Published online: June 20, 2016

Science & Society

EMBO reports

Microbial forensics: next-generation sequencing as catalyst

The use of new sequencing technologies to analyze whole microbial communities could become a powerful tool for forensic and criminal investigations

Irene Kuiper

When Antonie van Leeuwenhoek experimented with his microscopes in the late 17th century, he observed, among many other small things, bacteria from the mouth cavity. van Leeuwenhoek's detailed descriptions of bacteria, spermatozoa, and cells marked the beginning of microbiology; his successors Robert Koch and Louis Pasteur stoked further interest in the microbial world with their discoveries of the role of bacteria in disease and fermentation. The light microscope was eventually augmented by ever more powerful tools—electron microscopes, fluorescence microscopy, RNA arrays, and DNA sequencers—to study the members of the microbial world and their interactions with each other and their host organisms, particularly humans.

One technique in particular has revolutionized microbiology: next-generation sequencing (NGS) or massive parallel sequencing [1]. The ability to sequence a mixture of millions of DNA molecules within one analytical run has created new opportunities to analyze whole communities of microbes rapidly and efficiently, including many species that cannot be cultured in the laboratory. New methods for RNA analysis have also tremendously improved microbiologists' ability to look at the molecular activities not just within single cells but whole bacterial communities.

In addition to inspiring new research in microbiology, NGS will likely have a major impact on forensic science and eventually become a new tool in police work and court cases. The forensic value of microbiology became apparent during the investigation of the anthrax attacks 1 week after the terrorist attack on September 11, 2001. Letters that contained spores of *Bacillus anthracis* were mailed to two US Democratic senators and various media offices in the USA. Twenty-two persons who opened these letters were infected with *B. anthracis* and five people died. The major challenge for forensic science, as part of the FBI's investigation, was to attribute the source of the bacterial spores in order to identify the perpetrator. Microbiologists all over the world understood that their subject could become part of a criminal investigation and used as trace evidence for a forensic examination.

> *"The forensic value of microbiology became apparent during the investigation of the anthrax attacks 1 week after the terrorist attack on September 11, 2001."*

Investigating the anthrax attacks took many years, partly because new techniques, including NGS, had to be validated for forensic use, which led to new forensic questions [2]. These included the choice of the technique best suited for securing forensic evidence, the availability of source and reference materials, notably different anthrax strains, and assessing the evidential value of the results.

The FBI formally closed its investigation in 2011 and, despite lingering doubts over the conclusion, the anthrax attacks and their investigation are regarded as an example for the use of microbiology in forensics. In the years since, many papers on forensic investigations of microbial samples have been published. Microbial forensics was born and its tasks now range from distinguishing arterial blood from respiratory blood, soil comparisons, human individualization, strain attribution in deliberate infections, estimating the human niche of epithelial cells, establishing the time of contact with water in drowning cases, and many others [3–5]. Despite the enormous possibilities, much groundwork is still needed before it can generate reliable forensic evidence for criminal law cases.

One of microbes' greatest advantages for forensic science is the fact that microbial communities are highly diverse and ubiquitous and can be found nearly everywhere on Earth. Molecular ecology and medical health research have extensively studied and described the composition and role of microbial consortia in the environment, in animals, and in humans [6,7]. Using NGS to sequence total DNA extracts from any sample makes it possible to identify different bacterial taxa and strains to gain an overview of the microbial population present. Older methods, such as DGGE, t-RFLP, and AFLP, also generated such population profiles, but NGS overcomes many limitations of these techniques, as it requires no additional fragment analysis and directly generates sequence information. The costs have also been decreasing rapidly, which makes the analysis of a larger number of samples possible.

Netherlands Forensic Institute, The Hague, The Netherlands. E-mail: i.kuiper@nfi.minvenj.nl
DOI 10.15252/embr.201642794

Along with the growth in knowledge about microbial consortia came new applications. Microbiologists and forensic experts realized that knowing how to reliably describe microbiological communities, or microbiomes, how these develop, how they are maintained and influenced by external factors, could be valuable for forensic investigations. Another feature why microbes are interesting for forensics is the fact that they are very small and present in huge numbers in or on limited trace material. This provides the opportunity to carry out a reliable and robust analysis and look at the whole community structure, which is the subject of many papers with forensic application [8,9].

The use of NGS to analyze a microbial community, and produce a forensic profile, opens new possibilities: Can we tell where the sample comes from: is it a fecal stain or soil? Can we tell whom it came from by comparing personal microbiomes? Can we establish were in the body it came from? When applying microbial forensics within casework, two important questions arise. First, how reliable is the technique used to analyze a microbiome? This requires a technical validation. Second, is the microbiome under analysis suited to answer the forensic questions in that it has the required spatial resolution? Can we reliably discriminate between individuals based on their microbiome? Can we distinguish between different microbial communities collected from different soil types or body niches? This requires a biological and forensic validation, which strongly depends on the applied technique, but also includes the application of this technique on real-case samples.

> *"One of microbes' greatest advantages for forensic science is the fact that microbial communities are highly diverse and ubiquitous and can be found nearly everywhere on Earth."*

Technical validation includes, among others, correct data handling. Although many NGS applications require a first PCR step to select the target molecules and although sequence length is limited, the amount of data, which can be extracted from a sample, is tremendous. 30,000–500,000 sequence reads of 50–700 nucleotides per run, depending on the chosen NGS platform, challenge the subsequent data processing, which brings us to the maybe most complex part of NGS. We strongly depend on bioinformatics, not only to produce useful data, but also to evaluate the quality of the NGS platform. Within forensics, transparent data processing is crucial and although many other fields of research require it as well, forensic standards are particularly challenging. Decisions on what data to accept or how to distinguishing background noise from signals can be challenged in the court and are often assessed by independent experts in some juridical systems.

To apply a technique in a reliable manner, information about general quality aspects such as detection limit, reproducibility, robustness, and so on must be established. For NGS, this comes down to how effective the platform separates single molecules, sequence quality in terms of the number of wrongly incorporated nucleotides, chimera formation in amplicon sequencing, filtering rare reads, and so on. Some of these aspects are inherent to the chosen technique, others depend on the bioinformatics techniques used to process the data. All together, this comes to a technical validation of any analytical method, which is a key element in many forensic labs working under certification or accreditation. Without this type of validation, one cannot even think of applying it to forensic casework.

Once this is in place, the biological validation must be established. If a particular microbial population, profiled with a suitable phylogenetic marker, is able to distinguish between individual persons, just claiming that a particular microbiome is unique for each person will not be sufficient. Biological validation requires demonstrating that indeed, each person has his or her individual microbiome. This can be done by analyzing a representative number of microbiomes and building a database with relevant samples. Such databases are crucial for forensic science, not only to show the biological validation in a probabilistic approach by analyzing a representative subset, but also to develop decision models and apply statistics on possible matches, partial matches, and mismatches.

But this is still not enough. After the biological validation, we need information on if and how the method can be applied to forensic samples: so-called trace evidence. Since trace evidence can be very limited depending on the case and the crime site, it affects the analysis of microbiome on the sample. In addition, trace evidence left at a crime scene is often exposed to a cascade of events—environmental exposure, mixing with other microbiomes from other sources, preservation after sampling, and so on—which generate additional differences between the microbiome present in forensic traces, and the microbiome sampled from a suspect. It requires assessing how much difference can be accepted and/or how much overlap is required to report a match between two profiles.

To further challenge the forensic validation of, for example, the spatial resolution of the samples, different scenarios can explain why the evidence was found at that particular site. Analyzing the microbiome sampled from a tablet computer and comparing it with the microbiome sampled from the hands of the person suspected of stealing it could be challenged by a statement that the suspect's brother stole it, but hid it in the suspect's bed. Can we distinguish both scenarios by comparing the mixed microbiome from the tablet with the microbiome from the hands of the suspect and from his brother?

> *"Within forensics, transparent data processing is crucial and although many other fields of research require it as well, forensic standards are particularly challenging."*

Another example from actual casework would involve soil from the bottom of a suspected burglar's shoe. If a partial shoeprint was found at the crime site, the prosecutor will need to know if the soil from the shoeprint originated from the suspect's shoes. It is relatively easy and straightforward to analyze the microbiomes from the shoe. But can we compare it reliably with the soil from the shoeprint, since both samples were not collected at the same time? Moreover, as the suspect walked over

different grounds and collected soil residues from other places, we would expect a different microbiome from the bottom of the shoe. If the microbiome from the shoe only partially matches the microbiome from the shoeprint, how can we qualify this match for the court to weigh this evidence?

> *"The combination of criminalistics, microbiology and technical expertise is needed to make microbiome analysis a valuable forensic tool."*

Microbiome analysis can also be used to exclude hypothesis about a crime. For example, as the vaginal microbiome differs substantially from the microbiome on the skin or in the gut [10], determining the absence of vaginal or fecal microbiota can be crucial in sexual assault cases. The power of such exclusions—of course taking into account other reasons for its absence such as washing—is often underestimated since the focus is usually on finding matches.

These examples show that many types of validation are required before a novel technique (NGS) combined with novel trace evidence (the microbiome) can be applied to forensic casework. This is of course common for all techniques and traces and a standard aspect of development of applied forensic science. It should certainly not hold us back from using it, but it is prudent and necessary to analyze and balance limitations and benefits.

The combination of criminalistics, microbiology, and technical expertise is needed to make microbiome analysis a valuable forensic tool. Exploring the possibilities and validating the methods, step by step, will clarify which questions can be reliably answered by the technique. It will surely take longer to reliably answer the question of whether the microbiome found on a suspect's hands can be linked to the microbes found on an item, rather than determining whether a stain on a crime scene comes from feces or from soil. The experts who apply microbial forensics in casework have a responsibility to address all the requirements discussed here and to acknowledge and clearly communicate its possibilities and limitations. The end-users of forensic investigations—prosecutors, judges, and juries—must be informed accurately and in a transparent way so they can evaluate the results and make proper decisions.

Even though legal systems around the world require different standards of evidence, one generally accepted criterion for scientific evidence is the "Daubert standard": the technique should be validated, peer-reviewed, and generally accepted by the relevant scientific community. Defining the relevant community can however be challenging, especially when new techniques such as NGS and new types of trace evidence such as microbiomes are combined into new applications. Nonetheless, this does not stop us from discussing, exploring, challenging, and finally applying microbiome analysis in forensics. With time, the technique will hopefully answer ever more questions and help criminal investigators solve puzzling cases.

Conflict of interest

The author declares that she has no conflict of interest.

References

1. Metzker ML (2010) Sequencing technologies; the next generation. *Nat Rev Genet* 11: 31 – 46
2. National Research Council of the Aational Academies (2011) *Review of the Scientific Approaches Used During the FBI's Investigation of the 2001 Anthrax Letters.* Washington, DC, USA: National Academies Press
3. Kakizaki E, Ogura Y, Kozawa S, Nishida S, Uchiyama T, Hayashi T, Yukawa N (2012) Detection of diverse aquatic microbes in blood and organs of drowning victims: first metagenomic approach using high-throughput 454-pyrosequencing. *Forensic Sci Int* 220: 135 – 146
4. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R (2010) Forensic identification using skin bacterial communities. *Proc Natl Acad Sci USA* 107: 6477 – 6481
5. Meadow JF, Altrichter AE, Baternan AC, Stenson J, Brown GZ, Green JL, Bohannan BJM (2015) Humans differ in their personal microbial cloud. *PeerJ* 3: e1258
6. The Human Genome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486: 207 – 214
7. Gilbert JA, Jansson JK, Knight R (2014) The earth microbiome project: successes and aspirations. *BMC Biol* 12: 69
8. Quaak FCA, Kuiper I (2011) Statistical data analysis of bacterial t-RFLP profiles in forensic soil comparisons. *Forensic Sci Int* 210: 96 – 101
9. Benschop CCG, Quaack FCA, Boon ME, Sijen T, Kuiper I (2012) Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? *Int J Legal Med* 126: 303 – 310
10. Costello EK, Lauber CL, Harnady M, Fierer N, Gordon JI, Knight R (2009) Bacterial community variation in human body habitats across space and time. *Science* 326: 1694 – 1697