Computational and Structural Biotechnology Journal 22 (2023) 50–57

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Research article

Visualization of the individual blood microbiome to study the etiology of sarcoidosis

Yordan Hodzhev [a], Borislava Tsafarova [a], Vladimir Tolchkov [a], Vania Youroukova [b], Silvia Ivanova [b], Dimitar Kostadinov [b], Nikolay Yanev [b], Maya Zhelyazkova [c], Stefan Tsonev [d], Reni Kalfin [e,f], Stefan Panaiotov [a,*]

[a] Microbiology Department, National Center of Infectious and Parasitic Diseases, Yanko Sakazov 26 Blvd., Sofia 1504, Bulgaria
[b] Department of Pulmonary Diseases, University Hospital for Pulmonary Diseases "St. Sofia", Medical University of Sofia, Akad. Ivan Evstratiev Geshov 17 Blvd., Sofia 1431, Bulgaria
[c] Faculti of Mathematics and Informatics, Sofia University St. Kliment Ohridski, 5 James Bourchier Blvd., 1164 Sofia, Bulgaria
[d] Agrobioinstitute (ABI), 8 Dragan Tsankov, Blvd, Sofia 1164, Bulgaria
[e] Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria
[f] Department of Health Care, South-West University "Neofit Rilski", Blagoevgrad 2700, Bulgaria

ARTICLE INFO

Keywords:
Lung diseases
Sarcoidosis
Microbiome
Metagenome analysis
Visualization
Sankey diagram

ABSTRACT

Introduction: Single microbial pathogens or host-microbiome dysbiosis are the causes of lung diseases with suspected infectious etiology. Metagenome sequencing provides an overview of the microbiome content. Due to the rarity of most granulomatous lung diseases collecting large systematic datasets is challenging. Thus, single-patient data often can only be summarized visually.
Objective: To increase the information gain from a single-case metagenome analysis we suggest a quantitative and qualitative approach.
Results: The 16S metagenomic results of 7 patients with pulmonary sarcoidosis were compared with those of 22 healthy individuals. From lysed blood, total microbial DNA was extracted and sequenced. Cleaned data reads were identified taxonomically using Kraken 2 software. Individual metagenomic data were visualized with a Sankey diagram, Krona chart, and a heat-map. We identified five genera that were exclusively present or significantly enhanced in patients with sarcoidosis - *Veillonella, Prevotella, Cutibacterium, Corynebacterium*, and *Streptococcus*.
Conclusions: Our approach can characterize the blood microbiome composition and diversity in rare diseases at an individual level. Investigation of the blood microbiome in patients with granulomatous lung diseases of unknown etiology, such as sarcoidosis could enhance our comprehension of their origin and pathogenesis and potentially uncover novel personalized therapeutics.

1. Introduction

Granulomatous disorders encompass numerous nosological entities that share the histological feature of granuloma formation. Granulomatous lung diseases (GLD) in humans affect the lungs and can cause non-necrotizing granulomas, which may result in organ failure and death [1,2]. Granuloma is a particular type of inflammatory response that involves the formation of a mass of immune cells known as granulocytes, which form a protective barrier around a foreign substance or body [3]. Sarcoidosis is a granulomatous multisystem inflammatory disease and has both immunological and genetic components [4] however, recent studies on bronchoalveolar lavage (BAL) indicate that the disease may also be linked to lung microbiota alterations [5,6] or antigen-detected pathogens as a risk factor [7]. It is hypothesized that the etiology of sarcoidosis is associated with an autoimmune reaction, genetic predisposition, bacterial infection, or host-microbiome dysbiosis [8]. Among the various forms of the disease, pulmonary sarcoidosis is the most common. Sarcoidosis is typically diagnosed by observations of bilateral hilar adenopathy, and the presence of non-caseating granulomas in the lung tissue [9–11]. Fig. 1A outlines the current examination

* Corresponding author.
E-mail address: spanaiotov@yahoo.com (S. Panaiotov).

https://doi.org/10.1016/j.csbj.2023.10.027
Received 19 June 2023; Received in revised form 14 October 2023; Accepted 16 October 2023
Available online 20 October 2023

protocol for GLD, including a broad range of methods due to the challenges of diagnosis and limited treatment options [10]. Evidence supporting the hypothesis of microbial infection in lung sarcoidosis suggests the presence of microorganisms [12]. Members of the *Mycobacterium* and *Cutibacterium* genera as well as fungi have been linked to sarcoidosis, but not confirmed as causative agents [13,14]. Recently, *Atopobium* and *Fusobacterium* species have been suspected as etiological agents of sarcoidosis due to their identification in BAL samples from patients with pulmonary sarcoidosis [5].

The classical culturing approach cannot fully characterize the host microbiome due to a significant number of nonculturable microbial species. Metagenome analysis of lung and lymphatic nodule biopsy samples from patients with sarcoidosis and other GLD utilizing bacterial 16 S rDNA and fungal ITS taxonomy suggests no single microbial pathogen is associated with granulomatosis progression [5,12,15]. There is evidence suggesting that individuals with GLDs exhibit alterations in the quantity not quality of lung tissue microbiota content [12, 15]. It would be valuable to conduct in-depth qualitative and quantitative metagenomic evaluation of the affected tissues. In this context, sarcoidosis, as a multifaced disease, is contingent on the equilibrium between the host and microbiome. Thus, the hypothesis of dysbiosis in the host-microbiome relationship appears to be a feasible trigger for sarcoidosis or other GLDs.

The incidence of sarcoidosis in Bulgaria and southeastern Europe, and many Asian countries ranges from 1 to 10 cases per 100,000 population. Group data collection that meets the requirements for statistical analysis is severely limited [4,16]. Analysis and data visualization of one or a few samples with no statistical support is challenging. Therefore, methods for assessing metagenomic data at the individual level are necessary.

Visualization of metagenomic data is a rapidly developing field that introduces novel methods and tools for creating and evaluating new biological hypotheses [17,18]. Sankey diagrams were first used by Kennedy and Sankey in 1898 to represent machine energy flow [19]. They remain a useful tool for visualizing complex data in contemporary metagenomics research. The chart is composed of arrows, whose width represents the flow magnitude (such as microbial read counts) and nodes, that depict hierarchy levels of microbial taxa. In the context of metagenomics, Sankey plots are an appropriate visualization method for single-sample metagenomic data. Thus, they are an effective tool for visualizing complex metagenomic data [20,21].

Human blood is generally believed to be free of microorganisms due to the protective activity of the immune system. The occurrence of bacteria in the bloodstream is defined as bacteremia and is related to sepsis or chronic latent infection [22,23]. However, studies have demonstrated the presence of blood microbiota by culture resuscitation [2,24–27], and by 16 S metagenomic sequencing [28,29] in healthy individuals, indicating the presence of a self-sustainable microbial community in the blood. The analysis of the blood microbiome has become a potent tool for research and diagnosis, and utilization of liquid biopsy samples has enabled noninvasive investigations of infectious diseases [30]. The circulatory system, in general, is deemed a viable host-microbiome niche, and microbial dysbiosis may be associated with sarcoidosis and other GLDs [27,31,32]. Thus, evaluating the blood microbiome may be an appropriate line of research and a diagnostic approach for sarcoidosis and GLDs. Timely detection of microbial profiles in patient blood samples is crucial for accurate clinical assessment. Unfortunately, there is still a lack of thorough research dedicated specifically to the blood microbiome in patients with sarcoidosis.

As previously stated, obtaining a significant number of samples for statistical analysis in cases of rare diseases is challenging [33,34]. Therefore, the only approach is to analyze individual samples. It is imperative to develop and test methods and techniques that have universal applicability and serve as a foundation for the ever-evolving science of biomedical informatics. Our aim was to evaluate an advanced visualization method for assessing the qualitative and quantitative features of blood microbiomes from individual or small sample-size data sets of sarcoidosis patients. To enhance the information retrieval and address the existing constraints for analyzing small patient groups with rare diseases, we examined the 16 S metagenomic results of 7 patients with pulmonary sarcoidosis and 22 healthy controls. Here, we suggest monitoring the composition of the blood microbiome as a novel parameter in the diagnosis of sarcoidosis.

2. Materials and methods

The study was approved by the Institutional review board/ Institutional ethics committee (IRB/IEC) Number IRB00006384; Protocol Number 3/27.07.2020).

2.1. Patients

Between November 2020 and November 2022, a pool of 7 patients were suspected of having pulmonary sarcoidosis. The diagnostic criteria for sarcoidosis were applied [10,35]. Patients were subjected to transbronchial biopsy, followed by histopathological assessment (Fig. 1A), to

Table 1
Characteristics of study samples. Abbreviations: AH - Arterial hypertonia; SN - Struma nodosa; SA - Sleep apnea; DM - Diabetes mellitus; NA - not applicable.

Characteristics	Sarcoidosis	Controls
Sample size	7	22
Mean age years (± SD)	48.33 (± 19.14)	49.21 (± 17.45)
Males/females	4/3	12/10
Blood type distribution A/B/AB/0	2/2/1/2	7/4/6/5
Sarcoidosis type II	7	0
Number of histologically confirmed samples	7	NA
Corticosteroid therapy	0	0
Never/former/current smokers	4/1/2	10/4/2
Comorbidity – AH/SN/SA/DM	1/1/1/0	0/0/0/1

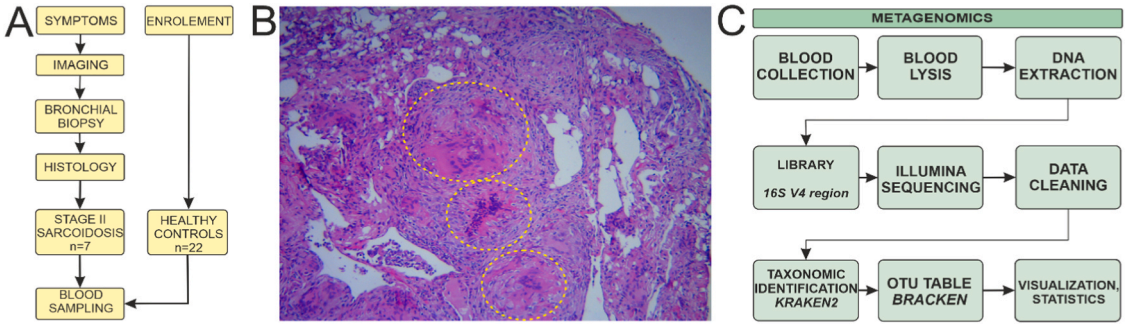

Fig. 1. (A) Diagnostic protocol applied for sarcoidosis. (B) Light microscopy of multiple non-caseous granulomas from bronchial wall of stage II sarcoidosis patient - marked in yellow circles. (C) Main steps of the metagenomic analysis to test for dysbiosis in the blood microbiome of sarcoidosis patients.

detect non-caseating epithelioid-cell granulomas (Fig. 1B). All patients (Table 1.) were positively diagnosed with stage two pulmonary sarcoidosis [36] and subsequently included in this study. Blood samples from patients and 22 clinically healthy individuals were collected in 5 ml Vacutainer tubes containing K_3EDTA as an anticoagulant (Vacutainer K3E, BD, USA). None of the patients were receiving antibiotics or any other medication related to sarcoidosis during sample collection.

2.2. DNA extraction

Three milliliters of whole blood cells were lysed in 10 ml of dH_2O that was filtered through a 0.2 μm filter and autoclaved. Prior to filtration and autoclaving of the dH_2O, DNase I was added at a concentration of 1 U/ml and incubated for 1 h at 37 ºC. DNase I was added to remove cell-free and human DNA. After centrifugation at 4000 rpm for 20 min microbial cells were collected and the pellet was washed twice with 10 ml dH_2O. Each wash was followed by centrifugation. The cell pellet was resuspended in 1 ml of lysis buffer containing 500 mM NaCl, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA, and 4% sodium dodecyl sulfate. The resulting cell suspension was vigorously vortexed or homogenized with 0.1/0.3 mm silica/zirconium beads (Biospec Products, Bartlesville, OK, USA) using a bead beater (Benchmark Scientific, Sayreville, NJ, USA) for 3 min at 4000 rpm. After performing microbial cell lysis we isolated the DNA using a previously described procedure [29]. Extracted DNA was resuspended in 100 μL of sterile DNA/RNA free dH_2O (Sigma-Aldrich, Burlington, VT, USA). The typical yield of DNA was > 150 ng/μL, with a 260/280 nm ratio > 1.7.

2.3. Metagenomic sequencing

Fig. 1C illustrates the key steps involved in our metagenomic analysis for preparing 16 S rDNA amplicon libraries. We used the NEXTFLEX® 16 S V4 Amplicon-Seq Kit 2.0 for Illumina Platforms (PerkinElmer, Inc. Waltham, MA, USA) to sequence the V4 hypervariable region. Sequencing was done using the Illumina NovaSeq 6000 (Illumina Inc., San Diego, CA, USA).

Identification of metagenomes was based on clean data after quality control (Fig. 1 B). The Illumina MiSeq Reporter software version 2.5.1.3 (Illumina Inc., San Diego, CA, USA) and the Illumina Sequence Analysis Viewer version 2.1.8 (Illumina Inc., San Diego, CA, USA) were used for imaging and evaluating of the sequencing run performance. MiSeq Reporter performed primary data analysis and quality control, including signal processing and de-multiplexing, with removal of low-quality sequences and human DNA from subsequent analysis. Kraken 2 software (version v2 2.1.2; SILVA database, version 2022–02–02T162959Z) was utilized for taxonomic profiling [37]. Additionally, we applied Bayesian Reestimation of Abundance with Kraken (Bracken) to accurately calculate the microbial abundance levels for each metagenomic sample based on the taxonomy reports generated by Kraken 2 [38]. The Pavian package (version 20) was used to generate an OTUs table, which provides genus-level microbiome abundance information [20]. Sequence clustering was performed using a 97% similarity index [39]. Microbial genera present in negative controls, such as water, reagents, air, washings of bronchoscope equipment, were excluded from the analysis. Individual OTUs were normalized via rarefaction to ensure an equivalent number of OTU reads in all samples [40]. The OTU tables underwent statistical evaluation to determine microbial composition and abundance.

2.4. Visualizing individual blood microbiome data in sarcoidosis

We used the Sankey [19] and Krona diagrams to visualize the microbial composition at individual and group level of sarcoidosis patients and controls. The data were processed with the Pavian package (R-3.6.0) for Sankey diagram and Krona tool for metagenomic visualization (KronaTools-2.8, https://github.com/marbl/Krona/releases/tag/v2.8) [41]. Sankey diagrams show actual abundance, while Krona diagrams display relative abundance of microbial reads in percentages.

2.5. Statistics

The MicrobiomeAnalyst web-based service (https://www.microbiomeanalyst.ca; accessed on January 9, 2023) was used to statistically compare the microbiome composition and abundance in sarcoidosis patients with the control group. The service is based on the MicrobiomeAnalyst R package for statistical, visual, and functional analysis of the microbiome [42]. For quantitative comparison of the microbial similarity in each sample (alpha diversity) a Shannon diversity index was calculated [43]. Group variations in alpha diversity were evaluated utilizing a parametric T-test. The normal distribution of the data was assessed with the Kolmogorov-Smirnov goodness of fit test. We applied principal coordinate analysis (PCoA) for the evaluation of between-group (beta) diversity [44], using the Bray-Curtis dissimilarity index [45]. Two-dimensional PCoA plots were generated for statistical comparison of samples by computing the first two principal components, which then served as the coordinates of the plots. To test the statistical significance of differences between the two sample groups, permutation analysis of variance (PERMANOVA) was utilized. Normalization of the data was performed to address the dispersion of the sampling depth. Rarefaction was applied since the sequencing number of reads in the metagenomic libraries per sample exceeded 10-fold [40]. Hierarchical clustering analysis was performed using the Euclidean distance measure and the Ward clustering algorithm [46]. The individual normalized abundance at the genus level was demonstrated via a Heat map diagram [47]. The study compared mean differences in major genera, which were log-transformed, using the Mann-Whitney U test. To control the false discovery rate (FDR), P-values were adjusted for multiple comparisons employing the Benjamini-Hochberg method [48]. Two-tailed P-values (including False Discovery Rate - adjusted P-values) of < 0.05 were deemed statistically significant.

3. Results

During a two-year period from November 2020 to November 2022, blood samples were taken from 7 sarcoidosis patients and 22 healthy individuals. Samples were sequenced by a targeted 16 S metagenomic approach. A total of 294736 raw sequences were obtained after quality control and processing, which were then normalized to 900 sequences per sample. Taxonomic analysis identified 1541 OTUs, with 118 OTUs having an abundance threshold of over 10 reads. Table 2 displays information on the distribution of the reads among the key taxa.

Microbiome data of each sarcoidosis patient was visually represented at phylum and genus level by two alternative graphical approaches – a Sankey diagram (Fig. 2 A) and a concentric Krona pie chart (Fig. 2 D). Figs. 2B and 2E display the microbiome abundance represented by the mean number of bacterial reads within the sarcoidosis group, whereas Figs. 2C and 2F depict the same analysis for the group of healthy individuals. These methods provided an easy-to-use color visual evaluation of the taxonomic content including higher levels such as phylum,

Table 2
A total number of taxonomically indexed bacterial taxa and their distribution in sarcoidosis patients and healthy controls.

Taxon	Control group	Sarcoidosis
Phylum	13	13
Family	70	70
Genus	118	102
Total number of reads	55451	6558
Number of reads per sample (SD)	2520 (293)	937 (58)

SD – standard deviation

Visual microbiome analysis of individual and small sample groups in sarcoidosis

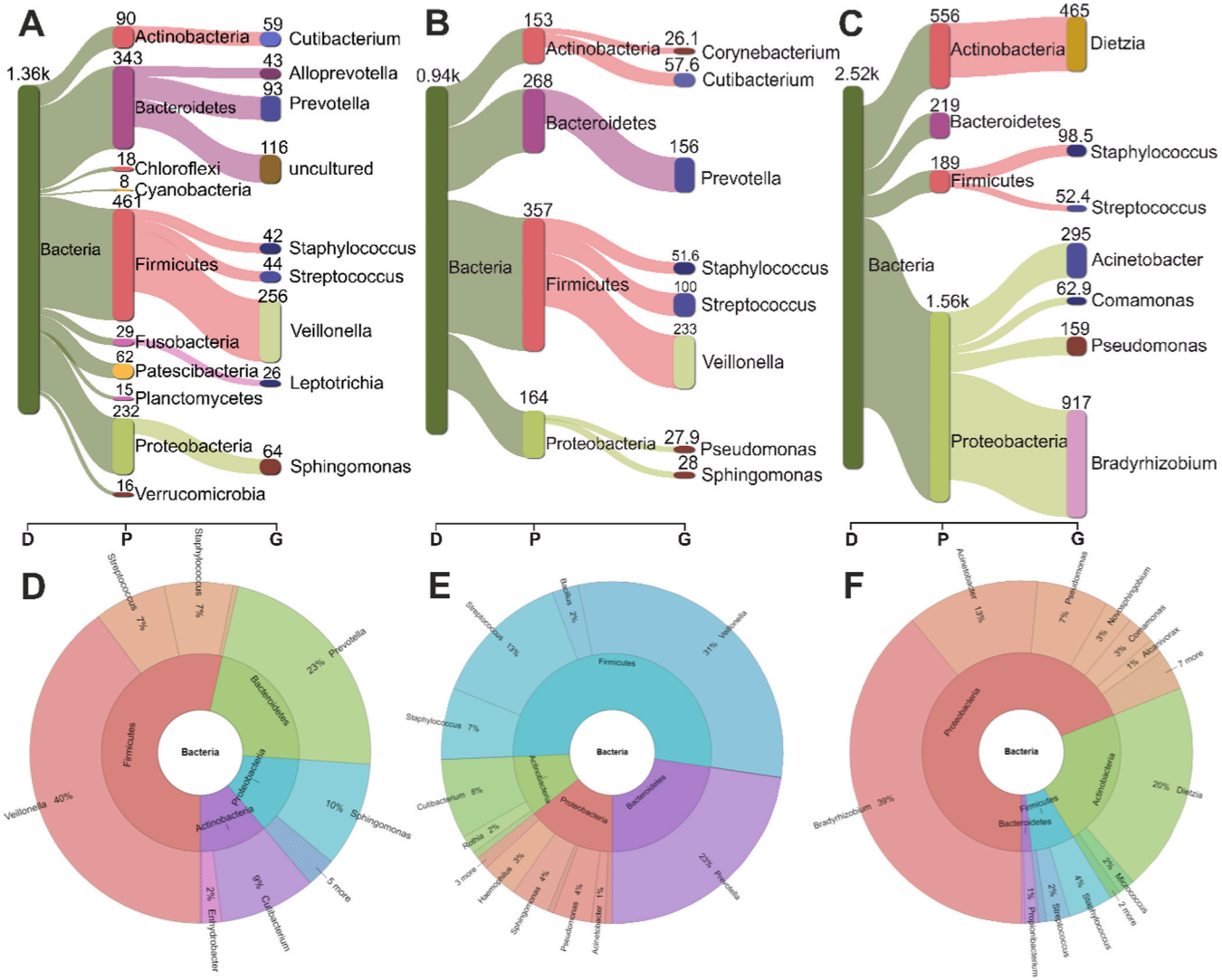

Fig. 2. Microbial composition and read abundance visualization of individual and group blood microbiomes at phylum and genus level by Sankey and Krona diagrams. (A and D) single patient with pulmonary sarcoidosis (Sarc1); (B and E) sarcoidosis group mean abundance; (C and F) healthy group mean abundance. (D – domain, P – phylum, G – genus).

and genus. The taxonomic abundance was represented by the thickness of branches in the Sankey diagram and the angular sector size in the Krona pie chart. The Sankey method displays the actual read abundance, while Krona illustrates the relative richness. Both diagrams showed four dominant phyla in similar proportions: Actinobacteria, Proteobacteria, Firmicutes, and Bacteroidetes. The sarcoidosis microbiomes revealed a set of specific genera: *Cutibacterium*, *Corynebacterium*, and *Veillonella*. Genera *Dietzia* and *Acinetobacter* were predominant among the healthy group, while *Streptococcus* was present in both groups with a slight increase observed in the sarcoidosis group.

3.1. Statistical characterization of the blood microbiome in sarcoidosis patients

The blood microbiome data of 7 patients with pulmonary sarcoidosis were statistically compared to the microbiome data of 22 healthy individuals. Statistical analysis showed a significant increase in bacterial alpha diversity richness in patients with sarcoidosis using the Shannon index (1.57 ± 0.63 vs. 2.39 ± 0.08; $F(1,28) = -4.95$; $P = 0.00015$) compared to healthy controls (Fig. 3. A). The Bray-Curtis dissimilarity index was used as an indicator of beta diversity in each subject's sample. PCoA analysis clearly illustrated a disconnect between the controls and sarcoidosis microbiome types (Fig. 2B) which accounted for 57.8% of the total variance (PERMANOVA, $F(1,28) = 10.92$; $P = 0.001$).

A heat map was generated to demonstrate the relative abundance of the dominant genera in sarcoidosis and healthy groups (Fig. 4). The results showed a clear differential taxonomic distribution of microbial genera based on the group. Furthermore, there was an increase in relative abundance of 36 microbial genera associated with sarcoidosis (located at the top right corner of the heat map diagram, Fig. 4).

Applying the Mann-Whitney test, we identified a significant increase in the microbiome abundance of five genera in sarcoidosis: *Veillonella, Prevotella, Cutibacterium, Corynebacterium,* and *Streptococcus* (Fig. 5).

4. Discussion

The blood microbiome has not been tested in sarcoidosis patients. To understand the potential microbial signatures linked to sarcoidosis, we compared the 16 S blood metagenome composition of sarcoidosis patients with that of healthy controls. Identification of microbial clusters may yield valuable information on sarcoidosis etiology. Currently, establishing a defined core blood microbiome in healthy individuals from diverse geographical regions is challenging. However, it is possible to outline indications of enriched or depleted microbial genera in

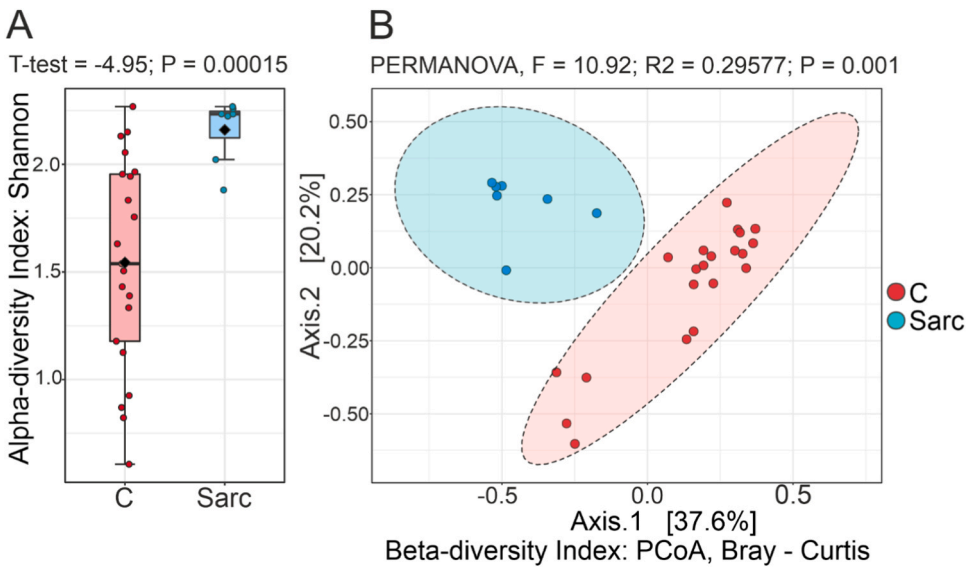

Fig. 3. Statistical comparison of the blood microbiomes. (A) Alpha diversity – median Shannon index in healthy control group and sarcoidosis group. (B) Cluster analysis of microbiome types calculated by principal coordinates analysis (PCoA), using Bray-Curtis beta diversity index, (abbreviations: C – controls, Sarc – sarcoidosis).

sarcoidosis patients.

This study characterizes the blood microbiome profiles of 7 patients with pulmonary sarcoidosis and 22 healthy individuals to evaluate strategies for visually examining individual microbiome data. We aimed to describe statistically significant alterations in the microbiota composition associated with sarcoidosis. Our results show the taxonomical composition and relative bacterial abundance for single or small sample sizes.

Sankey diagrams and Krona pie charts were proposed to visualize single-subject and small group microbiome data [20,41]. We applied both methods to illustrate the taxonomic composition in individual and group sarcoidosis patients in comparison to healthy controls (Fig. 2). Both methods demonstrated microbiome composition and the relative abundance of the core phyla - Actinobacteria, Proteobacteria, Firmicutes, and Bacteroidetes. Microbiome visualization methods can be useful in clinical practice aiding in diagnosis, monitoring disease progression, enabling personalized medicine, identifying potential pathogens and facilitating research. Thus, maximum valuable clinical data can be derived from the microbiome data of a single or a small group of patients.

We employed alpha and beta diversity measures and nonparametric univariate analysis to determine the blood microbial composition in both pulmonary sarcoidosis patients and a healthy control group. Five genera, namely *Veillonella, Prevotella, Cutibacterium, Corynebacterium,* and *Streptococcus* showed statistically significant abundancy differences in patients with sarcoidosis compared to the control group. These differences suggest the existence of sarcoidosis-specific microbial profiles, which were noticeable in individual patient data but became more apparent when analyzing data from a group of patients. Data analysis found potential microbial markers.

Monitoring the blood microbiome content allows for assessing sarcoidosis-related dysbiosis. Microbial dysbiosis can be marked by microbial species or genera that are either enriched or depleted due to disease. Recent metagenome analyses of BAL fluid [5,6,15] and tissue biopsy [12] have identified genera linked to microbial dysbiosis in lung disorders. The studies by Zimmermann et al. and Gupta et al. have provided valuable insights into the lung microbiome, studying BAL samples, associated with sarcoidosis [5,6]. Zimmermann et al. identified *Atopobium* and *Fusobacterium* as novel candidates for sarcoidosis-associated microbiota in the lung of the patients. On the other hand, Gupta et al. carried out a comparative analysis of the alveolar microbiome in various respiratory illnesses including sarcoidosis, revealing Actinobacteria and Proteobacteria to be significantly more abundant in sarcoidosis patients. In contrast, our blood microbiome results, showed a different profile in sarcoidosis patients. We observed a lower abundance of Actinobacteria and Proteobacteria, while Firmicutes were enriched. Our study validates the presence of such phyla both on an individual (Fig. 2) and group level (Fig. 4). Microbial alpha diversity was significantly higher in sarcoidosis blood samples compared to healthy individuals (Fig. 3A). The differences in microbial composition were supported by the increased beta diversity distance shown by PCoA (Fig. 3B). In sarcoidosis patients, enhanced bacterial abundance was associated with five genera - *Veillonella, Prevotella, Cutibacterium, Corynebacterium, Streptococcus*. Generally, identifying the presence of microbes in blood through DNA analysis may result in DNAemia due to microbial DNA translocating through the intestinal barrier [49]. In our study, we eliminated microbial cell-free circulating DNAs in blood during the DNA extraction process by treating the lysed blood with DNase I and washing the pellet multiple times. This process leaves the microbial cells intact for further DNA extraction. The five genera consist of human commensal or symbiotic species, as well as pathogens or species with pathogenic potential. Among the five genera, *Streptococcus* species have the highest infectious potential, while the *Cutibacterium spp.* have the highest potential for granuloma formation.

Cutibacterium acnes is a skin commensal, ubiquitously distributed among healthy individuals [5]. *Cutibacterium acnes* produces lipases which cause inflammation of the skin and other parts of the body [14, 50–52]. *Cutibacterium acnes* has also been detected in granuloma tissue via immunohistochemistry and is believed to be involved in the pathogenesis of sarcoidosis [7]. Gene homologies have been identified between *C. acnes* and other microbial species that have the potential to cause granuloma formation, such as *Yersinia pseudotuberculosis* and *Mycobacterium spp.*, through bioinformatic comparative analysis [53].

Streptococcus spp. were identified in patients with sarcoidosis [5]. Among streptococci, *S. pneumoniae* and *S. pyogenes* are the most pathogenic, but commensals with pathogenic potential have also been reported. Studies suggest that certain *Streptococcus* strains can contribute to granuloma formation by activating the immune system. It has been found that *Streptococcus spp.* can activate the Toll-like receptors (TLRs) resulting in production of pro-inflammatory cytokines that may contribute to granuloma formation [54]. *Streptococcus spp*. are known to be involved in the metabolism of sugars and amino acids and may have a

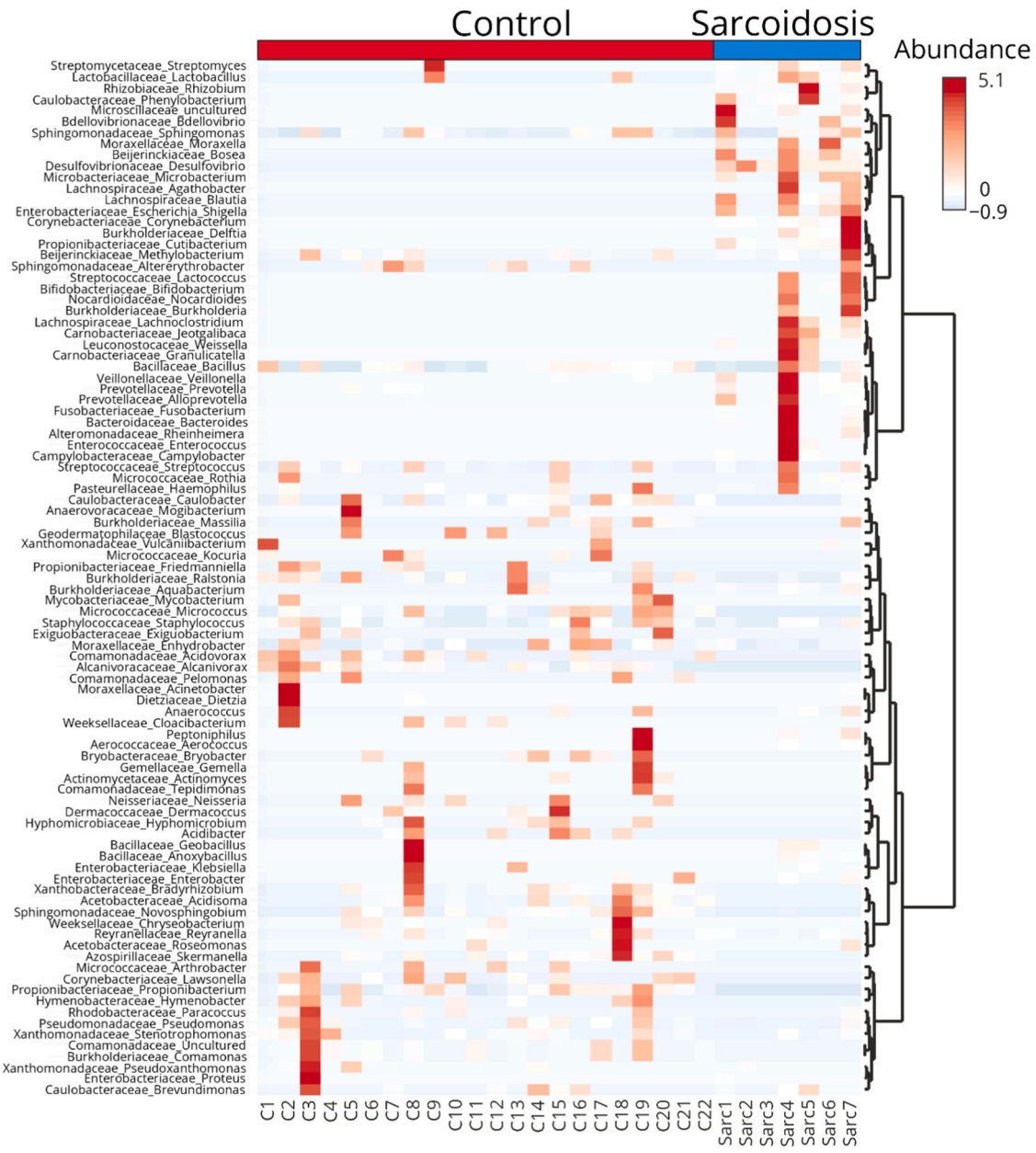

Fig. 4. Heat-map illustrating individual differences in taxonomic composition between the sarcoidosis patients (Sarc1- Sarc7) and the healthy control group (C1-C22) based on Euclidean distance and Ward's minimum variance method.

role in maintaining the balance of the blood microbiome [55].

Members of the *Corynebacterium* genus, particularly *C. diphtheria,* are lung pathogens that cause inflammation and trigger the immune response. Studies have indicated that *Corynebacterium spp.* can induce the formation granuloma by activating the NLRP3 inflammasome, which leads to the production of pro-inflammatory cytokines [56,57]. *Streptococci* and *Corynebacteria* are not commonly considered as etiological agents of sarcoidosis, but rather as companion lung infections. This notion is supported by observed enriched abundance in lung diseases other than sarcoidosis [5,6]).

Prevotella spp. and *Veillonella spp.* are mostly commensal species that have been recognized as opportunistic pathogens associated with internal and oral infections. Both genera demonstrated enriched abundance in BAL fluid of sarcoidosis patients [5,6]. Additionally, these genera have also been detected in the normal lung microbiota [58,59]. The role of *Veillonella* in granuloma formation is not yet fully understood. However, *Veillonella* species produce lactic and other fatty acids, altering the local environment's pH and causing an inflammatory response leading to granuloma formation. Certain species of *Prevotella* could be involved in sarcoidosis etiology through their contribution to granuloma formation by activating the immune system. Studies indicate that *Prevotella* species in the gut microbiome correlate with elevated pro-inflammatory cytokine levels [5,60,61]. Zimmerman et al. reported a disease-specific elevation of *Atopobium* and *Fusobacterium* genera in sarcoidosis patients [5]. This rise could be linked to unique reconfiguration of lung microbiome, but not necessarily reflected in blood.

The potential role of *Veillonella, Prevotella, Cutibacterium, Corynebacterium,* and *Streptococcus* species in sarcoidosis etiology remains unknown, and predicting their contribution is difficult. Bacteria can impact the host's immune system through two primary mechanisms - allergic inflammation or granuloma formation pathways [7].

The 7 drug naïve patients were diagnosed for the first time and were not treated for sarcoidosis with anti-inflammatory drugs. They also confirmed not having received any corticosteroid drugs for asthma that

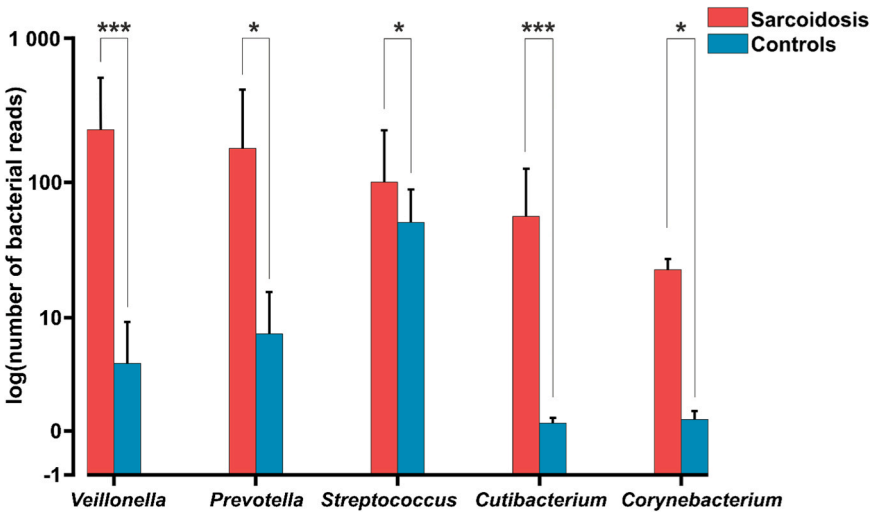

Fig. 5. Log-transformed average microbial abundance (± Standard Error of Mean) of the five genera *Veillonella, Prevotella, Cutibacterium, Corynebacterium* and *Streptococcus* specific to sarcoidosis etiology.

could potentially affect the composition of the microbiome.

The presence of over 95% of human DNA in the samples may mask the microbial content during the library amplification steps despite treatment with DNase I and washing steps included in the protocol prior to microbial DNA extraction (Materials and Methods). The amount of the microbial DNA in blood was estimated to be less than 0.2% [62]. The study's main limitation was the potential depletion of microbial DNA by human DNA.

5. Conclusion

Our study showed that Sankey and Krona diagrams can effectively illustrate the taxonomic diversity of blood microbiomes at the individual and group level. In our experience, the Sankey diagram is visually more intuitive and better suited for comparisons. Our approaches for visualizing metagenomic data provide practical value in diagnosing individual sarcoidosis cases and can be applied to studying rare lung diseases with limited access to large data sets. We observe an elevated abundance of five genera - *Veillonella*, *Prevotella*, *Cutibacterium*, *Corynebacterium*, and *Streptococcus,* but their roles in sarcoidosis and lung granuloma formation remain unclear. Additional research is mandatory to validate the outcomes and translate them into clinical practice. In the future, databases containing metagenomic, clinical and demographic data of sarcoidosis patients and healthy controls would enable the identification of distinct markers and quantitative thresholds for more effective diagnostic techniques based on statistical analysis.

Declaration of Interest

All authors notify that there's no financial/personal interest or belief that could affect our objectivity.

We state that potential competing interests don't exist.

The authors declare that generative AI and AI-assisted technologies were not used in the writing process.

Acknowledgements

This work was supported by the National Science Program VIHREN and the Bulgarian National Science Fund (BNSF) with grant number: KP-06-DV/10 - 21.12.2019.

Ethics statement

Study was approved by the Institutional review board/ Institutional ethics committee (IRB/IEC) Number IRB00006384; Protocol Number 3/ 27.07.2020).

References

[1] Baughman RP, Culver DA, Judson MA. A concise review of pulmonary sarcoidosis. Am J Respir Crit Care Med 2011;183:573–81. https://doi.org/10.1164/rccm.201006-0865CI.

[2] Tsafarova B, Hodzhev Y, Yordanov G, Tolchkov V, Kalfin R, Panaiotov S. Morphology of blood microbiota in healthy individuals assessed by light and electron microscopy. Front Cell Infect Microbiol 2023;12:1091341. https://doi.org/10.3389/fcimb.2022.1091341.

[3] Gerke AK. Treatment of sarcoidosis: a multidisciplinary approach. Front Immunol 2020;11:545413. https://doi.org/10.3389/fimmu.2020.545413.

[4] Arkema EV, Cozier YC. Epidemiology of sarcoidosis: current findings and future directions. Ther Adv Chronic Dis 2018;9:227–40. https://doi.org/10.1177/2040622318790197.

[5] Zimmermann A, Knecht H, Häsler R, Zissel G, Gaede KI, Hofmann S, et al. *Atopobium* and *Fusobacterium* as novel candidates for sarcoidosis-associated microbiota. Eur Respir J 2017;50:1600746. https://doi.org/10.1183/13993003.00746-2016.

[6] Gupta S, Shariff M, Chaturvedi G, Sharma A, Goel N, Yadav M, et al. Comparative analysis of the alveolar microbiome in COPD, ECOPD, Sarcoidosis, and ILD patients to identify respiratory illnesses specific microbial signatures. Sci Rep 2021;11:3963. https://doi.org/10.1038/s41598-021-83524-2.

[7] Eishi Y. Potential association of cutibacterium acnes with sarcoidosis as an endogenous hypersensitivity infection. Microorganisms 2023;11:289. https://doi.org/10.3390/microorganisms11020289.

[8] Kaiser Y, Eklund A, Johan Grunewald, Grunewald J. Moving target: shifting the focus to pulmonary sarcoidosis as an autoimmune spectrum disorder. Eur Respir J 2019;54:1802153. https://doi.org/10.1183/13993003.021532018.

[9] Brito-Zerón P, Pérez-Alvarez R, Pallarés L, Retamozo S, Baughman RP, Ramos-Casals M. Sarcoidosis: an update on current pharmacotherapy options and future directions. Expert Opin Pharmacother 2016;17:2431–48. https://doi.org/10.1080/14656566.2016.1258061.

[10] Crouser ED, Maier LA, Wilson KC, Bonham CA, Morgenthau AS, Patterson KC, et al. Diagnosis and detection of sarcoidosis. an official american thoracic society clinical practice guideline. Am J Respir Crit Care Med 2020;201:e26–51. https://doi.org/10.1164/rccm.202002-0251ST.

[11] Mitchell DN, Scadding JG, Heard BE, Hinson KF. Sarcoidosis: histopathological definition and clinical diagnosis. J Clin Pathol 1977;30:395–408. https://doi.org/10.1136/jcp.30.5.395.

[12] Clarke EL, Lauder AP, Hofstaedter CE, Hwang Y, Fitzgerald AS, Imai I, et al. Microbial lineages in sarcoidosis. a metagenomic analysis tailored for low–microbial content samples. Am J Respir Crit Care Med 2018;197:225–34. https://doi.org/10.1164/rccm.201705-0891OC.

[13] Suchankova M, Paulovicova E, Paulovicova L, Majer I, Tedlova E, Novosadova H, et al. Increased antifungal antibodies in bronchoalveolar lavage fluid and serum in pulmonary sarcoidosis. Scand J Immunol 2015;81:259–64. https://doi.org/10.1111/sji.12273.

[14] Yamaguchi T, Costabel U, McDowell A, Guzman J, Uchida K, Ohashi K, et al. Immunohistochemical detection of potential microbial antigens in granulomas in

the diagnosis of sarcoidosis. J Clin Med 2021;10:983. https://doi.org/10.3390/jcm10050983.
[15] Becker A, Vella G, Galata V, Rentz K, Beisswenger C, Herr C, et al. The composition of the pulmonary microbiota in sarcoidosis – an observational study. Respir Res 2019;20:46. https://doi.org/10.1186/s12931-019-1013-2.
[16] Tsafarova B. SARCOIDOSIS. Probl Infect Parasit Dis 2021;49:32–6. https://doi.org/10.58395/pipd.v49i1.55.
[17] O'Donoghue SI. Grand challenges in bioinformatics data visualization. Front Bioinforma 2021;1:669186. https://doi.org/10.3389/fbinf.2021.669186.
[18] Sudarikov K, Tyakht A, Alexeev D. Methods for the metagenomic data visualization and analysis. Curr Issues Mol Biol 2017:37–58. https://doi.org/10.21775/cimb.024.037.
[19] Kennedy ABW, Sankey HR. The thermal efficiency of steam engines. Report of the committe appointed to the council upon the subject of the definition of a standard or standards of thermal eficiency for steam engines: with an introductory note. Including appendixes and plate at back of volume. Minutes Proc Inst Civ Eng 1898; 134:278–312. https://doi.org/10.1680/imotp.1898.19100.
[20] Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics 2020;36(4):1303. https://doi.org/10.1093/bioinformatics/btz715.
[21] Platzer A, Polzin J, Rembart K, Han PP, Rauer D, Nussbaumer T. BioSankey: visualization of microbial communities over time. J Integr Bioinforma 2018;15. https://doi.org/10.1515/jib-2017-0063.
[22] Li Q, Wang C, Tang C, Zhao X, He Q, Li J. Identification and characterization of blood and neutrophil-associated microbiomes in patients with severe acute pancreatitis using next-generation sequencing. Front Cell Infect Microbiol 2018;8: 5. https://doi.org/10.3389/fcimb.2018.00005.
[23] Tourelle KM, Boutin S, Weigand MA, Schmitt FCF. Sepsis and the human microbiome. Just another kind of organ failure? a review. J Clin Med 2021;10: 4831. https://doi.org/10.3390/jcm10214831.
[24] Damgaard C, Magnussen K, Enevold C, Nilsson M, Tolker-Nielsen T, Holmstrup P, et al. Viable bacteria associated with red blood cells and plasma in freshly drawn blood donations. PLOS ONE 2015;10:e0120826. https://doi.org/10.1371/journal.pone.0120826.
[25] Domingue GJ, Schlegel JU. Novel bacterial structures in human blood: cultural isolation. Infect Immun 1977;15:621–7. https://doi.org/10.1128/iai.15.2.621-627.1977.
[26] Markova ND. L-form bacteria cohabitants in human blood: significance for health and diseases. Discov Med 2017;23:305–13.
[27] Panaiotov S, Filevski G, Equestre M, Nikolova E, Kalfin R. Cultural isolation and characteristics of the blood microbiome of healthy individuals. Adv Microbiol 2018;08:406–21. https://doi.org/10.4236/aim.2018.85027.
[28] Païssé S, Valle C, Servant F, Courtney M, Burcelin R, Amar J, et al. Comprehensive description of blood microbiome from healthy donors assessed by 16S targeted metagenomic sequencing: Blood microbiome 16S metagenomic sequencing. Transfus (Paris) 2016;56:1138–47. https://doi.org/10.1111/trf.13477.
[29] Panaiotov S, Hodzhev Y, Tsafarova B, Tolchkov V, Kalfin R. Culturable and non-culturable blood microbiota of healthy individuals. Microorganisms 2021;9:1464. https://doi.org/10.3390/microorganisms9071464.
[30] Duan H, Li X, Mei A, Li P, Liu Y, Li X, et al. The diagnostic value of metagenomic next-generation sequencing in infectious diseases. BMC Infect Dis 2021;21:62. https://doi.org/10.1186/s12879-020-05746-5.
[31] Inaoka PT, Shono M, Kamada M, Espinoza JL. Host-microbe interactions in the pathogenesis and clinical course of sarcoidosis. J Biomed Sci 2019;26:45. https://doi.org/10.1186/s12929-019-0537-6.
[32] Whittle E, Leonard MO, Harrison R, Gant TW, Tonge DP. Multi-method characterization of the human circulating microbiome. Front Microbiol 2019;9: 3266. https://doi.org/10.3389/fmicb.2018.03266.
[33] He Y, Wu W, Zheng H-M, Li P, McDonald D, Sheng H-F, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat Med 2018;24:1532–5. https://doi.org/10.1038/s41591-018-0164-x.
[34] Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Micro Inf Exp 2012;2:3. https://doi.org/10.1186/2042-5783-2-3.
[35] Statement on Sarcoidosis. Am J Respir Crit Care Med 1999;160:736–55. https://doi.org/10.1164/ajrccm.160.2.ats4-99.
[36] Miller BH, Rosado-de-Christenson ML, McAdams HP, Fishback NF. Thoracic sarcoidosis: radiologic-pathologic correlation. RadioGraphics 1995;15:421–37. https://doi.org/10.1148/radiographics.15.2.7761646.
[37] Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. Nat Protoc 2022;17: 2815–39. https://doi.org/10.1038/s41596-022-00738-y.
[38] Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Comput Sci 2017;3:e104. https://doi.org/10.7717/peerj-cs.104.
[39] Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol Microbiol 1994;44:846–9. https://doi.org/10.1099/00207713-44-4-846.
[40] Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 2017;5:27. https://doi.org/10.1186/s40168-017-0237-y.
[41] Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinforma 2011;12:385. https://doi.org/10.1186/1471-2105-12-385.
[42] Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat Protoc 2020;15: 799–821. https://doi.org/10.1038/s41596-019-0264-1.
[43] Reese AT, Dunn RR. Drivers of microbiome biodiversity: a review of general rules, feces, and ignorance. MBio 2018;9:e01294-18. https://doi.org/10.1128/mBio.01294-18.
[44] Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 2007;73:1576–85. https://doi.org/10.1128/AEM.01996-06.
[45] Bray JR, Curtis JT. An ordination of the upland forest communities of southern wisconsin. Ecol Monogr 1957;27:325–49. https://doi.org/10.2307/1942268.
[46] Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58:236–44. https://doi.org/10.1080/01621459.1963.10500845.
[47] Ling RL. A computer generated aid for cluster analysis. Commun ACM 1973;16: 355–61. https://doi.org/10.1145/362248.362263.
[48] Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol 2014;67:850–7. https://doi.org/10.1016/j.jclinepi.2014.03.012.
[49] Gargari G, Mantegazza G, Taverniti V, Del Bo' C, Bernardi S, Andres-Lacueva C, et al. Bacterial DNAemia is associated with serum zonulin levels in older subjects. Sci Rep 2021;11:11054. https://doi.org/10.1038/s41598-021-90476-0.
[50] Furusawa H, Suzuki Y, Miyazaki Y, Inase N, Eishi Y. Th1 and Th17 immune responses to viable Propionibacterium acnes in patients with sarcoidosis. Respir Investig 2012;50:104–9. https://doi.org/10.1016/j.resinv.2012.07.001.
[51] Kraaijvanger R, Veltkamp M. The role of cutibacterium acnes in sarcoidosis: from antigen to treatable trait? Microorganisms 2022;10:1649. https://doi.org/10.3390/microorganisms10081649.
[52] Yorozu P, Furukawa A, Uchida K, Akashi T, Kakegawa T, Ogawa T, et al. Propionibacterium acnes catalase induces increased Th1 immune response in sarcoidosis patients. Respir Investig 2015;53:161–9. https://doi.org/10.1016/j.resinv.2015.02.005.
[53] Cronan MR. In the thick of it: formation of the tuberculous granuloma and its effects on host and therapeutic responses. Front Immunol 2022;13:820134. https://doi.org/10.3389/fimmu.2022.820134.
[54] Ehrnström B, Beckwith KS, Yurchenko M, Moen SH, Kojen JF, Lentini G, et al. Toll-like receptor 8 is a major sensor of group b streptococcus but not Escherichia coli in human primary monocytes and macrophages. Front Immunol 2017;8:1243. https://doi.org/10.3389/fimmu.2017.01243.
[55] Willenborg J, Goethe R. Metabolic traits of pathogenic streptococci. FEBS Lett 2016;590:3905–19. https://doi.org/10.1002/1873-3468.12317.
[56] Anand PK, Malireddi RKS, Kanneganti T-D. Role of the Nlrp3 inflammasome in microbial infection. Front Microbiol 2011;2. https://doi.org/10.3389/fmicb.2011.00012.
[57] Huppertz C, Jäger B, Wieczorek G, Engelhard P, Oliver SJ, Bauernfeind F-G, et al. The NLRP3 inflammasome pathway is activated in sarcoidosis and involved in granuloma formation. Eur Respir J 2020;55:1900119. https://doi.org/10.1183/13993003.00119-2019.
[58] Garzoni C, Brugger SD, Qi W, Wasmer S, Cusini A, Dumont P, et al. Microbial communities in the respiratory tract of patients with interstitial lung disease. Thorax 2013;68:1150–6. https://doi.org/10.1136/thoraxjnl-2012-202917.
[59] Hogan DA, Willger SD, Dolben EL, Hampton TH, Stanton BA, Morrison HG, et al. Analysis of lung microbiota in bronchoalveolar lavage, protected brush and sputum samples from subjects with mild-to-moderate cystic fibrosis lung disease. PLOS ONE 2016;11:e0149998. https://doi.org/10.1371/journal.pone.0149998.
[60] Iljazovic A, Roy U, Gálvez EJC, Lesker TR, Zhao B, Gronow A, et al. Perturbation of the gut microbiome by Prevotella spp. enhances host susceptibility to mucosal inflammation. Mucosal Immunol 2021;14:113–24. https://doi.org/10.1038/s41385-020-0296-4.
[61] Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. Immunology 2017;151:363–74. https://doi.org/10.1111/imm.12760.
[62] Pietrzak B, Kawacka I, Olejnik-Schmidt A, Schmidt M. Circulating microbial cell-free DNA in health and disease. Int J Mol Sci 2023;24:3051. https://doi.org/10.3390/ijms24033051.