

Harnessing machine learning in phage selection workflows

Paolo Savoca, Bioinformatician, NexaBiome



Treating bacterial infections is becoming less about finding a single, broad solution and more about understanding the complex biological systems at play. Against this backdrop, bacteriophages (phages), viruses that specifically infect and kill bacteria, are attracting renewed attention for their potential to tackle hard-to-treat infections effectively and sustainably.

“Phages offer a unique advantage over chemical antibiotics: they target harmful bacteria with remarkable specificity while leaving the beneficial microbiome intact.”

Phages offer a unique advantage over chemical antibiotics as they act with remarkable specificity to target only harmful bacteria while leaving the beneficial microbiome intact. This precision, whilst effective in treating infections and helping to combat antimicrobial resistance (AMR), comes with a slight challenge for therapeutics. Phages are abundant and each one differs, so a phage that's effective against a bacterial strain in one patient may be ineffective in another, even when the infection appears identical under the microscope. Selecting the right phage or combination of phages (phage cocktail) therefore requires a balancing act between bacterial diversity, host interactions, and environmental factors.

With the rise of high-throughput sequencing and computational biology, researchers are exploring how machine learning can support and accelerate phage selection. By integrating large-scale genomic and phenotypic data, algorithms can uncover patterns that are difficult or impossible to detect manually, helping to predict which phages will be most effective against a given infection. This approach offers the promise of more efficient workflows, better reproducibility, and, ultimately, faster delivery of personalised therapies to patients.

The challenge of selecting the right phage

Phage selection is no straightforward task. Even within a single bacterial species, individual strains can possess unique defence mechanisms that affect susceptibility. Other factors, such as the presence of commensal bacteria, the patient's immune response, and the local tissue environment can further complicate the process. Broad-spectrum phage cocktails can target multiple strains, though they are not universally effective, so some infections require an even more tailored approach.

Traditional selection methods rely on manual assessment and laboratory expertise. For example, researchers will test phages against panels of bacteria, evaluate growth dynamics, and measure virulence to establish how each candidate performs. Whilst important, this selection process can be time-consuming for experts whose efforts can be better placed elsewhere.

Building detailed phage profiles

To improve selection, laboratories like ours are generating comprehensive datasets that capture

both phenotypic behaviour and genetic composition. Standardised phenotyping involves assessing host range, tracking infection cycles, and measuring virulence. These experiments produce a unique 'fingerprint' for each phage, allowing comparisons across strains and studies. High-throughput sequencing complements these efforts by revealing the full genomic blueprint of both phages and their bacterial hosts. Platforms such as Oxford Nanopore Technology enable rapid sequencing at scale, providing structured data that can be consistently analysed.

When combined, phenotypic and genotypic data offer a multidimensional view of each phage and its potential activity. However, the sheer volume and complexity of this information often exceed human capacity for analysis. Subtle correlations between genetic features and biological behaviour may go unnoticed without computational support. This is where machine learning becomes invaluable.

Machine learning meets phage therapy

As we know, machine learning excels in situations involving large, complex datasets with phage selection being no exception. By training models on known phage-bacteria interactions, algorithms can learn which genetic and phenotypic features predict successful infection. During model development, data is carefully curated to ensure consistency, and predictions are validated against independent datasets to test reliability.

Early applications have already demonstrated that certain aspects of phage activity, such as host range and virulence, can be estimated from genomic features alone. While experimental validation remains essential, predictive models can save time and resources by focusing laboratory efforts on the most promising candidates. In other words, machine learning can act as a guide, helping researchers navigate an otherwise overwhelming landscape of thousands of potential phage-bacteria combinations much more efficiently.

Integrating predictive models into workflows

The most effective approach blends computational prediction with traditional laboratory methods. A typical workflow begins with sequencing phages and bacterial isolates, followed by bioinformatic processing

to assemble genomes, annotate genes, and assess functional and phylogenetic characteristics. Phenotypic data are then integrated, creating a structured dataset that serves as the foundation for model training.

Once trained, machine learning models can identify phage candidates predicted to be effective against specific bacterial strains. These predictions are subsequently tested *in vitro*, confirming their validity against strains not included in the training dataset. Over time, as more data are collected, the models become increasingly refined, improving predictive accuracy to inform how we design phage cocktails.

These approaches are not purely theoretical as collaborative efforts are already underway to implement integrated phage selection workflows at scale today. By combining sequencing, structured data management and predictive modelling, laboratories such as NexaBiome Life Sciences are beginning to incorporate data-driven tools that support more consistent and efficient decision making into their everyday practices.

Advantages of AI-guided phage selection

Imagine a clinical scenario in which a bacterial isolate is sequenced, an algorithm identifies several phages with the highest predicted efficacy, and laboratory testing confirms their activity within days rather than weeks. Such a workflow would not only save time but also enable more precise, personalised interventions against infections that currently resist standard treatment.

Thus, integrating machine learning into phage workflows offers several clear benefits.

- Firstly, efficiency is improved. Traditional phage testing can take weeks, particularly when dealing with rare or resistant strains. Predictive models help prioritise candidates, allowing laboratories to focus on those most likely to succeed.
- Secondly, consistency is enhanced. Standardised datasets and model-driven recommendations reduce variability between operators and laboratories, supporting reproducibility.
- Furthermore, these approaches support personalised therapies. By analysing a patient's microbiome alongside bacterial genomic data, models may one day identify optimal phage cocktails tailored to individual infections. Predictive algorithms also aid in cocktail design, revealing synergistic combinations that might not be obvious through manual assessment alone.

Perhaps most importantly, machine learning does not replace scientific expertise but amplifies it. By

- Models must be validated rigorously against independent datasets and real-world conditions to ensure reliability.
- Transparency and interpretability are essential. Clinicians and researchers need to understand why a model recommends a particular phage or cocktail, making explainable AI a priority.
- Generating and maintaining large, high-quality datasets requires substantial infrastructure, both computational and experimental.

As the field grows, ongoing investment in databases, sequencing capacity, and bioinformatic tools will be necessary to realise the full potential of machine learning in phage therapy.

A vision for the future

Looking ahead, the integration of machine learning into phage selection workflows has the potential to boost the field. Rather than relying solely on manual

Machine learning does not replace scientific expertise - it amplifies it, enabling researchers to make more informed decisions and accelerate effective therapies to patients.

highlighting patterns and connections in complex datasets, algorithms can allow researchers to make more informed decisions, ultimately accelerating the pace at which effective therapies reach patients.

Challenges and considerations

Despite the promise of AI-guided phage selection, several hurdles must be addressed:

- Standardising data collection across laboratories is critical; inconsistent sequencing, annotation, or phenotyping could undermine predictive accuracy.

testing, researchers could use predictive tools to rapidly narrow the range of candidate phages, validate them experimentally, and deliver tailored therapies more efficiently.

By combining biological expertise with computational prediction, emerging workflows keep scientific judgement at the centre of phage selection while benefiting from increasingly powerful data-driven tools. As sequencing technologies and machine learning models mature, this integrated approach offers a practical and reproducible way to improve how phages are selected, bringing faster and more targeted options to help treat patients in need.