Aligning access to microbiome data and privacy considerations for better solutions for health and wellbeing of society and environments

By Frederik COPPENS
VIB-UGent, ELIXIR Belgium, Gent Belgium

Lene LANGE
BioEconomy Research & Advisory, Copenhagen, Denmark

And Kathleen D'HONDT
Department Economy, Science and Innovation, Flemish Government, Brussels, Belgium

There is a growing body of evidence that underpins the importance of microbiomes in biology. Understanding the functioning of microbiomes and their interaction with the environments will allow to develop novel interventions to support human, animal, and plant health as well as the environment. The potential that microbiomes can have to prevent the onset of non-communicable diseases is huge. This can only be developed when studying the impact of lifestyle, nutrition and environment in the context of the genetic content. As human microbiomes have been shown to be stable over time and can allow to identify the "carrier" of the microbiome, access to microbiome data has been questioned in the light of privacy protection and the General Data Protection Regulation. In this paper we discuss the potential of microbiomes in different areas and how microbiome data may be shared to support the concept of doing good.

Introduction

Microbiome research has taken a giant leap forward over the last years. It is now generally accepted that microbiomes are an integral part of our body and all biological ecosystems. The role of microbiomes on our health, wellbeing, development and to the health and resilience of whole ecosystems also becomes better documented.

The growing insights on the impact of microbiomes has created an important potential for the development of novel intervention, both in human and in animal health, as well as for ag-food production.

Understanding microbiome functioning is based on the analyses of large data sets, combined with learning from insights across different expertise domains. The importance of learning across microbiomes of man, animals, plants, soil and food was underlined by the calls from researchers to initiate global microbiome initiatives to accelerate the technologies for analyses and agree on standardisation in the different aspects of microbiome research (Alivisatos *et al.*, 2015; Dubilier *et al.* 2015).

Microbiomal fingerprints

Nevertheless, open access to microbiome data has become an issue of discussion as it was shown that metagenomic microbial fingerprints are sufficiently stable to be linked to individuals, even with follow-up samples (Franzosa *et al.*, 2015). The gut microbiome proved exceptionally stable, where over 80% of samples uniquely matched in follow up samples. Other body site-specific microbiomes reached over 30% positive matches. This makes microbiome-related findings a powerful clinical tool for care management, but raises privacy concerns (Chuong *et al.*, 2017). Indeed, associations between human host genotype and gut microbiome have been found (Hughes *et al.*, 2020), and the skin microbiome leaves a signature that allows to identify the person that last touched a surface (Elhaik *et al.*, 2021).

Privacy issues also arise when studying the health impact of lifestyle and diet through the microbiome. When combining such data to microbiome data and health information, analyses may reveal unexpected correlations on metabolic diseases, allergies or intolerance to certain foods. Understanding the interplay of nutrition, lifestyle, microbiome, and genome is believed to open novel manners for better health management and prevention. Nutrition and lifestyle have a major impact on the occurrence of non-communicable diseases (NCDs), which continue to have an increasing impact across the globe[(1)].

Combining nutrition and lifestyle information with genomes and microbiomes data

In the aftermath of the human genome mapping, the field of nutrigenomics was launched to understand the relationship between nutrients, diet and gene expression. More recently it became clear that the associated microbiome may be seen as an intermediate that translates food, nutrition and lifestyle into health effects. However, the multifactorial and long-term impact of diet and lifestyle on health, based on the interplay of genome-epigenome-microbiome is too complex for straightforward conclusions.

It needs large data sets gathered during lifetime to characterise how nutrition and lifestyle exert health impacts. Advanced functional prediction and artificial intelligence are expected to lead to novel pathways and interventions to prevent or postpone NCDs. Ultimately, personalised diets for better health outcomes may be defined and guidance for healthy diets targeting different types of societal groupings at different stages in life – from new-borne, childhood, adulthood to senior – as well as related to gender and/or ethnicities may be provided. The role of the gut microbiome for the onset and severity of NCDs, including mental conditions has become increasingly clear.

Accelerating this field will transform healthcare from disease to health management thereby contributing to the sustainability of our healthcare systems. In this respect, the concept of preventive personalised health, emphasising on the role of lifestyle, diet and microbiomes to realise health impacts, has been launched.

Industry perspectives based on microbiome insights

At the same time embracing the potential of microbiome research is also opening a huge potential for the industry, such as the diagnostics or food, feed and diet industry. Microbiome-based applications are also being developed in the bioag sector, mainly in the area of plant protection, soil resilience and improvement, yield improvements, waste management, and contributing to circular economy and organic farming. Finally, this data intensive field also creates extra activities for service providers.

Market prognoses by different consultancy companies estimate the global microbiome-based interventions for health to reach roughly over USD 1-2 billion with a compound annual growth rate (CAGR) of between 15%-20% or even up to 30% for the next 10 years, dependent on the inclusion criteria[(2)].

Also the global bioag industry is expected to be an important growth market with an expected CAGR of around 14%, to reach over USD 10 billion by 2025-2027[(3)].

Open big data effort needs

To realise these expectations, different types of data sets should be accessible and combined. The information buried in combined data sets holds enormous promise to develop personalised nutrition, lifestyle approaches or health interventions to conserve health and contribute to better wellbeing.

Open access to data has accelerated research and development considerably (Burgelman *et al.*, 2019). Data sets covering different microbiome domains and integrating the different data resources will lead to more profound insights and contribute to better solutions to address challenges in public health, climate change and food security (D'Hondt *et al.*, 2021).

Not only within the health field it is important to link the different microbiomes and view these as a continuum, also holds true for environmental microbiomes in soil, on and in animals, plants and crops. In view of the one-health concept, all microbiomes are in continuum interacting. The interaction among the microbiomes from different environments will give new insights in the biology and biodiversity. In addition, research is accelerated in a shared knowledge environment covering different microbiome areas. Open access to data resources is a strong means to accelerate knowledge distribution, but as outlined higher, the sharing of human microbiome data in open access is under discussion. As even summary statistics risk to expose privacy sensitive information, this warrants consideration and the development of guidelines for data release (Cho, 2021).

Open science and privacy protection are two sides of a coin that are in conflict to each other. These considerations are manifest in particular, in the realm of personalised medicine. Comparisons of genomes, documented with associated health data is instrumental to develop personalised medicine. The risk of such data being used by insurers, banks, employers to inform decisions whether or not to grant an insurance a loan or a job should be avoided. Several countries have adopted specific restrictions for the use of genomic information for crime investigation. Access to such information is granted on a case by case approach and jurisprudence.

(1) www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

(2) https://www.marketsandmarkets.com/Market-Reports/human-microbiome-market-37621904.html; https://www.mordorintelligence.com/industry-reports/human-microbiome-market; https://www.businesswire.com/news/home/20211221005468/en/Global-Human-Microbiome-Industry-Landscape-2021-2028---New-Drug-Requirements-for-Faecal-Microbiota-Presents-Opportunities---ResearchAndMarkets.com

(3) https://www.fortunebusinessinsights.com/industry-reports/agricultural-microbial-market-100412; https://www.marketsandmarkets.com/Market-Reports/agricultural-microbial-market-15455593.html

Privacy protection

Different options may be envisioned to deal with the privacy issues linked to the sharing of microbiome data. Microbiome fingerprints may be handled like a genomic passport or profile, i.e. for healthcare treatment this data should be available.

For secondary use, the sharing of sensitive data is currently typically based on contracts, after evaluation by a data access committee. As complete anonymisation is not possible, like in genomic analyses, rather than sharing the data, the analysis may be brought to the data. Through a federated infrastructure, the exact same analysis is performed locally on each dataset, and only the resulting (aggregated) data is shared, reducing or even eliminating privacy concerns[(4)]. This method can also alleviate the technical challenges associated to transferring the ever increasing volumes of data. However, this also imposes limitations which may require development of new algorithms. Trusted research environments are being established to provide a secure computing environment to access sensitive data [(5) and (6)].

Alternatively, analyses can be performed on encrypted data (Senf *et al.*, 2021). The Global Alliance for Genomics and Health, is developing standards to enable such innovative approaches[(7)]. These methodologies are being developed and deployed in the context of the 1-Million Genomes Project[(8)].

The field may also be advanced without compromising human health data protection, by drawing lessons from studies on farmed animals. In this setting feeding, age, gender and genetic background can be accurately tracked, while the impact on the microbiome composition and function can be analysed. The data collected in this way will increase the understanding of the functioning of microbiomes and role for animal health and welfare. It may allow to extrapolate insights from lifestock to approaches for human health.

The risk of privacy intrusions should not block research and hence the benefits that sharing of data may bring. Restricting access to human microbiome data will also impact other microbiome areas, as research on human linked microbiomes is most advanced, due to the potential to develop novel health interventions.

Policy and decision makers should engage to establish the right measures for correct use of data, in support of R&I, while preventing all types of misuse. Creating the right conditions to grant access to data is essential to ensure trust among citizens to agree on the sharing of their data.

(4) https://ega-archive.org/federated

(5) www.hdruk.ac.uk/access-to-health-data/trusted-research-environments/

(6) https://research.csc.fi/sensitive-data-services-for-research

(7) www.ga4gh.org

(8) https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes

Ownership of data

Another important aspect is the ownership of data and biological samples. According to the current legislation people own their health data, including their genome data and microbiome data. The governance of this data is with healthcare professionals, health institutions, insurers or governments. The individual is requested to give consent to use his or her data, but often lacks information about later use of the data. Nevertheless, people are most often interested to share their data and bio-sample, thereby supporting research as a token of good citizenship. Ownership of biobank samples is also a matter of debate.

There is no consensus yet whether microbiomes should be considered as part of the human body and hence be treated like human tissue. In particular, for faecal samples, which are generally considered waste in the first place, it becomes difficult, since it was shown that microbiomes are sufficiently unique to be linked to individuals.

Citizens should at all times have governance over their data, including the microbiome data. Informing people on what their data will be used for and on the outcomes of studies should become mandatory. At the same time, the raw data should not be the source for profit purposes. Data is referred to as the new gold, while it may be questionable whether it is acceptable to use the combined data that form the basis of our common co-evolution with microbiomes for profit and hence the basis to better understand the underlying biology, which will lead to novel interventions and treatments for better wellbeing. In contrast, novel interventions based on more profound understanding of the functioning of microbiomes may be offered for profit.

Building the right framework conditions to protect against misuse of data and ensure awareness building and informed citizens are essential.

Ethical considerations

Restrictions for use of microbiome data, due to the (interpretation of the) General Data Protection Regulation[(9)], data ownership and privacy interferes with getting to better understanding of microbiome compositions, roles and functions and represents an ethical dilemma as it limits the goals of doing good, based on the growing body of knowledge of the biology and co-evolution including the microbiomes in and around them.

The co-evolution and continuity of microbiomes is well illustrated in the bioag field. Enrichments of plant microbiomes of the root, leaf, or seed, may protect plants against pests and hence reduce the need for pesticides. In this way, microbiomes may contribute to higher yields and even more robustness against stress conditions.

In lifestock breeding, improvement of the gut microbiomes of farmed animals reduces the risk of inflammatory gut, which reduces the need for antibiotics.

(9) https://eur-lex.europa.eu/eli/reg/2016/679/oj

This contributes not only to better animal health, but more importantly, may significantly reduce the need for antibiotics and thus reduce the risk of antibiotic resistance. The spread of antibiotic resistant bugs is one of the most serious threats for future pandemics.

Microbiomes have potential to bring benefits in bioag, human health and environment. The protection of data should not be the limiting factor as it may be considered un-ethical not to share such knowledge.

Conclusions

Integrating the potential of microbiomes for human and animal, environmental and planetary health offers huge opportunities for health and wellbeing. The sharing of data is crucial to accelerate the insights on how microbiomes function and interact in the environments. It will lead to understanding what determines resilience and health of microbiomes and how this impacts health of living organisms and the environment.

However, access to health data should not be misused for unintended purposes. The data needed to build the knowledge should not be the source of profit unlike the development of novel interventions based on the analyses of these data.

Human microbiome data should be treated with similar privacy concerns like genomic data. Working in secured data computing environments, bringing the analyses to the data stored in federated infrastructures and sharing only the resulting aggregated data and using encrypted data may provide novel approaches to ensure privacy protection.

Bibliography

ALIVISATOS A. P., BLASER M. J., BRODIE E. L., CHUN M., DANGL J. L., DONOHUE T. J., DORRESTEIN P. C., GILBERT J. A., GREEN J. L., JANSSON J. K., KNIGHT R., MAXON M. E., MCFALL-NGAI M. J., MILLER J. F., POLLARD K. S., RUBY E. G. & TAHA S. A. (2015) – Unified Microbiome Initiative Consortium, "A unified initiative to harness Earth's microbiomes", *Sciences* 350, pp. 507-508, doi: 10.1126/science.aac8480. Epub 2015 Oct 28. PMID: 26511287.

BURGELMAN J. C., PASCU C., SZKUTA K., VONSCHOMBERG R., KARALOPOULOS A., REPANAS K. & SCHOUPPE M. (2019), "Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century", *Frontiers in big data* 2, p. 43, https://doi.org/10.3389/fdata.2019.00043

CHO J. C. (2021), "Human microbiome privacy risks associated with summary statistics", *PloS one* 16, e0249528, https://doi.org/10.1371/journal.pone.0249528

CHUONG K. H., HWANG D. M., TULLIS D. E., WATERS V. J., YAU Y. C., GUTTMAN D. S. & O'DOHERTY K. C. (2017), "Navigating social and ethical challenges of biobanking for human microbiome research", *BMC medical ethics* 18, p. 1, https://doi.org/10.1186/s12910-016-0160-y

D'HONDT K., KOSTIC T., MCDOWELL R., EUDES F., SINGH B. K., SARKAR S., MARKAKIS M., SCHELKLE B., MAGUIN E. & SESSITSCH A. (2021), "Microbiome innovations for a sustainable future", *Nat. Microbiol.* 6, pp. 138-142, doi: 10.1038/s41564-020-00857-w. PMID: 33510435.

DUBILIER N., MCFALL-NGAI M. & ZHAO L. (2015), "Microbiology: Create a global microbiome effort", *Nature* 526, pp. 631-634, https://doi.org/10.1038/526631a

ELHAIK E., AHSANUDDIN S., ROBINSON J. M., FOSTER E. M. & MASON C. E. (2021), "The impact of cross-kingdom molecular forensics on genetic privacy", *Microbiome* 9, p. 114, https://doi.org/10.1186/s40168-021-01076-z

FRANZOSA E. A., HUANG K., MEADOW J. F., GEVERS D., LEMON K. P., BOHANNAN B. J. & HUTTENHOWER C. (2015), "Identifying personal microbiomes using metagenomic codes", *Proceedings of the National Academy of Sciences of the United States of America* 112, pp. E2930-E2938, https://doi.org/10.1073/pnas.1423854112

HUGHES D. A., BACIGALUPE R., WANG J., RÜHLEMANN M. C., TITO R. Y., FALONY G., JOOSSENS M., VIEIRA-SILVA S., HENCKAERTS L., RYMENANS L., VERSPECHT C., RING S., FRANKE A., WADE K. H., TIMPSON N. J. & RAES J. (2020), "Genome-wide associations of human gut microbiome variation and implications for causal inference analyses", *Nature microbiology* 5, pp. 1079-1087, https://doi.org/10.1038/s41564-020-0743-8

SENF A., DAVIES R., HAZIZA F., MARSHALL J., TRONCOSO-PASTORIZA J., HOFMANN O. & KEANE T. M. (2021), "Crypt4GH: a file format standard enabling native access to encrypted data", *Bioinformatics (Oxford, England)* 37, pp. 2753-2754. Advance online publication, https://doi.org/10.1093/bioinformatics/btab087