AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Comparative Genomics of *Mycobacterium avium* Complex Reveals Signatures of Environment-Specific Adaptation and Community Acquisition

Eric C. Keen,[a,b] JooHee Choi,[a] Meghan A. Wallace,[b] Michelle Azar,[c] Carlos R. Mejia-Chew,[d] Shail B. Mehta,[d] Thomas C. Bailey,[d] Lindsay J. Caverly,[c] Carey-Ann D. Burnham,[b,d,e,f] Gautam Dantas[a,b,f,g]

[a]The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
[b]Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
[c]Department of Pediatrics, University of Michigan Medical School, Ann Arbor, Michigan, USA
[d]Department of Medicine, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
[e]Department of Pediatrics, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
[f]Department of Molecular Microbiology, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
[g]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri, USA

Carey-Ann D. Burnham and Gautam Dantas jointly supervised this work.

ABSTRACT Nontuberculous mycobacteria, including those in the *Mycobacterium avium* complex (MAC), constitute an increasingly urgent threat to global public health. Ubiquitous in soil and water worldwide, MAC members cause a diverse array of infections in humans and animals that are often multidrug resistant, intractable, and deadly. MAC lung disease is of particular concern and is now more prevalent than tuberculosis in many countries, including the United States. Although the clinical importance of these microorganisms continues to expand, our understanding of their genomic diversity is limited, hampering basic and translational studies alike. Here, we leveraged a unique collection of genomes to characterize MAC population structure, gene content, and within-host strain dynamics in unprecedented detail. We found that different MAC species encode distinct suites of biomedically relevant genes, including antibiotic resistance genes and virulence factors, which may influence their distinct clinical manifestations. We observed that *M. avium* isolates from different sources—human pulmonary infections, human disseminated infections, animals, and natural environments—are readily distinguished by their core and accessory genomes, by their patterns of horizontal gene transfer, and by numerous specific genes, including virulence factors. We identified highly similar MAC strains from distinct patients within and across two geographically distinct clinical cohorts, providing important insights into the reservoirs which seed community acquisition. We also discovered a novel MAC genomospecies in one of these cohorts. Collectively, our results provide key genomic context for these emerging pathogens and will facilitate future exploration of MAC ecology, evolution, and pathogenesis.

IMPORTANCE Members of the *Mycobacterium avium* complex (MAC), a group of mycobacteria encompassing *M. avium* and its closest relatives, are omnipresent in natural environments and emerging pathogens of humans and animals. MAC infections are difficult to treat, sometimes fatal, and increasingly common. Here, we used comparative genomics to illuminate key aspects of MAC biology. We found that different MAC species and *M. avium* isolates from different sources encode distinct suites of clinically relevant genes, including those for virulence and antibiotic resistance. We identified highly similar MAC strains in patients from different states and decades, suggesting community acquisition from dispersed and stable reservoirs, and we discovered a novel MAC

Citation Keen EC, Choi J, Wallace MA, Azar M, Mejia-Chew CR, Mehta SB, Bailey TC, Caverly LJ, Burnham C-AD, Dantas G. 2021. Comparative genomics of *Mycobacterium avium* complex reveals signatures of environment-specific adaptation and community acquisition. mSystems 6:e01194-21. https://doi.org/10.1128/mSystems.01194-21.

Editor Jack A. Gilbert, University of California San Diego

Address correspondence to Carey-Ann D. Burnham, cburnham@wustl.edu, or Gautam Dantas, dantas@wustl.edu.

Received 30 September 2021
Accepted 1 October 2021
Published 19 October 2021

species. Our work provides valuable insight into the genomic features underlying these versatile pathogens.

KEYWORDS comparative genomics, genomic epidemiology, *Mycobacterium*, *Mycobacterium avium* complex, nontuberculous mycobacteria, whole-genome sequencing

Nontuberculous mycobacteria (NTM), or mycobacterial species other than *Mycobacterium tuberculosis* complex species and *Mycobacterium leprae*, are ubiquitous in natural environments and important pathogens of humans and animals (1). Globally distributed in waters and soils, these diverse organisms cause equally diverse infections, from systemic mycobacteriosis in birds to gastrointestinal wasting in ruminants to soft tissue abscesses, chronic lymphadenitis, osteomyelitis, and disseminated disease in humans (2, 3). NTM are most notorious as pulmonary pathogens, and in patients with underlying lung conditions (e.g., cystic fibrosis or bronchiectasis), these infections are challenging to eradicate and sometimes fatal (4, 5). NTM infections are also increasing in frequency worldwide (6). For example, clinical NTM isolations in Ontario, Canada, nearly doubled from 1998 to 2010 (7), and in Taiwan, the incidence of NTM pulmonary disease increased more than 6-fold from 2000 to 2008 (8). Today, the prevalence of NTM disease is higher than that of tuberculosis in many countries, including the United States (9, 10). Thus, NTM represent an urgent and growing threat to global public health.

More than 160 NTM species have been described to date, but most NTM pulmonary disease is caused by *Mycobacterium avium* and nine closely related species, which collectively comprise the *M. avium* complex (MAC) (11). These 10 *Mycobacterium* species, *M. avium*, *M. intracellulare*, *M. colombiense*, *M. arosiense*, *M. vulneris*, *M. bouchedurhonense*, *M. timonense*, *M. marseillense*, *M. paraintracellulare*, and *M. lepraemurium*, are defined and differentiated genomically, and proposals for taxonomic rearrangement are ongoing (11, 12). However, despite the emergence of MAC and the utility of whole-genome sequencing (WGS) in characterizing clinically relevant pathogens, key aspects of MAC biology remain unresolved. Here, we leveraged comparative genomics and a unique cohort of isolates to explore three questions central to MAC pathogenesis, ecology, and evolution. First, we investigated whether the two most prominent MAC species, *M. avium* and *M. intracellulare* (13, 14), differ from each other and other complex members in their repertoires of biomedically significant genes, such as virulence factor (VF) genes and antibiotic resistance genes (ARGs). Second, we assessed whether MAC genomes display signatures of niche-specific adaptation, reflecting their versatility as free-living organisms and pathogens. Finally, we examined whether genomic epidemiology could provide new insights into MAC acquisition and transmission, as was recently described for another pathogenic NTM, *Mycobacteroides abscessus* complex (15).

To test these and related questions, we first profiled conserved genomic regions to establish population structures and phylogenetic relationships across MAC members and within *M. avium*. Next, we compared isolates' accessory genomes to identify specific genes, including VF genes and ARGs, associated with particular MAC species and environmental niches. We then inspected MAC genomes for signatures of horizontal gene transfer (HGT) to infer how mobile genetic elements (MGEs) have shaped MAC genotypes and phenotypes. Finally, we utilized two unpublished clinical cohorts, accumulated from two hospitals over nearly a decade, to quantify intra- and interpatient MAC dynamics with strain-level resolution. This work represents the most comprehensive genomic characterization of the MAC to date and provides numerous testable hypotheses for further investigation *in vitro* and *in vivo*.

RESULTS

Species distribution of the MAC cohort. Our MAC cohort contained 126 high-quality published genomes and an additional 44 newly sequenced genomes from two clinical cohorts: 27 genomes from 18 patients associated with Barnes-Jewish Hospital and Washington University School of Medicine in St. Louis (WUMAC isolates) and 17

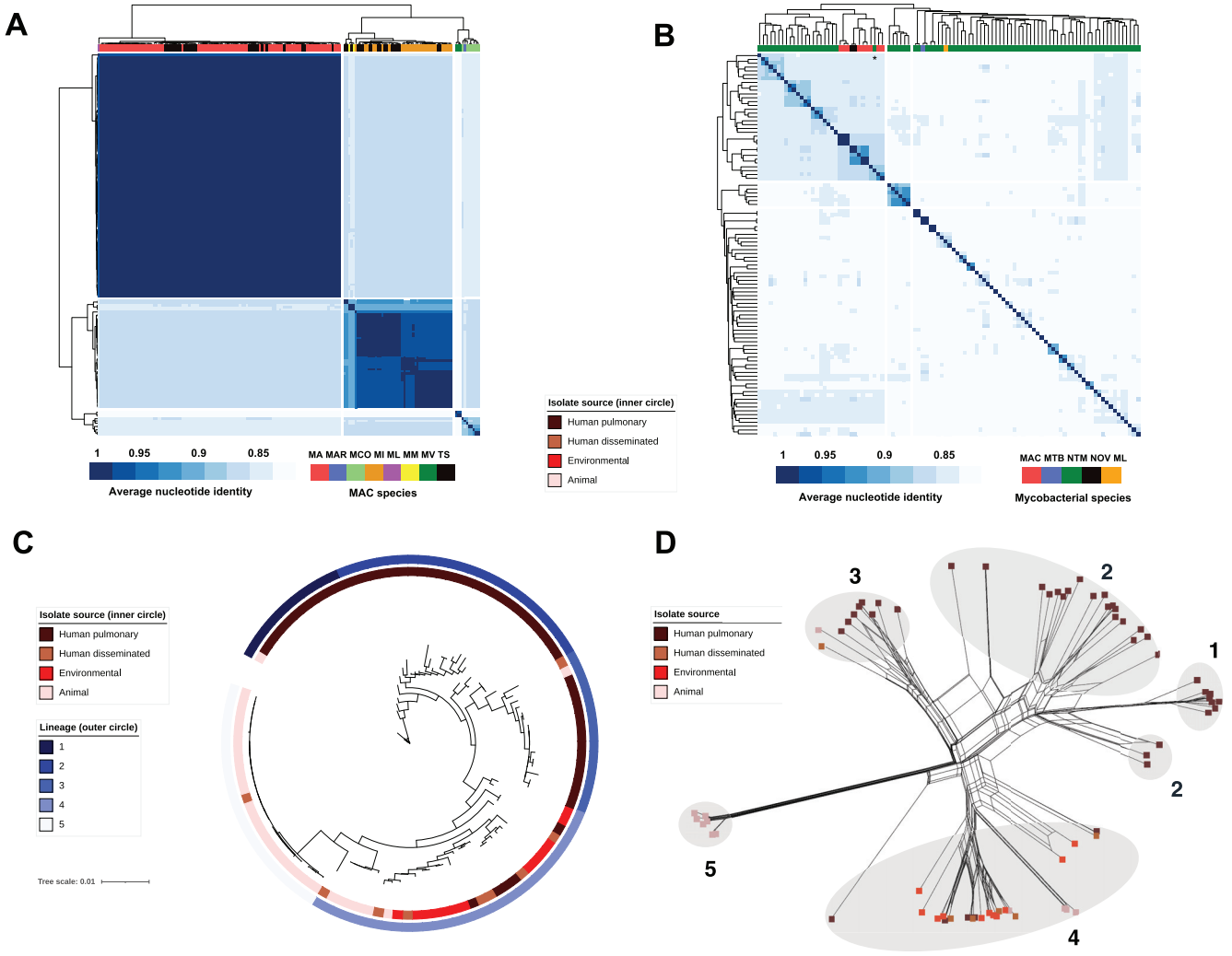

FIG 1 *M. avium* core genomes cluster by source of isolation. (A) Pairwise average nucleotide identity (ANI) matrix of all 170 MAC isolates in this study. MA, *M. avium*; MAR, *M. arosiense*; MCO, *M. colombiense*; MI, *M. intracellulare*; ML, *M. lepraemurium*; MM, *M. marseillense*; MV, *M. vulneris*; TS, novel isolates from WUMAC and FLAC cohorts described in this study. (B) ANI matrix of all 96 available mycobacterial representative/reference strains and 2 WUMAC isolates with <95% ANI to any other isolate in panel A. MAC, *M. avium* complex; NTM, all other nontuberculous mycobacteria; MTB, *M. tuberculosis*; ML, *M. leprae*, NOV, WUMAC isolates which represent a putative novel MAC genomospecies. An asterisk denotes *M. mantenii*. (C) Core genome phylogeny of 109 *M. avium* isolates as determined by Roary and RAxML. The scale bar represents the number of substitutions per site. (D) Core genome phylogenetic network of 109 *M. avium* isolates as determined by SplitsTree4. Branch lengths represent uncorrected *P* values. Numbers represent lineages as shown in panel C. Overlapping nodes were not annotated with the isolate source.

genomes from 14 patients associated with the University of Michigan Medical School (FLAC isolates) (see Materials and Methods and Tables S1 and S2 in the supplemental material). All WUMAC and FLAC isolates were obtained from clinically significant pulmonary MAC infections. We first calculated pairwise average nucleotide identity (ANI), the genomic gold standard for defining microbial species (16), across all 170 isolates. The resulting ANI distribution revealed a cohort dominated by *M. avium* (109 genomes; 64%) and *M. intracellulare* (44 genomes; 26%) and containing five additional MAC species: *M. vulneris*, *M. colombiense*, *M. marseillense*, *M. arosiense*, and *M. lepraemurium* (Fig. 1A). Of our 44 newly sequenced isolates, 25 were identified by ANI as *M. avium*, 16 as *M. intracellulare*, and 1 as *M. marseillense*, closely mirroring matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry (MS) predictions for these isolates (Table S2).

Identification of novel MAC genomospecies. Intriguingly, two additional isolates collected ~26 months apart from the same WUMAC patient, WUMAC-027 and WUMAC-065, showed <93.8% ANI to all other 168 MAC genomes. A threshold of 95%

ANI typically demarcates different species (16), implying that these isolates represent a distinct mycobacterial genomospecies. To confirm this finding, we downloaded representative or reference genomes for all available *Mycobacterium* and *Mycobacteroides* species (n = 96) and calculated pairwise ANIs for WUMAC-025 and WUMAC-067 (Fig. 1B). Despite the increased diversity of mycobacterial index genomes for comparison, the closest match for WUMAC-025 and WUMAC-067 was *M. intracellulare* ATCC 13950, with 92.9% ANI. We therefore suggest that isolates WUMAC-025 and WUMAC-067 are not closely related to a non-MAC species of mycobacteria but instead likely represent a novel genomospecies of the MAC. Our whole-genome ANI analysis, in contrast to previous comparisons of marker gene sequences (17), also supports the placement of *Mycobacterium mantenii* as an additional species within the MAC (Fig. 1B).

***M. avium* core genomes cluster by isolate source.** Unlike both *M. tuberculosis* complex and numerous other NTM species, *M. avium* is relatively unique in its ability to cause life-threatening pulmonary and disseminated infection, infect other mammals and birds, and thrive in water, soil, and other natural environments. Our *M. avium* cohort included isolates from human pulmonary and disseminated infections (56 and 9 isolates, respectively), animals (31 isolates), and soil (13 isolates). Given its prevalence within the MAC (13, 14) and its ability to occupy diverse niches, we selected *M. avium* for further genomic characterization. We annotated all 109 *M. avium* genomes and generated a core genome alignment from 3,425 core genes, which we then used to construct a maximum likelihood phylogenetic tree (Fig. 1C). The resulting phylogeny yielded five distinct lineages of isolates (Fig. S1A). As expected, lineages differed by subspecies, with all *M. avium* subsp. *paratuberculosis* (MAP) isolates forming their own discrete lineage (lineage 5). Intriguingly, we also observed strong clustering by *M. avium* isolate source (Fig. 1C; Fig. S1A). Lineages 1 to 3 comprised human pulmonary isolates almost exclusively (51 of 54 isolates), while lineage 5 was dominated by genomes of animal origin (23 of 24 isolates). In contrast, lineage 4 displayed substantial heterogeneity and included all 13 environmental isolates, 6 animal isolates, and 7 and 5 isolates from human disseminated and pulmonary infections, respectively. These human pulmonary isolates from lineage 4 (n = 5) bore numerous genes which were significantly less abundant in human pulmonary isolates from the more homogenous lineages 1 through 3 (n = 51) (Table S3). Lineage composition was significantly nonrandom ($P < 0.0001$; chi-square test), suggesting that in *M. avium*, genomic diversity accompanies habitat diversity and that specific genotypes are associated with specific niches.

Phylogenetic networks complement traditional phylogenetic trees by incorporating intertree variation and by better accounting for reticulate evolution, including horizontal gene transfer (HGT), hybridization, and duplication (18). Therefore, we conducted neighbor-net analysis on all 109 *M. avium* core genomes (Fig. 1D). Consistent with previous reports (19), we observed diminished genomic diversity within MAP relative to that of other *M. avium* subspecies. Critically, the resulting phylogenetic network closely resembled the genomic relationships depicted in Fig. 1C, suggesting a lack of significant ancestral recombination between isolate sources and lineages, consistent with largely independent evolutionary histories.

***M. avium* pangenomes cluster by isolate source and encode niche-specific genes.** We next considered the possibility that *M. avium* pangenomes, like core genomes, cluster by isolate source. Consistent with previous reports (20), we observed an open pangenome in *M. avium* (Fig. S1B) despite the addition of 25 newly sequenced genomes from this study. The *M. avium* accessory genome was considerably larger than the core genome (10,883 and 3,425 genes, respectively), implying that numerous gene functions may be differentially abundant across *M. avium* isolates and lineages. Indeed, upon unsupervised ordination of all isolates' pangenome contents, we again observed strong clustering by isolate source (Fig. 2A). When grouped by isolate source, *M. avium* pangenomes also displayed significantly greater within-source similarity (Jaccard index) than between-source similarity ($P < 0.01$; Kruskal-Wallis test with the Benjamini-Hochberg correction) (Fig. 2B). These results collectively demonstrate that in *M. avium*, both core genomes and pangenomes differ by environment of origin and,

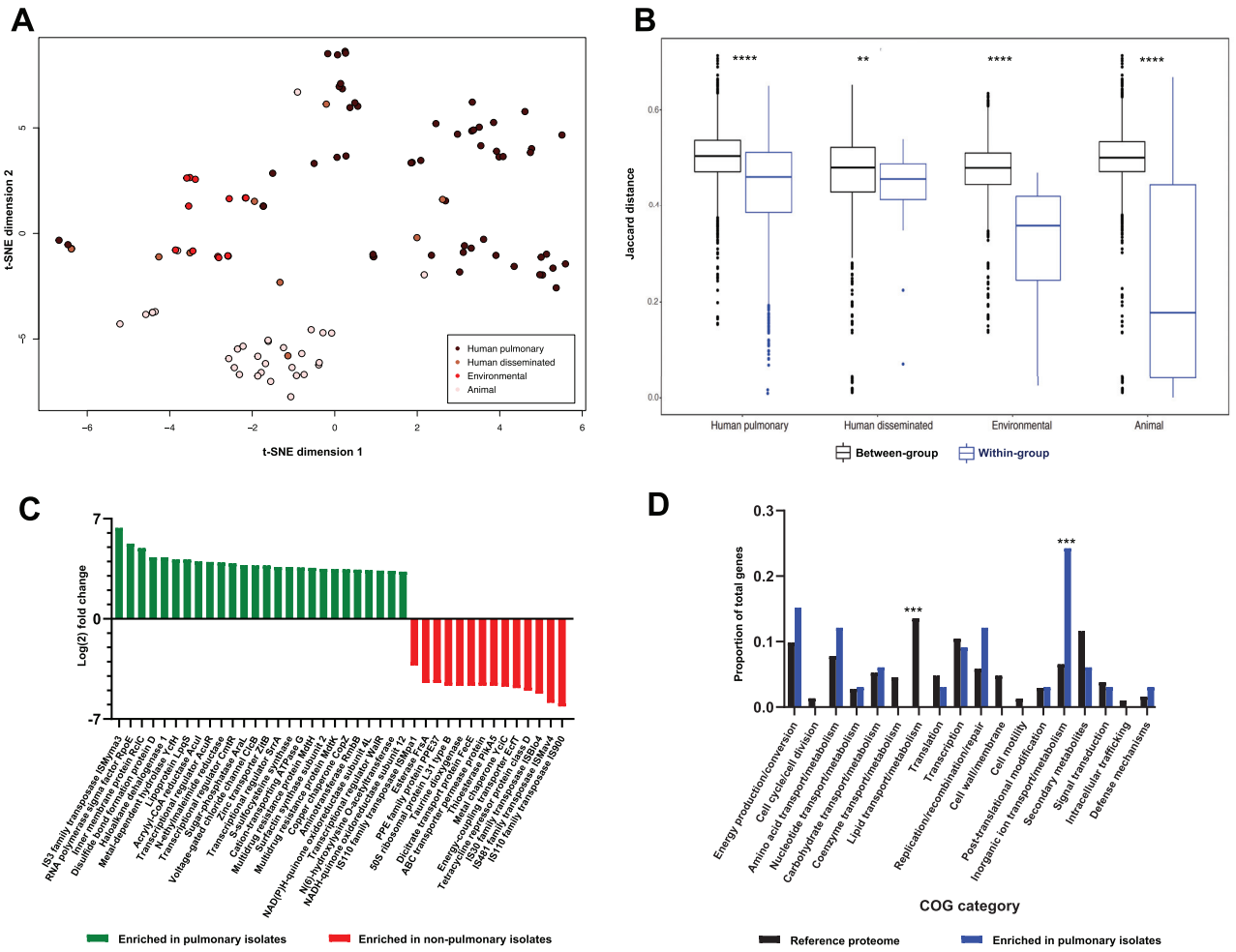

FIG 2 *M. avium* pangenomes cluster by source of isolation. (A) Pangenome ordination of 109 *M. avium* isolates by t-distributed stochastic neighbor embedding (t-SNE). (B) Between- and within-group Jaccard dissimilarity of *M. avium* isolate sources, comprising 109 total isolates. ****, $P < 0.0001$; **, $P < 0.01$ (Kruskal-Wallis test with the Benjamini-Hochberg correction). (C) Genes differentially abundant in pulmonary ($n = 56$; green) and nonpulmonary ($n = 53$; red) *M. avium* isolates as identified by Scoary. All genes with substantial ($\log_2$ fold change > 3) and significant ($P < 0.05$; Fisher's exact test with the Benjamini-Hochberg correction) enrichment are shown. (D) COG profiles of enriched genes in panel C. Categories of general or unidentified hits or those with no hits are not shown. The reference proteome represents the COG distribution of proteins in the representative *M. avium* strain OCU464. ***, $P < 0.001$ (Fisher's exact test with the Benjamini-Hochberg correction).

additionally, imply that *M. avium* bears accessory genes which facilitate niche-specific adaptation.

Given these associations between genomes and habitats, we hypothesized that the presence or absence of specific accessory genes might distinguish human pulmonary isolates from human disseminated, animal, and environmental isolates. We therefore quantitatively profiled the pangenomic contents of our 56 pulmonary and 53 nonpulmonary *M. avium* genomes. We identified numerous genes which were significantly enriched in human pulmonary genomes relative to their representation in isolates from other sources ($P < 0.05$; Fisher's exact test with the Benjamini-Hochberg correction) (Fig. 2C; Table S4). These gene products included a lipoprotein associated with virulence in *M. tuberculosis* (LpqS) (21), regulators of virulence and immune evasion in other Gram-positive pathogens (WalR and SrrA) (22, 23), multidrug resistance proteins (MdtH and MdtK) (24), components of metal transport machinery (CopZ and ZitB) (25, 26), and a membrane protein associated with resistance to reactive chlorine species (RclC) (27). Intriguingly, three separate subunits of bacterial NAD(P)H-quinone oxidoreductase, an enzyme thought to assist the intracellular survival of *M. tuberculosis* via

detoxification of host-derived metabolites and/or energy generation under hypoxic conditions (28), were also significantly enriched in human pulmonary isolates. These and other genes (Table S4) represent plausible candidates for *in vitro* and *in vivo* characterization in the context of *M. avium* pulmonary pathogenesis.

To gain further insight into gene functions enriched in pulmonary isolates, we binned differentially abundant genes into clusters of orthologous protein groups (COGs). Although this enriched gene set and the proteome of the representative *M. avium* strain OCU464 had generally similar COG distributions, proteins putatively involved in lipid transport and metabolism (COG category I) were entirely absent among differentially abundant hits despite being well represented in the *M. avium* OCU464 proteome (13.5% of annotations) (Fig. 2D). Conversely, annotations for inorganic ion transport and metabolism (COG category P) were significantly overrepresented in Scoary hits (22.9% of annotations) relative to those of the *M. avium* OCU464 proteome as a whole (5.2%) ($P < 0.001$; Fisher's exact test with the Benjamini-Hochberg correction) (Fig. 2D). These results indicate a nonrandom pattern of gene enrichment and are consistent with numerous reports that metal acquisition is essential for intracellular persistence and virulence in mycobacterial lung infections (29, 30).

Horizontal gene transfer varies by MAC species and *M. avium* isolate source. Given the presence of several transposases as top differentially abundant genes (Fig. 2C) in our data set, as well as the variable importance of HGT in the evolution of various mycobacteria (31–34), we next profiled HGT dynamics in MAC species. We observed clear differences in HGT burden across MAC species, with *M. intracellulare* containing significantly more foreign DNA than *M. avium* and *M. colombiense* ($P < 0.0001$; one-way analysis of variance [ANOVA] with Tukey's *post hoc* correction) (Fig. 3A). In human and environmental *M. avium* isolates, foreign DNA represented a fraction (4 to 5%) of the genome comparable to the 4.5% reported for *M. tuberculosis* (35, 36). Surprisingly, however, animal isolates encoded significantly less foreign DNA than all other *M. avium* isolate sources ($P < 0.05$; one-way ANOVA with Tukey's *post hoc* correction) (Fig. 3A). This reduced evidence of HGT in animal isolates, particularly but not exclusively in *M. avium* subsp. *paratuberculosis*, is consistent with the tight clustering of lineage 5 in Fig. 1D and with previous reports (19), indicating potential differences in the ecologies and/or genome biologies of animal isolates.

We next characterized these regions of foreign DNA on the basis of their mobility genes. The vast majority of foreign regions were associated with putative transposases, recombinases, or bacteriophage-derived integrases. Although the genomic abundance of these mobilization elements did not vary significantly by isolate source ($P > 0.05$; one-way ANOVA with Tukey's *post hoc* correction) (Fig. 3B), we found that different elements were enriched for distinct genes (Fig. 3C). For example, of the 10 nonhypothetical proteins most commonly associated with transposases, the most abundant MGE class in our genomes, only 3 were among the top 10 associated with recombinases or integrases. We observed that PPE proteins, an abundant and enigmatic mycobacterial protein family linked to immune evasion (37) and nutrient transport (38), were commonly colocalized with transposases but rarely with integrases or recombinases (Fig. 3C). PPE genes have previously been associated with insertion sequences in *M. tuberculosis* (39) but never, to our knowledge, in *M. avium*. Additionally, a cluster of metal-related genes, including the arsenate reductase gene *arsC* and cadmium-associated protein gene *cadI*, were strongly associated with integrases, consistent with the previous discovery of *arsC* in diverse temperate phage genomes (40). These results suggest that different MGEs encode overlapping but largely distinct gene repertoires in *M. avium*.

We further speculated that MGE-borne genes, like genomic content as a whole (Fig. 2), might also vary by source of isolation in *M. avium*. We therefore binned putative foreign regions by isolate source and compared gene annotations across bins. As predicted, we observed different MGE-encoded gene profiles in different isolate sources (Fig. 3D), albeit with greater overlap across groups than when binned by MGE type

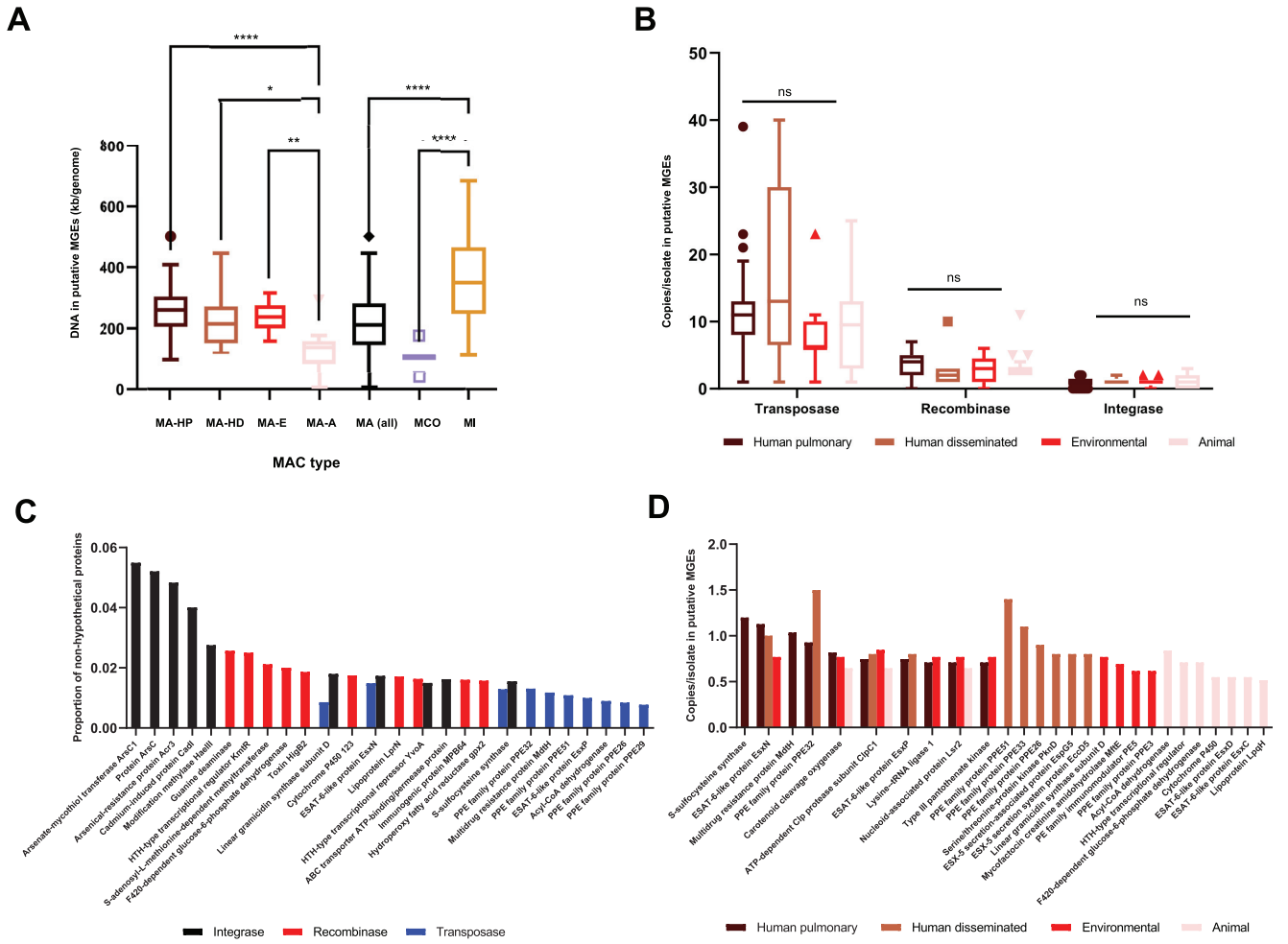

FIG 3 Foreign DNA signatures vary by MAC species and *M. avium* isolate source. (A) Total numbers of kilobases of putative foreign DNA in MAC genomes as identified by IslandPath-DIMOB. MA-HP, human pulmonary *M. avium* isolates; MA-HD, human disseminated *M. avium* isolates; MA-E, environmental *M. avium* isolates; MA-A, animal *M. avium* isolates; MI, *M. intracellulare*; MCO, *M. colombiense*. ****, $P < 0.0001$; **, $P < 0.01$; *, $P < 0.05$ (one-way ANOVA with Tukey's *post hoc* correction). (B) Total numbers of copies of mobility elements within foreign DNA in *M. avium*, binned by isolate source. ns, not significant (one-way ANOVA with Tukey's *post hoc* correction for each mobility element). (C) Top 10 most abundant genes within each mobility element in foreign DNA in *M. avium*. Mobility and hypothetical proteins are not shown. (D) Top 10 most abundant genes within foreign DNA for each *M. avium* isolate source. Mobility and hypothetical proteins are not shown.

(Fig. 3C). Consistent with the paradigm of genome islands being VF hot spots, including in *M. tuberculosis* (35), we identified several canonical mycobacterial VF genes within these regions, including PPE and ESAT-6-like proteins (41). Several other top hits, including *S*-sulfocysteine synthase (*cysK2*) and carotenoid cleavage oxygenase (*Rv0654*), have putative but largely unexplored roles in mycobacterial infection (42, 43), and our findings support their further investigation. More broadly, our results suggest that HGT profiles vary by species, habitat, and MGE type across the MAC, with potentially wide-ranging implications for phenotypic variation and pathogenesis.

Virulence factors vary by MAC species and *M. avium* isolate source. Like other bacterial pathogens, NTM encode a variety of VFs which collectively enable immune evasion and host antagonism (41, 44). Given the genomic distinctions described in Fig. 1 and 3, we hypothesized that virulence gene profiles would also differ between various MAC species and isolate sources. We therefore generated a custom BLAST database of 9,114 protein sequences from three virulence factor databases (45–47) and queried proteomes from all 170 MAC isolates against this database. With thresholds of 80% sequence identity and coverage, we identified 244 VF homologs present in at least one MAC isolate. Of these, 47 were significantly differentially abundant between

M. avium and all other MAC species ($P < 0.05$; Fisher's exact test with the Benjamini-Hochberg correction) (Table S5), and we observed clear clustering both by MAC species and by *M. avium* isolate source (Fig. S2). Within *M. avium*, 13 VFs were significantly differentially abundant between human pulmonary and nonpulmonary isolates ($P < 0.05$; Fisher's exact test with the Benjamini-Hochberg correction) (Fig. S3). Consistent with the results shown in Fig. 2D, several of these proteins, including the siderophore transporter IrtA (48) and the siderophore acyltransferase MbtK (49), are involved in metal acquisition and transport in mycobacteria. Intriguingly, a protein uniquely present in nonpulmonary *M. avium* isolates, the heparin-binding hemagglutinin HbhA, is also required for extrapulmonary dissemination in *M. tuberculosis* (50). In a mouse model of tuberculosis, *hbhA* mutants were severely deficient (200-fold) in spleen colonization, but not in lung colonization, following intranasal inoculation (50). Our results support a similar putative function for HbhA in *M. avium* and, more broadly, illuminate additional genes which potentially underlie different forms of infection by this versatile pathogen.

***M. vulneris* encodes unique antibiotic resistance genes.** Both pulmonary and disseminated MAC infections are notoriously refractory to antibiotic treatment (5, 6). Along with having innate resistance to entire classes of antibiotics, NTM can acquire antibiotic resistance via mutation and, less commonly, through HGT (51–53). These factors usually necessitate the adoption of lengthy multidrug treatment regimens involving some combination of macrolides, aminoglycosides, rifampin, and ethambutol (54, 55). We therefore used two complementary approaches to profile antibiotic resistance in MAC genomes. First, we manually inspected each 16S (*rrs*) and 23S (*rrl*) rRNA gene sequence for point mutations in positions 1406 to 1408 and 2058 to 2059, which confer resistance to aminoglycosides and macrolides, respectively (56, 57). We found these mutations to be relatively uncommon overall (in 5 and 7 of 170 total isolates, respectively), most common in *M. intracellulare* (in 3 and 4 of 44 isolates, respectively), and present only in human pulmonary isolates (Fig. S4). Next, we queried the MEGARes 2.0 resistance gene database (58) to identify ARGs more comprehensively. Nearly all MAC isolates encoded the same three intrinsic ARGs: the genes for an RNA polymerase binding protein associated with rifampin tolerance (*rbpA*) (59), an efflux pump (*efpA*) (60), and an efflux regulator (*mtrA*) (61) (Fig. S4). We also detected several additional ARGs, including the class A β-lactamase *blaF*, in all three isolates of *M. vulneris*, a recently described MAC species (62). To our knowledge, this is the first report of *blaF* in a MAC genome (63) and the first comprehensive characterization of ARGs across all MAC species.

Long-term and geographically dispersed reservoirs seed community acquisition of MAC organisms. An important aspect of this work, and one which distinguishes it from other recent MAC comparative genomics studies (20, 64, 65), is our collection of longitudinal isolates from individual patients from multiple cohorts (Table S2). Our WUMAC and FLAC cohorts collectively contained pairs of pulmonary isolates from five patients with *M. avium* and six patients with *M. intracellulare*, separated by an average of 339 $\pm$ 244 days, which we leveraged to investigate within-host strain dynamics. We first sought to characterize the diversity of these paired isolates within their broader genomic context by generating core genome phylogenies of all *M. avium* pulmonary isolates ($n = 56$; $n = 25$ from this study) and all *M. intracellulare* isolates ($n = 44$; $n = 16$ from this study). Our WUMAC/FLAC isolates spanned the genomic breadth of both species, but while all *M. intracellulare* isolate pairs aligned together (Fig. 4A), paired isolates from WUMAC *M. avium* patients 1 and 3 were surprisingly diverse (Fig. 4B).

To quantify these dynamics with greater genomic resolution, we calculated pairwise single-nucleotide polymorphism (SNP) distances for all intrapatient *M. avium* and *M. intracellulare* isolate pairs in our WUMAC and FLAC cohorts. We observed a substantial gradient of SNP distances, with six highly similar isolate pairs (two *M. avium* and four *M. intracellulare* isolate pairs; SNP distance, $\leq$20) and five pairs (three *M. avium* and two *M. intracellulare* isolate pairs) comprising more distantly related strains (SNP distance, $>$400) (Fig. 4C). Because we did not sequence metagenomic sweeps or multiple

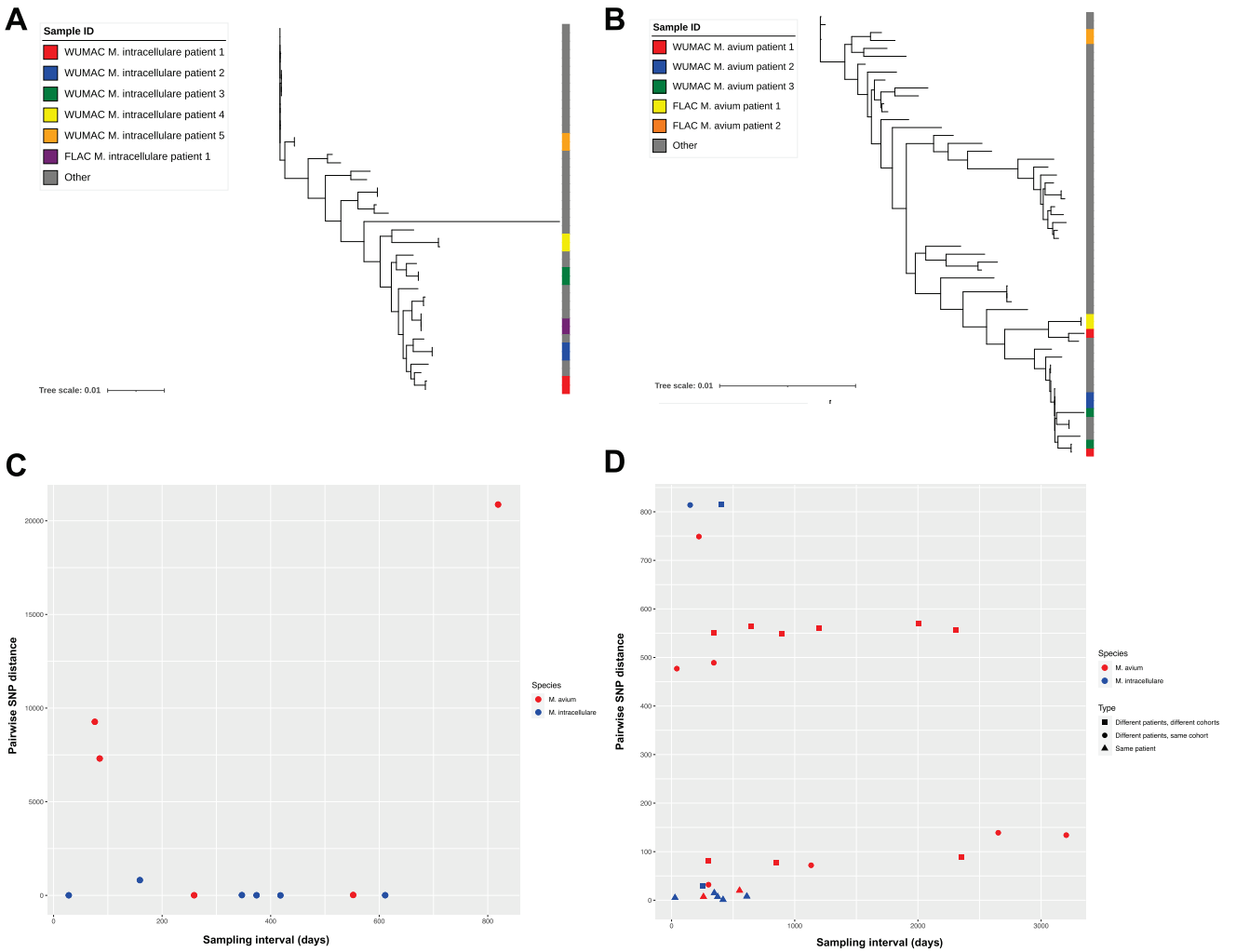

FIG 4 Similar MAC strains in distinct patients and cohorts. (A) Core genome phylogeny of all WUMAC and FLAC *M. intracellulare* isolates (n = 16) and all other *M. intracellulare* isolates (n = 28). Unpaired isolates are from patients for whom only a single isolate was available. The scale bar represents the number of substitutions per site. (B) Core genome phylogeny of all WUMAC and FLAC *M. avium* isolates (n = 25) and all other *M. avium* human pulmonary isolates (n = 31). Unpaired isolates are from patients for whom only a single isolate was available. The scale bar represents the number of substitutions per site. (C) Pairwise single-nucleotide polymorphism (SNP) distances for all intrapatient isolate pairs from WUMAC/FLAC *M. avium* (n = 5) and *M. intracellulare* (n = 6) patients. The sampling interval represents the number of days between isolate collections. (D) All intrapatient and interpatient pairwise core genome SNP distances that were <1,000 for all possible WUMAC/FLAC *M. avium* and *M. intracellulare* isolate pairs. Pairwise SNP distances of >1,000 are not shown. The sampling interval represents the number of days between isolate collections.

colonies per sample, we cannot exclude the possibility of monoclonal infection followed, in some patients, by strain replacement. However, these results are more parsimonious with long-standing evidence of polyclonal and even multispecies mycobacterial lung disease in NTM patients (66–69).

Next, we examined pairwise core genome SNP distances for all *M. avium* and *M. intracellulare* isolates, regardless of patient, from our WUMAC and FLAC cohorts. Consistent with the results in Fig. 4A and B, we observed clear discrepancies by species: while 5 of the 8 most related *M. intracellulare* pairwise comparisons were from the same patients, 17 of the 19 closest *M. avium* pairings were from different patients (Fig. 4D). Importantly, many of these closely related interpatient isolates were also obtained across considerable time and space. For example, *M. avium* isolates WUMAC-026 and WUMAC-062 differed by <75 SNPs but were collected from discrete patients in January 2013 and March 2016, respectively. *M. avium* isolates WUMAC-20 and WUMAC-35 (SNP distance of 134) were collected from a 43-year-old patient living with cystic fibrosis and a 71-year-old patient living with rheumatoid arthritis, respectively,

more than 8 years apart. Perhaps most remarkably, *M. intracellulare* isolates WUMAC-033 and FLAC0181 differed by just 24 SNPs despite being collected in different years from different hospitals more than 400 miles apart. Indeed, of the 26 isolate combinations with pairwise core SNP distances of <1,000, just 7 (27%) represented multiple isolates from the same patient (Fig. 4D).

Finally, we considered the possibility that despite possessing congruent core genomes (Fig. 4), some pairs of interpatient and intercohort isolates might encode divergent accessory genomes, thereby complicating assumptions of genomic relatedness. We therefore aligned reads from all WUMAC/FLAC *M. avium* isolates (n = 25) and *M. intracellulare* isolates (n = 16) to their respective representative NCBI strains (*M. avium* OCU464 and *M. intracellulare* ATCC 13950, respectively) and generated whole-genome SNP alignments. Critically, the resulting SNP distance distributions (Fig. S5) closely mirrored those from core genome alignments (Fig. 4D), further suggesting that comparable MAC isolates can be separated by considerable space and time. The striking similarity of isolates from such disparate sources raises intriguing questions about MAC reservoirs and acquisition in these cohorts and beyond, which we discuss below.

DISCUSSION

The *M. avium* complex is an increasing threat to public health. The vast majority of NTM lung disease in the United States is caused by MAC strains, with annual prevalence nearly doubling between 2008 and 2015 (70). MAC infections are often chronic, multidrug resistant, and fatal, with an all-cause 5-year mortality of ~25% for infected patients (4, 6). Several MAC subspecies, most notably MAP, are also prominent animal pathogens and cost hundreds of millions of dollars annually in culls and lost productivity (71, 72). Despite this obvious significance, however, MAC genomics remains understudied. Fewer than 300 MAC genomes, many of which are of low quality (high sequence heterogeneity or contamination and/or low completeness), have been deposited in GenBank, compared to >1,800 for the *M. abscessus* complex and >6,500 for the *M. tuberculosis* complex. In this work, we contribute an additional 44 high-quality MAC genomes, including two from a putative novel MAC genomospecies, and provide important insight into the comparative genomics of these emerging pathogens.

Despite substantial genotypic and phenotypic similarity, different MAC species vary in important ways. For instance, the complex's two most medically significant species, *M. avium* and *M. intracellulare*, differ both by geographic distribution (55) and by clinical outcomes (13, 14). Here, we extend and contextualize these observations by identifying differences in three key elements of MAC genomes, namely, VF genes, ARGs, and regions of foreign DNA (Fig. S2 and S4; Fig. 3). These differences were most pronounced for *M. vulneris*, a recently discovered and heretofore genomically undescribed MAC species (62), which harbored more ARGs but fewer virulence genes than its relatives. We note that 3 of the 10 MAC species represented in the NCBI, *M. bouchedurhonense*, *M. timonense*, and *M. paraintracellulare*, were not included in this study because no high-quality genomes were available, while another 3 species (*M. arosiense*, *M. marseillense*, and *M. lepraemurium*) were represented by two or fewer genomes. Our discovery of an additional novel MAC genomospecies (Fig. 1B) further illustrates that many more isolates, from many more species, will be required to fully capture the genomic intricacies of the microorganisms in this complex.

We focused deeper comparative analyses on *M. avium* because of its clinical importance, its preponderance within the MAC, and its diversity of occupied habitats. A central finding of this work is that *M. avium* isolates from different sources—human pulmonary and disseminated infections, animals, and free-standing environments—are genomically distinct by multiple measures of comparison. These parameters include core genomes (Fig. 1C and D; Fig. S1B), accessory genomes (Fig. 2A to B), and gene profiles (Fig. 2C and D), including for VF genes (Fig. S3) and mobile genetic elements (Fig. 3). The observed patterns cannot be explained simply by the presence of different *M. avium* subspecies in different environments, since pulmonary and nonpulmonary

isolates of *M. avium* subsp. *hominissuis* were well represented in our cohort. Although *M. avium* genomes have previously been shown to differ by source, most prior work has involved variable number tandem repeat fingerprinting (73–77), which provides limited genomic resolution and minimal biological insight compared to whole-genome sequencing. Of the few studies to employ WGS (20, 64, 65), ours is the largest to date and the only one to capture the diversity of niches that *M. avium* may occupy. By leveraging WGS, we also identified numerous specific genes, many of which are potentially clinically significant, which differentiate human pulmonary isolates from others (Fig. 2C; Table S4). While these findings would be strengthened by a broader collection of genomes (most human isolates were from North America and Asia, while environmental isolates were predominantly European), this work provides novel insights into *M. avium* biology and represents a rich substrate for future investigation *in vitro* and *in vivo*. For instance, our study implicates multiple individual genes, including *hbhA*, *lpqS*, and *cysK2*, as being particularly important for pulmonary MAC infection, a hypothesis which could be verified via gene knockouts and experimental infections of cells or animals. More broadly, comparative dual RNA sequencing (RNA-seq) profiling (i.e., host-pathogen transcriptomics in parallel) (78) upon infection by representative clinical and environmental isolates could shed valuable light on the mechanisms which underlie and restrict MAC disease.

Along with extensively analyzing publicly available genomes, our study incorporated longitudinal samples from two unpublished clinical cohorts, which we used to compare intra- and interpatient MAC dynamics. A limitation of this work is that we were restricted by upstream clinical collection procedures to a single colony per patient sample. Since NTM patients can be coinfected by multiple mycobacterial strains or even multiple species (66–69), such sampling is insufficient to accurately recapitulate microbial diversity. Thus, our finding of higher intrapatient diversity for *M. avium* than for *M. intracellulare* (Fig. 4C) should be interpreted with considerable caution, while the presence of highly related isolates in highly dissimilar patients (Fig. 4D) may be the rule rather than the exception. As was recently described (69), future studies may employ deep metagenomic sequencing and analysis of metagenome-assembled genomes, rather than selecting individual colonies, in order to fully capture mycobacterial diversity in samples of interest. These data sets will be instrumental in addressing key outstanding questions, including whether MAC mutation rates differ by species and host features (79), which genes are most frequently mutated during chronic infection, and how horizontal gene transfer shapes features of polymicrobial lung communities.

The striking similarity that we observed between clinical isolates from unrelated WUMAC/FLAC patients is particularly intriguing in light of MAC transmission dynamics. Unlike *M. tuberculosis* infections, MAC infections are typically thought to arise from environmental exposure, not human-to-human transmission (1–3). Multiple fingerprinting studies have demonstrated high genomic similarity between paired clinical and environmental isolates from the same household (80–83), but until recently, this paradigm had not been supported by WGS data (84). In pioneering work, Lande et al. described matched respiratory and built environment isolates differing by $<$100 SNPs and implicated municipal water supplies as likely *M. avium* reservoirs (84). Since this study encompassed a very narrow patient cohort (adult women living $<$20 miles apart in Pennsylvania, USA), however, its generalizability was unknown. By demonstrating genomic similarities between isolates collected up to 8 years apart from patients without direct contact (Fig. 4D), our data provide important support for this model and imply long-term colonization of community-accessible reservoirs by multiple *M. avium* strains. Moreover, since some of our most closely related isolates were collected from separate medical centers in different states, our results also suggest that highly similar MAC strains are not only temporally stable within reservoirs but also geographically dispersed across reservoirs. Our work thus provides valuable insights into MAC

acquisition dynamics and may facilitate interventions to slow the emergence of these versatile pathogens.

MATERIALS AND METHODS

Bacterial culturing and MALDI-TOF mass spectrometry. WUMAC isolates were recovered from clinical specimens submitted to the Barnes-Jewish Hospital microbiology laboratory as part of routine clinical care between 2006 and 2019. All isolates were recovered from respiratory specimens, including sputum, tracheal aspirates, bronchial wash fluids, and bronchial alveolar lavage fluids, and were stored at −80°C prior to analysis. Isolates were cultured from freezer stocks onto Middlebrook 7H11 agar (Hardy Diagnostics, Santa Maria, CA) and incubated at 35°C in air. Following sufficient growth, the identity of each isolate was confirmed using Vitek MALDI-TOF MS with Knowledge Base 3.0 (bioMérieux, Durham, NC, USA). Briefly, each isolate was picked using a sterile loop, spotted onto a target, and overlaid with 1 μl of formic acid. Once dry, 1 μl of the matrix was overlaid and allowed to dry. Isolates that could not be identified by Vitek MS were analyzed using the MALDI Biotyper with the MBT Mycobacteria module (Bruker, Billerica, MA, USA) using the same spotting procedure. Outgrowths positively identified as MAC (*M. avium*, *M. avium-M. intracellulare*, or *M. chimaera-M. intracellulare*) by either method were collected by swabbing the growth on agar plates and eluting it into molecular-grade water for storage at −80°C prior to extraction. In total, 27 WUMAC isolates were obtained from clinical samples from 18 patients.

FLAC isolates were obtained from clinical specimens associated with the University of Michigan Medical School as previously described (85). In total, 17 FLAC isolates were obtained from clinical samples from 14 patients.

DNA extraction, sequencing, and assembly. WUMAC isolate suspensions were thawed and mechanically lysed for 2 min (Mini-Beadbeater-24; BioSpec, Bartlesville, OK, USA), followed by DNA extraction (QIAamp BiOstic bacteremia DNA kit; Qiagen, Germantown, MD, USA) and quantification (Qubit HS; Thermo Fisher Scientific; Waltham, MA, USA) according to manufacturers' protocols. Genomic DNA (0.5 ng) was used as the input in the preparation of sequencing libraries (Nextera XT kit, Illumina, San Diego, CA, USA) as previously described (86). Libraries were pooled and sequenced to a depth of ~2.5 million paired-end reads (2× 150 bp) on a NextSeq500 high-output platform (Illumina, San Diego, CA, USA). Adapters were removed from demultiplexed reads with Trimmomatic 0.38 (leading, 10; trailing, 10; sliding window, 4:15; minimum length, 60) (87). Paired ends were fixed with custom Python scripts. Trimmed reads were assembled into genomes with Unicycler 0.4.7 with default parameters (88).

FLAC genomes were sequenced and assembled as previously described (89). All WUMAC and FLAC assemblies were queried for contamination and completeness with CheckM 1.0.7 (90), and all assemblies with >95% completeness, <5% strain heterogeneity, and <2% contamination were retained.

Retrieval of publicly available genomes. To assemble a comprehensive MAC cohort for comparative genomic analysis, all 278 publicly available genomes for *M. avium*, *M. intracellulare* (including those formerly classified as *M. chimaera* and *M. yongonense* [12]), *M. colombiense*, *M. arosiense*, *M. vulneris*, *M. bouchedurhonense*, *M. timonense*, *M. marseillense*, *M. paraintracellulare*, and *M. lepraemurium* were retrieved via the GenBank FTP portal. Genomes excluded from RefSeq (e.g., due to "fragmented assembly" or "many frameshifted proteins") were discarded. All assemblies were queried for contamination and completeness with CheckM (91), and those with <95% completeness, >5% strain heterogeneity, or >2% contamination were discarded. A total of 126 high-quality genomes were retained (see Table S1 in the supplemental material). To assemble a database of mycobacterial index strains, all 96 representative and reference strains from the genera *Mycobacterium* and *Mycobacteroides* were similarly retrieved via the GenBank FTP portal. No genomes from this data set were discarded. Existing GenBank nomenclature for MAC species and *M. avium* substrains was retained.

Species and strain profiling of the MAC cohort. Pairwise average nucleotide identity (ANI) was calculated for all MAC isolates and between WUMAC-027/WUMAC-065 and all mycobacterial representative/reference strains with pyani 0.2.7 (MUMmer mode; ANIm) (92). MAC genome assemblies were annotated with Prokka 1.13.7 (minimum contig length, 500) (93). ANI matrices were visualized with the R package pheatmap. Using annotated proteomes (Prokka .gff files) as the input, core genome alignments and gene presence/absence matrices were generated with Roary 3.12.0 (PRANK alignment; no splitting of paralogs; core gene threshold, 99%) (94). From this core genome alignment, genome lineages were identified with fastGEAR (95), and a phylogenetic tree was built with RAxML 8.2.11 (GTRGAMMA model, 100 rapid bootstrap searches, 20 maximum likelihood searches) (96). Phylogenetic trees were visualized and annotated with metadata in iTOL v4 (89). Using the same core genome alignment as the input, a phylogenetic network was constructed with SplitsTree4 (NeighborNet mode) (18). Ordination of isolates by pangenome content was conducted with the R package Rtsne (perplexity, 30), using a Roary gene/presence absence matrix as the input. Pangenome plots were generated in R with the supplemental Roary script create_pan_genome_plots. Jaccard distance was calculated from the Roary gene presence/absence matrix using the R package vegan.

Functional profiling of the MAC cohort. Genes differentially abundant in human pulmonary isolates, relative to those of MAC isolates from all other sources, were identified with Scoary 1.6.16 (97). To avoid spurious false positives, counts of paralogs (e.g., XerC_1, XerC_2, XerC_3, etc.) were collapsed into single annotations (e.g., tyrosine recombinase XerC) prior to calculation of enrichment. Scoary hits were also validated via sequence similarity. A BLAST database was built from protein sequences of genes enriched in pulmonary isolates, and nonpulmonary proteomes (Prokka .faa files) were queried against this database, which yielded only expected matches. Top Scoary hits were binned into clusters of orthologous groups (COGs) (98) with eggNOG mapper v2 (91, 99) and compared to the annotated proteome of the representative *M.*

avium strain OCU464. General or unidentified hits (COG categories R and S) were discarded. Regions of foreign DNA within MAC genomes were identified with IslandPath-DIMOB (100) within the IslandViewer4 suite (101), using each species' representative NCBI strain (*M. avium* OCU464, *M. intracellulare* ATCC 13950, or *M. colombiense* CECT3035) as the reference for contig alignment. Predicted regions of foreign DNA were parsed to compare types of mobility regions (counts of annotated transposases, integrases, and recombinases in these regions), the total burdens of foreign DNA, and the distributions of specific genes across MAC species, as determined by Prokka (93). Virulence factors were identified by querying MAC isolates' proteomes against a custom VF database containing all representative VF sequences from PATRIC_VF, VFDB, and Victors (45–47). Briefly, VF protein sequences were downloaded from PATRIC (45), clustered with MMseqs212–113e3 (identity, 0.9; coverage, 0.8) (102), and converted into a BLAST database. Annotated proteomes (Prokka .faa files) were queried against this database, and for each MAC protein, the top hit was retained if it exceeded 80% coverage and identity. High-quality hits with a corresponding PATRIC annotation were retained and visualized with the R package pheatmap. Antibiotic resistance genes were identified by screening MAC genomes against sequences in the MEGARes 2.0 database (58) with ABRicate (103). Published point mutations in *rrs* and *rrl* conferring resistance to aminoglycosides and macrolides, respectively (56, 57), were identified by manual inspection of these gene sequences for each isolate.

Strain tracking of WUMAC isolates. Phylogenies for all pulmonary *M. avium* isolates ($n = 56$) and all *M. intracellulare* isolates ($n = 44$) were generated with RAxML (96) and visualized with iTOL (89) as described above. Separate core genome alignments were created with Roary for all *M. avium* isolates ($n = 25$) and all *M. intracellulare* isolates ($n = 16$) from the WUMAC and FLAC cohorts. Pairwise SNP distance matrices were created from these WUMAC/FLAC core genome alignments with snp-dists (104) and visualized with the R package ggplot2. To construct whole-genome SNP alignments, reads from all WUMAC/FLAC *M. avium* and *M. intracellulare* isolates were aligned with the representative strains *M. avium* OCU464 and *M. intracellulare* ATCC 13950, respectively, with Snippy (105).

Data availability. WUMAC and FLAC genomes have been deposited in the NCBI with BioProject accession numbers PRJNA682417 and PRJNA315990, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.03 MB.
FIG S2, PDF file, 0.2 MB.
FIG S3, PDF file, 0.03 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.02 MB.
TABLE S1, DOCX file, 0.03 MB.
TABLE S2, DOCX file, 0.02 MB.
TABLE S3, DOCX file, 0.03 MB.
TABLE S4, DOCX file, 0.03 MB.
TABLE S5, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank Robert Thänert, Sanjam Sawhney, Skye Fishbein, and David Sibley for valuable discussion, Jessica Hoisington-López and MariaLynn Jaeger for sequencing support, and Eric Martin and Brian Koebbe for computational support.

This work was supported in part by awards to G.D. from the National Institute for Occupational Safety and Health of the U.S. Centers for Disease Control and Prevention (R01 OH011578) and from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R01 AI123394) and to L.J.C. from the Cystic Fibrosis Foundation (CAVERL17A0) and the National Heart, Lung, and Blood Institute of the NIH (K23HL136934). E.C.K. is supported by a Graduate Research Fellowship from the National Science Foundation (DGE-1143945).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

REFERENCES

1. Faria S, Joao I, Jordao L. 2015. General overview on nontuberculous mycobacteria, biofilms, and human infection. J Pathog 2015:809014. https://doi.org/10.1155/2015/809014.
2. Honda JR, Virdi R, Chan ED. 2018. Global environmental nontuberculous mycobacteria and their contemporaneous man-made and natural niches. Front Microbiol 9:2029. https://doi.org/10.3389/fmicb.2018.02029.
3. Nishiuchi Y, Iwamoto T, Maruyama F. 2017. Infection sources of a common non-tuberculous mycobacterial pathogen, Mycobacterium avium complex. Front Med (Lausanne) 4:27. https://doi.org/10.3389/fmed.2017.00027.
4. Diel R, Lipman M, Hoefsloot W. 2018. High mortality in patients with Mycobacterium avium complex lung disease: a systematic review. BMC Infect Dis 18:206. https://doi.org/10.1186/s12879-018-3113-x.

5. Jhun BW, Kim SY, Moon SM, Jeon K, Kwon OJ, Huh HJ, Ki CS, Lee NY, Shin SJ, Daley CL, Koh WJ. 2018. Development of macrolide resistance and reinfection in refractory Mycobacterium avium complex lung disease. Am J Respir Crit Care Med 198:1322–1330. https://doi.org/10.1164/rccm.201802-0321OC.
6. Daley CL, Winthrop KL. 2020. Mycobacterium avium complex: addressing gaps in diagnosis and management. J Infect Dis 222:S199–S211. https://doi.org/10.1093/infdis/jiaa354.
7. Marras TK, Mendelson D, Marchand-Austin A, May K, Jamieson FB. 2013. Pulmonary nontuberculous mycobacterial disease, Ontario, Canada, 1998–2010. Emerg Infect Dis 19:1889–1891. https://doi.org/10.3201/eid1911.130737.
8. Lai CC, Tan CK, Chou CH, Hsu HL, Liao CH, Huang YT, Yang PC, Luh KT, Hsueh PR. 2010. Increasing incidence of nontuberculous mycobacteria, Taiwan, 2000–2008. Emerg Infect Dis 16:294–296. https://doi.org/10.3201/eid1602.090675.
9. Kendall BA, Varley CD, Choi D, Cassidy PM, Hedberg K, Ware MA, Winthrop KL. 2011. Distinguishing tuberculosis from nontuberculous mycobacteria lung disease, Oregon, USA. Emerg Infect Dis 17:506–509. https://doi.org/10.3201/eid1703.101164.
10. Namkoong H, Kurashima A, Morimoto K, Hoshino Y, Hasegawa N, Ato M, Mitarai S. 2016. Epidemiology of pulmonary nontuberculous mycobacterial disease, Japan. Emerg Infect Dis 22:1116–1117. https://doi.org/10.3201/eid2206.151086.
11. van Ingen J, Turenne CY, Tortoli E, Wallace RJ, Jr, Brown-Elliott BA. 2018. A definition of the Mycobacterium avium complex for taxonomical and clinical purposes, a review. Int J Syst Evol Microbiol 68:3666–3677. https://doi.org/10.1099/ijsem.0.003026.
12. Tortoli E, Meehan CJ, Grottola A, Fregni Serpini G, Fabio A, Trovato A, Pecorari M, Cirillo DM. 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.
13. Azar M, Zimbric M, Shedden K, Caverly LJ. 2020. Distribution and outcomes of infection of Mycobacterium avium complex species in cystic fibrosis. J Cyst Fibros 19:232–235. https://doi.org/10.1016/j.jcf.2019.07.007.
14. Boyle DP, Zembower TR, Reddy S, Qi C. 2015. Comparison of clinical features, virulence, and relapse among Mycobacterium avium complex species. Am J Respir Crit Care Med 191:1310–1317. https://doi.org/10.1164/rccm.201501-0067OC.
15. Bryant JM, Grogono DM, Rodriguez-Rincon D, Everall I, Brown KP, Moreno P, Verma D, Hill E, Drijkoningen J, Gilligan P, Esther CR, Noone PG, Giddings O, Bell SC, Thomson R, Wainwright CE, Coulter C, Pandey S, Wood ME, Stockwell RE, Ramsay KA, Sherrard LJ, Kidd TJ, Jabbour N, Johnson GR, Knibbs LD, Morawska L, Sly PD, Jones A, Bilton D, Laurenson I, Ruddy M, Bourke S, Bowler IC, Chapman SJ, Clayton A, Cullen M, Daniels T, Dempsey O, Denton M, Desai M, Drew RJ, Edenborough F, Evans J, Folb J, Humphrey H, Isalska B, Jensen-Fangel S, Jonsson B, Jones AM, et al. 2016. Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. Science 354:751–757. https://doi.org/10.1126/science.aaf8156.
16. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126–19131. https://doi.org/10.1073/pnas.0906412106.
17. van Ingen J, Lindeboom JA, Hartwig NG, de Zwaan R, Tortoli E, Dekhuijzen PN, Boeree MJ, van Soolingen D. 2009. Mycobacterium mantenii sp. nov., a pathogenic, slowly growing, scotochromogenic species. Int J Syst Evol Microbiol 59:2782–2787. https://doi.org/10.1099/ijs.0.010405-0.
18. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030.
19. Bannantine JP, Conde C, Bayles DO, Branger M, Biet F. 2020. Genetic diversity among Mycobacterium avium subspecies revealed by analysis of complete genome sequences. Front Microbiol 11:1701. https://doi.org/10.3389/fmicb.2020.01701.
20. Uchiya KI, 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 7:39750. https://doi.org/10.1038/srep39750.
21. Sakthi S, Palaniyandi K, Gupta UD, Gupta P, Narayanan S. 2016. Lipoprotein LpqS deficient M. tuberculosis mutant is attenuated for virulence in vivo and shows protective efficacy better than BCG in guinea pigs. Vaccine 34:735–743. https://doi.org/10.1016/j.vaccine.2015.12.059.
22. Delaune A, Dubrac S, Blanchet C, Poupel O, Mader U, Hiron A, Leduc A, Fitting C, Nicolas P, Cavaillon JM, Adib-Conquy M, Msadek T. 2012. The WalKR system controls major staphylococcal virulence genes and is involved in triggering the host inflammatory response. Infect Immun 80:3438–3453. https://doi.org/10.1128/IAI.00195-12.
23. Mashruwala AA, Boyd JM. 2017. The Staphylococcus aureus SrrAB regulatory system modulates hydrogen peroxide resistance factors, which imparts protection to aconitase during aerobic growth. PLoS One 12:e0170283. https://doi.org/10.1371/journal.pone.0170283.
24. Li XZ, Nikaido H. 2009. Efflux-mediated drug resistance in bacteria: an update. Drugs 69:1555–1623. https://doi.org/10.2165/11317030-000000000-00000.
25. Garcia SS, Du Q, Wu H. 2016. Streptococcus mutans copper chaperone, CopZ, is critical for biofilm formation and competitiveness. Mol Oral Microbiol 31:515–525. https://doi.org/10.1111/omi.12150.
26. Huang K, Wang D, Frederiksen RF, Rensing C, Olsen JE, Fresno AH. 2017. Investigation of the role of genes encoding zinc exporters zntA, zitB, and fieF during Salmonella Typhimurium infection. Front Microbiol 8:2656. https://doi.org/10.3389/fmicb.2017.02656.
27. Parker BW, Schwessinger EA, Jakob U, Gray MJ. 2013. The RclR protein is a reactive chlorine-specific transcription factor in Escherichia coli. J Biol Chem 288:32574–32584. https://doi.org/10.1074/jbc.M113.503516.
28. Zheng Q, Song Y, Zhang W, Shaw N, Zhou W, Rao Z. 2015. Structural views of quinone oxidoreductase from Mycobacterium tuberculosis reveal large conformational changes induced by the co-factor. FEBS J 282:2697–2707. https://doi.org/10.1111/febs.13312.
29. Agoro R, Mura C. 2019. Iron supplementation therapy, a friend and foe of mycobacterial infections? Pharmaceuticals (Basel) 12:75. https://doi.org/10.3390/ph12020075.
30. Sritharan M. 2016. Iron homeostasis in Mycobacterium tuberculosis: mechanistic Insights into siderophore-mediated iron uptake. J Bacteriol 198:2399–2409. https://doi.org/10.1128/JB.00359-16.
31. Panda A, Drancourt M, Tuller T, Pontarotti P. 2018. Genome-wide analysis of horizontally acquired genes in the genus Mycobacterium. Sci Rep 8:14817. https://doi.org/10.1038/s41598-018-33261-w.
32. Wang J, Behr MA. 2014. Building a better bacillus: the emergence of Mycobacterium tuberculosis. Front Microbiol 5:139. https://doi.org/10.3389/fmicb.2014.00139.
33. Eldholm V, Balloux F. 2016. Antimicrobial resistance in Mycobacterium tuberculosis: the odd one out. Trends Microbiol 24:637–648. https://doi.org/10.1016/j.tim.2016.03.007.
34. Reva O, Korotetskiy I, Ilin A. 2015. Role of the horizontal gene exchange in evolution of pathogenic mycobacteria. BMC Evol Biol 15(Suppl 1):S2. https://doi.org/10.1186/1471-2148-15-S1-S2.
35. Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P. 2007. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. Mol Biol Evol 24:1861–1871. https://doi.org/10.1093/molbev/msm111.
36. Jang J, Becq J, Gicquel B, Deschavanne P, Neyrolles O. 2008. Horizontally acquired genomic islands in the tubercle bacilli. Trends Microbiol 16:303–308. https://doi.org/10.1016/j.tim.2008.04.005.
37. Brennan MJ. 2017. The enigmatic PE/PPE multigene family of mycobacteria and tuberculosis vaccination. Infect Immun 85:e00969-16. https://doi.org/10.1128/IAI.00969-16.
38. Wang Q, Boshoff HIM, Harrison JR, Ray PC, Green SR, Wyatt PG, Barry CE, III. 2020. PE/PPE proteins mediate nutrient transport across the outer membrane of Mycobacterium tuberculosis. Science 367:1147–1151. https://doi.org/10.1126/science.aav5912.
39. Fishbein S, van Wyk N, Warren RM, Sampson SL. 2015. Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. Mol Microbiol 96:901–916. https://doi.org/10.1111/mmi.12981.
40. Tang X, Yu P, Tang L, Zhou M, Fan C, Lu Y, Mathieu J, Xiong W, Alvarez PJ. 2019. Bacteriophages from arsenic-resistant bacteria transduced resistance genes, which changed arsenic speciation and increased soil toxicity. Environ Sci Technol Lett 6:675–680. https://doi.org/10.1021/acs.estlett.9b00600.
41. Bottai D, Stinear TP, Supply P, Brosch R. 2014. Mycobacterial pathogenomics and evolution. Microbiol Spectr 2:MGM2-0025-2013. https://doi.org/10.1128/microbiolspec.MGM2-0025-2013.
42. Steiner EM, Both D, Lossl P, Vilaplana F, Schnell R, Schneider G. 2014. CysK2 from Mycobacterium tuberculosis is an *O*-phospho-L-serine-dependent *S*-sulfocysteine synthase. J Bacteriol 196:3410–3420. https://doi.org/10.1128/JB.01851-14.
43. Scherzinger D, Scheffer E, Bar C, Ernst H, Al-Babili S. 2010. The Mycobacterium tuberculosis ORF Rv0654 encodes a carotenoid oxygenase

mediating central and excentric cleavage of conventional and aromatic carotenoids. FEBS J 277:4662–4673. https://doi.org/10.1111/j.1742-4658.2010.07873.x.
44. Claeys TA, Robinson RT. 2018. The many lives of nontuberculous mycobacteria. J Bacteriol 200:e00739-17. https://doi.org/10.1128/JB.00739-17.
45. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res 48:D606–D612. https://doi.org/10.1093/nar/gkz943.
46. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res 47:D687–D692. https://doi.org/10.1093/nar/gky1080.
47. Sayers S, Li L, Ong E, Deng S, Fu G, Lin Y, Yang B, Zhang S, Fa Z, Zhao B, Xiang Z, Li Y, Zhao XM, Olszewski MA, Chen L, He Y. 2019. Victors: a web-based knowledge base of virulence factors in human and animal pathogens. Nucleic Acids Res 47:D693–D700. https://doi.org/10.1093/nar/gky999.
48. Ryndak MB, Wang S, Smith I, Rodriguez GM. 2010. The Mycobacterium tuberculosis high-affinity iron importer, IrtA, contains an FAD-binding domain. J Bacteriol 192:861–869. https://doi.org/10.1128/JB.00223-09.
49. Krithika R, Marathe U, Saxena P, Ansari MZ, Mohanty D, Gokhale RS. 2006. A genetic locus required for iron acquisition in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103:2069–2074. https://doi.org/10.1073/pnas.0507924103.
50. Pethe K, Alonso S, Biet F, Delogu G, Brennan MJ, Locht C, Menozzi FD. 2001. The heparin-binding haemagglutinin of M. tuberculosis is required for extrapulmonary dissemination. Nature 412:190–194. https://doi.org/10.1038/35084083.
51. Falkinham JO, III. 2018. Challenges of NTM drug development. Front Microbiol 9:1613. https://doi.org/10.3389/fmicb.2018.01613.
52. Parker H, Lorenc R, Ruelas Castillo J, Karakousis PC. 2020. Mechanisms of antibiotic tolerance in Mycobacterium avium complex: lessons from related mycobacteria. Front Microbiol 11:573983. https://doi.org/10.3389/fmicb.2020.573983.
53. Uchiya K, Takahashi H, Nakagawa T, Yagi T, Moriyama M, Inagaki T, Ichikawa K, Nikai T, Ogawa K. 2015. Characterization of a novel plasmid, pMAH135, from Mycobacterium avium subsp. hominissuis. PLoS One 10:e0117797. https://doi.org/10.1371/journal.pone.0117797.
54. Shulha JA, Escalante P, Wilson JW. 2019. Pharmacotherapy approaches in nontuberculous mycobacteria infections. Mayo Clin Proc 94:1567–1581. https://doi.org/10.1016/j.mayocp.2018.12.011.
55. Kwon YS, Koh WJ, Daley CL. 2019. Treatment of Mycobacterium avium complex pulmonary disease. Tuberc Respir Dis (Seoul) 82:15–26. https://doi.org/10.4046/trd.2018.0060.
56. Jamal MA, Maeda S, Nakata N, Kai M, Fukuchi K, Kashiwabara Y. 2000. Molecular basis of clarithromycin-resistance in Mycobacterium avium intracellulare complex. Tuber Lung Dis 80:1–4. https://doi.org/10.1054/tuld.1999.0227.
57. Huh HJ, Kim SY, Shim HJ, Kim DH, Yoo IY, Kang OK, Ki CS, Shin SY, Jhun BW, Shin SJ, Daley CL, Koh WJ, Lee NY. 2019. GenoType NTM-DR performance evaluation for identification of Mycobacterium avium complex and Mycobacterium abscessus and determination of clarithromycin and amikacin resistance. J Clin Microbiol 57:e00516-19. https://doi.org/10.1128/JCM.00516-19.
58. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, Belk KE, Noyes NR, Morley PS. 2020. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res 48:D561–D569. https://doi.org/10.1093/nar/gkz1010.
59. Dey A, Verma AK, Chatterji D. 2010. Role of an RNA polymerase interacting protein, MsRbpA, from Mycobacterium smegmatis in phenotypic tolerance to rifampicin. Microbiology (Reading) 156:873–883. https://doi.org/10.1099/mic.0.033670-0.
60. Doran JL, Pang Y, Mdluli KE, Moran AJ, Victor TC, Stokes RW, Mahenthiralingam E, Kreiswirth BN, Butt JL, Baron GS, Treit JD, Kerr VJ, Van Helden PD, Roberts MC, Nano FE. 1997. Mycobacterium tuberculosis efpA encodes an efflux protein of the QacA transporter family. Clin Diagn Lab Immunol 4:23–32. https://doi.org/10.1128/cdli.4.1.23-32.1997.
61. Rouquette C, Harmon JB, Shafer WM. 1999. Induction of the mtrCDE-encoded efflux pump system of Neisseria gonorrhoeae requires MtrA, an AraC-like protein. Mol Microbiol 33:651–658. https://doi.org/10.1046/j.1365-2958.1999.01517.x.
62. van Ingen J, Boeree MJ, Kosters K, Wieland A, Tortoli E, Dekhuijzen PN, van Soolingen D. 2009. Proposal to elevate Mycobacterium avium complex ITS sequevar MAC-Q to Mycobacterium vulneris sp. nov. Int J Syst Evol Microbiol 59:2277–2282. https://doi.org/10.1099/ijs.0.008854-0.
63. Timm J, Perilli MG, Duez C, Trias J, Orefici G, Fattorini L, Amicosante G, Oratore A, Joris B, Frere JM. 1994. Transcription and expression analysis, using lacZ and phoA gene fusions, of Mycobacterium fortuitum beta-lactamase genes cloned from a natural isolate and a high-level beta-lactamase producer. Mol Microbiol 12:491–504. https://doi.org/10.1111/j.1365-2958.1994.tb01037.x.
64. Jeffrey B, Rose SJ, Gilbert K, Lewis M, Bermudez LE. 2017. Comparative analysis of the genomes of clinical isolates of Mycobacterium avium subsp. hominissuis regarding virulence-related genes. J Med Microbiol 66:1063–1075. https://doi.org/10.1099/jmm.0.000507.
65. Yano H, Iwamoto T, Nishiuchi Y, Nakajima C, Starkova DA, Mokrousov I, Narvskaya O, Yoshida S, Arikawa K, Nakanishi N, Osaki K, Nakagawa I, Ato M, Suzuki Y, Maruyama F. 2017. Population structure and local adaptation of MAC lung disease agent Mycobacterium avium subsp. hominissuis. Genome Biol Evol 9:2403–2417. https://doi.org/10.1093/gbe/evx183.
66. Wallace RJ, Jr, Zhang Y, Brown BA, Dawson D, Murphy DT, Wilson R, Griffith DE. 1998. Polyclonal Mycobacterium avium complex infections in patients with nodular bronchiectasis. Am J Respir Crit Care Med 158:1235–1244. https://doi.org/10.1164/ajrccm.158.4.9712098.
67. Fujita K, Ito Y, Hirai T, Kubo T, Maekawa K, Togashi K, Ichiyama S, Mishima M. 2014. Association between polyclonal and mixed mycobacterial Mycobacterium avium complex infection and environmental exposure. Annals ATS 11:45–53. https://doi.org/10.1513/AnnalsATS.201309-297OC.
68. Kimizuka Y, Hoshino Y, Nishimura T, Asami T, Sakakibara Y, Morimoto K, Maeda S, Nakata N, Abe T, Uno S, Namkoong H, Fujiwara H, Funatsu Y, Yagi K, Fujie T, Ishii M, Inase N, Iwata S, Kurashima A, Betsuyaku T, Hasegawa N, Non-Tuberculous Mycobacteriosis-Japan Research Consortium (NTM-JRC). 2019. Retrospective evaluation of natural course in mild cases of Mycobacterium avium complex pulmonary disease. PLoS One 14:e0216034. https://doi.org/10.1371/journal.pone.0216034.
69. Operario DJ, Pholwat S, Koeppel AF, Prorock A, Bao Y, Sol-Church K, Scheurenbrand M, Poulter M, Turner S, Parikh HI, Mathers A, Houpt ER. 2019. Mycobacterium avium complex diversity within lung disease, as revealed by whole-genome sequencing. Am J Respir Crit Care Med 200:393–396. https://doi.org/10.1164/rccm.201903-0669LE.
70. Winthrop KL, Marras TK, Adjemian J, Zhang H, Wang P, Zhang Q. 2020. Incidence and prevalence of nontuberculous mycobacterial lung disease in a large U.S. managed care health plan, 2008–2015. Annals ATS 17:178–185. https://doi.org/10.1513/AnnalsATS.201804-236OC.
71. Whittington R, Donat K, Weber MF, Kelton D, Nielsen SS, Eisenberg S, Arrigoni N, Juste R, Saez JL, Dhand N, Santi A, Michel A, Barkema H, Kralik P, Kostoulas P, Citer L, Griffin F, Barwell R, Moreira MAS, Slana I, Koehler H, Singh SV, Yoo HS, Chavez-Gris G, Goodridge A, Ocepek M, Garrido J, Stevenson K, Collins M, Alonso B, Cirone K, Paolicchi F, Gavey L, Rahman MT, de Marchin E, Van Praet W, Bauman C, Fecteau G, McKenna S, Salgado M, Fernandez-Silva J, Dziedzinska R, Echeverria G, Seppanen J, Thibault V, Fridriksdottir V, Derakhshandeh A, Haghkhah M, Ruocco L, Kawaji S, et al. 2019. Control of paratuberculosis: who, why and how. A review of 48 countries. BMC Vet Res 15:198. https://doi.org/10.1186/s12917-019-1943-4.
72. Dhama K, Mahendran M, Tiwari R, Dayal Singh S, Kumar D, Singh S, Sawant PM. 2011. Tuberculosis in birds: insights into the Mycobacterium avium infections. Vet Med Int 2011:712369. https://doi.org/10.4061/2011/712369.
73. Ichikawa K, van Ingen J, Koh WJ, Wagner D, Salfinger M, Inagaki T, Uchiya KI, Nakagawa T, Ogawa K, Yamada K, Yagi T. 2015. Genetic diversity of clinical Mycobacterium avium subsp. hominissuis and Mycobacterium intracellulare isolates causing pulmonary diseases recovered from different geographical regions. Infect Genet Evol 36:250–255. https://doi.org/10.1016/j.meegid.2015.09.029.
74. 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 12:846–852. https://doi.org/10.1016/j.meegid.2011.06.018.
75. Iakhiaeva E, Howard ST, Brown Elliott BA, McNulty S, Newman KL, Falkinham JO, III, Williams M, Kwait R, Lande L, Vasireddy R, Turenne C, Wallace RJ, Jr. 2016. Variable-number tandem-repeat analysis of respiratory and household water biofilm isolates of "Mycobacterium avium

subsp. hominissuis" with establishment of a PCR database. J Clin Microbiol 54:891–901. https://doi.org/10.1128/JCM.02409-15.
76. Radomski N, Thibault VC, Karoui C, de Cruz K, Cochard T, Gutierrez C, Supply P, Biet F, Boschiroli ML. 2010. Determination of genotypic diversity of Mycobacterium avium subspecies from human and animal origins by mycobacterial interspersed repetitive-unit-variable-number tandem-repeat and IS1311 restriction fragment length polymorphism typing methods. J Clin Microbiol 48:1026–1034. https://doi.org/10.1128/JCM.01869-09.
77. Uchiya KI, Asahi S, Futamura K, Hamaura H, Nakagawa T, Nikai T, Ogawa K. 2018. Antibiotic Susceptibility and genotyping of Mycobacterium avium strains that cause pulmonary and disseminated infection. Antimicrob Agents Chemother 62:e02035-17. https://doi.org/10.1128/AAC.02035-17.
78. Westermann AJ, Forstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Muller L, Reinhardt R, Stadler PF, Vogel J. 2016. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature 529:496–501. https://doi.org/10.1038/nature16547.
79. Liu Q, Wei J, Li Y, Wang M, Su J, Lu Y, Lopez MG, Qian X, Zhu Z, Wang H, Gan M, Jiang Q, Fu YX, Takiff HE, Comas I, Li F, Lu X, Fortune SM, Gao Q. 2020. Mycobacterium tuberculosis clinical isolates carry mutational signatures of host immune environments. Sci Adv 6:eaba4901. https://doi.org/10.1126/sciadv.aba4901.
80. De Groote MA, Pace NR, Fulton K, Falkinham JO, III. 2006. Relationships between Mycobacterium isolates from patients with pulmonary mycobacterial infection and potting soils. Appl Environ Microbiol 72:7602–7606. https://doi.org/10.1128/AEM.00930-06.
81. Fujita K, Ito Y, Hirai T, Maekawa K, Imai S, Tatsumi S, Niimi A, Iinuma Y, Ichiyama S, Mishima M. 2013. Genetic relatedness of Mycobacterium avium-intracellulare complex isolates from patients with pulmonary MAC disease and their residential soils. Clin Microbiol Infect 19:537–541. https://doi.org/10.1111/j.1469-0691.2012.03929.x.
82. Thomson R, Tolson C, Carter R, Coulter C, Huygens F, Hargreaves M. 2013. Isolation of nontuberculous mycobacteria (NTM) from household water and shower aerosols in patients with pulmonary disease caused by NTM. J Clin Microbiol 51:3006–3011. https://doi.org/10.1128/JCM.00899-13.
83. Tichenor WS, Thurlow J, McNulty S, Brown-Elliott BA, Wallace RJ, Jr, Falkinham JO, III. 2012. Nontuberculous mycobacteria in household plumbing as possible cause of chronic rhinosinusitis. Emerg Infect Dis 18:1612–1617. https://doi.org/10.3201/eid1810.120164.
84. Lande L, Alexander DC, Wallace RJ, Jr, Kwait R, Iakhiaeva E, Williams M, Cameron ADS, Olshefsky S, Devon R, Vasireddy R, Peterson DD, Falkinham JO, III. 2019. Mycobacterium avium in community and household water, suburban Philadelphia, Pennsylvania, USA, 2010–2012. Emerg Infect Dis 25:473–481. https://doi.org/10.3201/eid2503.180336.
85. Caverly LJ, Spilker T, LiPuma JJ. 2016. Complete genome sequence of Mycobacterium abscessus subsp. bolletii. Genome Announc 4:e00543-16. https://doi.org/10.1128/genomeA.00543-16.
86. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS One 10:e0128036. https://doi.org/10.1371/journal.pone.0128036.
87. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
88. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
89. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.
90. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
91. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314. https://doi.org/10.1093/nar/gky1085.
92. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24. https://doi.org/10.1039/C5AY02550H.
93. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
94. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421.
95. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. 2017. Efficient inference of recent and ancestral recombination within bacterial populations. Mol Biol Evol 34:1167–1182. https://doi.org/10.1093/molbev/msx066.
96. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
97. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 17:238. https://doi.org/10.1186/s13059-016-1108-8.
98. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269. https://doi.org/10.1093/nar/gku1223.
99. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148.
100. Bertelli C, Brinkman FSL. 2018. Improved genomic island predictions with IslandPath-DIMOB. Bioinformatics 34:2161–2167. https://doi.org/10.1093/bioinformatics/bty095.
101. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman FSL. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45:W30–W35. https://doi.org/10.1093/nar/gkx343.
102. Steinegger M, Soding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35:1026–1028. https://doi.org/10.1038/nbt.3988.
103. Seemann T. 2020. ABRicate. GitHub https://github.com/tseemann/abricate.
104. Seemann T. 2020. snp-dists. GitHub https://github.com/tseemann/snp-dists.
105. Seemann T. 2015. Snippy: fast bacterial variant calling from NGS reads. GitHub https://github.com/tseemann/snippy.