MICROBIAL GENOMICS

RESEARCH ARTICLE

Kumar *et al.*, *Microbial Genomics* 2026;12:001687

DOI 10.1099/mgen.0.001687

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Comparative genomics and molecular insights into smooth and rough clinical isolates of *Mycobacterium abscessus*

Rajender Kumar[1], Clara Braian[1], Gabrielle Fröberg[2,3], Lina Davies Forsman[3,4] and Thomas Schön[1,5,6,*]

Abstract

Background. Our recent findings indicate that the rough form of the *Mycobacterium abscessus* complex (MABC) is associated with worse clinical outcomes compared to the smooth variant. The glycopeptidolipid (GPL) locus has been considered vital for this transition; however, this has only been shown during *in vitro* selection. Thus, we aimed to investigate genetic differences between the morphotypes among clinical isolates.

Methods. Whole-genome sequencing (WGS) by Oxford Nanopore was applied to rough and smooth clinical MABC strains (*n*=10), including laboratory strains and *in vitro* exposure to nitric oxide (NO). Comparative genomic analysis, methylation and pangenome analysis were used to investigate the orthologs and unique protein clusters in relation to morphotype.

Results. WGS analysis showed high average nucleotide sequence identity (>98%), but genetic variants compared to the reference genome ranged from 1,433 to 27,025, and these differences did not correlate with morphotype. Among clinical isolates, smooth MABC clustered and were separated from the rough isolates. In both rough and smooth phenotypes, random genetic variations that could not separate the morphotypes were detected mainly in the *nrps* of the GPL loci. The major protein-cluster differences between smooth (*n*=123) and rough (*n*=22) clinical isolates were observed in key genes outside the GPL locus, such as *fabH*, *pks15*, *fadD29*, *otaA*, *mabA*, *lysX* and *acdA*, primarily involved in fatty acid biosynthesis and transporter systems. Repeated NO exposure, serving as a proxy for host-related stress, induced a rough phenotype by altering the acyl desaturase DesA1 protein, which is involved in fatty acid synthesis.

Conclusion. By analysing clinical MABC isolates, we show that morphotype variation cannot be explained by the GPL locus only. Unique genes involved in fatty acid biosynthesis were identified, informing further investigation into mechanisms distinguishing smooth and rough morphotypes of clinical MABC isolates.

Impact Statement

Mycobacterium abscessus complex is a rapidly growing, non-tuberculous mycobacterium that is particularly associated with lung infections. Patients with the rough morphotype experience worse clinical outcomes compared to those with the smooth morphotype. Until now, the genetic mechanisms underlying this transition have been inferred mainly from laboratory-engineered

[*Continued on next page*]

Received 09 December 2025; Accepted 11 March 2026; Published 06 May 2026

Author affiliations: [1]Department of Biomedical and Clinical Sciences, Division of Infection and Inflammation, Linköping University, Linköping, Sweden; [2]Department of Laboratory Medicine, Division of Clinical Microbiology, Karolinska Institutet, Stockholm, Sweden; [3]Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden.; [4]Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden; [5]Department of Infectious Diseases, Kalmar County Hospital, Linköping University, Kalmar, Sweden; [6]Department of Infectious Diseases in Östergötland, Linköping University, Linköping, Sweden.

***Correspondence:** Thomas Schön, thomas.schon@liu.se

Keywords: clinical isolate; glycopeptidolipids; *Mycobacterium abscessus*; nanopore sequencing; rough; smooth; whole-genome sequencing.

Abbreviations: ANI, average nucleotide identity; GPL, glycopeptidolipid; IPGA, integrated prokaryotic genome and pangenome analysis tool; LPG, lysylphosphatidylglycerol; MABC, *Mycobacterium abscessus* complex; NCBI, National Center for Biotechnology Information; NO, nitric oxide; NRPS, non-ribosomal peptide synthases; NTM, non-tuberculous mycobacterium; OADC, oleic acid, albumin, dextrose and catalase; ONT, Oxford Nanopore Technologies; PKS, polyketide synthase; TPP, trehalose polyphleate; WGS, whole-genome sequencing.

All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary figures and one supplementary table are available with the online version of this article.

mutants and site-directed mutagenesis studies. This study is the first to apply long-read whole-genome sequencing to both clinical and laboratory isolates with their morphotypes, smooth and rough. Results indicate that rough variants accumulate broader genomic variation involving multiple genes, extending beyond the previously implicated glycopeptidolipid locus. Our work reveals that this clinically significant shift is driven by a multifactorial process. These findings highlight new molecular targets such as *fabH*, *pks15*, *fadD29*, *otaA*, *mabA*, *lysX*, *acdA*, *mmpL4*, *mmpS4*, *mps1*, *mps2* and *htm (Rv0560c)* that may enable the development of improved diagnostic and therapeutic strategies against *M. abscessus*.

DATA SUMMARY

The whole-genome sequences have been deposited at the National Center for Biotechnology Information (NCBI) under BioProject accession number PRJNA1263120 (individual accession numbers are listed in Table 1). ANI values and other supplementary data are provided as supporting information.

INTRODUCTION

Mycobacterium abscessus complex (MABC) is a rapidly growing, non-tuberculous mycobacterium (NTM) that is particularly associated with pulmonary infection, especially in individuals with pre-existing lung conditions, such as cystic fibrosis or those who are immunocompromised [1]. MABC infections are notoriously difficult to treat due to the organism's intrinsic resistance to many antibiotics, driven by both spontaneous genetic mutations and its distinctive structural features [1, 2]. The MABC includes three subspecies*: Mycobacterium abscessus* subsp. *abscessus*, *Mycobacterium abscessus* subsp. *bolletii* and *Mycobacterium abscessus* subsp. *massiliense* [3]. During infection of host cells, the bacteria belonging to the MABC can change from a smooth variant, which is more linked to colonization, motility and biofilm formation, to a rough variant, which lacks motility and biofilm capacity due to restricting glycopeptidolipid (GPL) levels on the mycobacterial outer membrane [4]. Both smooth and rough variants can co-exist within the host and adapt independently under host immune pressure [5]. Many *in vitro* studies using site-directed mutagenesis have reported that this smooth-to-rough transition is associated with the GPL locus [6, 7]. In *vitro* and *in vivo* mutational studies indicated that deletion of *gtf3* in the GPL locus, with a putative function of transferring a rhamnose residue, uncovered the prominent role of the extra rhamnose in enhancing mannose receptor-mediated internalization of *M. abscessus* by macrophages [6]. Another study showed that the deletion of the GPL locus gene MAB_4111 c, which is involved in the dTDP-L-rhamnose into dTDP-6-deoxy-L-talose, influenced the GPL production [7]. The GPL locus genes encode proteins involved in glycosylation, methylation and acetylation to modify the GPL core structure, along with transport proteins that export the final products to the cell surface [4, 8–10].

Genomic and transcriptomic studies have revealed that alterations of DNA in the rough forms are involved, particularly genes encoding non-ribosomal peptide synthases (NRPS) and membrane proteins involved in GPL transport, such as *mmpl4b* [9, 11, 12]. Recently, it was discovered that MABC and closely related species possess a second GPL-associated locus, consisting of two adjacent non-ribosomal peptide synthetase genes, MAB_4690c and MAB_4691c. Deletion of MAB_4690c (ΔMAB_4690c) resulted in the loss of production of a previously uncharacterized lipid, named GL8P (glycosylated lipooctapeptid) [9]. Only selected proteins were examined, and the mechanism behind the transition from smooth to rough could not be explained by the second GPL locus only. Another study, which explored whole-genome recombination and pangenome analysis, examined the synergistic gain/loss of accessory genes that contribute to different metabolic profiles and the ability to adapt to oxidative stress, thereby facilitating strain adaptation to host environments, which can also be involved in infection pathogenesis [13]. Exposure to host-derived oxidative stress, such as nitric oxide, induces a rough morphotype, as indicated by reduced GPL expression levels in MABC [14].

To date, only a few studies have been conducted on whole-genome sequencing of clinical MABC smooth and rough isolates, as well as their comparative genomic analysis and study of genetic differences, making it a largely understudied area of research. Only limited investigations have been performed, focusing mainly on the GPL locus or trehalose polyphleate (TPP) locus and without a comparative genomic and protein analysis. In this study, we applied long-read whole-genome sequencing and analysis of smooth and rough isotypes of clinical MABC, including nitric oxide (NO) exposure, to identify genome-wide variations, not only the GPL locus in mutated isolates.

METHODS

Isolates and DNA extraction

Seven clinical MABC isolates originating from patients without non-cystic fibrosis or prior treatment against NTM were collected at Karolinska University Hospital, Stockholm, Sweden, and Linköping University Hospital, Sweden. The patients did not have any known infectious disease besides NTM nor an immune deficiency at the time of diagnosis. The median age was 58

(range 34–81), and 4/7 were females. The isogenic rough (CP194527) and smooth (CP194528) morphotype strains were received from Dr Laurent Kremer at Institut de Recherche en Infectiologie de Montpellier, France, sequenced initially and published [15]. The internal control lab strain was derived from the ATCC19977 reference strain (ATCC.org). The rough morphotype NO-exposed strain was attained by passaging of the ATCC19977 lab strain cultured in Middlebrook 7H9+10% oleic acid, albumin, dextrose and catalase (OADC) and 0.1 mM Deta-NONOate (Cayman Chemicals) every 3 days for a total of six passages. Smooth and rough morphotype definitions were based on macroscopic colony morphology by light microscopy on 7H10 agar supplemented with OADC as previously described [16]. After visible growth (5–7 days) on Löwenstein Jensen agar, single colonies were collected in glass beads and G2 buffer (QIAGEN). Samples were vortexed (30 s), sonicated (5 min) and then heat-killed at 80 °C for 20 min in a water bath before automated DNA extraction using the QIAGEN EZ1 DNA Tissue Kit. DNA was quantified using the Qubit dsDNA *HS* Assay Kit.

Library preparation and sequencing

Library preparation was performed using Oxford Nanopore Technologies (ONT) Ligation Sequencing kit SQK-LSK12 or SQK-LSK14, with the option of using Short Fragment Buffer in the clean-up step after adapter ligation to purify all DNA fragments equally. Libraries were sequenced using ONT FLO-MIN112 or FLO-MIN114 flow cells on an ONT MinION Mk1C device. All procedures were performed as recommended by the latest available Ligation Sequencing DNA protocol for SQK-LSK12 or SQK-LSK14 kits found online on the Nanopore website (https://nanoporetech.com) [17, 18].

Genome assemblies and annotations

Nanopore sequencing reads were generated using ONT, and the raw reads were base-called using the Guppy program. To enhance the overall quality of the input dataset, reads shorter than 1,000 bp and those with low-quality scores (Q-score <7) were removed using the Filtlong program (https://github.com/rrwick/Filtlong) as previously described [19]. The filtered database for each sample was subjected to generate a complete and high-quality consensus assembly using the Trycycler pipeline [20]. Trycycler uses multiple assembly programs, such as Flye [21], Miniasm [22] and Raven [23], on the same genome as input and produces a single consensus assembly. First, each filtered dataset was split into subsamples, and genome assemblies were generated for each subsample. Generated genome assemblies from each subset were clustered using the Trycycler cluster program, and the best cluster was selected after examining the clustering results. In the next step, the contigs were reconciled after an initial check to ensure that the contigs appeared sufficiently similar to each other. Finally, a consensus contig sequence for each cluster was generated based on multiple sequence alignments (Trycycler MSA) of the reads to the alternative sequences, and the option with the best-read alignment scores was chosen [20]. Additionally, we performed polishing of consensus sequences to increase their accuracy further using the Medaka program (https://github.com/nanoporetech/medaka), which was developed explicitly for the long-read ONT sequencer. This comprehensive Trycycler pipeline was applied to all ten samples to get complete genome assemblies. However, during the genome assembly, plasmid sequences were removed, so no plasmid sequence information is included. Finally, we used two different National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline [24] and Bakta programs [25] for gene prediction and annotation.

Genetic variation analysis

For genetic variation analysis, the Snippy 4.6.0 version (https://github.com/tseemann/snippy) was used to identify SNPs and small insertions/deletions (indels) from each isolate of MABC. *M. abscessus* subsp. *abscessus* ATCC 19977 was used as a reference for alignment. The functional impact of identified variants was assessed using SnpEff 4.3T [26]. Further results were analysed to check mutations and their predicted functional impacts. For the detection of methylation, the program Dorado was used to detect modified bases using built-in methylation models (https://github.com/nanoporetech/dorado). The reads were aligned to the reference genome using the Minimap2 program, and further sorting and indexing were performed using samtools [27]. The Modkit tool (https://github.com/nanoporetech/modkit) was used for modifying bases, specifically for converting modBAM to bedMethyl files and generating methylation frequency summaries, which were then used for statistical analysis.

Comparative genomics and pangenome analysis

In downstream analysis, we conducted a whole-genome comparison at the genomic level to assess genetic variability, focusing on overall nucleotide identity, SNPs and phylogenetic analysis based on whole-genome comparison. The average nucleotide identity (ANI) was estimated by using the Orthologous Average Nucleotide Identity Tool [28]. Furthermore, pangenome analysis was performed by applying an integrated prokaryotic genome and pangenome analysis tool (IPGA) [29]. For orthologous consideration, a cutoff of 70% sequence identity was set, while other parameters were kept at their default values.

GPLs loci and trehalose polyphleate locus analysis

The GPL loci sequences were obtained from the reference *M. abscessus* ATCC 19977 genome sequence (for the first GPL locus 4141888–4181959, genome assembly: GCF_000069185.1) [30]. This DNA sequence information was used as a query to retrieve

all sequenced genomes based on sequence identities using the MultiGeneBlast: Combined BLAST searches for operons and gene clusters [31]. All that information was further used for multiple sequence alignment analysis, and a phylogenetic tree was constructed based on the aligned sequence. The same approach was applied for the recently reported GPL locus second [9], as well as the trehalose polyphleate locus analysis [32].

Orthologous protein cluster identifications

For the identification of orthologs and singletons of protein between the smooth and rough morphotype isolates, the Ortho Venn3 program was used [33]. Gene-coded protein sequence files (smooth and rough surface isotypes) were subjected to compute the orthologs using the OrthoFinder algorithm [34]. The OrthoFinder algorithm performs all-vs-all sequence alignment to measure the sequence similarity. For protein sequence similarity, the *E*-values <1e−5 and sequence identity >30 were used as a threshold to consider the orthologs. A further similarity graph was constructed based on the homology relationship between proteins, where proteins are represented as nodes in a graph, as well as the unique protein clusters.

RESULT AND DISCUSSION

Whole-genome sequencing of *M. abscessus*

We applied a long-read ONT sequencing approach to complete the genome sequences of 10 MAB isolates (both smooth and rough morphotypes) to investigate their genetic variation. In the NCBI genome assembly database (September 2025), a total of 2,452 genome sequence records are publicly available (https://www.ncbi.nlm.nih.gov/datasets/genome/); however, most of these genome sequences are from the subspecies *M. abscessus* subsp. *abscessus* (a total of 1,112 genomes, of which 35 are complete genomes), while for subsp. *bolletii* and *massiliense* are only 132 and 439, respectively. In the NCBI gene database, strains of *M. abscessus* subsp. *abscessus*, such as ATCC:19977, CCUG:20993, CIP:104536 and DSM:44196, are recorded. So far, the morphotype is recorded for only four smooth and one rough isolate, along with complete genome information in the NCBI genome database (https://www.ncbi.nlm.nih.gov/datasets/genome). Only a few studies have carried out genome sequencing of *M. abscessus* smooth and rough clinical isolates and their comparative genomic analysis and genetic differences [35–37]. Most of the studies (described in more detail below) have been carried out to focus on targeted mutation, mainly in the GPL locus, other membrane transporter proteins and signalling proteins *in vitro* and *in vivo* models [9, 38].

All isolated genome assemblies achieved 100% coverage of the *M. abscessus* genome and were circularized. The genome sizes of all isolates fell between 4.8 and 5.1 Mb, which are comparable to the reference (accession number: NC_010397) *M. abscessus* ATCC 19977 genome size of 5.1 Mb (see Table 1). The genome assemblies of all clinical isolates had an average G+C content of 64.2%; no differences were found between the different morphotypes of MABC. There was no significant change in genome size and G+C content despite the predicted number of coding genes varying among the MABCs isolates (see Table 1). All the sequenced genomes comprised the complete three rRNA genes, including 5S, 16S and 23S, which were predicted. As previously reported in a 2019 study, when two different isotypes of *M. abscessus* clinical strains, smooth and rough, were compared, no major differences were observed in their genomic content [35]. However, our analysis indicates that, among the clinical isolates investigated so far (n=7), there are changes in their genetic content that alter the functionality of their coding regions, compared to previous studies.

Genetic variation analysis

The results of genetic variation analysis indicate varying degrees of genomic divergence in *M.* abscessus isolates. The number of total genetic variants compared to the reference genome varies widely among isolates, ranging from 1,433 (CP194519: Mab7_R) to 27,025 (CP194523: M1_R); see details in Table 2. However, the higher variant counts (>20,000) were observed in isolates M1_R (CP194523), M3_R (CP194522), M4_R (CP194521) and M7_S (CP194519), suggesting substantial genetic divergence from the reference genome in these clinical isolates. While the laboratory isolates (M24_S: CP194528 and M24_R: CP194527) have very low numbers of variant counts, 2,027 and 1,580, respectively, implying potential clonality. SNPs dominate across all isolates; however, the number of insertions/deletions varies, but generally contributes a smaller fraction, ~10–15% of the total variation, particularly in clinical isolates. Further, mutation effect analysis showed that missense mutations are the most frequent, often outnumbering silent mutations, while nonsense mutations are relatively rare, as expected. Furthermore, the Ts/Tv (transition/transversion) ratio varies markedly, which is also used as a quality control and understanding patterns of DNA sequence evolution; a very high ratio (>20) was observed in isolates M24_S (CP194528) and M24_R (CP194527), which suggests a strong transition bias typical of closely related isolates with less mutational noise. In contrast, lower ratios (~2.5–2.7) in isolates like M1_R, M3_R, M4_R and M7_S suggest a more balanced mutation spectrum and genomic diversification.

Furthermore, results indicate that some genes have a higher number of missense mutations in clinical isolates, while having a lower mutation number in laboratory isolates, such as non-ribosomal peptide synthetase (reference genome locus tag: MAB_RS23740 and MAB_RS16820) and type 1 polyketide synthase (locus tag: MAB_RS04905). It was observed that some genes have a higher number of missense mutations, mainly in both smooth and rough clinical isolates, such as DNA binding protein (MAB_RS10715), hypothetical protein (MAB_RS13600), non-ribosomal peptide synthetase (MAB_RS16820), FAD-dependent

Table 1. List of *M. abscessus isolated* clinical strains, genome sequenced, their genome assembly and annotation information

Genome ID/ accession no.	Clinical	Morphotype	Strain/isolate	Genome size (bp)	G+C content (%)	No. of genes	Reference
NC_010397	Ref. strain	Smooth	*M. abscessus* ATCC 19977	5,067,172	64.1	4,951	NCBI genome data
CM132931 (Int. control)	Lab strain	Smooth	*M. abscessus* derived from ATCC 19977	5,067,123	64.1	4,987	Our study
CP194528 (M24_S)	–	Smooth	*M. abscessus*	5,072,683	64.1	4,988	Our study [15]
CP194527 (M24_R)	–	Rough	*M. abscessus*	5,072,543	64.1	4,978	Our study [15]
CP194523 (M1_R)	Clinical	Rough	*M. abscessus*	4,961,325	64.3	4,836	Our study
CP194522 (M3_R)	Clinical	Rough	*M. abscessus*	5,153,539	64.2	5,077	Our study
CP194521 (M4_R)	Clinical	Rough	*M. abscessus*	4,910,125	64.1	4,784	Our study
CP194520 (Mab6_R)	Clinical	Rough	*M. abscessus*	4,870,434	64.2	4,778	Our study
CP194519 (Mab7_R)	Clinical	Rough	*M. abscessus*	5,099,870	64.1	5,022	Our study
CP194525 (M6_S)	Clinical	Smooth	*M. abscessus*	4,998,865	64.2	4,928	Our study
CP194524 (M7_S)	Clinical	Smooth	*M. abscessus*	5,108,508	64.2	5,028	Our study
CM132630 abs_R*)	Lab strain	Rough	*M. abscessus*	5,066,682	64	5,048	Our study

Note: R represents rough, and S represents smooth. M. abs_R*: derived from *M. abscessus* spp. *abscesus* strain ATCC 19977.

monooxygenase (MAB_RS13595), replication initiator protein (MAB_RS10720), cydC protein (MAB_RS13370), acyltransferase activity (MAB_RS18695) and mftG protein. A smaller number of missense mutations were observed in the no-ribosomal peptide synthetase (MAB_RS20780) type 1 polyketide synthase (MAB_RS07795) and ABC transporter (MAB_RS0775) genes in all isolates, including clinical, laboratory and NO-exposed. The SNP analysis concluded a general difference between laboratory and clinical isolates, which is not dependent on the morphotype.

The results of the methylation differentiation analysis indicate that some genetic regions undergo hypermethylation when comparing rough vs. smooth clinical isolates. A proposed significance threshold of $P<9\times10^{-8}$ adequately controls the false positive rate for EPIC array DNA methylation studies [39]. Considering this threshold parameter, we identified some genes *bdcA* (cyclic-di-GMP-binding protein), *gdh* (NAD-specific dehydrogenase) and *rocD* (ornithine aminotransferase), which all indicate highly significant differential methylation in rough isolates. Additional genes, such as *mmcO*, *mftC*, *ftsZ* and *smc*, showed low *P*-values ($\sim10^{-6}$) and may be considered significant at a nominal level in rough isolates. Genes *pccB* (0.000627), *dnaB* (0.000840), *geoB* (0.001124) and *ephA* (0.001956) with *P*-values between 0.001 and 0.00001 show weaker evidence that can be considered nominally significant in a standard test. Some other genes, such as *mmsB*, *mmpl4*, *atpD* and *argJ*, are also found to be more methylated in rough isolates. The membrane exporter protein coding genes, such as *mmpl4*, are located in the GPL locus. These findings reveal distinct differences between rough and smooth methylation signatures across metabolic and regulatory genes, suggesting a potential role for DNA methylation in modulating gene expression *in vivo* and contributing to the transition from smooth to rough morphotypes in clinical MABC isolates.

Genome diversity and pangenome analysis

Whole-genome sequence comparison was performed for all smooth and rough clinical and laboratory MABC isolates to identify and delineate their strain and subspecies (Fig. 1). For whole-genome sequence analysis, average nucleotide sequence identities (ANI) were estimated for all isolates, which showed an ANI of >98%, indicating that all isolates are very closely related to

Table 2. Summary genomic variation of all MABC isolates, including SNPs, multiple nucleotide polymorphism (MNPs), insertions/deletions (INS/DEL), mutation types (missense, nonsense, silent) and transition/transversion (Ts/Tv) ratio

Isolates (MABC)	Total variant	SNP	MNP	INS/DEL	Missense/nonsense	Silent	Ts/Tv ratio
CP194528 (*M24_S*)	2,027	877	341	374/121	724/0	107	37.130
CP194527 (*M24_R*)	1,580	660	254	274/169	557/0	81	24.384
CP194523 (*M1_R*)	27,025	23,773	2,493	168/470	4,877/8	17,009	2.627
CP194522 (*M3_R*)	20,267	18,248	1,346	132/458	3,885/4	12,991	2.697
CP194521 (*M4_R*)	23,903	21,152	1,831	147/659	4,425/8	15,085	2.634
CP194520 (*Mab6_R*)	20,812	18,283	1,378	239/811	3,934/4	12,896	2.629
CP194519 (*Mab7_R*)	1,433	747	84	283/180	587/1	129	13.094
CP194525 (*M6_S*)	20,569	17,939	1,368	274/834	3,971/6	12,579	2.532
CP194524 (*M7_S*)	24,521	21,445	1,922	196/811	4,297/6	15,555	2.634
CM132931 (Int. Control)	3,024	1,419	595	352/223	1,167/1	203	31.25
CM132630 *(M. abs_R)**	4,307	1,909	923	330/212	1,569/1	274	26.666

Notes: *Derived from *M. abscessus* spp. abscesus strain ATCC 19977.

each other and belong to the *M. abscessus* subsp. *abscessus*, which has been taken as a reference for comparative analysis. The pairwise sequence identities are detailed in Table S1 (available in the online Supplementary Material). For the identification of new species and strains, as well as assessing species relationships, the ANI method is more accurate compared to conventional methods, including 16S rRNA, line probe assays, such as HAIN NTM DR+ used in clinical microbiological routine [40] and other multiple-gene marker approaches [41–43].

To estimate the pangenome and core-genome (genes shared by all individual isolates) sizes of the MABC strains, we analysed all sequenced genome isolates along with the NCBI reference genome *M. abscessus* ATCC 19977. Our analysis revealed that as additional genomes were incorporated, the estimated size of the pangenome continued to increase, while the size of the core genome decreased. Since the core-/pangenome ratio did not plateau sharply, this indicates that MABC possesses an open pangenome (Fig. 1a). The estimated pangenome and core-genome sizes are ~7,740 and 4,119 genes, respectively, as shown in Fig. 1b. In addition, pangenome size was evaluated using the Panaroo program [44], and the obtained results were highly consistent, with core gene numbers of ~4,200 at sequence identity thresholds ranging from 90 to 95%. Minor differences in gene counts are expected and can be attributed to the total number of genes predicted by Prokka, since Panaroo relies on Prokka annotations as input. Furthermore, the final gene counts are influenced by the sequence identity threshold applied during clustering. In the IPGA pangenome analysis, we used a 70% identity cutoff. Although this threshold is relatively low for within-species pangenome studies, particularly morphotype variation, it was applied to define broader protein families and true orthologs within the species. A large number (>30%) of genes that are part of the core genome are poorly characterized (Fig. 1c). Earlier studies have revealed that these are mostly hypothetical proteins whose function is not yet known. A previous pangenome analysis of 40 *M. abscessus* genomes isolated from patients across diverse geographical regions identified 12,656 pangenomic genes and 3,354 core gene clusters [45]. Similarly, a more recent study analysing 50 globally distributed *M. abscessus* isolates estimated a pangenome comprising 13,477 genes, including 3,734 core gene clusters [46]. In comparison, the pangenome and core genome sizes observed in the present study differ from those in the earlier reports, likely because our analysis was restricted to distinct colony morphotypes (smooth and rough isolates) of *M. abscessus*.

A whole-genome phylogenetic tree is constructed to describe evolutionary relationships among species. Clinical and laboratory isolates are clustered into separate groups as seen in the phylogenetic tree (Fig. 1d). As expected, the result shows that laboratory strains (accession numbers CP194528 and CP194527) and the reference strain are clustered together, with the smooth laboratory isolate (accession number CP194528) being closer to the reference strain. However, the clinical isolates form a separate group, but they exhibit more variation among them as compared to laboratory strains. The clinical isolate cluster indicates that the smooth

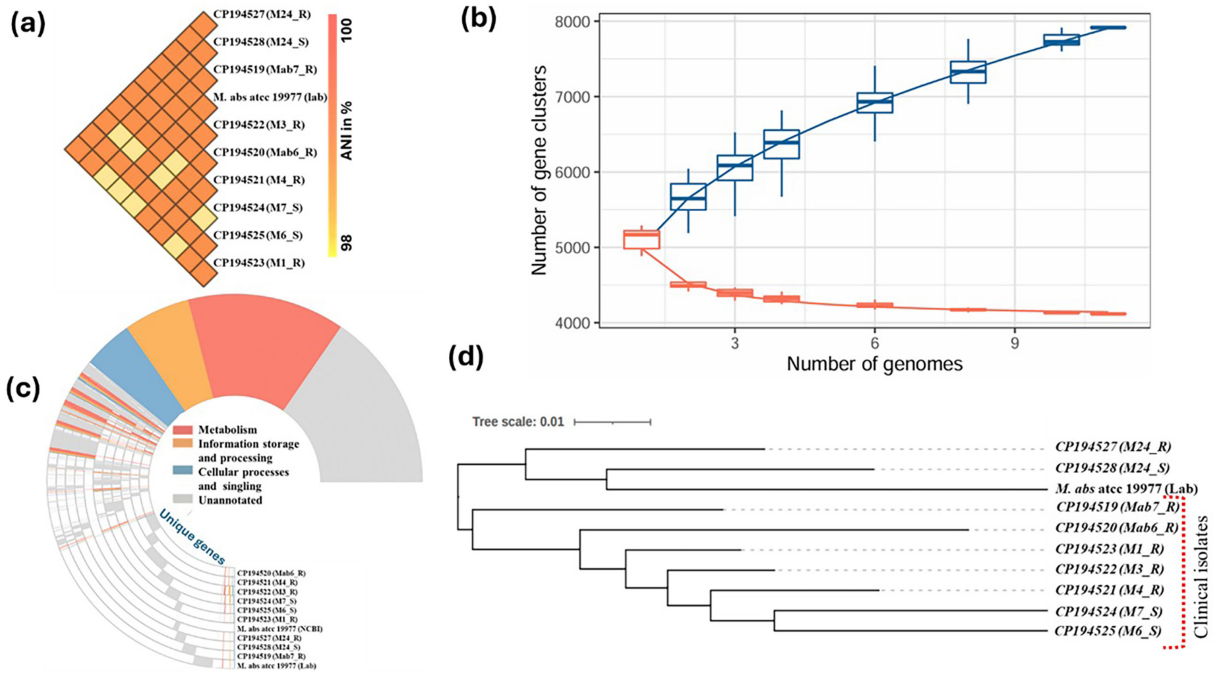

Fig. 1. Pangenome of *M. abscessus* strains. (a) ANI analysis. (b) The estimated pangenome and core genome. (c) The pangenome profile of all type strains in *M. abscessus* strains, functional analysis and unique genes. (d) Phylogenetic tree analysis of isolate *M. abscessus* subsp. *abscessus* based on single-copy genes reveals the evolutionary relationships and distances among the different isolates. A phylogenetic tree based on the identification of highly conserved single-copy genes was constructed to describe the evolutionary relationships between species. MUSCLE was used to perform sequence alignment, trimAl was used to extract and trim the conserved sequences and FastTree was used to infer the phylogenetic tree between species using the maximum-likelihood method. The SH test method was used to assess the credibility of each phylogenetic node.

isolates are closely related to each other; however, the rough isolates are not as well clustered. The rough isolates adopted more genetic variation among them. It was also noted that all the rough strains likely undergo a transition via different pathways, where genetic changes occur, but not solely via a unique target gene or a particular locus, which argues against the previous focus on the GPL locus.

Many studies have been carried out to understand the molecular basis mechanisms of the transition from smooth to rough morphotypes of MABC, particularly focusing on the glycopeptidolipid synthesis and the proteins involved in the export mechanism [6, 9, 30, 32, 47, 48]. In MABC, GPL biosynthesis is governed by the *msp1*, *msp2*, *gap*, *mmpS4*, *mmpL4a* and *mmpL4b* genes. A CG insertion in *msp1* disrupts transcription of the *msp1–msp2–gap* operon, halting GPL production [11]. Additional insertions, deletions, and single-nucleotide polymorphisms in these genes can also cause a loss of GPL synthesis, leading to the rough morphotype observed in clinical isolates [12]. Furthermore, an *in vitro* study demonstrated that aminoglycosides can trigger a smooth-to-rough transition in *M. abscessus* without complete GPL depletion; exposure to subinhibitory amikacin concentrations induced a rough morphology and upregulated MAB_3508 c, a homologue of *Mycobacterium tuberculosis* whiB7, which mediates antibiotic resistance [49]. Recently, to gain deeper insight into the genetic basis underlying this phenomenon, researchers employed transposon insertion site sequencing (Tn-seq) to identify genes essential for the survival of *M. abscessus* smooth and rough morphotypes [50]. The analysis revealed that each morphotype exhibits distinct genetic requirements for *in vitro* growth, including several genes associated with responses to environmental stressors. In a murine infection model, the divergence in Tn-seq essential gene profiles between smooth and rough MABC became even more noticeable. This variation appears to be influenced, at least in part, by the differing host immune responses triggered by each morphotype [50].

Comparative studies of the GPL locus

Several studies using site-mutagenesis, including laboratory strains, have reported that the GPL locus and other associated proteins play a key role in the transition between smooth and rough morphotypes of *M. abscessus*. In this study, we observed that the genomic content of the GPL locus showed high sequence identities (>99.95%) between the isolates and smooth and rough morphotypes. However, the genes are highly organized and ordered, and their coding proteins are highly conserved (Fig. 2a). Notably, the MmpL proteins showed sequence identity ranging from 98 to 100% among the clinical and laboratory strains. A few proteins of the GPL loci did exhibit small differences (>2% protein sequence differences). However, no consistent patterns

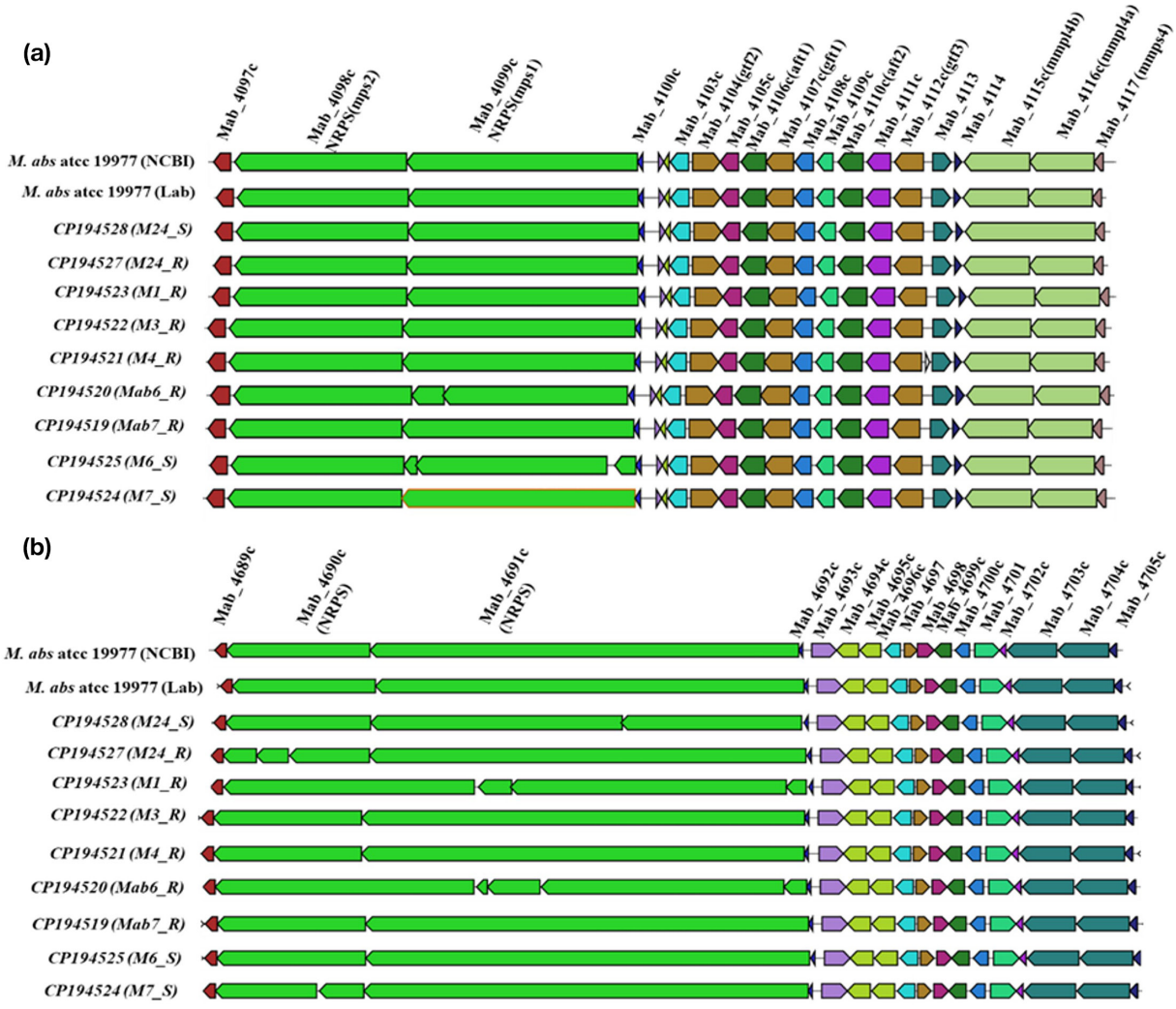

Fig. 2. Genomic structural organization of GPL loci (loci 1 and 2) in patient isolates of *M. abscessus*. (a) The classical GPL locus and its protein-coding genes are highly organized and conserved in different isolates. The nonribosomal peptide synthetases (*mps1* and *mps2*) exhibit SNP and small missense mutations, while the other proteins are highly conserved. (b) Like the GPL-like biosynthetic gene cluster (GPL locus 1), gene clusters similar to it included NRPS genes, transporter (*mmpL*) genes, glycosyltransferase genes and regulatory proteins (GPL locus 2).

were observed within the smooth and rough morphotypes, and the differences did not distinguish between morphotype groups. Significant sequence differences, including SNPs and small insertions/deletions, were identified in the GPL locus proteins Mps1 and Mps2 of smooth and rough isolates. Only a few random amino acid changes (~7–10 amino acid sites) were observed in the GPL locus proteins *gtf2*, *gtf3* and *mmpl4b* when comparing smooth and rough isolates. However, the laboratory strains' GPL locus proteins showed higher sequence identity. In contrast, clinical strains showed protein sequence differences (sequence identity ranging from 60 to 100%) mainly in the NRPS (*mps1* and *mps2*) proteins, which are non-ribosomal peptide synthetases that play a crucial role in the biosynthesis of GPLs in *Mycobacterium* species. In addition, concerning the relationship between these mutations and the corresponding protein structure and function, preliminary structural analyses and comparative modelling revealed no significant conformational alterations, with overall RMSD values below 1 Å. Although no major structural differences were detected in the static models, the mutations may still affect protein functionality, which can be further investigated using comprehensive approaches such as molecular dynamics simulations.

A previously reported study found that SNPs or small insertions/deletions in the non-ribosomal peptide synthase gene cluster proteins *mps1*, *mps2* and *mmpl4b* affect GPL production in clinically persistent MABC strains. Some studies have reported that the transition from smooth to rough typically results from mutations in genes within the GPL biosynthetic pathway, but these studies used mutagenesis [51]. However, the exact mechanisms are not yet fully understood. It is thought to be triggered by stress

conditions encountered during host colonization. Others have shown that post-synthetic modifications of GPL may further influence MABC's ability to adhere to and invade macrophages [30, 52].

Recently, a second, homologous GPL locus in *M. abscessus* was identified [9]. Like the canonical GPL locus, this newly discovered region contains two adjacent NRPS genes, MAB_4690 c and MAB_4691 c, along with several other associated genes [9]. However, further functional analysis revealed that a deletion mutant lacking MAB_4690 c (ΔMAB_4690 c) was unable to synthesize a previously undescribed lipid, termed GL8P (glycosylated lipooctapeptide). This novel molecule features an acylated octapeptide backbone modified with mono- or di-O-rhamnosyl substituents [9]. Similarly, in our study, we have also found a highly conserved second GPL locus, which not only contains NRPS and MmpL proteins but also exhibits a gene cluster organization in a similar pattern to canonical GPL (Fig. 2b). It was observed that sequence variations in NRPS could not be decisively attributed to the rough and smooth morphotypes.

Trehalose polyphleate locus analysis

As previously reported, the TPPs protein-coding genes are also involved and play a key role in the mycobacterial cell wall, particularly the cell surface [4]. We examined the trehalose polyphleate gene cluster and its comparison among the sequenced isolates. The TPP biosynthetic gene cluster (DAT/PAT locus) encodes polyketide synthases (*PKS*), acyltransferases (*PapA*), fatty acyl-AMP ligases (*FadD*) and *cpl*dentical. In contrast, the *papA3* and *Pks* genes exhibited some sequence alterations, including deletions and nonsense mutations (Fig. 3). No significant sequence variations in the TPP locus were noted that could be associated with rough and smooth morphotypes; however, the level of expression of the TPP locus protein may affect the morphotypes. The TPPs are high-molecular-weight glycolipids found primarily in mycobacteria, consisting of a trehalose core esterified with long-chain polyunsaturated fatty acids known as phleic acids [32]. These lipids are non-toxic, do not induce proinflammatory cytokines and do not inhibit phagolysosome fusion. It produces a high amount of octoacylated trehalose, containing seven phleic acid-like substituents and one C14–C19 fatty acid residue [32]. TPP biosynthesis begins in the cytoplasm with a 2,3-diacyl trehalose precursor, which is modified by the addition of a C14–C19 fatty acid. The transporter MmpL10 exports this intermediate to the cell surface, where an acyltransferase (PE) catalyses transesterification with a C36–C40 phleic acid, forming mature TPP [32]. TPP is a significant surface lipid in rough morphotype colonies of MABC and plays a key role in cell aggregation [53].

Orthologous protein clustering analysis

To identify orthologous and unique proteins between two morphotypes, smooth and rough isolates of MABC, protein comparisons and evolutionary analyses were performed. Whole-genome sequence information based on predicted coding proteins in the isogenic laboratory strains M24_S (CP194528) and M24_R (CP194527) were highly orthologous; only 1.3% proteins were identified as singleton proteins that did not detect any ortholog with other isolates, as well as those that are single proteins in the genome. In this analysis, a total of 4,513 ortholog clusters, including 4,178 single-copy clusters, were identified in the isogenic laboratory strains. A total of 83 and 45 singleton proteins were found in smooth and rough isolates, respectively (Fig. S1).

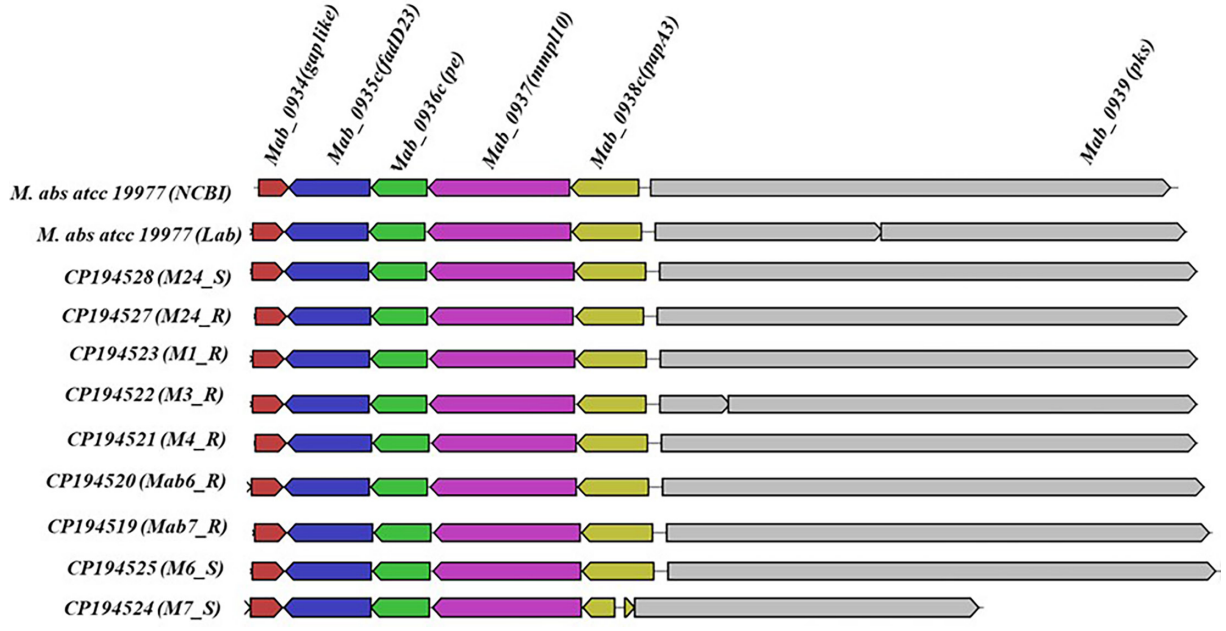

Fig. 3. Trehalose polyphleate locus gene cluster order and comparative analysis based on their protein sequences.

Notably, only a few key cluster proteins were identified, including acyl-carrier-protein (desaturase DesA1), which is involved in fatty acid biosynthesis, and showed differences between the isogenic isolates. Another unique protein cluster, consisting of two uncharacterized proteins, was found only in the rough isolate. Furthermore, the NCBI BLAST search of these cluster proteins showed 78% sequence identity but less query coverage (38%) with the L,D-transpeptidase. However, the identified proteins cluster L,D-transpeptidase, acyl-carrier-protein desaturase DesA1 and singleton proteins that may be distinct and play a key role in different morphotypes among clinical and non-clinical isolates [54, 55].

When the NO-exposed rough strain of MABC was compared to the parental smooth strain of MABC, a total of 4,321 orthologs were identified, which is a smaller number compared to the laboratory isogenic strains (Fig. S2). The NO-exposed rough strain comprises ~600 singleton proteins and 11 unique protein clusters. The unique protein clusters are mainly involved in fatty acid biosynthesis. However, the six protein clusters were observed only in reference smooth isolates, which mostly showed functional enrichment primarily in the plasma membrane (Table S2). However, comparing the laboratory MABC isolates revealed that the fatty acid biosynthesis pathway, involving the essential DesA1 enzyme, showed key differences. In mycobacterial species, this essential enzyme is involved in the fatty acid biosynthesis pathway, specifically in the modification of acyl-ACP substrates to create monounsaturated fatty acids, which are then incorporated into the mycolic acid cell wall. Though this DesA1 enzyme is not studied in detail in its function in MABC, similar enzymes in *M. tuberculosis* are known to be essential for cell wall integrity [56], and a *Mycobacterium smegmatis* strain overexpressing *M. tuberculosis* DesA1 showed Ca^{2}-dependent changes in its surface phenotype, suggesting a link to cell surface properties and function [54, 57].

When we compared the protein sequences of clinical smooth strains (CP194425: M6_S and CP194524: M7_S) and rough strains (CP194523: M1_R and CP194522: M3_R), ~97–98% of the encoded proteins were found to be orthologous. A total of 4,198 ortholog clusters were found across all four (two smooth and two rough), and 542 proteins were predicted as singletons, which is 2.7% of the total protein. However, a total of 123 protein clusters were found to be unique only to smooth isolates (M6_S and M7_S), while 22 protein clusters were found only in rough isolates, as shown in Fig. 4. Only a few protein clusters were found in particular isolates: smooth isolates 4–7 and rough isolates 4–22. Further functional enrichment analysis of unique protein clusters

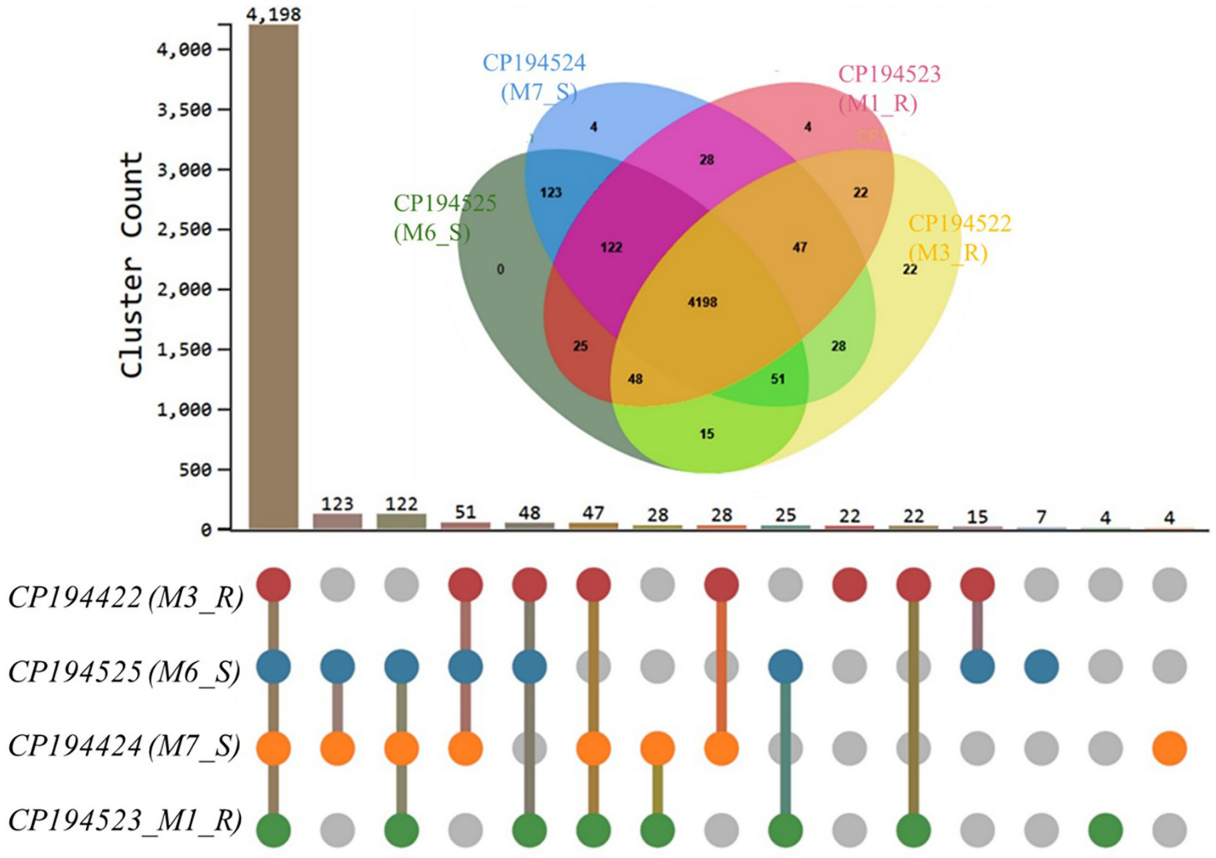

Fig. 4. Proteomic comparison of MABC between smooth (CP194424 and CP194525) and rough (CP194522 and CP194523) isotypes. An overall display shows orthologous and unique protein clusters between two different morphotypes (CP194523). Here, the vertical lines show intersecting sets.

of smooth isolates mainly involved plasma membrane, oxidoreductase activity, fatty acid biosynthetic process, transmembrane transport, response to antibiotics and oxidation–reduction process. In rough isolates, unique protein clusters are mainly involved in the plasma membrane, in all-trans-retinol 13,14-reductase activity and in fatty acid beta-oxidation via acyl-CoA dehydrogenase pathways (Table 3). In both cases, unique protein clusters in the smooth and rough isolates are predicted to be uncharacterized or hypothetical proteins, as well as those of very short length.

Table 3. List of functionally characterized proteins that are found as key differential proteins between smooth and rough isolates from orthologous analysis based on their protein sequence similarity

Gene	Uniprot hits ID	Protein name	Biological process
*mmpL4***	P9WJV2	Siderophore exporter MmpL4	Exporter system
int	P04890	Integrase	DNA integration
ephD	P66778	Probable oxidoreductase EphD	
mabA	P0A5Y5	3-Oxoacyl-[ACP] reductase MabA	Fatty acid elongation
SCO3172	Q9RKB5	Baeyer–Villiger monooxygenase	
TMP	E7DNB6	Tape measure protein (TMP)	Viral tail assembly
htm	P9WKL4	Methyltransferase	Methylation
ubiE	A4XPM7	C-methyltransferase UbiE	Aerobic respiration
ABC1K7	B9DGY1	BC1 complex kinase 7	Oxidative stress
Dpp8	Q80YA7	Dipeptidyl peptidase 8	Negative regulation of programmed cell death
*dnaJ1***	Q73XZ6	Chaperone protein DnaJ 1	Protein refolding
ephA	I6YGS0	Epoxide hydrolase A	Response to a toxic substance
*lysX***	B1MAX9	Lysylphosphatidylglycerol biosynthesis bifunctional protein	Lipid metabolic process
Sama_2741	A1S987	Methyltransferase	Methylation
mll6982	Q987N5	2-Dehydropantoate 2-reductase	Pantothenate biosynthetic process
mmr	P11545	Methylenomycin A resistance	Response to the antibiotic
lgrE	Q70LM8	Linear gramicidin dehydrogenase	Antibiotic biosynthetic process
pdhC	P21883	Acetyltransferase	Tricarboxylic acid cycle
fabH	A5VC86	Beta-ketoacyl-[ACP] synthase III	Fatty acid biosynthetic process
*pks15/1***	B2HIL7	Phenolphthiocerol synthesis polyketide synthase type I Pks15/1	Actinobacterium-type cell wall biogenesis
otaA	A0A1R3RGK0	Polyketide synthase otaA	Fatty acid biosynthetic process
acdA	P45867	Acyl-CoA dehydrogenase	Fatty acid beta-oxidation
fadD29	B2HIL4	4-Hydroxyphenylalkanoate adenylyltransferase	Actinobacterium-type cell wall biogenesis
mpl4	Q1ERI2	Aldehyde dehydrogenase mpl4	
*mmpS4**	P0A5K3	Transport accessory protein MmpS4	Exporter system
*pgi**	A0R3N9	Glucose-6-phosphate isomerase	Gluconeogenesis
*ligA**	A1UTB5	DNA ligase	DNA repair
*Acad11**	B3DMA2	Acyl-CoA dehydrogenase family member 11	Fatty acid beta-oxidation
*ephA**	I6YGS0	Epoxide hydrolase A	Response to a toxic substance
*htm MT0586**	P9WKL4	Methyltransferase	Methylation
*sauT **	Q0K844	Sulfoacetate–CoA ligase	Fatty acid metabolic process
*Retsata_zgc**	Q5BLE8	All-trans-retinol 13,14-reductase	Retinol metabolic process

Note: * represents protein found only in rough isolates and ** represents those that overlap in both strains.

Furthermore, we also examined protein clustering within each morphotype. Among the smooth strains, a high degree of similarity (protein clusters) was identified across three smooth isolates; only a small number were unique to individual strains. In contrast, the rough strains shared fewer orthologous protein clusters compared to smooth strains and exhibited a higher number of unique protein clusters, either present in more than one strain or restricted to a single isolate.

Overall, we observed more variations in protein sequences between smooth and rough clinical isolates, compared to the laboratory isolates. In the clinical isolates, many proteins showed differences between smooth and rough morphotypes, and some of them are previously reported, which play a key role in smooth and rough morphotypes [5, 58]. In mycobacteria, MmpL proteins, which are a part of the GPL locus, are essential components and also involved in the biosynthesis of the complex cell envelope. The comparative genomic analysis of a spontaneous rough morphotype variant of *M. abscessus* subsp. *bolletii* revealed a conserved tyrosine residue critical for MmpL protein function [5]. Alteration in this conserved residue led to a defect in glycopeptidolipid production or transport in the rough variant [5]. A molecular genetics and biochemical analysis study of MmpS4 protein in *M. smegmatis* revealed that the protein is required for the production and export of large amounts of cell surface glycolipids but is dispensable for biosynthesis [58]. There are many key enzymes that are related to the fatty acid biosynthetic and cell wall biosynthesis processes that are studied in mycobacteria. For example, LysX bifunctional enzyme involved in lysylphosphatidylglycerol (LPG) biosynthesis, which is an important modification of the LPG lipid core in mycobacteria [59].

In this study, comparative genomic and protein analyses were performed to explore differences between smooth and rough morphotypes, aiming to uncover factors beyond the glycopeptidolipid locus that influence morphotype variation. Although the ANI values indicate that overall genetic content among clinical MABC isolates appeared constrained, notable variations were detected in specific coding regions. When comparing smooth and rough laboratory strains, as well as the NO-exposed rough strain with its parental smooth strain, only a few proteins are observed that differ particularly in acyl-[acyl-carrier-protein] desaturase DesA1 protein. Overall, laboratory strains and the NO-exposed strain exhibited less genetic variation compared to clinical isolates of MABC. However, in clinical strains, the genetic differences were primarily found in the GPL-related non-ribosomal peptide synthetase genes (*mpS1* and *mps2*), as well as in *PapA3*, *mmpl10* and *pks* genes associated with trehalose polyphleate loci. Additional differences were also observed in proteins involved in cell wall biosynthesis, enzymes related to fatty acid and PKS pathways and mycobacterial membrane proteins, which suggests their potential role in morphotype variation. A recent study found that clinical *M. abscessus* isolates exhibit considerably more heterogeneity in morphology, glycopeptidolipid composition and macrophage infection efficiency compared to laboratory strains [60]. In *M. abscessus*, genetic alterations are frequently associated with pathogenic evolution and host adaptation [61]. Comparative genomic and variation analyses showed that clinical isolates differ in protein-coding regions within the same subspecies and across morphotypes, suggesting roles in pathogenicity, virulence and adaptation to the host environment. Whole-genome-based phylogenetics and SNP analysis revealed that smooth isolates were genetically closer to one another, whereas rough isolates showed greater genetic diversity, suggesting that transitions from smooth to rough morphotypes may occur via multiple genetic routes rather than a single locus.

The findings from this study strongly suggest that morphotype variation in MABC is a multifactorial process not limited to mutations in the GPL locus. Instead, a range of genes involved in cell wall and lipid biosynthesis contribute to these phenotypic transitions. The greater genetic diversity among rough isolates suggests that the smooth-to-rough transition occurs through diverse molecular pathways, highlighting the genetic complexity underlying morphotype adaptation in clinical MABC isolates.

Funding information
This work was funded by the Swedish Heart and Lung Foundation, the research council of South-East Sweden (FORSS) and the Swedish Research Council (2023-01895).

Author contributions
Conceptualization: all authors. Sample collection: G.F., L.D.F. and C.B. ONT sequencing: C.B. and R.K. Formal analysis: R.K., C.B. and T.S. Funding acquisition: T.S. Methodology: R.K., C.B. and T.S. Supervision: T.S. Visualization: R.K. and C.B. Writing – original draft: R.K., C.B. and T.S. Writing – review and editing: all authors.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
The study was approved by the Swedish Ethical Review Authority (Dnr 2024-03834-01).

References

1. Blake RJ, McChesney GR, Williams HJ, Holland SM, Lastinger AM. Unusual presentation of *Mycobacteroides abscessus* thigh mass: a case report. *Case Rep Infect Dis* 2025;2025:6970929.
2. Ferrell KC, Johansen MD, Triccas JA, Counoupas C. Virulence mechanisms of mycobacterium abscessus: current knowledge and implications for vaccine design. *Front Microbiol* 2022;13:842017.

3. Diricks M, Merker M, Wetzstein N, Kohl TA, Niemann S, *et al*. Delineating Mycobacterium abscessus population structure and transmission employing high-resolution core genome multilocus sequence typing. *Nat Commun* 2022;13:4936.
4. Parmar S, Tocheva EI. The cell envelope of mycobacterium abscessus and its role in pathogenesis. *PLoS Pathog* 2023;19:e1011318.
5. Bernut A, Viljoen A, Dupont C, Sapriel G, Blaise M, *et al*. Insights into the smooth-to-rough transitioning in *Mycobacterium bolletii* unravels a functional Tyr residue conserved in all mycobacterial MmpL family members. *Mol Microbiol* 2016;99:866–883.
6. Daher W, Leclercq L-D, Johansen MD, Hamela C, Karam J, *et al*. Glycopeptidolipid glycosylation controls surface properties and pathogenicity in *Mycobacterium abscessus*. *Cell Chem Biol* 2022;29:910–924..
7. Aguilera-Correa JJ, Wei F, Leclercq L-D, Tasrini Y, Mullapudi E, *et al*. A dTDP-L-rhamnose 4-epimerase required for glycopeptidolipid biosynthesis in *Mycobacterium abscessus*. *J Biol Chem* 2024;300:107852.
8. Hershko Y, Adler A, Barkan D, Meir M. Glycopeptidolipid defects leading to rough morphotypes of *Mycobacterium abscessus* do not confer clinical antibiotic resistance. *Microbiol Spectr* 2023;11:e0527022.
9. Leclercq LD, Le Moigne V, Daher W, Cortes M, Viljoen B, *et al*. A glycosylated lipooctapeptide promotes uptake and growth of Mycobacterium abscessus in the host. *Nat Commun* 2025;16:3326.
10. Kim B-R, Kim B-J, Kook Y-H, Kim B-J. Phagosome escape of rough *Mycobacterium abscessus* strains in murine macrophage via phagosomal rupture can lead to type I interferon production and their cell-to-cell spread. *Front Immunol* 2019;10:125.
11. Pawlik A, Garnier G, Orgeur M, Tong P, Lohan A, *et al*. Identification and characterization of the genetic changes responsible for the characteristic smooth-to-rough morphotype alterations of clinically persistent *Mycobacterium abscessus*. *Mol Microbiol* 2013;90:612–629.
12. Park IK, Hsu AP, Tettelin H, Shallom SJ, Drake SK, *et al*. Clonal diversification and changes in lipid traits and colony morphology in *Mycobacterium abscessus* clinical isolates. *J Clin Microbiol* 2015;53:3438–3447.
13. Chen Y, Bao R, Li N, Fang T, Yin X, *et al*. Whole-genome recombination and dynamic accessory genomes drive the phenotypic diversity of *Mycobacterium abscessus* subspecies. *Ann Clin Microbiol Antimicrob* 2025;24:44.
14. Nandanwar N, Gibson JE, Neely MN. Growth medium and nitric oxide alter *Mycobacterium abscessus* morphotype and virulence. *Microbiol Res* 2021;253:126887.
15. Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, *et al*. Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*. *PLoS One* 2009;4:e5660.
16. Hedin W, Fröberg G, Fredman K, Chryssanthou E, Selmeryd I, *et al*. A Rough colony morphology of *Mycobacterium abscessus* is associated with cavitary pulmonary disease and poor clinical outcome. *J Infect Dis* 2023;227:820–827.
17. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, *et al*. Evaluation of Oxford nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Sci Rep* 2018;8:10931.
18. Žemaitis L, Palepšienė R, Juzėnas S, Alzbutas G, Burgi P-Y, *et al*. High-performance protocol for ultra-short DNA sequencing using Oxford Nanopore Technology (ONT). *PLoS One* 2025;20:e0318040.
19. Boostrom I, Portal EAR, Spiller OB, Walsh TR, Sands K. Comparing long-read assemblers to explore the potential of a sustainable low-cost, low-infrastructure approach to sequence antimicrobial resistant bacteria with oxford nanopore sequencing. *Front Microbiol* 2022;13:796465.
20. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, *et al*. Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 2021;22:266.
21. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
22. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 2016;32:2103–2110.
23. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci* 2021;1:332–336.
24. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
25. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, *et al*. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom* 2021;7:000685.
26. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, *et al*. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 2012;6:80–92.
27. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
28. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 2016;66:1100–1103.
29. Liu D, Zhang Y, Fan G, Sun D, Zhang X, *et al*. IPGA: a handy integrated prokaryotes genome and pan-genome analysis web service. *Imeta* 2022;1:e55.
30. Illouz M, Leclercq L-D, Dessenne C, Hatfull G, Daher W, *et al*. Multiple *Mycobacterium abscessus* O-acetyltransferases influence glycopeptidolipid structure and colony morphotype. *J Biol Chem* 2023;299:104979.
31. Medema MH, Takano E, Breitling R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol* 2013;30:1218–1223.
32. Burbaud S, Laval F, Lemassu A, Daffé M, Guilhot C, *et al*. Trehalose polyphleates are produced by a glycolipid biosynthetic pathway conserved across phylogenetically distant *Mycobacteria*. *Cell Chem Biol* 2016;23:278–289.
33. Sun J, Lu F, Luo Y, Bie L, Xu L, *et al*. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res* 2023;51:W397–W403.
34. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;16:157.
35. Chhotaray C, Wang S, Tan Y, Ali A, Shehroz M, *et al*. Comparative analysis of whole-genome and methylome profiles of a smooth and a rough *Mycobacterium abscessus* clinical strain. *G3* 2020;10:13–22.
36. Gutiérrez AV, Baron SA, Sardi FS, Saad J, Coltey B, *et al*. Beyond phenotype: the genomic heterogeneity of co-infecting Mycobacterium abscessus smooth and rough colony variants in cystic fibrosis patients. *J Cyst Fibros* 2021;20:421–423.
37. Lewin A, Kamal E, Semmler T, Winter K, Kaiser S, *et al*. Genetic diversification of persistent *Mycobacterium abscessus* within cystic fibrosis patients. *Virulence* 2021;12:2415–2429.
38. Kam JY, Hortle E, Krogman E, Warner SE, Wright K, *et al*. Rough and smooth variants of Mycobacterium abscessus are differentially controlled by host immunity during chronic infection of adult zebrafish. *Nat Commun* 2022;13:952.
39. Mansell G, Gorrie-Stone TJ, Bao Y, Kumari M, Schalkwyk LS, *et al*. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics* 2019;20:366.
40. Epperson LE, Davidson RM, Kammlade SM, Hasan NA, Nick SE, *et al*. Evaluation of the GenoType NTM-DR line probe assay for nontuberculous mycobacteria using whole genome sequences as reference standard. *Diagn Microbiol Infect Dis* 2024;110:116526.
41. Rodriguez-R LM, Conrad RE, Viver T, Feistel DJ, Lindner BG, *et al*. An ANI gap within bacterial species that advances the definitions of intra-species units. *mBio* 2024;15:e0269623.

42. Palmer M, Steenkamp ET, Blom J, Hedlund BP, Venter SN. All anis are not created equal: implications for prokaryotic species boundaries and integration of anis into polyphasic taxonomy. *Int J Syst Evol Microbiol* 2020;70:2937–2948.
43. Kumar R, Bröms JE, Sjöstedt A. Exploring the diversity within the genus *Francisella* - An integrated pan-genome and genome-mining approach. *Front Microbiol* 2020;11:1928.
44. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, *et al*. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020;21:180.
45. Choo SW, Wee WY, Ngeow YF, Mitchell W, Tan JL, *et al*. Genomic reconnaissance of clinical isolates of emerging human pathogen *Mycobacterium abscessus* reveals high evolutionary potential. *Sci Rep* 2014;4:4061.
46. Trisakul K, Sirichoat A, Nithichanon A, Rotcheewaphan S, Phelan J, *et al*. Genomic insights into persistent infections, reinfections, and subspecies diversity of *Mycobacteroides abscessus*: a whole-genome sequencing study of Thai and global isolates. *Infect Genet Evol* 2025;135:105838.
47. Das L, Chaudhary M, Bhardwaj D, Verma S, Tirunagari P, *et al*. Application of genomics in novel microbial species identification. In: *Bioinformatics and Computational Biology*. Chapman and Hall/CRC, 2023. pp. 270–281.
48. Gutiérrez AV, Viljoen A, Ghigo E, Herrmann J-L, Kremer L. Glycopeptidolipids, a double-edged sword of the *Mycobacterium abscessus* complex. *Front Microbiol* 2018;9:1145.
49. Tsai S-H, Lai H-C, Hu S-T. Subinhibitory doses of aminoglycoside antibiotics induce changes in the phenotype of *Mycobacterium abscessus*. *Antimicrob Agents Chemother* 2015;59:6161–6169.
50. Ross BN, Evans E, Diggle F, Briaud P, Whiteley M. The role of colony morphotype in shaping gene essentiality in *Mycobacteroides abscessus Proc Natl Acad Sci U S A* 2025;122:e2500719122.
51. Li B, Ye M, Zhao L, Guo Q, Chen J, *et al*. Glycopeptidolipid genotype correlates with the severity of *Mycobacterium abscessus* lung disease. *J Infect Dis* 2020;221:S257–S262.
52. Daher W, Leclercq L-D, Viljoen A, Karam J, Dufrêne YF, *et al*. *O* -Methylation of the Glycopeptidolipid acyl chain defines surface hydrophobicity of *Mycobacterium abscessus* and macrophage invasion . *ACS Infect Dis* 2020;6:2756–2770.
53. Llorens-Fons M, Pérez-Trujillo M, Julián E, Brambilla C, Alcaide F, *et al*. Trehalose polyphleates, external cell wall lipids in *Mycobacterium abscessus*, are associated with the formation of clumps with cording morphology, which have been associated with virulence. *Front Microbiol* 2017;8:1402.
54. Yeruva VC, Savanagouder M, Khandelwal R, Kulkarni A, Sharma Y, *et al*. The *Mycobacterium tuberculosis* desaturase DesA1 (Rv0824c) is a Ca^{2+} binding protein. *Biochem Biophys Res Commun* 2016;480:29–35.
55. Savanagouder M, Mukku RP, Kiran U, Yeruva CV, Nagarajan N, *et al*. Dissecting the Ca^{2+} dependence of DesA1 function in *Mycobacterium tuberculosis*. *FEBS Lett* 2024;598:1620–1632.
56. Simcox BS. Adaptation to hypoxia and nitrate/nitrite assimilation require species-specific regulation by DosR and NnaR in *Mycobacterium abscessus*; 2024
57. Singh A, Varela C, Bhatt K, Veerapen N, Lee OYC, *et al*. Identification of a desaturase involved in mycolic acid biosynthesis in *Mycobacterium smegmatis*. *PLoS One* 2016;11:e0164253.
58. Deshayes C, Bach H, Euphrasie D, Attarian R, Coureuil M, *et al*. MmpS4 promotes glycopeptidolipids biosynthesis and export in *Mycobacterium smegmatis*. *Mol Microbiol* 2010;78:989–1003.
59. Boldrin F, Cioetto Mazzabò L, Lanéelle M-A, Rindi L, Segafreddo G, *et al*. LysX2 is a *Mycobacterium tuberculosis* membrane protein with an extracytoplasmic MprF-like domain. *BMC Microbiol* 2022;22:85.
60. Pichler V, Dalkilic L, Shoaib G, Shapira T, Rankine-Wilson L, *et al*. The diversity of clinical *Mycobacterium abscessus* isolates in morphology, glycopeptidolipids and infection rates in a macrophage model. *J Med Microbiol* 2024;73:001869.
61. Bryant JM, Brown KP, Burbaud S, Everall I, Belardinelli JM, *et al*. Stepwise pathogenic evolution of *Mycobacterium abscessus Science* 2021;372:eabb8699.