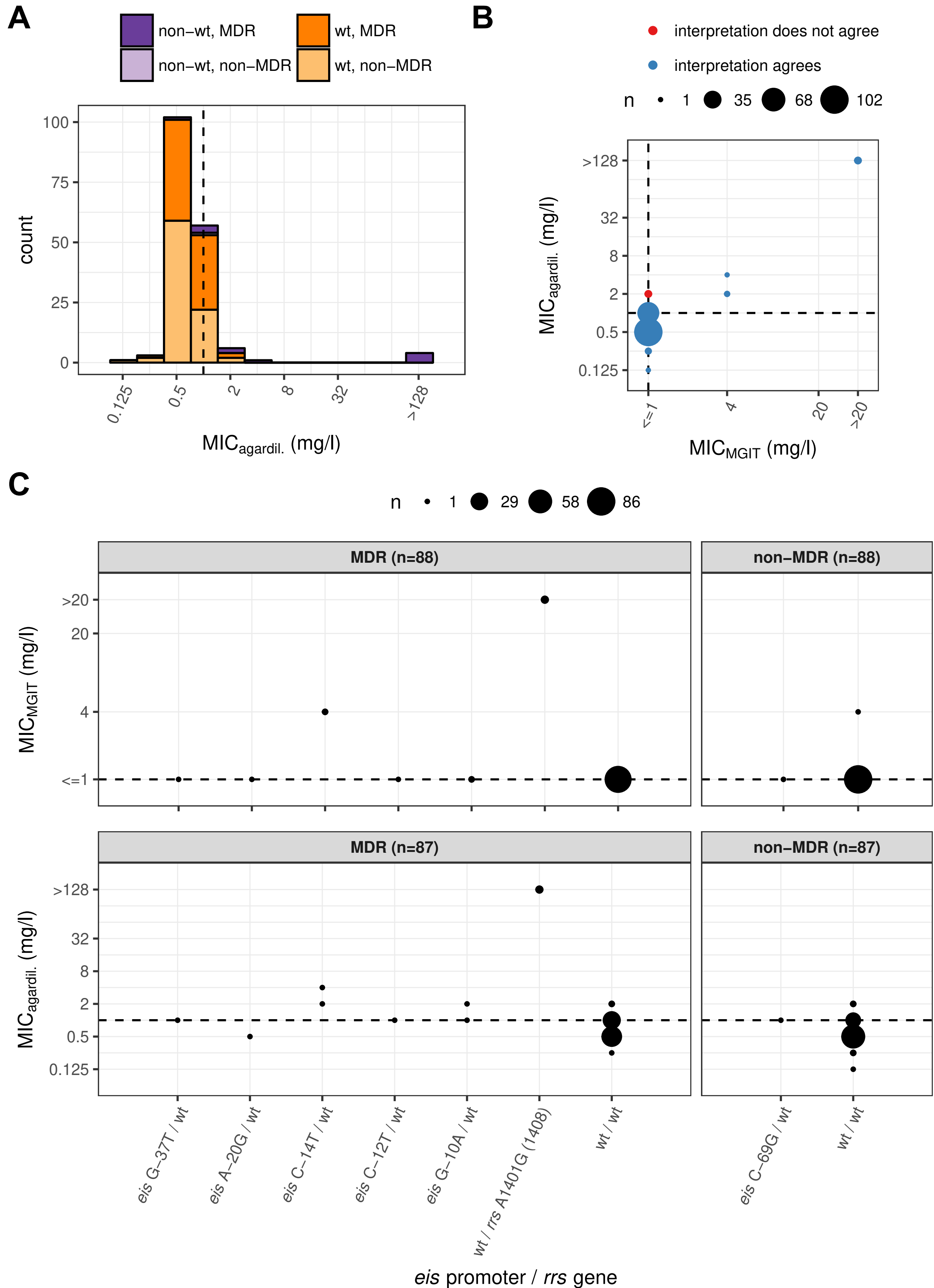
Supplementary Figure S4 - **A** Histogram of 7H10 agar dilution MICs for **capreomycin** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Numbers in parentheses indicate the corresponding *Escherichia coli* nucleotide numbering. Top panel MGIT 960, bottom panel 7H10 agar dilution. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Kanamycin A (KAN)



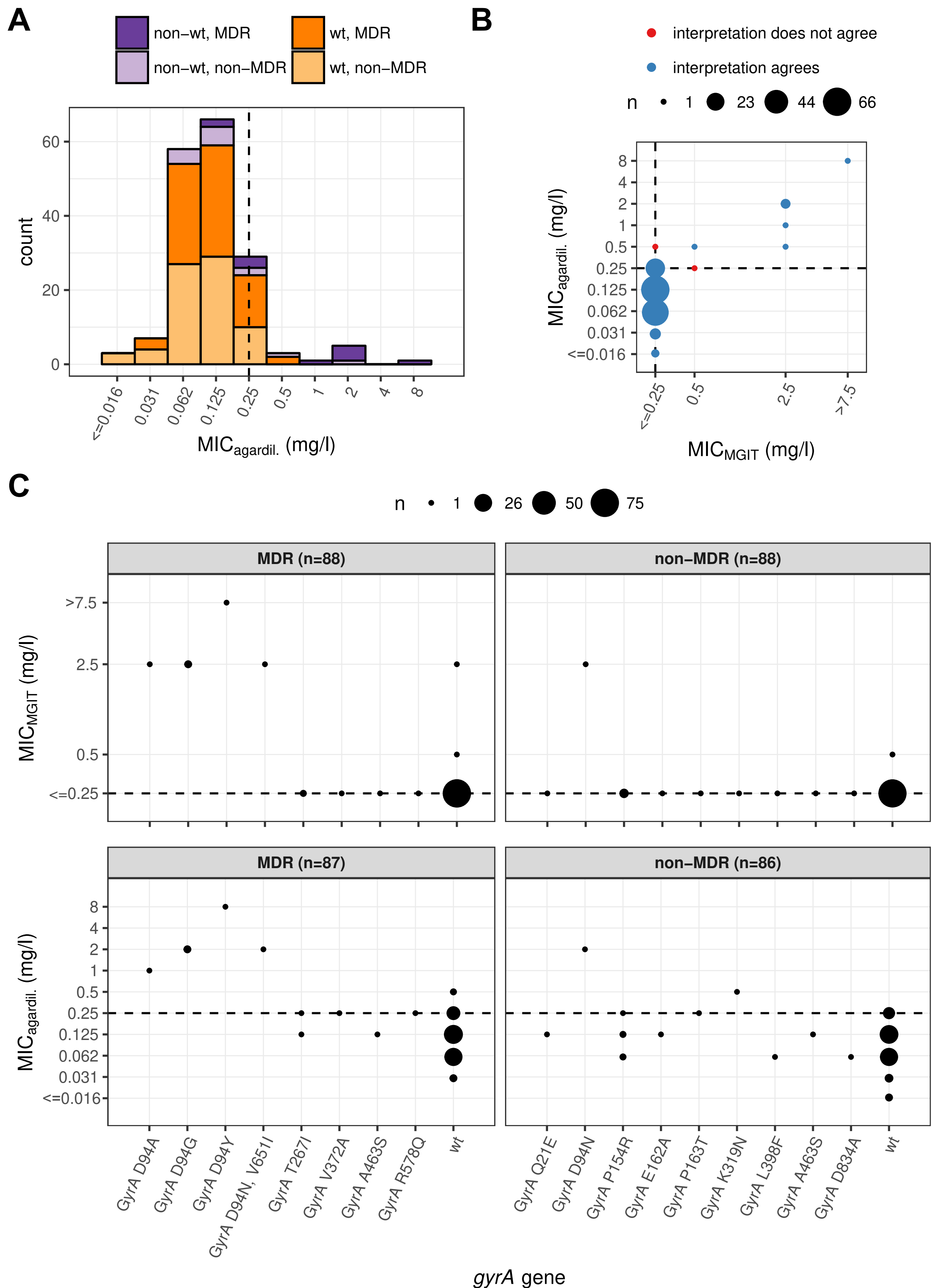
Supplementary Figure S5 - **A** Histogram of 7H10 agar dilution MICs for **kanamycin A** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Numbers in parentheses indicate the corresponding *Escherichia coli* nucleotide numbering. Top panel MGIT 960, bottom panel 7H10 agar dilution. Numbers in parentheses indicate the corresponding *Escherichia coli* nucleotide number. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Amikacin (AM)



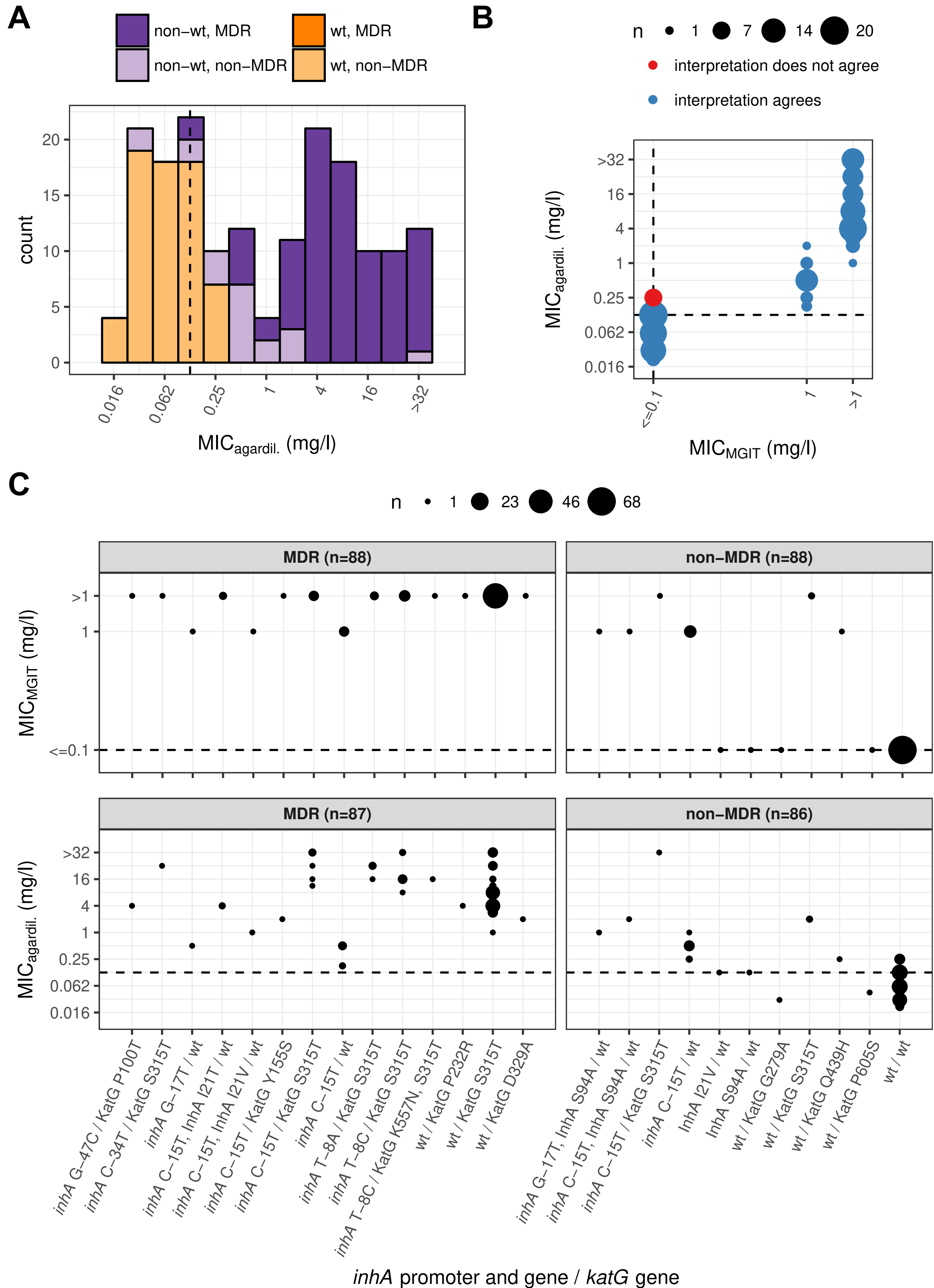
Supplementary Figure S6 - **A** Histogram of 7H10 agar dilution MICs for **amikacin** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Numbers in parentheses indicate the corresponding *Escherichia coli* nucleotide numbering. Top panel MGIT 960, bottom panel 7H10 agar dilution. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Moxifloxacin (MOX)



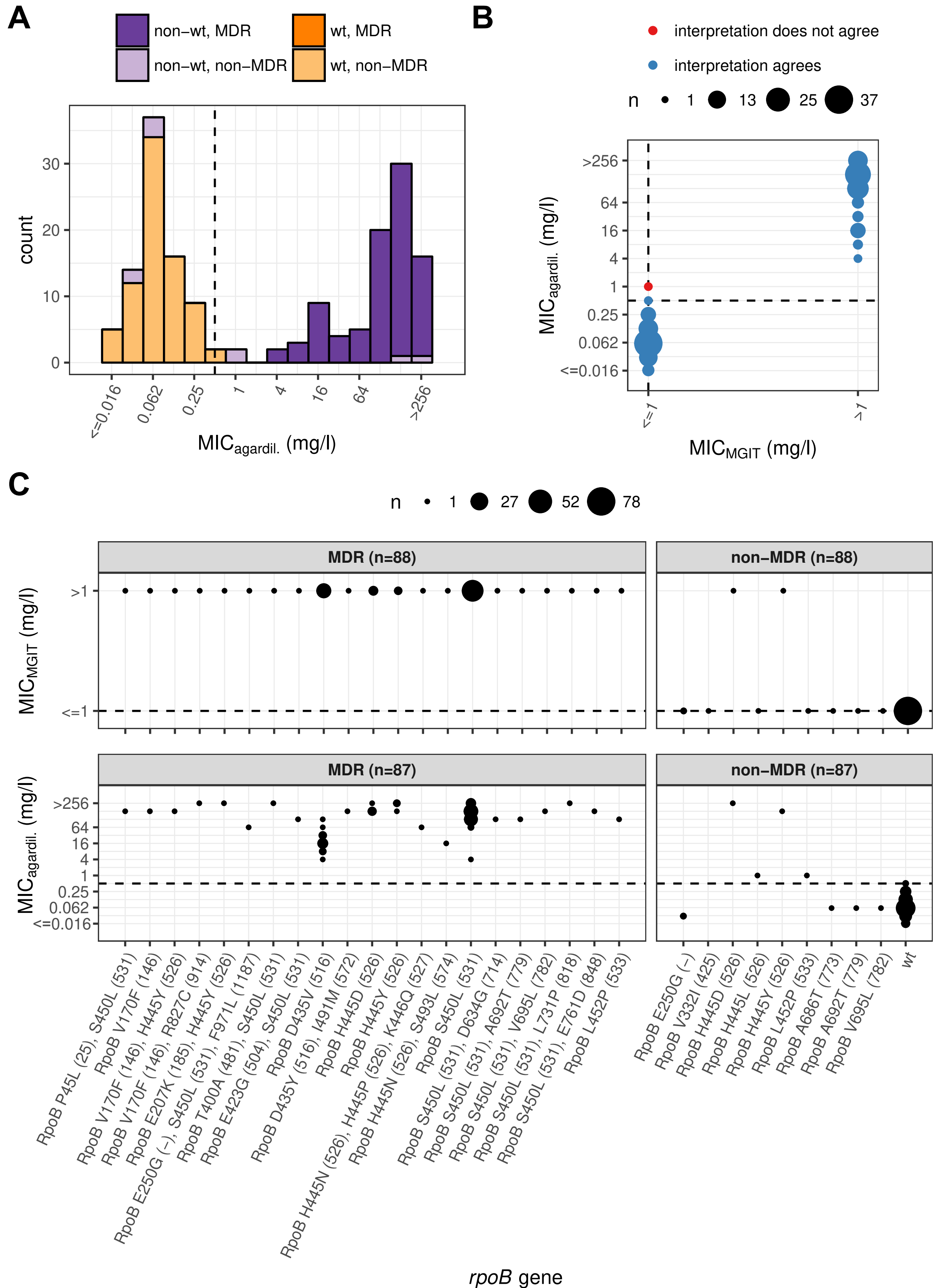
Supplementary Figure S7 - **A** Histogram of 7H10 agar dilution MICs for **moxifloxacin** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Top panel MGIT 960, bottom panel 7H10 agar dilution. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Isoniazid (INH)



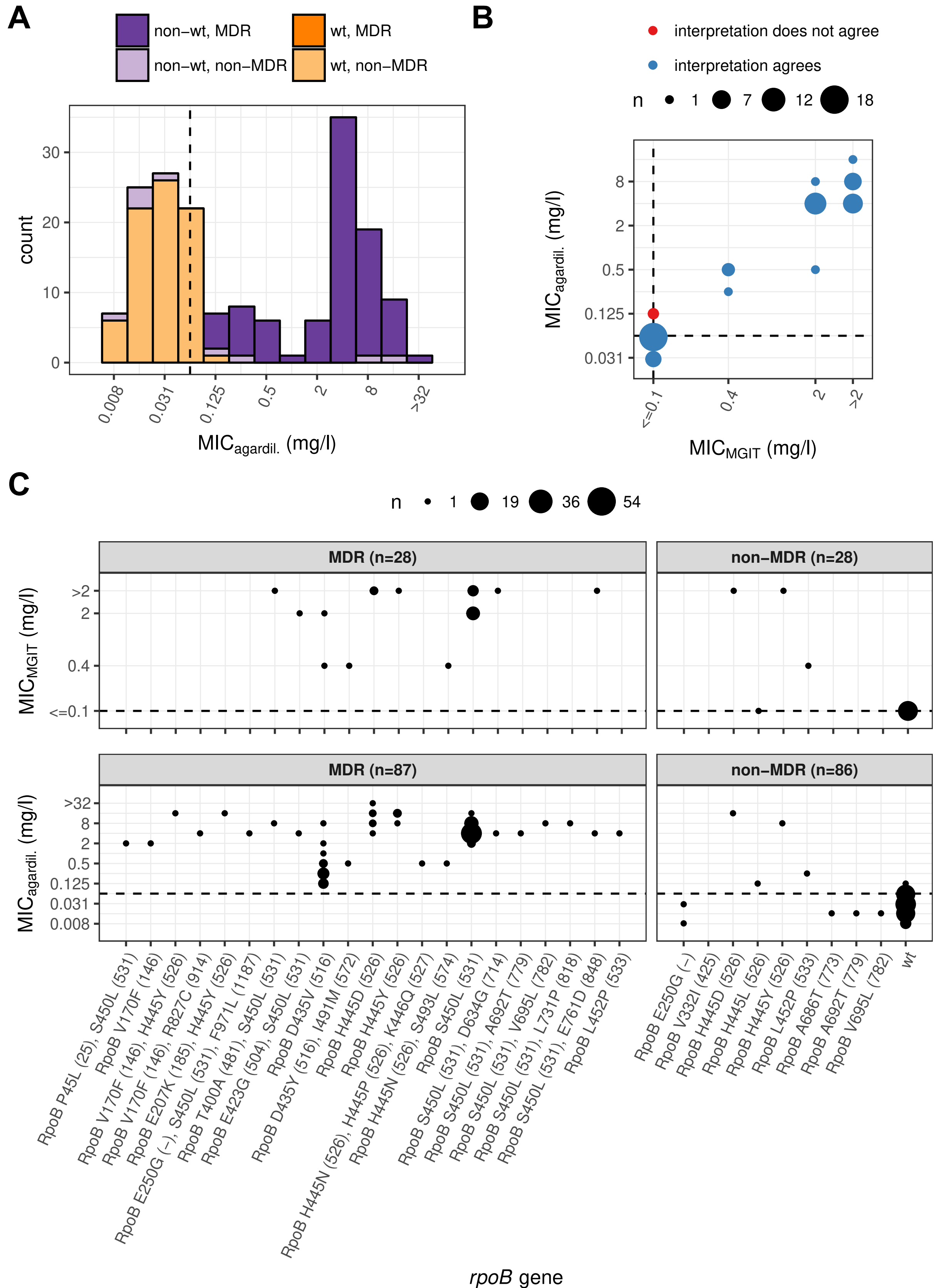
Supplementary Figure S8 - **A** Histogram of 7H10 agar dilution MICs for **isoniazid** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Top panel MGIT 960, bottom panel 7H10 agar dilution. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Rifampin (RMP)



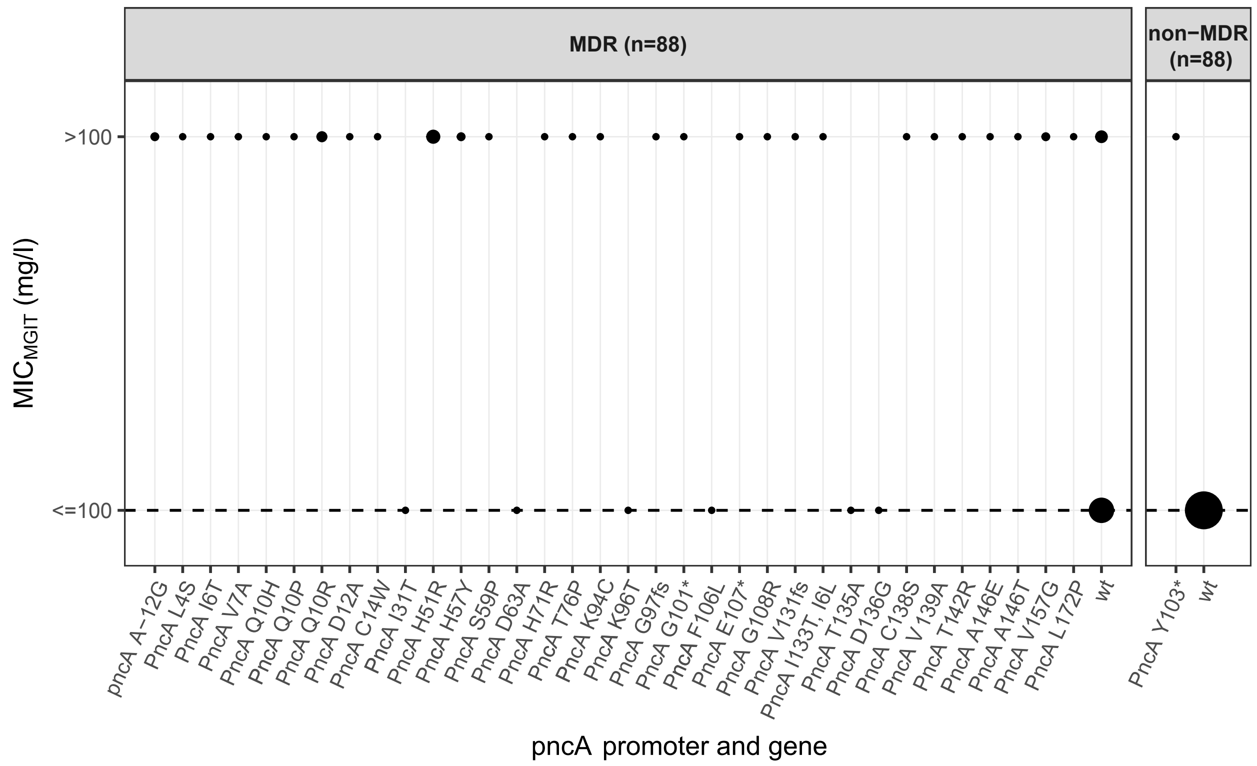
Supplementary Figure S9 - **A** Histogram of 7H10 agar dilution MICs for **rifampin** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Top panel MGIT 960, bottom panel 7H10 agar dilution. Numbers in parentheses indicate the corresponding Escherichia coli codon numbers. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Rifabutin (RBT)



Supplementary Figure S10 - **A** Histogram of 7H10 agar dilution MICs for **rifabutin** resistance. **B** method agreement between phenotypic DST by MGIT 960 and 7H10 agar dilution. **C** MICs of *M. tuberculosis* strains harboring resistance mutations in target genes or for wt. Numbers in parentheses indicate the corresponding *Escherichia coli* codon numbers. Top panel MGIT 960, bottom panel 7H10 agar dilution. In all panels, dashed lines indicate the ECOFF (see Table 1 in the main text).

Pyrazinamide (PZA)



Supplementary Figure S11 - **A** MICs of M. tuberculosis strains harboring resistance mutations in the target gene for **pyrazinamide** resistance or for wt. Only results for MGIT 960 are available. "*" indicates a stop-gain variant. Dashed line indicates the ECOFF (see Table 1 in the main text).