High-throughput sequencing of 16S rDNA amplicons characterizes bacterial composition in bronchoalveolar lavage fluid in patients with ventilator-associated pneumonia

Xiao-Jun Yang¹,*, Yan-Bo Wang²,3,*, Zhi-Wei Zhou¹,*, Guo-Wei Wang², Xiao-Hong Wang¹, Qing-Fu Liu¹, Shu-Feng Zhou⁴, Zhen-Hai Wang²,3

¹Department of Intensive Care Unit, ²Neurology Center, General Hospital of Ningxia Medical University, Yinchuan, Ningxia, People’s Republic of China; ³Key Laboratory of Brain Diseases of Ningxia, Yinchuan, Ningxia, People’s Republic of China; ⁴Department of Pharmaceutical Sciences, College of Pharmacy, University of South Florida, Tampa, FL, USA

*These authors contributed equally to this work

Correspondence: Zhen-Hai Wang Neurology Center, General Hospital of Ningxia Medical University, Key Laboratory of Brain Diseases of Ningxia, 804 Shenglí Street, Yinchuan, Ningxia 750004, People’s Republic of China Tel +86 951 6744 205 Fax +86 951 4082 981 Email 13995089189@163.com

Shu-Feng Zhou Department of Pharmaceutical Sciences, College of Pharmacy, University of South Florida, 12901 Bruce B Downs Boulevard, Tampa, FL 33612, USA Tel +1 813 974 6276 Fax +1 813 905 9885 Email szhou@health.usf.edu

Abstract: Ventilator-associated pneumonia (VAP) is a life-threatening disease that is associated with high rates of morbidity and likely mortality, placing a heavy burden on an individual and society. Currently available diagnostic and therapeutic approaches for VAP treatment are limited, and the prognosis of VAP is poor. The present study aimed to reveal and discriminate the identification of the full spectrum of the pathogens in patients with VAP using high-throughput sequencing approach and analyze the species richness and complexity via alpha and beta diversity analysis. The bronchoalveolar lavage fluid samples were collected from 27 patients with VAP in intensive care unit. The polymerase chain reaction products of the hypervariable regions of 16S rDNA gene in these 27 samples of VAP were sequenced using the 454 GS FLX system. A total of 103,856 pyrosequencing reads and 638 operational taxonomic units were obtained from these 27 samples. There were four dominant phyla, including Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes. There were 90 different genera, of which 12 genera occurred in over ten different samples. The top five dominant genera were Streptococcus, Acinetobacter, Limnohabitans, Neisseria, and Corynebacterium, and the most widely distributed genera were Streptococcus, Limnohabitans, and Acinetobacter in these 27 samples. Of note, the mixed profile of causative pathogens was observed. Taken together, the results show that the high-throughput sequencing approach facilitates the characterization of the pathogens in bronchoalveolar lavage fluid samples and the determination of the profile for bacteria in the bronchoalveolar lavage fluid samples of the patients with VAP. This study can provide useful information of pathogens in VAP and assist clinicians to make rational and effective therapeutic decisions.

Keywords: bioinformatics, drug resistance, OTU, PCR, DNA sequencing

Introduction

Increasing evidence shows that ventilator-associated pneumonia (VAP), a type of nosocomial pneumonia, has emerged as an important challenge in the intensive care unit (ICU) that lead to a considerable morbidity and mortality, placing a substantial heavy burden on an individual and society.¹ VAP occurs in patients who receive mechanical ventilation and is usually acquired in the hospital setting ~48–72 hours after mechanical ventilation.² The overall rate of VAP is 13.6 per 1,000 ventilator-day, according to the International Nosocomial Infection Control Consortium (http://www.injcc.org/english/index.php). The incidence of VAP varies within a range from 13–51 per 1,000 ventilation-day according to the patient group and hospital setting; the mean duration of occurrence of VAP is ~5–7 days. The VAP mortality ranges from 24%–76%,...
and there is even higher mortality rate among critically ill patients. In USA, VAP accounts for over 25% of all ICU-acquired infections, and there are over 100,000 cases annually. In addition, VAP represents more than one-half of all antibiotic use in the ICU.

The pathogenesis of VAP is complex but typically relates with colonization of the aerodigestive tract with pathogens, micro-aspiration of contaminated secretions, and formation of biofilms. In particular, there are several bacteria including Pseudomonas aeruginosa, Klebsiella pneumoniae, Serratia marcescens, Enterobacter, Citrobacter, Stenotrophomonas maltophilia, Acinetobacter, Burkholderia cepacia, and Methicillin-resistant Staphylococcus aureus, that can cause drug resistance in the treatment of VAP. These pathogens make the VAP treatment more challenging. Recently, the commonly used strategies to reduce the development of drug resistance in the treatment of VAP include de-escalation therapy, truncated courses of antibiotics, dosing regimens that account for patient-antibiotic pharmacokinetics and pharmacodynamics, antibiotic cycling, and surveillance cultures. However, the approaches for pathogen monitoring and to avoid drug resistance are still limited, and VAP still is a life-threatening infectious disease with challenges in the diagnosis and poor prognosis. Thus, it needs advanced methods to detect the full array of pathogens in patients with VAP to avoid the compromise of therapy and achieve maximum therapeutic outcome in clinical practice.

Growing evidence shows that the accurate and comprehensive identification of the causative pathogen can facilitate the VAP therapy in clinical practice. 16S rDNA gene sequencing approach possesses the capability of fast and accurately revealing the identity of the pathogens, because it can overcome the limitations of conventional culture-based bacterial detection method. With advances in the sequencing technology, the feasibility of 16S rDNA analysis using 454 GS FLX system has already been proven in the research of microbiota in the oral cavity, wound, urine, and gastrointestinal tract; and the massive data generated by 454 GS FLX system make it possible to analyze the diversity of the bacterial communities. The employment of the 16S rDNA gene sequencing approach can provide a global spectrum of the composition of the pathogens in samples, which will have a great clinical importance in the optimization of the therapy, leading to the maximum therapeutic outcome.

In the present study, in order to reveal the full array of the pathogens of VAP and help the optimization of VAP therapy, we explored the complexity of the bacterial communities in bronchoalveolar lavage fluid samples of patients with VAP using 16S rDNA amplicon 454 pyrosequencing.

Materials and methods

Patients

A total number of 27 patients from the department of ICU in the General Hospital of Ningxia Medical University were enrolled. All patients were diagnosed with VAP according to the clinical criteria provided by the China Health Ministry Guidelines. Clinical signs and symptoms of VAP included fever or low body temperature, new purulent sputum, and hypoxemia (Tables 1 and 2). Consent forms were obtained from all enrolled patients with VAP. The protocol was approved by the Ethics Committee of the General Hospital of Ningxia Medical University. All procedures were conducted in accordance with the criteria of the Declaration of Helsinki.

Clinical specimens and clinical laboratory work

The bronchoalveolar lavage fluid samples were collected from the enrolled patients. The collected bronchoalveolar lavage fluid samples were aliquoted into sterile Eppendorf tubes. Following the collection, the bacterial culture was performed. The remaining samples were stored at −80°C for subsequent metagenomic sequencing. The sample collection and transportation were carried out in strict accordance with the sterile operating procedures to avoid contamination.

DNA extraction from the bronchoalveolar lavage fluid samples

The bacterial DNA was isolated from bronchoalveolar lavage fluid samples using QIAamp DNA Micro Kit (Qiagen NV, Venlo, the Netherlands) according to the manufacturer’s instructions. Briefly, 2 mL of bronchoalveolar lavage fluid samples was required for the DNA extraction. The samples were centrifuged at 8,000 rpm for 15 minutes, and the supernatant was removed. The pellet was resuspended and digested using 200 µL of 20 mg/mL lysozyme at 37°C for 1 hour. Following the digestion, the supernatant was collected via centrifugation, and the sample was then resuspended in 200 µL PBS buffer containing 20 µL of proteinase K. Then, a quota of 200 µL AL was added to each diluted sample and incubated at 56°C for 2 hours. The resultant bacterial and human DNA were collected and transferred to sterilized 1.5 mL EP tubes for further extraction using the QIAamp DNA Micro Kit, and the DNA purification was performed according to the manufacturer’s instructions. The DNA
### Table 1 The clinical data of 27 patients with severe pneumonia

| Sample ID | Sex | Age (yr) | Main symptom            | Peripheral blood* | APACHE II | CPIS | Hospital stay (days) | Prognosis |
|-----------|-----|----------|--------------------------|-------------------|-----------|------|----------------------|-----------|
| T1        | M   | 62       | Dyspnea, wheeze          | 5.50 89.0         | 17        | 5    | 15                   | Upturn    |
| T2        | F   | 73       | Dyspnea, fever           | 8.59 79.5         | 15        | 5    | 31                   | Upturn    |
| T3        | M   | 68       | Dyspnea, fever           | 10.88 89.7        | 14        | 5    | 9                    | Upturn    |
| T4        | M   | 73       | Unconsciousness          | 9.15 82.8         | 17        | 1    | 8                    | Upturn    |
| T5        | M   | 61       | Dyspnea, fever           | 12.09 92.0        | 21        | 8    | 5                    | Upturn    |
| T6        | M   | 76       | Dyspnea, wheeze          | 12.78 82.0        | 29        | 9    | 10                   | Upturn    |
| T7        | M   | 72       | Dyspnea, fever           | 13.02 94.1        | 20        | 6    | 12                   | Upturn    |
| T8        | F   | 70       | Dyspnea                  | 8.83 94.1         | 26        | 6    | 5                    | Death     |
| T9        | F   | 57       | Unconsciousness, fever   | 19.38 89.7        | 20        | 7    | 10                   | Death     |
| T10       | M   | 70       | Dyspnea, fever           | 2.42 94.0         | 20        | 7    | 42                   | Upturn    |
| T11       | M   | 76       | Dyspnea, fever           | 12.54 95.6        | 15        | 8    | 12                   | Upturn    |
| T12       | M   | 62       | Dyspnea, fever           | 8.11 85.4         | 22        | 8    | 22                   | Upturn    |
| T13       | F   | 37       | Dyspnea                  | 9.66 93.2         | 7         | 5    | 17                   | Upturn    |
| T14       | F   | 52       | Dyspnea, fever           | 12.45 86.5        | 16        | 6    | 15                   | Upturn    |
| T15       | F   | 67       | Unconsciousness          | 7.58 92.2         | 21        | 5    | 9                    | Upturn    |
| T16       | M   | 70       | Unconsciousness          | 2.18 83.0         | 24        | 6    | 13                   | Upturn    |
| T17       | M   | 67       | Dyspnea, fever           | 4.15 91.3         | 17        | 5    | 10                   | Upturn    |
| T18       | F   | 65       | Dyspnea, fever           | 20.59 89.6        | 17        | 9    | 9                    | Death     |
| T19       | M   | 61       | Dyspnea, fever           | 17.39 83.2        | 12        | 9    | 16                   | Upturn    |
| T20       | M   | 61       | Unconsciousness          | 9.78 87.0         | 20        | 5    | 15                   | Upturn    |
| T21       | M   | 86       | Dyspnea                  | 9.98 93.9         | 20        | 7    | 5                    | Death     |
| T22       | M   | 81       | Dyspnea                  | 13.85 87.8        | 19        | 7    | 15                   | Death     |
| T23       | M   | 47       | Dyspnea, fever           | 17.60 86.2        | 25        | 8    | 24                   | Upturn    |
| T24       | F   | 47       | Dyspnea                  | 18.33 90.8        | 11        | 8    | 7                    | Upturn    |
| T25       | M   | 69       | Dyspnea                  | 13.41 88.8        | 14        | 6    | 2                    | Death     |
| T26       | M   | 40       | Dyspnea, fever           | 15.58 85.4        | 8         | 9    | 27                   | Upturn    |
| T27       | M   | 30       | Dyspnea, fever           | 18.09 77.2        | 20        | 10   | 46                   | Upturn    |

Notes: *Peripheral blood test before hospitalization. **White blood cells count of peripheral blood; *neutrophils count of peripheral blood.

Abbreviation: CPIS, clinical pulmonary infection score.
Table 2 Simplified version of the clinical pulmonary infection score used in this study

| Component                  | Value       | Points |
|----------------------------|-------------|--------|
| Temperature °C             | ≥36.5 and ≤38.4 | 0      |
|                            | ≥38.5 and ≤38.9 | 1      |
|                            | ≥39.0 and ≤36.0 | 2      |
| Blood leukocytes per mm³   | ≥4,000 and ≤11,000 | 0      |
|                            | <4,000 or >11,000 | 1      |
| Tracheal secretions        | Few         | 0      |
|                            | Moderate    | 1      |
|                            | Large       | 2      |
|                            | Purulent    | 1      |
| Oxygenation PaO₂/FiO₂ mmHg | >240 or presence of ARDS | 0      |
|                            | ≤240 and absence of ARDS | 2      |
| Chest radiograph           | No infiltrate | 0      |
|                            | Patchy or diffuse infiltrate | 1      |
|                            | Localized infiltrate | 2      |

Abbreviation: ARDS, acute respiratory distress syndrome.

The design of barcoded primers

The polymerase chain reaction (PCR) enrichment of the 16S rDNA V3-V5 hypervariable region was performed using the primers as shown in Table 3. The primers contained a ten-base oligonucleotide tag at the 5′ terminal. The sequence after the hyphen was able to pair with the sequences of the end region. The 27 pairs of primers that contained 27 different ten-base oligonucleotides and identical following sequences were used to lead the PCR enrichment (Table 4). The barcoded primers were synthesized by the Shanghai Sangon Biological Engineering Technology & Service Co., Ltd. (Shanghai, People’s Republic of China).

PCR enrichment of the V3–V5 region

The extracted and purified DNA from 27 bronchoalveolar lavage fluid samples was used for PCR assay. Each PCR reaction system consisted of 12.5 μL PCR master Mix (Dream Taq PCR Master Mix, Fermentas, Burlington, Canada), 0.5 μL forward primer, 0.5 μL reverse primer, 1 μL DNA, and 10.5 μL nuclease-free water. Touchdown PCR conditions were comprised of initial denaturation for 5 minutes at 95°C, denaturation for 30 seconds at 95°C, annealing for 30 seconds at 56°C, and extension for 30 seconds at 72°C. The resultant products were stored at −20°C after the reaction. The pooled tagged single-stranded pyrosequencing library underwent emPCR and pyrosequencing using a Roche 454 GS FLX Pyrosequencer (Hoffman-La Roche Ltd., Basel, Switzerland) according to the manufacturer’s instructions.

Bioinformatic analysis

The high-throughput pyrosequencing reads from 27 bronchoalveolar lavage fluid samples were reassigned to samples according to barcodes. The resultant sequences were clustered into operational taxonomic units (OTUs). The OTUs that reached 97% similarity level were used for alpha diversity analysis that analyzed the species diversity in the single sample by the evaluation of Chao, abundance-based coverage estimators (ACE), Shannon, and Simpson parameters; and the rarefaction curve was also analyzed using the Mothur software v1.27.0 program.18–21 Following the alpha diversity analysis, taxonomy-based analyses were performed through the classification of each sequence using the Naïve Bayesian classifier program in Ribosomal Database Project (RDP) at the Center for Microbial Ecology in Michigan State University (http://rdp.cme.msu.edu; MI, USA).20,21 The confidence level was of 95%. The sequences were assigned until the genus level in bacteria domain was collected and screened from 27 bronchoalveolar lavage fluid samples.18–21 Each read was assigned a phylum, class, family, and genus. The taxonomic assignment was unambiguous within an 80% confidence threshold, which has been estimated to taxonomically assign reads with over 98% accuracy at genus level. Furthermore, the beta diversity analysis was performed to assess the distribution and content of bacteria and evaluate the total diversity in different samples based on the bacterial profile. Sequences were clustered at 97% nucleotide identity over 90% sequence alignment length using the Mothur software. For this analysis, sequences over 97% identical were considered to correspond to the same OTUs, representing a group of reads that likely belong to the same species.22

Results

High-throughput sequencing reveals 103,856 pyrosequencing reads

In order to identify the pathogens, we first performed the high-throughput sequencing to examine the possible bacterial DNA sequences in 27 bronchoalveolar lavage fluid samples from patients with VAP. There were a total of

Table 3 The sequences of the primers for 16S rDNA V3-V5 hypervariable region

| 16S rDNA | Forward: 5′-CCATCTCATTCCGGTGCTCCTCCGACTCAGAGAGTGCCTCGCTCGTACTTCTTACGGTAC-3′ |
|          | Reverse: 5′-CCATCTCATTCCGGTGCTCCTCCGACTCAGAGAGTGCCTCGCTCGTACTTCTTACGGTAC-3′ |
103,856 pyrosequencing reads that were identified through 454 pyrosequencing (Table 5). The sequences with insufficient quality or sequences that could not be adequately assigned were not included, such as chimera sequences and a small amount of non-target sequences. For the identified pyrosequencing reads, the average length of the sequences was 550 bp after trimming the primers. Taken together, the high-throughput sequencing approach shows a capability of identifying the bacterial DNA sequences from bronchoalveolar lavage fluid samples from patients with VAP, which may be clinically helpful for the treatment of VAP.

Classification of the bacteria found in bronchoalveolar lavage fluid samples from patients with VAP

We next classified the obtained pyrosequencing reads using the RDP classifier at a confidence level of 95% and assigned taxonomic classifications to the sequences for biological analysis. The tag sequences of the identified DNA in 27 bronchoalveolar lavage fluid samples from patients with VAP were analyzed using the RDP to annotate species. The data showed that most of the bacterial reads were assigned to genus level and a small number of bacterial reads were assigned to species level (Figures 1 and 2). There were 638 OTUs

Table 4 The barcode sequences of 27 pairs of primers

| Sample ID | Barcode sequence | Forward | Reverse |
|-----------|------------------|---------|---------|
| T1        | ACGAGTGCGT       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T2        | ACGCTCGACA       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T3        | AGACGCACC       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T4        | AGCAGCTTGA       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T5        | ATCAGACAG       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T6        | ATATCGCGGAG     | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T7        | CGTCTGCTCTA     | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T8        | CTCCGGTGTC       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T10       | TCTCTATGCG      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T12       | TACCTGAGCTA     | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T13       | CATAGTAGT       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T14       | CGAGAGATAC      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T15       | ATACGACGTA      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T16       | TCACGTACTA      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T17       | CGCTAGTAC       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T18       | TCTACGTAGC      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T20       | ACGACTACAG      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T22       | TACAGTATG       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T23       | TACCTCCTGTG     | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T24       | TACCTGAGCTA     | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T25       | CATAGTAGT       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T26       | CGAGAGATAC      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T27       | ATACGACGTA      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T28       | TCACGTACTA      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T29       | CGCTAGTAC       | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T30       | TCTACGTAGC      | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |
| T31       | TGACTACTCTC     | CGTCAATTCTTTGAGTTT | ACTCTACGGGAGGCACGAG |

Table 5 Sampling depth found by 454 sequencing from 27 bronchoalveolar lavage fluid samples

| Sample ID | Raw tag number | Final tag number |
|-----------|----------------|-----------------|
| T1        | 2,539          | 1,163           |
| T2        | 2,896          | 1,793           |
| T3        | 3,900          | 1,284           |
| T4        | 3,122          | 1,094           |
| T5        | 3,567          | 1,120           |
| T6        | 3,549          | 1,735           |
| T7        | 3,129          | 1,328           |
| T8        | 3,203          | 939             |
| T10       | 3,077          | 1,398           |
| T12       | 3,304          | 1,019           |
| T13       | 3,636          | 944             |
| T14       | 3,017          | 1,599           |
| T15       | 3,106          | 811             |
| T16       | 3,230          | 995             |
| T17       | 2,528          | 1,467           |
| T18       | 2,671          | 1,257           |
| T20       | 2,590          | 1,098           |
| T22       | 2,352          | 967             |
| T23       | 2,559          | 1,184           |
| T24       | 7,024          | 3,258           |
| T25       | 6,555          | 2,066           |
| T26       | 5,764          | 2,155           |
| T27       | 5,012          | 1,647           |
| T28       | 5,882          | 2,509           |
| T29       | 4,411          | 1,578           |
| T30       | 4,838          | 2,745           |
| T31       | 5,195          | 2,716           |
that were obtained from the 27 bronchoalveolar lavage fluid samples from patients with VAP using Mothur, according to the tag annotation information of species for OTU comments. The Mothur analysis results are able to comment on the classification level under each OTU number (Figure 2), the classification level of each OTU ratio was phylum, class, order, family, genus, and species at 91.4%, 84.3%, 79.0%, 72.7%, 58.2%, and 31.5%, respectively.

Species richness and complexity estimation of microbiota in the bronchoalveolar lavage fluid samples from patients with VAP

Following the classification of the obtained pyrosequencing reads, the species richness complexity were analyzed using alpha diversity analysis. We performed bioinformatic analysis of the large number of pyrosequencing reads to evaluate the species richness and diversity of microbiota in the bronchoalveolar lavage fluid samples from patients with VAP. Based on the OTU data, we calculated alpha diversity. The indices of bacterial richness and diversity of OTUs at a 3% sequence dissimilarity level were summarized in Table 6. The bronchoalveolar lavage fluid sample containing low number of OTU indicated that it had a relatively low diversity in the cerebrospinal fluid sample compared with other environmental species (Figure 3). The Chao and ACE values and the Shannon and Simpson index were calculated, and the results showed that Chao value was ~80, ACE value was ~130, Shannon index was 1.25, and Simpson index was 0.46 (Figure 3). Moreover, the richness of bacterial communities in each bronchoalveolar lavage fluid sample was estimated as presented by the rarefaction curve. The trend of the rarefaction curves also confirmed that there was a low richness in each bronchoalveolar lavage fluid sample, and the saturated shape of the rarefaction curves indicated that bacterial richness of each bronchoalveolar lavage fluid sample was completely sampled (Figure 4).

Comparative analysis of multiple samples

Further, the comparison of the pathogens among the 27 bronchoalveolar lavage fluid samples from patients with VAP was performed. The pathogen profile at phylum and genus level was analyzed (Figures 5 and 6). The relative abundance of each pathogen was determined in the 27 bronchoalveolar lavage fluid samples (Figures 5A and 6A). At phylum level, there were four dominant phyla, including Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes (Figure 5B), to a lesser extent, Fusobacteria, Deinococcus-Thermus, Spirochaetes, Nitrospira, Planctomycetes, Synergistetes, and Tenericutes. Of note, Proteobacteria and Firmicutes were widely distributed in all of the samples (Figure 5A); Bacteroidetes and Actinobacteria occurred in half of the samples (Figure 5A). At genus level, these...
sequences represented 90 different genera (Figure 6A), and there were 12 genera occurring in over ten different samples (Figure 6B). The top five dominant genera, which contained the largest number of sequences, were *Streptococcus*, *Acinetobacter*, *Limnohabitans*, *Neisseria*, and *Corynecbacterium* (Figure 6A). The most widely distributed genera were *Streptococcus*, *Limnohabitans*, and *Acinetobacter* in the 27 bronchoalveolar lavage fluid samples, and *Prevotella*, *Sphingomonas*, *Aquabacterium*, *Corynebacterium*, *Klebsiella*, *Pseudomonas*, *Peptostreptococcus*, *Porphyromonas*, and *Tepidimonas* were distributed to a lesser extent (Figure 6B; ≥10 samples).

Additionally, in order to evaluate the total diversity and assess the distribution and content of bacteria in these 27 bronchoalveolar lavage fluid samples from patients with VAP, the beta analysis was performed. As shown in Figure 7A and B and Figure 8A–F, there was a substantial difference in the species distribution in the 27 bronchoalveolar lavage fluid samples. The beta diversity of 27 bronchoalveolar lavage fluid samples were evaluated by weighted- and unweighted-UniFrac index (Figure 8A–F), which showed a similar results to that of Whittaker index. Collectively, the results show a comparable difference in the diversity of species distribution and evolution in the 27 bronchoalveolar lavage fluid samples from patients with VAP.

### Discussion

Due to the substantial VAP-associated morbidity and mortality and the resultant considerable heavy burden on an individual and society, it is of great importance to identify the cause of VAP for the optimization of the treatment of VAP in clinical practice. The complex of pathogen-induced VAP is still a major challenge for the use of current therapeutics. Advances in the identification and profiling of the bacteria have facilitated the revealing of the global view of

---

**Table 6** Comparison of phylotype richness and diversity estimation at 3% dissimilarity from the pyrosequencing analysis

| Sample ID | OTU number | Chao   | ACE    | Jackknife | Shannon | Nps Shannon | Simpson |
|-----------|------------|--------|--------|-----------|---------|-------------|---------|
| T1        | 33         | 96.250 | 175.163| 126.492   | 1.247   | 1.332       | 0.365   |
| T10       | 41         | 83.167 | 82.105 | 87.537    | 0.776   | 0.860       | 0.727   |
| T12       | 9          | 9.333  | 11.074 | 0.000     | 1.115   | 1.127       | 0.374   |
| T13       | 26         | 44.200 | 83.238 | 42.104    | 1.222   | 1.288       | 0.466   |
| T14       | 15         | 15.429 | 16.650 | 0.000     | 0.593   | 0.610       | 0.742   |
| T15       | 16         | 26.500 | 51.897 | 25.083    | 1.450   | 1.483       | 0.290   |
| T16       | 40         | 107.667| 375.474| 114.976   | 0.489   | 0.636       | 0.863   |
| T17       | 23         | 28.600 | 41.421 | 31.000    | 1.037   | 1.067       | 0.548   |
| T18       | 14         | 21.500 | 37.082 | 20.472    | 0.366   | 0.391       | 0.879   |
| T2        | 15         | 24.333 | 39.182 | 23.815    | 0.128   | 0.155       | 0.966   |
| T20       | 58         | 112.375| 167.734| 117.955   | 2.292   | 2.386       | 0.166   |
| T22       | 40         | 66.250 | 112.928| 63.055    | 1.192   | 1.291       | 0.563   |
| T23       | 53         | 65.214 | 77.867 | 72.000    | 2.081   | 2.157       | 0.230   |
| T24       | 54         | 124.125| 208.976| 121.862   | 0.823   | 0.880       | 0.690   |
| T25       | 75         | 296.000| 671.875| 1186.636  | 1.645   | 1.750       | 0.300   |
| T26       | 87         | 392.000| 707.645| 367.753   | 1.755   | 1.870       | 0.301   |
| T27       | 93         | 221.333| 376.515| 251.031   | 2.215   | 2.343       | 0.200   |
| T28       | 69         | 111.500| 124.030| 111.733   | 1.517   | 1.588       | 0.359   |
| T29       | 113        | 386.000| 896.042| 387.827   | 2.515   | 2.677       | 0.159   |
| T3        | 11         | 14.333 | 30.143 | 16.000    | 0.263   | 0.286       | 0.905   |
| T30       | 47         | 83.111 | 126.877| 86.082    | 0.514   | 0.571       | 0.840   |
| T31       | 48         | 106.000| 151.567| 109.196   | 0.981   | 1.037       | 0.518   |
| T4        | 47         | 74.273 | 156.638| 73.418    | 1.288   | 1.396       | 0.538   |
| T5        | 60         | 153.500| 259.506| 152.614   | 1.680   | 1.799       | 0.415   |
| T6        | 4          | 4.000  | 5.878  | 0.000     | 0.049   | 0.054       | 0.985   |
| T7        | 25         | 51.250 | 129.015| 52.288    | 1.028   | 1.081       | 0.471   |
| T8        | 20         | 42.000 | 109.198| 40.727    | 1.282   | 1.334       | 0.330   |

**Abbreviations:** ACE, abundance-based coverage estimators; OTU, operational taxonomic units.
Figure 3 Alpha diversity analysis of the species distribution in 27 bronchoalveolar lavage fluid samples from patients with VAP.

Notes: Alpha diversity analysis is a species diversity of single sample, including the Chao and ACE value, Shannon and Simpson index. Chao and ACE are used to estimate species richness; Shannon index and Simpson index are a diversity index. Shannon index is larger, and Simpson index is closer to 0. It means there are abundant species in the sample. The results showed that Chao1 value was ~80 (A), ACE value was ~130 (B), Shannon index was 1.25 (C), and Simpson index was 0.46 (D).

Abbreviations: ACE, abundance-based coverage estimators; OTU, operational taxonomic units; VAP, ventilator-associated pneumonia.

Figure 4 Rarefaction curves of the microbial communities of 27 bronchoalveolar lavage fluid samples from patients with VAP.

Notes: Rarefaction curves were used to estimate whether the number of sequence is sufficient to cover all species and estimate species richness for the 27 bronchoalveolar lavage fluid samples from patients with VAP (in this case, the number of taxonomy at a 3% dissimilarity level). The vertical axis shows the number of operational taxonomic units that would be expected to be found after sampling the numbers of tags shown on the horizontal axis. Lines of different colors represent 27 bronchoalveolar lavage fluid samples. The saturated shapes of the rarefaction curves indicate that sequencing depth has covered all species in the sample. The unsaturated shapes of the rarefaction curves indicate that bacterial richness of the sample is high and there are many undetected species.

Abbreviations: OTU, operational taxonomic units; VAP, ventilator-associated pneumonia.
Figure 5 Bacterial community composition of 27 bronchoalveolar lavage fluid samples at phylum level.

Notes: It shows the bacterial community composition of each sample at phylum level. (A) The vertical axis represents relative bacterial abundance of corresponding phylum. The horizontal axis represents 27 bronchoalveolar lavage fluid samples. Different color components represent different phylum. Taxonomic name of higher abundance is shown in these charts. Those failed to meet the lower abundance is assigned the label “Others”. “Unknown” indicates that the taxonomy level is unable to be defined according to the corresponding abundance. (B) The vertical axis represents the number of samples and the horizontal axis represents phylum of bacterial. The most abundant sequences belonged to Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria.

causative pathogens of VAP. Metagenomic studies have dramatically expanded our understanding of the microbial world without the cultivation of microorganisms and can overcome the shortages of the conventional culture-based approach.17,18,21,23,24 In the present study, we have performed metagenomics high-throughput sequencing to analyze the pathogens that were present in 27 bronchoalveolar lavage fluid samples from patients with VAP. We have identified most of the bacterial pathogens in these 27 samples from patients with VAP. The percentage of sequences belonging to each bacterial genus has been calculated for each patient that may facilitate the optimization of the therapy for VAP treatment and the achievement of maximum therapeutic outcome in clinical practice.

Emerging evidence shows the advantage of high-throughput sequencing in the identification of the pathogens to potentially improve the specificity of the diagnosis of VAP and the consequent unnecessary antibiotic use and its
Figure 6 Bacterial community composition of 27 bronchoalveolar lavage fluid samples at genus level.

Notes: It shows bacterial community composition of each sample at genus level. (A) The vertical axis represents relative bacterial abundance of corresponding phylum. The horizontal axis represents 27 bronchoalveolar lavage fluid samples. Different color components represent different genus. Taxonomic name of higher abundance is shown in these charts. Those failed to meet the lower abundance was assigned the label “Others”. “Unknown” indicates that the taxonomy level is unable to be defined according to the abundance. (B) The vertical axis represents the number of samples, and the horizontal axis represents genus of bacterial. There were 15 major genera and much mixed infection in the samples.

In our study, the findings showed that the high-throughput sequencing approach provided a fast, accurate, and global pyrosequencing reads, which can expand our understanding of the diversity of bacteria present in the respiratory tract of patients with VAP. On the other hand, we detected part of the sequences that belong to unknown species, and there are completely unknown species in 22 of 27 samples in the present study. The observation of these results may be due to the associated problems, such as drug resistance.9,17,25 The advent of 454 pyrosequencing technology has greatly accelerated studies on the profiling and deciphering the complexity of the human microbiota composition,26,27 leading to an improved understanding in the causative pathogens in the pathogenesis of infectious diseases. The utilization of barcoded primers and 454 pyrosequencing in metagenomics generates a global and comprehensive profiling of the microbiota.28,29
non-homogenous lysis of bacterial cells, primer mismatches, and the presence of mixed bases in PCR process. These reasons can lead to the failure of the species functional annotation. Therefore, it is critical to eliminate these factors.

Due to the complicated composition of pathogens in the development of VAP, it is of great importance to discriminate the composition of the pathogens at different levels. This will be clinically helpful for the proper therapeutics selection and therapy optimization in the treatment of VAP. The causative bacteria for an early-onset VAP include *Streptococcus*, *Hemophilus influenzae*, methicillin-sensitive *S. aureus*, antibiotic-sensitive enteric Gram-negative bacilli, *Escherichia coli*, *K. pneumoniae*, *Enterobacter* species, *Proteus* species, and *S. marcescens*, and the bacteria causing late VAP are typically multiple drug-resistant bacteria, such as methicillin-resistant *S. aureus*, *Acinetobacter*, *P. aeruginosa*, and extended-spectrum beta-lactamase producing bacteria. Our findings showed that the most common isolates were *Streptococcus*, *Acinetobacter*, and *Limnohabitans* at genus level. In addition, the most prevalent representative sequence belonged to *Streptococcus*. They belong to opportunistic pathogens, which are widely distributed in nature. They exist in the form of carrier state, generally not pathogenic, but can cause disease when immunity decreased. The *Streptococcus* encompasses both commensal and pathogenic Gram-positive bacteria that inhabit various human body sites, such as *Streptococcus pneumoniae*, a well-known pathogen associated with pneumonia. In recent years, *Acinetobacter baumannii* has become increasingly prevalent. It has been the main pathogenic bacteria in hospital-acquired infection, particularly in VAP. Thus, the global identification of the causative pathogens in VAP is important for therapeutics selection.

In the treatment of VAP in the absence of risk factors for multidrug-resistant bacteria, the empirical therapy can be selected for the causative pathogens of *Streptococcus pneumoniae, Haemophilus influenzae*, methicillin-sensitive *S. aureus*, and antibiotic-sensitive gram-negative enteric organisms. The antibiotics used to treat drug-sensitive VAP include ceftriaxone, quinolones (levofloxacin, moxifloxacin, or ciprofloxacin), ampicillin/sulbactam, or ertapenem. However, the mixed infections with drug-resistant bacteria represent a major challenge in the treatment of VAP, which often occurs. Therefore, in the treatment of VAP in the presence of risk factors for multidrug-resistant organisms, the clinician must consider not only empirical therapies but also the therapies for the drug-resistant causative pathogens that include *P. aeruginosa, Klebsiella, Enterobacter, Serratia, Acinetobacter, Stenotrophomonas maltophilia, B. cepacia*, and methicillin-resistant *S. aureus*. The therapeutics should be broadened to achieve the therapeutic effect. This includes 1) either an antipseudomonal cephalosporin (cefepime or ceftazidime), an antipseudomonal
Figure 8 Beta diversity analysis of 27 bronchoalveolar lavage fluid samples from patients with VAP indicated by UniFrac index.

Notes: UniFrac index was used to evaluate the difference in species evolution among different samples. The higher index indicates more difference. (A) The weighted UniFrac index of species distribution of 27 bronchoalveolar lavage fluid samples from patients with VAP. (B) The PCoA analysis of weighted UniFrac for 27 bronchoalveolar lavage fluid samples from patients with VAP. (C) The cluster dendrogram based on weighted-UniFrac distance analyzed by Ward of hclust in R program. (D) The unweighted UniFrac index of species distribution of 27 bronchoalveolar lavage fluid samples from patients with VAP. (E) The PCoA analysis of unweighted UniFrac for 27 bronchoalveolar lavage fluid samples from patients with VAP. (F) The cluster dendrogram based on unweighted-UniFrac distance analyzed by Ward of hclust in R program.

Abbreviation: VAP, ventilator-associated pneumonia.
carbenem (imipenem or meropenem), or a β-lactam/β-lactamase inhibitor (piperacillin-tazobactam), and 2) an antipseudomonal fluoroquinolone (ciprofloxacin or levofloxacin) or an aminoglycoside (amikacin, gentamicin, or tobramycin) plus linezolid or vancomycin.

**Conclusion**
In summary, the present study shows that the high-throughput sequencing is a clinically valuable approach to analyze the pathogens of VAP with the advantage of overcoming the limitations of convention approaches. The global view of the bacterial composition can provide a better understanding of pathogens in VAP, which can assist clinicians to make rational and effective therapeutic decisions to treat drug-sensitive or resistant bacteria, and achieve maximum therapeutic effect in the treatment of VAP in clinical practice.

**Acknowledgments**
The authors appreciate the technical support from BGI Inc. (Shenzhen, People’s Republic of China). This work was supported by grants from the Natural Science Foundation of China (Grant numbers: 81450031, 81160151) and the Natural Science Foundation of NingXia (Grant number: NZ14142).

**Disclosure**
The authors declare that there are no conflicts of interest in this work.

**References**
1. Koenig SM, Truwit JD. Ventilator-associated pneumonia: diagnosis, treatment, and prevention. *Clin Microbiol Rev*. 2006;19(4):637–657.
2. Vallés J, Martin-Löeches I, Torres A, et al. Epidemiology, antibiotic therapy and clinical outcomes of healthcare-associated pneumonia in critically ill patients: a Spanish cohort study. *Intensive Care Med*. 2014; 40(4):572–581.
3. Zilberberg MD, Shorr AF. Ventilator-associated pneumonia: the clinical pulmonary infection score as a surrogate for diagnostics and outcome. *Clin Infect Dis*. 2010;51(suppl 1):S131–S135.
4. Morrow BM, Argent AC, Jeena PM, Green RJ. Guideline for the diagnosis, prevention and treatment of paediatric ventilator-associated pneumonia. *S Afr Med J*. 2009;99(4 pt 2):255–267.
5. Klevens RM, Edwards JR, Richards CL Jr, et al. Estimating healthcare-associated infections and deaths in U.S. hospitals, 2002. *Clin Microbiol Rev*. 2007;103(3):352–365.
6. Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. *BMC Microbiol*. 2011;11:244.
7. Chen CY, Wen TY, Wang GS, Cheng HW, Lin YH, Lien GW. Determining estrogenic steroids in Taipei waters and removal in drinking water treatment using high-flow solid-phase extraction and liquid chromatography/tandem mass spectrometry. *Sci Total Environ*. 2007;378(3):352–365.
8. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol*. 2008;6(11):e280.
9. Ahmadian A, Ghazizadeh B, Gustafsson AC, et al. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem*. 2000;280(1):103–110.
30. Fredricks DN, Smith C, Meier A. Comparison of six DNA extraction methods for recovery of fungal DNA as assessed by quantitative PCR. *J Clin Microbiol*. 2005;43(10):5122–5128.
31. Kuczynski J, Lauber CL, Walters WA, et al. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet*. 2012;13(1):47–58.
32. Shin S, Lee TK, Han J, Park J. Regional effects on chimera formation in 454 pyrosequenced amplicons from a mock community. *J Microbiol*. 2014;52(7):566–573.
33. Guillamet CV, Kollef MH. Ventilator associated pneumonia in the ICU: where has it gone? *Curr Opin Pulm Med*. 2015;21(3):226–231.
34. Kalanuria AA, Zai W, Mirski M. Ventilator-associated pneumonia in the ICU. *Crit Care*. 2014;18(2):208.
35. Johanson WG Jr, Pierce AK, Sanford JP. Changing pharyngeal bacterial flora of hospitalized patients. Emergence of Gram-negative bacilli. *N Engl J Med*. 1969;281(21):1137–1140.
36. Johanson WG Jr, Pierce AK, Sanford JP, Thomas GD. Nosocomial respiratory infections with Gram-negative bacilli. The significance of colonization of the respiratory tract. *Ann Intern Med*. 1972;77(5):701–706.
37. Bonten MJ, Bergmans DC, Ambergen AW, et al. Risk factors for pneumonia, and colonization of respiratory tract and stomach in mechanically ventilated ICU patients. *Am J Respir Crit Care Med*. 1996;154(5):1339–1346.
38. Nobbs AH, Lamont RJ, Jenkinson HF. Streptococcus adherence and colonization. *Microbiol Mol Biol Rev*. 2009;73(3):407–450, table of contents.
39. Magnotti LJ, Croce MA, Zarzaur BL, et al. Causative pathogen dictates optimal duration of antimicrobial therapy for ventilator-associated pneumonia in trauma patients. *J Am Coll Surg*. 2011;212(4):476–484; discussion 484–486.
40. Memish ZA, Shibr AM, Kambal AM, Ohaly YA, Ishaq A, Livermore DM. Antimicrobial resistance among non-fermenting Gram-negative bacteria in Saudi Arabia. *J Antimicrob Chemother*. 2012;67(7):1701–1705.