


Gaming bacterial metabolism

Jacqueline A. Valeri, Aarti Krishnan & James J. Collins

 Check for updates

Many microorganisms remain understudied due to the challenges and complexities of culturing. An integrated lab automation and machine learning platform called BacterAI could be the future.

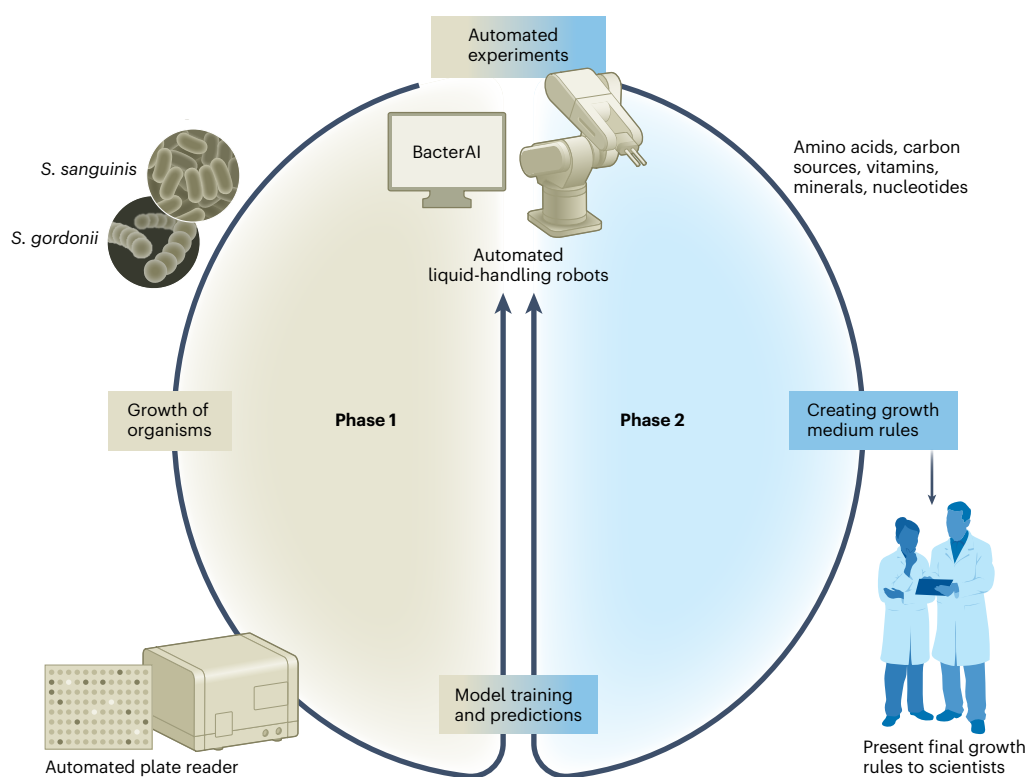
Advances in sequencing technology have greatly expanded our ability to identify and survey the staggeringly diverse and ever-expanding microbial world¹. Newly discovered species could be an untapped source of novel bioactive molecules such as antibiotics²; however, the potential of this vast microbial diversity has not been fully realized. Fewer than 2% of all microbial species are currently culturable³. This is due to a lack of knowledge about growth requirements^{3,4}, and the tedious, labour-intensive experiments that stand in the way of this essential knowledge.

The integration of lab automation and machine learning (ML) techniques could help with the culturing of understudied bacteria. These approaches have already been leveraged to improve the speed and throughput of experimentation. For example, artificial intelligence (AI)-guided colony selection based on colony morphology was used to isolate diverse strains from microbiome samples⁵, and efficient

metabolic engineering of biofuel-producing strains has been enabled by automated recommendations for which promoters should be manipulated⁶. A drawback of these approaches is that they can require large datasets to train predictive models. One exception to the data-hungry ML trend is reinforcement learning (RL), an ML technique that requires no prior training data. Reinforcement learning models learn through trial and error: positive outcomes are 'rewarded' with predefined rules, which serve as the input for these techniques. Reinforcement learning has been used to play and win games such as Chess and Go⁷, but it can also be co-opted for data-limited problems in microbiology.

Now writing in *Nature Microbiology*, Dama et al.⁸ describe an elegant RL-based platform, called BacterAI, that learns the growth requirements of bacteria with no prior knowledge. The robust RL pipeline is integrated with an automated experimental set-up to study bacterial growth requirements within a timeframe of two weeks. Researchers input a list of media ingredients, such as amino acids, to a computerized platform that provides liquid-handling instructions to lab equipment and human technicians. By treating the media composition as a 'game' with the goal of minimizing the number of ingredients, BacterAI's RL algorithm selects the most informative experiments to perform. Under the hood of the ML model, BacterAI is rewarded for removing as many ingredients as possible from media that still support bacterial growth. As BacterAI plays the 'game', hundreds of exploratory

Fig. 1 | Learning appropriate culture conditions as a game. BacterAI, an automated liquid-handling and experiment-planning platform, learns and reports growth rules of bacteria with no prior information. The platform operates in two phases: phase 1 (grey) automates iterative data collection with a RL method, similar to those used for playing games such as Chess; and phase 2 (blue) simplifies the data interpretation by using a separate algorithm to present logical growth rules to microbiologists.



experiments are executed daily, and the results are used to iteratively train an underlying neural network. The neural network then predicts bacterial growth given a defined media composition. Rather than a traditional ML approach where data collection and model building are isolated processes directed by a human expert, BacterAI automates both and can iterate until a predictive model is acquired.

The authors take BacterAI for a test drive on two bacteria: *Streptococcus gordonii* and a closely related species, *Streptococcus sanguinis*. Both are part of the oral microflora, a notoriously difficult environment for culturing. In less than two weeks, BacterAI predicted that *S. gordonii* and *S. sanguinis*, although closely related, each require a different combination of amino acids to support their growth (Fig. 1, phase 1). The generated models are more than 90% accurate, meaning they predict media as either growth supportive or growth unsupportive substantially better than a random coin flip.





Machine learning efforts for biology are often hampered by ‘black-box’ neural networks⁹ that are difficult to understand and contain thousands of parameters that render them uninterpretable by humans. BacterAI is unique in that, on completion of the automated experiments and collection of sufficient data, it automatically generates a human-readable interpretation of its findings (Fig. 1, phase 2). The output of BacterAI is a recipe for the precise, minimal combination of amino acids, carbohydrates, vitamins, minerals, and salts required to support bacterial growth. One cost of a readable output is accuracy, because a recipe with only a few ingredients will inherently be less accurate than the complex neural network that represents non-linear, fractional combinations of all ingredients. However, BacterAI makes up for these drawbacks with its user-friendly advantages.

Notably, BacterAI affords a high degree of automation in executing microbiology experiments. While a technician is still needed to upload the final growth measurements from the previous day and routinely inspect quality controls, BacterAI automatically generates a list of experiments to perform, produces a liquid-handling plan, and collects and processes all data. Soon, it may be possible that a single technician will be able to supervise an extensive set of experiments (for example, on hundreds of microbial species) running simultaneously using automation capabilities. In the future, the authors also foresee BacterAI’s deployment directly in chemically complex, species-rich environments such as the human gut or soil.

To harness BacterAI’s potential, other labs would require a similar lab set-up to use the open-source experiment-planning software, which could be a cost- and resource-prohibitive barrier. One way around this drawback would be the deployment of BacterAI’s integrated platform to a Cloud lab¹⁰ or virtual bench, where multiple research labs could participate and conduct experiments.

While BacterAI can learn from a blank slate, prior information improves model performance and the speed of experimentation. Dama et al. demonstrate that transfer learning, or the process of reusing information from one of BacterAI’s previous, related tasks, can speed up the RL game. For example, when an *S. gordonii* model is programmed with the final parameters of an *S. sanguinis* model, only 4 days of experiments are needed to achieve a model accuracy of 97.6%, compared with the 7 days required to obtain the initial *S. sanguinis* model. As genome sequencing and metabolic pathway reconstruction have become more cost effective and powerful over the past decade, future studies could incorporate additional layers of biological information to construct more accurate and context-aware models.

With minimal human time and labour, BacterAI can provide meaningful biological insight into understudied bacteria, catalysing further scientific studies. By automating the rapid identification of appropriate culture conditions for less-studied bacteria, platforms such as BacterAI can turn otherwise arduous research problems into easy wins.

Jacqueline A. Valeri ^{1,2,3,4,5}, Aarti Krishnan ^{1,2,3,5} & James J. Collins ^{1,2,3,4} 

¹Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ²Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA, USA. ³Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. ⁵These authors contributed equally: Jacqueline A. Valeri, Aarti Krishnan.

 e-mail: jimjc@mit.edu

Published online: 2 June 2023

References

1. Louca, S., Mazel, F., Doebeli, M. & Parfrey, L. W. *PLoS Biol.* **17**, e3000106 (2019).
2. Sekurova, O. N., Schneider, O. & Zotchev, S. B. *Microb. Biotechnol.* **12**, 828–844 (2019).
3. Wade, W. J. *R. Soc. Med.* **95**, 81–83 (2002).
4. Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z. & Ettema, T. J. G. *Nat. Rev. Microbiol.* **19**, 225–240 (2021).
5. Huang, Y. et al. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01674-2> (2023).
6. Radivojević, T., Costello, Z., Workman, K. & Garcia Martin, H. *Nat. Commun.* **11**, 4879 (2020).
7. Silver, D. et al. *Science* **362**, 1140–1144 (2018).
8. Dama, A. C. et al. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-023-01376-0> (2023).
9. Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yu, B. *Proc. Natl Acad. Sci. USA* **116**, 22071–22080 (2019).
10. Arnold, C. *Nature* **606**, 612–613 (2022).

Competing interests

The authors declare no competing interests.