Microbial contamination screening and interpretation for biological laboratory environments

Xi Li^{1, #}, Xue Zhu¹, Wenjie Wang¹, Kang Ning^{1, *}

¹ Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[#] These authors contributed equally to this work

^{*} Corresponding author. E-mail: ningkang@hust.edu.cn

Abstract

Advances in microbiome researches have led us to the realization that the composition of microbial communities of indoor environment is profoundly affected by the function of buildings, and in turn may bring detrimental effects to the indoor environment and the occupants. Thus investigation is warranted for a deeper understanding of the potential impact of the indoor microbial communities. Among these environments, the biological laboratories stand out because they are relatively clean and yet are highly susceptible to microbial contaminants. In this study, we assessed the microbial compositions of samples from the surfaces of various sites across different types of biological laboratories. We have qualitatively and quantitatively assessed these possible microbial contaminants, and found distinct differences in their microbial community composition. We also found that the type of laboratories has a larger influence than the sampling site in shaping the microbial community, in terms of both structure and richness. On the other hand, the public areas of the different types of laboratories share very similar sets of microbes. Tracing the main sources of these microbes, we identified both environmental and human factors that are important factors in shaping the diversity and dynamics of these possible microbial contaminations in biological laboratories. These possible microbial contaminants that we have identified will be helpful for people who aim to eliminate them from samples.

Key words: Biological laboratory; Microbial contamination; Metagenomics; Screening and interpretation.

32 **Importance**

33 Microbial communities from biological laboratories might hamper the conduction of molecular
34 biology experiments, yet these possible contaminations are not yet carefully investigated. In this
35 work, a metagenomic approach has been applied to identify the possible microbial contaminants
36 and their sources, from the surfaces of various sites across different types of biological laboratories.
37 We have found distinct differences in their microbial community compositions. We have also
38 identified the main sources of these microbes, as well as important factors in shaping the diversity
39 and dynamics of these possible microbial contaminations. The identification and interpretation of
40 these possible microbial contaminants in biological laboratories would be helpful for alleviate their
41 potential detrimental effects.

42

43 **Introduction**

44 Indoor environments are important since most of us spend his/her time indoor for the most part of
45 his/her life[1]. The microbial communities of these environments are of particular interests; in-
46 depth studies of environmental microbes in the last decade have shed light on the subtle effects they
47 have on human health[2]. For example, a chronic exposure to some fungi can cause asthma, but
48 early life exposure to various mold and its derivatives can protect children from allergic and
49 autoimmune diseases[3]. A growing number of studies have helped us estimate the microbial
50 diversity in various indoor environments, and revealed that microbial diversity is closely related to
51 the geographic locations[4], weather conditions[5, 6], populations[7], functions[8], and internal
52 ventilation conditions[9].

53 Ironically, the microbial compositions from indoor environment in various types of biological
54 laboratories are less well-understood. While microbial contaminants generally exist in molecular
55 biology laboratories[10], few studies have been dedicated to study their microbial compositions.
56 Biological laboratory contamination screening is an important task. Once a site is contaminated
57 during the sampling process or the experiment procedure, the contaminants of the reagent or the
58 environmental microbes may proceed to affect other samples, leading to biases in the results. It
59 would also be intriguing to examine the hypothesis that each laboratory has a relatively stable
60 microbial contamination, determined by various factors including the research subjects (such as
61 animals, plants or microbes), personal factors, as well as macroscopic environment. Each type of

62 microbial composition can then be used to characterize its associated type of laboratories, and help
63 simplify future studies.

64 There are several approaches in the identification and quantification of microbial contaminants. The
65 most commonly used technique is based on PCR amplification and sequencing of the genes which
66 encode small subunit ribosomal RNA (16S rRNA). The alternative is the metagenomics approach,
67 which sequences the DNA of the entire microbial community as a whole. Compared to culture-
68 based approaches, metagenomic approaches are better for identifying novel organisms with
69 unknown growth conditions[11]. High-throughput sequencing allows metagenomic approach to
70 obtain all the genome information of the community in one experiment, enabling us to study the
71 complex molecular interactions among species.

72 However, there are several difficulties in our application of the metagenomic approach. First,
73 significant amount of microbial contaminants may be introduced during sample preparation,
74 especially when sample has low microbial biomass. Second, unlike other well-studied environments,
75 there is no catalog for quick screening of possible microbial contaminations from biological
76 laboratory. Hence, it is imperative for us to design methods that could accurately identify microbial
77 contaminants, trace the pollution source, and uncover their potential adverse effects.

78 To work out these problems, we collected samples from surfaces of several important sites (lab
79 outlet, platform and the major public areas) of three types of biological laboratories (animal, plant
80 and microbe), screened and annotated the microbial contaminants, identified the difference between
81 sampling sites/laboratories, as well as discovered the microbial biomarkers for different types of
82 biological laboratories. We also identified possible sources of these microbes, as well as the
83 possible effects they may have on their occupants.

84

85 **Results and Discussions**

86 **Compositions of microbial communities from different laboratories and different sampling** 87 **sites**

88 We obtained 759,612 high-quality 16s rRNA sequences in total for 37 samples. 724,126 sequences
89 were retained after quality filtering, and all samples have reached the saturation plateau for
90 sequencing, indicating enough sequences for 16s rRNA profiling. Among all sequencing data,
91 432,092 sequences were from microbiology laboratory (ML), 137,575 from animal laboratory (AL)

and 154,460 from plant laboratory (PL). Then all sequences were clustered into 1,234 Operational Taxonomic Units (OTUs) at 97% similarity threshold. In order to ensure enough sequencing depth, we generated the rarefaction curves for each sample. At around 1,800 sequences per sample, most rarefaction curves showed saturation, suggesting that the depth of samples sequencing covered enough extent of taxonomic diversity.

To compare the microbial composition of all microbial contaminant samples from the animal, plant and microbe laboratories, the taxonomies at phylum- and genus- level were illustrated (**Figure 1**). The microbial communities are composed mainly of 6 different bacterial phyla, including *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes* and *Fusobacteria*, with differentiated proportions in each sample. *Actinobacteria* was the most abundant phylum across all samples (**Figure 1a**). At the genus level, *Proteus*, *Prevotella*, *Chryseobacterium*, *Methylobacterium*, *Acinetobacter*, *Enterobacter*, *Micrococcus*, *Rhodococcus*, *Stenotrophomonas* and *Staphylococcus* were the dominant components (**Figure 1b**). The microbial communities from various sites at genus level were very diverse, even from the same type of laboratory.

The relationship of microbial community composition, laboratory type and sampling sites

The type of laboratories carry more weight than sampling sites in the differentiation of microbial community samples. Alpha diversity analysis was performed (**Supplementary Table 1**), followed by the analysis of variance (ANOVA), to detect differences among samples from different sites and laboratories (**Figure 2**). Chaos indices showed that there is significant differences between AL and ML (**Figure 2a**). Shannon indices showed that significant difference in the platform between ML and PL (**Figure 2b**). Furthermore, the number of OTUs determined by the Observed_OTUs revealed a clear difference from the major public areas between AL and PL (**Figure 2c**). **Figure 2** also shows that the samples from different types of laboratories could always be distinguished, whereas the samples within the same type of laboratories are usually indistinguishable except for the lab outlet and platform of PL. Thus, the differences in microbial community composition of samples across different types of laboratories are clear, while within laboratory differences are relatively small.

To gain further insights into the differences between laboratories, a comparison of samples from the same type sampling site across different types of biological laboratories was conducted. The results showed that these samples composed of many similar genus, but the proportion of each genus was different (**Figure 3; Table 1**). *Pseudomonas*, *Cinetobacter*, *Enterobacter* and *Micrococcal* were

124 ubiquitous bacterial genus with dominant occurrence on the platform and lab outlet (**Figure 3a-c**).
125 In addition, while the total number of detected genus are similar among lab outlet (76), public area
126 (81) and platform (79), the number of shared genus is largest in public area (39), and smallest in lab
127 outlet (22) (**Figure 3d-f**). Moreover, for either of the sampling sites including lab outlet, public area
128 and platform, PL has much less laboratory-specific genus compared to AL and ML (**Figure 3d-f**).
129 Therefore, we speculated that while public areas shared by experimenters might have largest
130 number of shared genus, key sites such as lab outlet and platform has their specific sets of genus as
131 potential contaminations.

132 We next compared the relative abundances of representative genus from three main sites within the
133 laboratory. *Pseudomonas*, *Acinetobacter* and *Enterobacter* were most abundant among all sampling
134 sites (**Figure 4a-c**). In addition, the number of all identified genus in AL (84) and ML (87) were
135 much more than those in PL (53) (**Figure 4d-f**). Moreover, the platform of AL has the highest
136 number of site-specific genus (**Figure 4d-f**). These results again confirm that the richness of
137 microbial communities of platform and lab outlet depended heavily on the type of laboratory.

138

139 **Possible sources and microbial biomarkers for different types of laboratories**

140 We then performed literature mining to identify the possible sources of these microbial
141 contaminations, referencing various sources. We categorized the sources into laboratory reagent
142 microbe, human-introduced microbe, and basic environmental microbe. Interestingly, laboratory
143 reagents and human daily activities might play very important roles in introducing these possible
144 microbial contaminations (**Table 1**).

145 To obtain a characteristic set of microbial contaminants, or biomarkers, for each type of biological
146 laboratory, we used LDA Effect Size (LEfSe) to discover the biomarkers at each taxonomic level.
147 29 taxa (7, 15 and 7 taxa from AL, ML and PL respectively; **Figure 5**) were detected with high
148 LDA scores. For samples from AL, *Bacteroidetes*, *Flavobacteriaceae* and *Gemmata* were
149 identified as biomarkers. *Enterobacteriales* and *Enterobacteriaceae* were identified as biomarkers
150 for ML. *Pseudomonas*, *Pseudomonadaceae* and *Pseudomonadales*, which belong to *Pseudomonas*,
151 were identified as biomarker for PL with high confidence (**Figure 5a**). The evolutionary
152 relationship between these bacteria at different taxonomic levels is shown in **Figure 5b**.

153

154 To further explore the characteristics of the biomarkers for different laboratories, we screened the
155 genera with a relative abundance of $> 1/1000$ within the same type of laboratory. This identified ML
156 to contain a greater variety of bacteria (65 genera) than AL (59) and PL (48). The population of the
157 overlap between the detected genera of the three types of laboratories was 39 (**Table 1**), the highest
158 was found between ALs/MLs (9 shared genera) and followed by ALs/PLs (3) and MLs/PLs (1), 16
159 specific genera in MLs, more than ALs (8) and PLs (5). Comparing against the references tables
160 including reagents[12] (**Supplementary Table 2**), residential[13] (including daily residential areas,
161 office and classroom; **Supplementary Table 3**) and detected in ICU[14] contamination table
162 (**Supplementary Table 4**), the shared genera exhibited significant overlap, while the specific
163 genera did not. For laboratory-specific genera, only *Methyloversatilis* and *Psychrobacter* from AL
164 was detected in reagent (representing laboratory reagent microbes) and ICU contamination table
165 (representing basic environmental microbes) as mentioned above, while *Bacillus* from PL was
166 observed in three reference tables, and *Flavisolibacter* from ML was only present in ICU
167 contamination table. Therefore, we speculated that the overlapping specific- and shared- genera
168 should be ubiquitous bacteria in the environment, lab reagents contaminants or external bacteria
169 introduced by human activities.

170 Through literature mining, we assessed the possible effects of laboratory-specific genera (**Table 2**)
171 without any overlap with the three reference tables. Specific bacteria of laboratory will assert
172 adverse effects on researchers or experimental materials. To illustrate, the *Jeotgalicoccus* of AL as a
173 pathogen, can be transmitted via air or surfaces contact and hence infects hosts; *Moraxella* of PL
174 could influence the onset of bronchitis or pneumonia. Other microbes are less harmful; for instance,
175 *Psychrobacte* of AL is a probiotic of fish, and its highest diversity was detected in sample A1B1,
176 corresponding to the incubator of the zebrafish laboratory by backtracking analysis. *Buchnera* of PL,
177 a symbiotic bacterium of aphids is specifically associated with the tissue culture process.
178 *Flavisolibacter* of ML, which improves nitrogen fixation in rhizosphere of plants, has the highest
179 abundance in sample M1W12, which was from cultivated plants on the windowsill in the M1
180 laboratory. Together, these results showed high concordance between the characteristics of the
181 laboratory and the sampling site, demonstrating that the compositions of microbial communities
182 have profound association with their hosting laboratories.

183 As already known, the present of these contaminants can bring inconvenient for our experiment
184 more or less, so caution and preciseness must be followed throughout the whole experiment. And
185 the use of blank control during sampling, DNA extraction and sequencing is also necessary for
186 detecting contamination. Furthermore, he contaminants are associated with the use of different kits
187 or batches, which can introduce variation in reagent contamination[12], therefore, it would be best

188 to use the same kits in one experiment and disentangle batch effects. Additionally, we should
189 catalogue the laboratory microbial contaminants better, and thus, as if we know the contaminants,
190 antibiotic treatment can be executed before experiment to mitigate the experimental bias caused by
191 these microbial contaminants.

192

193 **Conclusion**

194 In this work, a metagenomic approach has been applied to identify the possible microbial
195 contaminants and their sources, from the surfaces of various sites across different types of
196 biological laboratories. The possible microbial contaminants that we have identified will be helpful
197 for people who aim to eliminate them from samples.

198 As far as we know, our work is the first investigation on the composition of microbial communities
199 in biological laboratories. We found several interesting patterns in these compositions. First, there
200 are significant differences in the structures of the microbial communities from the three types of
201 laboratories. Factors such as sampling sites (including lab outlet, platform and the major public
202 areas) and laboratory types (for animal, plant and microbe), have influenced the compositions of
203 indoor microbial communities: the number of microbial genus in animal and microbial laboratories
204 are significantly higher than those in plant laboratories, while key sites such as lab outlet and
205 platform have their specific sets of genus as potential contaminations for each type of laboratory.
206 These differences are highly related to the functions of the laboratories. Second, the type of
207 laboratories has more influence than sampling sites in the differentiation of microbial community
208 samples. Third, while public areas shared by experimenters may have the largest number of shared
209 genus, key sites such as lab outlet and platform have their specific sets of genus as potential
210 contaminations for each type of laboratory. This suggests that while general human activities have
211 the most effect on the microbial community structure of the laboratory, the microbial communities
212 of platform and lab outlet depends more heavily on the type of the laboratory. Finally, by tracking
213 the possible sources of laboratory microbes, we found that laboratory reagents and human daily
214 activities might play very important roles in introducing these possible microbial contaminations.
215 These microbes are intimately connected with the experimental materials, and will also assert
216 negative effects on the experiment process as well as on experimenters.

217 We would like to suggest two directions in future analysis of possible microbial contaminations
218 from laboratories. First, better profiles of the microbial compositions in the biological laboratories

219 are needed. They would help in devising countermeasures to mitigate the experimental bias caused
220 by these microbial contaminants. Second, we hope that longitudinal studies would help to confirm
221 our findings, since our samples were collected from the same building in summer and may not
222 reflect the seasonal dynamics of the microbial communities.

223

224 **Materials and Methods**

225 **Experimental design and sample collection.** We selected 8 laboratories (3 animal laboratories, 3
226 microbial laboratories, 2 plant laboratories) from the College of Life Science and Technology,
227 Huazhong University of Science and Technology in Wuhan, Hubei province of China. All
228 laboratories are in the same building, but distributed at different floors. We collected samples from
229 the lab outlet (e.g. doors, windows) with high air mobility, the platform (e.g. processing table, clean
230 bench), and the major public areas (e.g. floor, pool and preprocessing pond). We conducted all
231 sampling and genome extraction in July of 2017 to avoid the influence of environmental and
232 climate factors. The overall experimental design, and main methods of our project are shown in
233 **Figure 6.**

234 All samples were wiped on the selected surface areas and devices with 4 to 5 swabs that were
235 moistened with a 15 mL centrifugal tube containing 2.5 mL of normal saline. All sampling locations,
236 primer used and their characteristics are listed in **Supplementary Table 5**. During sampling, all the
237 staff and devices were in full operation (normal status). After sampling, the swabs were kept at 4 °C.
238 Afterwards, the genome of all samples extracted by using biological sampling kit was stored at -
239 20 °C.

240 **DNA extraction.** First, 1 mL sample, in total, and 1 mL buffer was added into centrifuge tube, and
241 the mixture was stirred gently. After water bath of 2 h 65 °C, mixing by hand every 30 min, the
242 suspension was vortexed for 10 seconds. The tube was placed on ice for 10 min and centrifuged
243 afterwards (100 g, 5 min, 4 °C). The supernatant was transferred into another tube, and an equal
244 volume of phenol: chloroform: isoamyl alcohol in a ratio of 25:24:1 was added. The suspension was
245 mixed gently and centrifuged at 15°C/1000g for 5 min. The aqueous layer was transferred into a
246 new tube. Then, the same volume of isopropanolis was added to cause the DNA to precipitate out of
247 the aqueous solution. After incubation at -20 °C overnight, the suspension was centrifuged at
248 4°C/13500g for 30 min. After removing the supernatant, the precipitate was rinsed with 1 mL of

249 70% ethanol, and centrifuged repeated at 4 °C/13500 g for 30 min until the precipitate was
250 completely dried and re-dissolved in 20 µL of PCR-grade water for easy handling and storage.

251 **16S rRNA gene amplification and Illumina sequencing.** For Illumina sequencing, 16 rRNA gene
252 was amplified in the PCR reaction mixture (20 µl), which contained 1 µl Taq polymerase, 0.25 µl of
253 forward primer, 0.25 µl of reverse primer, 0.5 µl of Dntp, 1 µl of template DNA, 5 µl 10×buffer and
254 12 µl ddH₂O. To reduce the nonspecific amplification, the PCR system was made up on the ice box.
255 The amplification process is as follows: 95 °C for 5 min, 25 cycles of 94 °C for 30 s, 54 °C for 40 s,
256 72 °C for 1 min, then followed by 30 cycle of 72 °C for 10 min and 4 °C hold. Amplification
257 products were visualized with e gel. After quality filtering, the products was purified using the kits,
258 and restored at -20 °C, then sent to company for Illumina sequencing. All sequencing data are
259 deposited to NCBI SRA with project accession number PRJNA490598.

260 **Bioinformatics and statistical analysis of sequencing results.** After obtaining the sequencing data
261 for these samples, we used FastQC(<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)[15]
262 to perform preliminary quality control and filtering of the data.

263 QIIME (Quantitative Insights Into Microbial Ecology; <http://qiime.org/>)[16] is used for 16S rRNA
264 profiling. Using the QIIME script join_paired_ends.py to process the double-ended sequences,
265 merge them, and make the Mappingfile containing SampleID, BarcodeSequence,
266 LinkerPrimerSequence, ReversePrimer, Description information. Then, we used
267 validate_mapping_file.py to check the Mappingfile, and marked the wrong locations in the finally
268 Mappingfile.html. Based on the extracted barcode information by referencing the Mappingfile with
269 extract_barcodes.py, we split the sample by split_libraries_fastq.py, where the quality threshold was
270 set to 20 (99% accuracy), then removed chimeras and length-marginized sequences.

271 Four common alpha diversity metrics and pick_de_novo_otus.py were used to extract OTUs from
272 the Fasta file, removed the single reads from OUTs and obtained the rarefaction curve of the sample
273 to determine the depth of the sequencing by filter_otus_from_otu_table.py and alpha_rarefaction.py.
274 Biome summarize-table for counting the number, average number, and total number of sequences
275 contained in each sample, alpha_diversity.py and beta_diversity_through_plots.py for analyzing the
276 diversity of samples. Statistical analysis and visualization were then performed in R (<https://www.r-project.org/>)[17] using the package ggplot2. We then used SPSS (<https://www.ibm.com/analytics/datascience-/predictive-analytics/spss>)[18] to perform ANOVA on the alpha diversity results of
278 samples to compare the difference of microbial community composition among the three sites.
279

280 LEfSe (LDA Effect Size; <http://huttenhower.sph.harvard.edu/galaxy/>) is used to find the biomarkers
281 in the sample. The input file was obtained by `summarize_taxa.py`. In each group, the biomarker,
282 LDA values and the hierarchical relationship between individual biomarkers were shown by
283 `run_lefse.py`, `plot_cladogram.py` and `plot_features.py` severally.

284

285 **Conflict of Interests**

286 The authors declare that they have no competing interests.

287

288 **Acknowledgments**

289 This work was partially supported by National Science Foundation of China grant 31871334 and
290 31671374, and Ministry of Science and Technology's grant 2014AA021502 and 2018YFC0910502.

291

292 **References**

- 293 1. Kelley ST, Gilbert JA: **Studying the microbiology of the indoor environment.** *Genome*
294 *Biology* 2013, **14**(2):1-9.
- 295 2. O'Hara NB, Reed HJ, Afshinnkoo E, Harvin D, Caplan N, Rosen G, Frye B, Woloszynek S,
296 Ounit R, Levy S: **Metagenomic characterization of ambulances across the USA.** *Microbiome*
297 2017, **5**(1):125.
- 298 3. Rook GA: **Review series on helminths, immune modulation and the hygiene hypothesis:**
299 **the broader implications of the hygiene hypothesis.** *Insect Science* 2010, **126**(1):3-11.
- 300 4. Chase J, Fouquier J, Zare M, Sonderegger DL, Knight R, Kelley ST, Siegel J, Caporaso JG:
301 **Geography and Location Are the Primary Drivers of Office Microbiome Composition.**
302 *Msystems* 2016, **1**(2):e00022-00016.
- 303 5. Proctor CR, Dai D, Edwards MA, Pruden A: **Interactive effects of temperature, organic**
304 **carbon, and pipe material on microbiota composition and Legionella pneumophila in hot**
305 **water plumbing systems.** *Microbiome* 2017, **5**(1):130.
- 306 6. Rintala H, Pitkäranta M, Toivola M, Paulin L, Nevalainen A: **Diversity and seasonal dynamics**
307 **of bacterial community in indoor environment.** *Bmc Microbiology* 2008, **8**(1):56.

- 308 7. Meadow JF, Altrichter AE, Kembel SW, Kline J, Mhuireach G, Moriyama M, Northcutt D,
309 O'Connor TK, Womack AM, Brown GZ: **Indoor airborne bacterial communities are**
310 **influenced by ventilation, occupancy, and outdoor air source.** *Indoor Air* 2014, **24**(1):41-48.
- 311 8. Korpi A, Pasanen AL, Pasanen P: **Volatile compounds originating from mixed microbial**
312 **cultures on building materials under various humidity conditions.** *Appl Environ Microbiol*
313 1998, **64**(8):2914-2919.
- 314 9. Gilbert JA, Stephens B: **Microbiology of the built environment.** *Nat Rev Microbiol* 2018.
- 315 10. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman
316 NJ, Walker AW: **Reagent and laboratory contamination can critically impact sequence-**
317 **based microbiome analyses.** *Bmc Biology* 2014, **12**(1):87.
- 318 11. Hugenholtz P, Goebel BM, Pace NR: **Impact of culture-independent studies on the emerging**
319 **phylogenetic view of bacterial diversity.** *Journal of Bacteriology* 1998, **180**(18):4765-4774.
- 320 12. Goffau MCD, Lager S, Salter SJ, Wagner J, Kronbichler A, Charnockjones DS, Peacock SJ,
321 Smith GCS, Parkhill J: **Recognizing the reagent microbiome.** *Nature Microbiology* 2018.
- 322 13. Meadow JF, Altrichter AE, Kembel SW, Moriyama M, O'Connor TK, Womack AM, Brown GZ,
323 Green JL, Bohannan BJM: **Bacterial communities on classroom surfaces vary with human**
324 **contact.** *Microbiome* 2014, **2**(1):7.
- 325 14. Oberauner L, Zachow C, Lackner S, Högenauer C, Smolle KH, Berg G: **The ignored diversity:**
326 **complex bacterial communities in intensive care units revealed by 16S pyrosequencing.** *Sci*
327 *Rep* 2013, **3**(3):1413.
- 328 15. Brown J, Pirrung M, Mccue LA: **FQC Dashboard: integrates FastQC results into a web-**
329 **based, interactive, and extensible FASTQ quality control tool.** *Bioinformatics* 2017, **33**(19).
- 330 16. Lawley B, Tannock GW: **Analysis of 16S rRNA Gene Amplicon Sequences Using the**
331 **QIIME Software Package.** *Methods in Molecular Biology* 2017, **1537**:153.
- 332 17. Sunil Bhavsar PD, Shantilal Tank: **R software package based statistical optimization of**
333 **process components to simultaneously enhance the bacterial growth, laccase production**
334 **and textile dye decolorization with cytotoxicity study.** *Plos One* 2018, **13**(5):e0195795.
- 335 18. Gouda MA: **Common Pitfalls in Reporting the Use of SPSS Software.** *Medical Principles &*
336 *Practice* 2015, **24**(3):300-300.
- 337
338
339
340

341 **Figures and Tables**

342

343 **Figure 1. The relative abundance of the top 15 genera detected in samples across all**
344 **laboratories.** Each column represents a single sample, and sections **a** and **b** describe the same
345 samples at different classification levels. **a**, At phylum level. **b**, At genus level. ‘Others’ indicates all
346 other phyla or genera except for the top 15 genera.

347

348 **Figure 2. Alpha diversity comparisons for samples from all sampling sites/ biological**
349 **laboratories.** **a**, Chao1. **b**, Shannon index. **c**, Observed_OTUs. **d**, PD_whole_tree. Where Chao1
350 and Observed_OTUs estimate the number of OTUs in the community, and a higher Shannon index
351 indicates greater abundance with a more even representation and PD_whole_tree adds the
352 evolutionary relationship between species to compare its diversity. All samples have been compared
353 with each other, categorized by different laboratories and sampling sites. The line indicates the
354 difference between two sites, and * $p < 0.1$, ** $p < 0.01$. ‘Other’ indicates the major public areas.

355

356 **Figure 3. The composition of microbial samples at same types of sampling sites among all**
357 **biological laboratories.** The relative abundance of different species from the lab outlet (**a**), major
358 public areas (**b**), and platform (**c**). Overlaps between the laboratories are indicated by Venn diagram
359 showing the detected bacterial genera from lab outlet (**d**), major public areas (**e**), and platform (**f**).

360

361 **Figure 4. Differences of microbial samples at different sites within the same type of biological**
362 **laboratory.** The relative abundance of different species from animal laboratory (**a**), microbiology
363 laboratory (**b**) and plant laboratory (**c**). Venn diagram showing the overlap between identified
364 microbial genera observed in animal laboratory (**d**), microbiology laboratory (**e**), plant laboratory (**f**).
365 Where colors in **a**, **b**, **c** indicate various microbial genera, while ‘other’ in **d**, **e**, **f** indicates the major
366 public areas.

367

368 **Figure 5. Biomarker for samples among three types of biological laboratories.** **a**, Microbial
369 richness that has significant differences between three types of laboratories ($LDA > 2$). **b**, The
370 phylogenetic relationships of these microbes.

371

372 **Figure 6. Schematic workflow including sampling site selection, DNA sequencing and**
373 **computational approaches.** Illustration of the main steps involved in sampling from lab outlet,
374 platform and major public areas of animal, microbe and plant laboratories, extracting DNA,
375 Illumina sequencing, bioinformatics analysis and interpretation. Finally, we compared the detected
376 genera with the publicly available common contaminants in the reagent, ICU microbe table and the
377 basic microbes of the environment to annotate the bacteria and trace the possible pollution source.

378

379

380 **Table 1. The relative abundance of shared genera across all laboratories and their possible**
 381 **sources.** Although these bacteria exist in all laboratories, they exist in different proportions in each
 382 laboratory. Where 'a', 'b', 'c' represents the microbes that may be prevalently contaminated in
 383 laboratory reagents (a), introduced by human daily activities (b), and basic environmental microbe
 384 (c), respectively.

Phylum	Genus	Relative abundance			Possible source
		Animal laboratory	Microbiological laboratory	Plant laboratory	
Actinobacteria	Corynebacterium	0.01043194	0.020823612	0.005222726	a, b, c
	Brevibacterium	0.004147394	0.00654011	0.001078781	a
	Micrococcus	0.043620243	0.059833667	0.014597697	a, b
	Actinomyces	0.00419055	0.011584026	0.001514026	
	Kocuria	0.002390392	0.004253708	0.001226562	a
	Propionibacterium	0.003290537	0.010990419	0.008227309	a, b
	Rhodococcus	0.013184083	0.019453764	0.004975993	a
	Microbacterium	0.001041871	0.003651781	0.001070219	a
Bacteroidetes	Sphingobacterium	0.003109507	0.001733227	0.001112906	b, c
	Flavobacterium	0.003071589	0.001690131	0.001168052	a, b
	[Prevotella]	0.002920677	0.001850549	0.001664687	
	Prevotella	0.027960465	0.016342249	0.015979757	
	Capnocytophaga	0.004329217	0.001414413	0.004057175	
	Chryseobacterium	0.034069126	0.004402164	0.012406579	a, b
	Sediminibacterium	0.006750533	0.002524042	0.003997523	
	Porphyromonas	0.001674581	0.001180926	0.001548046	
Firmicutes	Streptococcus	0.020726981	0.017232988	0.009853175	a, b, c
	Bulleidia	0.00260196	0.001167639	0.001166914	
	Lactobacillus	0.007177304	0.003449047	0.002745995	b, c
	Staphylococcus	0.011641965	0.010651629	0.006485062	b, c
	Veillonella	0.0060055	0.003390379	0.001033465	
Fusobacteria	Leptotrichia	0.006702987	0.003816631	0.004186437	
	Fusobacterium	0.010868195	0.004798143	0.007385495	
Proteobacteria	Neisseria	0.00857373	0.004033369	0.00663029	
	Stenotrophomonas	0.013283321	0.015911227	0.005654317	a, b
	Pseudoxanthomonas	0.00734037	0.004651302	0.005937601	a
	Sphingomonas	0.004221489	0.00850698	0.003254122	a, b
	Acinetobacter	0.085566067	0.10519708	0.074663258	a, b
	Agrobacterium	0.001620153	0.001257811	0.001000628	
	Sphingobium	0.006496551	0.003954288	0.004470411	a, b
	Paracoccus	0.002275995	0.003590857	0.00553695	a, c

	Haemophilus	0.004815064	0.002972872	0.001998964	
	Methylobacterium	0.016558476	0.014545819	0.008581459	a, b, c
	Pseudomonas	0.272891893	0.191455024	0.538318946	a, b, c
	Brevundimonas	0.006619573	0.009237963	0.002830279	a, b, c
	Enhydrobacter	0.070222079	0.032529829	0.074842363	a, b, c
	Novosphingobium	0.00188908	0.001385483	0.001607949	a
	Lysobacter	0.001891199	0.002709745	0.001717886	
Thermi	Deinococcus	0.008771836	0.008731487	0.005717905	

385

386

387 **Table 2. The specific genera of the three types of biological laboratories.** We labeled the basic
388 features of these bacteria, identified these potential effects through literature mining, and marked
389 the samples in which the species were most abundant for subsequent studies.

Laboratory (number)	Phylum	Genus	Possible Source	Sample ID	Annotations
Animal laboratory (8)	Actinobacteria	Atopobium			Bacterial vaginosis
					Anaerobic bacteria; Gram-positive bacteria; rod-shaped or oval;
		Gordonia			Degradation hydrocarbons
	Firmicutes	Jeotgalicoccus		A1B1 A2F1 A3B1 A3T1	Pathogens spread in the air of poultry farms
					The cells are globular facultative anaerobic; Gram-positive bacteria; Halophilic salttolerant bacteria;
		Dialister			Pneumonia, bacteremia
					Strict anaerobic
	Proteobacteria	Methylotenera			
		Methyloversatilis	a		Common in activated sludge
		Psychrobacter	b	A1B1	Can cause endocarditis and peritonitis; At the same time, it is also a probiotic for some fishes in the ocean.
				Gram-negative bacteria; Strong permeability; Oxidase positive; Psychrophilic or cold-tolerant aerobic bacteria;	
	Spirochaetes	Treponema			Syphilis
					Spiral bacteria
Plant laboratory (5)	Firmicutes	Bacillus	a, b, c	P1F1 P1PT1 P2F1	Anthrax, an important pathogen, can lead to food poisoning; At the same time, it can also promote plant rhizosphere growth.
					Gram-positive bacteria, rod-shaped;
	Proteobacteria	Moraxella			Easily cause human infectious bovine keratoconjunctivitis, tracheitis, pneumonia, otitis media, sinusitis, eyelid conjunctivitis
					Gram-negative bacteria
		Cellvibrio			Degrading polysaccharide
					Gram-negative bacteria; a slender bent rod shape;
	Buchnera		P1F1	Aphid symbiotic bacteria	
		Lautropia			

Microbial laboratory (16)	Actinobacteria	Rothia		Easy to cause lower respiratory tract infection Environmental plants, located in the air of the environment
		Leucobacter		Degradable herbicide Often appear in cow dung
		Modestobacter		
	Bacteroidetes	Spirosoma		Make contributins to the degradation of pollutants and the circulation of elements.
		Flavisolibacter	b M1W12	Improving plant rhizosphere carbon source
		Hymenobacter		Gram-negative bacteria; Acinetobacter;
	Chlamydiae	Waddlia		prone to cause sepsis and septicemia. a chlamydia-like replication cycle.
	Cyanobacteria	Chroococcidiopsis		Reduce nitrogen in the atmospheric. One of the most primitive cyanobacteria, photosynthetic bacteria, coccidiosis and extreme bacteria.
	Firmicutes	Anoxybacillus		Mostly in hot springs, manure and milk processing plants.
	Proteobacteria	Tepidimonas		Contribute to the degradation of industrial wastewater. rod-shaped, are mild thermophilic bacteria, Gram-negative bacteria, strict aerobic bacteria, oxidase and catalase positive.
		Erwinia		Plant pathogenicity, often infect woody plants, apples, pears and other Rosaceae crops, and is easy to cause fire disease and cucumber bacterial wilt.
		Hylemonella		Gram-negative coryneform bacteria
		Kaistobacter	M1W12	increasing of soil fertility
		Skermanella		Arsenic and antimony resistant bacteria in soil
		Rhodobaca		
	Thermi	Truepera		Degrading petrochemical components

Spherical cells, its optimum growth temperature is about 50 °C and have strong resistance to ionizing radiation.

390

391

variable

p_Acidobacteria_iii1.15
 p_Actinobacteria.f_Nocardioidaceae
 p_Acidobacteria.f_RB40
 p_Acidobacteria.f_Hoophagaceae.
 p_Actinobacteria.f_Propionibacteriacee
 p_Crenarchaeota.g_Candidatus.Nitrososphaera
 p_Actinobacteria.g_Streptomyces
 p_Bacteroidetes.g_Dysgonomonas
 p_Actinobacteria.g_Kocuria

p_Actinobacteria.g_Brachybacterium
 p_Actinobacteria.g_Rothia
 p_Actinobacteria.g_Friedmanniella
 p_Actinobacteria.g_Varibaculum
 p_Bacteroidetes.g_Dyadobacter
 p_Actinobacteria.g_Actinomadura
 p_Proteobacteria.g_Arenimonas
 p_Actinobacteria.g_Actinobacillus
 p_Proteobacteria.g_Sphingobium

p_Actinobacteriag_Propionimonas
 p_Actinobacteria.g_Modestobacter
 p_Actinobacteria.g_Williamsia
 p_Armatimonadetes.g_Fimbrillimonas
 p_Acidobacteria.g_Candidatus.Solibacter
 p_Proteobacteria.g_Sphingopyxis
 p_Actinobacteria.g_Collinsella
 Others

a

b