*Epidemiology and Infection*

Consequences of organ choice in describing bacterial pathogen assemblages in a rodent population.

P. Villette, E. Afonso, G. Couval, A. Levret, M. Galan, C. Tatard, J.F. Cosson, P. Giraudoux

**Supplementary Material 1 - Sequencing methods**

PCR amplification, indexing, multiplexing and demultiplexing followed Kozich et al. [1] with modifications detailed by Galan et al. [2]. Negative controls and general procedures for removing biases and ensuring the quality of the data for later interpretation followed the guidelines in Table 1 of Galan et al. [2]; we used *Mycoplasma putrefaciens*, *Mycoplasma mycoides*, *Borrelia burgdorferi,* and *Bartonella* *taylorii*, for positive PCR amplification controls.

We used primers 16S-V4F/16S-V4R dual-indexed with 8bp-indices (i5 and i7) and Nextera Illumina adapters to perform two PCR replicates for each sample in 5 μL of 2 × Multiplex PCR Kit (Qiagen), 4 μL of combined forward and reverse primers (2.5 μM) and 2 μL of DNA extract. The PCR program consisted of an initial denaturation step at 95°C for 15 minutes, followed by 40 cycles of denaturation at 95°C for 20s, annealing at 55°C for 15s, and elongation at 72°C for 5 minutes, followed by a final extension step at 72°C for 10 minutes. PCR amplification was verified by electrophoresis in a 1.5% agarose gel. Pooled amplicons were size-selected by excision after gel electrophoresis with 1.25% agarose gel in 1% TAE buffer, and were purified using the NucleoSpin Gel clean-up kit (Machery-Nagel Sarl, France). We performed quantitative PCR with the KAPA library quantification kit (KAPA Biosystems) as per [2] to quantify the DNA of the final library, and then loaded the library on the Illumina MiSeq flow cell using a 500 cycle reagent cartridge and 2 x 251bp paired-end sequencing, which yielded high-quality sequences through the reading of each nucleotide of the V4 fragments twice after the assembly of reads 1 and reads 2. The raw sequence reads are available in FASTQ format in the Dryad Digital Repository.

Sequence analysis and taxonomic classification were performed using the mothur program package [3] and following the standard operation procedure of Patrick D. Schloss [1]. We used the Silva SSU Reference database v119 [4] for taxonomic assignment. To discard false negatives introduced by biases in the procedure described above, we filtered the read-abundance table generated by mothur for each PCR product and each OTU following Galan et al. (2016). We used negative controls for both DNA extraction (1 well per 96-well extraction plate) and PCR amplification (1 well per 96-well PCR plate) to estimate cross-contamination rates for individual OTUs, and we used a false-assignment rate of 0.02%to filter read abundances. To account for potential false-positives, we considered a sample positive for a given OTU only if that OTU was present in both PCR replicates after the filtering steps described above were applied. Our mycoplasma positive controls were assigned to Entomoplasmataceae rather than Mycoplasmataceae due to a frequent taxonomic error reflected in most databases, including Silva (23).

We classified OTUs as potentially pathogenic using the 2nd edition of Bergey's Manuel of Systematic Bacteriology, Volumes 1 to 5 [5–9], Pathogenesis of Bacterial Infections in Animals [10] and a general search of the literature. If we found documentation implicating a genus in an infection or disease in mammals, all OTUs in that genus were designated as potentially pathogenic. If an OTU could not identified to the genus level, and was assigned to a clade that has no known pathogenic members, these OTUs were discarded from the analysis as contaminants, commensals, or rare taxa of unknown function. For the remaining unclassified OTUs that were assigned to clades that contain pathogenic members, we compared individual sequences from these OTUs to the NCBI’s GenBank database using BLAST [11], and potential pathogenicity status was assigned when a sequence matched a known pathogen, or matched a sequence implicated in an animal infection or disease.

Summaries of the number of reads produced and retained during the analysis are provided in Supplementary Table 1. The total number of reads retained in each sample after data filtering procedures were performed are summarized in Supplementary Table 2. A brief summary of the OTUs removed from the data set during data filtering procedures is provided in Supplementary Table 3.

**REFERENCES**

1. **Kozich JJ et al**. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*. 2013; **79**: 5112–5120.

2. **Galan M et al**. 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife: the importance of cleaning post-sequencing data before estimating positivity, prevalence and co-infection. *mSystems*. 2016; **1**: e00032-16.

3. **Schloss PD et al**. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*. 2009; **75**: 7537–7541.

4. **Quast C et al**. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*. 2013; **41**: 590–596.

5. **Boone DR, Castenholz RW, Garrity GM (eds)**. *Bergey’s Manual of Systematic Bacteriology. Volume 1: The Archaea and the deeply branching and phototrophic bacteria*. New York, New York: Springer; 2009.

6. **Brenner DJ et al (eds)**. *Bergey’s Manual of Systematic Bacteriology. Volume 2: The Proteobacteria, Part C: The Alpha-, Beta-, Delta-, and Epsilon-proteobacteria*. New York, New York: Springer; 2005.

7. **De Vos P et al (eds)**. *Bergey’s Manual of Systematic Bacteriology. Volume 3: The Firmicutres*. New York, New York: Springer; 2009.

8. **Krieg NR et al (eds)**. *Bergey’s Manual of Systematic Bacteriology. Volume 4: The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*. New York, New York: Springer; 2010.

9. **Goodfellow M et al (eds)**. *Bergey’s Manual of Systematic Bacteriology. Volume 5: The Actinobacteria, Part A and B*. New York, New York: Springer; 2012.

10. **Gyles CL et al (eds)**. *Pathogenesis of bacterial infections in animals*. Ames, Iowa, USA: Wiley-Blackwell; 2010.

11. **Altschul SF et al**. Basic local alignment search tool. *Journal of Molecular Biology*. 1990; **215**: 403–410.

|  |  |
| --- | --- |
| Supplementary Table 1: Summary of the number of reads produced and retained by our sequencing protocol and post-sequencing data filtering and analysis procedures. | |
| Total number of reads before filtering | 1154658 |
| Total number of reads removed after filtering procedures | 897725 (77.7%) |
| Total number of reads retained after filtering procedures | 256933 (22.3%) |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Supplementary Table 2. Read totals for each sample after data filtering. | | | | | | |
| Location | Animal | Heart | Liver | Lung | Kidneys | Spleen |
| A | 1 | 7658 | 0 | 1784 | 2042 | 622 |
| A | 2 | 44 | 40 | 0 | 1148 | 22 |
| A | 3 | 2062 | 0 | 0 | 11 | 17277 |
| A | 4 | 9974 | 3160 | 5008 | 1805 | 1872 |
| A | 5 | 24929 | 10550 | 14762 | 8640 | 13649 |
| A | 6 | 1296 | 1396 | 955 | 763 | 40888 |
| A | 7 | 0 | 0 | 0 | 8 | 0 |
| A | 8 | 4403 | 343 | 335 | 6144 | 653 |
| A | 9 | 35 | 0 | 20950 | 0 | 0 |
| C | 10 | 44 | 105 | 0 | 121 | 698 |
| C | 11 | 279 | 103 | 2769 | 20500 | 1132 |
| C | 12 | 2114 | 3031 | 6057 | 10216 | 31 |
| C | 13 | 4207 | 0 | 259 | 10 | 29 |

|  |  |
| --- | --- |
| Supplementary Table 3: The number of OTUs per bacterial phylum that were removed from the data set during data filtering procedures. | |
| Phylum | Number of OTUs |
| Acidobacteria | 45 |
| Actinobacteria | 131 |
| Bacteroidetes | 302 |
| Candidate\_division\_OD1 | 2 |
| Candidate\_division\_WS3 | 2 |
| Chlamydiae | 2 |
| Chlorobi | 1 |
| Chloroflexi | 10 |
| Cyanobacteria | 18 |
| Deferribacteres | 3 |
| Deinococcus-Thermus | 2 |
| Elusimicrobia | 3 |
| Firmicutes | 474 |
| Gemmatimonadetes | 7 |
| Nitrospirae | 1 |
| Planctomycetes | 42 |
| Proteobacteria | 249 |
| SHA-109 | 1 |
| Spirochaetae | 1 |
| Tenericutes | 25 |
| TM6 | 5 |
| unclassified | 100 |
| Verrucomicrobia | 30 |
| WCHB1-60 | 1 |