Phage Therapy: Leveraging Machine Learning and Big Data techniques to Unveil the Evolution, Innovations, and Global Landscape in the Fight Against Antibiotic Resistance

Maxime Descartes Mbogning Fonkou  
York University  https://orcid.org/0000-0001-5581-3891

Jude Dzevela Kong  (jdkong@yorku.ca)  
York University  https://orcid.org/0000-0002-7557-5672

Method Article

Keywords:

Posted Date: January 19th, 2024

DOI: https://doi.org/10.21203/rs.3.pex-2537/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Phage therapy patent analysis, spanning from 1955 to 2023, reveals a substantial growth in the field, particularly since the early 2000s, with a peak in patent filings in 2021. The study identified key thematic areas, with “Compositions for Bacterial Infection Treatment” being the most prominent among the ten identified topics. Geographically, Asia, especially China, emerged as a leader in patent production, although U.S. patents showed greater citation impact, indicating a significant influence on global research. The analysis also highlighted diverse sectoral contributions with commercial and academic entities actively involved, and the importance of collaborative dynamics in driving innovation. This underscores a dynamic, evolving landscape in phage therapy, marked by regional disparities and a shift in research focus towards broad-spectrum medical applications and treatments for bacterial infections.

Introduction

The rapidly evolving field of microbiology has seen a surge in interest in bacteriophages (phages), particularly in light of the growing global issue of antibiotic resistance. These viruses, which are integral to phage therapy, represent a promising solution to the problem of resistant bacterial infections. The scope of phage therapy extends beyond the identification of new phage types to include the enhancement of delivery methods and diagnostic technologies. The importance of this research is highlighted by the increasing prevalence of ‘superbugs’ - bacteria that have developed resistance to drugs due to the overuse of antibiotics. In 2019, these superbugs were linked to over 1.27 million deaths globally.

The overuse of antibiotics not only targets harmful bacteria but also beneficial microbes, resulting in a reduction in gut microbiome diversity. This realization has led to a search for alternatives to antibiotics, with phage therapy emerging as a potential solution. Phage therapy employs bacteriophages to specifically target bacterial infections, providing accuracy and effectiveness without the wide-ranging impact of conventional antibiotics. Recent scientific breakthroughs have played a key role in renewing interest in phage therapy. Enhanced techniques for identifying and matching phages to specific bacteria have been introduced, including traditional methods like microfluidic PCR and PhageFISH, as well as computational methods that use machine learning algorithms. These technological advancements have facilitated the identification of effective phage-bacterial interactions, laying the groundwork for more targeted and efficient treatments.

In addition, the genetic modification of phages is being investigated to improve their effectiveness against specific bacterial strains. Techniques such as high-throughput diversification of receptor-binding proteins and synthetic biology approaches are being used to produce phages with superior antibacterial properties. The progress in machine learning is playing a crucial role in supporting these developments.

However, phage therapy faces considerable obstacles in terms of production and regulatory approval. Despite these challenges, clinical trials are underway, and there is increasing acknowledgment of the need for alternatives to traditional antibiotics, especially in the wake of the COVID-19 pandemic. This study aims to provide a comprehensive understanding of the evolution, current state, and future prospects of phage therapy. By analyzing patent data, we seek to illuminate the collaborative efforts, thematic trends, and regional disparities that define this field, offering insights into its potential to revolutionize healthcare and address critical medical challenges in the 21st century.
Equipment

Procedure

Methods

1. Data collection: In this study, we employed a systematic and exhaustive approach to collect data on patents relevant to phage therapy. Utilizing the comprehensive and publicly available database at Lens.org, we amassed a dataset of patent documents spanning from 1955 to 2023. The initial search, unconstrained by chronological limits, yielded a total of 5,753 patents. To ensure the focus on substantial and unique advancements in phage therapy, we adopted a patent family-based analysis. This strategy groups interconnected patents, minimizing redundancies and more accurately reflecting the field's innovation trajectory. Following this approach, we distilled the dataset to 2,365 distinct patent families for in-depth analysis. Our search strategy was meticulously crafted to capture patents pivotal to phage therapy, encompassing the discovery of novel phage types and breakthroughs in delivery methods. We employed a combination of specific keywords, such as 'phage therapy', 'bacteriophage', and 'treatment', targeting the title and abstract sections of the documents. The precise keyword combinations and search parameters are systematically outlined in Supplementary Table 1.

The decision to use Lens.org was driven by its comprehensive, up-to-date, and freely accessible patent database. Lens.org is recognized for its extensive coverage of global patent data, making it an ideal resource for our study. Furthermore, the platform's advanced search and analysis tools enabled us to efficiently collect and process a large amount of data. To further refine our analysis, we drew inspiration from the World Intellectual Property Organization (WIPO) classification system. We used a Python script to split the IPCR classifications into IPCR subclasses. This allowed us to analyze the patents at a more granular level, providing deeper insights into the specific areas of technology each patent pertains to. Working with patent families, as opposed to individual patents, provided several advantages. Firstly, it reduced the risk of double-counting the same invention reported in multiple patents. Secondly, it allowed us to capture the collective contribution of related patents, providing a more holistic view of the innovation landscape. Lastly, analyzing patent families helped us identify key players (inventors and applicants) and trends in the field of phage therapy.


Data Preprocessing: Utilizing the refined dataset of 2,365 patent families, we excluded patents lacking either titles or abstracts, resulting in a focused corpus of 2,348 patents. This corpus, formed from the titles and abstracts, served as the foundation for our topic modeling analysis. The preprocessing was conducted using Python 3.12.0 and NLTK 3.8.1. We removed all punctuation and used NLTK's comprehensive stopwords list for all languages present in our dataset. Lemmatization was performed using NLTK's WordNetLemmatizer to standardize word forms.

Vectorization: We utilized the Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer (TfidfVectorizer) to convert the preprocessed text into a matrix of TF-IDF features. This transformation was essential to reflect the importance of words in relation to the entire corpus, enhancing the topic modeling process.

Optimization of Topic Number: To determine the optimal number of topics, we experimented with a range of topic numbers from 5 to 40. We evaluated the coherence of each model using the 'metric_coherence_gensim' from the
'tmtoolkit.topicmod.evaluate' module. Our approach involved averaging coherence scores (mean) for the top 30 words (top_n=30) within the topics, enabling us to determine the optimal number of topics for the model. This evaluation metric helped us in identifying the number of topics that best captured the thematic structure of the dataset. The coherence scores indicated that a model with 10 topics provided the most meaningful and distinct thematic divisions (Supplementary Fig. 1).

**Latent Dirichlet Allocation (LDA) Modeling**: For the topic modeling, we employed the Latent Dirichlet Allocation (LDA) algorithm, specifically the 'LatentDirichletAllocation' function from the 'decomposition' module. Our LDA model parameters included 'n_components' equal to the determined optimal topic number, 'learning_method' set to 'online' for efficient processing of our large dataset, 'learning_offset' at 50 to control learning rate, 'n_jobs' set to -1 for parallel computing, and a 'random_state' for reproducibility. LDA is a generative probabilistic model that assigns topics to documents and words to topics, facilitating the discovery of latent thematic structures in large text corpora.

**Topic Analysis and Visualization**: After fitting the LDA model, we conducted a topic similarity analysis using the Jaccard Similarity measure and performed Hierarchical Clustering of Topics based on this similarity. This approach enabled us to understand the relationships and overlaps between the different topics. A similarity matrix was constructed and visualized using a heatmap, created with matplotlib and seaborn (Supplementary Fig. 2). Hierarchical clustering was implemented using the 'linkage' function from scipy with an "average" method, applied to the similarity matrix (Supplementary Fig. 3).

For each of the 10 topics, we extracted the top 50 words (Supplementary Table 2), providing a detailed overview of the thematic content. These top words were then visualized using word clouds, offering an intuitive representation of the topic composition (Supplementary Fig. 4). Additionally, we tabulated the topic distributions in a pandas DataFrame for a more structured analysis of the thematic landscape (Supplementary Fig. 5).

3. Disease Entity Extraction Process:
In our study, we systematically extracted disease entities from a corpus of 2,348 phage therapy patent families, using advanced natural language processing techniques.

**Data Preprocessing**: Initially, we preprocessed the patent data by removing punctuation and creating multiple case variations of the titles and abstracts. This step was crucial to accommodate potential case sensitivity in the subsequent entity recognition process. We then concatenated these text transformations to form a comprehensive dataset for analysis.

**Disease Name Extraction Using SciSpacy**: Leveraging the capabilities of SciSpacy, specifically the 'en_ner_bc5cdr_md' (a spaCy NER model trained on the BC5CDR corpus) model version 0.5.3, we employed a custom Python function, `extract_diseases`, to accurately identify and extract disease entities from the concatenated text. This function processed each patent's text through the NER model, capturing entities labeled as 'DISEASE'. The extracted disease names were stored in a new column, 'Diseases', in our dataset. This meticulous approach enabled us to efficiently and accurately distill relevant disease entities from the extensive patent corpus, providing a foundational dataset for our in-depth analysis of trends and focuses in phage therapy research.

**Data Refinement**: Following the extraction of disease entities from the phage therapy patent data, we embarked on a rigorous cleaning and classification process. The extracted disease names, now compiled in an Excel file, were first subjected to a thorough review to remove any terms not related to diseases. This initial cleansing was crucial
to ensure the relevance and accuracy of our dataset. We then utilized OpenRefine\textsuperscript{9}, a powerful tool for working with messy data, to further refine and standardize the list of disease names. This step allowed us to correct inconsistencies, merge duplicates, and format the data for uniformity.

**Data Standardization:** The final and most significant phase of our methodology involved classifying each disease according to the International Classification of Diseases 11th Revision (ICD-11) published by the World Health Organization\textsuperscript{10}. This classification provided a standardized framework to categorize the diseases, facilitating a more organized and meaningful analysis. By aligning our dataset with ICD-11, we ensured that our study's findings were both globally relevant and comparable, setting a solid foundation for our subsequent analysis of phage therapy trends and applications.

4. Keywords Analysis:

**Bigram Keywords Extraction:** In our study, we conducted a comprehensive analysis of bigram keywords to uncover prevalent themes and patterns within the phage therapy patent dataset. Utilizing Python's Natural Language Toolkit (NLTK)\textsuperscript{4}, we implemented a methodical approach to extract meaningful bigram keywords from the abstracts of the patents. The process began with the tokenization of text in each abstract using NLTK's `word_tokenize` function. Subsequently, we filtered out predefined blacklisted words — specific terms deemed irrelevant or too common for our analysis — to refine the focus of our keyword extraction. The cleaned tokenized text was then processed to generate bigrams, pairs of consecutive words, using the `ngrams` function from NLTK. These bigrams were carefully compiled for each abstract and stored in a new column, 'bigrams', in our dataset. This method allowed us to identify and analyze frequently occurring word pairs, providing valuable insights into the focal points and trends within the corpus of phage therapy patents.

**Visualization of Bigram Keywords:** To effectively visualize and interpret the patterns in our bigram keyword analysis, we employed VOSviewer\textsuperscript{11}, a tool renowned for its capability to create and display bibliometric networks. Specifically, we focused on mapping the bigram keywords extracted from the phage therapy patent abstracts. In our approach, we set a threshold: only bigram keywords that appeared a minimum of 10 times in the dataset were included in the visualization. This criterion was chosen to ensure that our analysis concentrated on the most prevalent and significant themes within the corpus. Upon importing the relevant data into VOSviewer, we generated a network map that visually represented the relationships and clustering of these keywords. The map provided an intuitive overview of the interconnectedness of concepts, highlighting the dominant topics and potential areas of emerging interest in the field of phage therapy. This visual representation was instrumental in identifying key focal points and trends in the patent data, offering a clear and comprehensive perspective on the landscape of phage therapy research.

5. Cleaning of Inventor and Applicant Names:

**Objective and Rationale:** Our study necessitated the accurate and consistent representation of inventor and applicant names for reliable analysis. The accuracy in naming data is crucial, especially when evaluating individual contributions and affiliations within the phage therapy patent landscape.

**Data Export and Initial Processing:** The first step involved exporting the inventor and applicant names from our dataset into an Excel file. This file formed the basis for our cleaning and standardization process, ensuring that the names were prepared for detailed analysis.

**Use of OpenRefine for Cleaning:** To clean and standardize the names, we utilized OpenRefine, a tool particularly
effective in handling and refining messy data. OpenRefine's capabilities in data transformation and cleaning are well-suited for the intricate nature of our dataset, which included names from diverse sources and formats.

**Cleaning Process in OpenRefine:** Within OpenRefine, our cleaning process encompassed several key operations:

- **Trimming Whitespace:** We removed any leading or trailing spaces in the names. These spaces often lead to inconsistencies in data analysis and aggregation.

- **Standardizing Case:** All names were converted to a consistent case format (such as title case), ensuring uniformity across the dataset.

- **Removing Duplicates:** We identified and merged duplicate records that referred to the same individual or entity. This step was vital to prevent skewed analysis due to repeated entries.

- **Correcting Typos:** Wherever possible, obvious typographical errors in the names were manually corrected. This step required careful attention to detail to maintain the integrity of the data.

**Manual Review and Verification:** Following the automated cleaning process, we conducted a manual review of a random sample of the cleaned data. This verification was essential to confirm the effectiveness of our cleaning efforts and to ensure the data's readiness for subsequent analytical procedures.

6. **Analysis of Key Players:** Framework for Assessing Individual Impact and Participation Dynamics

In our comprehensive analysis of the phage therapy patent landscape, we employed a multi-faceted approach to evaluate individual contributions and impacts. Key metrics included Patent Count and Total Citations, Centrality Measures, Duration of Activity, and Patterns of Participation. This methodology ensures a robust and nuanced understanding of the roles and influences of entities in phage therapy research.

**Patent Count and Total Citations:** The objective was to quantify each entity's contribution and impact within the phage therapy patent landscape. The patent count serves as an indicator of productivity, while the total number of citations reflects the influence and recognition of an entity's work. To assess the patent count and total citations, we developed a Python function in the pandas library to process the dataset, counting patents and summing citations for each entity, be it an applicant or inventor. This approach provided insights into the quantity and impact of contributions made by each entity.

**Centrality Measures in Network Analysis:** The objective was to assess the position and influence of entities within the phage therapy research network. Centrality measures provide insights into an entity's prominence, connectivity, and mediating roles within the network. Employing NetworkX, we constructed a network graph from the dataset. Each entity pair within a patent formed an edge, enabling the computation of centrality measures (degree, betweenness) for each node. The outcomes were collated in a DataFrame, associating each entity with its centrality scores.

**Duration of Activity:** The aim was to determine the temporal engagement of entities in phage therapy research. The duration of activity highlights longstanding contributors and emerging entities in the field. A custom Python function was designed to identify the first and last years of patent applications for each entity, thus calculating the duration of their involvement in the field.

**Patterns of Participation:** The aim was to categorize entities based on their participation over time. Following the typology introduced by Price and Gürsey, we developed a Python-based classification system to categorize entities as Transients, Newcomers, Terminators, or Continuants:
- Transients: Authors publishing in a specific year but neither before nor after. This category helps identify entities with brief, one-time involvement in the field.

- Newcomers: Authors who begin publishing in the given year and continue thereafter, indicating new entrants to the field.

- Terminators: Authors who have published before and in the given year, but not subsequently, highlighting entities that have ceased their involvement.

- Continuants: Authors with ongoing participation, publishing before, in, and after the given year, reflecting a sustained commitment to the field.

The function iterates through the dataset, tracking each entity's participation across different periods. Based on their activity patterns, entities are classified into the aforementioned categories, offering insights into the degree of continuous participation and the incorporation of new researchers over time.

7. Classification of Applicants:
The aim was to categorize the applicants in our dataset as individual, academic, government, or commercial entities. This classification provides insights into the diversity of entities contributing to phage therapy patents and helps in understanding the different sectors driving innovation in this field. We adopted a keyword-based classification approach similar to the one used in a previous study of Mohajel et al.\textsuperscript{15}.

- \textbf{Academic Entities:} Identified through keywords such as 'univ', 'college', 'acad', 'inst', 'school', 'foun', 'center', 'hospital', and 'res'.

- \textbf{Commercial Entities:} Marked by terms like 'inc', 'Ltd', 'corp', 'llc', 'gmbh', 'as', 'ag', 'company', 'lab', 'pharma', 'therapeutics', 'pty', 'plc', and related variations.

- \textbf{Governmental Entities:} Detected using keywords such as 'gov', 'army', 'minister', 'federal', 'agency', and 'government'.

- \textbf{Individuals:} Classified by the absence of the above keywords in the applicant field.

For applicants that could not be classified through these keywords, we conducted internet searches to finalize their categorization.

\textbf{Data Processing:} Using Python, we iteratively scanned the 'applicant' field in our dataset for these keywords. The presence of specific keywords led to the categorization of an applicant into one of the predefined groups.

\textbf{Accuracy and Verification:} To ensure the accuracy of our classification, a subset of the data was manually reviewed. Additionally, for ambiguous cases, supplementary internet research was conducted to confirm the applicant's category.

8. Innovation Hotspot Analysis by Jurisdiction:
\textbf{Jurisdiction and Applicant Country Data:} To identify and analyze innovation hotspots in phage therapy research, we focused on the 'jurisdiction data' and 'applicants countries' columns in our dataset. These columns provided
crucial geographical insights, reflecting the global distribution of phage therapy patent activities.

**Data Transformation and Enrichment:**

- **Jurisdiction Initials to Names:**
  - Initially, we converted jurisdiction initials into their full names. This transformation was essential for clarity and ease of understanding, allowing us to accurately represent each jurisdiction in subsequent analyses.

- **Adding Regional Information:**
  - We further enriched our dataset by adding a column for the region associated with each jurisdiction and applicant country. This step involved mapping countries to their respective geographical regions, thereby facilitating a more comprehensive analysis of the data.

- **Country and Region Identification for Applicants:**
  - For all previously characterized applicants, we determined their respective countries and regions. This identification process was critical for understanding the geographical diversity of entities contributing to phage therapy patents and for pinpointing regions with significant research activity.

**Purpose of the Analysis:** This geographical analysis aimed to uncover the innovation hotspots in phage therapy research. By examining the jurisdictions and countries of the applicants, we gained insights into the region's leading in phage therapy innovations and identified potential areas of concentrated research and development.

**Ensuring Comprehensive Coverage:** This approach ensured that our analysis encompassed a global perspective, highlighting not only the countries with high patent activity but also the regions emerging as new centers of innovation in phage therapy research.

9. **Social Network Analysis of Inventors and Applicants in Phage Therapy Patents:**

**Objective and Rationale:** Our study embarked on a Social Network Analysis (SNA) to unravel the collaboration patterns and network dynamics among inventors and applicants in phage therapy. This analysis aimed to dissect both the structure and composition of the network and its cohesive properties.

**Network Structure and Composition:**

- **Purpose:** The primary objective was to analyze the vertices (entities such as inventors and applicants) and links (their relationships) within the network.

- **Implementation:** Utilizing a custom Python function create_graph, we constructed a graph representing the network. This process involved iterating through the dataset, identifying entities within the 'entity_column', and forming edges between them for each pairwise combination, with weights assigned based on collaboration frequency. Additionally, the calculate_edge_weights function was employed to determine the weight of these edges, indicative of the strength and frequency of collaborative interactions.

**Network Cohesion:**

- **Objective:** A key aspect of our analysis was assessing network density and other indicators of cohesion.
· **Method:** We utilized the `calculate_graph_metrics` function to compute various network metrics. These included the number of vertices and links, network density, and clustering coefficient. These metrics quantitatively captured the network's cohesion, reflecting the degree of interconnectedness and collaboration among the entities.

**Visualization with VOSviewer:**

· **Tool for Representation:** For a comprehensive visual interpretation of the network data, we employed VOSviewer. This tool facilitated the graphical depiction of the network, highlighting key players and illustrating the intensity of their collaborative ties. This visual representation was instrumental in providing intuitive insights into the network's overall structure and key collaboration hubs.

**Robustness and Accuracy:**

· **Ensuring Rigorous Analysis:** Our methodology was underpinned by meticulous data processing in pandas and network construction and analysis in NetworkX. These tools were pivotal in managing the intricacies of social network data, ensuring a robust and accurate analysis.

**Troubleshooting**

**Time Taken**

**Anticipated Results**

**References**


4. nltk: Natural Language Toolkit.


7. scispacy: A full SpaCy pipeline and models for scientific/biomedical documents.


**Acknowledgements**

I would like to express my deepest appreciation to Zahra Movahedi Nia for her invaluable contributions to this research. Her expertise and guidance in topic modeling and identification of the optimal number of topics have been instrumental in the success of this work. Her insights have not only enriched this research but also deepened my understanding of these complex areas. Thank you, Zahra, for your significant role in this endeavor.

**Figures**

![LDA Model Metrics](image)

**Figure 1**

**Supplementary Fig. 1 | Evaluation of Topic Coherence in LDA Models for Phage Therapy Research.**

This figure presents a line graph showing the coherence scores of Latent Dirichlet Allocation (LDA) models across different numbers of topics in phage therapy research. The x-axis displays the number of topics evaluated in each
LDA model, while the y-axis indicates the coherence scores associated with these models.

Coherence scores measure the semantic similarity of high-scoring words within each topic, providing an indication of the interpretability and relevance of the topics generated. Higher scores suggest that the topics are more meaningful and consist of words that are contextually related.

The graph demonstrates that the model with 10 topics achieved the highest coherence score, indicating that this number of topics optimally balances the comprehensibility and thematic detail in the LDA analysis. This optimal point guided the selection of 10 topics as the focus for further detailed analysis of the phage therapy research landscape, ensuring meaningful and interpretable results.

![Jaccard Similarity Matrix of LDA-Derived Topics in Phage Therapy Research](image)

**Figure 2**

**Supplementary Fig. 2 | Jaccard Similarity Matrix of LDA-Derived Topics in Phage Therapy Research.**

This figure presents a heatmap of the Jaccard similarity matrix, depicting the pairwise similarities between topics identified in phage therapy research through Latent Dirichlet Allocation (LDA). Each cell in the matrix represents the Jaccard similarity score between two topics, with the score ranging from 0 to 1. Higher scores (warmer colors) indicate greater similarity between topic pairs, suggesting shared thematic content or closely related research areas.

The matrix is symmetric, with both axes representing Topic IDs corresponding to different themes identified within the phage therapy field. The annotations within the cells provide exact similarity scores, allowing for precise interpretation of topic relationships. This visualization highlights the interconnectedness of topics, revealing which areas of phage therapy research are thematically aligned and potentially overlapping.
Figure 3

Supplementary Fig. 3 | Hierarchical Clustering of Phage Therapy Research Topics Based on Jaccard Similarity.

This figure presents a dendrogram resulting from a hierarchical clustering analysis of topics identified in phage therapy research via Latent Dirichlet Allocation (LDA). The analysis employs Jaccard similarity to measure the closeness of topics based on their word distributions. Each branch of the dendrogram represents a topic, with the linkage points illustrating the degree of similarity between them.

The vertical axis denotes the 'distance' or dissimilarity between clusters of topics, with lower values indicating greater similarity. The horizontal axis lists the Topic IDs, corresponding to different thematic areas within phage therapy research. The structure of the dendrogram allows for the visualization of topic groupings, showcasing how individual topics are related to each other and form larger clusters based on thematic similarities.

Figure 4

Supplementary Fig. 4 | Word Clouds Representing Key Themes in Phage Therapy Research Topics.
This figure displays a series of word clouds generated from Latent Dirichlet Allocation (LDA) analysis, each representing one of the identified topics in phage therapy research. The word clouds are based on the frequency of terms within each topic, with the size of each term in the cloud corresponding to its relative prevalence. Each word cloud (labeled as ‘Topic X’) visually encapsulates the essence of the respective topic, highlighting the most prominent and recurring terms identified in the LDA analysis. This visualization provides an intuitive understanding of the key themes and concepts that dominate each topic, offering a snapshot of the thematic focus areas in phage therapy research. The series of word clouds together portray the diversity and range of research interests in the field, from clinical applications to technological advancements.

Figure 5

Supplementary Fig. 5 | Temporal Evolution of Topics in Phage Therapy Research (1955-2023).

This figure presents a series of line graphs, each corresponding to a specific topic identified in phage therapy research through Latent Dirichlet Allocation (LDA). The subfigures, organized in a grid format, trace the frequency of each topic over time, from 1955 to 2023. Each subplot represents the temporal evolution of one topic, with the x-axis indicating the year and the y-axis showing the frequency of the topic in the research literature.

The evolution curves in each subplot provide insights into the historical and current prominence of specific topics within the field of phage therapy. Rising and falling trends across different years highlight how certain topics have gained or lost research interest over time. The collective presentation of these plots illustrates the dynamic nature of research focus in phage therapy, revealing shifts in scientific priorities and emerging areas of interest.

By charting the trajectory of each topic, this figure offers a comprehensive view of the evolving landscape of phage therapy research, underlining the temporal patterns and shifts in thematic focus over nearly seven decades.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesPhageTherapyPatents.xlsx