Suddenly everyone is a microbiota specialist!

Stefan A. Boers
Ruud Jansen
John P. Hays

Clin Microbiol Infect 2016; 22: 581-582.

Recently there has been an explosion in the number of publications linking the human microbiota to various diseases. These microbiota profiles are obtained by either PCR amplification and sequencing of regions of the 16S ribosomal RNA (rRNA) gene of bacteria, or by performing shotgun metagenomics directly on sampled environments. As a simple guide to the critical analysis of microbiota-based publications, the authors present here the 'Ten-E' method. The majority of the described 'Es' can be readily applied to both 16S rRNA gene amplicon sequencing, as well as to shotgun metagenomics-based microbiota profiling studies. As a further note, the authors recommend the adoption of consistent and defined terms within the field of microbiome/microbiota research, as previously published.[1] The ten Es are presented in chronological order of a typical microbiota profiling project, starting with the E of Extraction.

Extraction (E1) – Different DNA extraction methods can seriously impact the final microbiota profiling results. As shown by Kennedy et al., there are significant differences in microbial composition when comparing microbiota profiles obtained from the same specimen using different DNA extraction kits.[2] Therefore caution is necessary when comparing microbiota studies that have used different DNA extraction methodologies.

Environment (E2) – Negative extraction controls should be included and analysed in the experimental protocol for low biomass specimens such as nose swabs, blood or other normally sterile sites. These controls are required to accurately assess the influence of contaminating DNA molecules that may be present in the experimental set-up. These contaminating DNA molecules may already be present in laboratory reagents or commonly used DNA extraction kits. Additionally, contaminating DNA molecules from the laboratory environment may be present on the surface of consumables used during PCR and/or metagenomic microbiota profiling experiments.[3]

Efficiency (E3) – During PCR amplification certain 16S rRNA gene sequences may be amplified more efficiently than others, biasing the resultant microbiota profiles. Amplification efficiency differences are prominent when applying standard PCR protocols but can be overcome by using clonal amplification by micelle PCR. In a micelle PCR, the template DNA molecules are separated into a large number of physically distinct PCR compartments, preventing amplification bias and increasing the accuracy of microbiota profiling methods.[4] Scientists should be aware of the potential for amplification bias during PCRs.

Exaggeration (E4) – Standard 16S rRNA gene PCRs will generate chimeric amplification products, whereby a single DNA amplicon comprises sequences that originate from multiple 16S rRNA genes. Importantly, the inclusion of chimeric sequences that were

not recognized by computational filtering software, leads to incorrect taxonomic identifications and an overestimated microbiota richness in the final microbiota profiling results. These chimeric sequences may be incorrectly identified as new bacterial species. Essentially, the prevention of chimeric sequences will prevent the microbiologist unwittingly becoming a 'bacterial creationist'. One method that can be used to reduce chimera formation is clonal amplification via the use of micelle PCR.[4]

Evaluation (E5) – The evaluation of sequence data by different clustering algorithms may lead to different microbiota results and this fact should be appreciated by scientists.[5] In addition, accurate taxonomic identification of 16S rRNA gene microbiota data depends on the quality and completeness of the reference databases used to identify and classify the sequence data produced, e.g. SILVA, RDP, GreenGenes and NCBI. Since most reference databases contain some unidentified and poorly annotated sequences, and are also inevitably incomplete, manual evaluation of the main sequencing results is to be encouraged. This to ensure that the taxonomic identification of 'key' bacterial genera and species within the microbiota profile are correct.

Elongation (E6) – In general, only short regions of bacterial 16S rRNA genes tend to be sequenced, meaning that these sequences may not have the discriminative power to identify bacteria to the species level. Though some bacterial genera may show sufficient inter-species 16S rRNA gene sequence diversity to allow their accurate identification (e.g. *Akkermansia muciniphila*), other genera may not have sufficient inter-species variation to allow their accurate speciation.[6] Additionally, the naming of species may vary over time.[7] In general, restricting sequence identification to the genus level (when using short 16S rRNA gene sequences), is recommended.

Equality (E7) – 16S rRNA gene sequencing does not generate accurate information regarding the quantification of bacterial species. Different bacterial species carry different numbers of 16S rRNA genes and copy numbers for all bacteria are not known. For example, the *Mycobacterium tuberculosis* genome carries one 16S rRNA gene copy, whereas the *Clostridium beijerinckii* genome carries up to 14 copies of the gene. Therefore, it is recommended that microbiota profiles are expressed as ratios or percentages of '16S rRNA gene copies' rather than ratios of 'species' (which would suggest that bacterial cell or genome copy numbers are being expressed). To provide an accurate number of bacterial genome copies, the use of methods such as calibrated quantitative PCR or digital PCR have to be employed.

Evidence (E8) – Microbiota profiles are generated using bioinformatics approaches and speculations about the clinical importance of the bacterial species usually ignore

Koch's postulates and/or the updated version of Koch's postulates for molecular diagnostics.[8,9] For example, a correlation between an operational taxonomic unit-associated disease, and its corresponding organism, should not be made without first fulfilling Koch's postulates. Currently, many potential disease-associated organisms discovered by microbiota analysis cannot be cultured (although this situation is slowly changing).[10] More effort should be spent on isolating these currently 'non-culturable' organisms before they can be truly associated with a particular disease or condition. Moreover, DNA-based studies do not allow for accurate differentiation between viable, non-viable or dead bacterial cells. This could be important for example, in specimens that have previously been treated with bacteriostatic antibiotics or in environmental samples where 'relic DNA' from dead cells can persist from weeks to years.[11] Therefore, scientists and stakeholders should remain sceptical regarding the scientific claims associated with a microbiota-based article.

Enrolment (E9) – Microbiota results are often obtained using small cohort-sized studies. However, the microbiota of many ecosystems and environments may be very complex and highly variable, even among similar samples. Many small-scale studies lack the statistical power to test microbiota-based hypotheses to a valid statistical conclusion. This lack of statistical evidence has resulted in a lack of agreement about the microbial composition of many studies published within the scientific literature.[12] Therefore, a larger sample of cohorts and/or meta-cohort analyses should be enrolled when generating conclusions regarding the 'typical' composition of a clinical or environmental sample.

Expectations (E10) – Be aware of possible conflicts of interest between sponsors of microbiota research and the researchers themselves in this highly competitive scientific field. Most journals specifically ask authors to state possible conflicts of interest in their manuscripts. However, readers should still be alert to potential funding biases that may skew published microbiota profiling results.

Finally, the authors hope that the 'Ten-E' protocol published here will aid microbiologists, clinicians, environmentalists, food technologists, journalists and even the general public, to be more critical of the scientific literature when it comes to the reporting of the results of microbiota profiling studies.

REFERENCES

1. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome* 2015; **3**: 31.
2. Kennedy NA, Walker AW, Berry SH, et al. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PloS one* 2014; **9**: e88982.
3. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; **12**: 87.
4. Boers SA, Hays JP, Jansen R. Micelle PCR reduces chimera formation in 16S rRNA profiling of complex microbial DNA mixtures. *Sci Rep* 2015; **5**: 14181.
5. Westcott SL, Schloss PD. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ* 2015; **3**: e1487.
6. Chakravorty S, Helb D, Burday M, et al. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 2007; **69**: 330e9.
7. Collins MD, Lawson PA, Willems A, et al. The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol* 1994; **44**: 812e26.
8. Fredricks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* 1996; **9**: 18e33.
9. Lipkin WI. Microbe hunting in the 21st century. *Proc Natl Acad Sci USA* 2009; **106**: 6e7.
10. Lagier JC, Hugon P, Khelaifia S, et al. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clin Microbiol Rev* 2015; **28**: 237e64.
11. Carini P, Marsden PJ, Leff JW, et al. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol* 2016; **2**: 16242.
12. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014; **15**: 382e92.