narra j

Original Article

SARS-CoV-2 lineages and naso-oropharyngeal bacterial communities in COVID-19 reinfection: A study in West Java, Indonesia

Alvira R. Sativa[1], Isnaini Z. Asyifa[2], Muhammad M. Adzdzakiy[3], Syam B. Iryanto[4], Herjuno A. Nugroho[5], Ari S. Wulandari[5], Nova D. Yanthi[5], Mukh F. Nasrulloh[5], Ema Rahmawati[6], Cut NC. Alamanda[6], Ryan B. Ristandi[6], Rifky W. Rachman[6], Rini Robiani[6], Dian F. Agustiyani[7], Popi H. Wisnuwardhani[7], Andri Wardiana[7], Ratih A. Ningrum[7], Anik B. Dharmayanthi[8], Anggia Prasetyoputri[5], Azzania Fibriani[1] and Sugiyono Saputra[5*]

[1]School of Life Science and Technology, Institut Teknologi Bandung, Bandung, Indonesia; [2]Master Program in Biomedical Science, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia; [3]Graduate Program of Bioscience, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret, Surakarta, Indonesia; [4]Research Center for Computation, National Research and Innovation Agency (BRIN), Bogor, Indonesia; [5]Research Center for Applied Microbiology, National Research and Innovation Agency (BRIN), Bogor, Indonesia; [6]West Java Health Laboratory, Bandung, Bandung, Indonesia; [7]Research Center for Genetic Engineering, Nasional Research and Innovation Agency (BRIN), Bogor, Indonesia; [8]Research Centre for Biosystematics and Evolution, National Research and Innovation Agency Republic of Indonesia (BRIN), Bogor, Indonesia

*Corresponding author: sugiyono.saputra@brin.go.id

Abstract

Continuous emergence of new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants may influence viral transmission dynamics and alter interactions with the respiratory microbiota, potentially increasing the risks of reinfection. This study investigated cases of coronavirus disease 2019 (COVID-19) reinfection in West Java, Indonesia, with the aim of identifying the SARS-CoV-2 variants involved, characterizing their genomic mutations, and profiling the nasal and oropharyngeal microbiota associated with reinfection. Naso-oropharyngeal swab samples were collected from 42 COVID-19 reinfection cases and nine new infection cases. Whole genome sequencing was performed using Oxford Nanopore Technologies (ONT) MinION Mk1C and variant analysis was conducted using ARTIC workflow. Nexstrain and PANGOLIN were used to determine the lineages. Phylogenetic trees were constructed using IQ-tree and FigTree. Key mutations were identified by Cov-GLUE. Additionally, 16s rRNA amplicon sequencing was conducted on nine samples from each group to analyze bacterial communities using EPI2ME and MicrobiomeAnalyst. All identified SARS-CoV-2 strains in this study were Delta variant (B.1.617.2), predominantly lineage AY.23 (n=46, 90%), followed by AY.24 (n=3) and AY.109 (n=2). No differences in SARS-CoV-2 lineages were observed between reinfection and new infection cases. Unique hotspot mutations found only in COVID-19 reinfections included NSP3, V220A, S_T676I, ORF7a_V82A, and ORF7a_TI20I. Bacterial community analysis revealed no significant diversity differences (alpha and beta) between the two groups. While the most dominant phylum remained Terrabacteria in both groups, *Streptococcus* was dominant in COVID-19 reinfections, whereas *Prevotella* was dominant in new infection cases. Notably, *Haemophilus parainfluenzae, Fusobacterium periodonticum, Fusobacterium nucleatum*, and *Leptotrichia buccalis* had significant increases in reinfection cases. Despite the similarity in SARS-CoV-2 lineages causing both COVID-19 reinfection and new infection cases, the presence of distinct key mutations and bacterial species suggest their potential as biomarkers within this group.

Keywords: COVID-19, SARS-CoV-2, lineages, reinfection, bacterial community

https://narraj.org/

Received: August 11, 2025 - Accepted: September 2, 2025 - Published online: October 17, 2025

Introduction

Coronavirus disease 2019 (COVID-19) is a pneumonia-causing illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which spreads through droplets and aerosols [1]. This virus has a positive single-stranded RNA genome (+ssRNA) and at the 5' end, it contains 16 nonstructural proteins (nsPs), while at the 3' end, it encodes four structural proteins: envelope (E), nucleocapsid (N), membrane (M), and spike (S) [2]. The S protein binds the viral envelope to host cells, the N protein forms the nucleocapsid, the M protein contributes to virus assembly during viral replication, while the E protein plays a role in virus production and maturation [3]. To date, the rate of virus evolution continues to increase, resulting in the emergence of several new variants that can impact the virus's transmission and spread [4].

SARS-CoV-2 has led to millions of new infections globally. However, a concerning phenomenon emerges when individuals who have previously recovered become infected again, a situation known as reinfection [5]. The criteria for identifying suspected SARS-CoV-2 reinfection encompass instances where the detection of SARS-CoV-2 RNA (Ct<33) occurs within 90 days after the initial detection, either presenting with symptoms or as asymptomatic cases [6]. Additionally, reinfection is considered when SARS-CoV-2 RNA reappears between 45 and 89 days, accompanied by relevant symptoms or close contact with a confirmed COVID-19 case [6]. Detecting SARS-CoV-2 reinfection currently hinges on molecular assessments of the virus at two distinct time points, alongside analysis of viral genetic sequencing data. This determination is influenced by factors such as higher viral exposure or the presence of two infections with different lineages and mutations within the viral genome [7,8].

In addition to the SARS-CoV-2 variants that can influence the severity of COVID-19, there have been reported cases of coinfection stemming from the identification of bacterial communities in patients. This condition entails the simultaneous occurrence of viral and bacterial infections, thereby exacerbating the disease burden on patients [9]. These microorganisms colonize the nasopharyngeal and oropharyngeal areas and can consequently impact viral infectivity. Despite the extensive documentation of clinical and epidemiological aspects of COVID-19, research concerning the bacterial communities present in COVID-19 patients, particularly in cases of reinfection, remains limited [10]. Prior investigations have indicated that predominant phyla in SARS-CoV-2 positive patients encompass Proteobacteria, Actinobacteria, and Firmicutes, while the most commonly identified genera include *Prevotella*, followed by *Leptotrichia* and *Streptococcus* [12,13].

Next-generation sequencing is widely employed to characterize the entire genome sequences of SARS-CoV-2, along with their associated epidemiological data. This approach facilitates the investigation of viral diversity, transmission pathways, and potential mutations [14]. Additionally, the application of 16S rRNA amplicon sequencing aids in the identification of bacterial communities by assessing the abundance of these communities. These techniques subsequently provide insights into the composition of the respiratory microbiome and its potential link to coinfections that might impact the prognosis of SARS-CoV-2 infections [15]. The aim of this study was to determine the lineages and key mutations of SARS-CoV-2 and profile the bacterial communities in nasopharyngeal and oropharyngeal swab samples collected from COVID-19 reinfected patients in West Java, Indonesia.

Methods

Clinical samples

A total of 51 naso-oropharyngeal swabs were collected from COVID-19 patients who had been tested positive. The samples were divided into reinfection cases (n=42) and newly infected cases (n=9). The sample collection was carried out by the West Java Regional Health Laboratory, Ministry of Health, between March and June 2021, during the second wave of the COVID-19 pandemic in Indonesia.

Whole genome sequencing of SARS-CoV-2

RNA extraction, cDNA synthesis and amplification

SARS-CoV-2 RNA isolation was performed using the Viral Nucleic Acid Extraction Kit II (Geneaid, New Taipei City, Taiwan) following the manufacturer's instructions. All procedures

were conducted at the Biosafety Level 3 Laboratory, National Research and Innovation Agency of Indonesia (BRIN), Cibinong, Indonesia.

Complementary DNA (cDNA) synthesis was achieved using the LunaScript RT SuperMix kit (New England Biolabs, Ipswich, USA), followed by amplification using nCoV-2019/V3 primers. Coverage of the #74 amplicon in the pool B reaction was ensured by the addition of a specific primer pair. PCR amplification was carried out under the following conditions: initial denaturation at 98°C for 30 seconds, followed by 30 cycles of denaturation at 98°C for 15 seconds and annealing/extension at 65°C for 5 minutes. Samples exhibiting distinct and intense DNA bands during electrophoresis were selected for subsequent whole genome sequencing.

Library preparation for whole genome sequencing

The genome library was prepared following the ARTIC nCoV-2019 sequencing protocol V3 [16], using the Oxford Nanopore Technologies (ONT) Ligation Sequencing Kit (Oxford Nanopore Technologies, Oxford, UK). PCR and nuclease-free water (NFW) volumes were adjusted according to Qubit (Thermo Fisher Scientific, Waltham, USA) quantification results to achieve a final concentration of 100 ng/µL. Native barcoding was performed using the ONT Native Barcoding Expansion Kit 1–24 (Oxford Nanopore Technologies, Oxford, UK). Up to 24 samples, which included a negative control, were individually barcoded, pooled, and subjected to solid-phase reversible immobilization (SPRI) bead-based purification (1:1). Following barcoding, adapter ligation was carried out using the Adapter Mix II (AMII) ONT(Oxford Nanopore Technologies, Oxford, UK) and T4 DNA ligase (New England Biolabs, Ipswich, USA). Approximately 15–20 ng of the final library was then loaded onto an ONT MinION Mk1B/Mk1C sequencer for sequencing.

Sequencing and genome analysis

The sequencing run was executed using the MinKNOW software (v20.06.4) (Oxford Nanopore Technologies, Oxford, UK) and was monitored through the RAMPART (v2.1.0) (The Genome Analysis Center (TGAC), Norwich, UK). Basecalling, which involved obtaining FASTQ files, was conducted using Guppy (v4.0.14) (Oxford Nanopore Technologies, Oxford, UK) with HAC mode. Variant calling and the subsequent generation of a consensus SARS-CoV-2 sequence was accomplished using Medaka (v1.0.3) workflows (Oxford Nanopore Technologies, Oxford, UK) in conjunction with bcftools (v1.10.2) (the Samtools project and is primarily associated with the city of Oxford, UK).

To determine the lineages of the 51 SARS-CoV-2 genomes, The Phylogenetic Assignment of Named Global Outbreak Lineages PANGOLIN v3.1.16 (the Centre for Genomic Pathogen Surveillance, Cambridge, UK) and Nexstrain were utilized. The construction of the phylogenetic tree was carried out using IQ-TREE (GNU General Public License), followed by visualization using FigTree (Evomics, Ballwin, USA).

16S rRNA amplicon sequencing

DNA extraction and amplification

The DNA isolation process for the selected samples (reinfection n=9 and newly infected n=9) was conducted using the QIAamp DNA Mini Kit (Qiagen, Aarhus, Denmark). ZymoBIOMICS Microbial Community Standards were employed as positive controls (ZymoResearch, USA). Amplification was carried out using the Q5 Hot Start Master Mix (New England Biolabs, Ipswich, USA), utilizing primers 27F and 1492 R with the following PCR conditions: initial denaturation at 98°C for 5 mins, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1 min, with a final extension at 72°C for 5 mins. Samples displaying a clear DNA band with an amplicon size of approximately 1,500 bp were selected for subsequent library preparation.

Library preparation for 16S rRNA sequencing

Library preparation was performed using the 16S barcoding kit 1–24 (SQK-16S024) (Oxford Nanopore Technologies ONT, Oxford, UK). Each sample with a minimum DNA concentration of 10 ng/µL was processed. The thermal cycler was programmed with the following temperature cycles: initial denaturation at 95°C for 1 min, followed by 25 cycles of denaturation at 95°C for 20 seconds, annealing at 55°C for 30 secs, and extension at 65°C for 2 mins, with a final extension at 65°C for 5 minutes.

Following the amplification, samples were subjected to purification using AMPure XP beads (Beckman Coulter Life Sciences, California, USA). An incubation step at room temperature with gentle mixing using a hula mixer (Thermo Fisher Scientific, Waltham, USA) for 5 mins was performed, after which Qubit quantification was conducted. The Qubit results were then utilized to determine the DNA volume to be pooled from each sample tube. To this pooled microtube, 1 µl of RAP (Rapid Adapter) was added, and the mixture could be directly used or stored at 4°C.

Sequencing and statistical analysis

The 16S rRNA sequencing was conducted using the MinION Mk1b/Mk1c sequencer (Oxford Nanopore Technologies ONT, Oxford, UK, and the sequencing run was executed using the MinKNOW software (v20.06.4) (Oxford Nanopore Technologies ONT, Oxford, UK. Analysis of the obtained genomic data was performed using the EPI2ME labs workflow (Oxford Nanopore Technologies ONT, Oxford, UK. Statistical analysis was subsequently conducted using MicrobiomeAnalyst (microbiomeanalyst.ca).

Results

Variations of lineages from reinfection patients

Analysis of the SARS-CoV-2 variants in this study involved 51 samples of nasopharyngeal and oropharyngeal swabs collected from reinfected patients and new cases who had been tested positive for COVID-19. These samples were collected between March and June 2021. The amplification of the SARS-CoV-2 genome was conducted using nCoV-2019/V3 primers, resulting in a 400 bp amplicon that covered a genome length of 28,999–29,782 bp.

All 42 samples exhibiting re-infection were subjected to PANGOLIN-based phylogenetic classification. As a result, 42 sequenced genomes were categorized as Delta variant B.1.617.2 (India), including lineages AY.23 (n=37, 88%), AY.24 (n=3, 7.14%), and AY.109 (n=2, 4.77%). Conversely, among the nine new infection samples (n=9), all belonged to the AY.23 lineage. The phylogenetic tree encompassing all the viruses was constructed utilizing IQ-TREE, with the visualization presented in **Figure 1**.

Overall, the distribution of lineages did not exhibit significant differences between reinfection and non-reinfection cases, as well as across different genders and age groups. In the reinfection group, the predominant lineage was AY.23, accounting for 96% in females and 82% in males, as compared to lineages AY.24 and AY.109 (**Figure 2A** and **2B**). Similarly, in terms of age categories, the dominance of lineage AY.23 was observed, encompassing over 50% of both the young adult (20–34 years) and adult (35–60 years) samples.

Mutation analysis

The highest number of mutations observed were in the S protein in both reinfection (58/240, 24%) and non-reinfection cases, (63/239, 26%), followed by NSP3 and N (**Figure 3A–3B**). Notably, among the sequenced viruses, a total of 18 carried mutations in the S protein, including T19R, L452R, T478K, D614G, P81R, D950N, V1264L, T676I, and S680F. Moreover, amino acid alterations were also identified in genes not encoding the S protein. These encompassed N, NSP2, NSP3, NSP4, NSP6, NSP12, NSP13, NSP14, ORF3a, M, Orf7a, ORF7b, and N. Distinct hotspot mutations exclusive to reinfection cases comprised NSP3_V220A, S_T676I, ORF7a_V82A, and ORF7a_TI20I (**Figure 4**).

Bacterial diversity

Among the 18 samples used for 16S rRNA sequencing, nine were from reinfected patients and nine from new COVID-19 patients; seven patients were male and 11 were female. In this study, alpha diversity was assessed using the Shannon index and compared across groups with statistical tests (T-test/ANOVA). No significant differences in alpha diversity were observed by type of infection (reinfected vs new COVID-19) (p=0.212), age (young group (20–34 years) vs the adult group (35–60 years) (p=0.853), or gender (p=0.599) (**Figure 5**).

Sativa et al. Narra J 2025; 5 (3): e2901 - http://doi.org/10.52225/narra.v5i3.2901

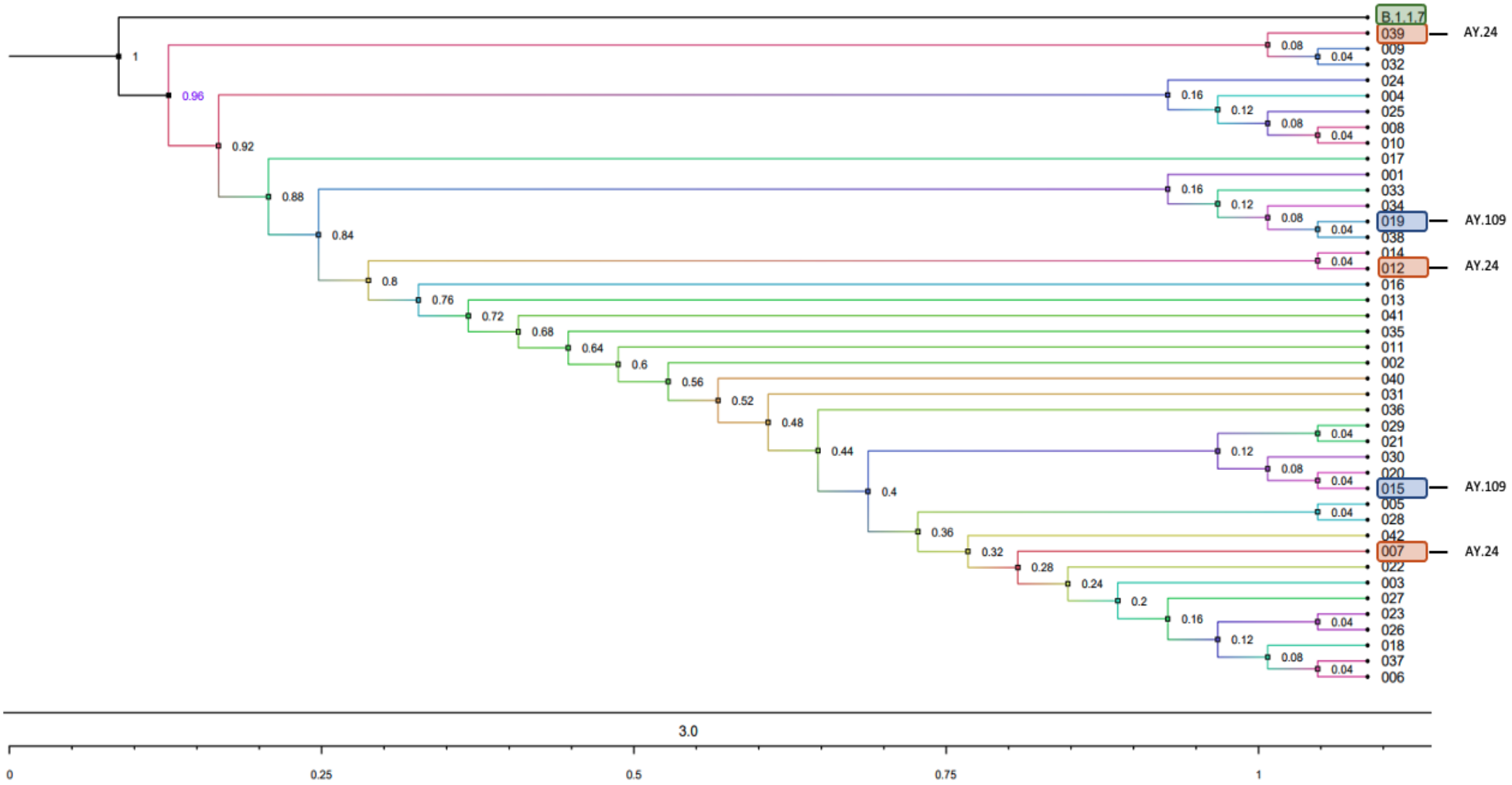

Figure 1. Phylogenetic tree of 42 genomes using Alpha variant as outgroup.

Sativa et al. Narra J 2025; 5 (3): e2901 - http://doi.org/10.52225/narra.v5i3.2901

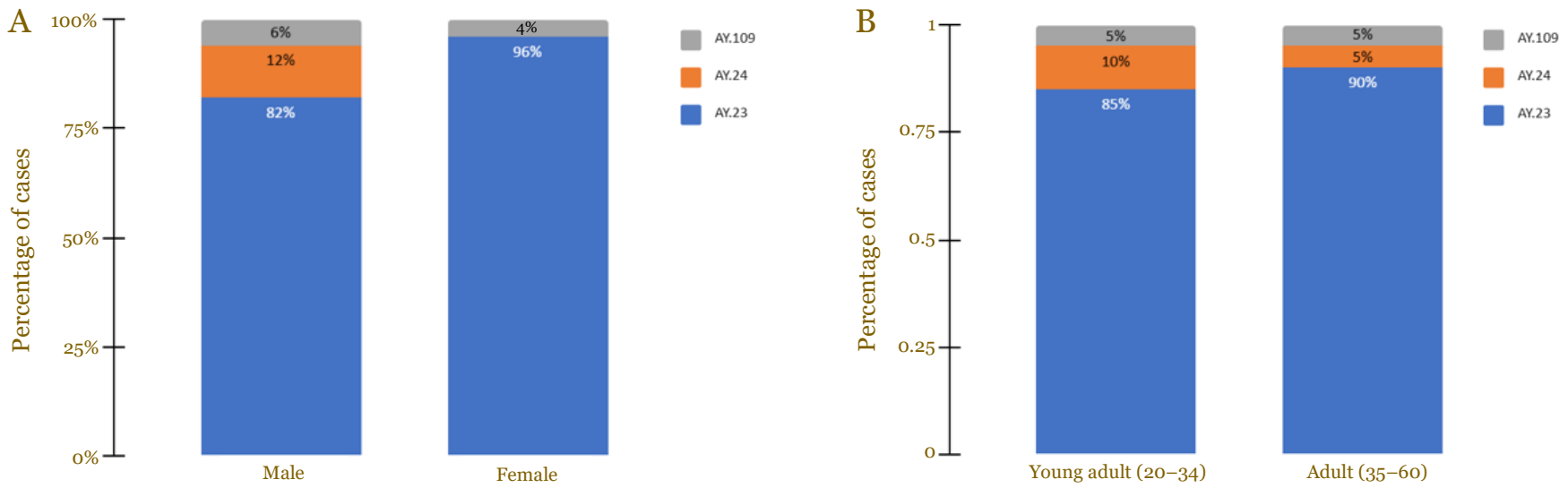

Figure 2. SARS-CoV-2 variants identified by whole genome sequencing (WGS). (A) Patients' distribution by gender and (B) SARS-CoV-2 variant distribution in reinfected and non-reinfected patients by age, showing AY.23 as the predominant lineage across all groups.

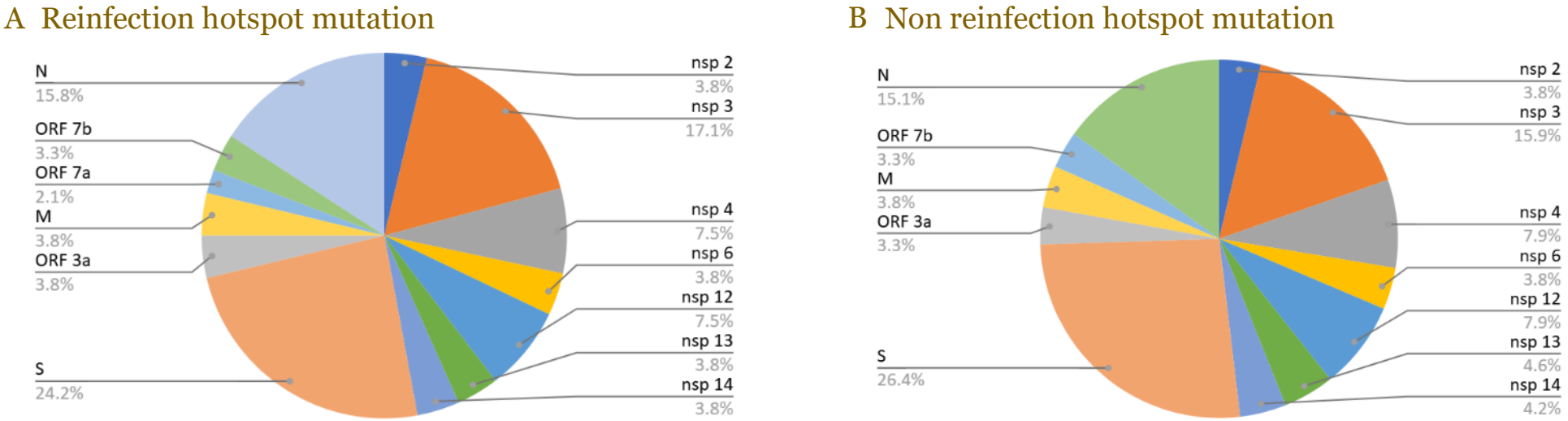

Figure 3. Hotspot mutations in reinfected patients with and new COVID-19 cases. (A) Hotspot mutations among nine reinfected COVID-19 patients. (B) Hotspot mutations among nine new COVID-19 patients.

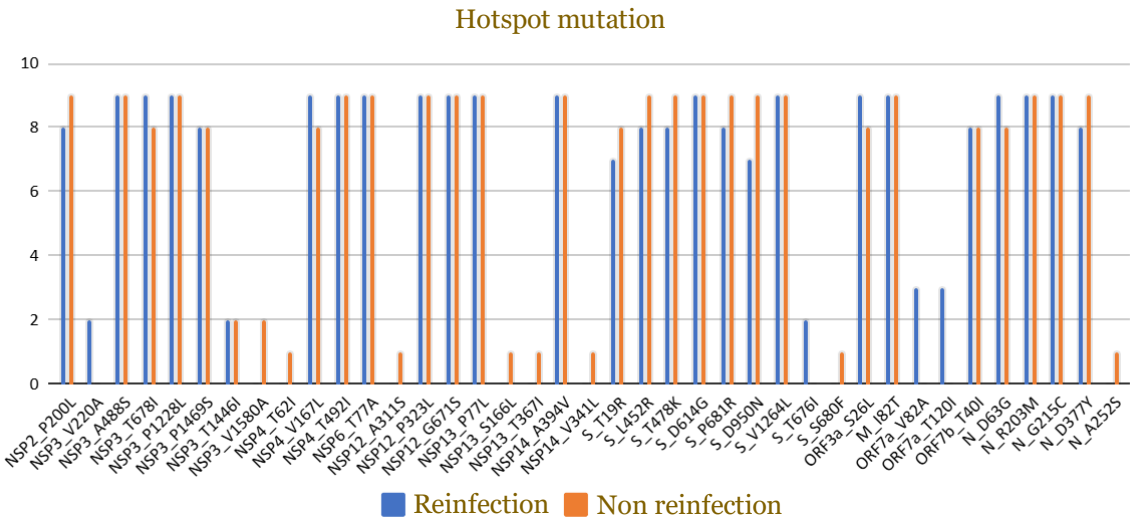

Figure 4. Comparison of amino acid substitutions among samples identified in this study.

Furthermore, beta diversity analysis was performed using the Bray-Curtis index, visualized through Principal Coordinates Analysis (PCoA), and assessed through the PERMANOVA statistical method. Beta diversity evaluation characterizes the species diversity between two communities. The analysis revealed that the beta diversity of the bacterial community in reinfection cases did not differ significantly from that in non-reinfection cases. Likewise, no significant differences were observed when samples were stratified by gender (male vs female) or age group (20–34 years vs 35–60 years) (p>0.05) (**Figure 6**).

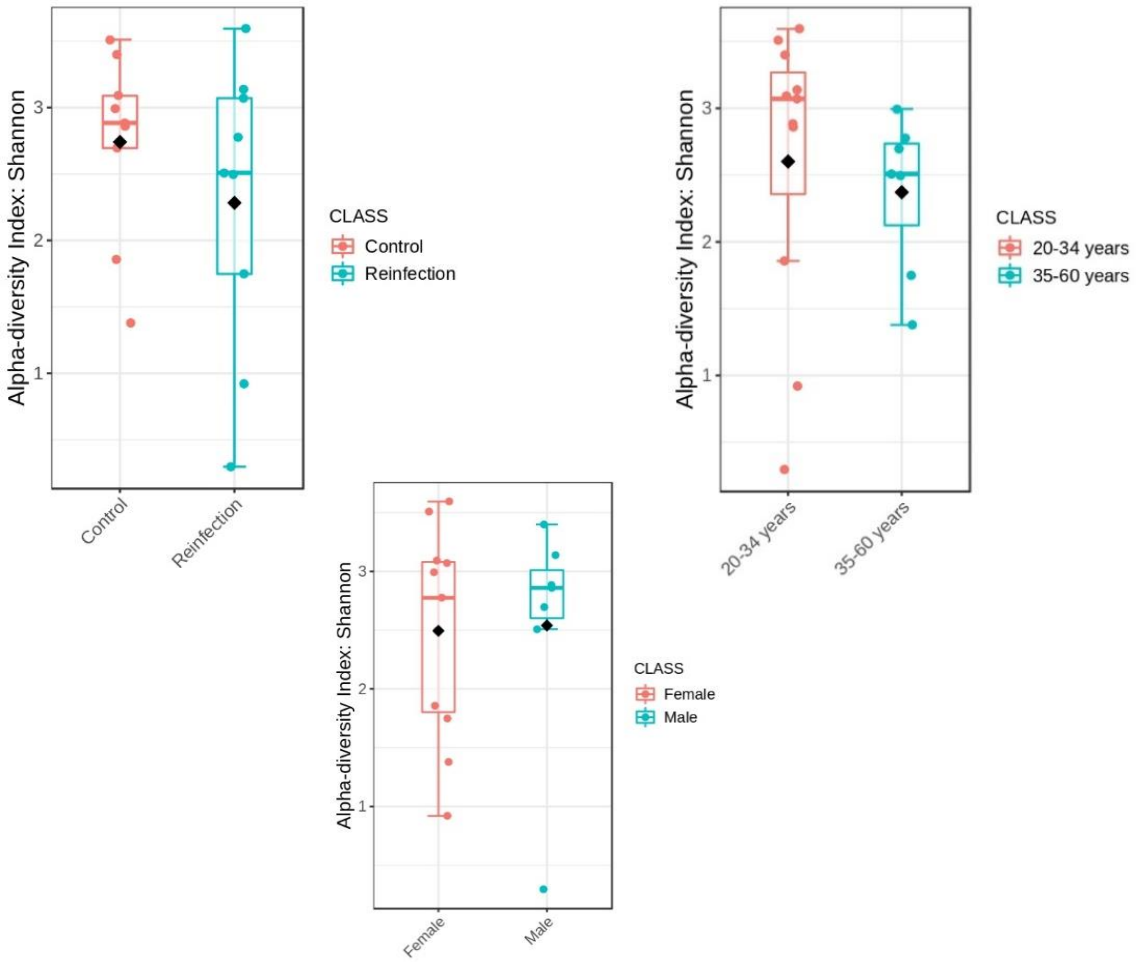

Figure 5. Alpha diversity analysis based on patient category, age and gender.

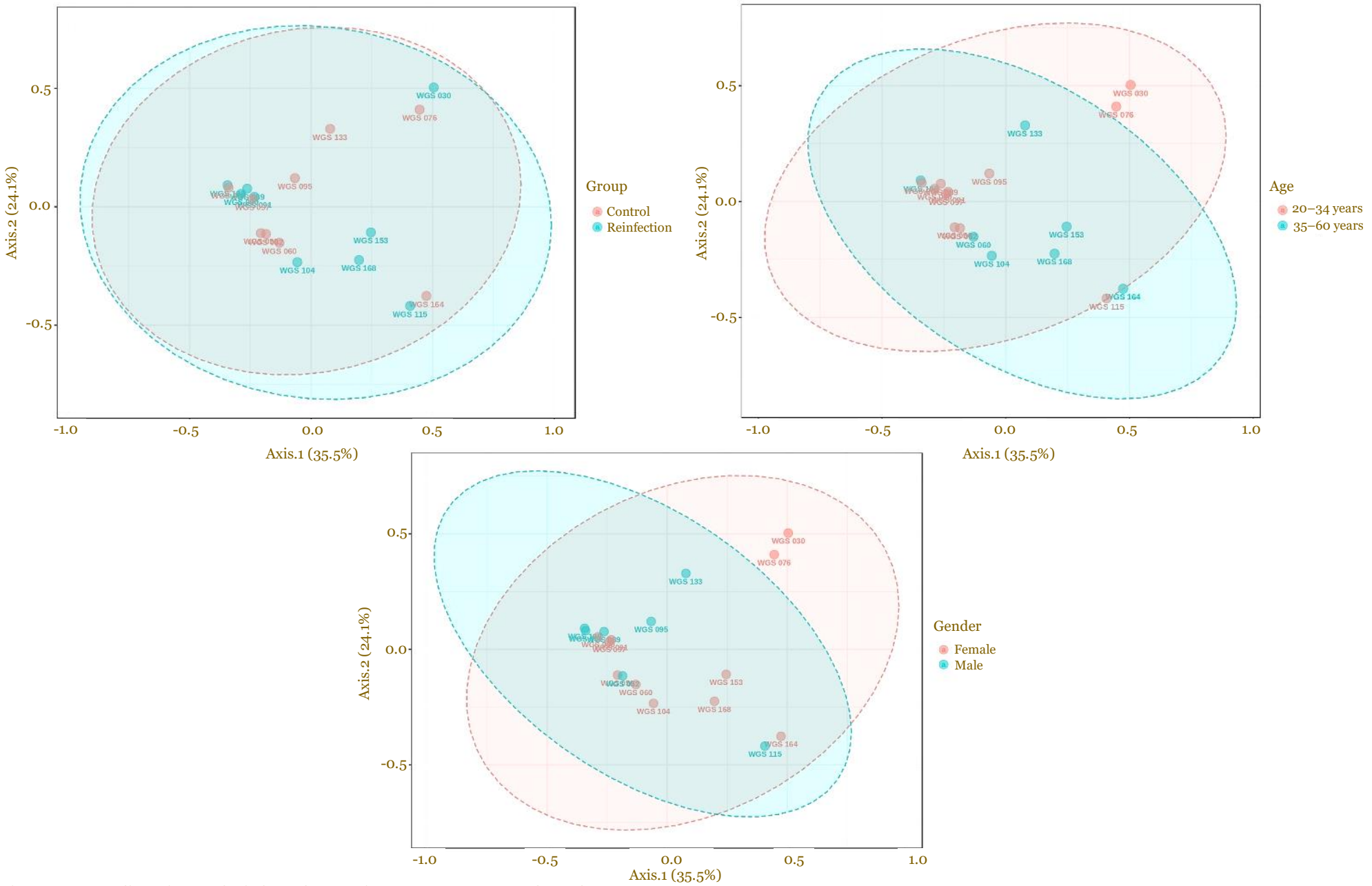

Figure 6. Beta diversity analysis based on patient category, age and gender.

Relative abundance analysis

At the phylum level, the predominant bacterial composition in the reinfection group encompassed *Terrabacteria* (52%), *Bacteroidetes* (27%), *Proteobacteria* (12%), and *Fusobacteria* (12%). This contrasted with the microbiota in non-reinfection COVID-19 cases, where those phyla constituted 55%, 30%, 7%, and 7%, respectively (**Figure 7A**). When analyzing the genus level, the dominant genus in both the reinfection and non-reinfection categories was *Streptococcus* (22% and 16%), followed by *Prevotella* (17% and 23%) (**Figure 7B**).

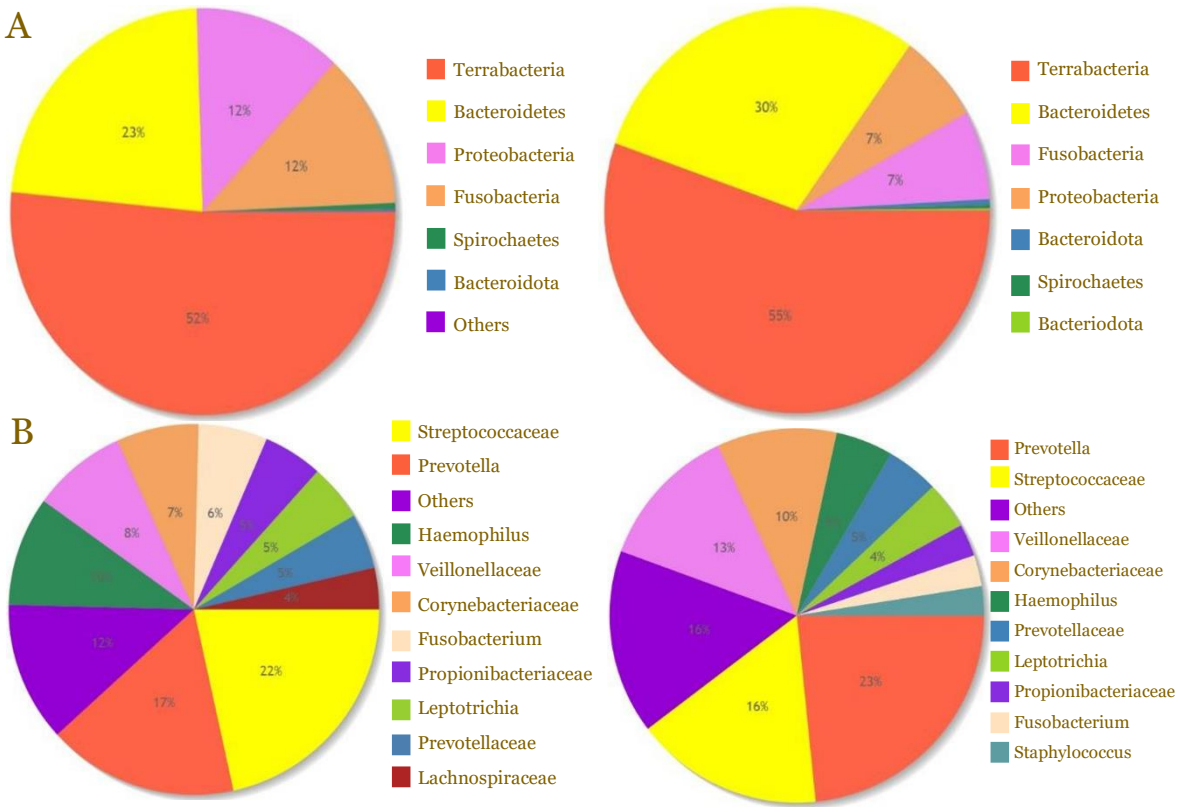

Figure 7. Analysis of bacterial abundance in reinfected and non-reinfected COVID-19 cases. (A) Relative abundance of phylum levels between reinfected and non-reinfected cases. (B) Relative abundance of genus level between reinfected and non-reinfected cases.

Through core Microbiome Analysis, several species emerged as dominant across both groups, including *Prevotella veroralis, Veillonella dispar, Veillonella parvula, Streptococcus pneumoniae*, and *Corynebacterium accolens* (**Figure 8**). At the genus level, several bacteria had higher abundance in reinfection cases, including *Streptococcus* (sample code of WGS 115), *Haemophilus* (WGS 153), *Corynebacterium* (WGS 030), and *Propionibacterium* (WGS 030). Conversely, other samples (WGS 050, WGS 052, WGS 054, WGS 060, and WGS 097) had a significantly higher proportion of *Prevotella, Veillonella*, and *Streptococcus*. In the non-reinfected samples, *Corynebacterium* was dominant in WGS 113, *Staphylococcus* in WGS 076, and Streptococcus and Haemophilus in WGS 164 (**Figure 9A**).

Upon comparing age and gender categories, this study revealed that the abundance of the genus Corynebacterium and Staphylococcus increased in female samples within the age range of 20–34 years. For male samples within the same age range, Streptococcus was dominant. Furthermore, for female samples within the adult group of 35–60 years, Haemophilus emerged as the dominant genus (**Figure 9B**).

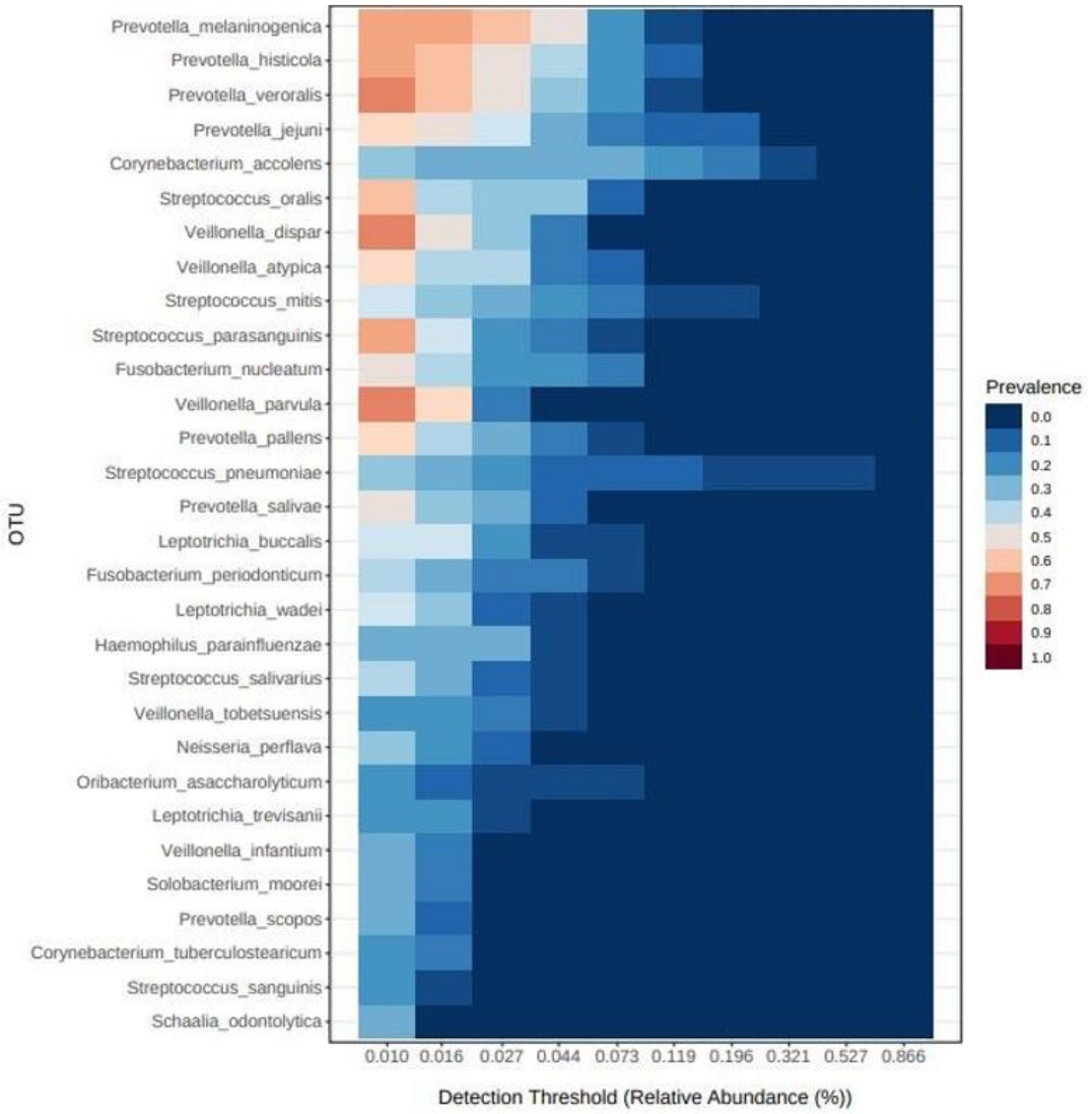

Figure 8. Care microbiome analysis based on relative abundance and sample prevalence of bacteria OTUs (Operational Taxonomic Units) group by species from reinfection and non-reinfection COVID-19 patients.

The MetagenomSeq analysis revealed that several species were significantly (p<0.05) more abundant in reinfection cases, such as *Haemophilus parainfluenzae, Fusobacterium periodonticum, Fusobacterium nucleatum, Leptotrichia buccalis mitis*, *Streptococcus pneumoniae, Streptococcus pseudopneumoniae*, and *Prevotella denticola* (**Figure 11–12**). In the male group, five species, including *Prevotella jejuni, Prevotella melaninogenica, Prevotella veroralis, Veillonella atypica*, and *Leptotrichia wadei* were significantly abundant compared to the female patients (p<0.05). Conversely, three species, *Streptococcus parasanguinis, Streptococcus oralis,* and *Corynebacterium pseudodiphtheriticum* were significantly more abundant in the female patients compared to the male patients (**Figure 13–14**).

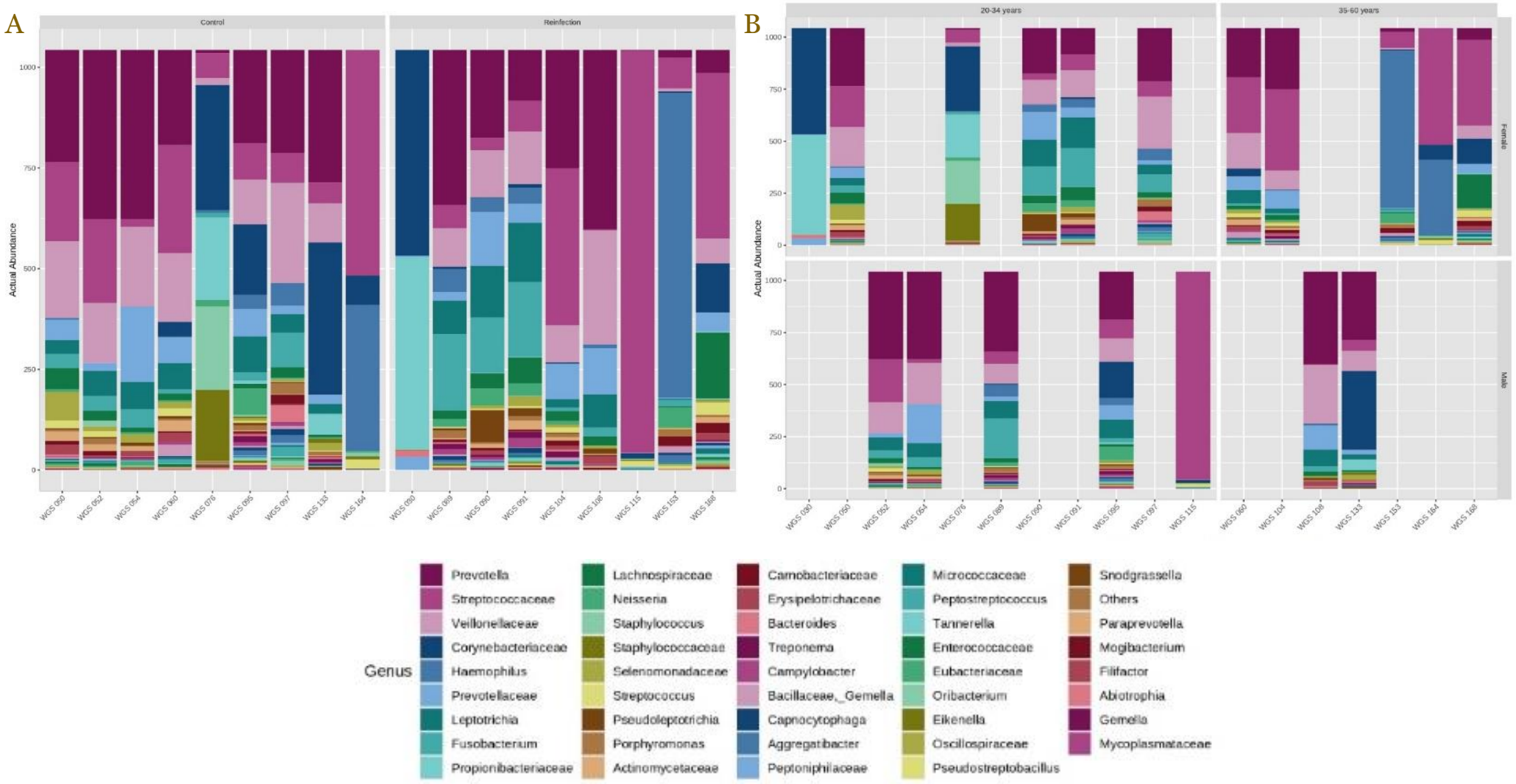

Figure 9. Analysis of the relative abundance of genus level for each sample. (A) *Genus level* bar plot by infection type. (B) Analysis of the relationship by age and gender.

Sativa et al. Narra J 2025; 5 (3): e2901 - http://doi.org/10.52225/narra.v5i3.2901

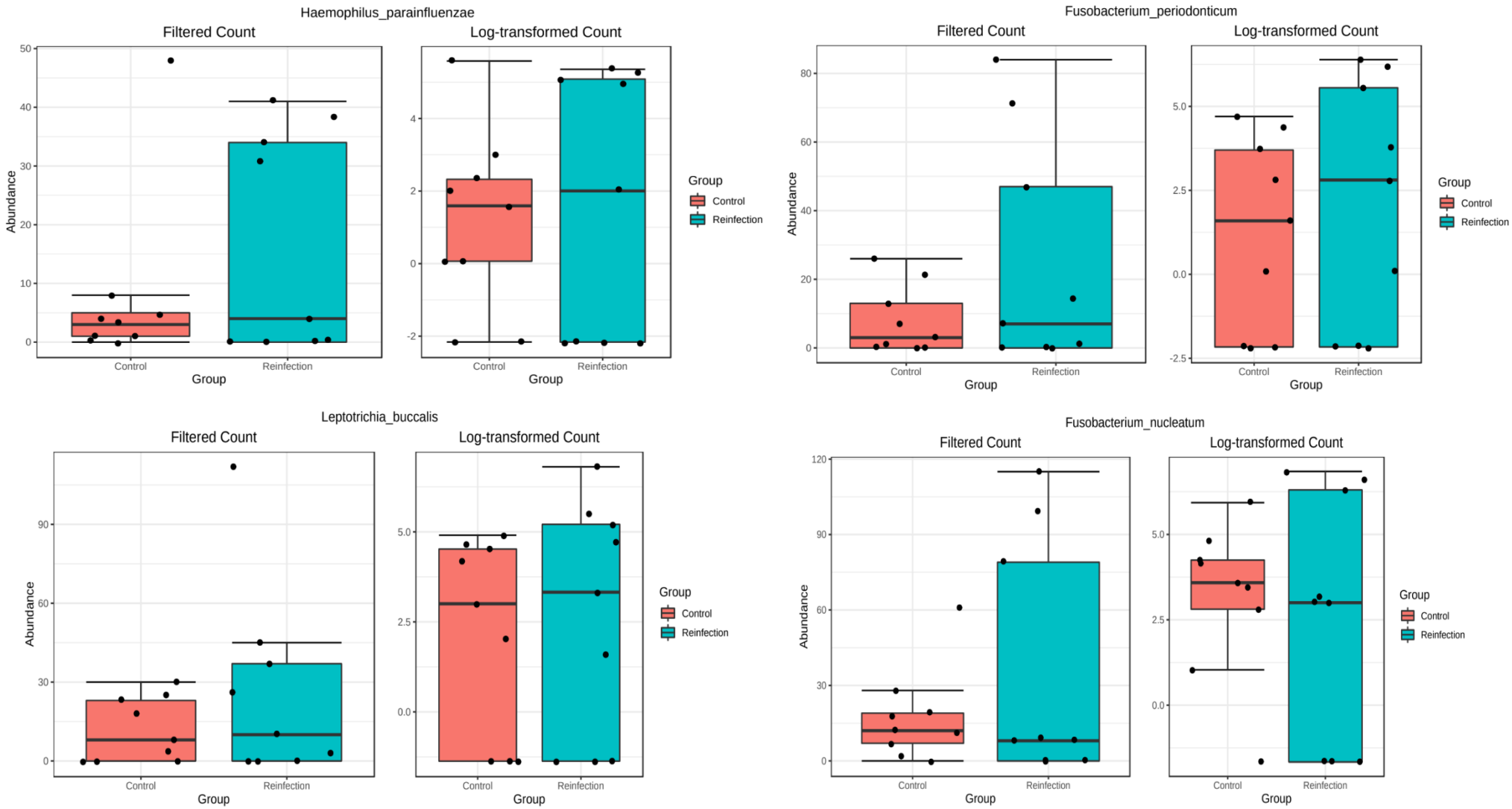

Figure 10. Visualization of metagenomSeq analysis based on false discovery rate (FDR) and *p*-value in various categories; several species that increased significantly in samples of reinfected patients were *Haemophilus parainfluenzae, Fusobacterium periodonticum, F. nucleatum and Leptotrichia buccalis*.

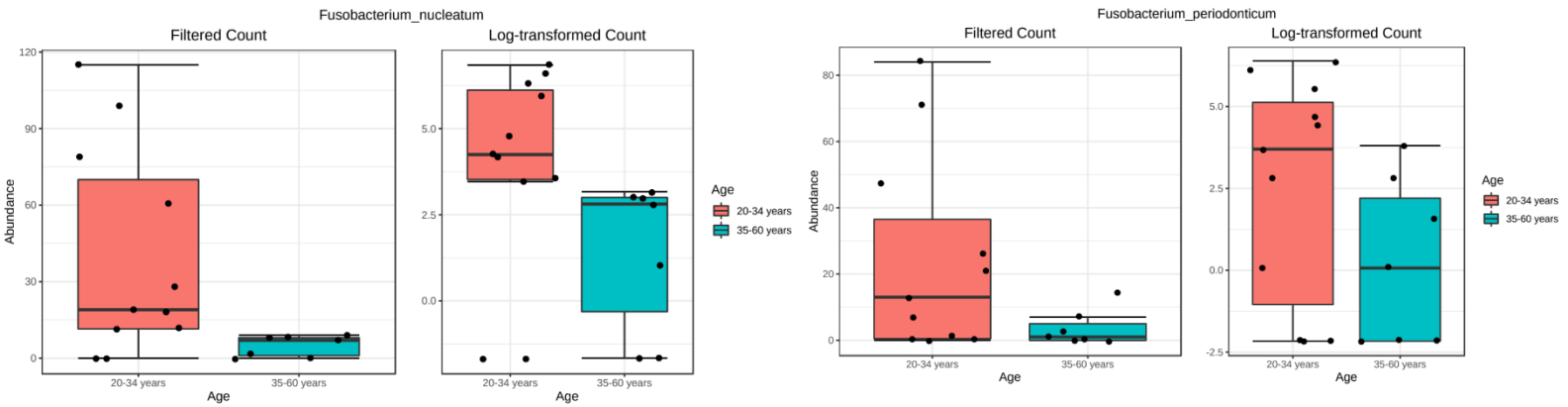

Figure 11. Visualization of metagenomSeq analysis based on false discovery rate (FDR) and *p*-value in various categories; significantly species visualization in young group.

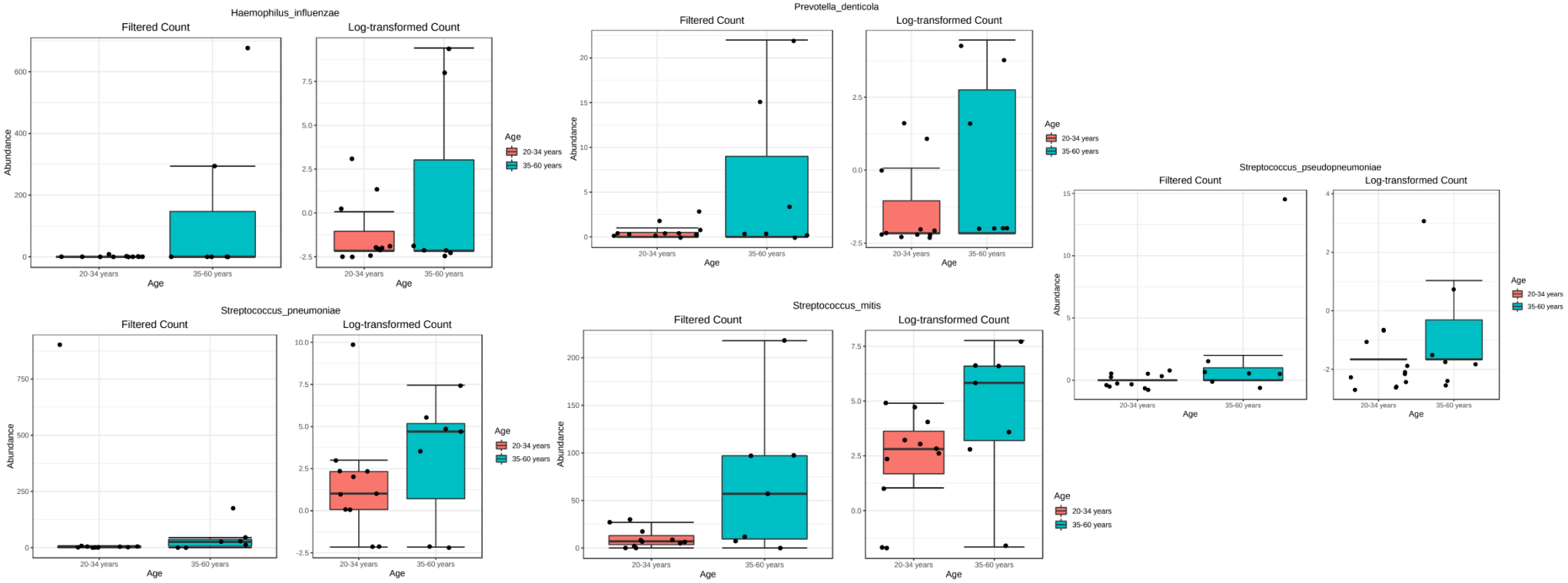

Figure 12. Visualization of metagenomSeq analysis based on false discovery rate (FDR) and *p*-value in various categories; in adult group.

Sativa et al. Narra J 2025; 5 (3): e2901 - http://doi.org/10.52225/narra.v5i3.2901

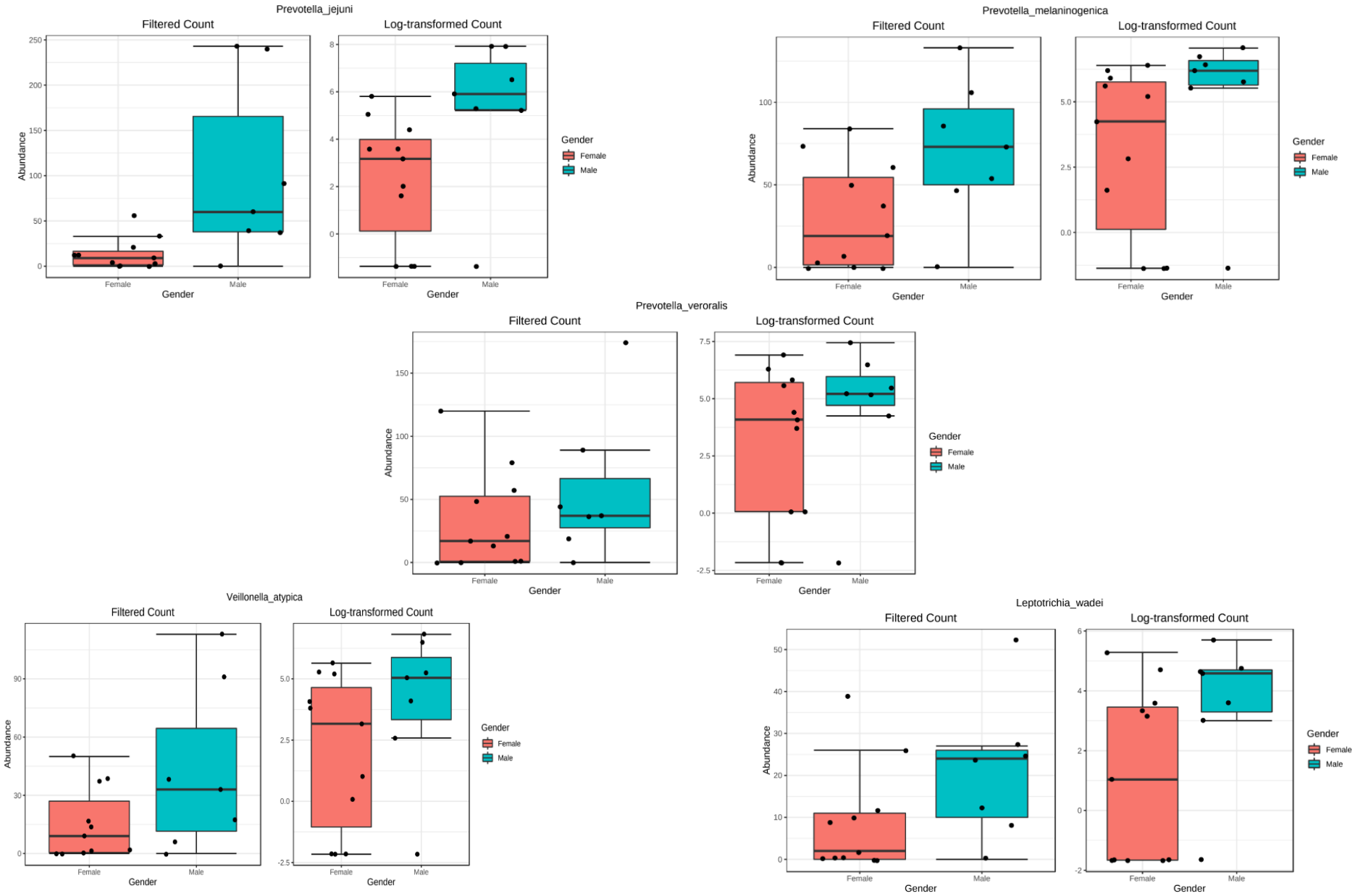

Figure 13. Visualization of metagenomSeq analysis based on false discovery rate (FDR) and *p*-value in various categories; analysis by male group.

Sativa et al. Narra J 2025; 5 (3): e2901 - http://doi.org/10.52225/narra.v5i3.2901

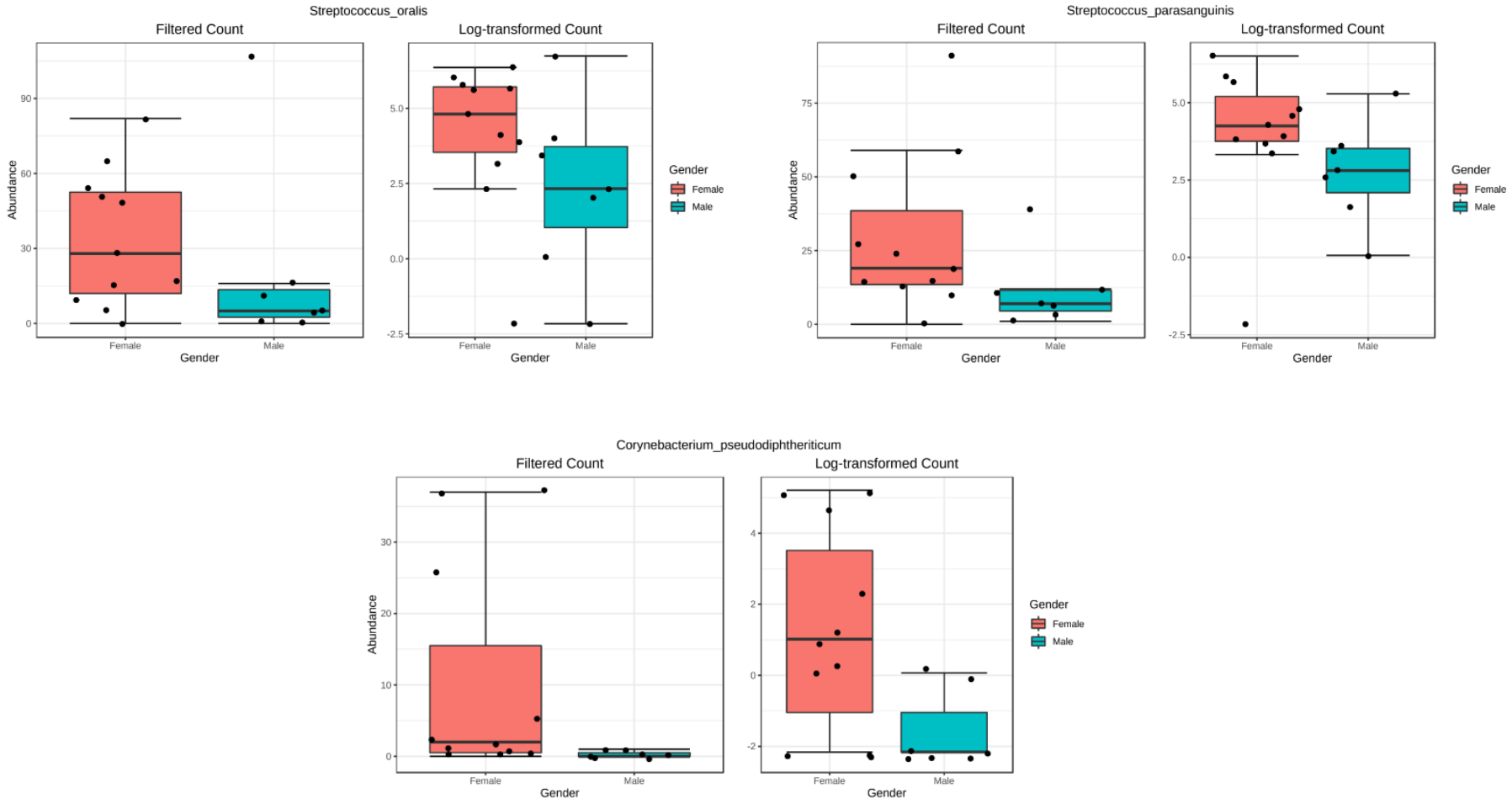

Figure 14. Visualization of metagenomSeq analysis based on false discovery rate (FDR) and *p*-value in various categories; found 3 species that increased significantly in the female group.

Discussion

Our research sought to characterize samples collected from COVID-19 reinfected patients in the West Java region using nanopore sequencing, yielding several noteworthy observations. Firstly, we examined the variants prevalent among the reinfected patients, revealing a marked predominance of the Delta variant. Secondly, we observed the mutations that manifested in SARS-CoV-2 genome. Lastly, we investigated the bacterial communities present within reinfected patients. These findings may have implications for COVID-19 transmissibility and are pertinent to the investigation of suspected reinfection cases [17].

Indonesia has been witnessing the prominence of various variants, including the Delta variant (B.1.671.2), Alpha variant (B.1.1.7), and the locally identified variant with lineages B.1.466.2 (WHO). Notably, during a four-month span of Delta variant outbreak in Indonesia, it was noteworthy that the lineage AY.23 emerged as the most dominant among Delta variants. This was subsequently followed by the AY.24 lineage, exerting an influence on the variant's prevalence during the outbreak period [18].

Our study, conducted using samples gathered between March and June 2021, underscores that the Delta variant predominantly existed with lineages AY.23, AY.24, and AY.109 (**Figure 1**). A resemblance can be drawn with a case in the United States (US) wherein, in April, the Delta variant accounted for a mere 0.1% of cases. This percentage surged to 1.3% in early May and remarkably soared to 20.6% in the subsequent month [18].

Furthermore, this comprehensive whole genome sequencing procedure assumes a vital role in tracking the emergence of mutations that may give rise to new variants, potentially amplifying virus transmissibility [17,20]. The Delta variant, marked by a notable mutation in the spike protein (S), perturbs the protein's ability to bind to the ACE-2 receptor in humans, a mechanism underpinning the rapid transmission capability of this variant [19].

Our study's findings underscore that in both reinfected and non-reinfected patients, the most prominent mutations were evident within the spike protein (S), followed by the NSP3 and Nucleocapsid protein (N) (**Figure 3A** and **3B**). Several pivotal mutations exist within distinct variants; the Delta variant, for instance, presents nine amino acid mutations within its spike protein (S), consequently impacting SARS-CoV-2 transmission dynamics, pathogenicity, and immunity [21]. Nonetheless, there remains a dearth of research regarding key mutations linked to reinfection cases. Therefore, our study's results unveil specific mutations exclusively present in certain reinfected samples when juxtaposed with non-reinfected patients. These include NSP3_V220A, S_T676I, ORF7a_V82A, and ORF7a_T120I (**Figure 4**).

Alpha diversity analysis, employed to assess the abundance of species and their distribution within a sample, provides insight into the overall species richness [22]. Our analysis results indicated that the obtained diversity levels exhibited relatively minor disparities across the reinfected and non-reinfected patient groups, as well as gender and age categories. This lack of significant differences between the observed groups is evident ($p>0.05$) (**Figure 5**). In a parallel study, similar Alpha diversity analysis was conducted, utilizing the Chao1, Shannon, and Simpson indices to juxtapose bacterial diversity between COVID-19 patients and controls. The outcome showcased a lack of statistical variance among all indices [23].

Generally, based on eta diversity analysis, an analogous trend emerges with no substantial differences noted ($p>0.05$). However, when segmenting the data according to age and gender categories, several samples were clustered differently from others (**Figure 6**). This observation underscores that samples positioned in close proximity exhibit greater similarity in their bacterial community composition [22].

In this investigation, the presence of SARS-CoV-2 infection did not exhibit a correlation with the relative abundance of bacterial communities present within the nasopharyngeal and oropharyngeal swabs of reinfected and non-reinfected patients. This outcome emerged from the observation of a shared dominance of the *Terrabacteria phylum* (52%; 55%) in both cases (**Figure 7A**). Similarly, the genera with the highest prevalence were *Streptococcus* (22%; 16%) and *Prevotella* (17%; 23%) (**Figure 7B**).

In a broader context, the most prevalent genus found among operational taxonomic units (OTUs) in COVID-19-positive patients is *Prevotella*, a common constituent of the oropharyngeal microflora frequently detected in cases of influenza. Overexpression of *Prevotella* may pose viral infections with higher clinical severity [24,25]. Furthermore, the outcomes of the MicrobiomeAnalyst analysis accentuate that among the entire sample set, *Prevotella veroralis,*

Veillonella dispar, and *Veillonella parvula* hold the highest prevalence. Meanwhile, *Streptococcus pneumoniae* and *Corynebacterium accolens* exhibit the highest abundance (**Figure 9**). These bacterial species are part of the normal upper respiratory tract microbiota and are predominantly found in saliva. However, they can trigger pneumonia among adults [25, 26, 27].

In comparison to healthy individuals, patients with underlying comorbidities linked to COVID-19 are at an elevated risk of suspected reinfection and are characterized by distinct genera associations. Among these, *Streptococcus* stands out, as it is frequently encountered in COVID-19 patients and exhibits a correlation with acute respiratory viral infections. Additionally, the presence of *Staphylococcus*, *Corynebacterium*, and *Haemophilus* genera has been linked to symptomatic manifestations in patients [25,28] (**Figure 10** and **11**). Furthermore, a metagenomSeq analysis pinpointed four species that yielded significantly heightened values: *Haemophilus parainfluenzae*, *Fusobacterium periodonticum, Fusobacterium nucleatum,* and *Leptotrichia buccalis* (**Figure 10**).

Among these, *Haemophilus parainfluenzae* and *Leptotrichia buccalis* are organisms naturally inhabiting the human upper respiratory tract, and their coinfection with COVID-19 can potentially exacerbate the patient's condition [29, 30]. Moreover, *Fusobacterium spp*., an oral bacterium, has been linked to SARS-CoV-2. This bacterium is implicated in the process of sialylation on cell surfaces, thereby serving as an alternative receptor for SARS-CoV-2 and influencing COVID-19 infection dynamics, potentially including the occurrence of reinfection [31, 32].

Furthermore, beyond comorbidities, various other risk factors have been documented to influence the outcomes for COVID-19 patients, including age and gender. Specifically, when considering the young age group, two species—Fusobacterium *periodonticum* and *Fusobacterium nucleatum*—were notably elevated. Conversely, within the adult age group, five species emerged: *H. influenzae, Streptococcus mitis, S. pneumoniae, S. pseudopneumoniae*, and *Prevotella denticola* (**Figure 11** and **12**).

This insight underlines the potential for age to reshape the relationship between specific bacterial taxa and SARS-CoV-2 infection [33]. The microorganisms identified in the older age group are often associated with respiratory disorders, including pneumonia. As humans age, their resistance to bacterial infections may wane, potentially leading to susceptibility to bacterial intrusion, protracted COVID-19 experiences, and even triggering scenarios of suspected reinfection [15, 26, 27].

Moving forward, our analysis examine into distinct male and female cohorts, uncovering several notable species. This finding underscores that both genders possess susceptibility to bacterial dysbiosis. In the male group, we identified five species that exhibited significant elevation: *Prevotella jejuni, P. melaninogenica, P. veroralis, Veillonella atypica*, and *Leptotrichia wadei*. The prevalence of these bacteria is often higher among men, potentially due to habits such as tobacco and alcohol consumption. For instance, *L. wadei*, detected in our results, is frequently isolated from the dental plaque of smokers [34].

Conversely, women's emphasis on dental and oral health contributed to the identification of only three species—*Streptococcus parasanguinis, S. oralis*, and *Corynebacterium pseudodiphtheriticum* (**Figure 13** and **14**). Notably, the presence of oral microbes, including *S. parasanguinis* and *S. oralis* were able to trigger an augmented cytokine response, potentially impacting lung immune homeostasis [35].

Moreover, numerous bacteria originating from the oral cavity have emerged. This phenomenon can be attributed to the oral environment, which offers a conducive setting for diverse microorganism growth [36]. As such, oral coinfection scenarios may arise, facilitated by the presence of the ACE2 receptor within tongue cells and oral epithelial cells [25]. Remarkably, despite the absence of existing reports detailing coinfection events in conjunction with COVID-19 reinfection cases, this study serves as a valuable guide for future investigations in this domain.

Nevertheless, our study does possess certain limitations. The sample size remains relatively small, and the absence of metadata for the initial infection, as well as comprehensive patient metadata, hampers our ability to ascertain patient lifestyles and underlying comorbidities.

Conclusion

Our observations reveal a lack of disparities in SARS-CoV-2 lineages between reinfection and non-reinfection cases, as well as within age and gender classifications. Despite both COVID-19 reinfection and new infection instances stemming from comparable SARS-CoV-2 lineages, noteworthy key mutations and distinct bacterial species potentially serve as biomarkers within both groups. This emphasizes the complex interaction between viral and bacterial communities that could contribute to the course of COVID-19 infections.

Ethics approval

All participants consented under Komisi Etik Penelitian Kesehatan FKUI/RSCM (20-10-1321_EXP).

Acknowledgments

We gratefully acknowledge the COVID-19 diagnostic testing team of BSL-3 and BSL-2 Laboratory, National Research and Innovation Agency (BRIN)-Indonesia and originating laboratories in West Java, Indonesia. ARS, IZA and MMA were the recipients of the LIPI Fellowship Program 2021 for the VenomCoV Project.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This work was supported by RISPRO-LPDP Funding Program for COVID-19 (07/FI/P-KCOVID-19.2B3/X/2020) under the project name “Surveilans Genom SARS-CoV-2” (VenomCoV Project) of LIPI/BRIN.

Underlying data

All 51 complete genomes of SARS-CoV-2 are available at GISAID : EPI_ISL_4004670, 72, 74, 76, 78-79 (n = 6) ; EPI_ISL_5536530-31, 34-36, 38-39 (n = 7) ; EPI_ISL_5536501-05, 07 (n – 6) ; EPI_ISL_5536540-42, 45, 49 (n = 5) ; EPI_ISL_4004705, 10-12 (n = 4) ; EPI_ISL_5536569 (n = 1) ; EPI_ISL_4004680, 84, 86-88 (n = 5) ; EPI_ISL_5536551-52, 55, 57 (n = 4) ; EPI_ISL_5536510, 13-14, 16, 19 (n = 5) ; EPI_ISL_4004693, 95 (n = 2) ; EPI_ISL_5536570, 72-73 (n = 3) ; EPI_ISL_5536520, 23-24 (n = 3).

Declaration of artificial intelligence use

We hereby confirm that no artificial intelligence (AI) tools or methodologies were utilized at any stage of this study, including during data collection, analysis, visualization, or manuscript preparation. All work presented in this study was conducted manually by the authors without the assistance of AI-based tools or systems.

How to cite

Sativa AR, Asyifa, IZ, Adzdzakiy MM, *et al*. SARS-CoV-2 lineages and naso-oropharyngeal bacterial communities in COVID-19 reinfection: A study in West Java, Indonesia. Narra J 2025; 5 (3): e2901 - http://doi.org/10.52225/narra.v5i3.2901.

References

1. Braun T, Halevi S, Hadar R, *et al*. SARS-CoV-2 does not have a strong effect on the nasopharyngeal microbial composition. Sci Rep 2021;11(1):8922.
2. Dubey A, Choudhary S, Kumar P, Tomar S. Emerging SARS-CoV-2 variants: Genetic variability and clinical implications. Curr Microbiol 2021;79(1):20.
3. Astuti I, Ysrafil. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes Metab Syndr 2020;14(4):407-412.
4. World Health Organization. Coronavirus disease (COVID-19). Available from: https://www.who.int/health-topics/coronavirus#tab=tab_1. Accessed: 18 January 2020.

5. Falahi S, Kenarkoohi A. COVID-19 reinfection: Prolonged shedding or true reinfection?. New Microbes New Infect 2020;38:100812.
6. Centers for Disease Control and Prevention. COVID-19: About reinfection. Available from: https://www.cdc.gov/coronavirus/2019-ncov/your-health/reinfection.html. Accessed: 20 January 2020.
7. Babiker A, Marvil CE, Waggoner JJ, *et al*. The importance and challenges of identifying SARS-CoV-2 reinfections. J Clin Microbiol 2021;59(4):e02769-20.
8. Fakhroo A, AlKhatib HA, Al Thani AA, Yassine HM. Reinfections in COVID-19 patients: Impact of virus genetic variability and host immunity. Vaccines 2021;9(10):1168.
9. Singh V, Upadhyay P, Reddy J, Granger J. SARS-CoV-2 respiratory co-infections: Incidence of viral and bacterial co-pathogens. Int J Infect Dis 2021;105:617-620.
10. Engen PA, Naqib A, Jennings C, *et al*. Nasopharyngeal microbiota in SARS-CoV-2 positive and negative patients. Biol Proced Online 2021;23(1):10.
11. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. J Med Virol 2020;92(4):418-423.
12. Ventero MP, Cuadrat RRC, Vidal I, *et al*. Nasopharyngeal microbial communities of patients infected with SARS-CoV-2 that developed COVID-19. Front Microbiol 2021;12:637430.
13. Brinkman S, Termorshuizen F, Dongelmans DA, *et al*. Comparison of outcome and characteristics between 6343 COVID-19 patients and 2256 other community-acquired viral pneumonia patients admitted to Dutch ICUs. J Crit Care 2022;68:76-82.
14. Mostafa HH, Fissel JA, Fanelli B, *et al*. Metagenomic next-generation sequencing of nasopharyngeal specimens collected from confirmed and suspect COVID-19 patients. mBio 2020;11(6):e01969-20.
15. Elouar I, Djekoun A. Oxford nanotechnologie: A new era for genome sequencing and precision diagnostics. Sci Technol 2019;4(1):9-14.
16. Quick J. nCoV-2019 sequencing protocol v3 (LoCost) V.3. Retrieved from Protokol Whole Genome Sequencing for SARS-CoV-2. Available from: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bp2l6n26rgqe/v3. Accessed: 11 August 2021.
17. Prasetyoputri A, Dharmayanthi AB, Iryanto SB, *et al*. The dynamics of circulating SARS-CoV-2 lineages in Bogor and surrounding areas reflect variant shifting during the first and second waves of COVID-19 in Indonesia. PeerJ 2022;10:e13132.
18. Alexandar S, Ravisankar M, Kumar RS, Jakkan K. A comprehensive review on COVID-19 delta variant. Int J of Pharmacology and Clin Research 2021;5(2):83-85.
19. Harvey WT, Carabelli AM, Jackson B, *et al*. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 2021;19(7):409-424.
20. Tian D, Sun Y, Zhou J, Ye Q. The global epidemic of the SARS-CoV-2 delta variant, key spike mutations and immune escape. Front Immunol 2021;12:751778.
21. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat Protoc 2020;15(3):799-821.
22. Nardelli C, Gentile I, Setaro M, *et al*. Nasopharyngeal microbiome signature in COVID-19 positive patients: Can we definitively get a role to *Fusobacterium periodonticum*?. Front Cell Infect Microbiol 2021;11:625581.
23. Lu HF, Li A, Zhang T, *et al*. Disordered oropharyngeal microbial communities in H7N9 patients with or without secondary bacterial lung infection. Emerg Microbes Infect 2017;6(12):e112.
24. Bao L, Zhang C, Dong J, *et al*. Oral microbiome and SARS-CoV-2: Beware of lung co-infection. Front Microbiol 2020;11:1840.
25. Könönen E, Gursoy UK. Oral *Prevotella* species and their connection to events of clinical relevance in gastrointestinal and respiratory tracts. Front Microbiol 2022;12:798763.
26. Haran JP, Bradley E, Zeamer AL, *et al*. Inflammation-type dysbiosis of the oral microbiome associates with the duration of COVID-19 symptoms and long COVID. JCI Insight 2021;6(20):e152346.
27. Teran RA, Walblay KA, Shane EL, *et al*. Postvaccination SARS-CoV-2 infections among skilled nursing facility residents and staff members - Chicago, Illinois, December 2020-March 2021. Am J Transplant 2021;21(6):2290-2297.
28. Ou X, Zhou L, Huang H, *et al*. A severe case with co-infection of SARS-CoV-2 and common respiratory pathogens. Travel Med Infect Dis 2020;35:101672.

29. Devi P, Maurya R, Mehta P, *et al*. Increased abundance of *Achromobacter xylosoxidans* and *Bacillus cereus* in upper airway transcriptionally active microbiome of COVID-19 mortality patients indicates role of co-infections in disease severity and outcome. Microbiol Spectr 2022;10(3):e0231121.
30. Yoneda S, Loeser B, Feng J, *et al*. Ubiquitous sialometabolism present among oral fusobacteria. PLoS One 2014;9(6):e99263.
31. Morniroli D, Gianni ML, Consales A, *et al*. Human sialome and coronavirus disease-2019 (COVID-19) pandemic: An understated correlation?. Front Immunol 2020;11:1480.
32. Hurst JH, McCumber AW, Aquino JN. Age–related changes in the nasopharyngeal microbiome are associated with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and symptoms among children, adolescents, and young adults. Clin Infect Dis 2022;75(1):e928-e937.
33. Eribe ERK, Olsen I. Leptotrichia species in human infections II. J Oral Microbiol 2017;9(1);1368848.
34. Khan S, Liu J, Xue M. Transmissions of SARS-CoV-2, required developments in research and associated public health concerns. Front Med 2020;7:310.
35. Ma S, Zhang F, Zhou F, *et al*. Metagenomic analysis reveals oropharyngeal microbiota alterations in patients with COVID-19. Signal Transduct Target Ther 2021;6(1):191.