Emerging Microbes & Infections
2025, VOL. 14, 2449087 (14 pages)
https://doi.org/10.1080/22221751.2024.2449087

Unravelling distinct patterns of metagenomic surveillance and respiratory microbiota between two P1 genotypes of *Mycoplasma pneumoniae*

Hailong You[a]*, Bin Yang[b]*, Huifang Liu[b]*, Wencai Wu[b], Fei Yu[c], Nan Lin[d], WenJiao Yang[b], Bingxue Hu[b], Yong Liu[a], Hongyan Zou[a], Sijia Hao[a], Yunping Xiao[a], Teng Xu[b] and Yanfang Jiang[a]

[a]Genetic Diagnosis Center, The First Hospital of Jilin University, Changchun, Jilin, People's Republic of China; [b]Center for Infectious Diseases, Vision Medicals Co., Ltd, Guangzhou, Guangdong, People's Republic of China; [c]Medical Research Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, People's Republic of China; [d]Department of Environmental Health, School of Public Health, Shanghai Jiao Tong University, Shanghai, People's Republic of China

ABSTRACT

To unravel distinct patterns of metagenomic surveillance and respiratory microbiota between *Mycoplasma pneumoniae* (*M. pneumoniae*) P1-1 and P1-2 and to explore the impact of the COVID-19 pandemic on epidemiological features, we conducted a multicentre retrospective study which spanned 90,886 pneumonia patients, among which 3164 cases *M. pneumoniae* were identified. Our findings revealed a concurrent outbreak of *M. pneumoniae*, with the positivity rate rising sharply to 9.62% from July 2023, compared to the 0.16% to 4.06% positivity rate observed during the 2020–2022 COVID-19 pandemic. P1-1 had a higher odds ratio of co-detecting opportunistic pathogens. However, no significant differences were observed in the co-detection odds ratio between children and other age groups in P1-2. This study is the first to demonstrate differences in relative abundance, diversity of respiratory microbiota and co-detection rate of opportunistic pathogen between *M. pneumoniae* P1-1 and P1-2. Through bronchoalveolar lavage (BAL) metagenomic and host transcriptomic analyses, we identified variations in co-detection rates of *M. pneumoniae* P1-1 genotype with opportunistic pathogens like *S. pneumoniae*, alterations in respiratory microbiota composition, lung inflammation, and disruption of ciliary function. Consistent with the results of host transcriptome, we found that P1-1 infections were associated with significantly higher rates of requiring respiratory support and mechanical ventilation compared to P1-2 infections (Fisher's exact test, *p*-value = 0.035/0.004). Our study provides preliminary evidence of clinical severity between *M. pneumoniae* strains, underscoring the need for ongoing research and development of targeted therapeutic strategies.

ARTICLE HISTORY Received 24 March 2024; Revised 4 December 2024; Accepted 29 December 2024

KEYWORDS *Mycoplasma pneumoniae*; P1 genotype; host immune response; clinical outcome; respiratory microbiota

Introduction

Mycoplasma pneumoniae (*M. pneumoniae*) is a common cause of lower respiratory tract infections (LRTI) in children, particularly affecting those aged 5–15 years and accounting for up to 40% of community-acquired pneumonia cases in this age group. While most infections are mild and resolve spontaneously, some can progress to severe or critical illness [1–3]. The factors influencing the severity of *M. pneumoniae* LRTI remain poorly understood. *M. pneumoniae* is primarily classified into two genetic groups, P1 type 1 (P1-1) and P1 type 2 (P1-2), based on nucleotide variations in two repetitive elements (RepMP2/3 and RepMP4) within the MPN141 gene, which encodes the P1 adhesion protein [4]. Previous studies have suggested that these genetic variants may differ in pathogenic potential and in vitro growth rates. However, whether these genetic groups contribute to different clinical manifestations of *M. pneumoniae* LRTI remains unclear [5]. Furthermore, although several studies have described the epidemiological characteristics of *M. pneumoniae*, it remains unclear whether different P1 genotypes of *M. pneumoniae* are associated with varying respiratory symptoms, such as co-infections, ICU staying days, and respiratory-related symptoms (needed mechanical ventilation) et al. Further research is needed to investigate there were different impacts of different P1 genotypes [6].

Immune responses play a pivotal role in determining the severity of diseases caused by *M. pneumoniae* infection [7]. Recent research has illuminated how interactions between microbes can influence the onset, progression, and prevalence of respiratory illnesses [8]. Increasing evidence suggests that microbial communities within the respiratory system can activate the

CONTACT Yanfang Jiang yanfangjiang@hotmail.com, yanfangjiang@jlu.edu.cn Genetic Diagnosis Center, The First Hospital of Jilin University, Changchun, Jilin 130021, People's Republic of China
*These authors contributed equally to this work.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/22221751.2024.2449087.

immune system, thereby enhancing its ability to defend against pathogen invasion [9]. The respiratory microbiota is crucial in modulating and directing immune responses within the lungs; even minor shifts in microbiota composition can significantly impact host immunity, due to the diverse structural ligands and metabolites produced by different microbial species [10]. While numerous studies have highlighted disruptions in the respiratory microbiota among children suffering from respiratory infections [11,12], the role of commensal bacteria in influencing *M. pneumoniae* pathogenesis remains under-explored. Understanding this interaction is essential for developing strategies to mitigate *M. pneumoniae* proliferation and infection in the respiratory system. Therefore, a comprehensive understanding of the respiratory microbiota and ecological characteristics of different *M. pneumoniae* P1 genotypes is fundamental for the prevention and management of *M. pneumoniae* infections.

In this study, we conducted a multicentre, retrospective analysis involving 90,886 pneumonia patients across three locations, identifying 3164 *M. pneumoniae*-positive cases through metagenomic next-generation sequencing (mNGS). A total of 183 draft genomes derived from metagenomic sequencing data were used for phylogenetic analysis, which was compared with 290 public reference genomes. Additionally, we analysed 1368 samples with reliable genotyping results to assess the relative abundance, diversity of respiratory microbiota, pathogen spectrum, and co-detection rates of opportunistic pathogens across different time periods (before, during, and after COVID-19) and age groups (children, school-age, adults, and older adults). Host transcriptomic analysis was performed on 174 samples with paired bronchoalveolar lavage (BAL) metatranscriptomic data. Clinical data from 118 patients were also collected to investigate the severity of infection associated with different P1 genotypes. Our results revealed the impact of non-pharmaceutical interventions (NPIs) on the national incidence of *M. pneumoniae* infection and explored difference between two P1 genotypes of *M. pneumoniae*. Through BAL metagenome and host transcriptome analyses, we observed variations in co-detection rates of *M. pneumoniae* P1 genotypes with opportunistic pathogens like *S. pneumoniae*, as well as differences in respiratory microbiota composition, lung inflammation, and ciliary function. Clinical outcome data further demonstrated that these variations contribute to differences in clinical severity among patients with *M. pneumoniae* pneumonia.

Methods

Patient selection and study design

A cohort of 90,886 pneumonia patients from three locations (Jilin, Shanghai, Guangdong) underwent bronchoalveolar lavage fluid (BAL)/sputum mNGS testing to investigate potential infections during their hospitalization between January 2019 and December 2023. A total of 3164 *M. pneumoniae*-positive patients were included in this study (Figure 1(A), Figure S1A). Pneumonia diagnosis followed the Chinese Thoracic Society (CTS) guidelines [13] with minor modifications. Patients with chest radiographs showing punctate, patchy, or uniform density opacity were considered to have radiographic evidence of pneumonia. The clinical diagnostic criteria for pneumonia included: (1) presence of clinical manifestations of pneumonia: ① new onset of cough or expectoration, or aggravation of existing symptoms of respiratory tract diseases, with or without purulent sputum, chest pain, dyspnoea, or haemoptysis; ② fever; ③ signs of pulmonary consolidation and/or moist rales; ④ peripheral WBC $> 10 \times 10^9$/L or $< 4 \times 10^9$/L, with or without a left shift, together with (2) chest radiograph showing new patchy infiltrates, lobar or segmental consolidation, ground-glass opacities, or interstitial changes, with or without pleural effusion. This study was approved by the institutional review board of the First Hospital of Jilin University (No. 2024-415). The requirement for patient informed consent was waived by the institutional review board, as the data used in this study did not contain personally identifiable information.

BAL and sputum DNA/RNA extraction, library construction and sequencing

BAL and sputum mNGS tests for microbial DNA were performed for each patient. 1 mL BAL sample was used for nucleic acid extraction as previously described. Briefly, host cells and DNA are removed using Tween 20 (Sigma) and Benzonase (Qiagen), and microbial DNA are extracted with the QIAamp® UCP Pathogen DNA Kit (Qiagen) following the manufacturer's instructions. After nucleic acid extraction, libraries were constructed using a Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA). For RNA sequencing, a 1 mL BAL sample was centrifuged at 12,000 rpm for 10 minutes. 800 μL of super was removed and the rest was lysed in TRIzol LS (Thermo Fisher Scientific, Carlsbad, CA, USA), and RNA was subsequently extracted using the Direct-zol RNA Miniprep kit (Zymo Research, Irvine, CA, USA). 10 μL of purified RNA was employed for cDNA synthesis and library preparation with the Ovation Trio RNA-Seq Library Preparation Kit (NuGEN, CA, USA). Library concentration was measured using a Qubit dsDNA HS Assay Kit, and library quality was evaluated with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) utilizing a High Sensitivity DNA kit. The libraries were sequenced on an Illumina Nextseq CN500 sequencer for 75 cycles

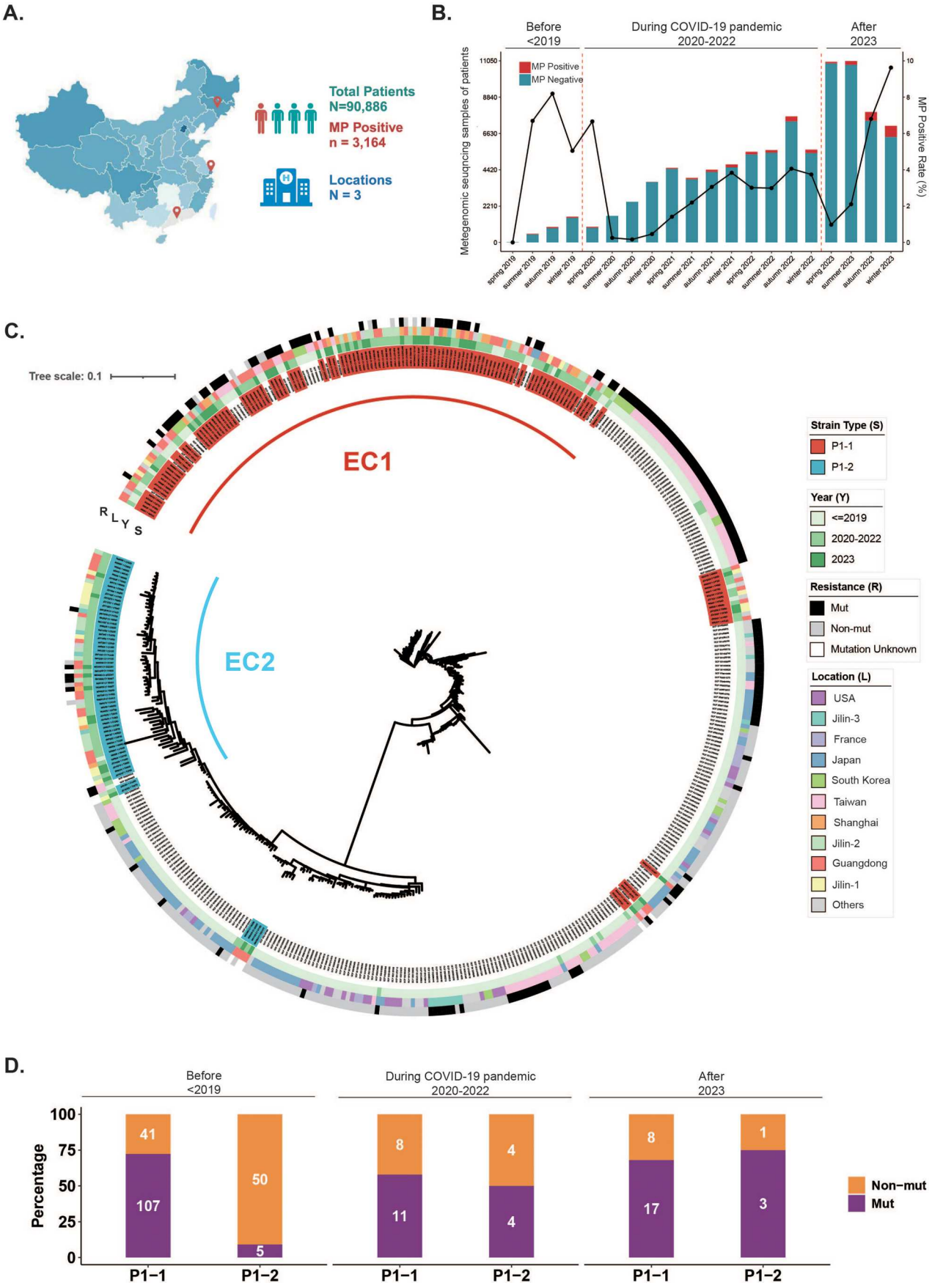


Figure 1. The epidemiology of *M. pneumoniae* and positive rate of 23s rRNA mutation in China. (A) The number and locations of bronchoscopy mNGS tests applied to patients with pneumonia from 2019 to 2023. Red represents *M. pneumoniae* positive and green represents *M. pneumoniae* negative in mNGS data. (B) The average number of mNGS tests per season for pneumonia patients and positive number and rate of *M. pneumoniae* infection during 2019-2023. (C) Phylogenetic tree revealed the presence of two primary epidemic clones of EC1 and EC2 in China. (D) The composition of *M. pneumoniae* changed dramatically with P1-2 Mut replacing non-mut strains after COVID-19 pandemic. Purple/Mut: mutations in the 23S rRNA gene (C2617T, A2063G, A2064G). Orange/Non-mut: there were no mutations in the 23S rRNA gene (C2617T, A2063G, A2064G).

of single-end sequencing, generating approximately 20 million reads per library. For quality control, both negative and positive samples were included in each batch [14–16].

Lung microbiome bioinformatic analysis

Adaptor contamination, low-quality reads, duplicate reads, low-complexity reads and reads shorter than 40 bp were removed by fastp (v0.23.4) default settings [17]. The human sequence data were located and eliminated by matching them to the hg38 reference genome utilizing Burrows–Wheeler Aligner software (BWA-0.7.17) [18]. Taxonomic profiles for all metagenomic samples were generated using Kraken 2 (v2.1.3 [19]) and Bracken v2.5 (https://doi.org/10.7717/peerj-cs.104) with default settings. The database used for quantifying taxonomic profiles was constructed using a combined database containing human, bacterial, fungal, archaeal, and viral genomes downloaded from NCBI RefSeq (https://benlangmead.github.io/aws-indexes/k2). Positive samples were identified based on microorganisms detected above a predefined threshold, determined by the highest area under the curve (AUC). For microorganisms lacking culture isolates, a positive cut-off value was established as the mean relative abundance plus three standard deviations (mean + 3SD) [14,20].

M. pneumoniae reads were aligned to the database using BWA software [18]. The reads with 90% identity of reference were defined as mapped reads. In addition, reads with multiple locus alignments within the same genus were excluded from the secondary analysis. Only reads mapped to the genome within the same species were considered [14]. Relative abundance data of *M. pneumoniae* at the species level were visualized. The quantitative skewness of *M. pneumoniae* relative abundance data was expressed as median (interquartile range) and analysed using the Wilcoxon–Mann–Whitney rank-sum test. The proportion of samples in which *M. pneumoniae* was the most dominant pathogen was analysed as categorical data using Fisher's exact test [11,15,21].

The vegan package [22] in R software was used to analyse the microbiome's alpha and beta diversity of the microbiome. Alpha diversity metrics (ACE, Chao1 estimator, Shannon and Simpson Indices) presented as median (interquartile range) and were analysed using the Wilcoxon–Mann–Whitney rank-sum test. Beta diversity was measured by the Bray–Curtis distance test. Principal coordinate analysis (PCoA) was used to visualize similar distances between samples [15]. Relative abundance differences of the microorganisms ranked in the top 10 between the *M. pneumoniae* at the species level were displayed. The samples were clustered based on the relative abundance of species in each sample and the Bray–Curtis distance. The differentially expressed species (DES) of the microbiome ranked in the top 10 between P1-1 and P1-2 types and among different periods (before, during and after COVID-19 pandemic) or age groups (child, school-age, adults and older-people) were selected using linear discriminant analysis (LDA) effect size (LEfSe) [23]. The DESs were determined by an LDA score > 2.0, and a $p < 0.05$ [15].

M. pneumoniae genomic analysis, genotyping and 23s rRNA resistance mutation calling

Metagenomic reads of samples with *M. pneumoniae* were assigned to P1-1 and P1-2 type with StrainGE [24] with default parameters. Also, metagenomic reads mapped to either the M129 reference genome (NCBI accession numbers: NC_000912.1 and CP010546.1, respectively) using Snippy v4.6.0 (https://github.com/tseemann/snippy) with default parameters. When mapping reads to a reference genome, Snippy could generate a consensus sequence of the reference genome. We annotated the consensus sequences with high sequencing depth (>5) and coverage (>95%) using Prokka v1.14.5 [25]. We defined 23s rRNA A2063G mutation with sequencing depth higher than 5, mutation frequency higher than 50% and at least 3 covered reads supporting that mutation. In parallel, a whole genome phylogeny was estimated by adding 290 *M. pneumoniae* genome sequences retrieved from public databases (Supplementary Table S2) with kSNP3 [26], with a Kchooser estimated k-mer length of 21, and a fraction of core k-mers of 0.46. Phylogenetic clusters were defined with ClusterPicker [27], initial and main support threshold of 0.9, a genetic distance threshold of 4.5, and a large cluster threshold of 20. The evolutionary relationship was analysed using kSNP3 and visualized by iTOL editor v5 [28].

Transcriptome of BAL samples

Genes with an FDR-corrected adjusted p value < 0.25 were considered significantly differentiated unless otherwise specified [29]. Pathway analysis using differentially regulated genes (FDR < 0.25) was done using KOBAS 3.0 [30]. GSEA was performed with differential genes (FDR < 0.25) for dataset comparison using R package fgsea v1.4.1 [31–33]. Normalized host transcriptome data as transcript per million (TPM) from the BAL metatranscriptome was subjected to digital cell-type quantification with CIBERSORTx [16,34]. Genes with at least one count per million in at least two samples were retained. The two-tailed Wilcoxon rank-sum test with Benjamini–Hochberg correction was computed between groups of samples for comparison.

Statistical analysis

Demographic and clinical data were summarized using descriptive statistics. Secondary infection rates were compared using Fisher's exact tests. Data analyses and visualizations were performed using R (version 4.1.2) with a p value ≤ 0.05 as the significance threshold, and all tests were two-tailed [15,16,35].

Results

Study population, positive rates and genomic surveillance of M. pneumoniae across COVID-19 pandemic

From 2019 to 2023, we obtained 90,886 metagenomic datasets (Figure 1(A), Figure S1A) derived from bronchoalveolar lavage fluid or sputum samples of pneumonia patients diagnosed according to the Chinese Thoracic Society (CTS) guidelines. Among these, 3164 samples tested positive for *M. pneumoniae* pneumonia. Consistent with previous findings, we observed a concurrent outbreak of *M. pneumoniae* during the winter of 2023 (Figure 1(B)). Notably, a marked resurgence of *M. pneumoniae* was recorded, with a positive rate of up to 9.62% since July 2023 (Figure 1(B)), compared to a lower incidence ranging from 0.16% to 4.06% during the COVID-19 period (2020–2022). To explore genomic characteristics of *M. pneumoniae*, we mapped all metagenomic data to reference genome (GCA_900660465.1, Strain M129). Furthermore, sequencing reads identified as *M. pneumoniae* were assigned to P1 type 1 (P1-1) or P1 type 2 (P1-2) using StrainGE (Methods). A total of 1,368 samples (42.96%) had over 10,000 reads, and 183 samples (5.78%) achieved coverage greater than 95% (Figure S1A–C). Demographic characteristics of patients infected with these two P1 genotypes are summarized in Supplementary Table S1. We selected data with >95% coverage of *M. pneumoniae* for further analysis.

A phylogenetic tree was subsequently constructed with public genomic data (Supplementary Table S2, $n = 290$) and showed the presence of two primary epidemic clones (EC) in China: EC1 in P1-1 and EC2 in P1-2 (Figure 1(C)). We found that P1-1 showed 72.30%, 57.89% and 68% A2063G mutation in 23S rRNA of *M. pneumoniae* before, during and after the COVID-19 pandemic (Figure 1(D)). We also found that P1-1 keeps a high 23s rRNA mutation rate (Figure 1(D), Fisher's exact test, p-value > 0.05). Meanwhile, mNGS-based genomic data also showed that P1-2 has 9.10%, 50% and 75% A2063G mutation rate in 23S rRNA before, during and after the COVID-19 pandemic. Notably, there was a significant increase in the A2063G mutation rate in P1-2 when the COVID-19 pandemic began (Figure 1(D), Fisher exact test, p-value = 0.01).

Variations in the prevalence and relative abundances of M. pneumoniae P1 genotypes

Our analysis revealed that P1-1 was the dominant genotype (accounting for more than 50%) from 2019 to 2023, with a few exceptions in certain seasons (Figure 2(A)). Additionally, we found a shift in the ratio of P1-1 to P1-2 during the COVID-19 pandemic (Figure 2(B)). Specifically, the ratio of P1-2 became comparable to that of P1-1 during the pandemic period (Fisher's exact test, p-value = 0.001, Before vs. During; p-value < 0.001, During vs. After). To investigate whether the distribution of P1-1 and P1-2 varied across different age groups during the COVID-19 pandemic, we stratified the samples with *M. pneumoniae* genotyping results into four age groups: children (<5y, $n = 204$), school-age (5–17y, $n = 726$), adult (18–59y, $n = 402$), and older-people (≥60, $n = 49$) groups (Table S1). Significant differences in the ratio of P1-1 to P1-2 were observed in the school-age group (Figure 2(B), Fisher's exact test, p-value < 0.001, school-age, Before vs During/During vs After). But no significant differences were found in children, adults and older people (Figure 2(B), Fisher's exact test, p-value > 0.05). Based on the prevalence of 23s rRNA A2063G mutation and clinical uses of macrolide, we hypothesize that P1-1 may be an earlier strain in East Asia and more adapted to the host immune system. To confirm this hypothesis, we examined the relative abundance of P1-1 and P1-2 across age groups. We found significant differences in *M. pneumoniae* reads per millions (RPM) between P1-1 and P1-2 in children and school-age individuals, who are likely to have lower immunity against *M. pneumoniae* (Figure 2(C), Mann–Whitney test, p-value = 0.001, 0.03). In contrast, there was less variation in the relative abundance between the two P1 genotypes in adults and older-people (Figure 2(C), Mann–Whitney test, p-value > 0.05).

Respiratory microbial diversity of different M. pneumoniae P1 genotypes across COVID-19 pandemic

Previous studies have shown that the lower respiratory tract (LRT) microbiota in pneumonia patients infected with *M. pneumoniae* can vary depending on disease severity. Based on this, we hypothesized that if P1-1 exhibits higher reproductive capacity in individuals with lower immune responses, we might observe disturbances or differences in the respiratory microbiota between P1-1 and P1-2. To explore this, we first compared microbial community differences (both α- and β-diversity) between P1-1 and P1-2 within specific age groups and time periods (Figure S2). No significant differences of α and β diversity between P1-1 and P1-2 in children, adults and older-people (Figure

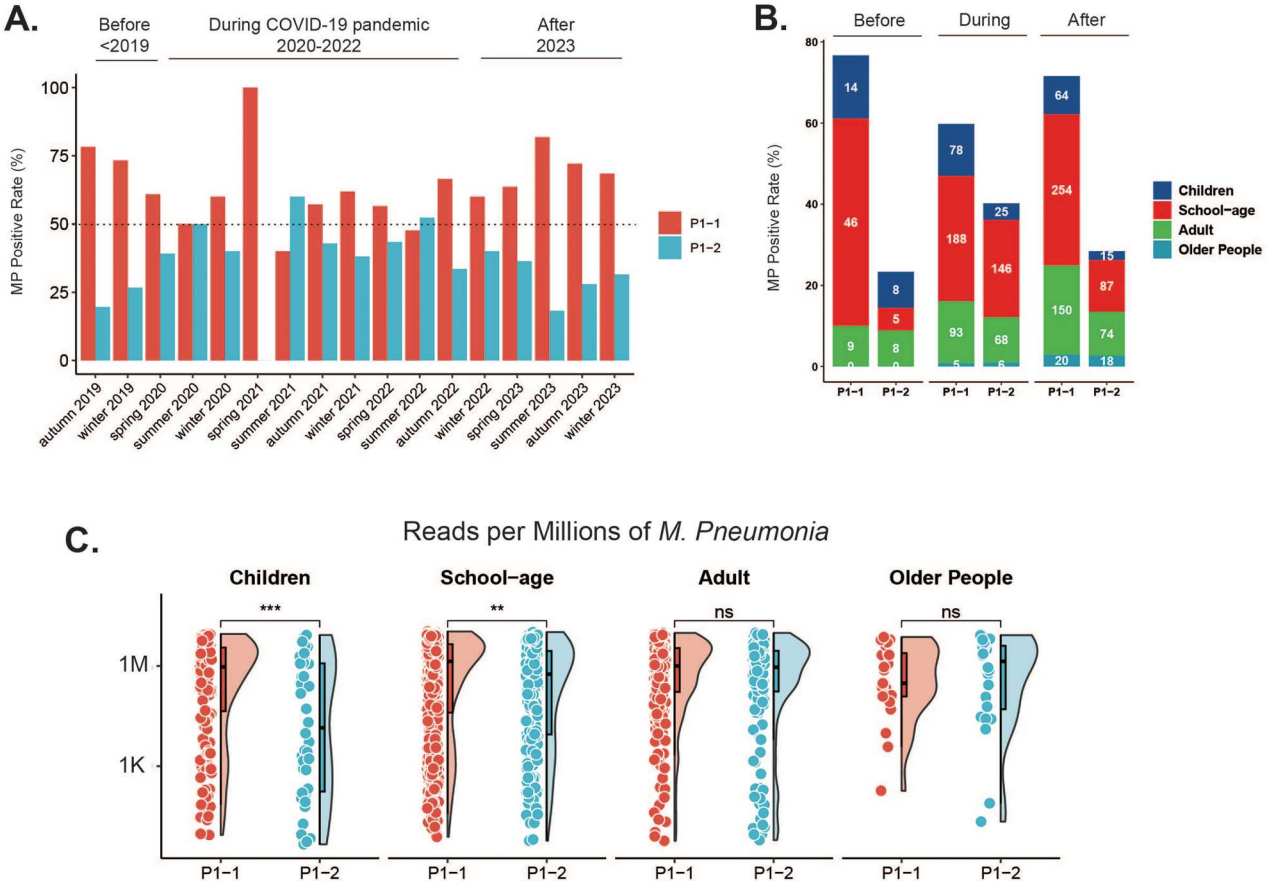


Figure 2. Genotype prevalence of *M. pneumoniae* and relative abundances of P1-1 and P1-2 among different periods and different ages. (A) Positive rate of *M. pneumoniae* P1 genotypes (P1-1 and P1-2) before, during, and after COVID-19 pandemic. (A) Positive rate of *M. pneumoniae* P1 genotypes (P1-1 and P1-2) of children, school-age, adults and older-people among different periods. (C) Reads per millions between two genotypes (P1-1 and P1-2) of *M. pneumoniae* across different age groups (children, school-age, adults and older-people). *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. "NS" represents "no significant difference."

S2B–E, p-value > 0.05). However, we found that the diversity of the microbial community, as measured by the Shannon index, was significantly lower in P1-1 compared to P1-2 after the COVID-19 pandemic (Figure S2A, Mann–Whitney test, p-value = 0.02). Additionally, we found that Shannon (Figure S2B, Mann–Whitney test, p-value = 0.001) and Simpson (Figure S3J, Mann–Whitney test, p-value = 0.001) indices in the P1-1 was lower than in P1-2 among school-age individuals. The PCoA based on Bray–Curtis distance calculating in species level, confirmed by the PERMANOVA test, revealed a significant difference in β-diversity between P1-1 and P1-2 among school-age (p-value = 0.027, Figure S2D). However, there was no significant differences of β-diversity in other groups when comparing P1-1 and P1-2 (p-value > 0.05, Figure S2C–E, Figure S4).

We also compared the respiratory microbiota composition between the two groups at the bacterial species level. LEfSe was performed on the bacterial taxa that showed statistically significant differences using multivariate analysis, confirmed by the Kruskal–Wallis sum-rank test. The results, ranked by LDA score histogram, indicated that *M. pneumoniae* was significantly enriched in P1-1 after the COVID-19 pandemic and among school-age individuals (Figure S5A,B). Besides, *Streptococcus* spp. were the most abundant genera among adults (Figure S5B).

Pathogen spectrum and co-detection pattern of M. pneumoniae *P1 genotypes in different age groups*

In addition to exploring changes in microbial structure between P1-1 and P1-2, we were also interested in the co-detection of commensal microbes potentially involved in *M. pneumoniae* pneumonia. Given the important role of commensal microbiota and variations in *M. pneumoniae* P1 genotypes, we first summarized the pathogen spectrum based on metagenomic sequencing data (Figure 3). Our analysis revealed that *E. coli* (15.56%, 59.11%, 91.94%), *P. aeruginosa* (33.33%, 72.09%, 65.98%) and *S. pneumoniae* (30%, 44.01%, 45.16%) were top three bacteria co-detected in *M. pneumoniae* pneumonia patients, irrespective of time (before, during, or after the COVID-19 pandemic) (Figure 3(A)). Notably, *E. coli* became nearly dominant in *M. pneumoniae* pneumonia patients after the COVID-19 pandemic (91.94%, Figure 3(A)). Additionally, *Aspergillus*

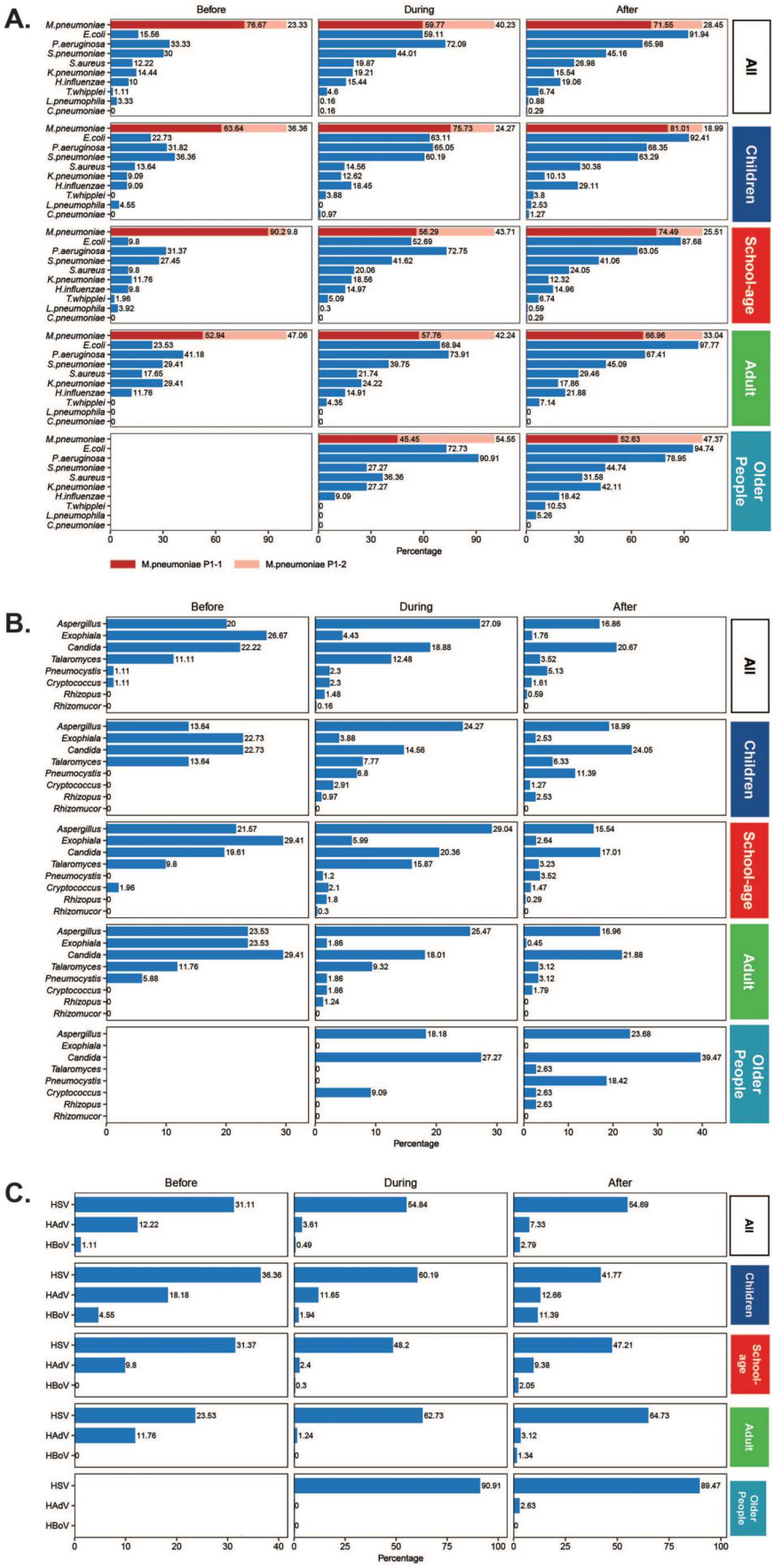


Figure 3. Changes in the composition of common respiratory pathogens in patients testing positive for *M. pneumoniae* before, during, and after COVID-19 pandemic. I, Detection rates of bacterial (a) and viral (b) pathogens. HAdv: human adenovirus, RSV: respiratory syncytial virus, HMPV: human metapneumovirus, HCoV: human coronavirus, IFV: influenza virus, HBoV: human bocavirus. II, Co-infection patterns in school-age group before (a), during (b), and after (c) COVID-19 pandemic. Bigger size and darker colour of the circles indicate higher co-infection rates between two pathogens.

(20%, 27.09%, 16.86%) and *Candida* (22.22%, 18.88%, 20.67%) the most frequently co-detected fungi, while HSV (31.11%, 54.84%, 54.89%) was the leading DNA virus (Figure 3(B,C)). Co-detection pattern was disrupted by COVID-19 pandemic in different ages. Specifically, the co-detection rates of *E. coli* and *P. aeruginosa* varied among children and school-age individuals (Figure 3(A), Figure S6A,B, Figure S7A, B). However, no significant changes were observed in the co-detection of fungi and DNA viruses across children, school-age individuals, adults, and older people (Figure 3(B,C), Figure S6, Figure S7).

To further investigate the hypothesis that P1-1 may exhibit higher reproductive capacity in individuals with lower immune responses compared to P1-2, we compared co-detection ratios of opportunistic pathogens with P1-1 and P1-2 across different age groups. It showed that P1-1 had a higher odds ratio of co-detecting opportunistic pathogen (Figure 4(A)). We found HAdV, *S. pneumoniae*, *Pneumocystis* (except when compared with older-people, OR = 1.04, p-value = 1) were significantly more likely to be co-detected in children infected with P1-1, compared to other age groups (Figure 4(A), Fisher's exact test, p-value < 0.05). Conversely, no significant differences were found between children and other age groups in P1-2 infections (Figure 4(B), Fisher's exact test, p-value > 0.05). Additional co-detection results for other microbes corroborated these findings (Figure S8A,B). To further validate these results, we compared the odds ratio of co-detection between P1-1 and P1-2 across different ages. We found that P1-1 was associated with a higher co-detection rate of several opportunistic pathogens in children, school-age individuals, and adults compared to P1-2 (Figure 4(C)). Such as *S. pneumoniae*, a key respiratory pathogen for children and school-age individuals, exhibited a 7.10, 2.79, and 2.61-fold increase in co-detection rates with P1-1 compared to P1-2 in children, school-age, and adult groups, respectively (Figure 4(C), Fisher's exact test, p-value < 0.001). Furthermore, the pattern of *S. pneumoniae* odds ratios closely mirrored the relative abundance of *M. pneumoniae* across different age groups (Figure 2(C), Figure 4(C)).

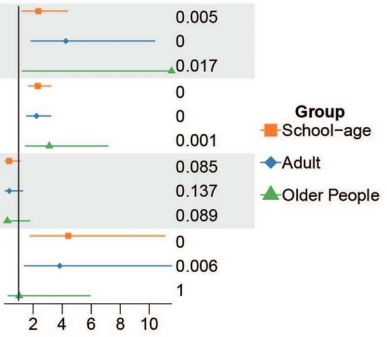
Clinical outcome and host immune responses of two M. pneumoniae *P1 genotypes*

To further investigate whether the P1-1 not only leads to co-detection of more opportunistic pathogens but also results in more severe clinical symptoms, we selected 118 patients from a cohort of 1368 individuals who had more comprehensive clinical outcome data (Figure S1A, Figure 5). Among these 118 patients, 61.9% (73/118) were male and 38.1% (45/118) were female, with no significant difference in gender distribution between the P1-1 and P1-2 groups (Fisher's exact test, p-value = 0.108). The age distribution of this cohort was primarily children and adolescents, with a median age of 8.00 years [4.00, 29.5] and there was no significant difference in age between two genotypes (Mann–Whitney test, p-value = 0.246). We analysed several clinical indicators related to disease severity: hospitalization stay (days), ICU admission, the need for any respiratory support, and mechanical ventilation. Interestingly, we found no significant differences in hospitalization stay or ICU admission rates between the P1-1 and P1-2 (Fisher's exact test, p-value = 0.564/0.057). However, consistent with the co-detection pattern of opportunistic pathogens, the P1-1 showed significantly higher rates of requiring respiratory support and mechanical ventilation compared to P1-2 (Fisher's exact test, p-value = 0.035/0.004). These clinical severity indicators suggest that the P1-1 has indeed higher reproductivity in lower immune response and is more likely to cause severe respiratory symptoms in the East Asian population.

To investigate potential differences in host immune response between two P1 genotypes of *M. pneumoniae*, we selected 174 BALF samples with paired metagenomic (DNA) and metatranscriptomic (RNA) data for host transcriptome analyses (Figure S1A, P1-1: n = 108, P1-2: n = 66). We identified 661 differentially expressed genes between the two P1 genotypes (|Fold change| ≥ 1.5, FDR < 0.25, Figure S9A) and found enrichment in 25 pathways via GSEA (adjusted p-value < 0.05, Figure S9B). First, to assess whether immune cell abundance varied between different clinical outcome groups, we estimated cell-type abundance from the host transcriptome with CIBERSORTx [34]. We found that activated NK cells were the only immune cell type showing a significant difference between the two genotypes (Figure S9G, Mann–Whitney test, p-value = 0.031). To further explore immune cell characteristics, we constructed GSEA gene sets based on the LM22 immune cell-specific genes. We found that specific genes of monocytes and activated NK cells were enriched in the BAL transcriptome from the P1-1 (Figure 6(C,D), adjusted p-value < 0.01), while specific genes of macrophage-M1 showed no preference between the two P1 genotypes. Interestingly, we noted that SFTPD, which encodes surfactant protein-D reflecting the severity of lung injury [36,37], and ULBP2, a mediator of NK cell-mediated immune recognition [38], were higher in the P1-1 BAL transcriptome (Figure 6(A,H,I), Mann–Whitney test, p-value < 0.001, <0.01). Furthermore, the lower airway transcriptomes of P1-1 showed upregulation of the phagosome signalling pathway, which regulates the production of inflammatory mediators during lung injury. Dysregulation of this process can lead to excessive inflammation, contributing to lung injury and diseases such as ARDS [39] (Figure 6(E), adjusted p-value < 0.001). The proteasome pathway, which contributes to inflammation in

A.

Pathogen pairs	Co-detection rate in Children(%)	Co-detection rate in Others(%)	OR(95% CI)	Compare outcome	P-value
M.pneumoniae-P1-1 HAdV	9.8	4.4	2.36 (1.25, 4.37)		0.005
		2.49	4.25 (1.85, 10.38)		0
		0	Inf (1.25, Inf)		0.017
M.pneumoniae-P1-1 S.pneumoniae	47.55	28.06	2.32 (1.67, 3.24)		0
		28.86	2.23 (1.55, 3.22)		0
		22.45	3.12 (1.46, 7.16)		0.001
M.pneumoniae-P1-1 T.whipplei	1.47	4.13	0.35 (0.07, 1.13)		0.085
		3.98	0.36 (0.07, 1.28)		0.137
		6.12	0.23 (0.03, 1.78)		0.089
M.pneumoniae-P1-1 Pneumocystis	6.37	1.51	4.42 (1.80, 11.09)		0
		1.74	3.83 (1.39, 11.53)		0.006
		6.12	1.04 (0.27, 5.94)		1



B.

Pathogen pairs	Co-detection rate in Children(%)	Co-detection rate in Others(%)	OR(95% CI)	Compare outcome	P-value
M.pneumoniae-P1-2 HAdV	2.94	1.79	1.66 (0.51, 4.76)		0.276
		0.25	12.11 (1.45, 558.70)		0.007
		2.04	1.45 (0.17, 68.26)		1
M.pneumoniae-P1-2 S.pneumoniae	11.27	12.24	0.91 (0.53, 1.50)		0.808
		13.43	0.82 (0.46, 1.41)		0.519
		18.37	0.57 (0.23, 1.50)		0.229
M.pneumoniae-P1-2 T.whipplei	1.96	1.51	1.30 (0.30, 4.45)		0.752
		1.74	1.13 (0.24, 4.50)		1
		2.04	0.96 (0.09, 48.25)		1
M.pneumoniae-P1-2 Pneumocystis	1.47	0.69	2.15 (0.33, 11.18)		0.383
		1	1.48 (0.22, 8.86)		0.693
		8.16	0.17 (0.02, 1.04)		0.028

C.

Pathogen	Co-detection rate with M.pneumoniae P1-1(%)	Co-detection rate with M.pneumoniae P1-2(%)	OR(95% CI)	Compare outcome	P-value
HAdV	9.8	2.94	3.58 (1.35, 11.13)		0.007
	4.4	1.79	2.53 (1.28, 5.29)		0.006
	2.49	0.25	10.21 (1.44, 444.13)		0.011
	0	2.04	0.00 (0.00, 39.00)		1
S.pneumoniae	47.55	11.27	7.10 (4.17, 12.47)		0
	28.06	12.24	2.79 (2.11, 3.72)		0
	28.86	13.43	2.61 (1.80, 3.82)		0
	22.45	18.37	1.28 (0.43, 3.94)		0.803
T.whipplei	1.47	1.96	0.75 (0.11, 4.48)		1
	4.13	1.51	2.80 (1.35, 6.24)		0.004
	3.98	1.74	2.34 (0.90, 6.79)		0.088
	6.12	2.04	3.10 (0.24, 167.55)		0.617
Pneumocystis	6.37	1.47	4.54 (1.22, 25.25)		0.019
	1.51	0.69	2.22 (0.71, 8.18)		0.208
	1.74	1	1.76 (0.44, 8.27)		0.546
	6.12	8.16	0.74 (0.10, 4.62)		1
Aspergillus	13.73	7.35	2.00 (0.99, 4.17)		0.052
	14.17	7.98	1.90 (1.34, 2.72)		0
	13.18	7.46	1.88 (1.15, 3.13)		0.01
	14.29	8.16	1.86 (0.44, 9.32)		0.524
Candida	14.22	4.9	3.21 (1.47, 7.60)		0.002
	11.28	7.43	1.58 (1.09, 2.32)		0.015
	13.68	6.97	2.12 (1.29, 3.55)		0.002
	28.57	8.16	4.43 (1.25, 20.14)		0.017
S.aureus	16.18	4.41	4.17 (1.88, 10.19)		0
	14.03	7.15	2.12 (1.47, 3.07)		0
	16.67	9.2	1.97 (1.26, 3.12)		0.002
	14.29	18.37	0.74 (0.21, 2.49)		0.785

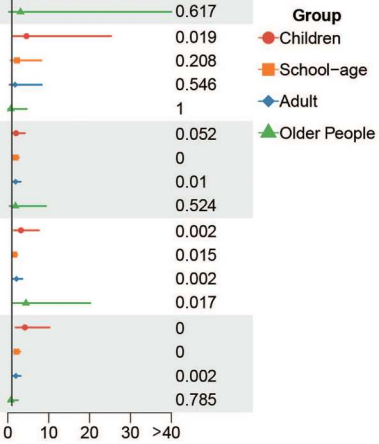


Figure 4. Odds ratio of co-detection rate between P1-1 and P1-2 among different age groups. (A) P1-1 of *M. pneumoniae* co-detection rate with important respiratory pathogen and odds ratio of comparing children to school-age, adults and older-people. (B) P1-2 of *M. pneumoniae* co-detection rate with important respiratory pathogen and odds ratio of comparing children to school-age, adults and older-people. (C) Odds ratio of co-detection rate between P1-1 and P1-2 among children, school-age, adults and older-people. HAdv: human adenovirus; HSV: human herpes simplex virus; HBoV human bocavirus.

Outcomes	ALL	P1−1	P1−2	Odds ratio (95% CI)	p−value
	N=118	N=61	N=57		
Gender:				0.50 [0.22;1.14]	0.108
Female	45 (38.1%)	28 (45.9%)	17 (29.8%)		
Male	73 (61.9%)	33 (54.1%)	40 (70.2%)		
Age	8.00 [4.00;29.5]	9.00 [4.00;45.0]	8.00 [3.00;15.0]	0.99 [0.98;1.01]	0.246
hospitalization stay, days	13.0 [10.0;23.8]	14.0 [10.0;26.0]	12.0 [10.0;19.0]	0.99 [0.97;1.01]	0.564
ICU admission:				2.16 [0.95;5.05]	0.057
N	74 (62.7%)	33 (54.1%)	41 (71.9%)		
Y	44 (37.3%)	28 (45.9%)	16 (28.1%)		
Any respiratory support:				2.58 [0.99;7.19]	0.035
N	89 (75.4%)	41 (67.2%)	48 (84.2%)		
Y	29 (24.6%)	20 (32.8%)	9 (15.8%)		
Mechanical ventilation:				5.05 [1.50;22.2]	0.004
N	97 (82.2%)	44 (72.1%)	53 (93.0%)		
Y	21 (17.8%)	17 (27.9%)	4 (7.02%)		

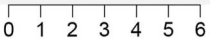


Figure 5. Odds ratio of clinical outcomes between P1-1 and P1-2.

acute lung injury by proteolytically cleaving proteins, leading to decreased anti-elastase activity and potentially predisposing individuals to excessive inflammation [40], was also upregulated in P1-1 (Figure 6(F)). In contrast, the transcriptomics of P1-2 showed upregulation of the calcium signalling pathway

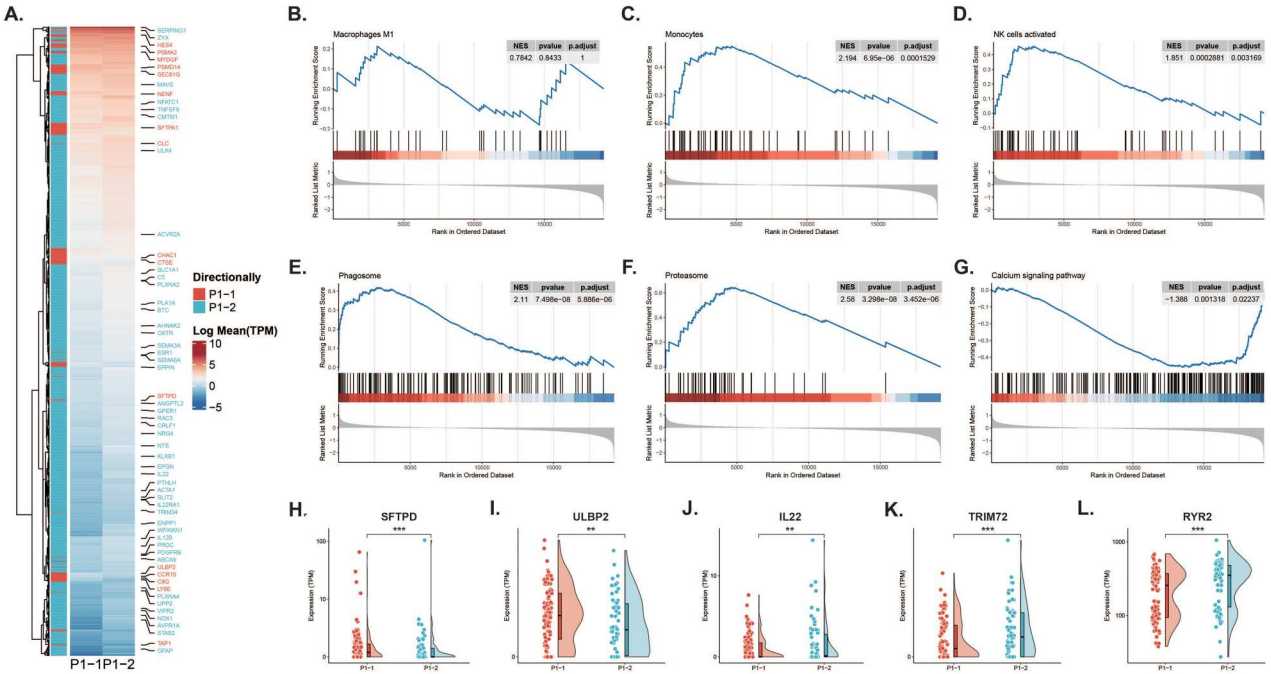


Figure 6. Host immune profiling between P1-1 and P1-2 BAL. (A) Heatmap of 62 significantly differential expressed genes between P1-1 and P1-2 involved in Interferon stimulated pathway, immune-related pathway, phagosome, proteasome and calcium signalling pathway. (B–G) GSEA of Macrophage M1, Monocytes, NK cell activated, phagosome, proteasome and calcium signalling pathway. Each line represents one particular gene set with unique colour, and up-regulated genes are located in the left approaching the origin of the coordinates, by contrast the down-regulated lay on the right of *x*-axis. Only gene sets with NOM *p*-value < 0.05 and FDR < 0.05 were considered significant. And only several leading gene sets were displayed in the plot. (H–L) Feature value of two P1 genotypes. Colour distinctions represent various groups associated with P1 type. The median is visually depicted by black lines. The *y*-axis in each panel was trimmed at the maximum value among all groups of 1.5*IQR above the third quartile, where IQR is the interquartile range. For each gene, we conducted formal comparisons among groups within the training cohort. Pairwise comparisons were performed with a Mann–Whitney test.

(Figure 6(G), adjusted p-value < 0.001). Evidence supports that TRIM72 (MG53) activates NFκB signalling, which is linked to an increase in intracellular calcium oscillation mediated by RYR2 [41] (ryanodine receptor, Figure 6(A,K,L), Mann–Whitney test, p-value < 0.01). Additionally, cilium-dependent cell motility and cilium movement were also upregulated in P1-2 (Figure S9C, p-value < 0.001). Lastly, IL22 and IL22RA1 [35,42] were down-regulated in the P1-1 genotype BALF transcriptome compared to P1-2 (Figure 6(A,J) and Figure S9A, Mann–Whitney test, p-value < 0.01). These findings suggest that at the host gene expression level, P1-1 triggers a more severe immune response in the respiratory tract, exacerbating respiratory symptoms and facilitating the colonization and infection of other opportunistic pathogens.

Discussion

M. pneumoniae is a significant respiratory pathogen in children, yet genomic surveillance and respiratory microbiota studies, particularly those comparing different P1 genotypes, have been limited [43,44]. To the best of our knowledge, this is the first study to reveal differences in the relative abundance, diversity of respiratory microbiota, and co-detection rates of opportunistic pathogens between the P1-1 and P1-2 genotypes of *M. pneumoniae*. Additionally, we provide a comprehensive overview of the impact of the COVID-19 pandemic on the epidemiological features of *M. pneumoniae*, offering insight into the pathogen's circulation in the context of NPIs. One major question in the field remains whether the clinical presentations of LRTI in children differ between the two *M. pneumoniae* P1 genotypes [45]. Some studies suggest that the P1 genotypes exhibit distinct pathogenic potentials, with infections caused by P1-2 potentially leading to more severe illness compared to those caused by P1-1 in children [46]. Conversely, other research suggests the opposite trend [47]. In alignment with findings from an East Asian study [47], our results indicate that P1-1 is associated with more severe clinical outcomes in this cohort. These findings are crucial for guiding future epidemic prevention and treatment strategies in East Asia.

M. pneumoniae is one of the leading bacterial pathogens found in children with pneumonia, responsible for 10-40% of community-acquired pneumonia cases [1–5,44]. In China, two *M. pneumoniae* pandemics occurred between 2019 and 2023. The positivity rate was high in 2019 (6.69% in summer, 8.19% in autumn, 5.04% in winter), despite fewer samples compared to later years. During the COVID-19 pandemic, the positivity rate dropped significantly (0.16% in autumn 2020, the lowest point; 4.06% in autumn 2022), but it rebounded in the autumn and winter of 2023 (6.80%/9.62%, Figure 1(B)). These trends are consistent with multiple studies indicating a global reduction in respiratory and gastrointestinal pathogen infections following the implementation of NPIs [43]. There was a notable decline in the incidence of invasive bacterial diseases caused by *S. pneumoniae*, *H. influenzae*, and *N. meningitidis*, which are spread through respiratory pathways, during the early months of the COVID-19 pandemic, likely due to disruptions in person-to-person transmission. These infection rates remained low even after schools reopened, with the exception of rhinovirus [48]. The COVID-19 pandemic, caused by the SARS-CoV-2 virus, coincided with the second *M. pneumoniae* pandemic, which saw an increase in macrolide resistance after the relaxation of China's "zero COVID" policy [6]. While the clinical outcomes associated with different *M. pneumoniae* genotypes have been well-documented [46,47], few studies have explored the changes in *M. pneumoniae* prevalence and genotypes among individuals with severe pneumonia before, during, and after the COVID-19 pandemic. Our findings highlight the increasing dominance of P1-1, aligning with reports from Japan, where P1-2 was predominant from 2017 (Figure 1(C)).

There was evidence that *M. pneumoniae* P1-1 has a higher growth rate than type 2 *in Vitro,* a finding that aligns with our results, which show a higher relative abundance of P1-1 in children and school-aged individuals (Figure 2(C)). Furthermore, previous studies have noted differences in P1 protein abundance between genotypes. The P1 protein is essential for adhesion to human respiratory epithelial cells and contributes to *M. pneumoniae* pathogenicity. The immunodominant regions of the P1 protein are located along its full length, with both the N-terminal and C-terminal regions exposed on the surface. Antibodies targeting these terminal regions can significantly hinder adhesion [49,50]. We suggested that there were structural differences between P1-1 and P1-2 may contribute to immune escape mechanisms that affect pathogenicity.

Recent research has shown that the respiratory microbiome can predict outcomes in various lung diseases [48,51]. Our study observed a reduction in alpha diversity, which measures the variety and distribution of taxa, in P1-1 compared to P1-2 (Figure S2A–E, Figure S4). Additionally, we discovered that the abundance of *M. pneumoniae* were significant different in child and school-age between two genotypes (Figure 2(C)). The initial detection of *M. pneumoniae* by innate immune cells and their subsequent activation is believed to be a primary cause of chronic inflammation and key symptoms and signs of *M. pneumoniae* infection [16]. The exposure level of *M. pneumoniae* to immune cells is crucial in determining the outcomes of the immune response. *M. pneumoniae*'s pathogenic mechanisms also involve direct harm to respiratory epithelia and immune cells, including damage from

adhesion, invasion, toxins, inflammation, and immune responses [11,12,44,46]. The mucus layer, which traps debris and pathogens such as *S. pneumoniae*, plays a vital role in protecting the respiratory tract by facilitating the expulsion of trapped bacteria via mucociliary action [51]. Studies on primary ciliary dyskinesia, a condition that impairs ciliary motion, show that individuals with this condition have higher rates of respiratory infections, including *S. pneumoniae* [52]. Our BAL transcriptome GSEA results indicated that genes of cilium motility or movement were suppressed in P1-1 (Figure S9C). This provides a comprehensive evidence chain explaining why, in our study, the co-detection rate between P1-1 and *S. pneumoniae* was higher than that of P1-2 across children, school-aged individuals, and adults (Figure 6(C)). Children exhibited the highest odds ratio In particular.

Our BAL transcriptome analysis also aligns with trends observed in COVID-19-related studies: both the P1-1 vs. P1-2 comparison and the <28 days vs. ≥28 days decreased comparison show a tendency for more differential genes in milder cases (<28 days/P1-2). We found that the severity-related genes SFTPD and ULBP2 are significantly upregulated in the P1-1. Surfactant Protein D (SP-D) has potential as both a biomarker and a therapeutic agent for lung injuries and diseases such as COPD, influenza A, COVID-19, and acute lung injury (ALI) [36,37]. Studies show SP-D inhibits lipid-laden macrophages in COPD, reduces influenza A virus replication and inflammation, and prevents SARS-CoV-2 replication by binding to its spike protein, thus blocking cell entry and reducing pro-inflammatory mediators. Additionally, SP-D mitigates inflammation in various lung injury models, supporting its potential use as a treatment for lung-related conditions [53]. ULBP2 plays a crucial role in mediating NK cell activity and is significantly associated with poor prognosis in various tumours [38,54]. Evidence from the BAL transcriptome, combined with clinical data on respiratory support and mechanical ventilation, explains the potential mechanisms of respiratory invasion in P1-1 patients: P1-1 *M. pneumoniae* leads to primary ciliary dyskinesia, more inflammation accumulation, and down-regulation of anti-inflammatory systems, providing space for opportunistic pathogens to proliferate. This exacerbates the patient's condition, resulting in more severe respiratory symptoms.

This study has several limitations. First, we were unable to analyse the phenotypic or genomic characteristics of *M. pneumoniae* due to the absence of positive culture results. Future research should address this limitation by incorporating genomic and phenotypic data. Second, incorporating mouse models in future studies could help further validate and enhance our findings.

In conclusion, our study provides preliminary evidence that P1-1 is more invasive than P1-2 and that this genotype may play a more significant role in exacerbating respiratory disease. This difference in pathogenicity warrants closer monitoring of vaccine efficacy and the emergence of neutralization escape variants in the future.

Ethics approval and consent to participate

This study was approved by the ethics committee of the central institutional review board in The First Hospital of Jilin University (No.2024-415). Written informed consent were provided by all enrolees or surrogates.

Acknowledgements

Y. Jiang, H. You and B. Yang conceived the project and acquired the grant funds; H. You, F. Yu, N. Lin, Y. Liu, H. Zou, S. Hao and Y. Xiao performed the experiments and acquired the data; B. Yang, H. Liu, F. Yu, N. Lin, W. Yang and W. Wu designed the protocol and conducted the analysis; B. Yang, N. Lin, B. Hu and T. Xu drafted and revised the manuscript; Y. Jiang, H. You and B. Yang had final responsibility for the decision to submit for publication. All authors reviewed the draft and approved the decision to submit it for publication.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program of China [grant numbers 2021YFC2701800, 2021YFC2701803]; the National Natural Science Foundation of China [grant number 22174137]; Jilin Province Science and Technology Development [grant number 2023C013]; Jilin Province Science and Technology Agency [grant numbers JJKH20211210KJ, JJKH20211164KJ, JLSWSRCZX2020-009, 20200901025SF and 20200403084SF]; Beijing Medical Award Foundation [grant number YXJL-2021-1097-0645]; and "Hongmian Plan" Project of Guangzhou [grant number HMJH-2020-0005].

Data availability statement

The datasets generated and/or analysed during the current study are not publicly available considering the privacy or ethical restrictions but are available from the corresponding author on a reasonable request. Microbial sequencing data are available at NGDC (https://ngdc.cncb.ac.cn/) Accession no. PRJCA024106. Host transcriptomic sequencing data are available at National Library of Medicine-National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/bioproject/) with a BioProject Accession no. PRJNA1144209.

ORCID

Yanfang Jiang http://orcid.org/0000-0002-2478-1521

References

[1] Atkinson TP, Waites KB. *Mycoplasma pneumoniae* infections in childhood. Pediatr Infect Dis J. 2014;33:92–94. doi:10.1097/INF.0000000000000171

[2] Kutty PK, Jain S, Taylor TH, et al. *Mycoplasma pneumoniae* among children hospitalized with community-acquired pneumonia. Clin Infect Dis. 2019;68:5–12. doi:10.1093/cid/ciy419

[3] Blasi F. Atypical pathogens and respiratory tract infections. Eur Respir J. 2004;24:171–182. doi:10.1183/09031936.04.00135703

[4] Kannan TR, Hardy RD, Coalson JJ, et al. Fatal outcomes in family transmission of *Mycoplasma pneumoniae*. Clin Infect Dis. 2012;54:225–231. doi:10.1093/cid/cir769

[5] Simmons WL, Daubenspeck JM, Osborne JD, et al. Type 1 and type 2 strains of *Mycoplasma pneumoniae* form different biofilms. Microbiology (Reading). 2013;159:737–747. doi:10.1099/mic.0.064782-0

[6] Conroy G. What's behind China's mysterious wave of childhood pneumonia? Nature. 2023. doi:10.1038/d41586-023-03732-w

[7] Waites KB, Talkington DF. *Mycoplasma pneumoniae* and its role as a human pathogen. Clin Microbiol Rev. 2004;17:697–728. doi:10.1128/CMR.17.4.697-728.2004

[8] Wu BG, Segal LN. The lung microbiome and its role in pneumonia. Clin Chest Med. 2018;39:677–689. doi:10.1016/j.ccm.2018.07.003

[9] Esposito S, Principi N. Impact of nasopharyngeal microbiota on the development of respiratory tract diseases. Eur J Clin Microbiol Infect Dis. 2018;37:1–7. doi:10.1007/s10096-017-3076-7

[10] Budden KF, Shukla SD, Rehman SF, et al. Functional effects of the microbiota in chronic respiratory disease. Lancet Respir Med. 2019;7:907–920. doi:10.1016/S2213-2600(18)30510-1

[11] Dai W, Wang H, Zhou Q, et al. The concordance between upper and lower respiratory microbiota in children with *Mycoplasma pneumoniae* pneumonia. Emerg Microbes Infect. 2018;7:92. doi:10.1038/s41426-018-0097-y

[12] Zhou Q, Xie G, Liu Y, et al. Different nasopharynx and oropharynx microbiota imbalance in children with *Mycoplasma pneumoniae* or influenza virus infection. Microb Pathog. 2020;144:104189. doi:10.1016/j.micpath.2020.104189

[13] Lai K, Shen H, Zhou X, et al. Clinical practice guidelines for diagnosis and management of cough—Chinese Thoracic Society (CTS) asthma consortium. J Thorac Dis. 2018;10:6314–6351. doi:10.21037/jtd.2018.09.153

[14] Diao Z, Lai H, Han D, et al. Validation of a metagenomic next-generation sequencing assay for lower respiratory pathogen detection. Microbiol Spectr. 2023;11:e0381222. doi:10.1128/spectrum.03812-22

[15] Ren L, Wang Y, Zhong J, et al. Dynamics of the upper respiratory tract microbiota and its association with mortality in COVID-19. Am J Respir Crit Care Med. 2021;204:1379–1390. doi:10.1164/rccm.202103-0814OC

[16] Sulaiman I, Chung M, Angel L, et al. Microbial signatures in the lower airways of mechanically ventilated COVID-19 patients associated with poor clinical outcome. Nat Microbiol. 2021;6:1245–1258. doi:10.1038/s41564-021-00961-5

[17] Chen S, Zhou Y, Chen Y, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–i890. doi:10.1093/bioinformatics/bty560

[18] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–1760. doi:10.1093/bioinformatics/btp324

[19] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257. doi:10.1186/s13059-019-1891-0

[20] Colquhoun D. An investigation of the false discovery rate and the misinterpretation of *p*-values. R Soc Open Sci. 2014;1:140216. doi:10.1098/rsos.140216

[21] Zhu Y, Luo Y, Li L, et al. Immune response plays a role in *Mycoplasma pneumoniae* pneumonia. Front Immunol. 2023;14:1189647. doi:10.3389/fimmu.2023.1189647

[22] Hu YJ, Satten GA. A rarefaction-without-resampling extension of PERMANOVA for testing presence-absence associations in the microbiome. Bioinformatics. 2022;38:3689–3697. doi:10.1093/bioinformatics/btac399

[23] Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12:R60. doi:10.1186/gb-2011-12-6-r60

[24] van Dijk LR, Walker BJ, Straub TJ, et al. StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities. Genome Biol. 2022;23:74. doi:10.1186/s13059-022-02630-0

[25] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–2069. doi:10.1093/bioinformatics/btu153

[26] Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics. 2015;31:2877–2878. doi:10.1093/bioinformatics/btv271

[27] Kassaye SG, Grossman Z, Vengurlekar P, et al. Insights into HIV-1 transmission dynamics using routinely collected data in the Mid-Atlantic United States. Viruses. 2023;15:68. doi:10.3390/v15010068

[28] Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–W296. doi:10.1093/nar/gkab301

[29] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol. 2014;15:550. doi:10.1186/s13059-014-0550-8

[30] Bu D, Luo H, Huo P, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. 2021;49:W317–w325. doi:10.1093/nar/gkab447

[31] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–15550. doi:10.1073/pnas.0506580102

[32] Kanehisa M, Furumichi M, Sato Y, et al. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51:D587–D592. doi:10.1093/nar/gkac963

[33] Burge S, Kelly E, Lonsdale D, et al. Manual GO annotation of predictive protein signatures: the InterPro approach to GO curation. Database (Oxford). 2012;2012:bar068. doi:10.1093/database/bar068

[34] Steen CB, Liu CL, Alizadeh AA, et al. Profiling cell type abundance and expression in bulk tissues

with CIBERSORTx. Methods Mol Biol. 2020;2117:135–157. doi:10.1007/978-1-0716-0301-7_7
[35] Yan Z, Chen B, Yang Y, et al. Multi-omics analyses of airway host-microbe interactions in chronic obstructive pulmonary disease identify potential therapeutic interventions. Nat Microbiol. 2022;7:1361–1375. doi:10.1038/s41564-022-01196-8
[36] King BA, Kingma PS. Surfactant protein D deficiency increases lung injury during endotoxemia. Am J Respir Cell Mol Biol. 2011;44:709–715. doi:10.1165/rcmb.2009-0436OC
[37] Arroyo R, Kingma PS. Surfactant protein D and bronchopulmonary dysplasia: a new way to approach an old problem. Respir Res. 2021;22:141. doi:10.1186/s12931-021-01738-4
[38] Duan S, Guo W, Xu Z, et al. Natural killer group 2D receptor and its ligands in cancer immune escape. Mol Cancer. 2019;18:29. doi:10.1186/s12943-019-0956-8
[39] Levine B, Mizushima N, Virgin HW. Autophagy in immunity and inflammation. Nature. 2011;469:323–335. doi:10.1038/nature09782
[40] Willemsen N, Arigoni I, Studencka-Turski M, et al. Proteasome dysfunction disrupts adipogenesis and induces inflammation via ATF3. Mol Metab. 2022;62:101518. doi:10.1016/j.molmet.2022.101518
[41] Sermersheim M, Kenney AD, Lin P-H, et al. MG53 suppresses interferon-β and inflammation via regulation of ryanodine receptor-mediated intracellular calcium signaling. Nat Commun. 2020;11:3624. doi:10.1038/s41467-020-17177-6
[42] Whittington HA, Armstrong L, Uppington KM, et al. Interleukin-22: a potential immunomodulatory molecule in the lung. Am J Respir Cell Mol Biol. 2004;31:220–226. doi:10.1165/rcmb.2003-0285OC
[43] Meyer Sauteur PM, Beeton ML, Uldum SA, et al. *Mycoplasma pneumoniae* detections before and during the COVID-19 pandemic: results of a global survey, 2017 to 2021. Euro Surveill. 2022;27:2100746. doi:10.2807/1560-7917.ES.2022.27.19.2100746
[44] Waites KB, Xiao L, Liu Y, et al. *Mycoplasma pneumoniae* from the respiratory tract and beyond. Clin Microbiol Rev. 2017;30:747–809. doi:10.1128/CMR.00114-16
[45] Xu M, Li Y, Shi Y, et al. Molecular epidemiology of *Mycoplasma pneumoniae* pneumonia in children, Wuhan, 2020–2022. BMC Microbiol. 2024;24:23. doi:10.1186/s12866-024-03180-0
[46] Rodman Berlot J, Krivec U, Mrvič T, et al. *Mycoplasma pneumoniae* P1 genotype indicates severity of lower respiratory tract infections in children. J Clin Microbiol. 2021;59:e0022021. doi:10.1128/JCM.00220-21
[47] Fan L, Li D, Zhang L, et al. Pediatric clinical features of *Mycoplasma pneumoniae* infection are associated with bacterial P1 genotype. Exp Ther Med. 2017;14:1892–1898. doi:10.3892/etm.2017.4721
[48] Oster Y, Michael-Gayego A, Rivkin M, et al. Decreased prevalence rate of respiratory pathogens in hospitalized patients during the COVID-19 pandemic: possible role for public health containment measures? Clin Microbiol Infect. 2021;27:811–812. doi:10.1016/j.cmi.2020.12.007
[49] Li S, Sun H, Liu F, et al. Two case reports: whole genome sequencing of two clinical macrolide-resistant *Mycoplasma pneumoniae* isolates with different responses to azithromycin. Medicine (Baltimore). 2016;95:e4963. doi:10.1097/MD.0000000000004963
[50] Chourasia BK, Chaudhry R, Malhotra P. Delineation of immunodominant and cytadherence segment(s) of *Mycoplasma pneumoniae* P1 gene. BMC Microbiol. 2014;14:108. doi:10.1186/1471-2180-14-108
[51] Beentjes D, Shears RK, French N, et al. Mechanistic insights into the impact of air pollution on pneumococcal pathogenesis and transmission. Am J Respir Crit Care Med. 2022;206:1070–1080. doi:10.1164/rccm.202112-2668TR
[52] Ferreira DM, Jambo KC, Gordon SB. Experimental human pneumococcal carriage models for vaccine research. Trends Microbiol. 2011;19:464–470. doi:10.1016/j.tim.2011.06.003
[53] Elmore A, Almuntashiri A, Wang X, et al. Circulating surfactant protein D: a biomarker for acute lung injury? Biomedicines. 2023;11. doi:10.3390/biomedicines11092517
[54] Yang X, Su X, Wang Z, et al. ULBP2 is a biomarker related to prognosis and immunity in colon cancer. Mol Cell Biochem. 2023;478:2207–2219. doi:10.1007/s11010-022-04647-2