AI-powered massive microbiome cultivation

A new data-driven machine-learning culturomics framework advances bacterial high-throughput isolation beyond current limits

Marta Selma-Royo [1], Nicola Segata [1,^], Liviana Ricci [1]

1. Department CIBIO, University of Trento, Italy
^ Correspondence: Nicola Segata nicola.segata@unitn.it

The study of the human microbiome is continuing to deliver important insights into the biological and translational implications of our microbial ecosystems. While metagenomics is a driver of such new studies, cultivation of human microbiome members remains key to identify and verify new molecular and causal mechanisms as well as to provide collections of reference genomic information. However, microbiome cultivation is technologically and intrinsically difficult to scale to a high number of samples and high number of isolates per sample. A new work in Nature Biotechnology by Huang et al. [1] presented a new machine-learning-based robotic system to systematically isolate a high number of different bacterial taxa present in biological samples. The workflow is a substantial advance to the field of high throughput cultivation as it integrates AI-informed colony picking into automatized isolation pipelines expanding the available toolsets for the exploration of the hidden microbiome diversity and the generation of more complete reference genome compendia.

Historically, bacterial isolation has been based on manual colony picking from isolation media. While this approach at cultivation is still efficient for well-characterized pathogens with validated protocols [2], it presents several challenges when applied to human microbiome members including the difficulty of batch reproducibility, the limitations of time-consuming protocols that are prone to experimental errors and contamination, and the lack of specific protocols for many taxa. Microbial culturomics for non-pathogens is however gaining momentum with technologies and approaches able to increase the isolation efficiency. Parallelization and miniaturization of assays with multiple culture conditions are now a possibility [3], and mass spectrometry identification is an option for quick screening within protocols. High throughput approaches to culturomics such as microfluidic systems [4] and FACS-based pipelines [5] have overcome some of the remaining hurdles that hinder traditional microbial isolation techniques from being applied on complex ecosystems, including the elimination of out-competition of fastidious, low abundance species by fast-growing or predominant taxa. Nonetheless, systematic methods for building comprehensive strain collection with integrated downstream identification of obtained isolate representatives are still lacking.

The new system - the Culturomics by Automated Microbiome Imaging and Isolation (CAMII) - integrates automatic picking of bacterial colonies maximizing morphological diversity with the high-throughput acquisition of single-colony genomic data. The method enables the creation of large-scale bacterial collections from each single microbiome sample. Images from each isolation plate are captured and automatically analyzed to segment the colonies and identify those that are unique based on morphological features including area, perimeter, circularity, convexity, among others. Then, an automated

protocol picks the list of colonies that have been selected by the software and transfers them into liquid media for downstream genomic DNA extraction and for long-term storage. The authors also developed a computational algorithm that exploits phenotypic information to improve colony isolation diversity maximizing morphological differences between isolates and allowing model-driven isolation of target taxa.

Compared with previously described automated methods for high-throughput microbial isolation, the framework by Huang et al. provides a database with objective and systematic morphological data (shape, size, color, etc.) which is used to train the machine-learning approach to drive the colony picking step. Morphological data are collected in an open-access comprehensive database (http://microbial-culturomics.com/) that promises to inform future targeted isolation efforts. By efficiently maximizing the diversity of the downstream characterized microbes, this process supports construction of extensive isolate collections, and this in turn facilitates both the description of new taxa and the exploration of the intra-species strain diversity of already known species. Both such tasks are theoretically possible also using metagenomics [6,7], but the possibilities that the CAMII system offers with its single-strain handling and sequencing enables a much higher accuracy that is needed for investigations of species variability, molecular evolution, horizontal gene transfer characterization, and ecology of multiple conspecific strain coexistence.

A key application area of the new system is the systematic effort at uncovering the "microbial dark matter". While metagenomics has vastly advanced the identification of new species, it is estimated that a large proportion of gut microbes remains yet to be isolated and cultured [8], and such proportion is much higher for non-human microbiomes. An efficient culturomics workflow may enable uncovering of the unknown (and thus its functional experimental characterization) by isolating species previously only pinpointed by metagenomic surveys. Cultivation of microorganisms identified by metagenomics ("the known unknown") is essential to unravel the interaction of specific bugs with the host and for the development of live biotherapeutics (e.g. next generation probiotics). The isolation capacity of this system and its increased efficiency to obtain previously uncultured species contribute to expanding both reference genomes databases and bacterial culture collections thus improving detection of the microbial dark matter in the future metagenomic studies.

Despite the promising applications of the system presented by Huang et al., further research is needed to address some limitations of the technology. While CAMII appears to increase the efficiency compared to traditional isolation methods, the system relies on systematic picking of a high number of colonies and shows significant species and strain redundancy. Also, it would be important to somehow expand the system to handle multiple culture conditions in parallel (in terms of media, substrates, diverse pH, not tested in the study) to further improve the diversity of the isolated species and avoid biasing the experiment toward a specific and single condition. Of note, colony morphology features may be highly specific to the implemented culture conditions [9,10] and may show high inter-donor variability. Thus, intensive data collection is still required to evaluate the efficiency of targeted isolation experiments.

In the new work, CAMII is also set up to explore interspecies interactions (e.g. competitive or cooperative) based on the spatial arrangement of colonies on the isolation plates. While

this is an extremely appealing approach to gain insights into the ecology of microbiomes, the reliability of these interactions and their relevance in the original complex microbiome need confirmation by other methodologies. Further, the use of traditional sample plating implemented in this framework does not fully eliminate common obstacles to successful cultivation of fastidious gut anaerobes such as out-competition by dominant bacteria. A combination of the approach with multiple cultivation media could thus further improve the likelihood that the new machine-learning system will be able to maximize the picking of low-abundant frequently outcompeted strains that are recalcitrant to traditional laboratory culture conditions.

The isolation and cultivation of human associated microbes represent an increasingly key component of the microbiome research field. By leveraging automation and AI, the CAMII platform offers a promising approach to encourage efficient isolation of diverse microbes from multiple samples as well as the compilation of an exhaustive isolate database, with multiple applications in the microbiome analysis that needs to be further explored.

1. Huang, Y. *et al.* High-throughput microbial culturomics using automation and machine learning. *Nat. Biotechnol.* (2023).
2. Larsen, J. *et al.* Evaluation of a widely used culture-based method for detection of livestock-associated meticillin-resistant Staphylococcus aureus (MRSA), Denmark and Norway, 2014 to 2016. *Euro Surveill.* **22**, (2017).
3. Lagier, J.-C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology* vol. 1 Preprint at https://doi.org/10.1038/nmicrobiol.2016.203 (2016).
4. Jiang, M.-Z. *et al.* Droplet microfluidics-based high-throughput bacterial cultivation for validation of taxon pairs in microbial co-occurrence networks. *Sci. Rep.* **12**, 18145 (2022).
5. Bellais, S. *et al.* Species-targeted sorting and cultivation of commensal bacteria from the gut microbiome using flow cytometry under anaerobic conditions. *Microbiome* **10**, 24 (2022).
6. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649–662.e20 (2019).
7. Tett, A. *et al.* The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host Microbe* **26**, 666–679.e7 (2019).
8. Thomas, A. M. & Segata, N. Multiple levels of the unknown in microbiome research. *BMC Biol.* **17**, 48 (2019).
9. Sousa, A. M., Machado, I., Nicolau, A. & Pereira, M. O. Improvements on colony morphology identification towards bacterial profiling. *J. Microbiol. Methods* **95**, 327–335 (2013).
10. Tasaki, S., Nakayama, M. & Shoji, W. Self-organization of bacterial communities against environmental pH variation: Controlled chemotactic motility arranges cell population structures in biofilms. *PLoS One* **12**, e0173195 (2017).