Received: 26 June 2024 | Accepted: 12 August 2024

DOI: 10.1111/1751-7915.70004

MICROBIAL BIOTECHNOLOGY Open Access

OPINION

Improving the odds: Artificial intelligence and the great plate count anomaly

Detmer Sipkema

Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

Correspondence
Detmer Sipkema, Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands.
Email: detmer.sipkema@wur.nl

Funding information
HORIZON EUROPE Innovative Europe, Grant/Award Number: H2020-FNR-2020-2-101000392

Abstract

Next-generation DNA sequencing has shown that the great plate count anomaly, that is, the difference between bacteria present in the environment and those that can be obtained in culture from that environment, is even greater and more persisting than initially thought. This hampers fundamental understanding of bacterial physiology and biotechnological application of the unculture majority. With big sequence data as foundation, artificial intelligence (AI) may be a game changer in bacterial isolation efforts and provide directions for the cultivation media and conditions that are most promising and as such be used to canalize limited human and financial resources. This opinion paper discusses how AI is or can be used to improve the success of bacterial isolation.

INTRODUCTION

Since the beginning of this century, biological dark matter has been brought to light at an ever-increasing speed. Metagenomic sequencing of environmental samples from across the globe have shown the same pattern over and over again: The large majority of bacteria as observed through the presence of their DNA has never been obtained in culture (Dat et al., 2021; Hug et al., 2016; Parks et al., 2018), also coined as the 'great plate count anomaly' (Staley & Konopka, 1985). There are of course success stories where previously uncultivable bacteria are isolated in the laboratory through just standard plating or more advanced methods, but the rate of discovery of microbes that are as yet uncultivable outpaces the rate of isolating novel species (Gutleben et al., 2018; Steen et al., 2019). As such, the percentage of cultivable bacteria is actually decreasing and a breakthrough is needed to diminish the growing gap between in situ (as demonstrated in silico with omics methods) and in vitro.

Millions of bacterial metagenome-assembled genomes (MAGs) and the hypotheses they give rise to, make isolation of bacteria more valuable than ever. It is the most reliable way for annotation and functional characterization of newly discovered and hypothetical genes. In addition, isolates allow to study bacterial metabolism at the biochemical level, which is important for biotechnological development. Interestingly, many MAGs of uncultured bacteria harbour biosynthetic gene clusters (encoding the machinery to produce specialized metabolites) that belong to different clades than their isolated counterparts (Paoli et al., 2022). The latter is particularly relevant for our quest to find new cures to beat old foes that have become resistant to clinically used antibiotics. Bacteria are the primary source of bioactive natural products, with over 50% of current pharmaceutical drugs directly derived from, or inspired by, microbial compounds (Newman & Cragg, 2020). However, preclinical development of novel drugs relies on the availability of milligrams to grams of the pure compound, which are difficult to obtain without being able to grow the bacterium in vitro. In addition, the outlook of not having even a basic production platform in place refrains companies from preclinical trials with these hard-to-get molecules. On the other hand, re-discovery upon re-isolation of the same old bacteria often leads to discovery of molecules that are only a variation on a theme, rather than finding truly novel chemical scaffolds and this endangers our

potential to treat future or already current infectious diseases, also in richer countries (Lockhart et al., 2023). Bacterial growth media compositions and cultivation conditions to isolate currently uncultivable bacteria can theoretically be varied almost infinitely and increased creativity on how to isolate bacteria has yielded a number of successes (Lewis et al., 2021). As time and money are rarely even close to infinite, a paradigm shift is needed to make the leap from trial and error on an agar plate to targeted isolation of members of the most wanted bacterial taxa. And it may be artificial intelligence (AI) to fuel this transition.

FIRST AI APPLICATIONS IN BACTERIAL CULTIVATION

Currently, AI is implemented only to a limited extent in bacterial isolation. In clinical microbiology laboratories, assessment of presence or absence of growth on agar plates and counting colonies is a substantial part of the routine workload. AI algorithms have been developed to automatically identify the appearance of bacterial colonies (Jacot et al., 2024) and enumerate them, greatly reducing time for standard operations (Makrai et al., 2023; Zhang et al., 2022). Further, future advantages are earlier than the human eye detection of growth thereby reducing the culture read-out time (Jacot et al., 2024). One step beyond is AI-based identification of bacterial species. Currently, species identification mostly relies on biochemical tests (for well-known subsets in a clinical setting) or sequencing of phylogenetic marker genes. However, both methods typically require 8–24 h and are rather costly (Tran et al., 2015). MALDI-TOF MS-based identification has greatly reduced costs and time needed for identification of known bacterial pathogens (Yo et al., 2022), but purchasing the equipment needed for this type of analysis is a too large capital cost for many laboratories and industry startups. In a recent study, an AI-based complex hierarchical network capable of handling bacterial identification, within a set of 32 urinary tract bacterial pathogens was generated that could assess single colonies or whole plates (Signoroni et al., 2023). Although these methods will probably not be feasible for bacterial identification of highly complex or more unknown sources due to the limited variation in morphological characteristics of bacteria, wide-range adoption in the clinical setting may be expected because universally available equipment, such as a cell phone camera offers sufficient resolution needed for AI-based image analyses.

AI-ASSISTED BACTERIAL ISOLATION

However, the Holy Grail would be AI-assisted isolation of members of the uncultivated majority. Biological dark matter is not as dark anymore as it used to be before the next generation sequencing revolution as MAGs can now routinely be generated for uncultured bacteria that make part of complex communities in environmental samples (Almeida et al., 2019; Ma et al., 2023; Nishimura & Yoshizawa, 2022). These MAGs have not only revealed diverse biosynthetic gene clusters, but also provide a glimpse of the putative lifestyle of these bacteria that can be used for generating genome-scale metabolic models (GEMs). A GEM is a mathematical representation of the metabolic repertoire of a microorganism that is based on (meta)genomics, (meta)transcriptomics, (meta)proteomics and/or metabolomics data (Ankrah et al., 2021). As such, they can be used to simulate the growth of currently uncultivable microorganisms. The avalanche of omics data of the uncultivated majority that has been generated during the last 10 years is awaiting systematic computational analyses to generate GEMs for bacterial species, genera and families that are currently without cultured representatives. GEMs are by no means magic wands that will bring bacteria into culture with a slight touch, because MAG quality is often limited by the suboptimal sequencing strategies applied (Benoit et al., 2024). In addition, a structural uncertainty of the models generated for non-model organisms is caused by poorer gene annotation, higher inaccuracy related to genome gap-filling (Chen et al., 2023) and limitations with respect to metabolic flux balance analyses as metabolite concentrations in environmental samples (if at all possible to measure) cannot easily be linked to individual species. A hierarchical AI-based network of algorithms (each with its own flaws) can be used to refine the prediction of the metabolism of the target bacterium in its natural environment (Figure 1). For example: (1) Plain GEMs can be used to predict growth substrates based on presence or absence of primary metabolic pathways, and when metabolic fluxes can be simulated the growth rate can be estimated as well (Passi et al., 2021); (2) Computational models using amino acid frequencies derived from (partial) genomes can be used to predict physicochemical requirements, such as oxygen tolerance, optimum growth temperature, salinity and pH for those species (Barnum et al., 2024); (3) Resistance to certain antibiotics of target bacterial species can be predicted based on its genome sequence using a machine learning approach (Ren et al., 2022), which can subsequently be exploited by adding those antibiotics to the cultivation medium for selective enrichment of the target; (4) Hidden Markov models to predict surface-exposed epitopes can be applied for reverse-genomics approaches to isolate target bacteria (Cross et al., 2019); (5) Computational models can be used to predict the growth state of the cell in the environment (whether it is actively dividing or 'viable-but-not-culturable'; and perhaps, still many types of available data are waiting for models to predict particular aspects relevant for the isolation of currently uncultivated bacteria).

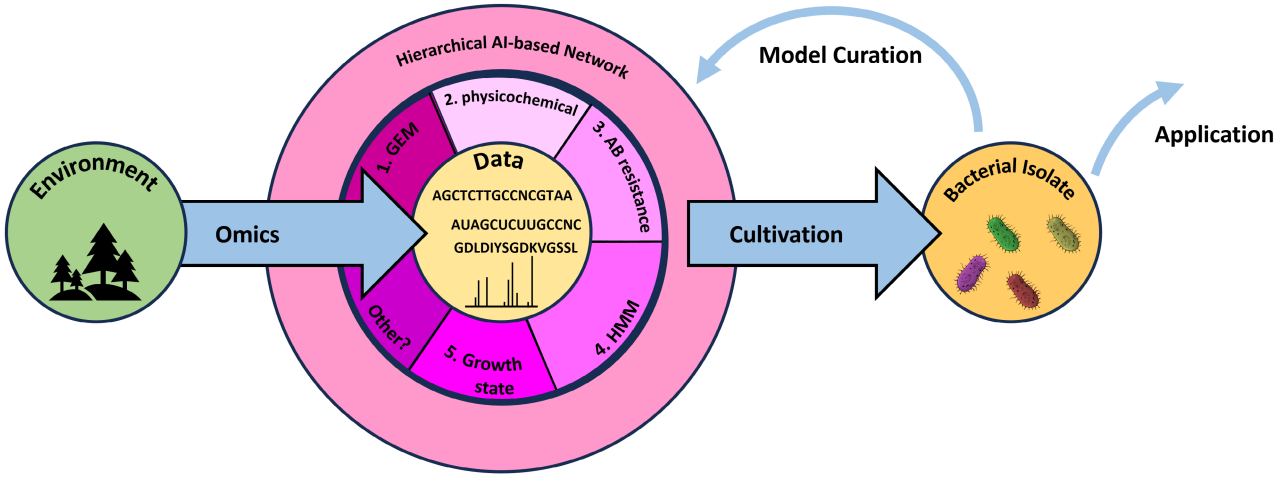

FIGURE 1 Omics-based data from any environmental sample is the foundation of a diverse suite of AI-based tools (GEM = Genome-scale metabolic model, AB resistance = antibiotic resistance and HMM = Hidden Markov Model) used to predict metabolism, growth conditions and the growth state in the natural environment of individual bacterial species. Integration of different models into a hierarchical AI-based network may lead to more accurate prediction of components and conditions to be tested in the laboratory to isolate currently uncultivable bacteria. Bacterial isolation results can be used for AI model curation and—if successful—for biotechnological application.

CONCLUDING REMARKS

We should not expect AI to provide us with clear recipes and instructions how to isolate the bacterium of the day. However, AI can provide new hypotheses that go beyond individual researcher's imagination and turn dark microbial matter into 'grey microbial matter' and thus reduce the search space with respect to the media compositions and cultivation conditions to be tested. In addition, the dirty laboratory work will remain to rely on expert intuition and tricks, for example, the use of baffled or non-baffled flasks that cannot easily be computed. A glimpse of what may be the future of bacterial isolation can already be seen in two recent studies where AI models were combined with robotics to increase the scale and speed of bacterial isolation, while relying on phenotype and genotype data (Dama et al., 2023; Huang et al., 2023). The CAMII model is based on automatic bacterial colony imaging and cultivation to maximize diversity and novelty of the obtained isolates (Huang et al., 2023), while BacterAI uses a reinforced learning algorithm to gamify a bacterial isolation experiment in an iterative manner to cross the boundary between no growth and growth for individual strains (Dama et al., 2023). It is computationally challenging to build a hierarchical model integrating the different levels of information that can be obtained from the different models, but the massive force of AI will undoubtedly improve our odds for isolating members of the uncultivated majority.

AUTHOR CONTRIBUTIONS

Detmer Sipkema: Conceptualization; investigation; writing – original draft; visualization; methodology; funding acquisition.

ACKNOWLEDGEMENTS

The authors acknowledge funding from European Commission; HORIZON EUROPE Framework Programme; HORIZON EUROPE Innovative Europe (H2020-FNR-2020-2-101000392). The authors declare no conflict of interests.

CONFLICT OF INTEREST STATEMENT

The author has no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Detmer Sipkema https://orcid.org/0000-0003-3836-219X

REFERENCES

Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A. et al. (2019) A new genomic blueprint of the human gut microbiota. *Nature*, 568, 499–504.

Ankrah, N.Y.D., Bernstein, D.B., Biggs, M., Carey, M., Engevik, M., García-Jiménez, B. et al. (2021) Enhancing microbiome research through genome-scale metabolic modeling. *mSystems*, 6, e0059921.

Barnum, T.P., Crits-Christoph, A., Molla, M., Carini, P., Lee, H.H. & Ostrov, N. (2024) Predicting microbial growth conditions from amino acid composition. *bioRxiv*, https://doi.org/10.1101/2024.03.22.586313

Benoit, G., Raguideau, S., James, R., Phillippy, A.M., Chikhi, R. & Quince, C. (2024) High-quality metagenome assembly from long accurate reads with metaMDBG. *Nature Biotechnology*. Available from: https://doi.org/10.1038/s41587-023-01983-6

Chen, C., Liao, C. & Liu, Y.-Y. (2023) Teasing out missing reactions in genome-scale metabolic networks through hypergraph learning. *Nature Communications*, 14, 2375.

Cross, K.L., Campbell, J.H., Balachandran, M., Campbell, A.G., Cooper, C.J., Griffen, A. et al. (2019) Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nature Biotechnology*, 37, 1314–1321.

Dama, A.C., Kim, K.S., Leyva, D.M., Lunkes, A.P., Schmid, N.S., Jijakli, K. et al. (2023) BacterAI maps microbial metabolism without prior knowledge. *Nature Microbiology*, 8, 1018–1025.

Dat, T.T.H., Steinert, G., Cuc, N.T.K., Smidt, H. & Sipkema, D. (2021) Bacteria cultivated from sponges and bacteria not yet cultivated from sponges—a review. *Frontiers in Microbiology*, 12, 3427.

Gutleben, J., Chaib De Mares, M., Van Elsas, J.D., Smidt, H., Overmann, J. & Sipkema, D. (2018) The multi-omics promise in context: from sequence to microbial isolate. *Critical Reviews in Microbiology*, 44, 212–229.

Huang, Y., Sheth, R.U., Zhao, S., Cohen, L.A., Dabaghi, K., Moody, T. et al. (2023) High-throughput microbial culturomics using automation and machine learning. *Nature Biotechnology*, 41, 1424–1433.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J. et al. (2016) A new view of the tree of life. *Nature Microbiology*, 1, 16048.

Jacot, D., Gizha, S., Orny, C., Fernandes, M., Tricoli, C., Marcelpoil, R. et al. (2024) Development and evaluation of an artificial intelligence for bacterial growth monitoring in clinical bacteriology. *Journal of Clinical Microbiology*, 62, e01651-23.

Lewis, W.H., Tahon, G., Geesink, P., Sousa, D.Z. & Ettema, T.J.G. (2021) Innovations to culturing the uncultured microbial majority. *Nature Reviews. Microbiology*, 19, 225–240.

Lockhart, S.R., Chowdhary, A. & Gold, J.A.W. (2023) The rapid emergence of antifungal-resistant human-pathogenic fungi. *Nature Reviews Microbiology*, 21, 818–832.

Ma, B., Lu, C., Wang, Y., Yu, J., Zhao, K., Xue, R. et al. (2023) A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. *Nature Communications*, 14, 7318.

Makrai, L., Fodróczy, B., Nagy, S.Á., Czeiszing, P., Csabai, I., Szita, G. et al. (2023) Annotated dataset for deep-learning-based bacterial colony detection. *Scientific Data*, 10, 497.

Newman, D.J. & Cragg, G.M. (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, 83, 770–803.

Nishimura, Y. & Yoshizawa, S. (2022) The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. *Scientific Data*, 9, 305.

Paoli, L., Ruscheweyh, H.-J., Forneris, C.C., Hubrich, F., Kautsar, S., Bhushan, A. et al. (2022) Biosynthetic potential of the global ocean microbiome. *Nature*, 607, 111–118.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. et al. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, 36, 996–1004.

Passi, A., Tibocha-Bonilla, J.D., Kumar, M., Tec-Campos, D., Zengler, K. & Zuniga, C. (2021) Genome-scale metabolic modeling enables in-depth understanding of big data. *Metabolites*, 12, 14.

Ren, Y., Chakraborty, T., Doijad, S., Falgenhauer, L., Falgenhauer, J., Goesmann, A. et al. (2022) Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics*, 38, 325–334.

Signoroni, A., Ferrari, A., Lombardi, S., Savardi, M., Fontana, S. & Culbreath, K. (2023) Hierarchical AI enables global interpretation of culture plates in the era of digital microbiology. *Nature Communications*, 14, 6874.

Staley, J.T. & Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, 39, 321–346.

Steen, A.D., Crits-Christoph, A., Carini, P., DeAngelis, K.M., Fierer, N., Lloyd, K.G. et al. (2019) High proportions of bacteria and archaea across most biomes remain uncultured. *The ISME Journal*, 13, 3126–3130.

Tran, A., Alby, K., Kerr, A., Jones, M. & Gilligan, P.H. (2015) Cost savings realized by implementation of routine microbiological identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology*, 53, 2473–2479.

Yo, C.H., Shen, Y.H., Hsu, W.T., Mekary, R.A., Chen, Z.R., Lee, W.J. et al. (2022) MALDI-TOF mass spectrometry rapid pathogen identification and outcomes of patients with bloodstream infection: a systematic review and meta-analysis. *Microbial Biotechnology*, 15, 2667–2682.

Zhang, B., Zhou, Z., Cao, W., Qi, X., Xu, C. & Wen, W.A. (2022) New few-shot learning method of bacterial colony counting based on the edge computing device. *Biology*, 11, 156.

How to cite this article: Sipkema, D. (2024) Improving the odds: Artificial intelligence and the great plate count anomaly. *Microbial Biotechnology*, 17, e70004. Available from: https://doi.org/10.1111/1751-7915.70004