Dissecting and tracing the gut microbiota of infants with botulism: A cross sectional and longitudinal study

Dai Wang
Xiamen University

Kexin Li
Beihang University

Lijuan Wang
Beijing Children's Hospital Capital Medical University

Zhongqiu Teng
Chinese Center for Disease Control and Prevention

Xia Luo
Chinese Center for Disease Control and Prevention

Hui Sun
Chinese Center for Disease Control and Prevention

Ying Huang
Chinese Center for Disease Control and Prevention

Songnian Hu
Chinese academy of sciences

Xuefang Xu (✉ xuxuefang@icdc.cn)
Chinese Center for Disease Control and Prevention

Ruiting Lan
University of New South Wales

Zilong He
Beihang University

Jianguo Xu
Chinese Center for Disease Control and Prevention

Research Article

Keywords: Infantile botulism, 16S rRNA sequencing, Whole genome sequencing, Bifidobacterium, Longitudinal study

Posted Date: August 31st, 2023
Abstract

Background. Infantile botulism is caused by botulinum neurotoxin (BoNT), which is mainly produced by *Clostridium botulinum*. However, there is a lack of longitudinal cohort study on infant botulism. Herein, we have constructed a cross-sectional and longitudinal cohort of infants infected with *C. botulinum*. Our goal was to reveal the differences in the intestinal microbiota of botulism-infected and healthy infants as well as the dynamic changes over time through multi-omics analysis.

Methods. We performed 16S rRNA sequencing and whole genome sequencing of 20 infants’ stools over a period of 3 months. Through bioinformatics analysis, we focused on the changes in the infants’ intestinal microbiota as well as function over time series.

Results. We found that *Enterococcus* was significantly enriched in the infected group and declined over time, whereas *Bifidobacterium* was significantly enriched in the healthy group and gradually increased over time. All isolates carried the type B botulinum toxin gene with identical sequence. *in silico* Multilocus sequence typing found that 20 *C. botulinum* isolates from the patients were typed into ST31 and ST32.

Conclusions. Differences of intestinal microbiota and functions in infants were found with botulism through cross-sectional and longitudinal studies and *Bifidobacterium* may play a role in the recovery of infected infants.

Introduction

Infant botulism is one of the 6 forms of botulism (infant botulism, foodborne botulism, wound botulism, adult intestinal colonization, iatrogenic botulism, and inhalational botulism) [1]. Botulism is caused by botulinum neurotoxin (BoNT), which is mainly produced by *Clostridium botulinum* and rarely, by neurotoxigenic *Clostridium baratii* or *Clostridium butyricum* [2]. The illness is usually found in infants less than one year of age. *C. botulinum* colonizes the infant’s large intestine, where they germinate, reproduce, and release BoNTs into the gut. The enteric toxin is absorbed via lymphatics into the blood and transported to the neuromuscular junction, where it binds with soluble N-ethylmaleimide-sensitive factor attachment protein receptor, blocks acetylcholine release, and leads to flaccid paralysis[3, 4]. Infant botulism accounted for around 70% of botulism cases over the past decade [5] and was found across the world [6–13]. Infant botulism was firstly reported in China in 1990 [14] and more cases have been reported in recent years with better diagnostic technologies [15–17].

There are a few studies on the infection source and epidemiology of infant botulism[18, 19]. Some studies showed that *C. botulinum* spores are prevalent in soil which is the most likely source of infant botulism. A recent study in Colorado, United States found that the isolates from infant botulism cases were classified into a cluster from soil and dust, indicating a close relationship in genome sequences and a possible source from soil[18]. Infant botulism is known to affect the gut microbiota. Shirey TB et al. found significant differences in *Proteobacteria*, *Firmicutes*, and *Enterobacteriaceae* abundances in the fecal microbiota of infants with botulism compared to non-confirmed cases samples [1]. However, there
is a lack of longitudinal cohort studies on gut microbiota variation in infant botulism in short and long term, especially in gut microbiota following the recovery.

In this study, we constructed a cross-sectional and longitudinal cohort of infants infected with *C. botulinum*. Stool samples for 16S rRNA sequencing were collected from 1 to 90 days of confirmed infected infants. Our goal was to reveal the differences in the intestinal microbiota and metabolism changes of botulism-infected infants compared with healthy infants as well as the dynamic changes over time through multi-omics analysis following treatment. The aim of the study was to gain a better understanding of the changes in intestinal microbiota caused by infant botulism and during recovery, to provide better theoretical basis for the treatment of the disease.

**Materials and Methods**

**Sample collection**

This study was conducted in accordance with the guidelines of the Helsinki Declaration and Rules of Good Clinical Practice. In compliance with human subjects’ exemption protocol (SHERLL 2019057) approved by Ethics Committee of the Capital Institute of Pediatrics. Stool samples for 16S rRNA gene sequencing were collected from 20 infants who were all confirmed as BoNT gene positive by quantitative PCR from 2015 to 2021 [20]. In addition, 10 stool samples were obtained from 10 healthy infants aged from 2–8 months for 16S rRNA gene sequencing analysis. Besides, at least one botulism isolate from each patient was cultured and sequenced. Detailed meta information for 16S rRNA gene sequencing samples and isolated strains is listed in Table 1.

**DNA extraction and Sequencing**

All the stool samples for 16S rRNA sequencing were collected and frozen at -80°C after sampling. DNA was extracted using the E.Z.N.A stool DNA kit as the manufacturer's instructions for 16S rRNA sequencing. The 16S rRNA gene were amplified for PCR amplification using universal primers 341F (5'-CCTAYGGGRBGCASCAG-3') and 806R (3’-GGACTACNNNGGTATCTAAT-5’). According to the manufacturer's instructions, the library was constructed using TruSeq DNA PCR-free sample preparation kit. Each PCR reaction was comprised of 5 µL of 10 × PCR buffer, 0.5 µL of dNTP (10 mM), 0.5 µL of PCR forward primer (50 µM), 0.5 µL of reverse Primer (50 µM), 0.5 µL of Platinum Taq (5 U/µL) (Thermo Fisher Scientific, USA) and 20 ng DNA in a total volume of 50 µL. PCR conditions comprised of initial denaturation at 95 °C for 5 min, followed by 29 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 60 s and elongation at 72 °C for 60 s, with the final extension at 72 °C for 7 min. Each sample were indexed and sequenced by using the Illumina MiSeq PE-300. For isolated whole genome sequencing samples, total DNA was purified from overnight culture using Qiagen Genomic-tip 100/G columns (Qiagen, Germantown, MD, United States) under the manufacturer's instructions. The purity and integrity of the DNA were tested by agarose gel electrophoresis and quantified by Qubit. Qualified DNA samples detected by electrophoresis were randomly broken into fragments. Libraries were prepared using a
Nextera XT DNA library preparation kit (Illumina, Inc., Cambridge, UK). Whole genome sequencing (WGS) was performed using Illumina NovaSeq 6000 with PE-150 model.

**Processing of 16S rRNA sequencing data**

All 16S rRNA sequencing data were processed with the Quantitative Insights into Microbial Ecology version 2 (QIIME2) software [21]. The DADA2 plugin was used to quality control, filter chimeric sequences, and assembly reads [22]. Q2-feature-classifier plugin with the Greengene database was used to do taxonomic annotation [23]. Alpha diversity and Principal Coordinates Analysis (PCoA) based on Bray-Curtis diversity were performed by vegan and ggplot2 package in R software (version 3.6.1). The LEfSe was applied to determine the microbial taxa with significantly differential abundance between groups (LDA > 3.0) [24]. Metabolic pathways were predicted by PICRUSt2 using the KEGG database [25]. The co-abundance networks of microbial taxa were visualized by Cytoscape version 3.72.

**Processing of whole genome sequencing data**

Low quality and adapter reads were removed by trimmomatic v0.39. The reads were assembled using SPAdes v3.14.0 with the “-careful” option [26]. QUAST v4.6 and BUSCO v5 were used to conduct a quality check of the genomes and infer the annotation completeness [27, 28]. Prokka v1.13 was used to predict and annotate the genome assemblies [29]. The pan genome analysis were performed using Roary v3.13 with default parameters [30]. Multilocus sequence typing (MLST) schemes were determined using the PubMLST C. botulinum database (http://pubmlst.org/cbotulinum/). The multi-sequence alignment was performed by mafft and visualized by MEGA X [31, 32].

**Statistical analysis**

R scripts with packages ggplot2 and vegan in R v3.6.1 were used for statistical analysis and visualization. The heatmap was drawn by the pheatmap package.

**Result**

**The overview of infant cohorts and sequencing data**

Stool samples were collected from a total of 30 infants for the study (10 healthy and 20 infected infants). The 30 infants came from northern cities in China and ranged in age from 2 months to 8 months, with a sex ratio of nearly 1:1 (Table 1). To investigate the ongoing effects of C. botulinum on infant gut microbiota, we collected the fecal samples in 12 of 20 infected infants at days 1, 7, 14, 30, 60 and 90. A total of 47 longitudinal cohort samples were collected for 16S rRNA analysis. On average, each 16S sample contained 71,427 ± 11050 reads across all samples. In addition, 20 strains of C. botulinum were isolated from infected infant fecal samples for whole genome sequencing. The total size of whole genomes sequencing for strains was 26.46 G, with an average sequencing coverage of 339 ± 122.36 X per strain. Following pre-processing, all samples were saturated by genome completion assessment and ready for subsequent analysis. The specific analysis flow is shown in Figure S1.
The composition of intestinal microbiota in healthy and C. botulinum-infected infants

Based on the standard 16S data analysis protocol, 16 phyla, 24 orders, 34 orders, 54 families, 138 genera and 195 species were identified in healthy and C. botulinum-infected infants. In terms of taxonomic composition, the top 3 phyla were Actinobacteriota (37.46% mean abundance), Firmicutes (29.65%) and Proteobacteria (26.20%). At the genus level (Fig. 1A), the most abundant genera were Bifidobacterium (31.58%), Escherichia (15.87%) and Enterococcus (10.24%). In terms of diversity, differences in the overall distribution between the healthy and infected groups were found using PCoA analysis based on the Bray-Curtis dissimilarity method (PERMANOVA, P value < 0.05) (Fig. 1B). The Shannon index and the Simpson index of the healthy group were higher than those of the infected group but the difference was not statistically significant (Fig. 1C).

The classification and functional differences in intestinal microorganisms between healthy and C. botulinum infected infant cohorts

To systematically assess the differences in intestinal microbiota between healthy and infected infants, we analyzed the microbial communities with different abundance and the function of differential microorganisms. We detected 15 differential abundant microbiota at the genus level (LDA > 2), with Bifidobacterium showing the greatest degree of difference among those with higher abundance in the healthy group (LDA > 4). Gemella, Lachnoanaerobaculum and Veillonella were also different in the two groups. Besides, differential abundant microbiota with higher abundance in the infected group were Enterococcus (LDA > 4) and Lactobacillus (LDA > 3) (Fig. 2A).

In terms of tertiary metabolic functions, the two groups differed in 21 pathways (p < 0.05) (Fig. 2B). The metabolic pathways with the greatest differences were mainly amino sugar and nucleotide sugar metabolism, fructose and mannose metabolism and pyruvate metabolism, which were significantly enriched in the infected group. Compared with control group, fatty acid metabolism, cysteine and methionine metabolism and nicotinate and nicotinamide metabolism decreased in infection group.

In addition, we constructed microbial co-abundance networks for the healthy group and the infected group of infants respectively (Fig. 2C and 2D). We found that the network in the healthy group was more tightly linked than the network in the infected group, with more connections between nodes. In the healthy group, the degree of all the differential abundant microbiota was less than or equal to 3, except for Veillonella, which had a linkage greater than 5. Flavonifractor (degree = 8) was the most linked taxon in the healthy group network (Fig. 2C). The overall degree of the infected group network was not high. The most linked taxa in the infected group network were Dysgonomonas, Bilophila, and Dialister, all of which had a linkage of 5 (Fig. 2D).

Longitudinal analysis reveals the dynamics of key microbiota from C. botulinum infected infants

In the differential abundant microbiota analysis described above, we found that Bifidobacterium and Enterococcus differed most in the two cohorts. To examine the temporal dynamics of these two groups,
we followed up some of the infected infants (12/20) and collected stool samples for longitudinal study. The cross-sectional 16S data analysis also showed significant differences in *Enterococcus* (P < 0.05) (Fig. 3A and B). The overall abundance of *Bifidobacterium* increased significantly after 30 days and the dispersion throughout decreased, whereas *Enterococcus* decreased rapidly after 14 days and was almost undetectable after 30 (Fig. 3C and D). We also compared the samples of patients at the initial stage of infection (Infection group) with the samples of patients at 60 days and 90 days after disease diagnosis (Recovery group). We found that the differential abundant bacteria between the two groups were also different between the infected group and the healthy group, with the exception of *Bifidobacterium* (Figure S2). Secondary metabolic pathway analysis revealed that the microbial-related metabolic pathways like Glycan biosynthesis and metabolism, amino acid metabolism, metabolism of cofactors and vitamins and biosynthesis of other secondary metabolites in the infected group gradually returned to the levels of the control group over time (Fig. 3E).

**Genomic analysis of C. botulinum isolates from infected infants**

The 20 *C. botulinum* isolates from the stool samples of the 20 infected infants with one isolate per case were sequenced using Illumina sequencing. The average genome length ranged from 3.8 to 4.3M, with an average N50 of around 846 kb (Table S1). All isolates carried the botulinum toxin gene with identical nucleotide sequence encoding the type B botulinum toxin (Figure S3). *in silico* MLST typing found that the 20 *C. botulinum* isolates from the patients were typed into two sequence types (ST), ST31 and ST32 with 10 isolates each. KEGG enrichment analysis based on the core genome of 20 strains of *C. botulinum* showed that ABC transporters and two-component system pathways had the highest gene enrichment degree (Figure S4).

**Discussion**

The initial sterile gut in the infant is an oxidized environment favorable to be colonized by facultative aerobes such as *Lactobacillus*, *Prevotella* and *Sneathia* sp. With less oxygen in the gut, it becomes more suitable for anaerobic bacteria to grow[33]. Therefore, *C. botulinum* infection mostly occurs in infants aged 0 to 8 months. Because the intestinal microbiota of infants is not stable, it is likely to carry spores into the gut through mouth, thus causing poisoning. At present, there are few studies on intestinal microbiota of infants with botulism. Only one study in the US reported intestinal microbial profiling of 14 infected infants [1]. Most botulinum toxin studies are associated with strains isolated from humans or the environment, but there is a lack of cohort comparison in cross sectional and longitudinal analysis of intestinal microbiota in infant botulism. In this study, we first collected the fecal samples of infants infected with *C. botulinum* from China. In addition, some of the isolates were isolated from infant feces and the surrounding environment. In order to study the dynamic changes of intestinal microbiota in infants infected with *C. botulinum* in time-scale, we tracked and collected fecal samples of some infants at multiple time points. The strategy of cross sectional and longitudinal study based on metagenome and whole genome sequencing of bacteria may help us better understand the infection process and mechanism of botulism.
Shirey et al. examined 14 infant fecal samples including 8 botulism samples and 6 non-confirmed which were used as control [1]. Due to the short of metadata of the samples, the non-confirmed samples might not be strictly healthy controls. In this study, we investigated 20 botulism samples which were not only confirmed by PCR but also by strain isolation. 10 healthy infant fecal samples were collected as control covering all the age spectrum of the confirmed samples. Shirey et al. illustrated that the abundance of Enterobacteriaceae in infant intestinal microbiota of botulism is significantly increased, which is consistent with our findings [1]. We found that the abundance of Escherichia and Enterococcus were all increased in the infected group compared with the healthy group. In the time scale analysis, we also found that the abundance of Enterococcus in the infected group gradually decreased after 14 days until it could not be detected, indicating that Enterococcus may be involved in the occurrence of the infection. Interestingly, it was also reported an increase in Enterococcus in children infected with Clostridioides difficile [34]. In addition, the LDA score of Bifidobacterium was very high in the healthy group compared with the infected group. It is well known that Bifidobacterium, as a probiotic, plays an important role in the regulation of intestinal microbiome [35]. Many studies have reported that Bifidobacterium plays a positive role in promoting intestinal health [36–38]. In our study, the abundance of Bifidobacterium in the healthy group was higher than that in the infected group (LDA > 4). The time scale analysis showed that the abundance of Bifidobacterium increased significantly after 30 days in the infection group, indicating that the abundance of Bifidobacterium decreased after the initial infection, and increased during the recovery of infection. This phenomenon is consistent with previous reports indicated that infected children progressively return to a “healthy microbiota status” [39, 40]. Hence, it may be indicated that the supplement of Bifidobacterium during the infection process may help the treatment of the infection. Furthermore, metabolic changes caused by microflora changes were also studied here. Amino acid metabolism including arginine, proline, cysteine and methionine were altered. We know that high protein concentration was needed in the growth of C. botulinum. Therefore, we infer that these amino acid changes are related to C. botulinum growth.

In this study, genomic analysis of C. botulinum isolates identified two STs, neither of which has been reported in previous studies. In terms of botulinum toxin identification, all the strains were found to carry the botulinum toxin gene. Interestingly, all the botulinum toxin genes identified were type B and the protein sequences were highly consistent, suggesting that the botulinum toxins prevalent in northern China are likely to be closely related as previously reported[41, 42].

In this project, we used a strategy of 16S sequencing and whole genome sequencing of single bacteria isolates to study C. botulinum infection. Unfortunately, 16S rRNA sequencing failed to obtain OTUs of C. botulinum, which may be due to the resolution of the 16S segments and thus the inability to accurately quantify the abundance of C. botulinum. Given the low microbial content of infant feces, the full-length 16S rRNA sequencing based on the third-generation sequencing may be more suitable. This technology can pinpoint the species-level abundance of most microorganisms without requiring much raw DNA volume compared to normal 16S rRNA sequencing, which may be a better choice for the study of intestinal microbes in infants in the future. In addition, a combined strategy of full-length 16S sequencing and whole genome sequencing is also recommended.
Conclusion

In conclusion, we have systematically analyzed the differences in gut microbes and functions in botulism-affected infants through cross-sectional and longitudinal studies. We found that *Enterococcus* was significantly enriched in the infected group and declined over time, whereas *Bifidobacterium* was significantly enriched in the healthy group and gradually increased over time following infection in infected infants. Pathway analysis revealed that metabolism of cofactors and vitamins, and amino acid metabolism gradually returned to normal levels over time. The findings provided a better understanding of the changes in intestinal microbiota caused by infant botulism and during recovery and a better theoretical basis for the management of the disease.

Declarations

**Ethics approval and Consent to participate**

This study was conducted in accordance with the guidelines of the Helsinki Declaration and Rules of Good Clinical Practice. In compliance with human subjects’ exemption protocol (SHERLL 2019057) approved by Ethics Committee of the Capital Institute of Pediatrics.

**Consent for publication**

Not applicable.

**Availability of data and materials**

Sequencing data were deposited in NCBI with accession numbers and PRJNA908234 and PRJNA931660.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This work was supported by the NSFC grants (82272349/81971905), Shenzhen Science and Technology Program (JCYJ20220530143409021) to D.W.

**Author contributions**

DW, KXL, LJW, ZQT, XL, HS, HY, XFX, ZLH, and RL performed the experiments, analyzed the data and wrote the manuscript. DW, XFX, ZLH, RL and JGX conceptualized and designed the study. LJW, XL, HS, YH, and SNH provided material and samples. All authors reviewed and edited the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**
We thank Xiushan Ge for providing the infant stool samples.

**Authors' information**

Not applicable.

**References**


Tables

Table 1 is available in the Supplementary Files section.

Figures
Figure 1

Diversity analysis based on 16S rRNA analysis. (a) Taxonomic classification at the genus level between control and infection groups (day1). (b) Principal coordinate analysis plot between control and infection groups based on Bray Curtis of bacterial communities (PERMANOVA, FDR, p < 0.05). (c) Shannon and Simpson indexes between Control (red) and Infection (blue) samples.
Figure 2

**Bacteria analysis and functional categories of the bacterial communities predicted by PICRUSt2 analysis.**

(a) Lefse analysis at the genus level between control and infection groups (Wilcoxon test, FDR, \(p < 0.05\)). (b) Functional categories of the differentiated pathways at KEGG level 3 between control and infection groups (Welch’s t-test, \(P < 0.05\)). The bars in the graph represent the mean proportion (%) of the functional categories. The 95% confidence intervals reflect the difference in mean proportions (%), and corrected \(P\) values are displayed. (c-d) Co-abundance networks of genera between control (c) and infection (d) groups. The red dots represent the differentially abundant genera for each group. The blue dots represent intestinal genera that were not differentially abundant genera. The size of the dots represents the degree. The links represents the interactions of genera. The correlations with \(r > 0.75\) and \(P < 0.05\) are shown.
Figure 3

Analysis of significant bacteria and functional categories in different groups. (a) *Bifidobacterium* in Control and Infection groups. (b) *Enterococcus* in Control and Infection groups. (c) Changing of *Bifidobacterium* in day 1, 7, 14, 30, 60, 90. (d) Changing of Enterococcus in day 1, 7, 14, 30, 60, 90. (e) Heatmap of the relative abundances of pathways in different groups predicted by PICRUSt2 analysis at KEGG level 3 based on z-scores.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.xlsx
- FigureS1.pdf
- FigureS2.pdf
- FigureS3.pdf
- Figure4.pdf
- TableS1.xlsx