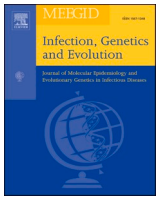
Infection, Genetics and Evolution 125 (2024) 105682

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Genome-scale analysis of *Mycobacterium avium* complex isolates from Portugal reveals extensive genetic diversity

Sofia Carneiro [a,b,1], Miguel Pinto [c,1], Joana Rodrigues [a], João Paulo Gomes [c,d], Rita Macedo [a,*]

[a] National Reference Laboratory for Mycobacteria, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal
[b] Department of Life Sciences, NOVA School of Science and Technology, NOVA University Lisbon, Caparica, Portugal
[c] Genomics and Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal
[d] Animal and Veterinary Research Center (CECAV), Faculty of Veterinary Medicine, Lusófona University-Lisbon University Centre, Lisbon, Portugal

ARTICLE INFO

Keywords:
Mycobacterium avium complex
Nontuberculous mycobacteria
Whole-genome sequencing
Genomic diversity
Phylogenetic core-SNP analysis

ABSTRACT

Opportunist infections caused by nontuberculous mycobacteria (NTM) have emerged as a significant public health problem. Among these, species of the *Mycobacterium avium* complex (MAC) are the main responsible for the increase in the number of human disease cases. In order to address the current needs in the detection and surveillance of MAC disease cases, we evaluated different species classification methodologies (BLASTn-based marker-gene approach, Kraken v2, rMLST and MLST databases) and their congruence with a core-SNP phylogenetic approach, based on whole genome sequencing (WGS) data. For this purpose, we used a collection of 142 MAC isolates from Portuguese patients diagnosed between 2014 and 2022. The marker-gene approach (based on the *rpoB*, *hsp65* and *groEL* genes), showed the best results, allowing the identification of the 142 MAC isolates to the species/subspecies level (*M. avium* subsp. *hominissuis, M. intracellulare, M. intracellulare* subsp. *chimaera, M. intracellulare* subsp. *yongonense, M. marseillence* and *M. colombiense*). Additionally, we performed drug susceptibility testing that confirmed clarithromycin efficacy as a first-line treatment for MAC disease, as 93 % of the Portuguese isolates were susceptible. Using a core-SNP approach we also performed an in-depth phylogenetic analysis within each identified species group, and despite the high genetic diversity within the MAC species, we were able to clearly distinguish all the species/subspecies and identify genetic clusters with epidemiological potential.

We highlight not only the need for the standardization of an appropriate genotyping approach for species identification and management of MAC disease, but also a more robust large-scale WGS data analysis, in a One Health perspective, in order to identify potential routes of transmission.

1. Introduction

Members of *Mycobacterium avium complex* (MAC) are the most common nontuberculous mycobacteria (NTM) causal agents of opportunistic human infections, representing a significant concern to human public health worldwide (Crilly et al., 2021; Van Ingen et al., 2018). These ubiquitous pathogens can be isolated from environmental sources such as soil and both natural and man-made water supply systems (Nishiuchi et al., 2017). Over the last few years, there has been an increase on the number of infections caused by MAC species (Dahl et al., 2022), which are the most common NTM species isolated in North America, Europe, Australia/ New Zealand and some parts of Asia (Dahl et al., 2023; Prevots and Marras, 2015; Zweijpfenning et al., 2018). In Portugal, according to a recent study conducted by our research group, around 40.8 % of NTM infections were caused by MAC members (Santos et al., 2024).

MAC was initially composed by two species, *Mycobacterium avium* and *Mycobacterium intracellulare* (Holt and Daley, 2019). Breakthroughs in molecular taxonomy have allowed the identification of additional/ new species and subspecies, although there is still no consensus in their nomenclature due to the high genetic similarity between some of these species (Mizzi et al., 2022). In 2018, Jakko and colleagues proposed that MAC was composed by twelve species: *Mycobacterium avium, Mycobacterium intracellulare, Mycobacterium chimaera, Mycobacterium*

* Corresponding author at: National Reference Laboratory for Mycobacteria, Department of Infectious Diseases, National Institute of Health, Avenida Padre Cruz, 1649-016 Lisbon, Portugal.
E-mail address: rita.macedo@insa.min-saude.pt (R. Macedo).
[1] both authors contributed equality

https://doi.org/10.1016/j.meegid.2024.105682
Received 16 July 2024; Received in revised form 17 October 2024; Accepted 18 October 2024
Available online 20 October 2024

colombiense, Mycobacterium arosiense, Mycobacterium vulneris, Mycobacterium bouchedurhonense, Mycobacterium timonense, Mycobacterium marseillense, Mycobacterium yongonense, Mycobacterium paraintracellulare and *Mycobacterium lepraemurium* (Van Ingen et al., 2018). Further studies, based on whole genome sequencing (WGS), have proposed that *M. yongonense* and *M. chimaera* are actually subspecies of *M. intracellulare* (Castejon et al., 2018), or even that *M. intracellulare* subsp. *yongonense* is a later heterotopic synonym *of M. intracellulare* subsp. *chimaera*, with the latter being the appropriate designation (Nouioui et al., 2018; Tortoli et al., 2019). Similar findings were observed in the case of *M. paraintracellulare*, suggesting that *M. intracellulare* consisted of only two subspecies: *M. intracellulare* subsp. *intracellulare* (absorbing *M. paraintracellulare*), and *M. intracellulare* subsp. *chimaera* (Tortoli et al., 2019). According to the literature, *M. avium* species are also divided into four subspecies: *M. avium* subsp. *avium*, *M. avium* subsp. *silvaticum*, *M. avium* subsp. *hominissuis* and *M. avium* subsp. *paratuberculosis* (Mijs et al., 2002; Thorel et al., 1990). In 2019, Tortoli and colleagues proposed that *M. avium* subsp. *hominissuis* should be considered a variant of *M. avium* subsp. *avium* and that *M. lapraemurium* should be included as a subspecies of *M. avium* (Tortoli et al., 2019).

Due to the high similarity between some MAC species, currently available techniques, such as hybridization probe assays, sequencing of genes such as *hsp65* and *rpoB*, mycobacterial interspersed repetitive units - variable number of tandem repeats (MIRU-VNTR) analysis, and the more recent matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), are not able to accurately identify and type isolates (Bull et al., 2003; Dauchy et al., 2010; Kim and Shin, 2018; Radomski et al., 2010; Rodriguez-Temporal et al., 2022). Nevertheless, a combined approach of the analysis of specific gene sequences normally used in the identification of NTM, such as *rpoB* and *GroEL* (Mizzi et al., 2022; Rindi and Garzelli, 2014; Telenti et al., 1993), could be an alternative to correctly classifying species and subspecies. In fact, Chawla and colleagues proved that the combined analysis of the *rpoB*, *GroEl-1* and *hsp65* (*GroEL-2*) genes were sufficient to obtain a correct identification of MAC members (Chawla et al., 2023). With an expansion of MAC genomes in public databases, WGS appears to be a better approach for species differentiation while simultaneously enabling us to understand transmission dynamics and/or identify pathogenic or specific genetic traits (e.g. resistance markers) of MAC species (Augusto et al., 2018; Keen et al., 2021; Mizzi et al., 2022). The management of the disease caused by MAC depends on the correct identification of subspecies since there is variability in the ability to cause lung infection, in the severity of the disease, and, consequently, in the prognosis (S.-Y. Kim et al., 2017). For this reason, due to its ubiquitous character and the fact that there are no reports of person-to-person transmission of MAC infections, molecular genotyping becomes essential to understand the epidemiology of MAC and identify sources of infection (Shin et al., 2020).Therefore, to address the current needs in the identification and management of MAC subspecies, we evaluated different classification methodologies and their congruence with a WGS-based phylogenetic approach. We highlight the genetic diversity of MAC strains circulating in Portugal, within a period of nine years (2014–2022), aiming to contribute to a better understanding on the genomic characteristics of this NTM species and its phylogeny.

2. Methods

2.1. Sample dataset characterization

The current study enrols a dataset of 142 MAC strains isolated in Portugal (PT) between 2014 and 2022 at the National Reference Laboratory for Tuberculosis (NRL-TB) of the Portuguese National Institute of Health (NIH). Positive cultures were identified using GenoType *Mycobacterium* CM/NTM-DR® (Hain Lifescience, Germany) according to manufacturer's instructions, or *hsp65* DNA sequencing, as previously described (Hain Lifescience, 2024a, Hain Lifescience, 2024b; Saifi et al., 2013). Social-demographic data were retrieved for epidemiological inferences and included age, gender, region from requesting hospital, and place of residence of the patients (defined according to Portuguese NUTS – Nomenclature of Territorial Units for Statistics). As clinical data was unavailable for most of the cases, case definition was based only on microbiological data. As such, three categories were defined: 1) Definite case – at least three positive NTM cultures, with identical identification, irrespective of sample anatomic source, or a single positive NTM culture isolated from naturally sterile anatomic sources (such as cerebrospinal fluid, biopsies, lavages); 2) Possible NTM disease – positive NTM cultures isolated from unknown anatomic source (e.g., cultures received at the NRL-TB for identification and susceptibility testing without reference of site of disease/infection); 3) NTM colonization – only one positive NTM culture isolated from a respiratory sample (e.g., sputum).

2.2. DNA extraction

Total DNA was extracted from solid/liquid cultures using cetyltrimethylammonium bromide procedure, according to the previously described protocol from the National Institute for Public Health and the Environment (RIVM) (RIVM, 2024) with slight adjustments. Namely, i) after lysozyme was added, suspensions were incubated overnight at 37 °C; and ii) after addition of SDS/proteinase K solution, suspensions were incubated at 65 °C until complete dissolution.

2.3. Drug susceptibility testing

Phenotypic drug susceptibility testing (pDST) was performed on 134 strains (eight isolates could not be recovered for further pDST), according to the Clinical and Laboratory Standards Institute, for clarithromycin (CLR), linezolid (LZD) and moxifloxacin (MXF) (Woods et al., 2011). Briefly, the inoculums were prepared in cation-adjusted Mueller-Hinton broth supplemented with 5 % of oleic albumin dextrose catalase and incubated at 37 °C for seven days. At this point, minimum inhibitory concentrations (MIC) were registered. When growth in the control tube (without antibiotic) was not observed, cultures were re-incubated, and readings registered at days 10 and 14 of incubation (Woods et al., 2011). Molecular susceptibility testing (mDST) was performed by screening mutations in known genetic resistance markers, namely the*16SrRNA* (*rrs;* T1406A, A1408G, C1409T, G1491C, G1491T, and C1496T) (Brown-Elliott et al., 2013; S.-Y. Kim et al., 2021) and *23SrRNA* (*rrl;* A2058N and A2059N*)* (Moon et al., 2016) genes using the GenoType *Mycobacterium* NTM-DR® kit (Hain Lifescience, Germany) (Hain Lifescience, 2024b) and Sanger sequencing (Park et al., 2020).

2.4. MIRU-VNTR typing

MIRU-VNTR typing was performed for isolates that were previously identified as *M. avium, M. intracellulare,* and *M. chimaera* as described above. *M. avium* isolates were genotyped using eight MIRU-VNTR loci as described by Thibault et al. (2007)), with slight modifications. Briefly, DMSO was used instead of betaine in all amplification reactions; for VNTRs 10 and 32, PCR mixtures were performed in a final volume of 25 µL, using 1 U of KAPA HotStart ReadyMix (KAPA Biosystems, USA), 1 µM of each primer and, 1 µL of DMSO (INRA, 2024; Thibault et al., 2007); and annealing temperatures were increased by one degree. For the characterization of *M. intracellulare* and *M. chimaera* isolates, amplification of MIRU-VNTR loci was performed as described by Dauchy et al. (2010). Fragment sizing was determined automatically by the QiaXcel software (Qiagen; Germany), according to the manufacturers' instructions. The genetic relations between isolates were analysed by constructing a minimum spanning tree based on the obtained MIRU-VNTR allelic profiles using the *GrapeTree* software, with the MSTv2 algorithm to account for missing data (i.e., loci for which amplification was unsuccessful) (Zhou et al., 2018).

2.5. Whole genome sequencing and genome assembly

Total genomic DNA of 142 MAC isolates were subjected to Nextera-aXT library preparation (Illumina, USA) before paired-end sequencing (2 × 250 bp or 2 × 150 bp) on either a MiSeq, NextSeq 550 or NextSeq 2000 instrument (Illumina, USA), according to the manufacturer's instructions (Supplementary Table S1). All genome sequences were assembled using the INNUca v4.2.2 pipeline (https://github.com/B-UMMI/INNUca), an integrative bioinformatics pipeline for read quality analysis and de novo genome assembly (Llarena et al., 2018). Briefly, read quality analysis and improvement are performed using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Trimmomatic v0.38 (Bolger et al., 2014), respectively. Genomes were assembled with SPAdes v3.14.0 (Bankevich et al., 2012) and subsequently polished using Pilon v1.23 (Walker et al., 2014), with Quality Assurance/Quality Control (QA/QC) statistics being monitored and reported throughout the analysis (details in Supplementary Table S1).

2.6. Species' classification and phylogenetic analysis

A multi-step approach was applied in order to classify the PT isolates within the MAC. All publicly available complete genomes of MAC (n = 103) were retrieved from Genbank (1st of February 2024; Supplementary Table S2). The nucleotide sequences of *hps65*, *rpoB* and *groEL1* were extracted from all genomes, individually BLASTn queried against the MAC and the reported consensus species at 100 % identity and 100 % query coverage were retrieved for each gene (Chawla et al., 2023). The consensus species classification (henceforth-named marker-gene based consensus classification) across all three genes was then used as the baseline for taxonomic classification. The same strategy was then applied to each novel PT genome. Additionally, for comparison purposes, species classification was also performed on genome assemblies with: i) Kraken v2 (Wood and Salzberg, 2014) (database "Standard-16 10/09/2023", available at https://benlangmead.github.io/aws-indexes/k2); ii) the PubMLST's ribossomal Multilocus Sequencing Typing (rMLST) (Jolley et al., 2012) (https://pubmlst.org/species-id); iii) the PubMLST's *Mycobacteria* spp. MLST (https://pubmlst.org/organisms/mycobacteria-spp) (Jolley et al., 2018). Core-SNP alignment subsets were constructed using parsnp v2.0.5 (Kille et al., 2024), taking advantage of the 103 public genomes and the 142 PT genomes, with either GCA_002219285 (*M. intracellulare* subsp. *chimaera*), GCA_009741445 (*Mycobacterium avium*) or GCA_023278525 (*Mycobacterium intracellulare*) as reference genomes depending on the analysed subset (see Results Section). Phylogenetic analysis was performed by calculating pairwise distance matrices for each core-SNP alignment with snp-dists (https://github.com/tseemann/snp-dists), and performing single-linkage hierarchical clustering using Reportree v2 (Mixão et al., 2023). Simultaneously, using Reportree, cluster compositions were retrieved at all possible SNP thresholds and cluster stability regions (in SNP threshold ranges) were assessed to compare the obtained species classifications with each isolate's phylogenetic clustering.

Since the NRL-TB does not have access to patients' clinical data, making it difficult to associate the molecular relations between the strains with possible epidemiological links, for this study's purpose, we used a conservative threshold of 50 SNPs to define a close related genetic cluster for each subspecies. Additionally, for same-patient samples an in-depth SNP analysis was performed using Snippy (https://github.com/tseemann/snippy), by mapping quality-filtered reads (after Trimmomatic) to one of the assembled genome of the patient pair.

2.7. Data availability

All reads generated for the present study were deposited in the European Nucleotide Archive (ENA) under the study accession number PRJEB57933. Complete details of all isolates and individual ENA run accession numbers are provided in Supplementary Table S1.

3. Results

3.1. Sample dataset characterization

During the study period, the NRL-TB received 1355 NTM-positive cultures, of which 616 were classified as MAC using the genotype kit (Hain Lifescience GmbH, Germany). These were retrieved from 476 patients (256 males and 220 females) with a median age of 67 ± 15 years old. For the present study, we were able to regrow and further analyse 142 MAC strains from 132 patients (64 females and 68 males) with a median age of 68 ± 14 years old. The majority of the cases were from the Lisbon Region (which includes the Lisbon Metropolitan Area and the Setúbal Peninsula), with 85 cases (59.86 %), followed by the Centro region, with 20 cases (14.08 %), North region with 19 cases (13.38 %), Algarve with seven cases (4.93 %), Alentejo with six cases (4.23 %), and the Azores Autonomous Region with five cases (3.52 %). Regarding "case definition", 61 (42.96 %) of the cases were classified as "definite NTM disease", 60 (42.25 %) as "possible NTM disease", and 21 (14.79 %) as "NTM colonization" (see methods section). The strains enrolled in this study (n = 142) were identified at the species level, by the molecular approaches used in the NRL-TB for routine diagnostics purposes (GenoType *Mycobacterium* CM and NTM-DR® kits, or *hsp*65 gene Sanger sequencing), as *M. intracellulare* (60 isolates), *M. avium* (35 isolates), *M. chimaera* (14 isolates) and MAC (33 isolates).

3.2. Drug susceptibility testing

For 134 strains out of the 142 (94 %) we were able to perform pDST (Table 1). MAC strains were mostly susceptible to macrolides (i.e., CLR) with only four strains showing a resistant phenotype. In contrast, for LNZ and MFX, the majority of the isolates presented intermediate or resistant phenotypes, with 29.10 % and 32.09 % of the isolates exhibiting a resistant phenotype, respectively.

mDST revealed that 139 out of 142 isolates had no mutations in the *rrl* gene, and as such 97.89 % of the isolates are presumably susceptible to macrolides (Supplementary Table S1). Still, two isolates (PTNTM_0047 and PTNTM_0107) presented the *rrl* A2059C alteration, suggesting resistance to macrolides. Curiously, one isolate (PTNTM_0037) carried a non-fixed mutation in *rrl* gene, A2058G, likely reflecting an infection with a mixed population. With regard to mutations in the *rrs* gene associated with resistance to aminoglycosides, none of the isolates revealed mutations and are therefore presumably susceptible to this class of antibiotics (Supplementary Table S1). Of note, strain PTNTM_0143 showed phenotypic resistance to CLR without the detection of mutations in the *rrs* gene by mDST. Additionally we observed, for LZD and MXF, higher rates of susceptibility in

Table 1
Phenotypic drug susceptibility testing (pDST) results for MAC isolates (n = 134) in Portugal.

	pDST (n/%)			MIC (μg/mL) Interpretation*		
Antimicrobial*	S	I	R	S	I	R
CLR	125 (93.28 %)	5 (3.73 %)	4 (2.99 %)	≤8	16	≥32
LNZ	66 (49.25 %)	29 (21.64 %)	39 (29.10 %)	≤8	16	≥32
MXF	56 (41.80 %)	35 (26.12 %)	43 (32.09 %)	≤1	2	≥4

* both drug selection and MIC values were retrieved from CLSI M24-A2 guidelines(M24-A2: Susceptibility Testing of Mycobacteria, Nocardiae, and Other Aerobic Actinomycetes; Approved Standard—Second Edition, 2024.). CLR – Clarithromycin; LNZ – Linezolid; MXF – Moxifloxacin; S – Susceptible; I – Intermediate resistance; R – Resistant.

M. intracellulare subsp. *chimaera* (71.4 % and 51.0 % respectively) compared to *M. avium* (32.7 % and 34.5 % respectively) and *M. intracellulare* species (39.3 % and 35.7 % respectively).

3.3. Global genetic diversity and isolate classification

In order to classify Portuguese MAC isolates at the subspecies level, we analysed 103 MAC publicly available reference genomes using three different genetic classifiers including Kraken v2, rMLST and the BLASTn consensus of the *hsp65*, *rpoB* and *groEL1* genes (see methods section). Single-linkage hierarchical clustering (SL-HC) of reference genomes was performed, enrolling 912,787 sites, corresponding to a core-genome of 18.4 % (Fig. S1). A high genetic diversity [172,637 single nucleotide variant sites (SNV), corresponding to 18.9 % polymorphic sites] between MAC species was observed, however, the classifiers returned different results. Both the rMLST approach and the Kraken v2 approach proved to be insufficient in the taxonomic classification of MAC, since it was not possible to differentiate at the subspecies level (Fig. S1). Indeed, only the marker-gene based consensus classification yielded taxonomic classification consistent with the phylogenetic structure, facilitating not only the differentiation of subspecies but also ensuring clear delineation between each of the groups. An extensive SNP distance threshold range was observed (from 4625 up to 6313 SNPs) where most genomes cluster congruently with the marker-gene consensus classification results, with the exception of *M. intracellulare/M. paraintracellulare* and *M. avium* subsp. *hominisuis/M. avium* subsp. *avium*. Curiously, at this threshold range, reference genome GCA_002287605 does not cluster with other genomes identified as *M. intracellulare* subsp. *chimaera* (Fig. S1).

Integration of the PT MAC genomes into the global dataset (i.e., 103 reference genomes, plus 142 PT isolates) reduced the core-genome size from 18.4 % to 9.9 % (912,878 to 492,441 sites), due to the introduction of more genetic diversity from distinct species/subspecies (Fig. S2). Still, this SL-HC analysis, together with the marker-gene consensus classification, allowed the potential classification of one isolate as *M. marseillence* (PTNTM_0033 / ~10,780 SNPs from *M. marseillense* reference genomes) and two isolates as *M. colombiense* (PTNTM_0098 / 18,963 SNPs and PTNTM_0106 / 2555 SNPs from *M. colombiense* reference genome), although they were very distinct from their respective reference genomes, suggesting, once again, high genetic diversity within species/subspecies of this complex.

In order to increase core-SNP resolution, underrepresent groups/members of MAC were removed from subsequent analysis, namely *M. columbiense*, *M. marseillense*, *M. senriense*, *M. mantenii* and *M. lepraemurium*. As such, core-genome size increased to 40.5 %, with 2,010,210 sites under analysis (276,082 SNV), and SL-HC revealed two major branches, separated by a mean pairwise distance of 191,431 ± 144 SNPs, one composed by isolates from the *M. avium* species and the other by isolates from *M. intracellulare* species (Fig. 1). As observed for the reference genome-based analysis (Fig. S1), an extensive SNP distance threshold range was identified displaying high congruence between genome clustering and species classification (from 12,707 up to 15,090 SNPs). As such, this allowed the classification of 31 isolates as *M. intracellulare*, 45 as *M. intracellulare* subsp. *chimaera*, one as *M. intracellulare* subsp. *yongonense* and 57 as *M. avium* subsp. *hominisuis*. The only non-congruent classification was observed for isolate PTNTM_0169 that clusters together with the *M. intracellulare* and *M. paraintracellulare* genomes. Of note, at this threshold range, four PT isolates (PTNTM_0143, PTNTM_0138, PTNTM_0092 and PTNTM_0172) clustered with reference genome GCA_002287605 (Fig. 1). Although these five genome were classified as *M. intracellulare* subps. *Chimaera*, their highly divergent genetic background suggests that these might correspond to a potential distinct subspecies of *M. intracellulare* (hence denominated subsp. *chimaera-like*). Noteworthy, these genomes also had ambiguous classifications results for the *hps65* gene (Fig. S2).

Aiming to further characterize the PT MAC species, the subsequent analysis focused on the two major phylogenetic groups, representing the two predominant species in the dataset: *M. intracellulare* (*M. intracellulare*, *M. intracellulare* subsp. *chimaera, M. intracellulare* subsp. *yongonense and M. intracellulare* subsp. *chimaera-like*) and *M. avium* (*M. avium* subsp. *hominissuis and M. avium* subsp. *avium*).

3.4. Mycobacterium avium species

A draft phylogenetic tree (SL-HC) was reconstructed using a dataset of 104 *M. avium* genomes (57 PT genomes and 47 reference genomes), which enrolled 3,825,777 sites corresponding to a core-genome of 77.2 % (Fig. S3), with 133,716 SNV responsible for the observed genetic diversity. All of our isolates were previously identified as *M. avium subp. Hominissuis* (Fig. 1), and the increase of resolution still showed the clear early separation of the *M. avium* subsp. *paratuberculosis* from other subspecies (mean pairwise distance of 33,918 ± 66 between groups). As a means to potentiate the genome group identified as *M. avium* subsp. *avium*, we identified the earliest clustering region (from 16,836 up to 19,335 SNPs) that separated these from other species groups (Fig. S3). As such, at this threshold range, four major groups were highlighted: Group A – exclusively composed by genomes classified as *M. avium* subsp. *hominissuis*; Group B – exclusively composed by *M. avium* subsp. *avium* genomes; Group C – composed by highly diverse *M. avium* subsp. *hominissuis* genomes (mostly unclustered at this threshold range); and Group D – exclusively composed by *M. avium* subsp. *paratuberculosis* genomes.

For a better understanding of the characteristics of the *M. avium* subsp. *hominissuis* strains circulating in Portugal, the phylogeny was reconstructed using only PT genomes ($N = 57$), which now enrolled 4,069,721 sites corresponding to a core-genome of 81.2 % (Fig. 2). Here, two different groups are included (indicated in Fig. 2 as A and C) with most isolates presenting high SNP distances (overall mean pairwise distance of 23,721 ± 30, and a mean pairwise distance of 29,676 ± 36 between both groups), which emphasises the high genetic diversity within this subspecies. Still, although group A is composed by much more isolates than group C (49 versus 8, respectively) it displays less diversity within it (mean pairwise distance of 21,256 ± 32 versus 30,852 ± 35, respectively). This highlights the high diversity of *M. avium* subsp. *hominissuis* isolates circulating in PT and, as most cases originated in the Lisbon Region (52.6 %, with 16 cases from the Setúbal Peninsula and 14 from the Lisbon Metropolitan Area), the diversity in this geographical region. Within group A, nine close related genetic clusters were identified (≤ 50 SNPs), with the largest cluster enrolling eight isolates susceptible to most antimicrobials tested (only one isolate showed resistance to LNZ), from patients diagnosed between 2017 and 2022, mostly residing in the North region. Five other clusters were composed by isolates collected in the same geographic region, and separated by a mean of one year (ranging from zero up to two years). Within group C, only one genetic cluster was observed, and both cases were from the Lisbon Metropolitan Area and diagnosed in 2022.

Regarding MLST, within this dataset, although most close related isolates possessed the same ST, we observed cases where ST was associated with a polyphyletic structure, namely for ST4 and ST9 (Fig. 2). Moreover, although 55 unique MIRU-VNTR profiles were generated for the 57 isolates (Fig. S4), these were not concordant with WGS clustering analysis. In fact, even the isolates from the same patient (PTNTM_0037 and PTNTM_0047) that only differed by less than 40 SNPs did not group with the same MIRU-VNTR code, proving some inefficacy for this molecular typing technique.

3.5. Mycobacterium intracellulare subspecies

As mentioned above, 82 PT isolates were classified as belonging to *M. intracellulare* species. These strains were isolated from 77 patients (41 males and 36 females), mostly from the Lisbon Region (37 cases from the Setúbal Peninsula and 14 from the Lisbon Metropolitan Area). A draft phylogenetic tree was reconstructed using 132 *M. intracellulare* genomes

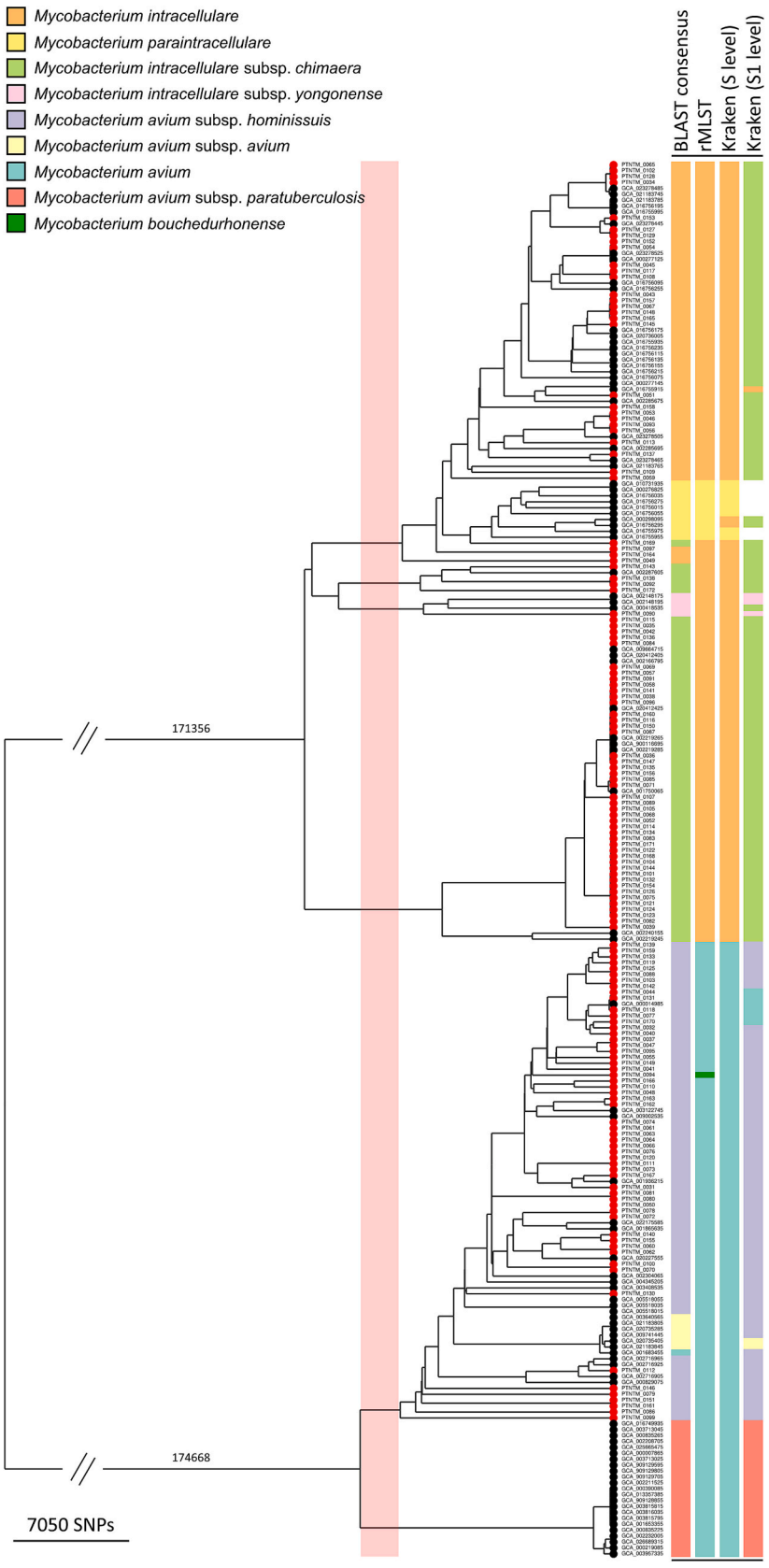


Fig. 1. Comparison of different *Mycobacterium avium complex* genomes classification methods with core-SNP-based single-linkage hierarchical clustering. Analysis enrols 97 reference genomes (black nodes) and 139 Portuguese MAC genomes (red nodes). Core-SNP alignment was reconstructed using GCA_009741445 as a reference. BLAST consensus, corresponds to the combined BLAST results for *hsp65*, *groEL1* and *rpoB* genes. Red box highlights the cluster congruence region with BLAST consensus species identification (corresponding to a SNP distance threshold of 12,707–15,090). Due to scaling cut, length values are presented for the first two branches. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

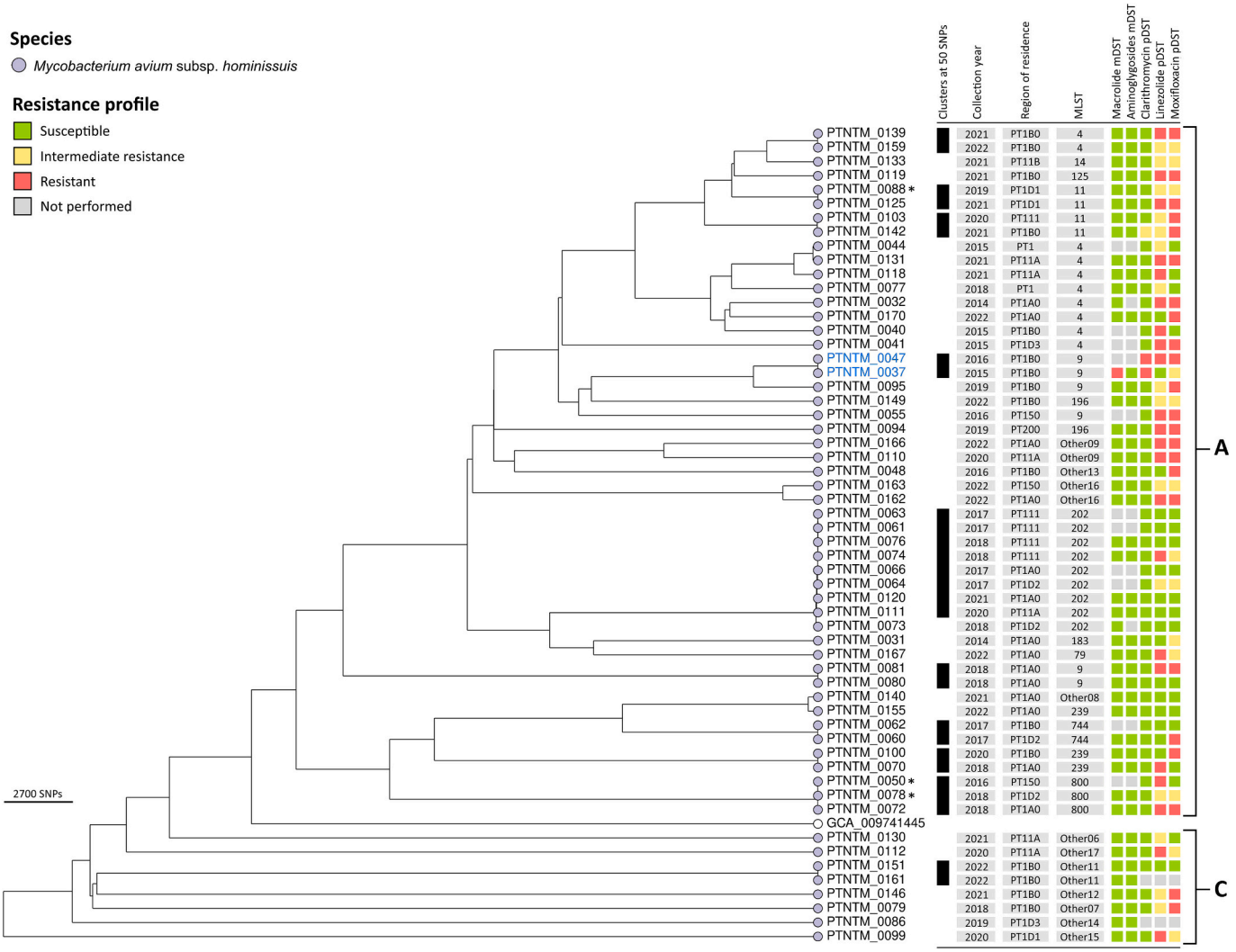


Fig. 2. Core-SNP-based single-linkage hierarchical clustering of Portuguese *Mycobacterium avium* species genomes ($N = 57$). Core-SNP alignment was reconstructed using GCA_009741445 as a reference. Coloured sample labels correspond to same patient samples. *Patient with a MAC reinfection with another strain or species in the dataset. Nodes are coloured according to marker-gene consensus identification. Group A and C refer to groups identified in Fig. S3. Regions of residence are presented according to the Portuguese NUTIII as: PT1B0 – Lisbon Metropolitan Area; PT11B – Alto Tamega; PT1D1 – Oeste; PT111 – Alto Minho; PT1 – Portugal Continent; PT11A – Porto Metropolitan Area; PT1D3 – Leziria do Tejo; PT1A0 – Setúbal Peninsula; PT150 – Algarve; PT200 – Azores Autonomous Region; PT1D2 – Medio Tejo.

(82 PT genomes and 50 closed public genomes), which enrolled 4,100,630 sites corresponding to a core-genome of 75.9 % (Fig. S5). Results revealed an extensive stable clustering region from 26,228 up to 39,550 SNP distance thresholds where four major groups could be identified: Group E – enrolling *M. intracellulare* and *M. paraintracellulare* genomes; Group F – exclusively enrolling close related *M. intracellulare* subsp. *chimaera* genomes; Group G – enrolling highly divergent *M. intracellulare* subsp. *chimaera* genomes (subsp. *chimaera-like*); and Group H – exclusively enrolling *M. intracellulare* subsp. *yongonense* genomes. Of note, isolate PTNTM_0049 (*M. intracellulare)* is the only genome that segregates early within the phylogeny (Fig. S5), suggesting a potential distinct *M. intracellulare* subspecies/lineage.

To inspect PT genomes classified as *M. intracellulare*, a SL-HC phylogenetic tree based on the 31 isolates that composed the group E (identified in Fig. S5) was reconstructed using GCA_023278525 as a reference genome, which enrolled 4,628,344 sites (corresponding to a core-genome of 85.6 %) and 113,876 SNVs (Fig. 3). Only one isolate of this group had been classified as *M. intracellulare* subsp. *chimaera* (PTNTM_0169) by the maker-based consensus method. Although this isolate seems to be genetically more related to the *M. intracellulare* group, and was therefore included in this analysis, it segregates in a more distant branch, being separated from the rest of the group by a mean of 41,101 ± 95 SNPs. Most isolates were collected between 2020 and 2022 (61.2.%) from the Lisbon Region (48.4 %, which includes the Lisbon Metropolitan Area and the Setúbal Peninsula). Within this subspecies, only one strain displayed an intermediate resistance phenotype to CLR, whereas twelve and ten strains exhibited resistant phenotypes to LNZ and MXF, respectively. Notably, none of the strains showed mutations within the *rrs* or *rrl* genes using mDST. Applying the conservative 50 SNP threshold defined above (see methods section) eight close related genetic clusters were revealed (Fig. 3), two of these enrolling known same-patient samples. Still, in general, all clusters included isolates from distinct geographic regions and spanning different collection years. MLST typing for *M. intracellulare* showed a high congruence with the obtained phylogeny at a SNP threshold of 9226 up to 13,390, highlighting its usefulness for long-term surveillance purposes. In contrast, MIRU-VNTR did not show good typing resolution, as isolates clustered by WGS were not grouped using MIRU-VNTR (Fig. S6).

Regarding *M. intracellulare* subsp. *chimaera*, using 45 PT isolates (Fig. S5; Group F) and the reference genome GCA_002219285, a draft

phylogenetic tree was constructed, relying on 5,038,386 sites (corresponding to a core genome of 85.9 %) and 29,189 SNV (Fig. 4). Two main genetic groups could be identified (F1 and F2 in Fig. 4), with two additional isolates displaying a mean pairwise distance of 15,687 ± 19 to these groups (PTNTM_0039 and PTNTM_0107). Group F1 and F2 were separated by a mean pairwise distance of 8715 ± 30 (mean within each group of 1333 ± 10 and 31 ± 2, respectively). Group F1 includes 22 isolates from 2014 up to 2022, mostly collected in the Setúbal Peninsula (54.5 %), and exhibits two major clusters separated by ~5000 SNPs. Close inspection of these shows a close related genetic cluster (< 50 SNPs), enrolling 17 and 3 isolates, respectively. None of the isolates in F1 showed resistance to CLR and only three and two isolates showed resistance to LNZ and MXF, respectively. As with F1, group F2 includes 21 isolates from 2015 up to 2022, mostly collected in the Setúbal Peninsula (85.7 %), and exhibits a single large close-related genetic cluster composed of 19 isolates. All the isolates were sensitive to CLR and only two showed phenotypic resistance to LNZ. Still, 38 % of isolates within F2 were resistant to MXF. Regarding aminoglycosides and macrolides mDST, none of the isolates from both groups harboured genetic alterations associated with resistance. In fact, only the divergent isolate PTNTM_0107 was presumed resistant to macrolides by both mDST and pDST. Of note, although three pairs of same-patient samples were identified in this dataset, only one (PTNTM_0154 / PTNTM_0123) clustered together at <50 SNPs. Additionally, all isolates shared the same ST81 (with the exception of PTNTM_0068) likely due to the close genetic relatedness within this subspecies. In fact, when analysing all isolates identified as *M. intracellulare* subsp. *chimaera* [i.e., including the additional four genetically divergent *chimaera-like* isolates (group G in Fig. S5)], all subspecies *chimaera-like* presented distinct MLST profiles (Supplementary Table S1). Again, this highlights MLST typing adequacy for long-term surveillance of *M. intracellulare* species, and supports the suggestion that the *chimare-like* isolates are genetically different from the rest of the *M. intracellulare* subsp. *chimaera* strains circulating in Portugal, as well as most reference genomes similarly classified. Nevertheless, once again, MIRU-VNTR showed poor sensitivity, as it showed low congruence with the identified clusters by WGS (Fig. S6).

3.6. Re-infection or persistence

Within our dataset, for ten patients we retrieved more than one sample, collected more than six months apart, in order to evaluate possible cases of persistence of infection. In four cases, re-infection was clear as each patient was infected with isolates belonging to different MAC species (highlighted in Figs. 2, 3 and 4, with details in Supplementary Table S1), i.e., for patient 34 (PTNTM_0087 / PTNTM_0049), patient 37 (PTNTM_0059 / PTNTM_0088), patient 43 (PTNTM_0126 / PTNTM_0050) and patient 136 (PTNTM_0065 / PTNTM_0078). Taking advantage of the obtained complete genomes, for the other six cases, a fine-tuned paired analysis was conducted revealing, as observed above in the subspecies analysis, ≤ 40 genetic alterations (either SNPs or indels, ranging from one up to 40) between isolates from five patients, namely patient 13 (PTNTM_0121 / PTNTM_0083), 20 (PTNTM_0087 /

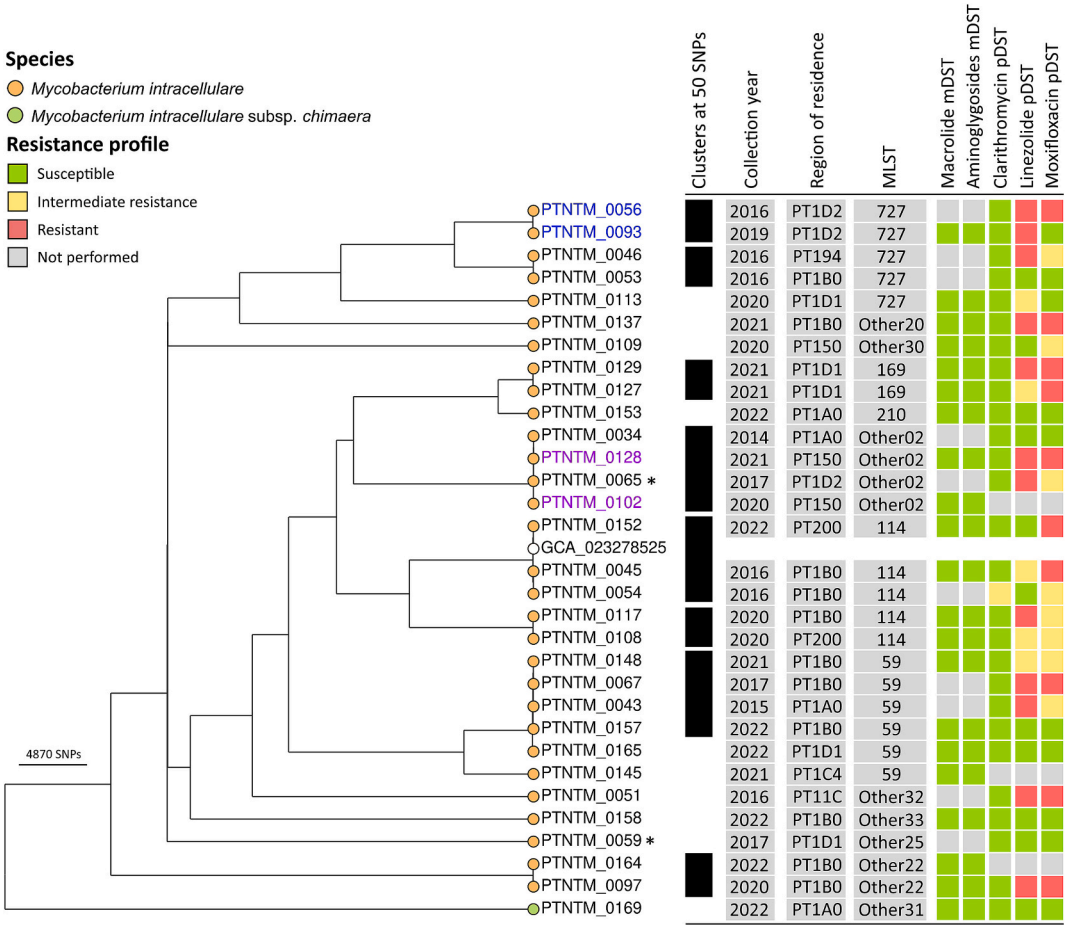


Fig. 3. Core-SNP-based single-linkage hierarchical clustering of Portuguese *Mycobacterium intracellulare* genomes ($N = 31$). Core-SNP alignment was reconstructed using GCA_023278525 as a reference. Coloured sample labels correspond to same patient samples. *Patient with a MAC reinfection with another strain or species in the dataset. Nodes are coloured according to marker-gene consensus identification. Regions of residence are presented according to the Portuguese NUTIII as: PT1D2 – Medio Tejo; PT1B0 – Lisbon Metropolitan Area; PT1D1 – Oeste; PT150 – Algarve; PT200 – Azores Autonomous Region; PT1A0 – Setúbal Peninsula; PT11C – Tamega e Sousa; PT194 – Viseu Dao Lafoes; PT1C4 – Central Alentejo.

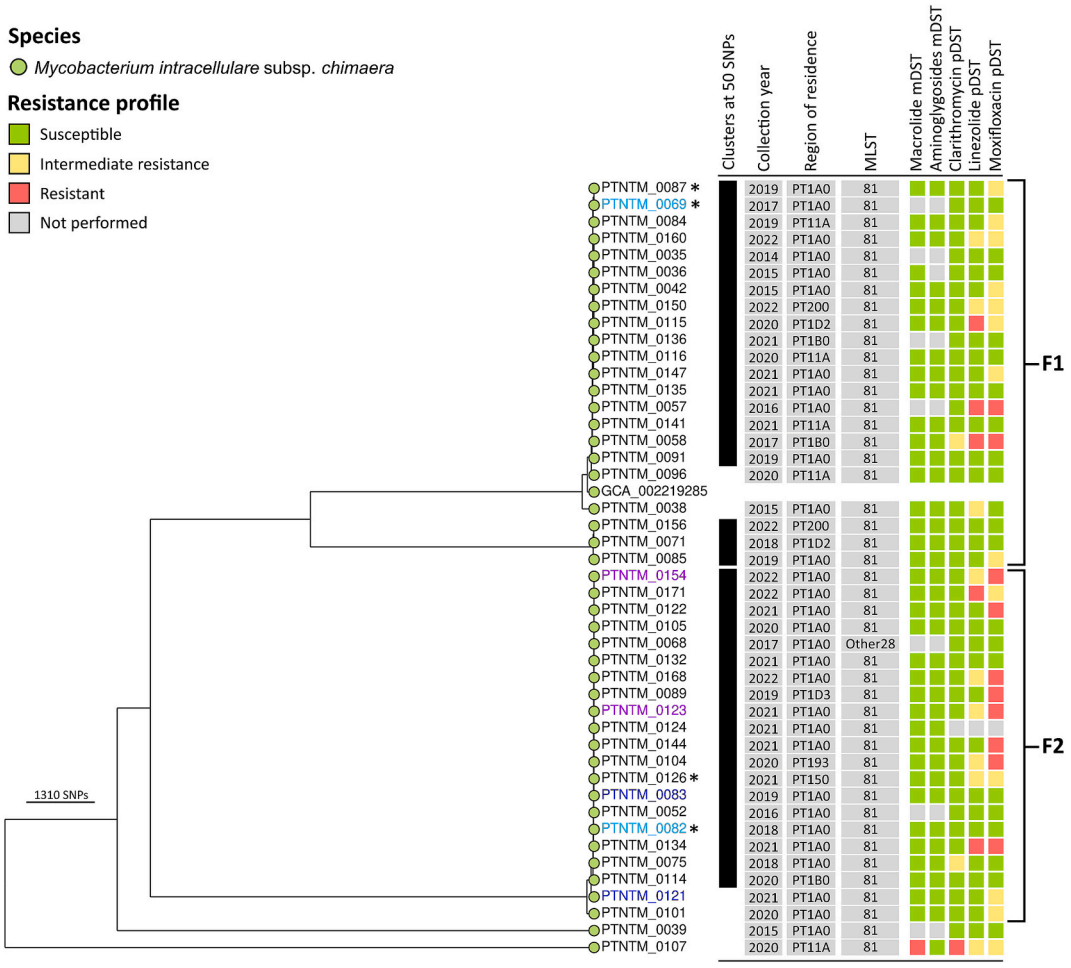


Fig. 4. Core-SNP-based single-linkage hierarchical clustering of Portuguese *Mycobacterium intracellulare* subsp. *chimaera* genomes ($N = 45$). Core-SNP alignment was reconstructed using GCA_002219285 as a reference. Coloured sample labels correspond to same patient samples. *Patient with a MAC reinfection with another strain or species in the dataset. Nodes are coloured according to marker-gene consensus identification. Regions of residence are presented according to the Portuguese NUTIII as: PT1B0 – Lisbon Metropolitan Area; PT1D2 – Medio Tejo; PT200 – Azores Autonomous Region; PT1A0 – Setúbal Peninsula; PT11A – Porto Metropolitan Area; PT1D3 – Leziria do Tejo; PT193 – Leiria Region; PT150 – Algarve.

PTNTM_0049), 30 (PTNTM_0154 / PTNTM_0123), 49 (PTNTM_0102 / PTNTM_0128), and 101 (PTNTM_0056 / PTNTM_0093). Considering that most paired isolates were collected within a one- or two-year interval, these data suggest that these cases may constitute persistent infections with strains displaying within-patient evolution. Of note, for patient 101, even though isolates were collected three years apart, only one SNP difference was observed between them. For patient 130 (PTNTM_0069 / PTNTM_0082) isolates showed >7000 genetic differences between them, suggesting that, although both isolates belonged to *M. intracellulare subsp. chimaera*, patient was, at some point, re-infected with a different strain (Fig. 4).

4. Discussion

Currently, there is still a huge lack of knowledge not only on the global distribution of MAC subspecies but also on methodologies used for their accurate identification. This study aimed to enrich comprehension on MAC diversity and characterize the population of MAC isolates circulating in Portugal by integrating them within the available reference MAC genomes and applying different methodological approaches for species classification. To achieve this, we analysed 142 isolates collected in Portugal between 2014 and 2022, mainly from male patients (51.5 %) from Lisbon region (59.86 %). MAC isolates were mostly susceptible to CLR (3.0 % of resistant isolates), reinforcing its use as a primary therapeutic option, whereas only about half of the isolates were susceptible to LNZ and MXF (49.3 % and 41.8 %, respectively).

As highlighted in Fig. 1, MAC seems to be composed by two species (*Mycobacterium avium* and *Mycobacterium intracellulare groups*), however, by performing a deeper analysis of each of this species we clearly see the emergence of smaller groups of additional major lineages within each of these major branches. Using different classification approaches (Kraken, rMLST and marker-gene based results) we were able to classify our isolates at the subspecies level. The results obtained with the marker-gene based consensus approach were the most congruent with the SNP tree analysis of each species (Fig. 1 and Fig. S2). Thus, we believe that, although there is still much to debate concerning taxonomic classification of MAC (Mizzi et al., 2022), this approach may be useful for quick classification of isolates at the species and subspecies level, proving its efficacy for identification purposes. Other studies have also reported the advantage of using the analysis of the *rpoB groEL-1 and groEL-2 (hsp65)* genes for NTM identification (Chawla et al., 2023; Rindi and Garzelli, 2014; Van Ingen et al., 2018). In fact, Chawla and colleagues, the first suggesting the use of a consensus analysis of these genetic markers, also reported sufficient discriminatory power for accurate MAC subspecies identification (Chawla et al., 2023). Still, the differences observed in classification for the distinct methodologies also highlight the need to have larger and more diverse genomic datasets, particularly relevant for underrepresented species (such as

M. lepraemurium or *M. mantenii*). Additionally, powerful software such as Kraken, which relies on known genomic information, would benefit from this to perform identification of MAC species (namely up to the subspecies level). Although we found a general agreement between the phylogenetic analysis and the taxonomic classification determined by the marker-gene consensus method, highly divergent strains were also observed. Thus, additional genome-scale data would strengthen our understanding on the population structure of MAC while helping in the establishment of a more robust taxonomic classification of the species of this complex.

Phylogenetic analysis of the PT *M. avium* isolates, revealed, not only that all analysed cases were caused by *M. avium* subsp. *hominissuis*, but also a division into two potential highly genetically distinct groups (Fig. 2, groups A and C). These results are in line with previous genomic studies that also showed great diversity among the isolates of this subspecies, which is probably related to the different environmental niches, different hosts and even different disease outcomes (Iwamoto et al., 2012; Mizzi et al., 2022; Uchiya et al., 2017). Noteworthy, we observed that PT *M. avium* isolates were more susceptible to LNZ than previous reported published data (32.7 % vs 24.7 %) (Cho et al., 2018). Although the majority of our cases of *M. avium* were collected in the Lisbon region and we observed at least 10 clusters of close related isolates ($\leq$ 50 SNPs), no conclusions regarding their linkage can be drawn since the LNR-TB does not have access to clinical/epidemiological data, thereby hindering the establishment of relations between molecular and potential epidemiological links.

It is still not established how to taxonomically classify *M. intracellulare* group and some authors even consider that *M. chimaera* is a subspecies of *M. intracellulare* (Nouioui et al., 2018). Our study revealed the existence of three phylogenetically distinct groups, separating the isolates of *M. intracellulare*, *M. intracellulare* subsp. *chimaera* and *M. intracellulare* subsp. *yongonense*, as well as a potential fourth group composed of *M. intracellulare* subsp. *chimaera-like* isolates (Fig. S5, groups E, F, H and G, respectively). However, although *M. intracellulare* subsp. *yongonense* genomes clearly diverge from the remaining *M. intracellulare* genomes, we highlight the lack of both PT isolates and publicly available reference genomes, hindering the clarification whether *M. intracellulare* subsp. *yongonense* is really a subspecies or just a variant of *M. chimaera* (Castejon et al., 2018; Tortoli et al., 2019; Van Ingen et al., 2018). Still, even with limited data, the phylogenetic analysis conducted on the present study supports the former being more likely than the latter given their early divergence from the large *M. intracellulare* subsp. *chimaera* (Fig. 1). It has already been described that isolates of *M. chimaera* are linked to outbreaks associated with hospital heater-cooler units (Quintás Viqueira et al., 2021; Uchiya et al., 2017), suggesting this subspecies is more likely to contaminate/inhabit these hospital niches than the other species of the complex, again suggesting well defined associations between origin and causative agent of infection. In fact, previous studies focusing on the genomic characterization of *M. intracellulare* subsp. *chimaera* isolates associated with the occasional contamination of heater-coolers, and subsequently cardiac surgery patients, revealed that those were caused by very genetically related strains (Lecorche et al., 2021; Schreiber et al., 2021) regardless of country of origin, and thus suggesting the existence of potential international highly fitted dominant lineages. This could lead to the assumption that, as already demonstrated for other NTM diseases, particularly, *M. abscessus* (Carneiro et al., 2023; Diricks et al., 2022), there may also exist Dominant Circulating Clones (DCC, terminology used for *M. abscessus*). In our dataset, we also observed low genetic diversity between *M. intracellulare* subsp. *chimaera* isolates, and only two major clusters (groups F1 and F2 in Fig. 4) were identified suggesting the existence of potential DCC in Portugal. Nevertheless, more studies focused on the genomic comparison of isolates collected from environmental sources and isolates from patients worldwide are needed. Regarding antimicrobial resistance, contrary to previous reports (Fernandez-Pittol et al., 2022), PT *M. intracellulare* subsp. *chimaera* isolates were more susceptible to MXF (52.03 %) compared to those of *M. avium* (34.5 %) and *M. intracellulare* (35.7 %). Once more, we believe that additional data are still needed to understand the differences in susceptibility patterns among MAC isolates across different geographical regions. It could be speculated that variations in antibiotic susceptibility patterns may be linked to delayed clinical diagnostics or to therapeutic schemes not guided by DST results, promoting the spread of more resistant strains.

WGS analysis allowed the identification of five cases of potential persistence of infection with the isolates showing less than 40 genetic alterations (including SNPs or indels) and five cases of clear reinfections, four of which with different species of MAC and one with the same subspecies. In fact, although poorly documented, cases of relapse with the same or different MAC species have also been reported (Boyle et al., 2016; Kwon et al., 2019; Wallace Jr. et al., 2002). This high number of cases of reinfection (~4 % of the total cases analysed) are also an evidence of the diversity of the MAC species in the environment and, consequently, high human exposure to potential sources of infection. Persistent cases, however, may be associated with treatment failure.

Regarding the limitations of this study, we highlight our lack of access to clinical data of the patients. We believe that it could have highly benefited from more detailed clinical and demographic relevant information in order to eventually relate the genomic data to distinct infection sources and/or disease outcomes of each patient, or to flag potential epidemiological links in the context of Public Health investigations. It is also of note that MAC disease is not mandatory notifiable, which also contributes to the lack of information, and there are no guidelines to the follow up on the diagnosed patients, leading, in some cases, to a potential chronic disease, that could otherwise be cured.

In order to achieve a global understanding of MAC in a One Health perspective, indiscriminate large-scale WGS data of MAC clinical and environmental isolates is needed both to establish associations between MAC species/subspecies and human disease outcomes and to reliably identify common sources of infection for epidemiological purposes. Moreover, establishing an appropriate genotyping approach for each MAC species is essential in managing disease progression, as it has been shown that identifying MAC subspecies is key for prioritizing patient treatment (Pan et al., 2021). We believe that by improving our understanding of the disease-subspecies dynamics in MAC infections, the application of WGS will have an impact on patient care that goes beyond laboratory diagnosis.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2024.105682.

CRediT authorship contribution statement

Sofia Carneiro: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Miguel Pinto:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Joana Rodrigues:** Writing – review & editing, Methodology, Investigation. **João Paulo Gomes:** Writing – review & editing, Validation, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Rita Macedo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All reads generated for the present study were deposited in the

European Nucleotide Archive (ENA) under the study accession number PRJEB57933. Complete details of all isolates and individual ENA run accession numbers are provided in Supplementary Table S1.

Acknowledgements

S. Carneiro was supported by the Portuguese Science and Technology Foundation (FCT) through grant 2020.08503.BD. This study was partially supported, by the Health Emergency Preparedness and Response (HERA) project (Grant/2021/PHF/23776) supported by the European Commission through the European Centre for Disease Control and Prevention and the project EU4H-2022-DGAMS-IBA-1 (https://www.insa.min-saude.pt/category/projectos/geneo/) supported by the European Union on behalf of EU4H programme.

References

Augusto, C.J., Carvalho, W.D.S., Almeida, I.N.D., Figueiredo, L.J.D.A., Dantas, N.G.T., Suffys, P.N., Miranda, S.S.D., 2018. Comparative study of RFLP-IS6110 and MIRU-VNTR from Mycobacterium tuberculosis isolated in the state of Minas Gerais, Brazil. Braz. J. Microbiol. 49 (3), 641–646. https://doi.org/10.1016/j.bjm.2017.04.015.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477. https://doi.org/10.1089/cmb.2012.0021.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Boyle, D.P., Zembower, T.R., Qi, C., 2016. Relapse versus reinfection of *Mycobacterium avium* complex pulmonary disease. Patient characteristics and macrolide susceptibility. Ann. Am. Thorac. Soc. 13 (11), 1956–1961. https://doi.org/10.1513/AnnalsATS.201605-344BC.

Brown-Elliott, B.A., Iakhiaeva, E., Griffith, D.E., Woods, G.L., Stout, J.E., Wolfe, C.R., Turenne, C.Y., Wallace, R.J., 2013. In vitro activity of amikacin against isolates of *Mycobacterium avium* complex with proposed MIC breakpoints and finding of a 16S rRNA gene mutation in treated isolates. J. Clin. Microbiol. 51 (10), 3389–3394. https://doi.org/10.1128/JCM.01612-13.

Bull, T.J., Sidi-Boumedine, K., McMinn, E.J., Stevenson, K., Pickup, R., Hermon-Taylor, J., 2003. Mycobacterial interspersed repetitive units (MIRU) differentiate *Mycobacterium avium* subspecies paratuberculosis from other species of the *Mycobacterium avium* complex. Mol. Cell. Probes 17 (4), 157–164. https://doi.org/10.1016/S0890-8508(03)00047-1.

Carneiro, S., Pinto, M., Silva, S., Santos, A., Rodrigues, I., Santos, D., Duarte, S., Vieira, L., Gomes, J.P., Macedo, R., 2023. Genome-scale characterization of *Mycobacterium abscessus* complex isolates from Portugal. Int. J. Mol. Sci. 24 (20), 15402. https://doi.org/10.3390/ijms242015402.

Castejon, M., Menéndez, M.C., Comas, I., Vicente, A., Garcia, M.J., 2018. Whole-genome sequence analysis of the *Mycobacterium avium* complex and proposal of the transfer of Mycobacterium yongonense to *Mycobacterium intracellulare* subsp. Yongonense subsp. Nov. Int. J. Syst. Evol. Microbiol. 68 (6), 1998–2005. https://doi.org/10.1099/ijsem.0.002767.

Chawla, R., Shaw, B., Von Bredow, B., Chong, C., Garner, O.B., Zangwill, K.M., Yang, S., 2023. Accurate subspecies-level identification of clinically significant *Mycobacterium avium* and *Mycobacterium intracellulare* by whole-genome sequencing. J. Microbiol. Methods 208, 106726. https://doi.org/10.1016/j.mimet.2023.106726.

Cho, E.H., Huh, H.J., Song, D.J., Moon, S.M., Lee, S.-H., Shin, S.Y., Kim, C.K., Ki, C.-S., Koh, W.-J., Lee, N.Y., 2018. Differences in drug susceptibility pattern between *Mycobacterium avium* and *Mycobacterium intracellulare* isolated in respiratory specimens. J. Infect. Chemother. 24 (4), 315–318. https://doi.org/10.1016/j.jiac.2017.10.022.

Crilly, N.P., Ayeh, S.K., Karakousis, P.C., 2021. The new frontier of host-directed therapies for *Mycobacterium avium* complex. Front. Immunol. 11, 623119. https://doi.org/10.3389/fimmu.2020.623119.

Dahl, V.N., Mølhave, M., Fløe, A., Van Ingen, J., Schön, T., Lillebaek, T., Andersen, A.B., Wejse, C., 2022. Global trends of pulmonary infections with nontuberculous mycobacteria: A systematic review. Int. J. Infect. Dis. 125, 120–131. https://doi.org/10.1016/j.ijid.2022.10.013.

Dahl, V.N., Laursen, L.L., He, Y., Zhang, Y.-A., Wang, M.-S., 2023. Species distribution among patients with nontuberculous mycobacteria pulmonary disease in Europe. J. Inf. Secur. 87 (5), 469–472. https://doi.org/10.1016/j.jinf.2023.03.010.

Dauchy, F.-A., Dégrange, S., Charron, A., Dupon, M., Xin, Y., Bébéar, C., Maugein, J., 2010. Variable-number tandem-repeat markers for typing *Mycobacterium intracellulare* strains isolated in humans. BMC Microbiol. 10 (1), 93. https://doi.org/10.1186/1471-2180-10-93.

Diricks, M., Merker, M., Wetzstein, N., Kohl, T.A., Niemann, S., Maurer, F.P., 2022. Delineating *Mycobacterium abscessus* population structure and transmission employing high-resolution core genome multilocus sequence typing. Nat. Commun. 13 (1), 4936. https://doi.org/10.1038/s41467-022-32122-5.

Fernandez-Pittol, M., Batista-Arnau, S., Román, A., San Nicolás, L., Oliver, L., González-Moreno, O., Martínez, J.A., Amaro-Rodríguez, R., Soler, N., Gené, A., González-Cuevas, A., Tudó, G., Gonzalez-Martin, J., 2022. Differences in drug-susceptibility patterns between *Mycobacterium avium*, *Mycobacterium intracellulare*, and *Mycobacterium chimaera* clinical isolates: prospective 8.5-year analysis by three laboratories. Antibiotics 12 (1), 64. https://doi.org/10.3390/antibiotics12010064.

Hain Lifescience, 2024a. GenoType Mycobacterium CM VER 2.0 (n.d.-a-A). https://www.hain-lifescience.de/en/products/microbiology/mycobacteria/ntm/genotype-mycobacterium-cm.html.

Hain Lifescience, 2024b. GenoType NTM-DR VER 1.0 (n.d.-b-B). https://www.hain-lifescience.de/en/products/microbiology/mycobacteria/ntm/genotype-ntm-dr.html.

Holt, M.R., Daley, C.L., 2019. *Mycobacterium avium* complex disease. In: Griffith, D.E. (Ed.), Nontuberculous Mycobacterial Disease: A Comprehensive Approach to Diagnosis and Management. Springer International Publishing, pp. 301–323. https://doi.org/10.1007/978-3-319-93473-0_11.

INRA, 2024. MAC-INMV-SSR Database. n.d. http://mac-inmv.tours.inra.fr/index.php.

Iwamoto, T., Nakajima, C., Nishiuchi, Y., Kato, T., Yoshida, S., Nakanishi, N., Tamaru, A., Tamura, Y., Suzuki, Y., Nasu, M., 2012. Genetic diversity of *Mycobacterium avium* subsp. Hominissuis strains isolated from humans, pigs, and human living environment. Infect. Genet. Evol. 12 (4), 846–852. https://doi.org/10.1016/j.meegid.2011.06.018.

Jolley, K.A., Bliss, C.M., Bennett, J.S., Bratcher, H.B., Brehony, C., Colles, F.M., Wimalarathna, H., Harrison, O.B., Sheppard, S.K., Cody, A.J., Maiden, M.C.J., 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158 (4), 1005–1015. https://doi.org/10.1099/mic.0.055459-0.

Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 3, 124. https://doi.org/10.12688/wellcomeopenres.14826.1.

Keen, E.C., Choi, J., Wallace, M.A., Azar, M., Mejia-Chew, C.R., Mehta, S.B., Bailey, T.C., Caverly, L.J., Burnham, C.-A.D., Dantas, G., 2021. Comparative genomics of *Mycobacterium avium* complex reveals signatures of environment-specific adaptation and community acquisition. mSystems 6 (5). https://doi.org/10.1128/mSystems.01194-21 e01194-21.

Kille, B., Nute, M.G., Huang, V., Kim, E., Phillippy, A.M., Treangen, T.J., 2024. Parsnp 2.0: Scalable Core-Genome Alignment for Massive Microbial Datasets. https://doi.org/10.1101/2024.01.30.577458.

Kim, S.H., Shin, J.H., 2018. Identification of nontuberculous mycobacteria using multilocus sequence analysis of 16S rRNA, *hsp65*, and *rpoB*. J. Clin. Lab. Anal. 32 (1), e22184. https://doi.org/10.1002/jcla.22184.

Kim, S.-Y., Shin, S.H., Moon, S.M., Yang, B., Kim, H., Kwon, O.J., Huh, H.J., Ki, C.-S., Lee, N.Y., Shin, S.J., Koh, W.-J., 2017. Distribution and clinical significance of *Mycobacterium avium* complex species isolated from respiratory specimens. Diagn. Microbiol. Infect. Dis. 88 (2), 125–137. https://doi.org/10.1016/j.diagmicrobio.2017.02.017.

Kim, S.-Y., Kim, D.H., Moon, S.M., Song, J.Y., Huh, H.J., Lee, N.Y., Shin, S.J., Koh, W.-J., Jhun, B.W., 2021. Association between 16S rRNA gene mutations and susceptibility to amikacin in *Mycobacterium avium* complex and *Mycobacterium abscessus* clinical isolates. Sci. Rep. 11 (1), 6108. https://doi.org/10.1038/s41598-021-85721-5.

Kwon, B.S., Shim, T.S., Jo, K.-W., 2019. The second recurrence of *Mycobacterium avium* complex lung disease after successful treatment for first recurrence. Eur. Respir. J. 53 (1), 1801038. https://doi.org/10.1183/13993003.01038-2018.

Lecorche, E., Daniau, C., La, K., Mougari, F., Benmansour, H., Kumanski, S., Robert, J., Fournier, S., Lebreton, G., Carbonne, A., Cambau, E., CNR-MyRMA, Alauzet, C., Cavalié, L., Chefson-Girault, C., Belmonte, O., Bounhiol Colin, C., Bourgoin, A., Bousseau, A., Verdet, C., 2021. *Mycobacterium chimaera* genomics with regard to epidemiological and clinical investigations conducted for an open chest postsurgical *Mycobacterium chimaera* infection outbreak. Open Forum Infect. Dis. 8 (6), ofab192. https://doi.org/10.1093/ofid/ofab192.

Llarena, A., Ribeiro-Gonçalves, B.F., Nuno Silva, D., Halkilahti, J., Machado, M.P., Da Silva, M.S., Jaakkonen, A., Isidro, J., Hämäläinen, C., Joenperä, J., Borges, V., Viera, L., Gomes, J.P., Correia, C., Lunden, J., Laukkanen-Ninios, R., Fredriksson-Ahomaa, M., Bikandi, J., Millan, R.S., Rossi, M., 2018. INNUENDO: A cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens. EFSA Supp. Publ. 15 (11). https://doi.org/10.2903/sp.efsa.2018.EN-1498.

M24-A2: Susceptibility Testing of Mycobacteria, Nocardiae, and Other Aerobic Actinomycetes, 2024. Approved Standard, Second edition (n.d.).

Mijs, W., De Haas, P., Rossau, R., Van Der Laan, T., Rigouts, L., Portaels, F., Van Soolingen, D., 2002. Molecular evidence to support a proposal to reserve the designation *Mycobacterium avium* subsp. Avium for bird-type isolates and '*M. avium* subsp. Hominissuis' for the human/porcine type of *M. avium*. Int. J. Syst. Evol. Microbiol. 52 (5), 1505–1518. https://doi.org/10.1099/00207713-52-5-1505.

Mixão, V., Pinto, M., Sobral, D., Di Pasquale, A., Gomes, J.P., Borges, V., 2023. ReporTree: A surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data. Genome Med. 15 (1), 43. https://doi.org/10.1186/s13073-023-01196-1.

Mizzi, R., Plain, K.M., Whittington, R., Timms, V.J., 2022. Global phylogeny of *Mycobacterium avium* and identification of mutation hotspots during niche adaptation. Front. Microbiol. 13, 892333. https://doi.org/10.3389/fmicb.2022.892333.

Moon, S.M., Park, H.Y., Kim, S.-Y., Jhun, B.W., Lee, H., Jeon, K., Kim, D.H., Huh, H.J., Ki, C.-S., Lee, N.Y., Kim, H.K., Choi, Y.S., Kim, J., Lee, S.-H., Kim, C.K., Shin, S.J., Daley, C.L., Koh, W.-J., 2016. Clinical characteristics, treatment outcomes, and resistance mutations associated with macrolide-resistant *Mycobacterium avium* complex lung disease. Antimicrob. Agents Chemother. 60 (11), 6758–6765. https://doi.org/10.1128/AAC.01240-16.

Nishiuchi, Y., Iwamoto, T., Maruyama, F., 2017. Infection sources of a common non-tuberculous mycobacterial pathogen, *Mycobacterium avium* Complex. Front. Med. 4. https://doi.org/10.3389/fmed.2017.00027.

Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J.P., Woyke, T., Kyrpides, N.C., Pukall, R., Klenk, H.-P., Goodfellow, M., Göker, M., 2018. Genome-based taxonomic classification of the phylum Actinobacteria. Front. Microbiol. 9, 2007. https://doi.org/10.3389/fmicb.2018.02007.

Pan, S.-W., Shu, C.-C., Feng, J.-Y., Chien, J.-Y., Wang, J.-Y., Chan, Y.-J., Yu, C.-J., Su, W.-J., 2021. Impact of different subspecies on disease progression in initially untreated patients with *Mycobacterium avium* complex lung disease. Clin. Microbiol. Infect. 27 (3), 467.e9–467.e14. https://doi.org/10.1016/j.cmi.2020.04.020.

Park, H.-E., Kim, S., Shim, S., Park, H.-T., Park, W.B., Im, Y.B., Yoo, H.S., 2020. 16S and 23S rRNA gene mutation independent multidrug resistance of non-tuberculous mycobacteria isolated from south Korean soil. Microorganisms 8 (8), 1114. https://doi.org/10.3390/microorganisms8081114.

Prevots, D.R., Marras, T.K., 2015. Epidemiology of human pulmonary infection with non-tuberculous mycobacteria: A review. Clin. Chest Med. 36 (1), 13–34. https://doi.org/10.1016/j.ccm.2014.10.002.

Quintás Viqueira, A., Pérez Romero, C., Toro Rueda, C., Sánchez Calles, A.M., Blázquez González, J.A., Alejandre Leyva, M., 2021. *Mycobacterium chimaera* in heater–cooler devices: an experience in a tertiary hospital in Spain. New Microbes New Infect. 39, 100757. https://doi.org/10.1016/j.nmni.2020.100757.

Radomski, N., Thibault, V.C., Karoui, C., De Cruz, K., Cochard, T., Gutiérrez, C., Supply, P., Biet, F., Boschiroli, M.L., 2010. Determination of genotypic diversity of *Mycobacterium avium* subspecies from human and animal origins by mycobacterial interspersed repetitive-unit-variable-number tandem-repeat and IS *1311* restriction fragment length polymorphism typing methods. J. Clin. Microbiol. 48 (4), 1026–1034. https://doi.org/10.1128/JCM.01869-09.

Rindi, L., Garzelli, C., 2014. Genetic diversity and phylogeny of Mycobacterium avium. Infect. Genet. Evol. 21, 375–383. https://doi.org/10.1016/j.meegid.2013.12.007.

RIVM, 2024. Protocol, Isolation of high molecular weight genomic DNA from Mycobacteria (CTAB procedure). n.d. https://www.rivm.nl/sites/default/files/2018-11/Protocol%2C%20Isolation%20of%20high%20molecular%20weight%20genomic%20DNA%20%28CTAB%29%20a1a.pdf.

Rodriguez-Temporal, D., Alcaide, F., Mareković, I., O'Connor, J.A., Gorton, R., Van Ingen, J., Van Den Bossche, A., Héry-Arnaud, G., Beauruelle, C., Orth-Höller, D., Palacios-Gutiérrez, J.-J., Tudó, G., Bou, G., Ceyssens, P.-J., Garrigó, M., González-Martin, J., Greub, G., Hrabak, J., Ingebretsen, A., Rodríguez-Sánchez, B., 2022. Multicentre study on the reproducibility of MALDI-TOF MS for nontuberculous mycobacteria identification. Sci. Rep. 12 (1), 1237. https://doi.org/10.1038/s41598-022-05315-7.

Saifi, M., Jabbarzadeh, E., Bahrmand, A.R., Karimi, A., Pourazar, S., Fateh, A., Masoumi, M., Vahidi, E., 2013. HSP65-PRA identification of non-tuberculosis mycobacteria from 4892 samples suspicious for mycobacterial infections. Clin. Microbiol. Infect. 19 (8), 723–728. https://doi.org/10.1111/j.1469-0691.2012.04005.x.

Santos, A., Carneiro, S., Silva, A., Gomes, J.P., Macedo, R., 2024. Nontuberculous mycobacteria in Portugal: trends from the last decade. Pulmonology 30 (4), 337–343. https://doi.org/10.1016/j.pulmoe.2022.01.011.

Schreiber, P.W., Kohl, T.A., Kuster, S.P., Niemann, S., Sax, H., 2021. The global outbreak of *Mycobacterium chimaera* infections in cardiac surgery—A systematic review of whole-genome sequencing studies and joint analysis. Clin. Microbiol. Infect. 27 (11), 1613–1620. https://doi.org/10.1016/j.cmi.2021.07.017.

Shin, J.-I., Shin, S.J., Shin, M.-K., 2020. Differential genotyping of *Mycobacterium avium* complex and its implications in clinical and environmental epidemiology. Microorganisms 8 (1), 98. https://doi.org/10.3390/microorganisms8010098.

Telenti, A., Marchesi, F., Balz, M., Bally, F., Böttger, E.C., Bodmer, T., 1993. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J. Clin. Microbiol. 31 (2), 175–178. https://doi.org/10.1128/jcm.31.2.175-178.1993.

Thibault, V.C., Grayon, M., Boschiroli, M.L., Hubbans, C., Overduin, P., Stevenson, K., Gutierrez, M.C., Supply, P., Biet, F., 2007. New variable-number tandem-repeat markers for typing *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* strains: comparison with IS *900* and IS *1245* restriction fragment length polymorphism typing. J. Clin. Microbiol. 45 (8), 2404–2410. https://doi.org/10.1128/JCM.00476-07.

Thorel, M.-F., Krichevsky, M., Vincent Levy-Frebault, V., 1990. Numerical taxonomy of Mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* subsp. Avium subsp. Nov., *Mycobacterium avium* subsp. Paratuberculosis subsp. Nov., and *Mycobacterium avium* subsp. Silvaticum subsp. Nov. Int. J. Syst. Bacteriol. 40 (3), 254–260. https://doi.org/10.1099/00207713-40-3-254.

Tortoli, E., Meehan, C.J., Grottola, A., Fregni Serpini, G., Fabio, A., Trovato, A., Pecorari, M., Cirillo, D.M., 2019. Genome-based taxonomic revision detects a number of synonymous taxa in the genus Mycobacterium. Infect. Genet. Evol. 75, 103983. https://doi.org/10.1016/j.meegid.2019.103983.

Uchiya, K., Tomida, S., Nakagawa, T., Asahi, S., Nikai, T., Ogawa, K., 2017. Comparative genome analyses of *Mycobacterium avium* reveal genomic features of its subspecies and strains that cause progression of pulmonary disease. Sci. Rep. 7 (1), 39750. https://doi.org/10.1038/srep39750.

Van Ingen, J., Turenne, C.Y., Tortoli, E., Wallace Jr., R.J., Brown-Elliott, B.A., 2018. A definition of the *Mycobacterium avium* complex for taxonomical and clinical purposes, a review. Int. J. Syst. Evol. Microbiol. 68 (11), 3666–3677. https://doi.org/10.1099/ijsem.0.003026.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9 (11), e112963. https://doi.org/10.1371/journal.pone.0112963.

Wallace Jr., R.J., Zhang, Y., Brown-Elliott, B.A., Yakrus, M.A., Wilson, R.W., Mann, L., Couch, L., Girard, W.M., Griffith, D.E., 2002. Repeat positive cultures in *Mycobacterium intracellulare* lung disease after macrolide therapy represent new infections in patients with nodular bronchiectasis. J. Infect. Dis. 186 (2), 266–273. https://doi.org/10.1086/341207.

Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15 (3), R46. https://doi.org/10.1186/gb-2014-15-3-r46.

Woods, G.L., Brown-Elliott, B.A., Conville, P.S., Desmond, E.P., Hall, G.S., Lin, G., Pfyffer, G.E., Ridderhof, J.C., Siddiqi, S.H., Wallace, R.J., Warren, N.G., Witebsky, F. G., 2011. Susceptibility Testing of Mycobacteria, Nocardiae, and Other Aerobic Actinomycetes, 2nd ed. Clinical and Laboratory Standards Institute http://www.ncbi.nlm.nih.gov/books/NBK544374/.

Zhou, Z., Alikhan, N.-F., Sergeant, M.J., Luhmann, N., Vaz, C., Francisco, A.P., Carriço, J. A., Achtman, M., 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res. 28 (9), 1395–1404. https://doi.org/10.1101/gr.232397.117.

Zweijpfenning, S., Ingen, J., Hoefsloot, W., 2018. Geographic distribution of nontuberculous mycobacteria isolated from clinical specimens: A systematic review. Semin. Respir. Crit. Care Med. 39 (03), 336–342. https://doi.org/10.1055/s-0038-1660864.