Curriculum and Teaching Methodology (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/curtm.2024.070829
ISSN 2616-2261 Vol. 7 Num. 8

Biblioshiny-R Application on Bioinformatics Education Research (2004-2023)

Fengmei Yang[a], Jinming Gao[b], Xianzhao Kan[c,*]

College of Life Sciences, Anhui Normal University, Wuhu, 241000, China
[a]fengmei@ahnu.edu.cn, [b]gaojming@ahnu.edu.cn, [c]xianzhao@ahnu.edu.cn
[*]Corresponding author

Keywords: Bioinformatics teaching, Education, Bibliometric analysis, Biblioshiny

Abstract: Bioinformatics education involves teaching the use of computer methods to solve biological problems. Bibliometrics is the quantitative study of academic literature, such as books, articles, and other publications. Moreover, the R-based package Bibliometrix is written in the R language, and offers many tools for quantitative bibliometric studies. Biblioshiny is a Shiny app providing a web-interface for Bibliometrix. In this study, we employed the Web of Science (WoS) Core Collection as our database and used Biblioshiny as our analysis tool to better understand bioinformatics education. A total of 369 documents from 106 sources between 2004 and 2023 were investigated, with a focus on annual scientific production, top journals, influential affiliations, and keyword trends. The results of this study can provide valuable insights into bioinformatics education.

1. Introduction

Bioinformatics is an interdisciplinary field that involves biology, computer science, information technology, physics, mathematics, and so on [1]. Bioinformatics education refers to the teaching of how to use computer methods for gathering, storing, and analyzing data to solve biological issues[2]. In the past few decades, there have been notable changes in the area of bioinformatics education. In the early days, this course mainly focused on programming and biological databases. However, today it also includes areas like machine learning, artificial intelligence, and big data analysis. For example, in the practical teaching of protein structure prediction at the College of Life Sciences, Anhui Normal University, China, we have added artificial intelligence methods, such as AlphaFold3 [3], RossTTAFold [4], trRosetta, and RaptorX. Due to the key role of bioinformatics education, it is urgent to conduct systematic research in this area.

Bibliometrics is the statistical analysis of academic literature, including books, articles, and other publications. These analytical methods come from mathematics, social sciences, and natural sciences. Nowadays, bibliometrics is widely used in research management and has become a truly interdisciplinary research field that covers nearly all scientific disciplines. Therefore, bibliometrics is crucial for evaluating the influence of academics.

In order to perform a better bibliometric analysis, the choice of databases and tools is essential. The databases include Web of Science (WoS), Scopus [5], Google Scholar, PubMed, Dimensions, Microsoft Academic, and CrossRef, with the first three being the most important data sources [6]. In

comparison with other databases, the WoS platform is owned and operated by Clarivate Analytics. It provides over 170 million records, including journals, books, and proceedings, and covers the Web of Science Core Collection, BIOSIS Citation Index, Data Citation Index, etc. Considering that our university can access WoS platform, we chose this commercial database for further analysis in this study. Furthermore, the tools of bibliometric analysis comprise the following types: (1) Java-based: CRExplorer, CiteSpace, Gephi, and VOSviewer; (2) Python-based: ScientoPyUI, pybliometrics, and python-bibtexparser; (3) R-based: Bibliometrix, biblionetwork, and cocorresp. Among these resources, the R-based Bibliometrix, available at http://www.bibliometrix.org, is written in the R language, and offers many tools for quantitative bibliometric studies [7]. Moreover, Biblioshiny is a shiny app providing a web-interface for bibliometrix.

In this study, we employed WoS as the database and Biblioshiny as the analysis tool, to address the following questions:

(1) What is the main information on bioinformatics education?
(2) Which sources and affiliations are most relevant in this field?
(3) How has authors' production changed over time?
(4) What insights can be gained from the Word Cloud analysis?
(5) What does the Country Network analysis reflect on international collaborations?

2. Materials and Methods

2.1. Document Selection

The Web of Science Core Collection on the WoS platform can search the world's leading publications, including Science Citation Index Expanded (SCI-EXPANDED, 1996-present), Social Sciences Citation Index (SSCI, 1996-present), Emerging Sources Citation Index (ESCI, 2019-present) and five other databases. In the present study, SCI-EXPANDED was selected as the search database, covering the period from 2004 to 2023. The first selection criterion for scientific documents was "Bioinformatics" OR "computational biology" (Topic), combined with "education" OR "teaching" OR "pedagogy" OR "curriculum" OR "Bioinformatics training" (Topic). The second selection criterion was Article OR Proceedings Paper OR Review (Document Type), along with English (Language). Finally, we removed records that were not highly relevant to bioinformatics education or lacked Keywords (DE) and Keywords Plus (ID).

2.2. Bibliometric Analysis

Biblioshiny is a powerful R-based tool that requires no coding from the user. It is developed in the Shiny environment (a user-friendly interface). Bibliometric analysis using Biblioshiny includes the following steps: (1) Run the latest RStudio (2024.04.2 Build 764); (2) Load bibliometrix with 'library(bibliometrix)'; (3) Load biblioshiny with 'biblioshiny()'; (4) Load Data: In the web interface, import a raw file from WoS (plain text format); (5) Perform an overview analysis, including Main information, Annual Scientific Production, Average Citations per Year, and Three-Field Plot; (6) Conduct other analyses, including Sources, Authors, Documents, and Clustering, etc.

3. Results and Discussion

3.1. Overview analysis

3.1.1 Main information on bioinformatics education

Through a systematic search in the SCI-EXPANDED of the Wos Core Collection database, a total

of 369 documents on bioinformatics education were selected from 106 sources. These documents have an average citation score of 8.37, and the number of authors is 1863. For detailed information, please see Table 1.

Table 1. Summary of the documents.

Description	Results
Timespan	2004-2023
Sources (Journals)	106
Documents	369
Document Average Age Average citations per docs	8.39 22.65
Author's Keywords (DE)	926
Keywords Plus (ID)	715
Authors	1863
International Co-Authorships %	19.51
Coauthors per docs	5.81

3.1.2 Annual scientific production and average citations per year

The data, from 2004 to 2023, shows the overall upward trend in annual scientific production on bioinformatics education (see Figure 1 (A)). Further analysis reveals that the number of publications was relatively low from 2004 to 2010. Subsequently, there was a significant increase, with a peak of 33 articles in 2021. In general, this growth trend indicates a rising interest in bioinformatics education over the years.

Furthermore, the average number of citations per year is calculated by dividing the total citations by the number of years since publication. From the Figure 1(B), we observed that the years 2006 and 2008 have notably high average citations per year. This may be due to a few highly influential articles published during those years.

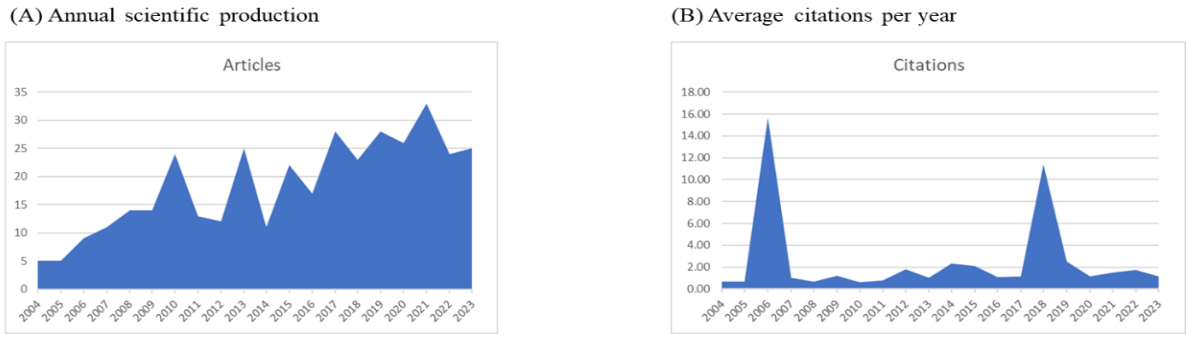

Figure 1: Annual scientific production (A) and average citations per year (B)

3.2. Most relevant sources and affiliations analyses

We listed the top 10 most relevant sources on the bioinformatics education in Figure 2. As the official journal of the international Union of Biochemistry and Molecular Biology (IUBMB), *Biochemistry and Molecular Biology Education* has published 95 relevant articles (ranked 1st), reflecting its significant influence in this area. It is noteworthy that *Briefings in Bioinformatics* (with 51 articles, ranked 2nd) and *PLOS Computational Biology* (with 27 articles, ranked 3rd), which are generally considered to publish scientific research, also include many articles related to education and training. Furthermore, three journals (*CBE-Life Sciences Education*, *American Biology Teacher*,

Journal of Biological Education) also contribute to the bioinformatics education. These three specialized education journals have played a key role in this field.

Furthermore, we conduct an analysis of the most relevant affiliations in bioinformatics education. The results show that the University of California System has the highest number of articles (32), followed by the University of California, Los Angeles (UCLA) with 22, and the European Molecular Biology Laboratory (EMBL) with 19. This finding emphasizes the core position of these institutions in the field of bioinformatics education.

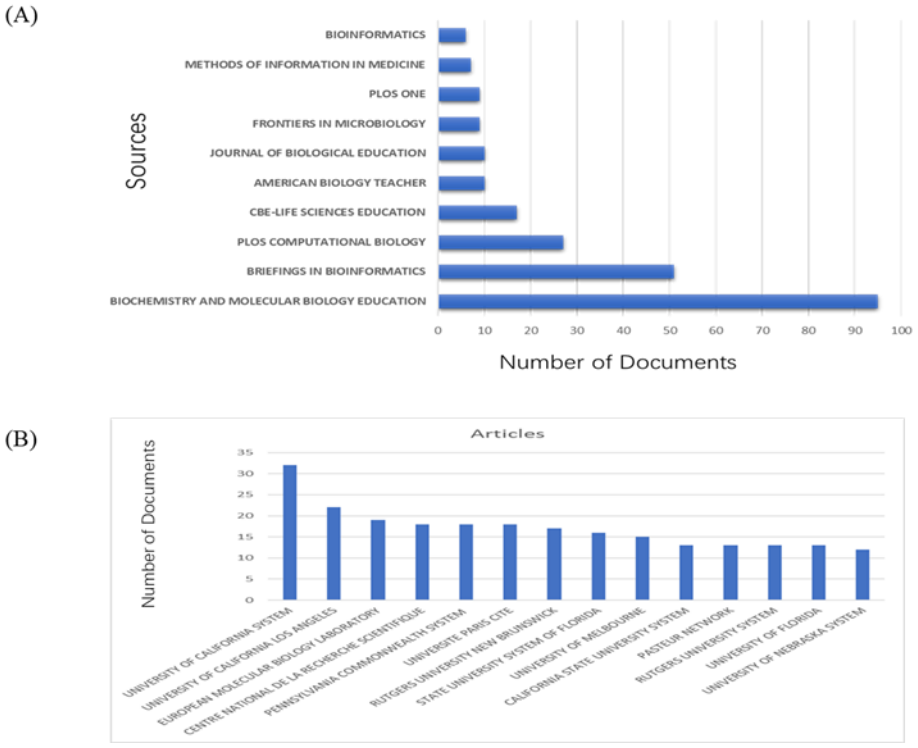

Figure 2: Analyses of most relevant sources (A) and most relevant affiliations (B)

3.3. Authors' production over time

To better understand the major contributors and trends, we analyzed the top authors' production over time. As depicted in Figure 3, the circle sizes denote the number of publications, while circle shading reflects yearly citation totals (TC/Y). Over the past two decades, a total of 62 papers have been published by the ten most prolific authors in the field of bioinformatics education, with peaks in 2013 (12 papers) and 2015 (11 papers). Schneider MV and colleagues from the European Bioinformatics Institute published 11 papers (2010–2019), focusing on evolving bioinformatics, data science training needs, infrastructure development, innovative programs, and teaching methods. Their publications are widely cited, with several exceeding 30 citations. Other notable contributors, such as Brazas MD and Attwood TK, have also authored impactful papers with high citation counts. All these works have greatly influenced bioinformatics education, providing key insights and resources for life science training.

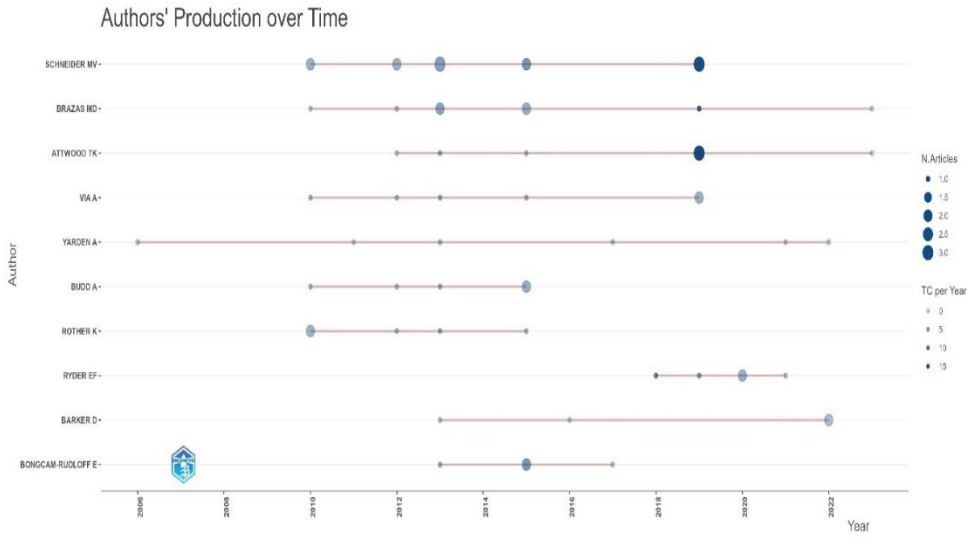

Figure 3: Top 10 authors' production over time

3.4. Keyword analysis

Word Cloud analysis can provide valuable insights into the key themes and emerging trends in their research area. In this study, we also conducted this type of analysis to illustrate the frequency of authors' keywords in bioinformatics education research. To better visualize the data, the word occurrences were adjusted using the square root method (see Figure 4). The results reveal that the most frequent word is "bioinformatics", appearing 117 times, followed by "computational biology" (43) and "education" (41). Key terms such as "genomics," "training," and "bioinformatics education" highlight the primary goal of integrating bioinformatics into education systems. Additionally, the presence of "laboratory exercises," "active learning," and "curriculum" indicates a strong focus on practical and instructional methods.

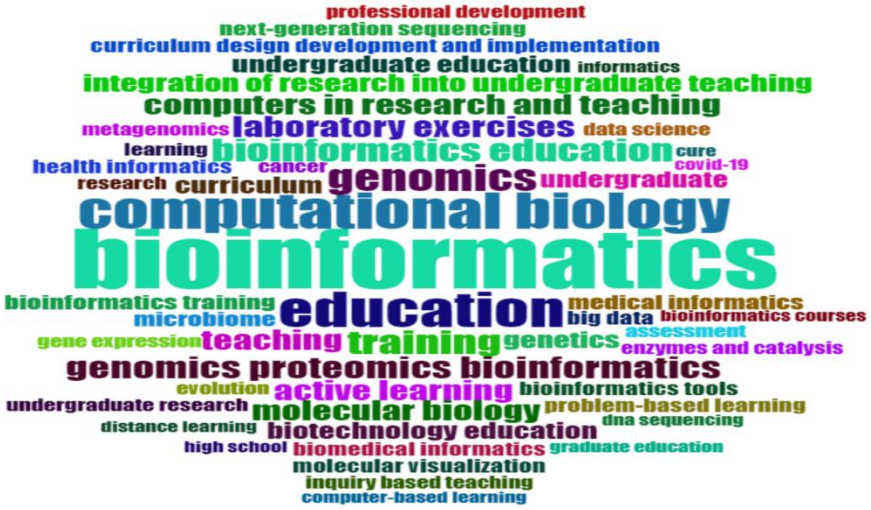

Figure 4: Word cloud of 50 author keywords

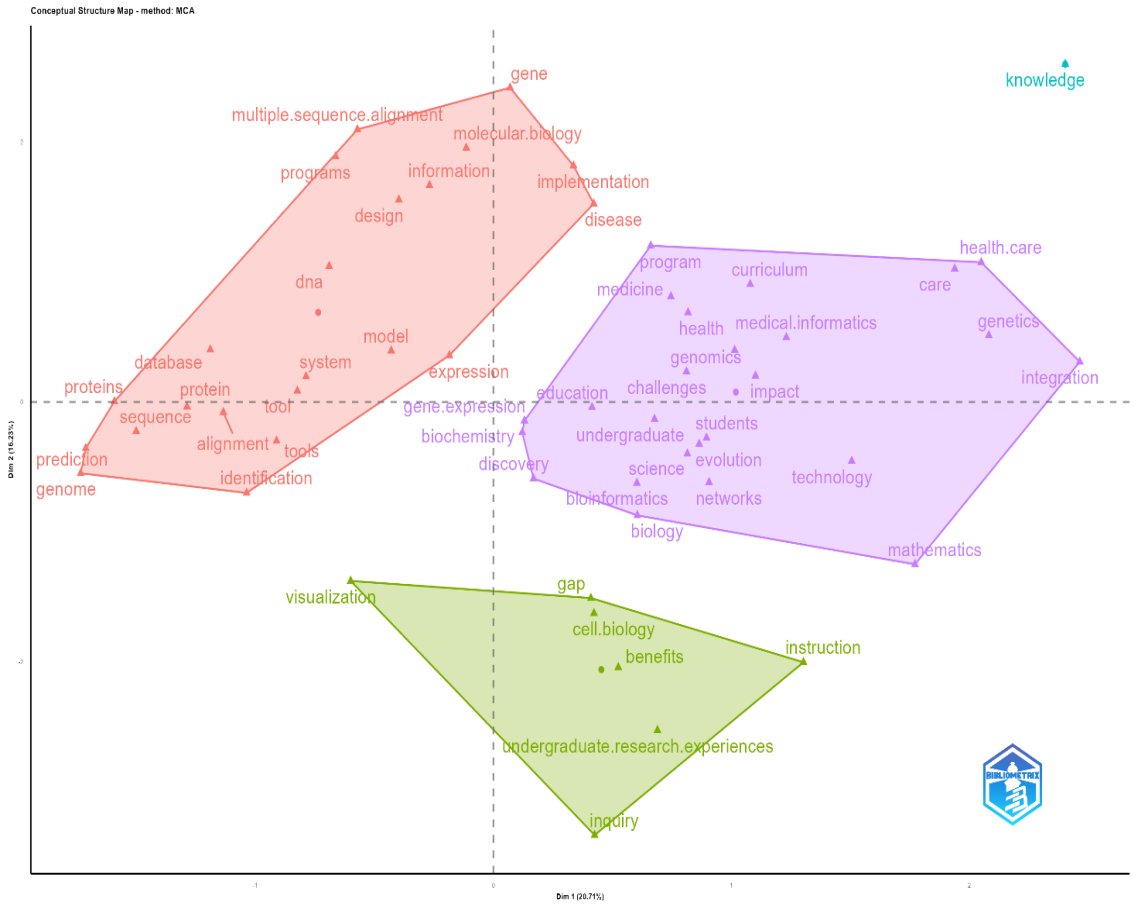

Figure 5: Factorial analysis of 50 author keywords

Moreover, to illustrate the conceptual structure of author keywords-plus in bioinformatics education, we performed a Factorial Analysis using Multiple Correspondence Analysis (MCA) in Biblioshiny. The results revealed three major clusters, each represented by a different color (see Figure 5). The red cluster includes keywords involved in Bioinformatics Fundamentals and Tools, such as genomics, proteins, sequence alignment, tool design, and molecular biology. The purple cluster comprises keywords focusing on Bioinformatics Applications and Impact, including medicine, disease, and health care. The green cluster emphasizes the Bioinformatics Education and Research, containing visualization, instruction, and research experiences. Overall, the diagram reveals the relationships between different research themes and activities within bioinformatics education, as represented by the spatial distribution of these keywords.

3.5. Country Network

We thoroughly analyzed the countries of the corresponding authors, and the results for the top 10 contributors are presented in Table 2. To assess international collaboration, the data are grouped into Single Country Publications (SCP) and Multiple Countries Publications (MCP). The USA has the highest number of publications (187), but also has the highest SCP rate, indicating a preference for internal research. In contrast, countries like the Netherlands and Israel exhibited higher rates of MCP, suggesting a greater emphasis on international collaboration. Furthermore, three countries, such as the United Kingdom, Germany, and Australia, made notable contributions to the total publication output and international collaboration. These results reveal the varied nature of research collaboration in bioinformatics education among different countries.

Table 2: Top 10 productive countries

Country	Articles	SCP	MCP	Freq	MCP_Ratio
USA	187	172	15	0.507	0.08
United Kingdom	26	20	6	0.07	0.231
Germany	18	11	7	0.049	0.389
Australia	13	11	2	0.035	0.154
China	10	6	4	0.027	0.4
Canda	9	8	1	0.024	0.111
Portugal	9	7	2	0.024	0.222
Israel	8	8	0	0.022	0
Netherlands	7	1	6	0.019	0.857
Spain	6	5	1	0.016	0.167

3.6. Conclusion

In this study, we employed the Web of Science (WoS) Core Collection as our database and used Biblioshiny as our analysis tool to better understand bioinformatics education. A total of 369 documents from 106 sources between 2004 and 2023 were investigated, with a focus on annual scientific production, top journals, influential affiliations, and keyword trends. These documents have an average citation score of 8.37, and the number of authors is 1863. Furthermore, the data shows the overall upward trend in annual scientific production on bioinformatics education. We observed that the years 2006 and 2008 have notably high average citations per year, possibly due to a few highly influential articles published during those years. The analysis also identified leading journals like *Biochemistry and Molecular Biology Education*, *Briefings in Bioinformatics*, and *PLOS Computational Biology* as key sources for bioinformatics education research. Additionally, the University of California System, UCLA, and EMBL emerged as influential institutions in this field. Over the past two decades, a total of 62 papers have been published by the ten most prolific authors in the field of bioinformatics education, with peaks in 2013 (12 papers) and 2015 (11 papers). Schneider MV and colleagues from the European Bioinformatics Institute published 11 papers (2010–2019). The word cloud analysis revealed that the most frequent word is “bioinformatics”, appearing 117 times, followed by “computational biology” (43) and “education” (41). The USA has the highest number of publications (187), but also has the highest SCP, indicating a preference for internal research. The results of this study can provide valuable insights into bioinformatics education.

Acknowledgements

This work was financially supported by Anhui Provincial Quality Engineering Project for Higher Education Institutions (2021xnfzxm040 and 2022jyxm538).

References

[1] J. B. Hagen (2000) The origins of bioinformatics. Nature Reviews Genetics. 1: 231-236.
[2] A. J. Magana, M. Taleyarkhan, D. R. Alvarado, M. Kane, J. Springer and K. Clase (2014) A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research. CBE-Life Sci. Educ. 13: 607-623.
[3] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard and J. Bambrick (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 630: 493-500.
[4] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch and R. D. Schaeffer (2021) Accurate prediction of protein structures and interactions using a three-track neural network. Science. 373: 871-876.

[5] M.-A. Vera-Baceta, M. Thelwall and K. Kousha (2019) Web of Science and Scopus language coverage. Scientometrics. 121: 1803-1813.
[6] M. Visser, N. J. Van Eck and L. Waltman (2021) Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. Quantitative science studies. 2: 20-41.
[7] M. Aria and C. Cuccurullo (2017) bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of informetrics. 11: 959-975.