Int J Clin Exp Med 2016;9(12):23491-23498
www.ijcem.com /ISSN:1940-5901/IJCEM0030860

Original Article

Comparison of fresh bronchoalveolar lavage fluid samples with frozen bronchoalveolar lavage fluid samples of microbial community measures based on metagenomic data

Yan Wen[1], Fei Xiao[2], Chen Wang[3]

[1]*Department of Intensive Care Unit, The First Affiliated Hospital of Xiamen University, Xiamen 361003, Fujian Province, China;* [2]*Beijing Institute of Geriatrics, Beijing Hospital, Beijing 100730, China;* [3]*Beijing Key Laboratory of Respiratory and Pulmonary Circulation Disorders, National Clinical Research Center of Respiratory Diseases, China-Japan Friendship Hospital, Beijing 100029, China*

Received December 7, 2015; Accepted September 29, 2016; Epub December 15, 2016; Published December 30, 2016

Abstract: Background: DNA from the host, and extraction biases can significantly interfere with microbiota assessment and increase the cost of high throughput sequencing. This study aimed to find a better method to enable a higher concentration of microbial DNA to be extracted from the bronchoalveolar lavage fluid (BALF) samples with a higher proportion of host cells. Methods: Two subjects were enrolled in this study. DNA was extracted from 3.0 ml of the fresh or frozen BALF samples of each subject using QIAamp DNA Microbiome Kit. DNA libraries were constructed according to the manufacturer's instructions (illumina). Cluster generation, template hybridization, isothermal amplification, linearization, blocking, denaturing, and hybridization of the sequencing primers were performed. Raw data was uploaded to MG-RAST v3 and analyzed. Results: Through preliminary analysis of sequence characteristics, it was found that the ratio of human genome in clean reads of each DNA sample from frozen BALF samples was significantly lower than those from fresh BALF samples (Patient 1: 59.80% vs. 68.60%, Patient 2: 47.91% vs. 66.89%, respectively). For each individual, numbers of hits from frozen samples by same database were obviously higher than those from fresh samples. Tiny difference on the top 50 most abundant classified phyla was observed between fresh BALF sample and frozen BALF sample from the same individual. The different classified phyla belonged to low abundance phyla of Eukaryota. No significant difference on the principal bacterial community structure between fresh BALF samples and frozen BALF samples by comparing the bacterial community structure of fresh BALF samples with frozen BALF samples. Conclusion: Cryopreservation of BALF samples enables conservation of a higher yield of microbial DNA from samples and a higher fraction of host cells. It will not affect the principal structure of bacterial community.

Keywords: BALF, microbial DNA, metagenomic data

Introduction

Metagenomics is adopted to consider the microbial population as a whole 'metagenome' by applying high-throughput shot-gun sequencing to the entire population to identify the community members' presentation and their genetically encoding functional capacity. As we know, the reports of metagenomic high throughput sequencing include host or background DNA sequence reads while supplying DNA sequence reads of microorganisms. Host DNA sequence reads are not included in the following analysis. Fewer data discard refers to the lower cost of sequencing and more accurate result.

To date there is a dearth of research on metagenomic explorations of lower respiratory tract microorganisms from bronchoalveolar lavage fluid (BALF) or lung tissue. As observed by Hilty et al. [1], the concentrations of bacteria in the alveoli and small airways are obviously lower than that of the mouth or lower digestive tract. Accordingly, it is a challenge to find a better method to enable a higher concentration of microbial DNA extracted from the BALF sam-

ples. The development of effective and efficient decontamination methods suitable for high-throughput use or development of ultrapure reagents could potentially further reduce the background DNA [2, 3]. Mao et al. [4] confirmed that, particularly in comparison restricted to a specific type of sample, technical differences in experimental protocols between laboratories, including the manner in which samples are obtained and stored, DNA extraction methods, and the instruments used to determine the nucleotide sequences, might all generate variability that could outweigh biological differences.

We therefore compared the taxon abundance and bacterial community structure of fresh BALF samples with frozen BALF samples. This study aimed to discover whether the manner in which samples were stored may affect the concentration of extracted microbial DNA and the bacterial community structure, thereby finding an optimized sample pretreatment strategy which could reduce the cost of sequencing and increase the accuracy of the analysis of the sequencing data.

Materials and methods

Patient information

All the enrolled individuals with an indication for bronchoscopy at Beijing Chao-Yang Hospital were ≥ 18 years of age and provided informed consent. A complete clinical, functional, and radiological evaluation of all patients was made.

The study conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Institutional Review Board of Beijing Chao-Yang Hospital.

BALF samples collection

BALF samples were taken from each individual. Three 50 ml-aliquots of sterile saline (0.9% w/v) were instilled to obtain the BALF samples, from the third generation bronchus of the middle lobe (lingual) or the area containing most lung infiltrates using a fibreoptic bronchoscope (Type 40; Olympus, Tokyo, Japan) under local anesthesia. Each aliquot was retrieved immediately by suction.

All specimens collected were delivered immediately from our hospital to the laboratory in an ice bag using insulating polystyrene foam container. In the laboratory, each specimen was divided into 1.5 ml aliquot. Two aliquots (3.0 ml) of each specimen were processed for DNA extraction immediately, while the left aliquots were stored at -80°C using Glycerol (10%) as cryoprotective agent until processing for DNA extraction.

DNA extraction

All DNA extraction was performed using 3.0 ml of the fresh samples or frozen samples. We adopted QIAamp DNA Microbiome Kit (Catalogue 51704, Qiagen, Hilden, Germany) to extract DNA. To increase DNA production, DNA was eluted with relatively small volume (30 μl) of recommended elution buffer. DNA was isolated according to the manufacturer's instruction. DNA concentration was measured by Qubit® 2.0 Fluorometer (Life Technologies, Invitrogen, USA).

For the negative control, the same procedure was applied with sterile water. No PCR products were detected in any experiment, indicating avoid of contamination from any reagents used.

DNA library construction and sequencing

DNA libraries were constructed according to the manufacturer's instruction (illumina). The same workflows from illumina were used to perform cluster generation, template hybridization, isothermal amplification, linearization, blocking, denaturing, and hybridization of the sequencing primers. We performed paired-end sequencing on 2 × 100 base pairs (base pair, bp) for all libraries. The base-calling pipeline (Casava 1.8.2 with parameters '-use-bases-mask y100n, I6n, Y100n, -mismatches 1, -adaptor-sequence') was used to process the raw fluorescent images and call sequences. The same insert size inferred by Agilent 2100 was used for all libraries.

Sequence processing and statistical analysis

After upload to MG-RAST v3, the data was preprocessed by using SolexaQA to trim low-quality regions from FASTQ data [5, 6]. MG-RAST v3 used DRISEE (Duplicate Read Inferred Sequencing Error Estimation) to analyze the sets of

Table 1. Comparison of sequence characteristics of fresh BALF samples with frozen BALF samples

Samples (BALF)		Raw reads	Raw bases	Clean reads	Clean bases	Ratio[1] (%)	Align_hg19[2]	Ratio[3] (%)
Patient 1	Fresh1	6,167,942	931,359,242	5,554,766	838,769,666	90.06	3,810,726	68.60
	Frozen1	6,794,296	856,081,296	6,288,128	792,304,128	92.55	3,760,308	59.80
Patient 2	Fresh2	5,950,386	749,748,636	5,563,678	701,023,428	93.50	3,721,586	66.89
	Frozen2	6,209,540	776,192,500	4,993,396	624,174,500	80.41	2,392,573	47.91

Notes: [1]values of column "Clean reads"/values of column "Raw reads"; [2]Reads of alignments to hg19 (human genome); [3]values of column "Align_hg19"/values of column "Clean reads", i.e. ratio of human genome in clean reads of each sample. Fresh1, specimen collected from patient 1 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately. Frozen1, specimen collected from patient 1 were stored at -80°C until processing for DNA extraction. Fresh2, specimen collected from patient 2 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately. Frozen2, specimen collected from patient 2 were stored at -80°C until processing for DNA extraction.

Artificial Duplicate Reads (ADRs) and determine the degree of variation among prefix-identical sequences derived from the same template [7, 8]. The data was compared to M5NR using the following parameters: a maximum e-value of 1e-5, a minimum identity of 60%, and a minimum alignment length of 15 measured in amino acid for protein and base pair for RNA databases. The displayed data was normalized to values between 0 and 1 to allow for comparison of differently sized samples based on abundance. The taxonomic profiles use the NCBI taxonomy. We used the best hit classification to report the functional and taxonomic annotation of the best hit in the M5NR for each feature. Raw sequences were analyzed using mothur v1.21 to remove sequences containing homopolymers greater than 8 bp, mismatches in the barcode or primer, one or more ambiguous bases, or an average quality score below 35 over a moving window of 50 bp [6]. Remaining sequences that were at least 200 bp but less than 590 bp in length were further curated to remove chimeric sequences using UCHIME and to reduce sequencing noise by a preclustering methodology before being assigned to operational taxonomic units (OTUs). An average neighbor algorithm at 0.03 dissimilarity cutoff was adopted [8, 9]. The consensus taxonomy of each OTU was identified at the genus level using the Bayesian method [10]. The total number of reads for each community was normalized to 498. The smallest number of reads among the samples included in the study, to control for differences in sequencing depth before alpha and beta diversity measures was calculated.

The graphs for comparison of numbers of hits from fresh sample with frozen sample were drawn by GraphPad Prism 5. Comparison of the principal bacterial community structure of fresh BALF samples with frozen BALF samples was calculated by Fisher's exact test using IBM SPSS 22.0 software package. $P < 0.05$ was considered statistically significant.

Results

Study subjects, DNA sequence characteristics

A total of 2 subjects were enrolled in this study. All DNA extractions were performed using 3.0 ml of the fresh or frozen BALF samples from each subject.

Through preliminary analysis of sequence characteristics, it was found that the ratio of human genome in clean reads of each DNA sample from frozen BALF samples was remarkably smaller than those from fresh BALF samples (Patient 1: 59.80% vs. 68.60%, Patient 2: 47.91% vs. 66.89%, respectively, **Table 1**). It indicated that a microbial DNA extraction with frozen BALF samples can generate a higher production of microbial DNA, compared with the fresh BALF samples.

Comparison of fresh BALF samples with frozen BALF samples of taxon abundance and bacterial community structure

For fresh sample from patient 1, 418,819 (66.8%) of the predicted protein features could be annotated with similarity to a protein of known function. 250,227 (59.7%) of these annotated features could be placed in a functional hierarchy. 13,402 (0.2%) of reads had similarity to ribosomal RNA genes. For frozen sample from patient 1, 543,383 (55.0%) of the predicted protein features could be annotated

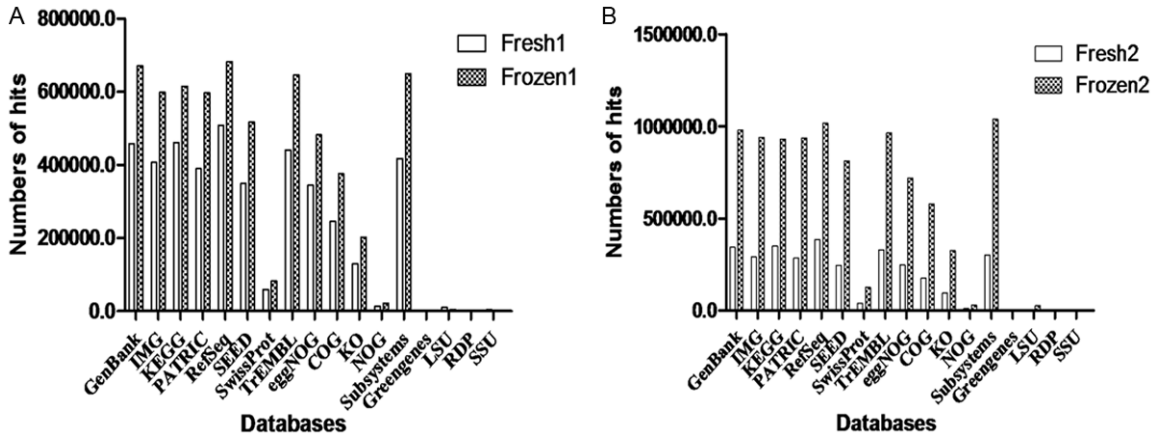


Figure 1. The numbers of hits annotated by the different databases for each sample. A. Comparison of numbers of hits from fresh sample with frozen sample of patient 1 annotated by the different databases. B. Comparison of numbers of hits from fresh sample with frozen sample of patient 2 annotated by the different databases. Fresh1: specimen collected from patient 1 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately; Frozen1: specimen collected from patient 1 were stored at -80°C until processing for DNA extraction; Fresh2: specimen collected from patient 2 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately; Frozen2: specimen collected from patient 2 were stored at -80°C until processing for DNA extraction.

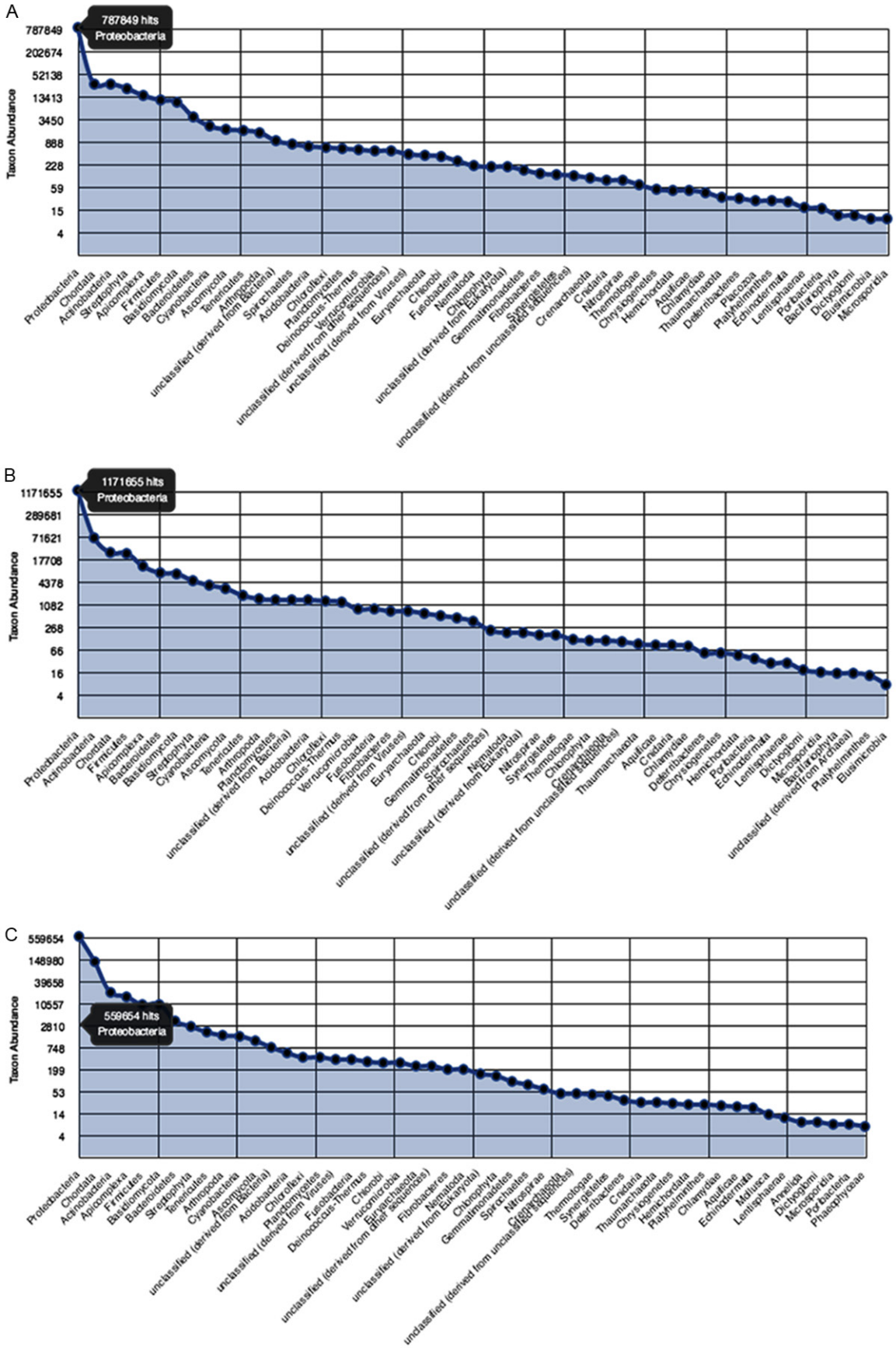
with similarity to a protein of known function. 370,488 (68.2%) of these annotated features could be placed in a functional hierarchy. 11,180 (0.2%) of reads had similarity to ribosomal RNA genes.

For the fresh sample from patient 2, 323,226 (57.7%) of the predicted protein features could be annotated with similarity to a protein of known function. 185,388 (57.4%) of these annotated features could be placed in a functional hierarchy. 10,092 (0.2%) of reads had similarity to ribosomal RNA genes. For frozen sample from patient 2, 657,903 (66.3%) of the predicted protein features could be annotated with similarity to a protein of known function. 459,704 (69.9%) of these annotated features could be placed in a functional hierarchy. 209,522 (4.2%) of reads had similarity to ribosomal RNA genes.

Figure 1 displayed the number of features in datasets that were annotated by the different databases (including protein databases, protein databases with functional hierarchy information, and ribosomal RNA databases). Different databases for each sample have different numbers of hits. For each individual, numbers of hits from frozen samples by same database were higher than those from fresh samples (Tables S1, S2, S3 and S4).

The top 50 most abundant phyla of BALF samples were shown in **Figure 2**. The abundances ordered from the most abundant to least abundant. The plot outline of frozen sample is similar with that of fresh sample from the same individual (Tables S5, S6, S7 and S8). We compared the top 50 most abundant phyla of fresh and frozen BALF samples, and found tiny difference on the top 50 most abundant classified phyla between fresh BALF sample and frozen BALF sample from the same individual (**Table 2**, Tables S5, S6, S7 and S8). For Patient 1, *Placozoa* with an abundance 25 only existed in fresh sample; for Patient 2, *Mollusca* with an abundance 12 only existed in fresh sample, *Bacillariophyta* with an abundance 28 only existed in frozen sample. This result suggested that these different classified phyla were all low abundance phyla of Eukaryota. No difference on principal bacterial community structure between fresh BALF sample and frozen BALF sample from the same individual was observed. The difference on numbers of hits between fresh BALF sample and frozen BALF sample can only impact on the order of different phyla. However, they shared the most principal bacterial phylum (*Proteobacteria*).

By comparing the bacterial community structure of fresh BALF samples with frozen BALF samples, there was no significant difference on


A
787849 hits
Proteobacteria
Taxon Abundance
787849
202674
52138
13413
3450
888
228
59
15
4
B
1171655 hits
Proteobacteria
Taxon Abundance
1171655
289681
71621
17708
4378
1082
268
66
16
4
C
559654 hits
Proteobacteria
Taxon Abundance
559654
148980
39658
10557
2810
748
199
53
14
4

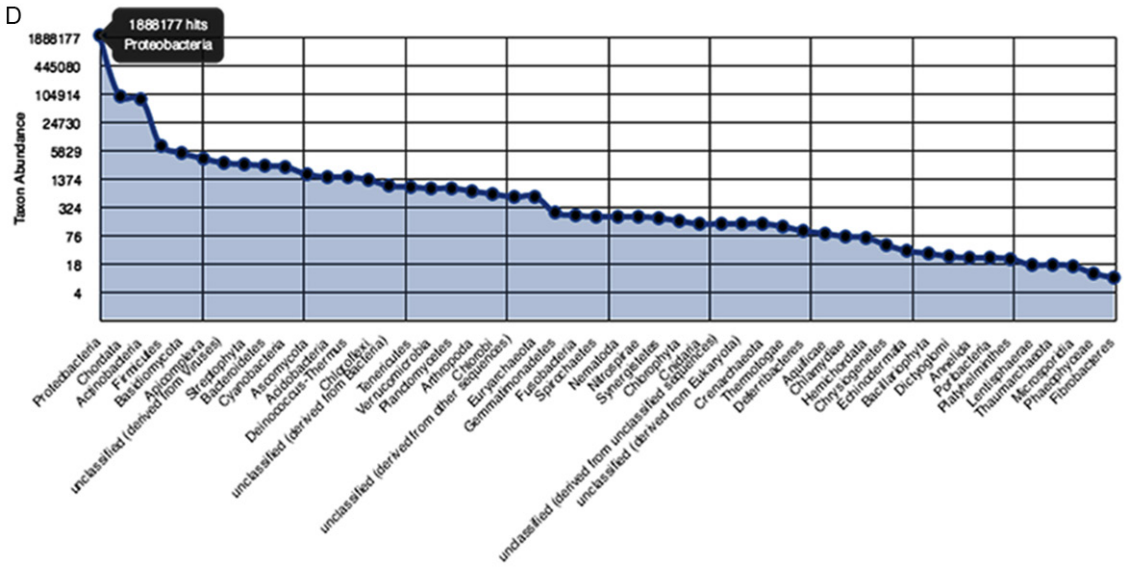


Figure 2. The top 50 most abundant phyla of each sample. A. The top 50 most abundant phyla of fresh BALF sample from patient 1. B. The top 50 most abundant phyla of frozen BALF sample from patient 1. C. The top 50 most abundant phyla of fresh BALF sample from patient 2. D. The top 50 most abundant phyla of frozen BALF sample from patient 2. The plots show the phylum abundances ordered from the most abundant to least abundant. Only the top 50 most abundant are shown. The y-axis plots the abundances of annotations in each phylum on a log scale. The rank abundance curve is a tool for visually representing taxonomic richness and evenness.

Table 2. Comparison of the top 50 most abundant classified phyla of fresh BALF samples with frozen BALF samples

Samples		Phylum	Taxon abundance*
Patient 1	Fresh1	Placozoa	25
	Frozen1	-	-
Patient 2	Fresh2	Mollusca	12
	Frozen2	Bacillariophyta	28

Notes: This table shows the classified phyla only existed in the top 50 most abundant phyla of fresh BALF sample or frozen BALF sample from the same individual. Fresh1, specimen collected from patient 1 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately. Frozen1, specimen collected from patient 1 were stored at -80°C until processing for DNA extraction. Fresh2, specimen collected from patient 2 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately. Frozen2, specimen collected from patient 2 were stored at -80°C until processing for DNA extraction. *abundance of the specific phylum.

the principal bacterial community structure between fresh BALF samples and frozen BALF samples (Patient 1: P=0.760, Patient 2: P=0.310, for the constituent ratio of Phyla of bacteria; Patient 1: P=0.777, Patient 2: P=0.435, for the constituent ratio of Classes of Proteobacteria; **Table 3**; Figures S1, S2, S3, S4, S5, S6, S7 and S8).

Discussion

In this study, we compared the metagenomic sequence characteristics, the taxon abundance, and bacterial community structure of fresh BALF samples with frozen BALF samples.

Through preliminary analysis of sequence characteristics, we found that the ratio of human genome in clean reads of each DNA sample from frozen BALF samples was markedly lower than those from fresh BALF samples. It indicated that a microbial DNA extraction from frozen BALF samples generated a higher production of microbial DNA.

With regard to cryoprotective conditions, animal cells, bacteria, bacteriophage, and spores (fungi) minimum share the similar cryoprotective agent Glycerol (10%). In this study, we chose Glycerol (10%) as cryoprotective agent, which is suitable to animal cells and microbial cells for cryopreservation. The minimum storage temperature is -60°C for bacteria, bacteriophage, and spores (fungi). Most of bacteria and spore-forming fungi may tolerate storage tem-

Table 3. Comparison of the principal bacterial community structure of fresh BALF samples with frozen BALF samples

Projects \ Samples		Patient 1		Patient 2	
		Fresh1	Frozen1	Fresh2	Frozen2
Ratio of Phylumin bacteria (%)	Proteobacteria	97%	95%	93%	97%
	Actinobacteria	2%	3%	2%	2%
	Others	1%	2%	5%	1%
	P value	0.760		0.310	
Ratio of Class in Proteobacteria (%)	Betaproteobacteria	66%	69%	73%	69%
	Alphaproteobacteria	32%	28%	24%	30%
	Gammaproteobacteria and others	2%	3%	3%	1%
	P value	0.777		0.435	

Notes: Fresh1, specimen collected from patient 1 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately. Frozen1, specimen collected from patient 1 were stored at -80°C until processing for DNA extraction. Fresh2, specimen collected from patient 2 were delivered immediately to the laboratory in an ice bag and were processed and performed DNA extractions immediately. Frozen2, specimen collected from patient 2 were stored at -80°C until processing for DNA extraction.

perature of -60°C to -80°C and can survive for a short period (shorter than 1 year). Most viruses can be frozed as cell-free preparations without difficulty and do not require controlled cooling [11]. However, more fastidious cells, such as mammalian tissue, must be maintained below -130°C. In present study, the storage temperature for BALF samples was -80°C as described in most of studies [12-17]. As we know, mammalian cells cannot survive under this storage temperature longer than 24 hours. Compared to microbial cells, animal cells are more fastidious, such as to pH, medium, cooling rate, and procedure of reconstitution (thawing). We performed the cryopreservation according to cryoprotective protocol of microbial cells, which could cause the death of most of human cells during cryopreservation or reconstitution.

In order to observe whether cryopreservation could impact on the bacterial community structure, we compared the taxon abundance and bacterial community structure of fresh BALF samples with frozen BALF samples. We found that number of hits from frozen samples upon the same database was higher than that from fresh samples for each individual. By comparing the top 50 most abundant phyla of fresh BALF samples with frozen BALF samples, there was tiny difference on the top 50 most abundant classified phyla between fresh BALF sample and frozen BALF sample from the same individual. The result suggested that these different classified phyla were all low abundance phyla of Eukaryota. As noted in MG-RAST Manual for version 3.6, "The system supports the analysis of the prokaryotic content of samples, analysis of viruses and eukaryotic sequences is not currently supported" [5]. Therefore, this difference is invalid. The difference on number of hits between fresh BALF sample and frozen BALF sample can only impact on the order of different phyla. However, they share the most principal bacterial phylum (*Proteobacteria*). We further compared the bacterial community structure of fresh BALF samples with frozen BALF samples. We found that there was no significant difference on the principal bacterial community structure between fresh BALF samples and frozen BALF samples. These findings indicated that cryopreservation could not affect the bacterial community structure.

Cryopreservation of BALF samples enables a higher yield of microbial DNA from samples with a higher fraction of host cells obtained, without affecting the bacterial community structure, thereby reducing the cost of sequencing and increasing the accuracy of the analysis of the sequencing data.

Conclusion

Cryopreservation of BALF samples enables a higher production of microbial DNA from samples with a higher fraction of host cells to be obtained, whereas shows no impact on the principal bacterial community structure.

Acknowledgements

This work was supported by a grant from the National High Technology Research and Development Program of China (No. 2012AA02A511). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure of conflict of interest

None.

Authors' contribution

Conceived and designed the experiments: YW FX CW. Performed the experiments: YW. Analyzed the data: YW. Wrote the paper: YW.

Address correspondence to: Chen Wang, Beijing Key Laboratory of Respiratory and Pulmonary Circulation Disorders, National Clinical Research Center of Respiratory Diseases, China-Japan Friendship Hospital, Beijing 100029, China. Tel: (011) 86-10-64222969; Fax: (011) 86-10-65911810; E-mail: cyh_birmw@sina.com

References

[1] Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF and Cookson WO. Disordered microbial communities in asthmatic airways. PLoS One 2010; 5: e8578.

[2] Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarski EB and Fox AJ. Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. J Clin Microbiol 2000; 38: 1747-1752.

[3] Klaschik S, Lehmann LE, Raadts A, Hoeft A and Stuber F. Comparison of different decontamination methods for reagents to detect low concentrations of bacterial 16S DNA by real-time-PCR. Mol Biotechnol 2002; 22: 231-242.

[4] Mao DP, Zhou Q, Chen CY and Quan ZX. Coverage evaluation of universal bacterial primers using the metagenomic datasets. BMC Microbiol 2012; 12: 66.

[5] Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J and Edwards RA. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 2008; 9: 386.

[6] Cox MP, Peterson DA and Biggs PJ. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 2010; 11: 485.

[7] Keegan KP, Trimble WL, Wilkening J, Wilke A, Harrison T, D'Souza M and Meyer F. A platform-independent method for detecting errors in metagenomic sequencing data: DRISEE. PLoS Comput Biol 2012; 8: e1002541.

[8] Gomez-Alvarez V, Teal TK and Schmidt TM. Systematic artifacts in metagenomes from complex microbial communities. ISME J 2009; 3: 1314-1317.

[9] Huse SM, Huber JA, Morrison HG, Sogin ML and Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 2007; 8: R143.

[10] Huson DH, Auch AF, Qi J and Schuster SC. MEGAN analysis of metagenomic data. Genome Res 2007; 17: 377-386.

[11] Simione FP. ATCC preservation methods: freezing and freeze drying. Rockville, Maryland: American Type Culture Collection; 1991.

[12] Salonen A, Nikkila J, Jalanka-Tuovinen J, Immonen O, Rajilic-Stojanovic M, Kekkonen RA, Palva A and de Vos WM. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J Microbiol Methods 2010; 81: 127-134.

[13] Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, Hwang J, Chen J, Berkowsky R, Nessel L, Li H and Bushman FD. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol 2010; 10: 206.

[14] Zoetendal EG, Ben-Amor K, Akkermans AD, Abee T and de Vos WM. DNA isolation protocols affect the detection limit of PCR approaches of bacteria in samples from the human gastrointestinal tract. Syst Appl Microbiol 2001; 24: 405-410.

[15] Zoetendal EG, Heilig HG, Klaassens ES, Booijink CC, Kleerebezem M, Smidt H and de Vos WM. Isolation of DNA from bacterial samples of the human gastrointestinal tract. Nat Protoc 2006; 1: 870-873.

[16] Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, Vazquez-Baeza Y, Jansson JK, Gordon JI and Knight R. Meta-analyses of studies of the human microbiota. Genome Res 2013; 23: 1704-1714.

[17] Garzoni C, Brugger SD, Qi W, Wasmer S, Cusini A, Dumont P, Gorgievski-Hrisoho M, Muhlemann K, von Garnier C and Hilty M. Microbial communities in the respiratory tract of patients with interstitial lung disease. Thorax 2013; 68: 1150-1156.

Table S1. Fresh1 hit source

Database	Numbers of hits
GenBank	458247
IMG	406874
KEGG	461651
PATRIC	389448
RefSeq	508631
SEED	349287
SwissProt	59877
TrEMBL	441061
eggNOG	344430
COG	245982
KO	129926
NOG	13816
Subsystems	417859
Greengenes	1283
LSU	10598
RDP	1326
SSU	4294

Table S2. Frozen1 hit source

Database	Numbers of hits
GenBank	671104
IMG	599300
KEGG	615925
PATRIC	597492
RefSeq	683154
SEED	517319
SwissProt	82331
TrEMBL	645753
eggNOG	482367
COG	375652
KO	202275
NOG	21529
Subsystems	649899
Greengenes	845
LSU	4060
RDP	939
SSU	2186

Table S3. Fresh2 hit source

Database	Numbers of hits
GenBank	344549
IMG	292018
KEGG	351203
PATRIC	285707
RefSeq	386175
SEED	245563
SwissProt	39392
TrEMBL	329319
eggNOG	248183
COG	175531
KO	94473
NOG	10322
Subsystems	300658
Greengenes	499
LSU	6005
RDP	513
SSU	1991

Table S4. Frozen2 hit source

Database	Numbers of hits
GenBank	980594
IMG	940179
KEGG	930669
PATRIC	936618
RefSeq	1016244
SEED	811453
SwissProt	125213
TrEMBL	964839
eggNOG	719296
COG	577935
KO	324531
NOG	29304
Subsystems	1038605
Greengenes	1357
LSU	25355
RDP	1464
SSU	4962

Table S5. The top 50 most abundant phyla of fresh1

Phylum	Abundance
Proteobacteria	787849
Chordata	26347
Actinobacteria	25597
Streptophyta	20660
Apicomplexa	13706
Firmicutes	10198
Basidiomycota	8844
Bacteroidetes	3591
Cyanobacteria	2144
Ascomycota	1774
Tenericutes	1622
Arthropoda	1427
Unclassified (derived from Bacteria)	895
Spirochaetes	708
Acidobacteria	637
Chloroflexi	600
Planctomycetes	568
Deinococcus-Thermus	523
Verrucomicrobia	486
Unclassified (derived from other sequences)	475
Unclassified (derived from Viruses)	390
Euryarchaeota	363
Chlorobi	354
Fusobacteria	262
Nematoda	193
Chlorophyta	187
Unclassified (derived from Eukaryota)	187
Gemmatimonadetes	155
Fibrobacteres	124
Synergistetes	113
Unclassified (derived from unclassified sequences)	108
Crenarchaeota	94
Cnidaria	82
Nitrospirae	81
Thermotogae	64
Chrysiogenetes	49
Hemichordata	46
Aquificae	44
Chlamydiae	39
Thaumarchaeota	30
Deferribacteres	28
Placozoa	25
Platyhelminthes	25
Echinodermata	23
Lentisphaerae	16
Poribacteria	15
Bacillariophyta	10
Dictyoglomi	10
Elusimicrobia	8
Microsporidia	8

Table S6. The top 50 most abundant phyla of frozen1

Phylum	Abundance
Proteobacteria	1171655
Actinobacteria	61814
Chordata	25765
Firmicutes	23281
Apicomplexa	10882
Bacteroidetes	7012
Basidiomycota	6871
Streptophyta	4356
Cyanobacteria	3281
Ascomycota	2625
Tenericutes	1767
Arthropoda	1460
Planctomycetes	1364
Unclassified (derived from Bacteria)	1328
Acidobacteria	1303
Chloroflexi	1260
Deinococcus-Thermus	1157
Verrucomicrobia	789
Fusobacteria	787
Fibrobacteres	686
Unclassified (derived from Viruses)	651
Euryarchaeota	580
Chlorobi	507
Gemmatimonadetes	434
Spirochaetes	343
Unclassified (derived from other sequences)	203
Nematoda	182
Unclassified (derived from Eukaryota)	173
Nitrospirae	152
Synergistetes	152
Thermotogae	117
Chlorophyta	108
Crenarchaeota	106
Unclassified (derived from unclassified sequences)	100
Thaumarchaeota	88
Aquificae	80
Cnidaria	79
Chlamydiae	75
Deferribacteres	50
Chrysiogenetes	49
Hemichordata	43
Poribacteria	35
Echinodermata	26
Lentisphaerae	26
Dictyoglomi	17
Microsporidia	15
Bacillariophyta	14
Unclassified (derived from Archaea)	14
Platyhelminthes	12
Elusimicrobia	7

Table S7. The top 50 most abundant phyla of fresh2

Phylum	Abundance
Proteobacteria	559654
Chordata	125297
Actinobacteria	18924
Apicomplexa	14461
Firmicutes	9432
Basidiomycota	9253
Bacteroidetes	3325
Streptophyta	2451
Tenericutes	1724
Arthropoda	1441
Cyanobacteria	1364
Ascomycota	1055
Unclassified (derived from Bacteria)	686
Acidobacteria	503
Chloroflexi	389
Planctomycetes	389
Unclassified (derived from Viruses)	349
Fusobacteria	341
Deinococcus-Thermus	288
Chlorobi	284
Verrucomicrobia	278
Euryarchaeota	229
Unclassified (derived from other sequences)	225
Fibrobacteres	182
Nematoda	181
Unclassified (derived from Eukaryota)	139
Chlorophyta	127
Gemmatimonadetes	90
Spirochaetes	73
Nitrospirae	56
Crenarchaeota	44
Unclassified (derived from unclassified sequences)	43
Thermotogae	40
Synergistetes	39
Deferribacteres	30
Cnidaria	25
Thaumarchaeota	25
Chrysiogenetes	24
Hemichordata	23
Platyhelminthes	23
Chlamydiae	21
Aquificae	20
Echinodermata	18
Mollusca	12
Lentisphaerae	10
Annelida	8
Dictyoglomi	8
Microsporidia	7
Poribacteria	7
Phaeophyceae	6

Table S8. The top 50 most abundant phyla of frozen2

Phylum	Abundance
Proteobacteria	1888177
Chordata	81677
Actinobacteria	72001
Firmicutes	6593
Basidiomycota	4853
Apicomplexa	3403
Unclassified (derived from Viruses)	2873
Streptophyta	2713
Bacteroidetes	2503
Cyanobacteria	2247
Ascomycota	1536
Acidobacteria	1419
Deinococcus-Thermus	1411
Chloroflexi	1219
Unclassified (derived from Bacteria)	869
Tenericutes	832
Verrucomicrobia	790
Planctomycetes	780
Arthropoda	668
Chlorobi	560
Unclassified (derived from other sequences)	518
Euryarchaeota	496
Gemmatimonadetes	227
Fusobacteria	199
Spirochaetes	187
Nematoda	183
Nitrospirae	179
Synergistetes	172
Chlorophyta	151
Cnidaria	131
Unclassified (derived from unclassified sequences)	130
Unclassified (derived from Eukaryota)	129
Crenarchaeota	126
Thermotogae	106
Deferribacteres	87
Aquificae	76
Chlamydiae	65
Hemichordata	61
Chrysiogenetes	44
Echinodermata	32
Bacillariophyta	28
Dictyoglomi	24
Annelida	23
Poribacteria	23
Platyhelminthes	21
Lentisphaerae	16
Thaumarchaeota	16
Microsporidia	15
Phaeophyceae	10
Fibrobacteres	8

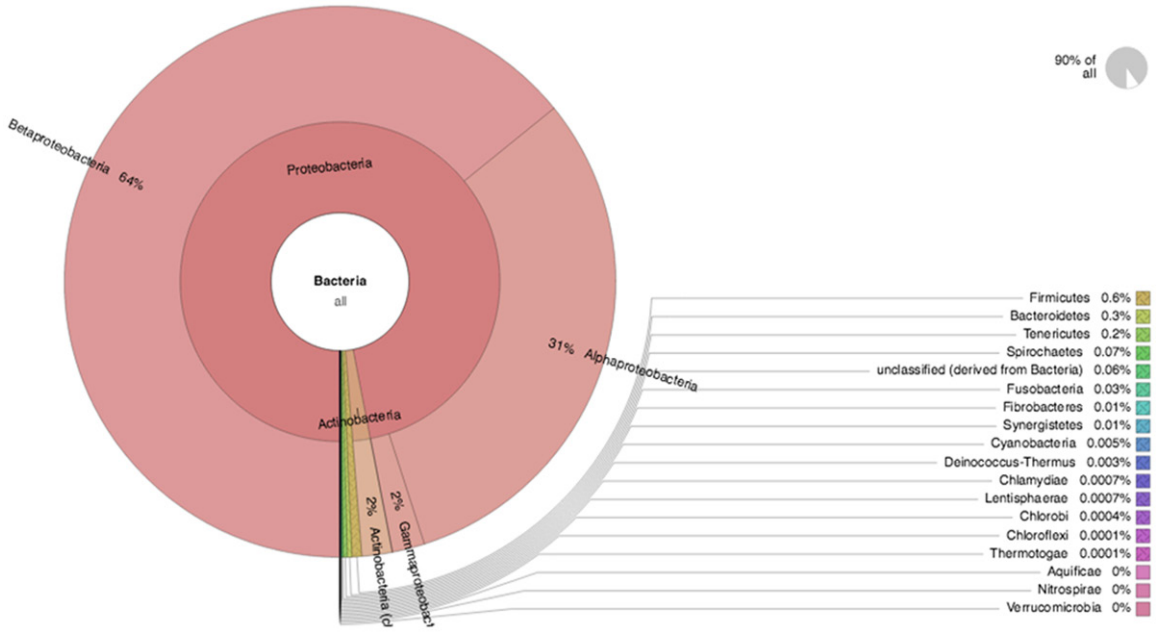


Figure S1. Ratio of Phylumin bacteria (%) of fresh sample from patient 1.

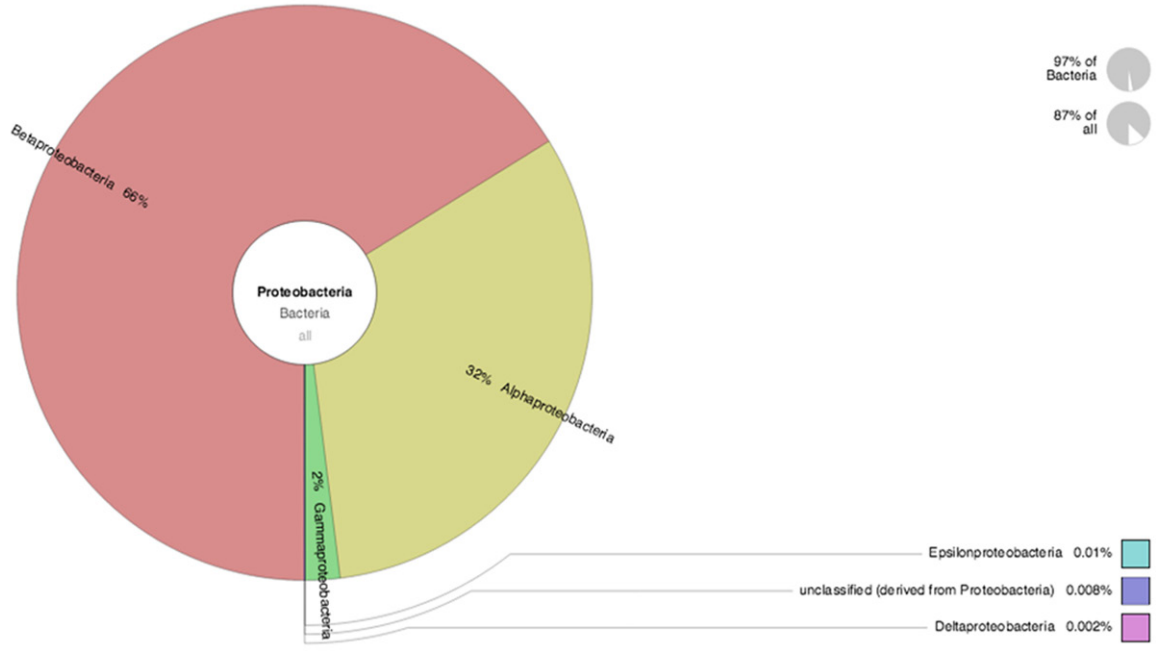


Figure S2. Ratio of Class in Proteobacteria (%) of fresh sample from patient 1.

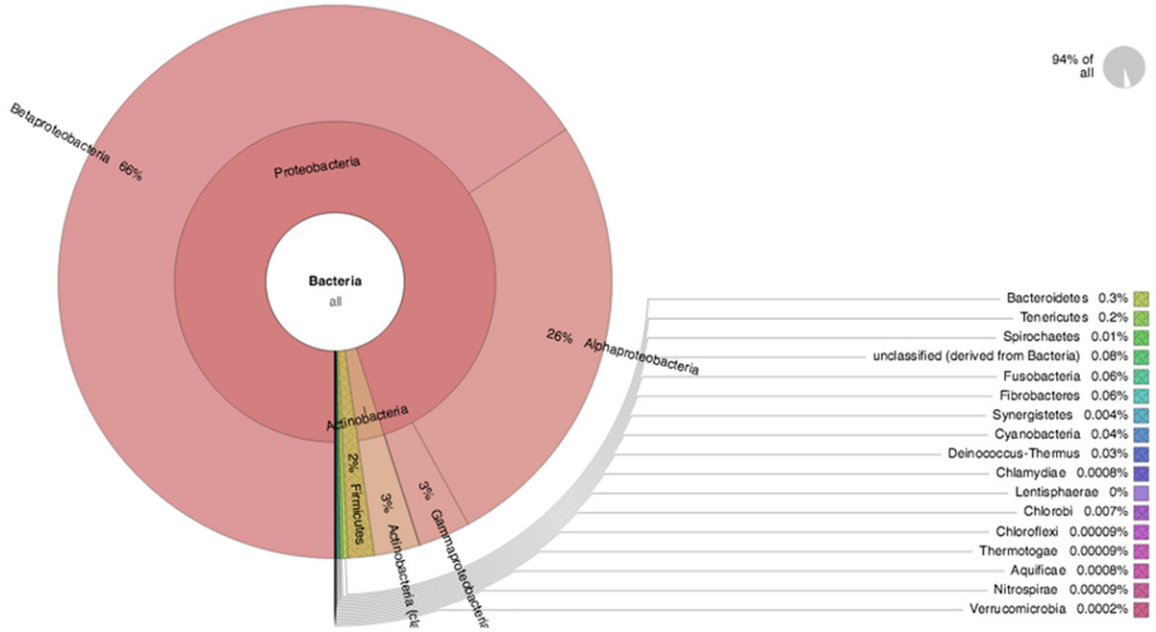


Figure S3. Ratio of Phylumin bacteria (%) of frozen sample from patient 1.

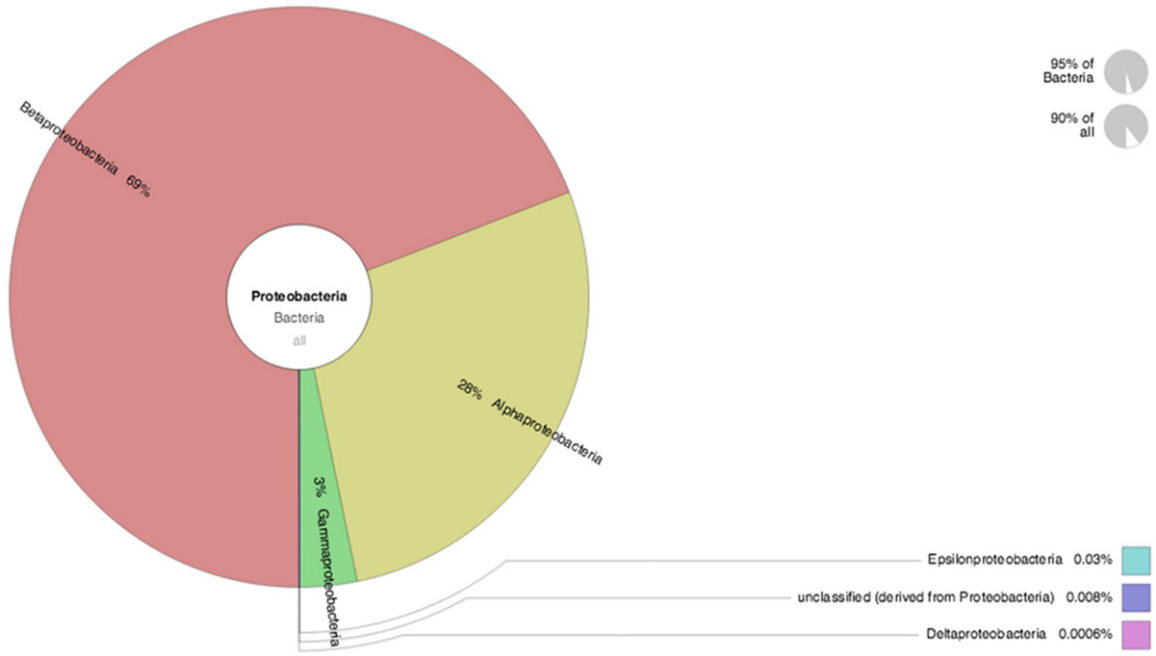


Figure S4. Ratio of Class in Proteobacteria (%) of frozen sample from patient 1.

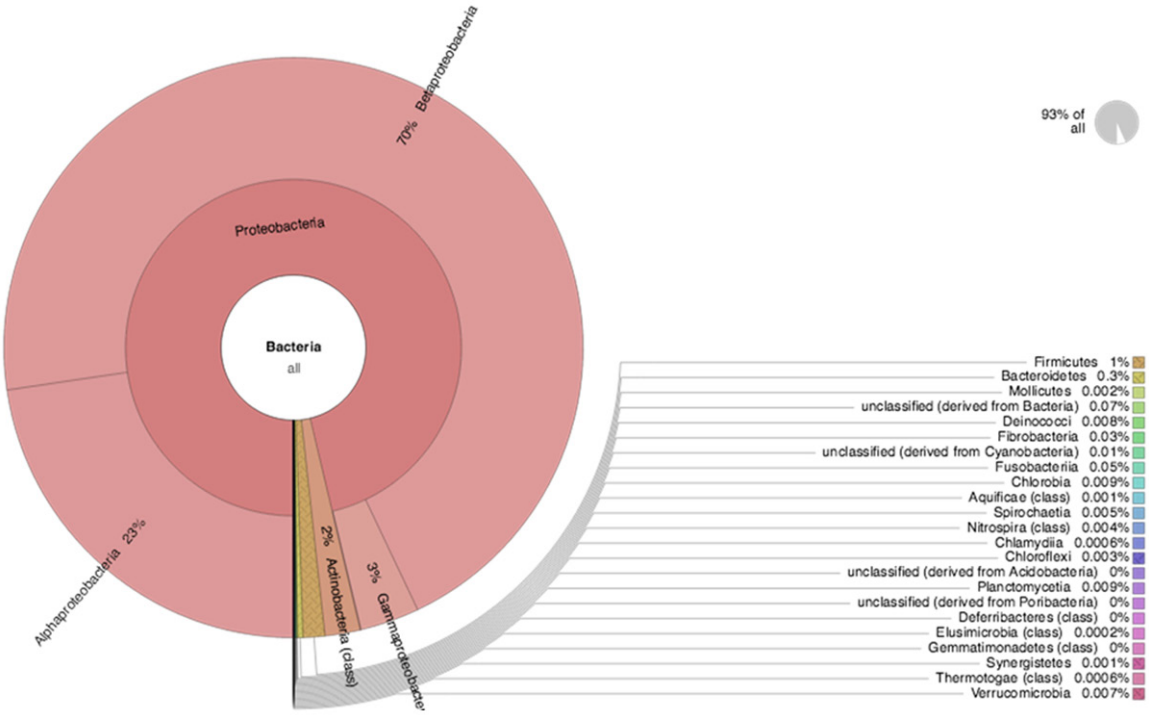


Figure S5. Ratio of Phylumin bacteria (%) of fresh sample from patient 2.

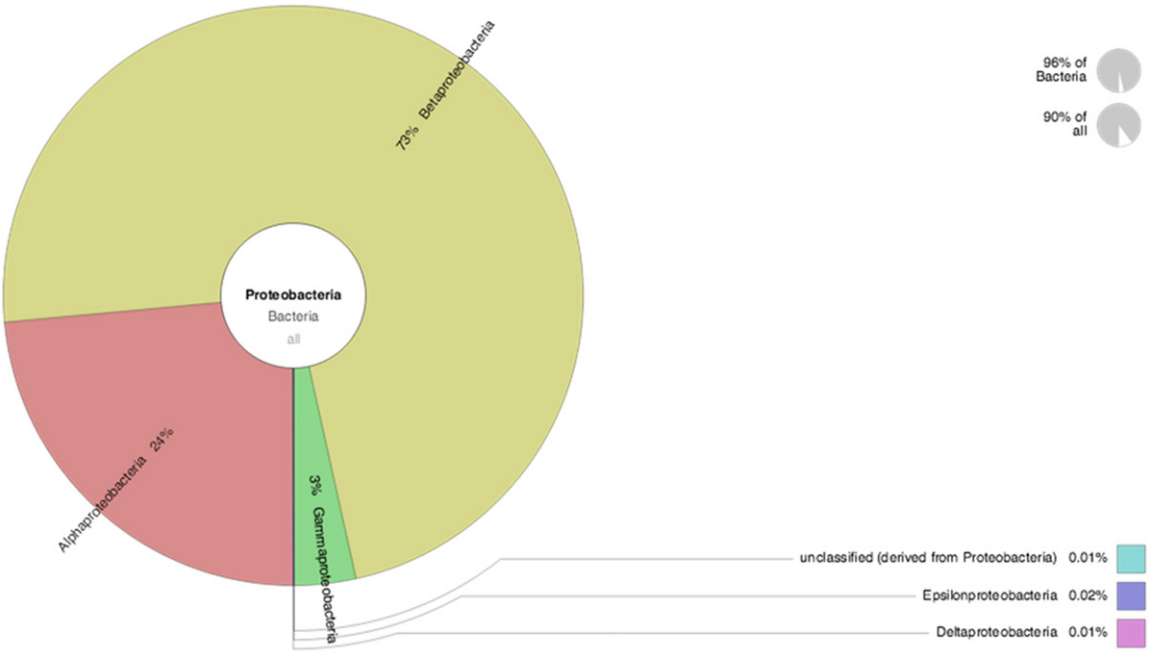


Figure S6. Ratio of Class in Proteobacteria (%) of fresh sample from patient 2.

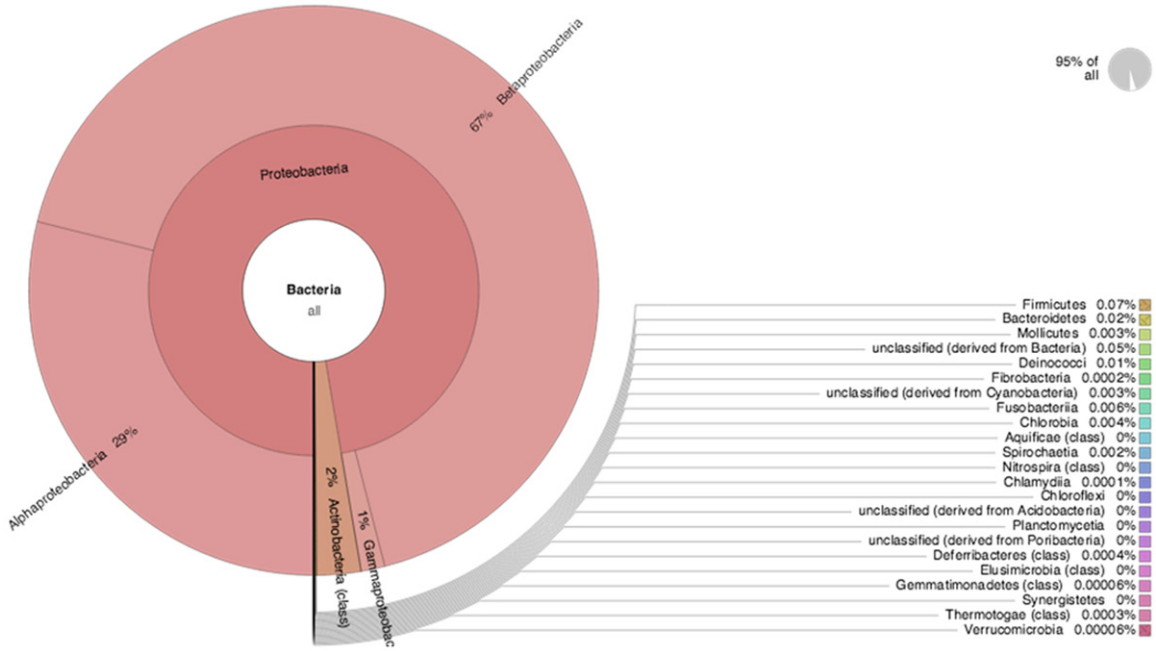


Figure S7. Ratio of Phylumin bacteria (%) of frozen sample from patient 2.

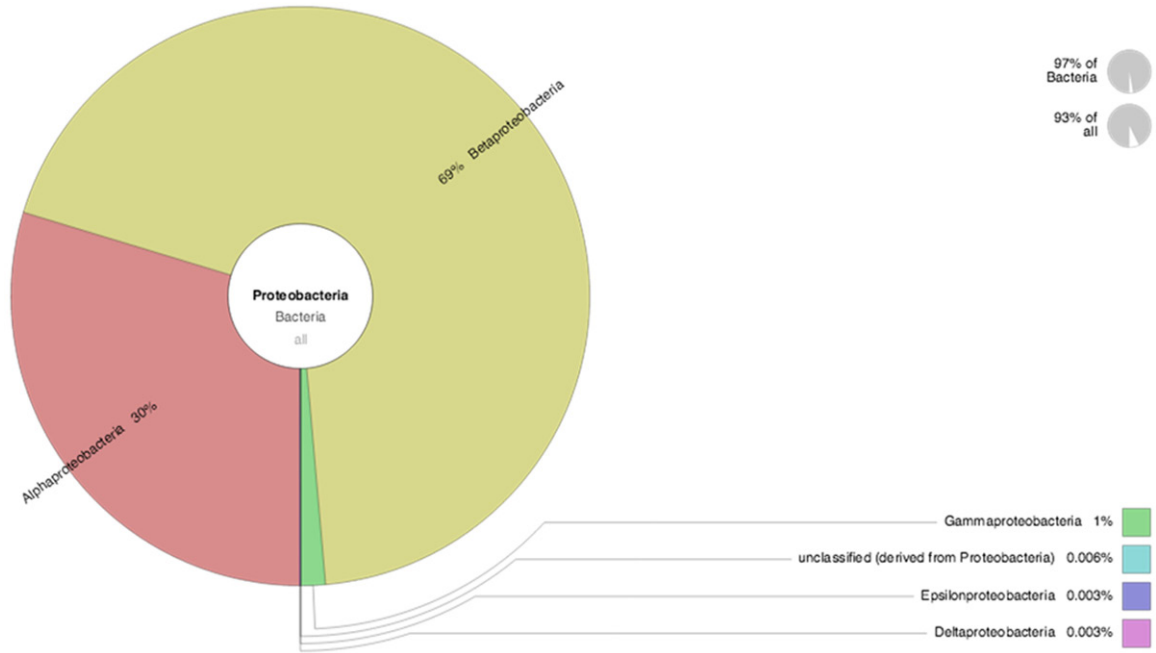


Figure S8. Ratio of Class in Proteobacteria (%) of frozen sample from patient 2.