

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,200

Open access books available

168,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Exploring the Knowledge Landscape of *Escherichia coli* Research: A Scientometric Overview

Andrej Kastrin and Marjanca Starčič Erjavec

Abstract

Escherichia coli (*E. coli*) has the hallmark of being the most extensively studied organism. This is shown by the thousands of articles published since its discovery by T. Escherich in 1885. On the other hand, very little is known about the intellectual landscape in *E. coli* research. For example, how the trend of publications on *E. coli* has evolved over time and which scientific topics have been the focus of interest for researchers. In this chapter, we present the results of a large-scale scientometric analysis of about 100,000 bibliographic records from PubMed over the period 1981–2021. To examine the evolution of research topics over time, we divided the dataset into four intervals of equal width. We created co-occurrence networks from keywords indexed in the Medical Subject Headings vocabulary and systematically examined the structure and evolution of scientific knowledge about *E. coli*. The extracted research topics were visualized in strategic diagrams and qualitatively characterized in terms of their maturity and cohesion.

Keywords: *Escherichia coli*, scientometric analysis, knowledge mapping, keyword analysis, co-word analysis

1. Introduction

Escherichia coli (*E. coli*) is widely known and one of the most studied microorganisms in the life sciences. Since its discovery in 1885 by Theodor Escherich [1], *E. coli* has been the subject of intense research. *E. coli* is believed to be one of, if not the, most important organisms in studies aimed at discovering fundamental biological principles and mechanisms, as well as biology field-specific research methods and techniques. However, very little is known about the knowledge landscape in *E. coli* research. In particular, how published empirical findings on *E. coli* have evolved over time and what scientific questions have been the focus of researchers' interest? Answering these questions motivates the present work.

Scientific achievements are traditionally published in the form of a journal article, a paper in a conference volume, or a book chapter. To illustrate the effect of the accumulation of research results, we show in **Figure 1** the annual number of publications in the period 1940–2021 with *E. coli* as the main research topic. Although it is questionable whether the number of publications is directly related to

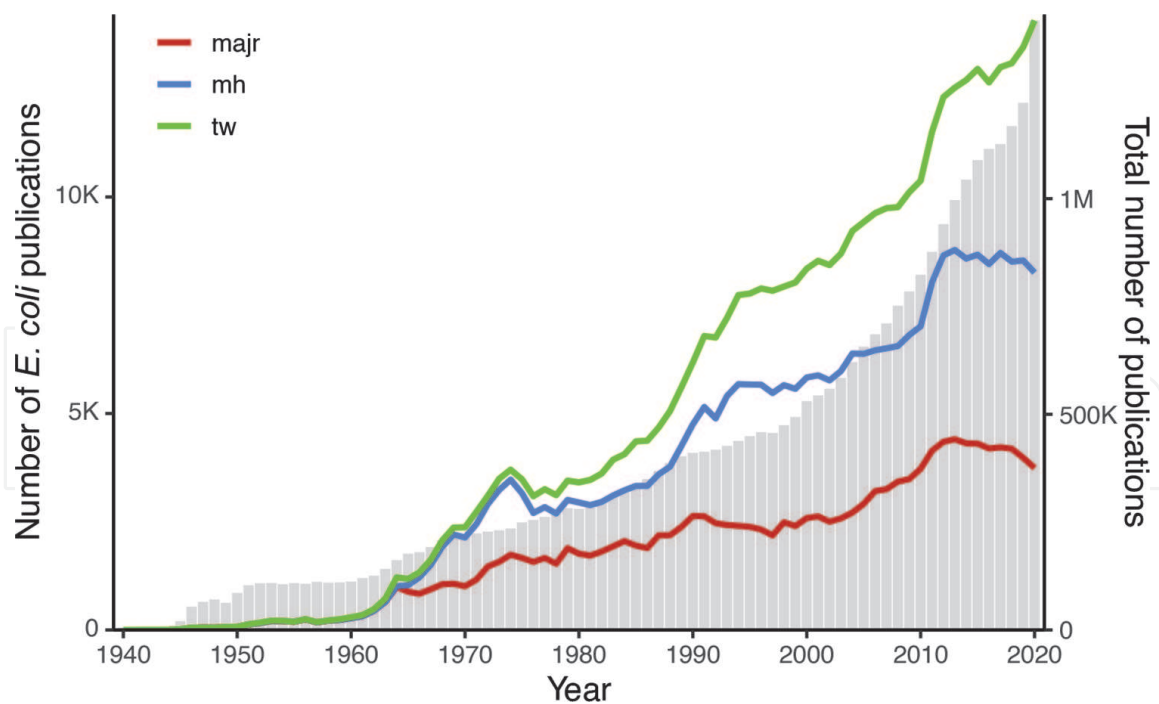


Figure 1.

The annual distribution of PubMed publications on *E. coli*. The different colored lines represent the PubMed field tag used to retrieve *E. coli* publications: MeSH heading, which is a major topic of an article (red), MeSH term in general (blue), and all words in the title, abstract, and other relevant fields (green). The gray bars of the histogram, quantified by the second y-axis, show the total number of publications in PubMed in a given year.

the amount of knowledge in a particular scientific field, we can at least use it as a proxy indicator of research activity in a particular area of interest.

However, the body of scientific literature is growing at an unimaginable pace [2, 3]. PubMed, for example, a leading bibliographic database in the life sciences, has indexed an average of 3800 new articles daily over the past 5 years [4]. Manual review of such a large body of new literature, even in a very specialized field of research, is therefore not only time-consuming, but virtually impossible. Fortunately, we can draw on a rich toolkit of automated methods and techniques offered by a modern scientometric approach to deepen our understanding of science itself [5, 6].

The main objective of this work is to examine the *E. coli* literature from an evolutionary point of view using a data-driven approach. Specifically, the aims are twofold: (i) to provide insights into research topics based on a quantitative text-mining analysis of a large number of *E. coli* papers from 1981 to 2021 indexed in PubMed, and (ii) to highlight the evolution of scientific knowledge in the field from a domain expert's perspective.

2. Background and related work

2.1 Science mapping

If we may paraphrase Ebbinghaus' famous statement, we could say that the field of scientometrics has a long past but a short history [7]. The study of scientific knowledge itself has been the subject of many famous works with great impact on the research community, including contributions by Lotka [8], Zipf [9], Price [10], Merton [11], Garfield [12], and later by Borner [13], Uzzi [14], Wang [15], Clauset [5], and Milojević [16].

Formally, scientometrics is an umbrella term for a set of approaches that aim to describe and understand the (relational) structure between researchers, their institutions, and scientific knowledge—operationalized through ideas, concepts, citations, and keywords—in order to identify and track the driving mechanisms of science [6]. One of these approaches, commonly used in the literature, is also a science map. A science map is a spatial and/or temporal representation of individual authors, their research groups, or the knowledge concepts they have written about [17].

The seminal studies that addressed the organization of scientific knowledge were driven by the study of citation networks, a type of analysis in which we seek to understand common patterns of citation links among articles in a collection of scientific literature [18]. The authors discovered several important structural features, including the famous small-world phenomenon [19, 20], the rich-get-richer mechanism [21], and the hierarchical organization of scientific knowledge [22]. We refer the reader interested in further details to the recently published monograph by Wang and Barabási [23].

Other authors argue that scientific knowledge could be represented more realistically with keywords as basic knowledge elements [24]. Keywords and key phrases are typically extracted from the title and/or abstract of each article using natural language processing tools, or parsed from a list of descriptors already provided by the authors. To overcome the challenges of normalizing keywords, many authors use controlled vocabularies such as Medical Subject Headings (MeSH) in the life sciences [25] or Mathematics Subject Classification in mathematics [26].

2.2 Co-word analysis

Co-word analysis is an improved version of pure keyword-based co-occurrence analysis. By combining various theoretical concepts from graph theory, co-word analysis allows a simple but efficient reduction of a massive network of co-occurring keywords to a higher-level network of clustered keywords. First described by Callon et al. [27] in the 1980s, co-word analysis is a powerful method for mapping the detailed intellectual structure of unstructured text data [17]. The method has been used in a variety of scientific fields, including microbiology [28]. However, to our knowledge, it has not yet been applied to elucidate the intellectual structure of knowledge about *E. coli*.

Technically, the input for co-word analysis is a network of keywords, as described in Section 2.2. In the next step, we use a type of cluster analysis—often referred to as community detection in the language of complex networks—to partition nodes into a smaller number of communities based on the similarity of their wiring patterns. The clustering algorithm is optimized with the objective of maximizing both homogeneity within communities and heterogeneity between communities. Finally, we create a strategic diagram to uncover and explore interesting patterns within the detected community structure based on a set of predefined heuristic rules [29].

3. Methods

In this study, we used a scientometric methodology to capture the structural and dynamic features of the knowledge landscape in *E. coli* research. In Section 3.1, we explain the details of compiling the dataset from the PubMed database. Then, in Sections 3.2 and 3.3, we explain the procedure for extracting keywords and creating

co-word networks. Finally, in Section 3.4, we describe a method for identifying broad research topics and interpreting them.

3.1 Data collection

The literature collection was created using an automated procedure from PubMed distribution. We retrieved all PubMed records indexed with the major MeSH descriptor “*Escherichia coli*” and restricted to the English language. Full bibliographic records were downloaded via the PubMed API and stored locally in XML format. To restrict a PubMed search result by the specified date range, we set the “datatype” parameter in an API call to “pdat”. The last query update was performed on October 1, 2022.

3.2 Keyword extraction

The co-word analysis presented here is based on MeSH terms to overcome problems with the normalization of plain keywords, as described previously in Section 2.1.

Each PubMed record is manually annotated by human indexers at the National Library of Medicine using the MeSH terms. MeSH is a controlled vocabulary consisting of biomedical terms at different levels of granularity. There are several types of MeSH terms, two of which are important for further understanding of the present work: Main MeSH headings (or descriptors) and MeSH subheadings (or qualifiers). Descriptors are the main elements of the thesaurus and denote the main topic of the paper. For example, the MeSH descriptors for a paper dealing with adherent-invasive *E. coli* pathovar strains in the context of Crohn’s disease might be “Bacterial Adhesion”, “Crohn’s Disease”, and “*Escherichia coli*”. Qualifiers are optionally assigned to the descriptors to express a particular aspect of the knowledge concept.

For further processing, we extracted all pairs of mesh heading/subheading terms along with the publication date of each bibliographic record.

3.3 Co-word network

In Section 2.2, we introduced the notion of co-word analysis, which aims to detect communities of keywords that frequently occur in conceptually similar articles. Formally, we first created a co-occurrence network based on the MeSH term lists from all retrieved documents. A node in the co-occurrence matrix refers to a particular MeSH heading/MeSH subheading pair, and a relationship between two nodes is established when both headings occur together in a particular document. In the following paragraphs, the phrase “MeSH heading/MeSH subheading” is referred to as “term” or simply “heading”.

In the next step, the co-occurrence network was weighted according to the number of observed pairs of MeSH headings. For example, if MeSH heading i and MeSH heading j appear together in 100 papers, the weight of their co-occurrence was set to 100. Finally, the raw edge weights were normalized to account for the unbalanced number of MeSH headings in the papers. For normalization, we used an association measure defined as

$$e_{ij} = \frac{c_{ij}^2}{c_i c_j}, \quad (1)$$

where c_{ij} is the number of co-occurrences of headings i and j [29]. Also, c_i and c_j are the numbers of occurrences of MeSH headings i and j , respectively. The normalized value is zero if the MeSH heading pair is not associated at all, and is equal to one if a given pair occurs together in each paper.

3.4 Identification of research topics

On a prepared co-occurrence network, we ran Louvain's community detection algorithm to identify clusters of homogeneous MeSH headings [30]. Each of the detected clusters groups together several contextually similar MeSH headings and plays the role of a research topic.

The interpretation of the research topics followed the procedure described by Callon et al. [27]. We calculated two measures, centrality and density, to represent a particular research topic in a two-dimensional plot called a strategic diagram. Centrality represents the relatedness of an observed research topic to other topics in a strategic diagram. The stronger this relatedness is, the more central the topic is in the observed network. In practice, we interpret centrality as the strength of a research topic in the entire scientific domain. Formally, the centrality of a topic is defined as

$$c = 10 \times \sum e_{kh}, \quad (2)$$

where k is a MeSH heading from the observed topic, h is a MeSH heading belonging to other topics, and e_{kh} is the normalized co-occurrence frequency of the pair of MeSH headings k and h according to Eq. (1).

Density, on the other hand, represents internal cohesion, i.e., how strongly an observed research topic is conceptually developed. Density is formally defined as

$$d = 100 \times \left(\frac{\sum e_{ij}}{w} \right), \quad (3)$$

where i and j are MeSH headings associated with a cluster, and e_{ij} is the normalized frequency of co-occurrence of the two MeSH headings. The w in denominator represents the total number of MeSH headings in a given research topic.

Finally, considering centrality and density, we created a strategic diagram to represent the structural landscape of knowledge. The diagram is centered by the median of the two axis values and divides the plot area into four quadrants characterized by different types of research topics [29]. A particular topic can be assigned a unique qualitative description based on its position in the diagram as follows:

1. The *motor* research topics in quadrant I are characterized by high centrality and high density. These topics are well defined, mature, and have been worked on over a long period of time by already well-developed research groups.
2. *Niche* topics in quadrant II have low centrality but high density. Such research topics are very homogeneous (i.e., they are characterized by strong internal linkages). However, they have weak external linkages and are therefore not well connected to other research topics.
3. *Emerging or declining* topics in quadrant III are defined by both low centrality and low density and refer to either new (i.e., emerging) or declining research topics.
4. *Basic* research topics in quadrant IV are characterized by high centrality but low density and thus combine transversal and very general research topics.

Although such topics are important to a particular research community, they are not well-developed.

4. Results

A total of 98,085 unique bibliographic records for the period 1981–2021 on the topic of *E. coli* were retrieved and considered for further processing. We identified 13,408 unique MeSH descriptors, which in turn yielded 54,663 unique combinations with MeSH qualifiers.

In the next sections, we present the descriptive results for each subperiod analyzed. In addition to the strategic chart, we have also included a table with a list of MeSH terms that define a particular research topic, as well as a brief qualitative description of the topics.

4.1 Period 1981–1990

For the time period 1981–1990 the performed survey resulted in 20,739 documents and the MeSH-based co-word analysis resulted in 10 different clusters. The strategic diagram is shown in **Figure 2**. The description of the identified research topics is summarized in **Table 1**.

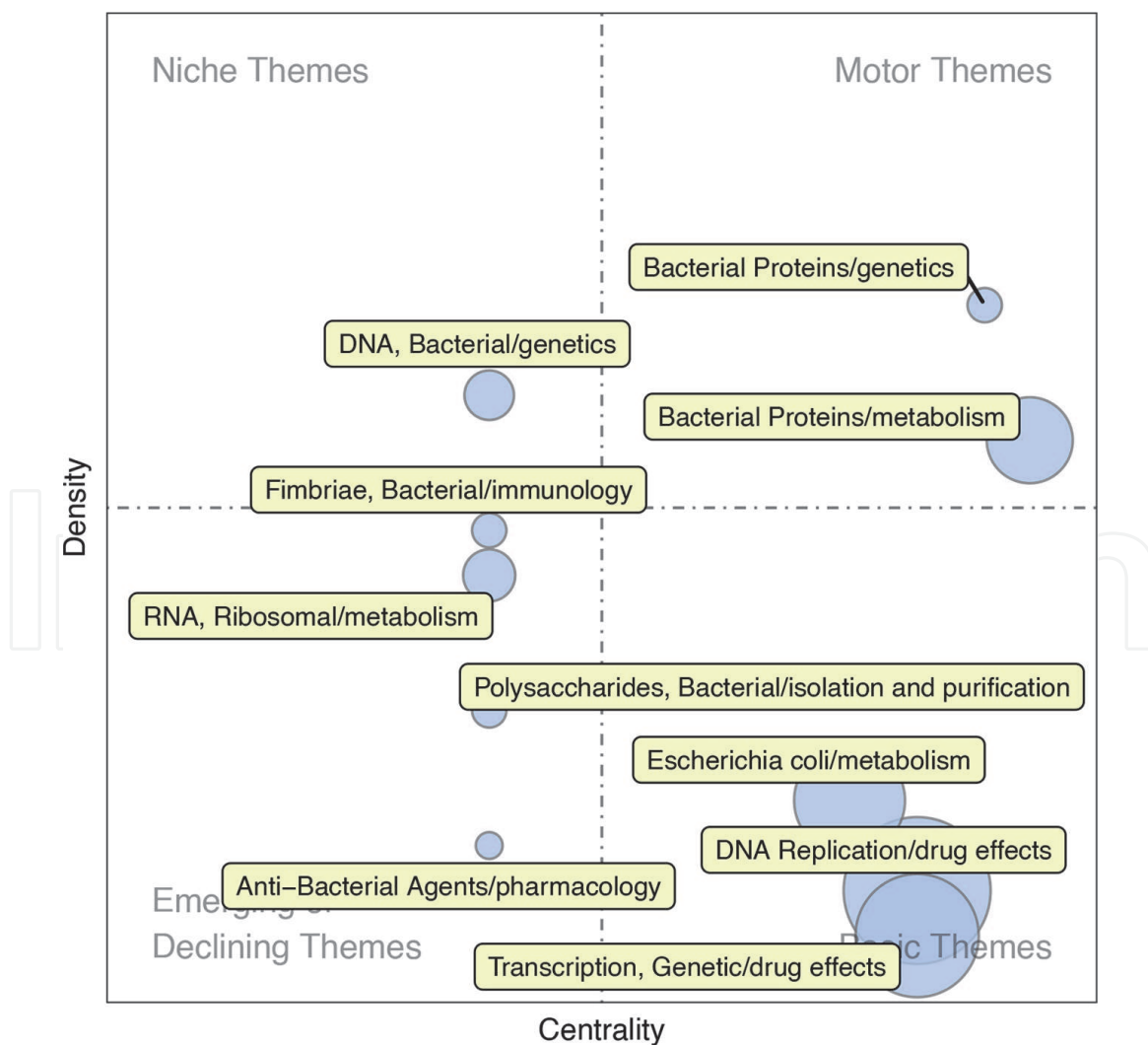


Figure 2.

Strategic diagram of the period 1981–1990. Each research topic is represented by a node and labeled with the most frequent pair MeSH heading/MeSH subheading. The size of the node is proportional to the number of MeSH heading/MeSH subheading pairs in each cluster.

ID	Size	MeSH headings
1	127	DNA Replication/drug effects, DNA Repair/radiation effects, Plasmids/drug effects, DNA Repair/drug effects, Genes, Bacterial/radiation effects, DNA Replication/radiation effects, Genes/radiation effects, Plasmids/radiation effects, Recombination, Genetic/radiation effects, Transformation, Bacterial/drug effects
2	81	Transcription, Genetic/drug effects, Genes, Bacterial/drug effects, Gene Expression Regulation/drug effects, Genes/drug effects, Protein Biosynthesis/drug effects, Operon/drug effects, Promoter Regions, Genetic/drug effects, Gene Expression Regulation, Bacterial/drug effects, Suppression, Genetic/drug effects
3	64	<i>Escherichia coli</i> /metabolism, <i>Escherichia coli</i> /drug effects, <i>Escherichia coli</i> /genetics, RNA, Transfer/genetics, <i>Escherichia coli</i> /isolation and purification, <i>Escherichia coli</i> /radiation effects, <i>Escherichia coli</i> /growth and development, <i>Escherichia coli</i> /immunology, Bacteriolysis/drug effects, RNA, Transfer/metabolism
4	40	Bacterial Proteins/metabolism, Bacterial Proteins/isolation and purification, Ribosomal Proteins/isolation and purification, Transcription Factors/metabolism, Repressor Proteins/metabolism, Ribosomal Proteins/analysis, Bacterial Proteins/immunology, DNA-Binding Proteins/isolation and purification, Flagellin/metabolism, Ribosomal Proteins/immunology
5	21	RNA, Ribosomal/metabolism, RNA, Bacterial/metabolism, RNA, Ribosomal/isolation and purification, Ribosomal Proteins/metabolism, RNA, Bacterial/isolation and purification, RNA, Transfer, Amino Acyl/metabolism
6	20	DNA, Bacterial/genetics, DNA, Bacterial/analysis, DNA, Bacterial/metabolism, DNA, Bacterial/isolation and purification, DNA/analysis, DNA/genetics, DNA, Superhelical/radiation effects
7	15	Bacterial Proteins/genetics, DNA-Directed DNA Polymerase/genetics, DNA Polymerase I/genetics, Viral Proteins/genetics
8	15	Fimbriae, Bacterial/immunology, <i>Escherichia coli</i> /ultrastructure, Fimbriae, Bacterial/ultrastructure, <i>Escherichia coli</i> /analysis
9	15	Polysaccharides, Bacterial/isolation and purification, Antigens, Bacterial/isolation and purification, Lipopolysaccharides/isolation and purification
10	13	Anti-Bacterial Agents/pharmacology, Anti-Bacterial Agents/metabolism

Table 1.
 Principal research topics related to *E. coli* research in the period 1981–1990.

In the period 1981–1990 the biggest cluster was the cluster “DNA Replication/drug effects” comprised of the MeSH headings dealing with drug or radiation effects on DNA replication and DNA repair. This cluster was found to be the basic theme in this time period. A typical representative article for this cluster is the article published by Fram et al. with the title “DNA repair mechanisms affecting cytotoxicity by streptozotocin in *E. coli*” [31]. A further important basic research time in this period was also encompassed in the cluster “Transcription, Genetic/drug effects” consisting of the MeSH headings relating to the gene expression and regulation of gene expression and drug effects on the gene expression and regulation of this expression. A typical representative article for this cluster is the article published by Goda and Greenblatt with the title “Efficient modification of *E. coli* RNA polymerase *in vitro* by the N gene transcription antitermination protein of bacteriophage lambda” [32]. The major motor theme in this period was covered by the cluster “Bacterial Proteins/metabolism” involving MeSH headings related to metabolism and isolation and purification of different bacterial proteins, e.g., ribosomal proteins, transcription factors, repressors, and other DNA-binding proteins. A typical representative article for this cluster is the article published by Thomas et al. with the title “Amplification and purification of UvrA, UvrB, and UvrC proteins of *Escherichia coli*” [33]. Only one niche theme was detected—the cluster

“DNA, Bacterial/genetics”—covering MeSH headings dealing mostly with isolation and purification of DNA and with DNA analysis. A typical representative article for this cluster is the article published by Klaer et al. with the title “The sequence of IS4” [34].

4.2 Period 1991–2000

For the period 1991–2000, the retrieval yielded 23,470 documents from PubMed. The top 10 clusters that emerged from the co-word analysis are presented in the form of a strategic diagram in **Figure 3**. The corresponding summary of the identified research topics is given in **Table 2**.

In the observed period, the biggest cluster was the cluster “Gene Expression Regulation, Bacterial/drug effects” comprised of the MeSH headings dealing with drug or radiation effects on the genes, their expression, and regulation. This cluster was found to be the basic theme in this time period. A typical representative article for this cluster is the article published by Lutz and Bujard with the title “Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements” [35].

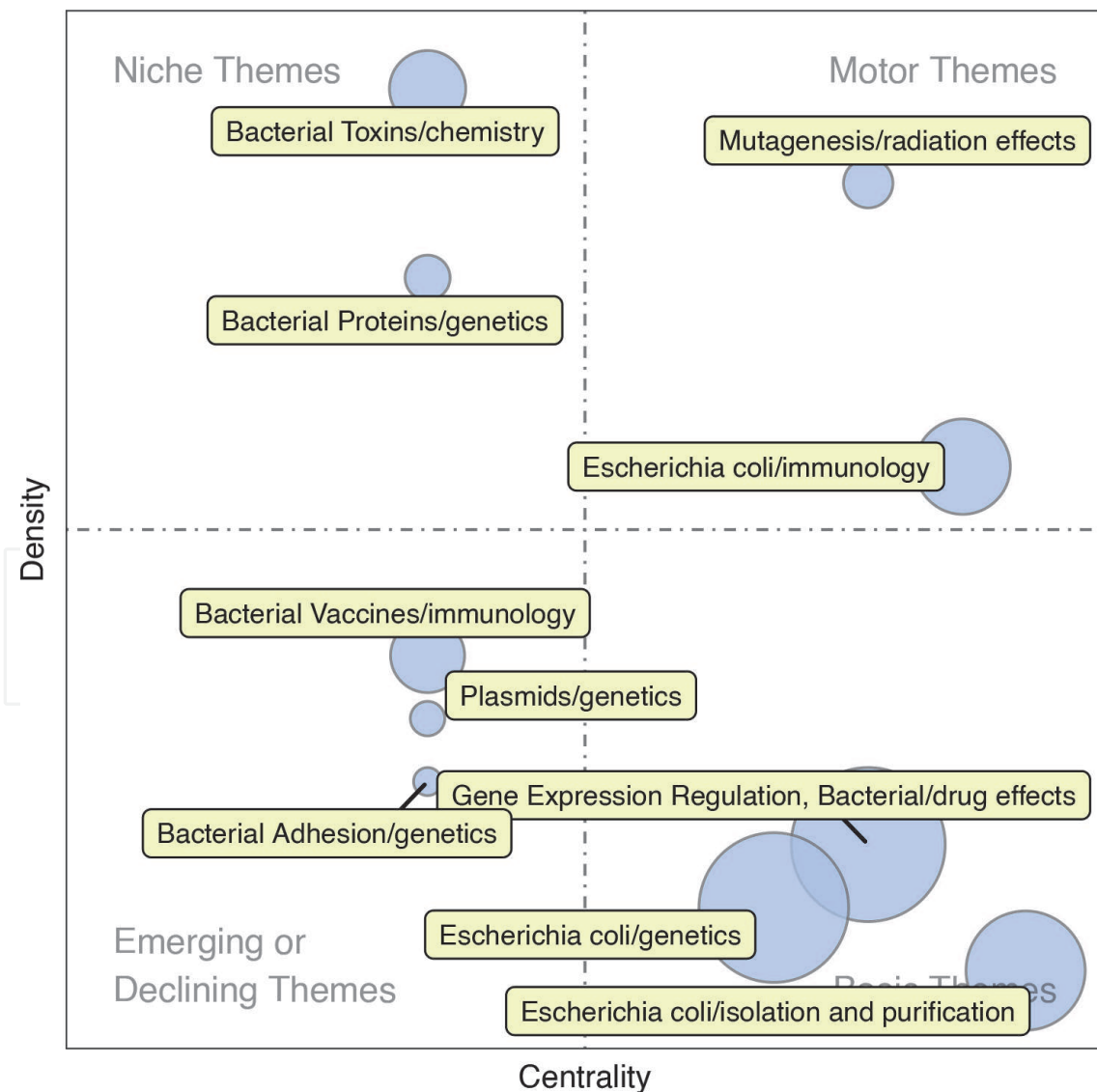


Figure 3. Strategic diagram of the period 1991–2000. Each research topic is represented by a node and labeled with the most frequent pair MeSH heading/MeSH subheading. The size of the node is proportional to the number of MeSH heading/MeSH subheading pairs in each cluster.

ID	Size	MeSH headings
1	122	Gene Expression Regulation, Bacterial/drug effects, Transcription, Genetic/drug effects, Protein Biosynthesis/drug effects, Gene Expression Regulation, Enzymologic/drug effects, Operon/drug effects, Genes, Bacterial/radiation effects, Protein Processing, Post-Translational/drug effects, SOS Response, Genetics/drug effects, SOS Response, Genetics/radiation effects, Gene Expression Regulation, Bacterial/radiation effects
2	110	<i>Escherichia coli</i> /genetics, <i>Escherichia coli</i> /metabolism, <i>Escherichia coli</i> /drug effects, <i>Escherichia coli</i> /chemistry, <i>Escherichia coli</i> /growth and development, <i>Escherichia coli</i> /ultrastructure, <i>Escherichia coli</i> /cytology, Bacterial Outer Membrane Proteins/genetics, <i>Escherichia coli</i> /enzymology, <i>Escherichia coli</i> /radiation effects
3	53	<i>Escherichia coli</i> /isolation and purification, <i>Escherichia coli</i> /classification, <i>Escherichia coli</i> /pathogenicity, <i>Escherichia coli</i> /physiology, Germ-Free Life/immunology
4	30	<i>Escherichia coli</i> /immunology, <i>Escherichia coli</i> Infections/diagnosis, <i>Escherichia coli</i> Infections/epidemiology, <i>Escherichia coli</i> Infections/microbiology, <i>Escherichia coli</i> Infections/therapy, <i>Escherichia coli</i> Infections/drug therapy, <i>Escherichia coli</i> Infections/physiopathology
5	19	Bacterial Toxins/chemistry, Bacterial Toxins/metabolism, Bacterial Toxins/toxicity, Bacterial Toxins/genetics, Enterotoxins/genetics, Enterotoxins/chemistry, Enterotoxins/metabolism, Enterotoxins/toxicity
6	18	Bacterial Vaccines/immunology, Bacterial Vaccines/administration and dosage, Bacterial Vaccines/toxicity, Bacterial Vaccines/standards
7	10	Mutagenesis/radiation effects, DNA, Bacterial/radiation effects, Frameshift Mutation/drug effects, Mutagenesis/drug effects
8	9	Bacterial Proteins/genetics, Bacterial Proteins/chemistry, Bacterial Proteins/metabolism, Bacterial Proteins/physiology
9	7	Plasmids/genetics, Plasmids/chemistry
10	6	Bacterial Adhesion/genetics, Bacterial Adhesion/immunology

Table 2.
 Principal research topics related to *E. coli* research in the period 1991–2000.

A further important basic research time in this period was also encompassed in the cluster “*Escherichia coli*/genetics” including the MeSH headings relating also to *E. coli* metabolism and enzymology, and growth and development. A typical representative article for this cluster is the article published by Hiraga with the title “Chromosome partition in *Escherichia coli*” [36]. The major motor theme in this period was covered by the cluster “*Escherichia coli*/immunology” involving also MeSH related to epidemiology, diagnosis, pathogenesis, and drug therapy of *E. coli* infections. A typical representative article for this cluster is the article published by Johnson with the title “Virulence factors in *Escherichia coli* urinary tract infection” [37]. The major niche theme was the cluster “Bacterial Toxins/chemistry”, covering different MeSH headings dealing with genetics, chemistry, metabolism, and toxicity of bacterial toxins and enterotoxins. A typical representative article for this cluster is the article published by Gyles with the title “*Escherichia coli* cytotoxins and enterotoxins” [38].

4.3 Period 2001–2010

For the period 2001–2010, the search strategy retrieved 24,266 documents. Selected research topics are shown in **Figure 4** (here we point out to the reader that we have included the 11 most important research topics, while the ranks of clusters 7–9 are tied). The contextual meaning of the clusters is summarized in **Table 3**.

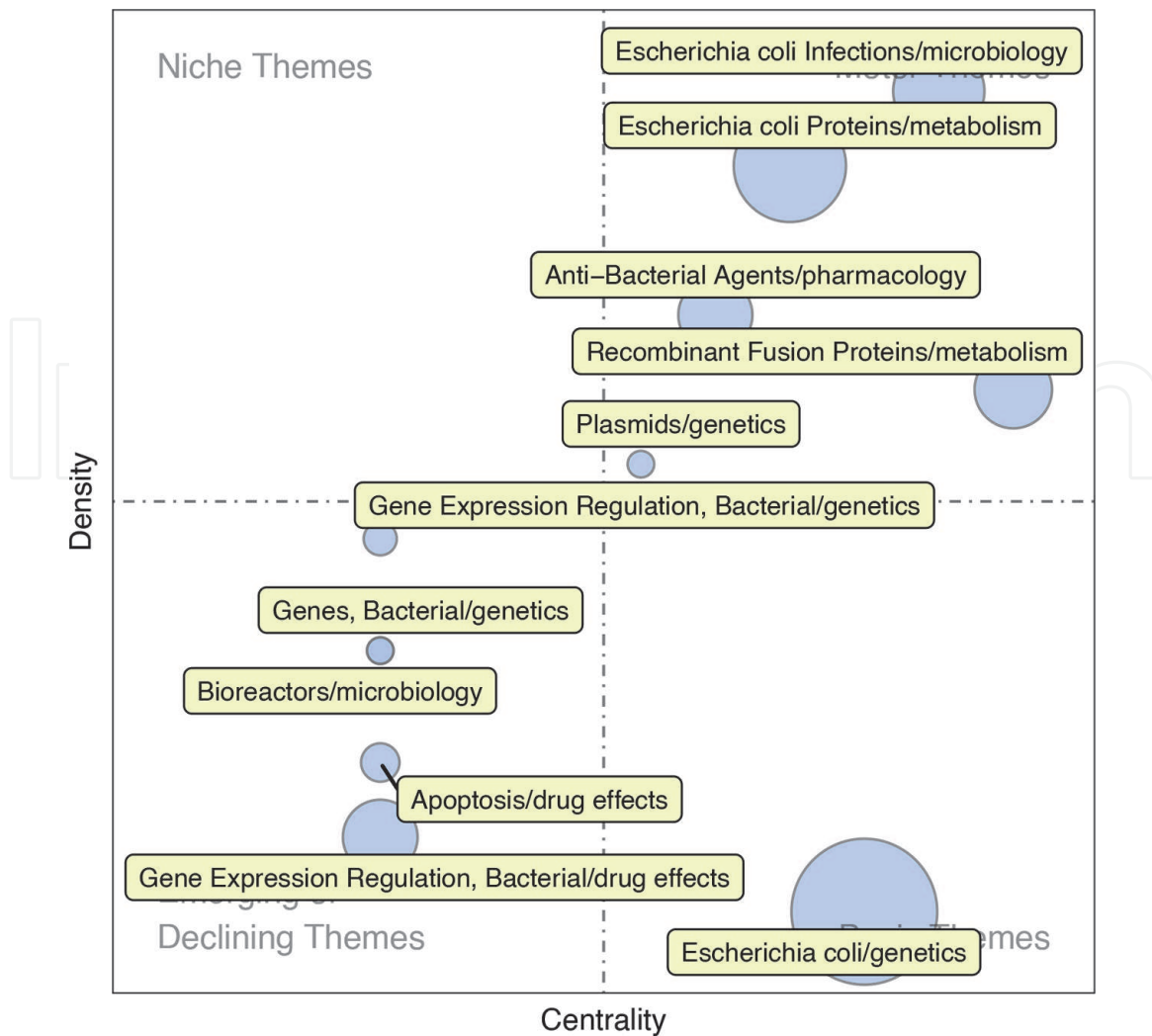


Figure 4.

Strategic diagram of the period 2001–2010. Each research topic is represented by a node and labeled with the most frequent pair MeSH heading/MeSH subheading. The size of the node is proportional to the number of MeSH heading/MeSH subheading pairs in each cluster.

In the period 2001–2010, the biggest cluster was the cluster “*Escherichia coli*/genetics” comprised of very diverse MeSH headings dealing with genetics and classification, but also with enzymology, metabolism and drug effects, isolation, purification and chemistry, cytology, growth, and development and also pathogenicity. This cluster was found to be the only basic theme in this time period. A typical representative article for this cluster is the article published by Tenaillon et al. with the title “The population genetics of commensal *Escherichia coli*” [39].

The major motor theme in this period was covered by the cluster “*Escherichia coli* Proteins/metabolism” involving MeSH headings related to metabolism and isolation and purification of different native, but also recombinant bacterial proteins. A typical representative article for this cluster is the article published by Bell with the title “Structure and mechanism of *Escherichia coli* RecA ATPase” [40]. Among motor themes, another cluster of very similar size was revealed—the cluster “*Escherichia coli* Infections/microbiology” covering the MeSH heading relating to important *E. coli* infection topics associated with pathogenic *E. coli*, e.g., urinary tract infections and bacteremia, but also antimicrobial agents with special emphasis on the topic of beta-lactamases. A typical representative article for this cluster is the article published by Croxen and Finlay with the title “Molecular mechanisms of *Escherichia coli* pathogenicity” [41]. In this time period no niche themes were found.

ID	Size	MeSH headings
1	666	<i>Escherichia coli</i> /genetics, <i>Escherichia coli</i> /metabolism, <i>Escherichia coli</i> /drug effects, <i>Escherichia coli</i> /enzymology, <i>Escherichia coli</i> /isolation and purification, <i>Escherichia coli</i> /chemistry, <i>Escherichia coli</i> /growth and development, <i>Escherichia coli</i> /pathogenicity, <i>Escherichia coli</i> /classification, <i>Escherichia coli</i> /cytology
2	185	<i>Escherichia coli</i> Proteins/metabolism, <i>Escherichia coli</i> Proteins/genetics, Drug Resistance, Bacterial/genetics, <i>Escherichia coli</i> Proteins/chemistry, Bacterial Proteins/chemistry, Bacterial Proteins/metabolism, Bacterial Proteins/genetics, Membrane Proteins/metabolism, <i>Escherichia coli</i> Proteins/isolation, and purification, Recombinant Proteins/genetics
3	83	<i>Escherichia coli</i> Infections/microbiology, <i>Escherichia coli</i> Infections/epidemiology, beta-Lactamases/genetics, Urinary Tract Infections/microbiology, <i>Escherichia coli</i> Infections/drug therapy, Anti-Infective Agents/pharmacology, beta-Lactamases/biosynthesis, Urinary Tract Infections/drug therapy, beta-Lactamases/metabolism, Bacteremia/microbiology
4	49	Recombinant Fusion Proteins/metabolism, Recombinant Fusion Proteins/genetics, Recombinant Fusion Proteins/isolation and purification, Antimicrobial Cationic Peptides/genetics, Antimicrobial Cationic Peptides/metabolism, Antimicrobial Cationic Peptides/pharmacology, Anti-Infective Agents/chemistry, Anti-Infective Agents/metabolism
5	44	Gene Expression Regulation, Bacterial/drug effects, Protein Biosynthesis/drug effects, Nucleic Acid Conformation/drug effects, Transcription, Genetic/drug effects
6	43	Anti-Bacterial Agents/pharmacology, Anti-Bacterial Agents/chemistry, Anti-Bacterial Agents/chemical synthesis, Anti-Bacterial Agents/pharmacokinetics, Anti-Bacterial Agents/therapeutic use, Anti-Bacterial Agents/administration and dosage, Fluoroquinolones/pharmacology, Oligopeptides/pharmacology
7	11	Apoptosis/drug effects, Phagocytosis/drug effects
8	9	Gene Expression Regulation, Bacterial/genetics, Gene Expression Regulation, Bacterial/physiology
9	7	Plasmids/genetics, Plasmids/metabolism
10	7	Genes, Bacterial/genetics, Mutation/drug effects
11	7	Bioreactors/microbiology, Industrial Microbiology/methods

Table 3.
 Principal research topics related to *E. coli* research in the period 2001–2010.

4.4 Period 2011–2021

In the last observed period, we analyzed 30,114 bibliographic records from PubMed. The co-word analysis revealed six thematic clusters, as shown in **Figure 5**. The corresponding details are presented in **Table 4**.

In the period 2011–2022 was again the biggest cluster the cluster “*Escherichia coli*/genetics” comprised of very diverse MeSH headings dealing with genetics and genetically modified microorganisms, but also with enzymology, metabolism, biosynthesis, purification and chemistry, and growth and development. This cluster was again the basic theme in this time period. A typical representative article for this cluster is the article published by Yang et al. with the title “*Escherichia coli* as a platform microbial host for systems metabolic engineering” [42]. A further important basic research time in this period was also encompassed in the cluster “*Escherichia coli*/drug effects” consisting of the MeSH headings relating to bacterial drug resistance. A typical representative article for this cluster is the article published by Da Silva and Mendonça with the title “Association between antimicrobial resistance and virulence in *Escherichia coli*” [43]. Only one major motor theme in this period was revealed, covered by the cluster “Anti-Bacterial Agents/pharmacology” covering MeSH headings related to different aspects, e.g., pharmacology, chemistry of antibacterial agents, including silver and antimicrobial cationic

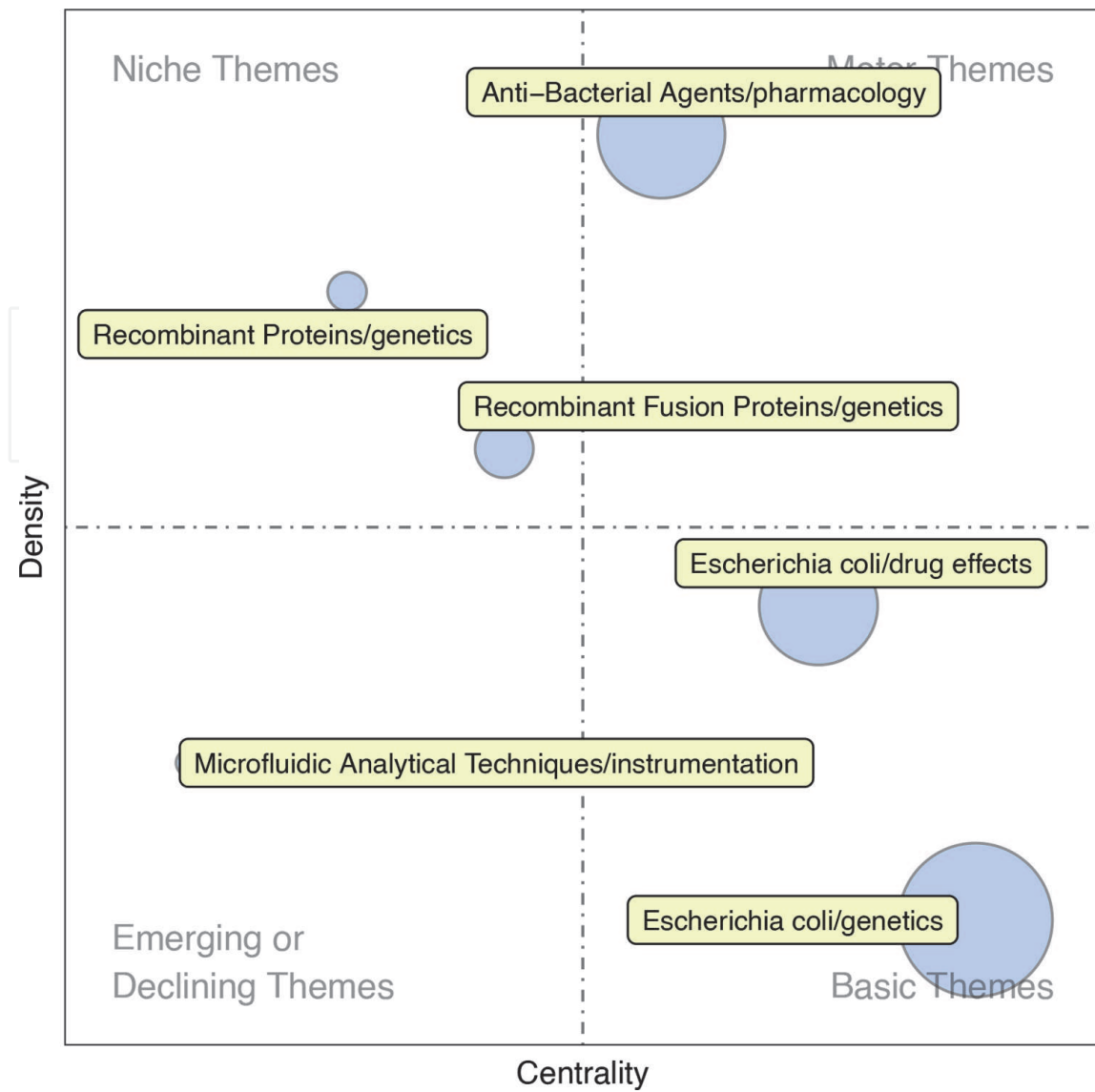


Figure 5. Strategic diagram of the period 2011–2021. Each research topic is represented by a node and labeled with the most frequent pair MeSH heading/MeSH subheading. The size of the node is proportional to the number of MeSH heading/MeSH subheading pairs in each cluster.

ID	Size	MeSH headings
1	2329	<i>Escherichia coli</i> /genetics, <i>Escherichia coli</i> /metabolism, <i>Escherichia coli</i> Proteins/genetics, <i>Escherichia coli</i> Proteins/metabolism, <i>Escherichia coli</i> /enzymology, Bacterial Proteins/genetics, <i>Escherichia coli</i> /growth and development, <i>Escherichia coli</i> /chemistry, Microorganisms, Genetically-Modified/genetics, Bacterial Proteins/biosynthesis
2	921	Anti-Bacterial Agents/pharmacology, Anti-Bacterial Agents/chemistry, Silver/chemistry, Silver/pharmacology, Anti-Infective Agents/pharmacology, Anti-Bacterial Agents/chemical synthesis, Anti-Infective Agents/chemistry, Antimicrobial Cationic Peptides/pharmacology, Antimicrobial Cationic Peptides/chemistry, Titanium/chemistry
3	673	<i>Escherichia coli</i> /drug effects, <i>Escherichia coli</i> /isolation and purification, <i>Escherichia coli</i> /physiology, Drug Resistance, Bacterial/genetics, <i>Escherichia coli</i> /pathogenicity, <i>Escherichia coli</i> Infections/epidemiology, <i>Escherichia coli</i> Infections/microbiology, <i>Escherichia coli</i> /classification, Drug Resistance, Bacterial/drug effects, <i>Escherichia coli</i> Infections/drug therapy
4	81	Recombinant Fusion Proteins/genetics, Recombinant Fusion Proteins/biosynthesis, Recombinant Fusion Proteins/isolation and purification, Recombinant Fusion Proteins/chemistry

ID	Size	MeSH headings
5	41	Recombinant Proteins/genetics, Recombinant Proteins/metabolism, Recombinant Proteins/chemistry
6	28	Microfluidic Analytical Techniques/instrumentation, Microfluidic Analytical Techniques/methods

Table 4.

Principal research topics related to *E. coli* research in the period 2011–2021.

peptides. A typical representative article for this cluster is the article published by Zhao et al. with the title “Synthesis of Ag/AgCl modified anhydrous basic bismuth nitrate from BiOCl and the antibacterial activity” [44]. The major niche theme in this period was covered by the cluster “Recombinant Fusion Proteins/genetics” wrapping the topics of genetics, biosynthesis, chemistry, and isolation and purification of recombinant fusion proteins. A typical representative article for this cluster is the article published by Jeffery with the title “Expression, solubilization, and purification of bacterial membrane proteins” [45]. The second cluster among the niche themes in this time period is the cluster “Recombinant Proteins/genetics” covering the topics of genetics, metabolism, and chemistry of recombinant proteins. A typical representative article for this cluster is the article published by Gopal and Kumar with the title “Strategies for the production of recombinant protein in *Escherichia coli*” [46].

5. Discussion

E. coli is known to be a versatile microorganism—it is a commensal in the gut microbiota of healthy hosts, but can be found also as a pathogen instigating intestinal but also extraintestinal infections [47]. *E. coli* is also a well-known probiotic bacterium, as some important probiotic drugs including *E. coli* are on the market [48–50]. Further, it is a very well-known model microorganism for Gram-negative bacteria, which was and still is used as a laboratory “workhorse” on which many basic topics of molecular biology, physiology, genetics, evolution, genetic engineering, and biotechnology were and still are studied [51–53].

So there is no surprise in finding many papers published on *E. coli*. Bibliometric co-word analysis has the potential to reveal the topic trends in *E. coli* research. The result of this kind of analysis is two-dimensional plots in which circles (i.e., nodes), whose size corresponds to the number of including MeSH terms, are partitioned into different quadrants. The top right quadrant depicts motor themes with strong centrality as well as high density. The upper left quadrant shows specialized themes, which refer to themes having a high density, but also having inadequate external interactions. The bottom-right quadrant shows the basic themes—these are themes that have a strong centrality, but low density. In the bottom, left quadrant themes are shown that are emerging or declining, as they have in general low density and centrality. For just one strategic diagram, it is usually not possible to determine whether a theme is emerging or declining, however, when data from several graphs, each from a certain period, are compared, for some themes that are found in more graphs a trend can be established. From our analysis, it can be assumed that the cluster “Anti-Bacterial Agents/pharmacology” which appeared in the 1981–1990 period in the quadrant of Emerging or declining themes was in that time period an emerging theme, as the same cluster can be also found in the graphs of the periods 2001–2010 and 2011–2022, namely in the quadrant of motor themes. A similar can

be stated for the cluster “Plasmids/genetics”, which appeared on the strategic diagram of the period 1991–2000 as an emerging theme, as the same cluster can be found also in the 2001–2010 diagram, namely in the quadrant of motor themes. An example of a declining cluster theme is the cluster “Gene Expression Regulation, Bacterial/drug effects” which is the major basic theme in the diagram of the time period 1991–2000, but moved to the emerging or declining themes quadrant in the plot of the time period 2001–2010. In the strategic diagram of the period 2011–2021 in the emerging or declining themes quadrant the cluster “Microfluidic Analytical Techniques/instrumentation” appeared, which is for sure an emerging cluster as much of the *E. coli* research is now moving into the area of single cell analysis which is enabled by the microfluidic techniques.

6. Conclusions

In the present study, we retrieved nearly 100,000 scientific articles on *E. coli* from the PubMed bibliographic database and investigated the intellectual structure and evolution using co-word analysis. To our knowledge, this is the first systematic knowledge mapping in the field of *E. coli* research. The analysis performed clearly revealed the main research topics in *E. coli* research over the last decades. Based on this analysis, major, niche, and basic topics in *E. coli* research were identified in each decade studied, and new topics are expected to emerge. The future in the field of *E. coli* research lies in single-cell analysis.

Acknowledgements

This work was supported by the Slovenian Research Agency (Grant No. J5-2552 (AK) and P1-0198 (MSE)).

Conflict of interest

The authors declare no conflict of interest.

IntechOpen

Author details

Andrej Kastrin¹ and Marjanca Starčič Erjavec^{2*}

¹ Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

² Biotechnical Faculty, Department of Microbiology, University of Ljubljana, Ljubljana, Slovenia

*Address all correspondence to: marjanca.starctic.erjavec@bf.uni-lj.si

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Escherich T. The intestinal bacteria of the neonate and breast-fed infant. *Reviews of Infectious Diseases*. 1988; **10**(6):1220-1225
- [2] Larsen PO, von Ins M. The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics*. 2010; **84**(3):575-603
- [3] Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*. 2015; **66**(11):2215-2222
- [4] National Library of Medicine (US). MEDLINE PubMed Production Statistics. 2021. Available from: https://www.nlm.nih.gov/bsd/medline_pub_med_production_stats.html [Accessed: November 4, 2022]
- [5] Clauset A, Larremore DB, Sinatra R. Data-driven predictions in the science of science. *Science*. 2017; **355**(6324): 477-480
- [6] Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, et al. Science of science. *Science*. 2018; **359**(6379):eaao0185
- [7] Ebbinghaus H. *Psychology: An Elementary Text-Book*. Boston, MA: D. C. Heath and Company; 1908
- [8] Lotka AJ. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*. 1926; **16**(12):317-323
- [9] Zipf GK. *Human Behavior and the Principle of Least Effort*. Oxford: Addison-Wesley Press; 1949
- [10] Price DJ. Networks of scientific papers. *Science*. 1965; **149**(3683):510-515
- [11] Merton RK. The Matthew effect in science: The reward and communication systems of science are considered. *Science*. 1968; **159**(3810):56-63
- [12] Garfield E. Citation indexing for studying science. *Nature*. 1970; **227**(5259):669-671
- [13] Börner K, Chen C, Boyack KW. Visualizing knowledge domains. *Annual Review of Information Science and Technology*. 2003; **37**(1):179-255
- [14] Uzzi B, Mukherjee S, Stringer M, Jones B. Atypical combinations and scientific impact. *Science*. 2013; **342**(6157):468-472
- [15] Wang D, Song C, Barabási AL. Quantifying long-term scientific impact. *Science*. 2013; **342**(6154):127-132
- [16] Wu L, Kittur A, Youn H, Milojević S, Leahey E, Fiore SM, et al. Metrics and mechanisms: Measuring the unmeasurable in the science of science. *Journal of Informetrics*. 2022; **16**(2): 101290
- [17] Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*. 2011; **5**(1):146-166
- [18] Klavans R, Boyack KW. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*. 2017; **68**(4):984-998
- [19] Watts DJ. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton, NJ: Princeton University Press; 2003

- [20] Newman M. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*. 2005;**46**(5): 323-351
- [21] Barabási AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;**286**(5439):509-512
- [22] Ravasz E, Barabási AL. Hierarchical organization in complex networks. *Physical Review E*. 2003;**67**(2):026112
- [23] Wang D, Barabási AL. *The Science of Science*. Cambridge, MA: Cambridge University Press; 2021
- [24] Yi S, Choi J. The organization of scientific knowledge: The structural characteristics of keyword networks. *Scientometrics*. 2012;**90**(3):1015-1026
- [25] Lipscomb CE. Medical subject headings (MeSH). *Bulletin of the Medical Library Association*. 2000; **88**(3):265-266
- [26] Dunne E, Hulek K. Mathematics subject classification 2020. *European Mathematical Society Magazine*. 2020; **115**:5-6
- [27] Callon M, Courtial JP, Turner WA, Bauin S. From translations to problematic networks: An introduction to co-word analysis. *Information (International Social Science Council)*. 1983;**22**(2):191-235
- [28] Moral-Munoz JA, Lucena-Antón D, Perez-Cabezas V, Carmona-Barrientos I, González-Medina G, Ruiz-Molinero C. Highly cited papers in microbiology: Identification and conceptual analysis. *FEMS Microbiology Letters*. 2018; **365**(20):fny230
- [29] Callon M, Courtial JP, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*. 1991;**22**(1):155-205
- [30] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;**2008**(10): P10008
- [31] Fram RJ, Mack SL, George M, Marinus MG. DNA repair mechanisms affecting cytotoxicity by streptozotocin in *E. coli*. *Mutation Research*. 1989; **218**(2):125-133
- [32] Goda Y, Greenblatt J. Efficient modification of *E. coli* RNA polymerase *in vitro* by the N gene transcription antitermination protein of bacteriophage lambda. *Nucleic Acids Research*. 1985;**13**(7):2569-2582
- [33] Thomas DC, Levy M, Sancar A. Amplification and purification of UvrA, UvrB, and UvrC proteins of *Escherichia coli*. *The Journal of Biological Chemistry*. 1985;**260**(17):9875-9883
- [34] Klaer R, Kühn S, Tillmann E, Fritz HJ, Starlinger P. The sequence of IS4. *Molecular and General Genetics*. 1981;**181**(2):169-175
- [35] Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*. 1997; **25**(6):1203-1210
- [36] Hiraga S. Chromosome partition in *Escherichia coli*. *Current Opinion in Genetics & Development*. 1993;**3**(5): 789-801
- [37] Johnson JR. Virulence factors in *Escherichia coli* urinary tract infection. *Clinical Microbiology Reviews*. 1991; **4**(1):80-128
- [38] Gyles CL. *Escherichia coli* cytotoxins and enterotoxins. *Canadian Journal of Microbiology*. 1992;**38**(7): 734-746

- [39] Tenaille O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology*. 2010;8(3): 207-217
- [40] Bell CE. Structure and mechanism of *Escherichia coli* RecA ATPase. *Molecular Microbiology*. 2005;58(2): 358-366
- [41] Croxen MA, Finlay BB. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature Reviews Microbiology*. 2010;8(1):26-38
- [42] Yang D, Prabowo CPS, Eun H, Park SY, Cho IJ, Jiao S, et al. *Escherichia coli* as a platform microbial host for systems metabolic engineering. *Essays in Biochemistry*. 2021;65(2):225-246
- [43] Da Silva GJ, Mendonça N. Association between antimicrobial resistance and virulence in *Escherichia coli*. *Virulence*. 2012;3(1):18-28
- [44] Zhao M, Hou X, Lv L, Wang Y, Li C, Meng A. Synthesis of Ag/AgCl modified anhydrous basic bismuth nitrate from BiOCl and the antibacterial activity. *Materials Science & Engineering: C*. 2019;98:83-88
- [45] Jeffery CJ. Expression, solubilization, and purification of bacterial membrane proteins. *Current Protocols in Protein Science*. 2016;83: 29.15.1-29.15.15
- [46] Gopal GJ, Kumar A. Strategies for the production of recombinant protein in *Escherichia coli*. *The Protein Journal*. 2013;32(6):419-425
- [47] Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*. 2004;2(2): 123-140
- [48] Sonnenborn U. *Escherichia coli* strain Nissle 1917—From bench to bedside and back: History of a special *Escherichia coli* strain with probiotic properties. *FEMS Microbiology Letters*. 2016;363(19):fnw212
- [49] Wassenaar TM, Zschüttig A, Beimfohr C, Geske T, Auerbach C, Cook H, et al. Virulence genes in a probiotic *E. coli* product with a recorded long history of safe use. *European Journal of Microbiology & Immunology*. 2015;5(1):81-93
- [50] Wassenaar TM. Insights from 100 years of research with probiotic *E. coli*. *European Journal of Microbiology & Immunology*. 2016;6(3):147-161
- [51] Blount ZD. The unexhausted potential of *E. coli*. *eLife*. 2015;4:e05826
- [52] Foster PL. Adaptive mutation: Implications for evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*. 2000;22(12):1067-1074
- [53] Tatum EL, Lederberg J. Gene recombination in the bacterium *Escherichia coli*. *Journal of Bacteriology*. 1947;53(6):673-684