ORIGINAL ARTICLE

Eastern Journal of Medicine

East J Med 28(4): 595-602, 2023
DOI: 10.5505/ejm.2023.53179

Global Trends in Genomic Data Analyses: A Bibliometric Study

Sadi Elasan

Van Yuzuncu Yil University, Faculty of Medicine, Department of Biostatistics, Van, Türkiye

ABSTRACT

Genomic data analysis is a field that deals with the processing and interpretation of large amounts of genetic data. Studies in this field offer researchers the opportunity to work in many fields. In addition, bibliometric analyses are an important tool in identifying publications and trends in the field of genomic data analysis. The purpose of this bibliometric review is to analyse the current status and trends to identify trends and prevalence of worldwide research in the field of genomic data analysis and to guide the advancement of research in this field.

In this bibliometric study, genomic data analysis studies conducted globally between 1993-2023 were examined. As a result of the searches made using the keyword "genomic data analysis", 258 studies were found and 173 articles were used by eliminating the unsuitable and non-article studies. The articles in the database were analyzed using information such as title, author names, publication year, journal name and number of citations. Web of Science and Scopus databases was used for systematic data collection and all text data of the publications included in the study were evaluated with VOSviewer software. These analyses were performed using text mining and data visualization methods (bubble maps and other graphical) to ensure the accuracy and reliability of the study.

In this article, information is given about 173 articles taken from the WOS and Scopus databases and 8706 citations to these articles. The average number of citations per article was calculated as 51 and the H index as 28. Since 2013, both the number of articles and the number of citations have increased. Most of the articles (60%) are devoted to mathematical biology, biochemical, genetic inheritance. The United States of America is the country that publishes the most (47%) articles on this subject. Most of the articles (59%) were published by publishers such as Springer Nature, Oxford University Press, Elsevier, Wiley. Most of the articles (91%) are in SCI-Expanded category.

The findings of our study show that many researchers are active in the field of genomic data analysis and that the researches in this field are increasing. This bibliometric analysis reveals global trends and important work in the field of genomic data analysis and provides important insights into the future directions of research in this field. Since this analysis only aims to identify trends and trends in the literature, it is not intended to evaluate the applicability of this topic.

Keywords: Genomic data analysis, bibliometric study, bioinformatics

Introduction

Genomic data can be defined as the information contained in the genetic material (DNA or RNA) of living things, recorded in a certain format or data structure. This data provides information about many different genetic traits, such as the sequence of genes, their variation, expression and function. Genomic data analysis is the process of processing, interpreting and extracting meaningful information from this data. Genomic data analysis allows researchers to conduct studies in many different areas, such as understanding the genetic structure of living things and developing the causes and treatments of diseases. (1, 2)

Genomic data analysis is the process of collecting, processing and interpreting large amounts of genetic data. These data contain genetic codes that provide information about the genomic components of an organism. Genomic data analysis can be used for many different purposes, such as identifying genetic variations, determining the relationship of genetic factors to disease, and understanding the development of an organism. These analyses are usually performed using high-performance computing technologies, statistical methods and machine-learning techniques. (3, 4)

Bibliometric analysis is concerned with the quantitative examination and evaluation of publications in a field. These analyses determine the trends and priorities of that field by evaluating the number of researches in a particular field with different parameters such as publication year, authors, institutions, keywords, journals and citation counts. In particular, bibliometric studies on genomic data analysis help determine the

---

***Corresponding Author:** Sadi Elasan, Yuzuncu Yil University, Faculty of Medicine, Department of Biostatistics, pbx:65040, Zeve Campus, Van, Türkiye
E-mail: sadielasan@yyu.edu.tr, Phone: +90 432 225 1701 / 25205
ORCID ID: Sadi Elasan: 0000-0002-3149-6462

Received: , Accepted:

trends and priorities of research in this field and provide predictions about the future direction of research in this field. Therefore, bibliometric analyses are an important tool for scaling and evaluating research in a field and help researchers identify trends and research topics in the literature. (5, 6)

This study aims to bibliometrically examine the worldwide trends in genomic data analysis. This review aims to reveal the publications on genomic data analysis made by researchers in different disciplines using the bibliometric analysis method. This study is the first example of bibliometric research on general genomic data analysis. The main goal is to explore the importance of publications on this topic and to review relevant trends and clusters.

Materials and Methods

A systematic data collection method, search strategy and network analysis software were used to ensure the reliability of our study and the accuracy of the results. The use of these methods enabled the collection and analysis of the most up-to-date and comprehensive data in the literature. In addition, the analysis of these data provides information about the current trends and trends of the scientific community in the field of genomic data analysis. Global publication trends in the field of genomic data analysis were determined by examining different factors such as the most influential researchers, countries and the most frequently used keywords.

Data Collection Method and Search Strategy: In this bibliometric study, global genomic data analysis studies conducted between 1993-2023 (last accessed: March 20, 2023) using the "Web of Science Core Collection (WOS, Clarivate Analytics, Philadelphia, PA, USA)" and "Scopus (Elsevier B.V.)" databases were examined. As a result of searches in the database using the keyword "genomic data analysis", 258 studies were found. The remaining 173 articles were used when unsuitable studies from these studies and studies other than the article were eliminated. The articles in the database were analyzed using information such as article title, authors' names, publication year, journal name and number of citations. The Hirsch (H) index was used as an indicator of publication effect. The materials were accessed using the online library and digital resources of Van Yuzuncu Yil University. The search language was English.

In the study, publications related to genomic data analysis using the WOS database were examined using bibliometric methods. WOS is a comprehensive database of academic articles published in many disciplines and subjects. This database is an important resource for interdisciplinary research. In this study, publications in the WOS database were collected using certain search terms and subjected to bibliometric analysis. Data were collected using the online interface of WOS and analyzed using various parameters. Data on post-growth, most active countries and institutions, and keyword matching were analyzed. All scanned articles were meticulously reviewed.

Network Analysis: In this bibliometric study, the "network of collaboration, highlights and future trends" was analyzed using VOSviewer (version 1.6.19, University of Leiden, The Netherlands) to identify global trends in genomic data analysis and key issues of research in this field. Web of Science (WOS) database was used for systematic data collection, and all text data of the publications included in the study were collected and evaluated with VOSviewer software. These analyses were performed using text mining and data visualization (bubble maps and other graphical) methods to ensure the accuracy and reliability of the study.

Figures: 1, 2, 3, 4

Bubble Maps: In the bibliometric analyses made with VOSviewer, the graphs called "Bubble Maps" show the grouping of the articles published in a research area according to the frequency of the keywords. Each keyword or group is represented as a "balloon", and the size of the balloon depends on the frequency of that keyword or group. The balloons are colour-coded, ensuring that keywords related to the same group or topic are close together.

The thickness and length of the lines between the balloons indicate how often these keywords are used together and with which other keywords. Thus, by using "bubble maps" in a bibliometric analysis, it is possible to identify the most important keywords or groups in the research area and visualize the relationships between these keywords.

Bubble maps, which are frequently used in bibliometric analysis, consist of small circles (bubbles), each representing a publication. The size of the bubbles is proportional to the number of citations the publication receives. Colours indicate sets or topics of publications. The thickness of the lines between two broadcasts

Fig. 1. International collaboration network map. (Collaboration between countries is shown by lines, with thickness indicating strength and circle/text size indicating the level of international collaboration.)

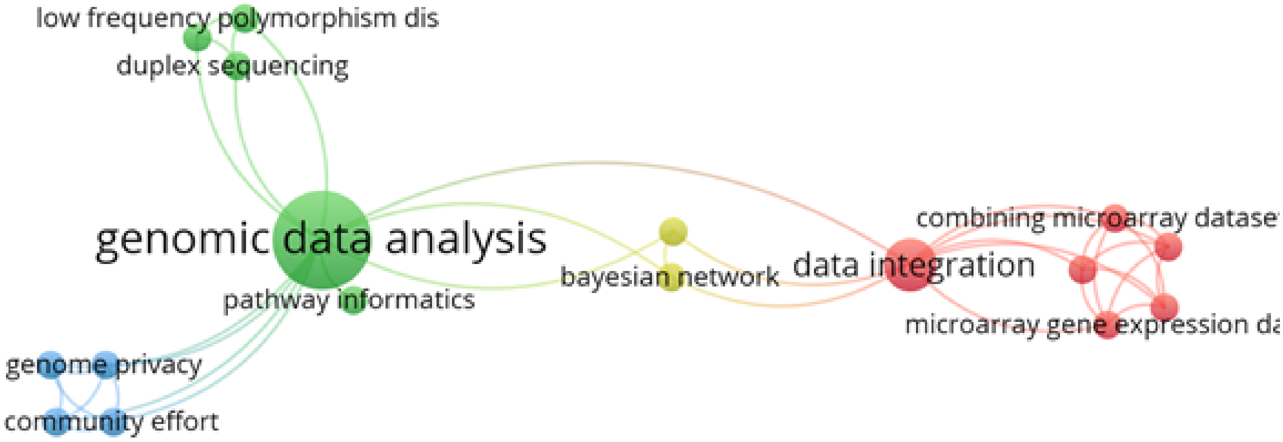

Fig. 2. Keyword analysis. (Shows which keywords the topic is associated with and how often those keywords are used.)

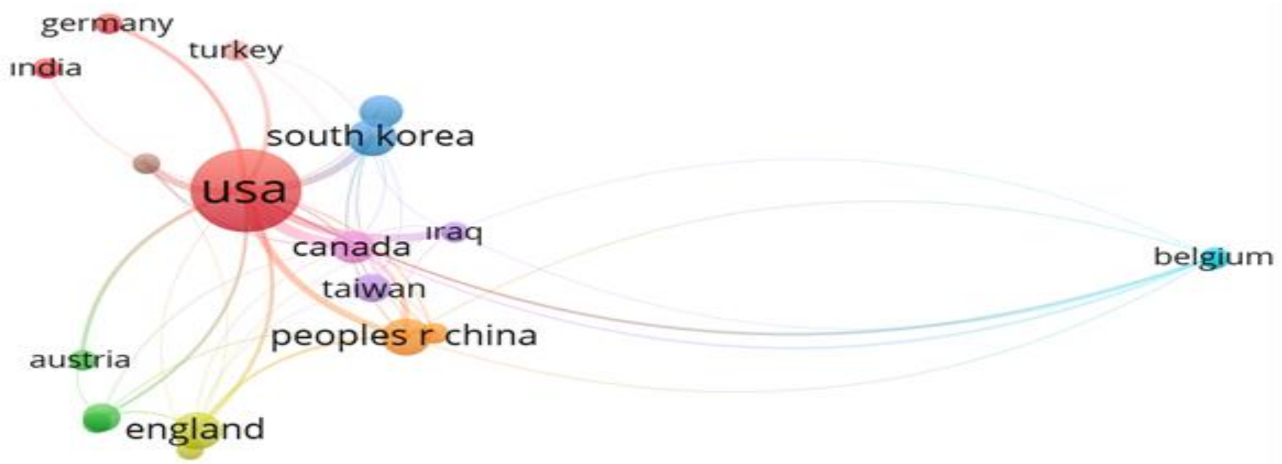

Fig. 3. Bibliographic coupling analysis for country. (The relatedness of items was determined based on the number of references they shared.)

reflects the co-citation frequency between those broadcasts. Bubble maps allow researchers to quickly and easily understand keywords, citations, authors, institutions, countries, and topics among publications on a particular topic.

Results

One hundred seventy-three published articles were retrieved from the WOS database. A total of 8722 citations were made to articles (8706 citations without self-.

Fig. 4. Bibliographic coupling analysis for organization. (The relatedness of items was determined based on the number of references they shared.)

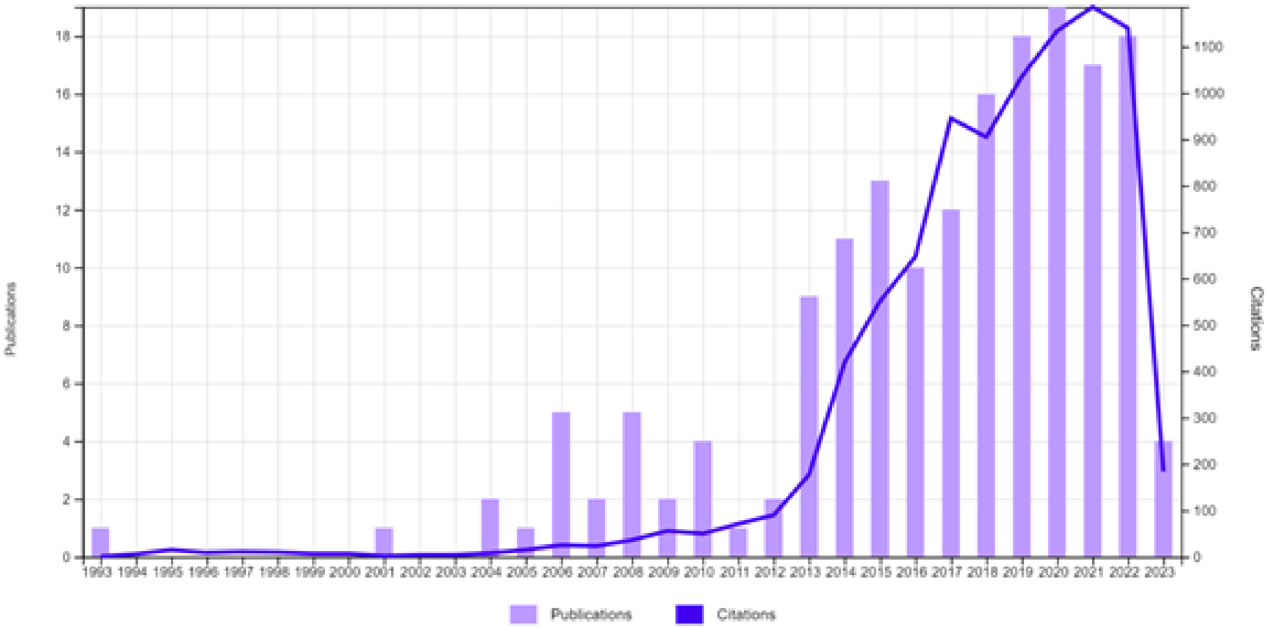

Fig. 5. Frequency of publications and citations by year

citations). The average number of citations per article was 51, and the H index was 28. Especially since 2013, both the number of citations and the number of articles showed an increasing trend. The distribution of publications and citations is shown in Figure 5. The first article and the first citation were made in 1993. Figure: 5

The number of articles was the highest in Mathematical Computational Biology (26.6%), Biochemical Research Methods (24.3%), Statistics Probability (21.4%), Biotechnology Applied Microbiology (20.9%), Computer Science Interdisciplinary Applications (16.8%), Genetics Heredity (16.2%)" and other areas. The distribution of publications by research area is shown in Table 1.

The United States (USA) ranks first in the number of articles published (n:82; 47.4%), followed by China (n:26; 15.0%); South Korea (n:14; 8.1%); England (n:10; 5.8%) and Italy (n:10; 5.8%) followed. There were broadcasts from 52 countries around the world, including these first five countries, and Turkey ranked 19th. Countries with more than two publications are listed in Table 2.

The University of California System (5.2%), the University of Pennsylvania (5.2%), and the Chinese Academy of Sciences (4.6%) were the leading institutions. Most of the leading links were from the USA. In this context, 10 of 492 records are shown in the table.

Most of the articles were published in Springer Nature (23.7%), Oxford University Press (13.9%), Elsevier (13.3%), Public Library Science (4.6%), and Wiley (3.5%). Other journals that have published the most articles on the subject are listed in Table 3.

Looking at the Web of Science Indexes, Most of the articles are in the "Science Citation Index Expanded (SCI-Expanded) (91.3%)" category, followed by "Book Citation Index (BKCI-S) (4.6%)", "Emerging Sources Citation Index (ESCI)" respectively. (4.1%) followed by "Conference Proceedings Citation Index (CPCI-S) (4.1%)" and "Social Sciences Citation Index (SSCI) (2.3%)" (Table 4).

Today, genomic research is advancing at a rapid pace and with new technologies, the collection, analysis and interpretation of genetic data is

Table 1. Publication Categories

Web of Science Categories	Record Count	% of 173
Mathematical Computational Biology	46	26.590
Biochemical Research Methods	42	24.277
Statistics Probability	37	21.387
Biotechnology Applied Microbiology	36	20.809
Computer Science Interdisciplinary Applications	29	16.763
Genetics Heredity	28	16.185
Multidisciplinary Sciences	15	8.671
Biochemistry Molecular Biology	13	7.514
Computer Science Information Systems	10	5.780
Computer Science Theory Methods	7	4.046
Microbiology	7	4.046
Biology	6	3.468
Medical Informatics	6	3.468
Plant Sciences	6	3.468
Cell Biology	5	2.890
Computer Science Artificial Intelligence	5	2.890
Health Care Sciences Services	5	2.890
Oncology	5	2.890

*Showing 18 out of 55 entries

getting easier. These developments have caused the analysis of genomic data to gain importance in many fields around the world. Genomic data analysis is used in many fields, such as medicine, agriculture, animal husbandry, biotechnology and the environment, and it allows fast and effective processing of information in large data sets.

Genomic data analysis is one of the most important applications of modern biotechnology. Genomic data analysis provides information about an organism's genomic characteristics through the processing, analysis, and interpretation of data from cells containing genetic material. These analyses are used in many areas, such as identifying genetic variations, understanding the causes and treatments of complex diseases, population genetics, evolution and development. It combines many different technologies and approaches, such as genomic data analysis, high-performance computing technologies, bioinformatics tools and databases, statistical methods and machine learning techniques. (7, 8)

Genomic data analysis is becoming increasingly important due to the large-scale production and storage of biological data. These data are increasing rapidly, especially thanks to DNA sequencing and microarray technologies. Genomic data analysis is used in many areas, such as the identification of genetic variations, understanding the causes of complex diseases, drug discovery and development, breed improvement in agriculture and livestock, and biological evolution through the processing and interpretation of an organism's genetic material. Such analyses are carried out using high-performance computing technologies, statistical methods, databases and bioinformatics tools. (9, 10)

Based on the results obtained, it has been observed that the number of studies on genomic data analysis has increased rapidly in recent years. This increase can be attributed to increased interest in the analysis of expanding genomic datasets. The results of the literature review show that different approaches and methods are used in the field of genomic data analysis. (11 12) These methods include machine learning, artificial neural networks, data mining, network analysis and statistical methods. (13) In addition, it has been determined that studies in this field mostly focus on human diseases, plant genomics, microbial genomes, animal genomics, genomics of prehistoric organisms, environmental genomics and biotechnology. (14)

Considering similar studies, Tirink (2022) analyzed genome-wide association studies (GWAS) using bibliometric methods. (15) Sweileh et al. (2016) present a bibliometric analysis of genomic research from 2000 to 2015. (16) Zhang and Lin

Table 2. Countries With At Least 2 Publications

Countries/Regions	Record Count	% of 173
USA	82	47.399
China	26	15.029
South Korea	14	8.092
England	10	5.780
Italy	10	5.780
Germany	9	5.202
Japan	8	4.624
Finland	6	3.468
Belgium	5	2.890
Canada	5	2.890
Russia	5	2.890
Spain	5	2.890
Australia	4	2.312
Egypt	4	2.312
France	4	2.312
Netherlands	4	2.312
Switzerland	4	2.312
Taiwan	4	2.312
Türkiye	4	2.312
Brazil	3	1.734
India	3	1.734
Israel	3	1.734
Poland	3	1.734

*Showing 23 out of 52 country

Table 3. Publishers With The Highest Number of Articles On Genomic Data Analyses

Publishers	Record Count	% of 173
Springer Nature	41	23.699
Oxford Univ Press	24	13.873
Elsevier	23	13.295
Public Library Science	8	4.624
Wiley	6	3.468
Frontiers Media Sa	5	2.890
Humana Press Inc	5	2.890
IEEE	5	2.890
Mary Ann Liebert Inc	5	2.890
NATURE PORTFOLIO	5	2.890
Taylor & Francis	4	2.312
Amer Soc Microbiology	3	1.734
MDPI	3	1.734

Showing 13 entries with >2 documents

(2023) examined scientific publications on exome sequencing between 2008 and 2020 with bibliometric methods. (17) Zhang et al. (2019) present a bibliometric analysis of scientific publications on genomic epidemiology between 2000 and 2017. (18) In this study, current and "general genomic data analysis" studies until 2023 were discussed.

Table 4. Web of Science Categories of Index

Web of Science Index	Record Count	% of 173
Science Citation Index Expanded (SCI-Expanded)	158	91.329
Book Citation Index – Science (BKCI-S)	8	4.624
Emerging Sources Citation Index (ESCI)	7	4.046
Conference Proceedings Cit. Index-Science (CPCI-S)	7	4.046
Social Sciences Citation Index (SSCI)	4	2.312

This article presents a bibliometric analysis examining worldwide trends and publication trends in genomic data analysis. The article is based on the analysis of 173 articles obtained as a result of a large literature review. In this analysis, the most influential countries, institutions, authors, journals and keywords in genomic data analysis were determined. Its results can be used to guide research in this area and provide a roadmap for research in the field of genomic data analysis.

In conclusion, this bibliometric analysis reveals global trends and important work in the field of genomic data analysis and provides important insights into the future directions of research in this field.

The highlights of these results can be summarized as follows: This study is a bibliometric review of worldwide trends in genomic data analysis. Between 2013-2023, the number of genomic data analysis studies and the number of citations increased significantly. The USA, China, South Korea and the U.K. are the countries that conduct the most genomic data analysis. The most widely used methods of genomic data analysis include RNA sequencing, genome variant analysis and gene expression analysis. Future research in the field of genomic data analysis may focus on the development of biological data analysis methods and increased collaboration between related disciplines.

The limitations of these results can be summarized as follows: It should be noted that this study does not make an assessment of the accuracy or quality of the original data, presenting only a bibliometric analysis. Therefore, more comprehensive studies should be carried out in related fields, and a similar analysis should be performed using different databases. Moreover, since this bibliometric analysis only aims to identify trends and trends in the literature, it is not intended to evaluate the applicability of this topic.

Declaration of Interests: The author declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding: The author received no financial support for the research, authorship, and/or publication of this article.

About ethical approval: This study is a bibliometric analysis that does not include human or animal experiments. Therefore, ethics committee approval is not required.

Author Contributions: All aspects of the article were done by S.E.

Guarantor: There is no guarantor in this study.

Acknowledgements: I would like to thank Grammarly (app.grammarly.com) for English language editing.

References

1. Kahn, S. D. (2011). On the future of genomic data. science, 331(6018), 728-729.
2. Bouaoun, L., Sonkin, D., Ardin, M., Hollstein, M., Byrnes, G., Zavadil, J., & Olivier, M. (2016). TP53 variations in human cancers: new lessons from the IARC TP53 database and genomics data. Human mutation, 37(9), 865-876.
3. Vaske, C. J., Benz, S. C., Sanborn, J. Z., Earl, D., Szeto, C., Zhu, J., & Stuart, J. M. (2010). Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics, 26(12), i237-i245.
4. Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. Genomics, 99(6), 323-329.
5. Joshi, M. A. (2014). Bibliometric indicators for evaluating the quality of scientific publications. The journal of contemporary dental practice, 15(2), 258.
6. Bornmann, L., Thor, A., Marx, W., & Schier, H. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. Journal of

the Association for Information Science and Technology, 67(11), 2778-2789.

7. Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. BMJ: British Medical Journal, 324(7344), 1018.
8. Alharbi, W. S., & Rashid, M. (2022). A review of deep learning applications in human genomics using next-generation sequencing data. Human Genomics, 16(1), 1-20.
9. Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., & Huang, X. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nature genetics, 50(2), 278-284. DOI: 10.1038/s41588-018-0041-z
10. Navarro, F. C., Mohsen, H., Yan, C., Li, S., Gu, M., Meyerson, W., & Gerstein, M. (2019). Genomics and data science: an application within an umbrella. Genome biology, 20, 1-11.
11. Williams, J. K., Tripp-Reimer, T., Daack-Hirsch, S., & DeBerg, J. (2016). Five-year bibliometric review of genomic nursing science research. Journal of Nursing Scholarship, 48(2), 179-186.
12. Daniels, H., Jones, K. H., Heys, S., & Ford, D. V. (2021). Exploring the use of genomic and routinely collected data: narrative literature review and interview study. Journal of Medical Internet Research, 23(9), e15739.
13. Di Franco, G., & Santurro, M. (2021). Machine learning, artificial neural networks and social research. Quality & quantity, 55(3), 1007-1025.
14. Xu, J. (2006). Invited review: microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. Molecular ecology, 15(7), 1713-1731.
15. Tirink, C. (2022). Bibliometric Analysis for Genome-Wide Association Studies in Animal Science. Black Sea Journal of Agriculture, 5(3), 234-239.
16. Sweileh, W. M., Al-Jabi, S. W., Sawalha, A. F., AbuTaha, A. S., & Zyoud, S. E. H. (2016). Bibliometric analysis of publications on Campylobacter:(2000–2015). Journal of Health, Population and Nutrition, 35(1), 1-12.
17. Zhang, J., & Lin, M. (2023). A comprehensive bibliometric analysis of Apache Hadoop from 2008 to 2020. International Journal of Intelligent Computing and Cybernetics, 16(1), 99-120.
18. Zhang, Q., Yue, Y., Shi, B., & Yuan, Z. (2019). A bibliometric analysis of cleft lip and palate-related publication trends from 2000 to 2017. The Cleft Palate-Craniofacial Journal, 56(5), 658-669.