

RESEARCH PAPER



## Running after ghosts: are dead bacteria the dark matter of the human gut microbiota?

Sara Bellali<sup>a</sup>, Jean-Christophe Lagier<sup>a,b</sup>, Matthieu Million<sup>a,b</sup>, Hussein Anani<sup>a</sup>, Gabriel Haddad<sup>a,b</sup>, Rania Francis<sup>a,b</sup>, Edmond Kuete Yimagou<sup>a</sup>, Saber Khelaifa<sup>a</sup>, Anthony Levasseur<sup>a,b</sup>, Didier Raoult<sup>a</sup>, and Jacques Bou Khalil<sup>a</sup> 

<sup>a</sup>IHU Méditerranée Infection, Marseille, France; <sup>b</sup>Aix-Marseille Université, IRD, MEPHI, Marseille, France, Marseille, France

### ABSTRACT

The human gut microbiota has been explored by a wide range of culture-dependent and culture-independent methods, revealing that many microbes remain uncharacterized and uncultured. In this work, we aimed to confirm the hypothesis that some of the species present in the human gut microbiota remain uncultured not because of culture limitations, but because all members of such species are dead before reaching the end of the gastro-intestinal tract.

We evaluate this phenomenon by studying the microbial viability and culturability of the human gut microbiota from the fresh fecal materials of eight healthy adults. For the first time, we applied fluorescence-activated cell sorting (FACS) combined with 16S metagenomics analysis and microbial culturomics.

We identified a total of 1,020 bacterial OTUs and 495 bacterial isolates through metagenomics and culturomics, respectively. Among the FACS metagenomics results, only 735 bacterial OTUs were alive, comprising on average 42% of known species and 87% of relative abundance per individual. The remaining uncultured bacteria were rare, dead, or injured.

Our strategy allowed us to shed light on the dark matter of the human gut microbiota and revealed that both metagenomics and culturomics approaches are needed for greater insight into the diversity and richness of bacteria in the human gut microbiota. Further work on culture is needed to enhance the repertoire of cultured gut bacteria by targeting low abundance bacteria and optimizing anaerobic sample conditioning and processing to preserve the viability of bacteria.

### ARTICLE HISTORY

Received 2 September 2020

Revised 8 February 2021

Accepted 18 February 2021

### KEYWORD


Human gut microbiota; cultured; not-yet-cultured; culturomics; metagenomics; fluorescence-activated cell sorting; viability; culturability

## Introduction

Almost one hundred and twenty years ago, it was postulated that there was a large discrepancy between initial cell counts and culturable bacteria on nutrient media from the same samples.<sup>1</sup> This mismatch was subsequently been confirmed in several recent studies.<sup>2–10</sup> The phenomenon was named “the great plate count anomaly” by Staley and Konopka in 1985.<sup>8</sup> The fact that an overwhelming majority of micro-organisms, especially intestinal bacteria, do not grow on nutrient agar plates under laboratory conditions, led to the concept of “unculturable bacteria”, which is highly speculative, and “uncultured bacteria”. Furthermore, with the advent of culture-independent approaches and the dawn of metagenomics, it became clear that this “great plate count anomaly” resulted from a large number of unknown micro-organisms.<sup>11,12</sup>

Metagenomics has been widely used to describe the diversity of culturable and not-yet-cultured bacteria in the human gut microbiota.<sup>13,14</sup> Indeed, only a fraction of intestinal bacteria can be accessed through standard cultivation techniques, although an estimated 80% are not-yet-cultured.<sup>15</sup> Several hypotheses have been suggested to explain the microbial unculturability of microbes. For example, certain bacteria have low prevalence and/or low abundance, while others may require long incubation time to form visible colonies. Moreover, a great majority of intestinal bacteria are sensitive to oxygen, and some bacteria require specific nutrients and physical conditions for growth. In an attempt to solve this anomaly and increase our knowledge of the bacterial community composition in the human gut, recent advances in microbial culture techniques were applied to cultivate previously uncultured bacteria.<sup>11,16–22</sup> One of the techniques

**CONTACT** Didier Raoult  [didier.raoult@gmail.com](mailto:didier.raoult@gmail.com); Jacques Bou Khalil  [boukhaliljacques@gmail.com](mailto:boukhaliljacques@gmail.com)  IHU - Méditerranée Infection, 19-21 Boulevard Jean Moulin, Marseille 13005, France

 Supplemental data for this article can be accessed on the [publisher's website](#).

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

that we have created to achieve this is “microbial culturomics”, a high-throughput culture method that has been developed in our lab. It consists of several culture conditions and nutrient media, followed by the identification of bacterial species using MALDI-TOF mass spectrometry and 16S rRNA gene sequencing.<sup>23</sup> To date, culturomics has enabled the isolation of more than 300 new bacterial species in the human gut that were previously believed to be uncultured.<sup>24</sup>

Despite the advances in culture techniques, a huge number of gut bacterial species remain uncultured. However, culturomics provides viable pure cultures, unlike molecular approaches that only give information on the species level, without any information regarding their viability status. In addition, microbial culture using selective media has made it possible to detect and cultivate minority bacterial populations that may pass undetected by genomic technologies.<sup>23</sup>

In attempt to resolve the “uncultured” enigma of the intestinal bacterial species, and to describe the known richness and relative abundance of the live known and unknown bacterial population in human gut microbiota, we developed an optimized strategy using fluorescence-activated cell sorting (FACS) technique followed by 16S metagenomics analysis. In addition, the metagenomics of FACS sorted live, injured, and dead bacterial populations, was compared for the first time to culturomics data, for the same samples. In this work, we aimed to test the hypothesis that some gut species remain uncultured not because of culture limitations, but because all members of such species are dead before reaching the end of the gastro-intestinal tract. In addition, among live species in feces, we compared cultured and not-yet-cultured species, and more specifically, we compared their abundance.

## Results

### Metagenomics analysis of sorted live, dead, and injured bacterial populations

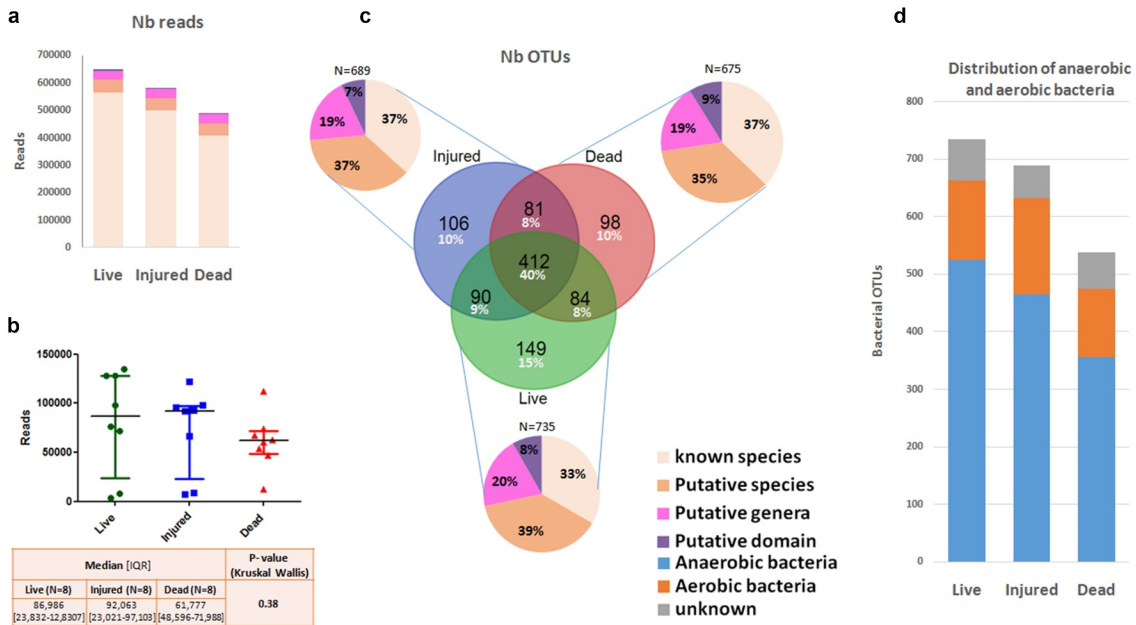
#### Global analysis

Analysis of metagenomics sequencing from eight sorted fecal samples generated 2,287,729 reads in the live, injured, and dead bacterial populations that were organized into 3,590 unique OTUs

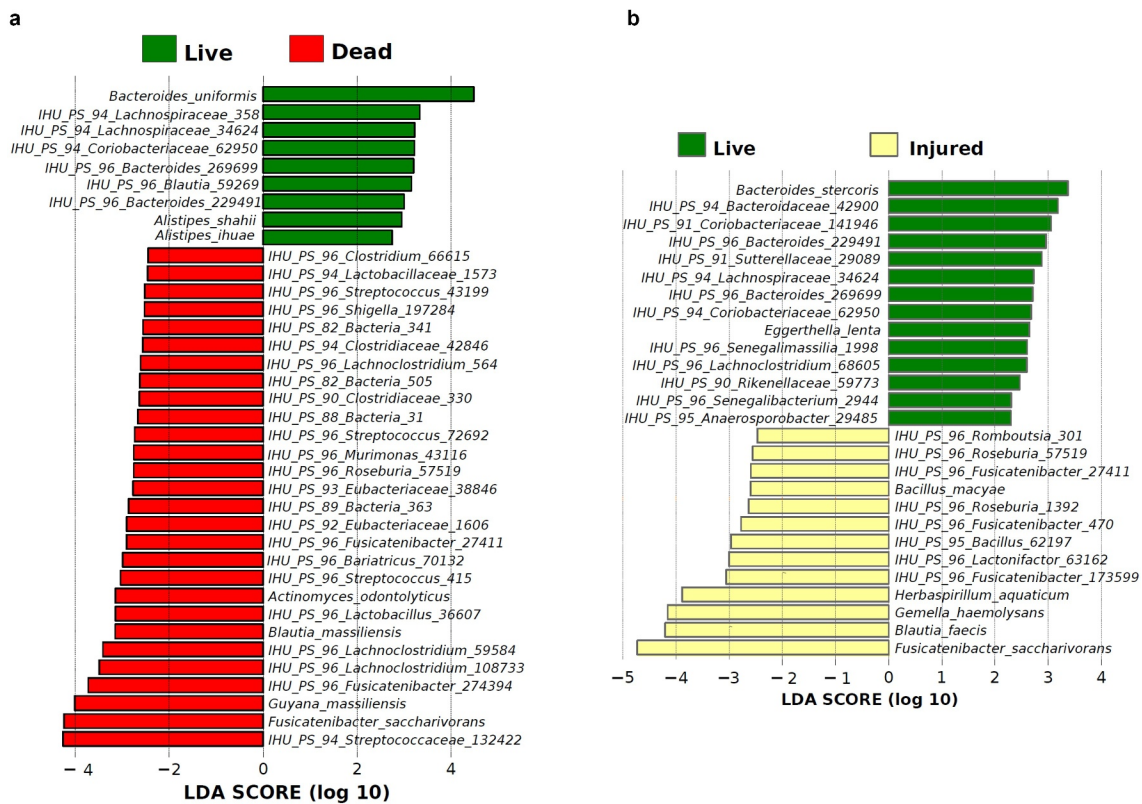
(Operational Taxonomic Units) (Supplementary Table 1). Of these 3,590 bacterial OTUs, 2,569 (567,583 reads) were not assigned to any taxonomic level (unclassified). We subsequently excluded all unassigned OTUs and focused on bacterial OTUs assigned at least to the bacterial domain, which represented 1,020 unique bacterial OTUs (1,720,146 reads) (Supplementary Table 2). Ultimately, we obtained 648,460 reads (37.70%) generated from the viable bacterial population, 490,400 reads (28.51%) from the injured bacterial population, and 581,286 reads (33.79%) from the dead bacterial population (Figure 1a). There was no statistically significant difference in total reads between live, injured, and dead bacterial groups (Figure 1b). Regarding the total numbers of OTUs in the three bacterial populations, 15% (n = 149 OTUs) were exclusively live, 28% (n = 285 OTUs) were exclusively injured and/or dead, and 57% (n = 586 OTUs) were shared between live, injured, and dead bacterial populations (Figure 1c). In addition, anaerobic bacteria outnumbered aerobic bacteria by a factor of almost three in the live (n = 525 OTUs; 71.43%), dead (n = 356 OTUs; 66.17%) and injured (n = 465 OTUs 67.49%) bacterial population (Figure 1d).

Distinct differences at the phylum level were observed between live, injured, and dead bacterial groups. A total of 11 assigned phyla were identified, with four phyla dominating all samples in the live, injured, and dead bacterial groups. These were respectively, *Firmicutes* (35.11%, 46.12% and 46.01%), *Bacteroidetes* (33.65%, 13.33% and 16.96%), *Proteobacteria* (22.38%, 35.75% and 31.26%) and *Actinobacteria* (7.88%, 3.72 and 4.68%) (Supplementary Figure 1). No significant differences were observed between live, injured, and dead bacterial groups at the phylum level (Supplementary Figure 2).

We used linear discriminant analysis (LDA) of effect size (LEfSe) to compare the enrichment analysis of OTUs at all taxonomic levels between the live, injured, and dead bacterial populations (Figure 2). We found that the majority of dead bacterial OTUs were not previously cultured. Of these unknown bacteria, *IHU\_PS\_94\_Streptococcaceae\_132422* presented the highest LDA score (LDA score > 4.0) (Figure 2a). However, among known dead bacteria, *Guyana marseillensis* and *Fusicatenibacter saccharivorans* (LDA



**Figure 1.** (a) Total read numbers in live, injured, and dead bacterial populations in eight fecal samples. (b) Graph representing Kruskal-Wallis test results on total number reads numbers in live, injured, and dead bacterial populations in eight fecal samples. (c) Pi chart representing the proportion of bacterial OTUs (known species, putative species, putative genera, and putative domain) in live, injured, and dead bacterial populations. (d) Distribution of anaerobic and aerobic bacteria in live, injured, and dead bacterial populations.



**Figure 2.** Histogram of the LDA scores computed for differentially abundant bacterial OTUs between live and dead bacterial groups (a) and live and injured bacterial groups (b).

score  $\geq 4.0$ ) were overrepresented. The bacterium *F. saccharivorans*, was also enriched in the injured bacterial group. In addition, *Gemella haemolysans* and *Blautia faecis* had the highest LDA score (LDA score  $> 4.0$ ) among the injured group (Figure 2b).

### Live microbial richness and relative abundance of known species in fecal samples

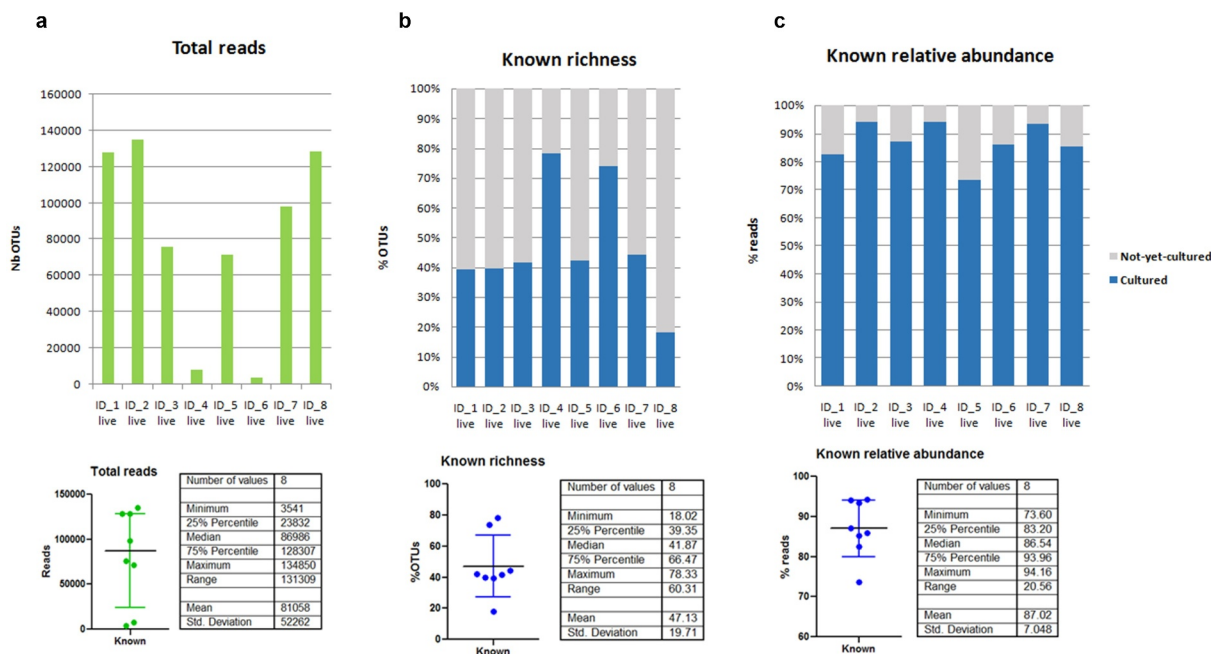
To quantify the richness and relative abundance corresponding to known species in each sample from the eight healthy donors, we excluded all dead and injured bacterial OTUs, and only considered OTUs from live bacteria (Figure 3a). We found 42% [83–94%] richness (Figure 3b) and 87% [39–66%] relative abundance corresponding to cultured species (Figure 3c) per individual (median [interquartile range]). These results suggest that majority species are more frequently cultured than minority species.

Thus, we assess a putative relationship between culturability (probability of being cultured) and abundance (sum of reads for a species in the eight sampled individuals). The abundance of cultured species was significantly higher than the abundance of not-yet-cultured species (Figure 4a). Moreover, we found that species abundance

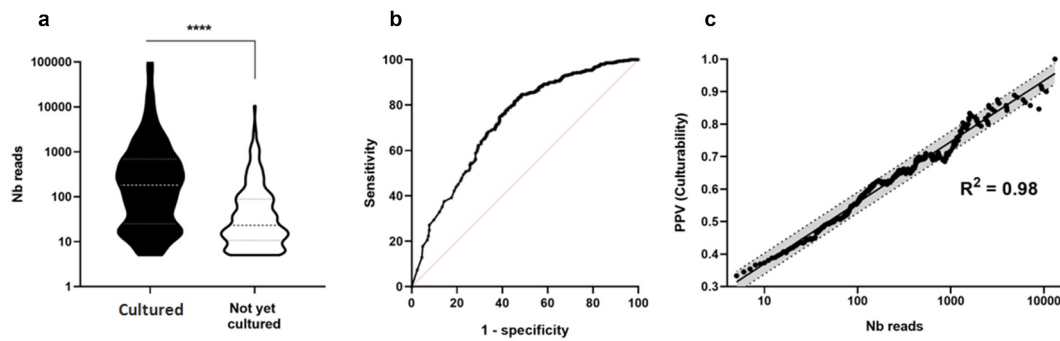
followed a log-normal distribution for both cultured and not-yet-cultured species. Interestingly, the Receiver Operating Characteristic (ROC) analysis confirmed a highly significant relationship ( $P = 2.8 \times 10^{-23}$ ) between the abundance of a species (number of reads) and its culturability (Figure 4b). The area under curves (AUCs) was 0.72 (95% confidence interval: 0.68–0.76). The ROC analysis also illustrated that species with total reads above 13,152 in the eight sampled individuals were all cultured (positive predictive value = 1 above 13,152 reads). Between 5 and 10,000 reads, the relationship between abundance and culturability was log-linear from 0.3 to 1 (linear regression after log10 transformation,  $R^2 = 0.98$ , Figure 4c).

### Viability and culturability: metagenomics of sorted live, injured, and dead bacteria versus microbial culturomics

The microbial culturomics approach yielded 495 bacterial isolates (334 anaerobic, 161 aerobic) (Supplementary Table 3) from fresh fecal samples of eight healthy donors, using 58 culture media under anaerobic and aerobic conditions. In



**Figure 3.** (a) **Total live read numbers** (known and unknown bacterial species) in eight fecal samples. (b) **Bacterial richness** of known and unknown bacterial species in the live bacterial population. (c) **Relative abundance** of known and unknown bacterial species in the live bacterial population.



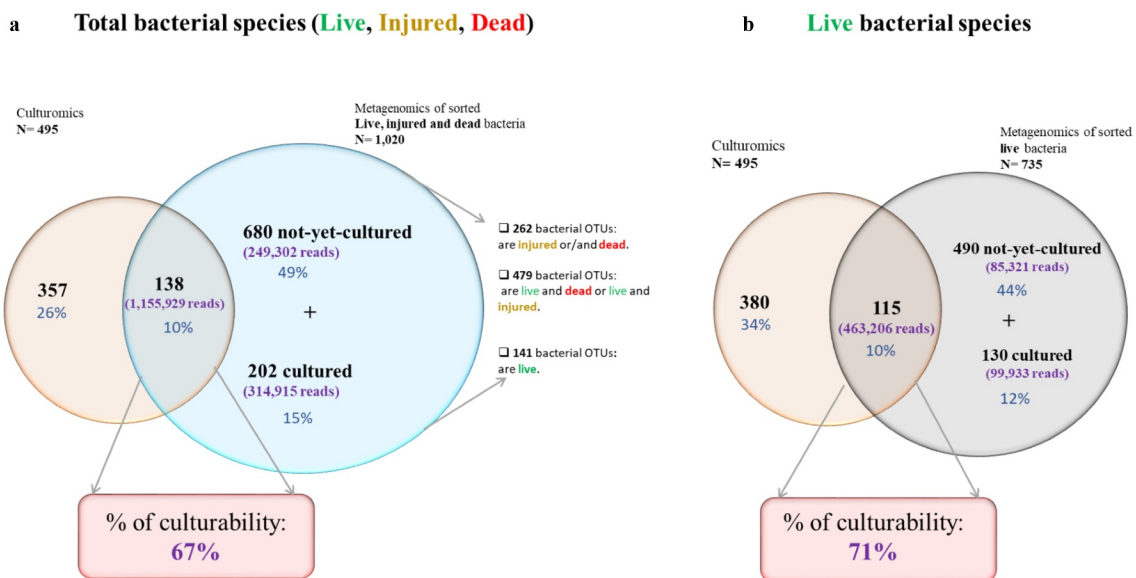
**Figure 4.** \*\*\*\*:  $p < .00005$ . PPV: positive predictive value. (a) The number of reads was significantly higher for cultured than for not-yet-cultured species (bilateral Mann-Whitney test). Medians and interquartile ranges are shown. (b) **ROC curve** between culturability (cultured versus not-yet-cultured species) and abundance (total number of reads). (c) **Linear regression** of the culturability (positive predictive value to be “cultured”) versus the log<sub>10</sub> of the total number of reads.

contrast, the metagenomics approach coupled with flow cytometry identified a total of 1,020 OTUs in the live, injured, and dead bacterial populations. Of them, 33.33% (340 OTUs; 1,470,844 reads) were assigned to known species, 37.35% (381 OTUs; 138,925 reads) were putative species, 20.29% (207 OTUs; 96,922 reads) were putative genus and 9.02% (92 OTUs; 13,455 reads) were putative domain.

We compared the bacterial populations cultivated by culturomics to metagenomics of sorted live, injured, and dead bacterial populations using a Venn diagram (Figure 5a). 357 cultured bacteria were exclusively identified by culturomics and 882

bacterial OTUs were exclusively identified by metagenomics. Of these missed bacteria (not isolated in the eight fecal samples using culturomics), 86 bacterial OTUs (34 anaerobic, 40 aerobic and 12 unknown bacterial OTUs) were exclusively dead, 141 bacterial OTUs (97 anaerobic, 22 aerobic and 22 unknown bacterial OTUs) were exclusively live, 99 bacterial OTUs (59 anaerobic, 31 aerobic and nine unknown bacterial OTUs) were exclusively injured and 556 bacterial OTUs were shared between live, dead and injured bacterial populations.

By comparing the metagenomics of live sorted bacteria and culturomics (Figure 5b), we found 115 known species (463,206 reads; 107 anaerobic, 8



**Figure 5.** Culturability and viability, percentages of culturable bacteria calculated from reads generated by metagenomics of sorted bacteria in eight fecal samples: (a) Percentages of **culturable** bacteria calculated from total live, injured, and dead bacterial populations. (b) Percentages of **culturable** bacteria calculated from total live bacterial populations.

aerobic) shared by both approaches. 380 known species (224 anaerobic, 158 aerobic) were exclusive to culturomics, and 620 bacterial OTUs (418 anaerobic, 130 aerobic and 72 unknown) were exclusively identified by metagenomics. Of these 620 bacterial OTUs (185,254 reads), 130 OTUs (99,933 reads) were assigned to known species (Figure 5). Of these 130 known species, 46 were isolated in previous culturomics studies, and 85 bacterial OTUs were not previously isolated using a culturomics approach, including 15 bacterial OTUs that were found in humans but not in the gut, 29 bacterial OTUs that were previously found to be associated with the human gut, and 41 bacterial OTUs that were not previously found in humans.

### Viability and culturability

Regarding the percentage of culturability, we determined the cultivable proportion of the human gut microbiota from total metagenomics reads of sorted live, injured, and dead bacteria. By including total bacterial OTUs (1,720,146 reads; 1,020 OTUs), the cultivable portion was 67% (1,155,929 reads) (Figure 5a). Moreover, when we only included live bacterial OTUs (648,460 reads; 735 OTUs), the percentage of culturability increased to 71% (463,206 reads) (Figure 5b).

Overall, comparing the culturomics and metagenomics of sorted live bacteria at the species level, culturomics ( $n = 495$  bacterial isolates) doubled the number of isolated species compared to metagenomics ( $n = 245$  known species). In addition, shared species were most abundant (463,206 reads; 115 known species) compared to the bacterial OTUs exclusively found by metagenomics (395,092 reads; 2322 OTUs).

### Discussion

The human gut microbiota harbors a wide range of micro-organisms which play different roles in human health and disease.<sup>25</sup> Advances in next generation sequencing, in particularly advances in metagenomics, have provided great insights into the diversity of microbial populations in the human gut. Metagenomics claimed to be able to detect “uncultured” populations, nevertheless this

technique is not able to determine whether the prokaryotes are alive or dead.<sup>24</sup> Here, for the first time, we performed a fluorescence activated cell sorting (FACS) combined with 16S metagenomics analysis and microbial culturomics to explore the diversity and richness of the known and unknown bacterial species in the human gut microbiota of eight healthy individuals. In doing so, we successfully discriminated between live, injured, and dead bacterial groups using stringent methods that include negative and positive controls, unlike previous metagenomics analysis of the human gut.<sup>13,14,23,26,27</sup> Our results showed that only about one third of reads were generated from the live bacterial population and the remaining bacteria were either injured or dead. Interestingly, within this live bacterial population, we found that known bacteria were the most abundant, so it will therefore be important to target minority populations using more specific culture media based on whole genome data along with a larger number of analyzed colonies.

This study revealed that the abundance of a species was a strong predictor of its culturability, with a biphasic relationship. This suggests that minority species constitute a large part of the dark matter of the live microbial population in human fecal samples. Future studies should include specific “kill the winner” strategies to isolate these minority species and/or use “high throughput” culturomics. These results also suggest that culturability is, at least in part, a stochastic or random process: the more colonies are identified, the more species would be found. In this way, single cell culturomics could be the next revolution using the same approaches developed for single cell sequencing: fluorescent-activated cell-sorting of individual cells collected by micromanipulation (serial dilution or nanotube), laser capture micro-dissection (LCM) and microfluidics.<sup>28,29</sup>

Moreover, by including total bacterial OTUs (live, injured and dead) at the genus level, we found that about 56% of bacterial OTUs were not-yet-cultured which is twice as high as previously estimated, 14 years ago.<sup>15</sup> The same results were found by Sunagawa *et al.* using a single-cell genomics approach, where about 58% of gut species-level OTUs were not-yet-cultured.<sup>27</sup> More recently, Nayefsh *et al.* found that 58% of gut species-level

OTUs with sequenced genomes were uncultured species.<sup>13</sup>

One of the most important results of our study was that 28% of bacterial OTUs in the total fecal samples were exclusively found to be dead and/or injured. Of these non-live bacteria, about two-thirds were not-yet-cultured, and a large amount were anaerobic. This may explain why the majority of bacterial species were missed in culture, such as the genus *Romboutsia* (three bacterial OTUs). In agreement with our study, two recent studies compared the genomic profile of cultured and not-yet-cultured gut bacteria which revealed that uncultured species were missing genes encoding antioxidant and redox functions.<sup>13,14</sup> These findings suggest that processing fecal samples under aerobic conditions may influence the mortality of these uncultured bacteria, and therefore their culturability. Recent studies have shown the impact of aerobic manipulations on the viability and diversity of fecal microbiota.<sup>30–32</sup> Greater effort is needed to optimize anaerobic sample conditioning and processing, and consequently, anaerobic culturing.

We performed a comparative analysis between isolated cultures and FACS-sorted metagenomics for the same samples in order to explain the divergence between both approaches. The study led us to interesting conclusions. The first is that shared bacterial species between culturomics, and metagenomics represented only ten percent of the total bacteria and they are more abundant compared to the bacterial population identified by metagenomics alone. Secondly, the majority of bacterial species missing from culturomics were not-yet-cultured bacteria, and moreover, they were less abundant. Third, the percentage of culturability increased from 67% to 71% when we excluded dead and injured metagenomics data, which means that only a minority of bacterial species remain uncultured. And finally, at the species level, culturomics doubled the number of isolated species compared to metagenomics.

One limitation of our approach is the need to dilute the fecal sample to avoid background noise and achieve a successful bacterial sort. This may result in the loss of some rare bacterial populations. However, this could be enhanced by optimizing the settings of the flow cytometry cell sorter for efficient

quantitative sorting, capable of analyzing more concentrated samples.

## Conclusions

We showed that fluorescence activated cell-sorting combined with 16S metagenomics analysis provides a way to distinguish between live, injured, and dead bacterial groups. Moreover, combining these culture-independent techniques to microbial culturomics, a valuable complementary approach, has enabled us to better understand the unculturability of some bacterial species. We aim to develop new culture techniques targeting rare bacteria that are still not-yet-cultured, as well as try filling the metagenomics gaps. Our process has made the link between culturomics and metagenomics. This bridge helped us answer many questions and resolve enigmas around the complementarity or discordance between the most widely used techniques to study the human gut microbiota.

## Materials & methods

### Stool sample collection

Fresh stool samples were obtained from eight healthy adults of different geographic origins: France (n = 2), Senegal (n = 2), Algeria (n = 2), Cameroon (n = 1) and Benin (n = 1) (Supplementary Table 4). All donors had lived in France for more than six months prior to sampling. They gave their written consent, and the project was approved by the ethics Committee of IHU Méditerranée Infection under number 2016–11. Donors had not received any antibiotics within the three months prior to sampling. Stool samples were collected in a sterile container and a GasPak generator (Becton Dickinson, Sparks, MD, USA) was immediately introduced. All containers were closed and kept in a plastic zipper bag until use.

### Microbial culturomics

The same eight fecal samples were cultured using 58 culture conditions that were detailed in a previous study carried out in our lab (Supplementary Table 5).<sup>33</sup> We studied the bacterial culturability of these fecal samples by comparing the culturomics

data to the data of the sorted metagenomics of live, injured, and dead bacterial populations for the same stool samples, as described below. The study strategy design is summarized in the flowchart represented in Figure 6.

### Metagenomics of sorted bacteria

#### Stool sample processing and staining conditions (BaLight staining)

Approximately 50 g of fecal sample were homogenized in 250 mL of normal saline solution (Fresenius Kabi, Sevres, France) using a commercial immersion blender (BOSCH Ultracompact 400 W) for five minutes. Fecal suspension was filtered using sterile coffee strainers to remove large aggregates. The slurry was centrifuged at  $6,000 \times g$  for 15 minutes, the supernatant was removed, and the pellet was re-suspended in 125 mL of normal saline solution. The homogenized stool solution was diluted with PBS (Life Technologies, Paisley, United Kingdom) to 1/1000. The staining step was carried out using the Live/Dead BaLight kit (Invitrogen, Eugene, USA) as per the manufacturer's recommendation: 1 mL of diluted fecal sample at the  $10^{-3}$  dilution, was stained with 1  $\mu\text{L}$  of SYTO9 and 1  $\mu\text{L}$  of propidium

iodide (PI) in a microbial safety cabinet and then incubated for 15 minutes in the dark at room temperature, before being processed in a cell FACS sorter (BD Biosciences) and then sequenced.

#### Flow cytometry and cell sorting

Live and dead dye optimization and protocols for live and dead assay were previously described by Bellali *et al.*<sup>32</sup> We proceeded to analyze and sort artificial mixed fractions, in order to assess the fecal samples. We used a BD FACS Aria™ Fusion Special Order (SORP) cell sorter cytometer (BD Biosciences) equipped with a combination of six lasers (355 nm, 405 nm, 488 nm, 561 nm, and 640 nm) in a biosafety cabinet for the best safety performance and sorting results. The flow sheath was 0.22  $\mu\text{m}$  in-line filtered. The pressure was stabilized for at least one hour before experiments began. A photodiode was used as a standard FSC-detector on this instrument, and triggering was based on the side scatter. Sorting was performed using a 70  $\mu\text{m}$  nozzle, pressure of 70 PSI and a frequency of 90 kHz, after discriminating between all populations based on forward scatter, side scatter, SYTO9 and PI. To analyze the purity of the sorted fraction, collected samples were processed under the same conditions. The purity calculation was

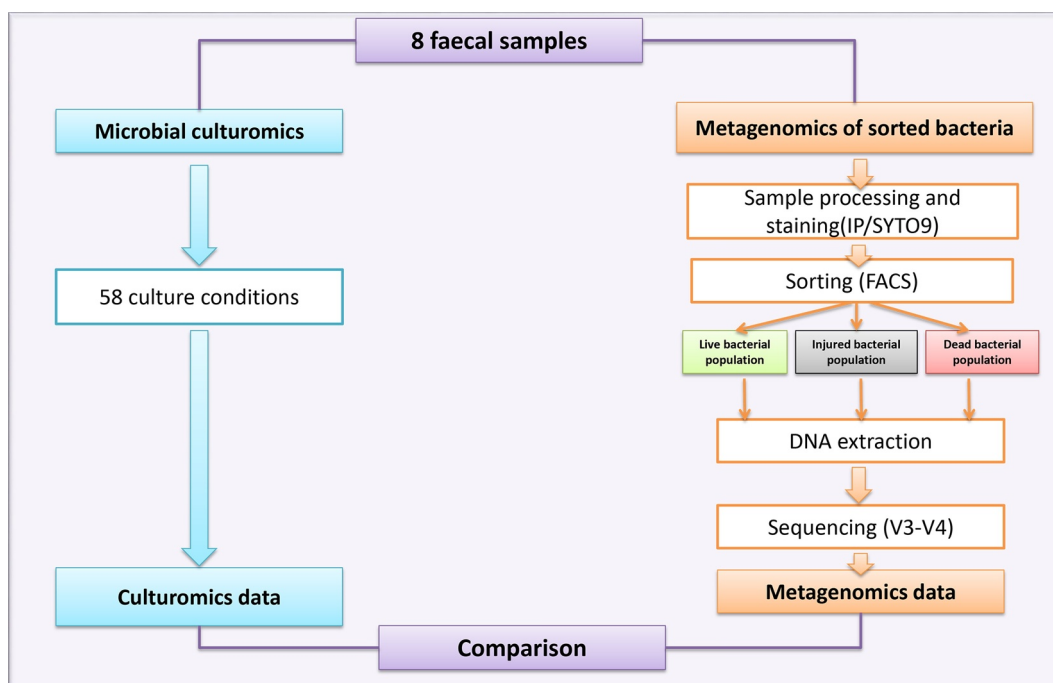


Figure 6. Workflow illustrating the overall strategy used in this study.



based on the ratio between the different population counts, by applying the following equation: [% Sorted fraction = (#Sorted fraction/# all fractions) x 100] after reanalysis of the sorted population. We managed to differentiate between three populations clearly separated on the dot plots, “live” bacterial cells (SYTO9- stained), “dead” bacterial cells (IP-stained), and “injured” or damaged bacterial cells (IP and SYTO9-stained) between the live and the dead bacterial populations. All three populations were sorted and then sequenced as described below. The efficiency and purity of our sorting was validated by culturing the sorted fractions in Columbia sheep blood agar plates (BioMérieux, Marcy l’Etoile, France) under aerobic and anaerobic conditions for 48 hours at 37°C, in order to demonstrate that dead and live bacteria had been correctly discriminated, and that there was no cross-contamination.

#### **DNA extraction and metagenomics sequencing**

DNA was extracted from all the dead, live and injured bacterial populations. Briefly, sorted samples were completely centrifuged and re-suspended in 750 µL of 0.22 micro-filtered PBS. DNA was then extracted according to the extraction methods 1 and 5 described by Angelakis *et al.*<sup>34</sup>

The extracted fractions (live, dead and injured) were then sequenced for 16S rRNA on MiSeq technology (Illumina, Inc, San Diego CA 92121, USA) as previously described by Angelakis *et al.*<sup>34</sup>

#### **Metagenomics analysis**

The paired-end reads were assembled into contigs using Pandaseq.<sup>35</sup> The high-quality sequences containing both primers (forward and reverse) were then selected for the next step. At the filtering stage, all sequences containing N and sequences which were shorter than 200nt were removed. Sequences longer than 500nt were trimmed. In addition, the forward and reverse primers were also removed from each of the sequences. To remove the chimeric sequences we applied an additional filtering step using UCHIME<sup>36</sup> by USEARCH.<sup>37</sup> The filtering steps were carried out using the QIIME pipeline.<sup>38</sup> The clustering of duplicate sequences (the dereplication process) was carried out on the filtered sequences, and they were subsequently sorted by decreasing abundance.<sup>39–41</sup> The clustering of OTUs was performed with 97%

identity for each metagenome. The Silva SSU and LSU database and release 132 from the Silva website were downloaded and integrated. From this, a local database of predicted amplicon sequences was built by extracting the sequences containing both primers. 16S sequences from 375 species isolated in our laboratory from diagnosis or culturomics were added to the database (Supplementary Table 6). A reference database of 14,459 sequences was thus generated. All the putative species of previous analysis were also added. Finally, the generated database contained 76,368 sequences ready to be used. Regarding taxonomic assignments, we applied at least five reads per OTU because our samples were diluted at 1/1,000, and after sorting we had 1,000,000 cells per tube for each population: live, dead, and injured bacteria. The OTUs were then searched against each database using BLASTN.<sup>42</sup> The sequences were assigned a taxonomic classification using the criteria of ≥ 77% identity for domain, ≥ 75% identity for phylum, ≥ 80% identity for class, ≥ 85% identify for order, ≥ 90% identity for family, ≥ 94% identity for genus and ≥ 97% sequence identity for species. The best match of ≥ 97% identity and 100% coverage for each of the OTUs was extracted from the reference database, and taxonomy was assigned up to the species level. Sequences identities below 77% were assigned to the “unclassified or unassigned” category. Finally, we counted the number of OTUs.

#### **Statistical analysis**

Differentially abundant OTUs between live, injured, and dead sorted bacteria were identified using the LDA Effect Size (LEfSe: Linear Discriminant Analysis Effect Size) algorithm available online at (<http://huttenhower.sph.harvard.edu/galaxy/root>)<sup>43</sup>. The threshold on the logarithmic discriminant analysis (LDA) score was set to 2.0 and the significance level was 0.05.

To test whether there was a relationship between “cultured species” and “not-yet-cultured species” and their abundance in the human gut microbiota, we used a ROC analysis and the area under curve was calculated. Any species previously reported cultured in the literature or in our culturomics approach was defined as “cultured”. OTUs in the “live” population that do not correspond to any cultured species, were defined as “not-yet-cultured species”. The quantitative

variable for microbial abundance was the total reads (sum of reads in the eight fecal samples for each species). A comparison of read numbers between live, injured, and dead bacterial group was carried out using the non-parametric Kruskal-Wallis test. Statistical analyses were performed using XLSTAT 2019.3.1 (Addinsoft, Paris, France) and Graph Pad (Prism v8.2.1) (GraphPad Software, San Diego, California USA).

### Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed

### Authors' contributions

JBK and DR conceived the project. JBK and SB developed the methodology. SB, RF, JBK, SK and EKY conducted the experiments. SB, JBK, MM, GH, HA and AL performed data analysis. SB wrote the draft manuscript. JBK, JCL, and DR revised the manuscript. All authors contributed to the revised version of the manuscript and read and approved the submitted version.

### Availability of data

All data generated or analysed during this study are included in this published article and its supplementary information files.

### Competing interests

The authors declare that they have no competing interests.

### Ethics approval and consent to participate

All donors recruited for the study signed an informed written consent, and the project received the IHU Méditerranée Infection ethics committee agreement under the number 2016-011.

### Funding

This study was supported by the “Fondation Méditerranée Infection” and by the French Government under the “Investissements d’avenir” (Investments for the Future) programme managed by the Agence Nationale de la Recherche (ANR, fr: National Agency for Research), (reference: Méditerranée Infection 10-IAHU-03). This work was also supported by the Région Provence-Alpes-Côte d’Azur and European funding FEDER PRIM1.

### ORCID

Jacques Bou Khalil  <http://orcid.org/0000-0002-4272-5208>

### References

1. Winterberg H. Zur methodik der bakterienzählung. *Med Microbiol Immunol.* 1898;29:75–93. doi:10.1007/BF02217377.
2. Eilers H, Pernthaler J, Glöckner FO, Amann R. Culturability and in situ abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol.* 2000;66:3044–3051. doi:10.1128/AEM.66.7.3044-3051.2000.
3. Xu H-S, Roberts N, Singleton F, Attwell R, Grimes D, Colwell R. Survival and viability of nonculturable *Escherichia coli* and *Vibrio cholerae* in the estuarine and marine environment. *Microb Ecol.* 1982;8:313–323. doi:10.1007/BF02010671.
4. Razumov A. The direct method of calculation of bacteria in water: comparison with the Koch method. *Mikrobiologija.* 1932;1:131–146.
5. Jannasch HW, Jones GE. Bacterial populations in sea water as determined by different methods of enumeration 1. *Limnol Oceanogr.* 1959;4:128–139. doi:10.4319/lo.1959.4.2.0128.
6. Waksman SA, Hotchkiss M. Viability of bacteria in sea water. *J Bacteriol.* 1937;33:389. doi:10.1128/JB.33.4.389-400.1937.
7. ZoBell CE. *Marine microbiology. A monograph on hydrobacteriology.* Chronica Botanica Co., Waltham, MA. 1946;240.
8. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol.* 1985;39:321–346. doi:10.1146/annurev.mi.39.100185.001541.
9. Butkevich N. Multiplication of sea bacteria depending on the composition of the medium and on temperature. *Microbiology (Moscow).* 1936;5:322–343.
10. Cholodny N. Zur Methodik der quantitativen Erforschung des bakteriellen Planktons. *Zentralbl Bakteriol Parasitenkd Infektionskr Hyg A.* 1929;77:179–193.
11. Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Aca. Sci.* 2011;108:6252–6257. doi:10.1073/pnas.1102938108.
12. Suau A, Bonnet R, Sutren M, Godon -J-J, Gibson GR, Collins MD, Doré J. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol.* 1999;65:4799–4807. doi:10.1128/AEM.65.11.4799-4807.1999.
13. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global

- human gut microbiome. *Nature*. 2019;568:505. doi:10.1038/s41586-019-1058-x.
14. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD. A new genomic blueprint of the human gut microbiota. *Nature*. 2019;568:499. doi:10.1038/s41586-019-0965-1.
  15. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. *science*. 2005;308:1635–1638. doi:10.1126/science.1110591.
  16. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature*. 2016;533:543. doi:10.1038/nature17645.
  17. Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Dubourg G. Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol*. 2016;1:16203. doi:10.1038/nmicrobiol.2016.203.
  18. Tanaka Y, Benno Y. Application of a single-colony coculture technique to the isolation of hitherto unculturable gut bacteria. *Microbiol Immunol*. 2015;59:63–70. doi:10.1111/1348-0421.12220.
  19. Lau JT, Whelan FJ, Herath I, Lee CH, Collins SM, Bercik P, Surette MG. Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med*. 2016;8:72. doi:10.1186/s13073-016-0327-7.
  20. Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, Dunn M, Mkandawire TT, Zhu A, Shao Y. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol*. 2019;37:186. doi:10.1038/s41587-018-0009-7.
  21. Zou Y, Xue W, Luo G, Deng Z, Qin P, Guo R, Sun H, Xia Y, Liang S, Dai Y. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol*. 2019;37:179. doi:10.1038/s41587-018-0008-8.
  22. Nichols D, Cahoon N, Trakhtenberg E, Pham L, Mehta A, Belanger A, Kanigan T, Lewis K, Epstein S. Use of ichip for high-throughput in situ cultivation of “uncultivable” microbial species. *Appl Environ Microbiol*. 2010;76:2445–2450. doi:10.1128/AEM.01754-09.
  23. Lagier J, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect*. 2012;18:1185–1193. doi:10.1111/1469-0691.12023.
  24. Lagier J-C, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, Levasseur A, Rolain J-M, Fournier P-E, Raoult D. Culturing the human microbiota and culturomics. *Nat Rev Microbiol*. 2018;16:540. doi:10.1038/s41579-018-0041-0.
  25. Lynch SV, Pedersen O. The human intestinal microbiome in health and disease. *N Engl J Med*. 2016;375:2369–2379. doi:10.1056/NEJMra1600266.
  26. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell*. 2019;176:649–662. doi:10.1016/j.cell.2019.01.001.
  27. Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger SA, Kultima JR, Coelho LP, Arumugam M, Tap J, Nielsen HB. Metagenomic species profiling using universal phylogenetic marker genes. *Nat Methods*. 2013;10:1196. doi:10.1038/nmeth.2693.
  28. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet*. 2013;14:618. doi:10.1038/nrg3542.
  29. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet*. 2016;17:175. doi:10.1038/nrg.2015.16.
  30. Martínez N, Hidalgo-Cantabrana C, Delgado S, Margolles A, Sánchez B. Filling the gap between collection, transport and storage of the human gut microbiota. *Sci Rep*. 2019;9:8327. doi:10.1038/s41598-019-44888-8.
  31. Papanicolaou LE, Choo JM, Wang Y, Leong LE, Costello SP, Gordon DL, Wesselingh SL, Rogers GB. Bacterial viability in faecal transplants: which bacteria survive? *EBioMedicine*. 2019;41:509–516. doi:10.1016/j.ebiom.2019.02.023.
  32. Bellali S, Lagier J-C, Raoult D, Khalil B, Yaacoub J. Among live and dead bacteria, the optimization of sample collection and processing remains essential in recovering gut microbiota components. *Front Microbiol*. 2019;10:1606. doi:10.3389/fmicb.2019.01606.
  33. Diakite A, Dubourg G, Dione N, Afouda P, Bellali S, Ngom II, Valles C, Million M, Levasseur A, Cadoret F, et al. Extensive culturomics of 8 healthy samples enhances metagenomics efficiency. *PLoS ONE*. 2019;14:e0223543. doi:10.1371/journal.pone.0223543.
  34. Angelakis E, Bachar D, Henrissat B, Armougom F, Audoly G, Lagier J-C, Robert C, Raoult D. Glycans affect DNA extraction and induce substantial differences in gut metagenomic studies. *Sci Rep*. 2016;6:26276. doi:10.1038/srep26276.
  35. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinform*. 2012;13:31. doi:10.1186/1471-2105-13-31.
  36. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27:2194–2200. doi:10.1093/bioinformatics/btr381.

37. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–2461. doi:10.1093/bioinformatics/btq461.
38. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335. doi:10.1038/nmeth.f.303.
39. Stoeck T, Behnke A, Christen R, Amaral-Zettler L, Rodriguez-Mora MJ, Chistoserdov A, Orsi W, Edgcomb VP. Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol*. 2009;7:72. doi:10.1186/1741-7007-7-72.
40. Mondani L, Piette L, Christen R, Bachar D, Berthomieu C, Chapon V. *Microbacterium lemovicicum* sp. nov., a bacterium isolated from a natural uranium-rich soil. *Int J Syst Evol Microbiol*. 2013;63:2600–2606. doi:10.1099/ijs.0.048454-0.
41. Boissière A, Tchioffo MT, Bachar D, Abate L, Marie A, Nsango SE, Shahbazkia HR, Awono-Ambene PH, Levashina EA, Christen R. Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. *PLoS Pathog*. 2012;8:e1002742. doi:10.1371/journal.ppat.1002742.
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–410. doi:10.1016/S0022-2836(05)80360-2.
43. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12:R60. doi:10.1186/gb-2011-12-6-r60.