

Scientometric Study of Bioinformatics Research based on Web of Science Database during 2010-2019

Dr. S. Baskaran

Assistant Librarian, Madras University Library,

University of Madras, Chennai 600 005

baskarnphd@gmail.com

ABSTRACT

The research paper is analyzing the global publication trends on bioinformatics research output based on the Web of Science database. During 2010-2019, the database contained 35,611 publications were published in the field. The average number of publications per year was 3561 and the highest number of publications 7043 was published in 2019. Zhang, Y is the most productive author with 444 (1.25%) publications followed by Wang, Y with 415 (1.17%) and Wang, J with 415 (1.17%). RGR is decreased from 0.75 in 2011 to 0.20 in 2018 and the Doubling time of the publications gradually increased from 0.92 in 2011 to 3.47 in 2018. Major source of publications on geophysics research is journal articles with 30,256 publications (84.96%).

KEYWORDS: Bioinformatics, annual growth rate, relative growth rate and doubling time

1 INTRODUCTION

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. Bioinformatics is an interdisciplinary field that develops methods and [software tools](#) for understanding biological data. As an inter-disciplinary field of science, bioinformatics combines [computer science](#), [statistics](#), [mathematics](#), and engineering to analyze and interpret [biological](#) data. Bioinformatics has been used for *insilico* analyses of [biological](#) queries using [mathematical](#) and statistical techniques.

Bioinformatics is both an [umbrella term](#) for the body of biological studies that use [computer programming](#) as part of their methodology, as well as a reference to specific analysis "pipelines" that are repeatedly used, particularly in the field of [genomics](#). Common uses of bioinformatics include the identification of candidate genes and [nucleotides](#) (SNPs). In developing country like India, bioinformatics has a key role to play in areas like

agriculture where it can be used for increasing the nutritional content, increasing the volume of the agricultural produce and implanting disease resistance etc. There are a large number of applications of bioinformatics in the fields of medicine, microbial genome applications, and agriculture. Therefore, the present study has been undertaken in order to know the growth and development of publications in the field of bioinformatics research as indexed in the web of science database.

2 OBJECTIVES FOR THE STUDY

The objective of the study was to perform a scientometric analysis of all bioinformatics publications in the world.

The parameters studied include:

- Form of Publications
- Annual Growth Rate, the compound growth rate of publications
- Most prolific authors
- Highly productive countries
- Highly productive institutes
- Most preferred source titles for publication
- Language-wise distribution of publications
- High productive subject areas

3 METHODOLOGY

The Web of Science database was used for retrieving data on bioinformatics during 2010-2019, using search terms namely 'bioinformatics' in 'topic filed'. Web of Science database is one of the very comprehensive bibliographic databases covering all aspects of science and technology. A total of 35611 publications were downloaded with no. of bibliographical data, the data were transferred to a spreadsheet application and analyzed the data as per the objectives of the study.

4 DATA ANALYSIS AND INTERPRETATIONS

4.1 Form of publications

Table 1 Form of publications

S. No.	Form of publications	No. of publications	Percentage
1	Journal articles	30256	84.96
2	Review	2888	8.11
3	Editorial Material	705	1.98
4	Meeting Abstract	695	1.95
5	Proceeding Papers	634	1.78
6	Book Chapter	243	0.68
7	Early Access	109	0.31
8	Letter	59	0.17
9	Book Review	22	0.06
Total		35611	100

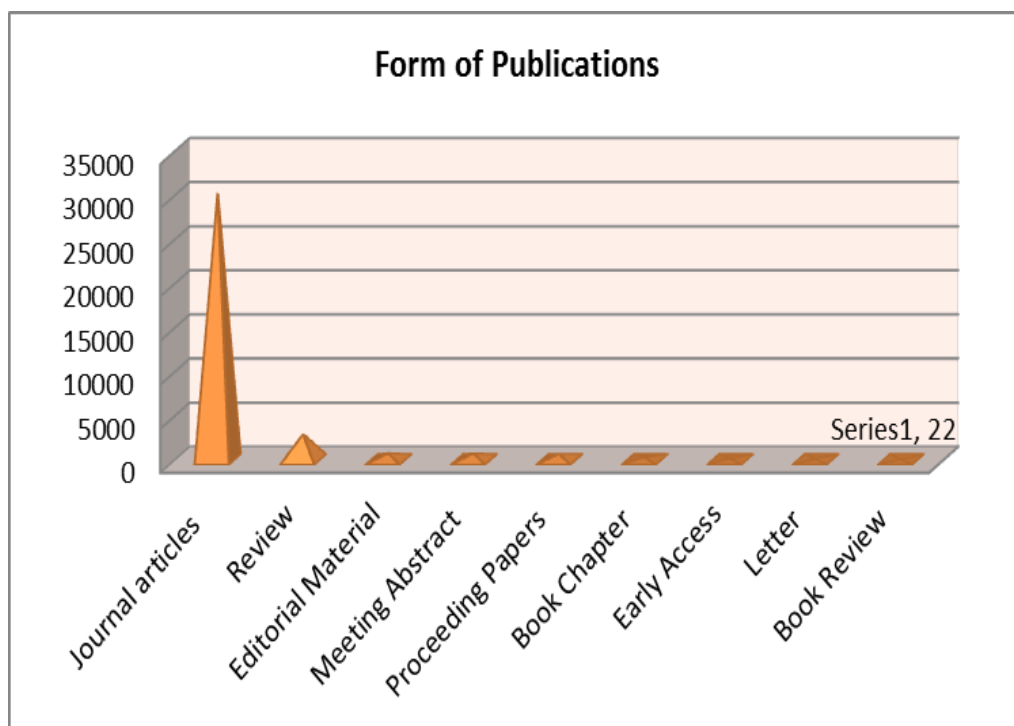


Figure 1 Form of publications

The table 1 reveals that the major source of publications covered by web of science databases on geophysics research is Journal Articles with 30,256 publications (84.96%) followed by Review with 2888 publications (8.11%). Editorial Material ranks the third position with 705 publications (1.98%), Meeting abstract with 695 publications (1.95%), Proceeding papers with 634 publications (1.78%) and remaining forms are less than one percentage as seen in the table. The results indicate that the research outputs on the subject of the period covered by the study are mostly published in the form of journal articles.

4.2 Growth of publications – RGR and DT

The Relative Growth Rate (RGR) is the increase in a number of articles or pages per unit of time. This definition derived from the definition of relative growth rates in the study of growth analysis in the field of bioinformatics. The mean relative growth rate (R) over the specific period of the interval can be calculated from the following equation.

Relative Growth Rate (RGR)

$$1 - 2R = \frac{\log W_2 - \log W_1}{T_2 - T_1}$$

Whereas

1-2 R- mean relative growth rate over the specific period of interval

$\log_e W_1$ - log of the initial number of articles

$\log_e W_2$ - log of the final number of articles after a specific period of interval

$T_2 - T_1$ - the unit difference between the initial time and the final time

The year can be taken here as the unit of time.

$$\text{Doubling Time (DT)} = 0.693/R$$

Table 2 Relative growth rate (RGR) and Doubling time (DT) of publications

Year	No. of Publications	Cumulative Total	W1	W2	RGR	DT
2010	1776	1776	-	7.48	-	-
2011	1967	3743	7.48	8.23	0.75	0.92
2012	2279	6022	8.23	8.70	0.47	1.47
2013	2633	8655	8.70	9.07	0.37	1.87
2014	3049	11704	9.07	9.37	0.30	2.31
2015	3384	15088	9.37	9.62	0.25	2.77
2016	3821	18909	9.62	9.85	0.23	3.01
2017	4395	23304	9.85	10.06	0.21	3.3
2018	5264	28568	10.06	10.26	0.20	3.47
2019	7043	35611	10.26	10.48	0.22	3.15

Table 2 indicates that the RGR is decreased from 0.75 in 2011 to 0.20 in 2018. The highest value 0.75 corresponds to the year 2011, whereas the lowest value 0.20 for the year 2018. Correspondingly, the Doubling Time of the publications gradually increased from 0.92 in 2011 to 3.47 in 2018. major source of publications covered by web of science databases on geophysics research is Journal Articles with 30,256 publications (84.96%)

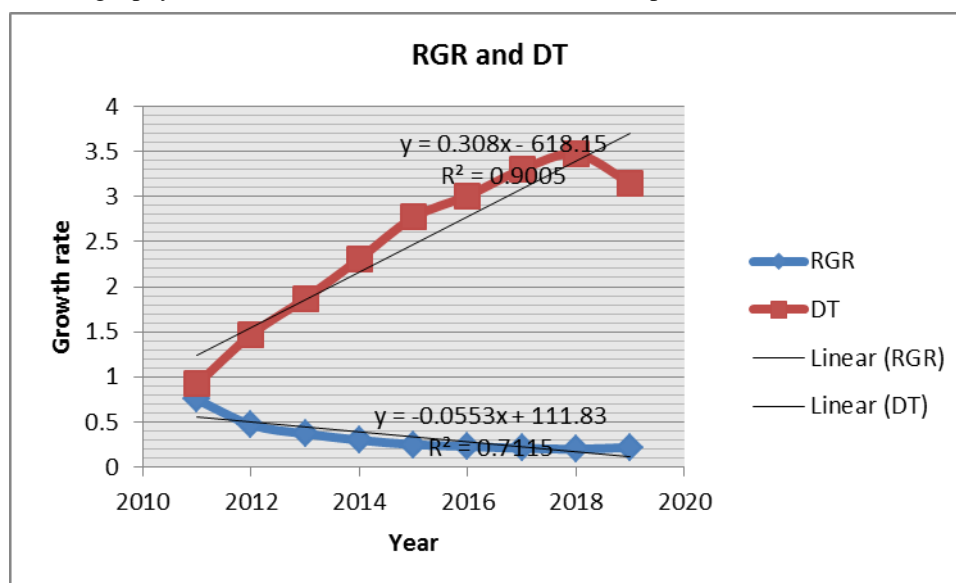


Figure 2 Relative growth rate and Doubling time

4.3 Most Prolific Authors

Table 3 Identification of Most Prolific Authors

Rank	Author	No. of publications	Percentage
1	Zhang Y	444	1.25
2	Wang Y	415	1.17
3	Wang J	367	1.03
4	Li Y	339	0.95

5	Li J	327	0.92
6	Wang L	283	0.79
7	Zhang J	283	0.79
8	Zhang L	280	0.77
8	Liu Y	272	0.76
10	Li L	255	0.72

The data on bioinformatics research publication during 10 years between 2010 and 2019 reveals that in total, 72,203 authors contributed to the publishing of the 35611 publications. The authors having 250 or more publications during 2010-2019 are shown in Table 3. Zhang, Y is the most productive author with 444 (1.25%) publications followed by Wang, Y with 415 (1.17%) publications, Wang, J with 367 (1.03%) publications, Li, Y with 339 (0.95%) publications, Li, J with 327 (0.92%) publications, Wang, L and Zhang, J each with 283 (0.79%) publications and Zhang, L with 280 (0.77%) publications, respectively.

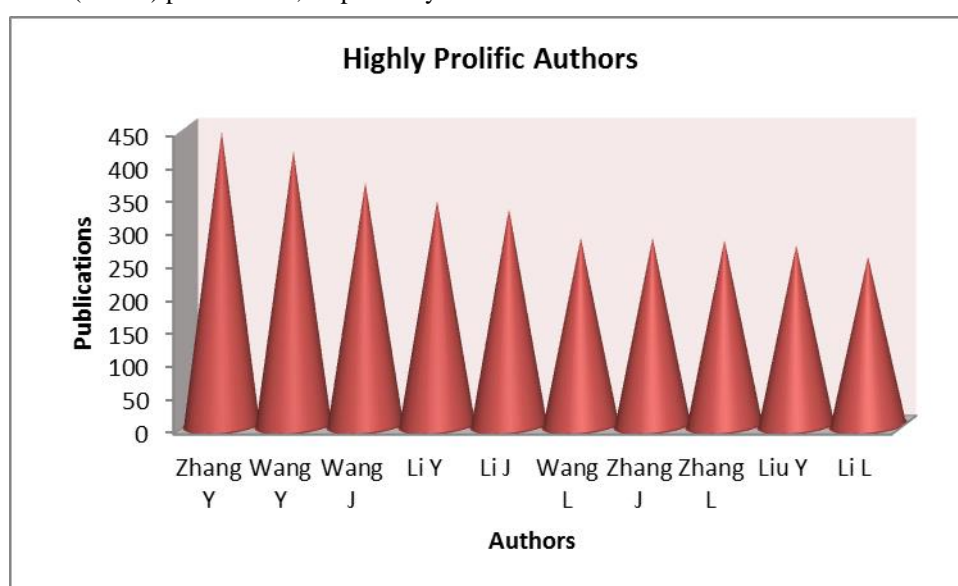


Figure 3 Highly prolific authors

4.4 Highly Productive institutions

Table 4 Highly Productive institutions

Rank	Institutions	Country	No. of Publications
1	University of California System	USA	1120 (3.15%)
2	Chinese Academy of Science	China	818 (2.30%)
3	Shanghai Jiao Tong University	China	640 (1.80%)
4	Centre National De La Recherche Scientifique CNRS	France	611 (1.72%)
5	Harvard University	USA	593 (1.67%)
6	University of Texas	USA	540 (1.52%)
7	National Institutes of Health NIH	USA	526 (1.48%)
8	Nanjing Medical University	China	513 (1.44%)
9	Fudan University	China	445 (1.25%)
10	University of London	UK	439 (1.23%)

Scientometric Study of Bioinformatics Research based on Web of Science Database during 2010-2019

A total of 15596 organizations are contributed entire research output of the study. The scientometric profile of top 10 institutions is presented in table 4. Findings revealed that University of California System, USA with 1120 (3.15%) publications is the most productive institutions in the field of bioinformatics research followed by Chinese Academy of Science, China with 818 (2.30%) publications, Shanghai Jiao Tong University, China with 640 (1.80%) publications, Centre National De La Recherche Scientifique CNRS, France with 611 (1.72%) publications, Harvard University, USA with 593 (1.67%) publications, University of Texas, USA with 540 (1.52%) publications, National Institutes of Health NIH, USA with 526 (1.48%) publications and Nanjing Medical University, China with 513 (1.44%) publications.

4.5 Highly Productive Countries

Table 5 Highly Productive Countries

Rank	Country	Total Publications (%)
1	China	13073 (36.71%)
2	USA	9765 (27.42%)
3	England	2188 (6.14%)
4	Germany	2158 (6.06%)
5	Canada	1416 (3.98%)
6	Italy	1382 (3.88%)
7	India	1378 (3.87%)
8	France	1295 (3.64%)
9	Australia	1196 (3.36%)
10	Spain	1096 (3.08%)

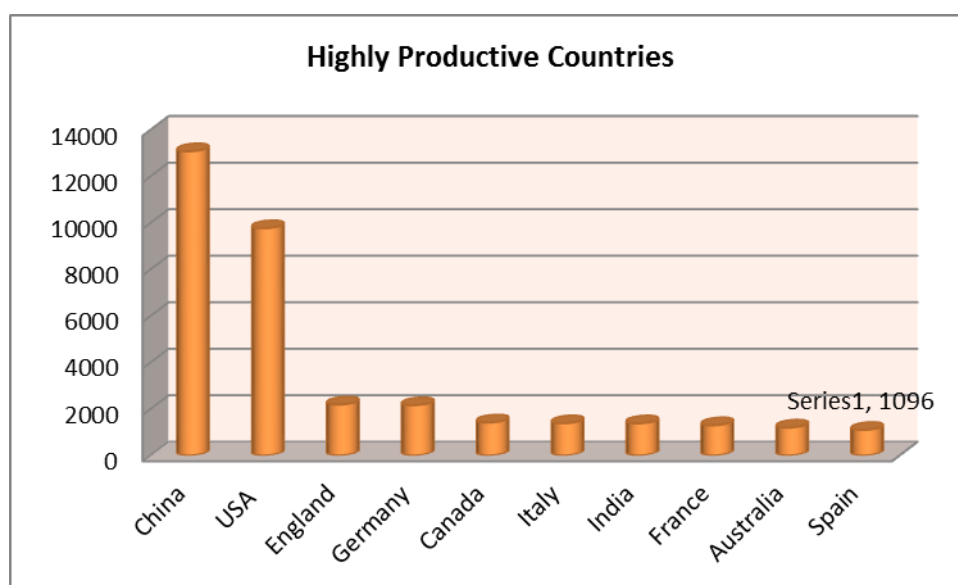


Figure 4 Highly productive countries

The publication share of highly productive countries (≥ 1000 publications) on bioinformatics is given in Table 5. In all, there were 148 countries involved in the research in bioinformatics; however, China topped the list with highest share (36.71%) of publications. USA ranked second with 27.42% share of publications followed by England 6.14% share of publications, Germany with 6.06% share of publications, Canada with 3.98% share of publications, Italy

with 3.88% share of publications and India with 3.87% share of publications and France with 3.64% share of publications.

4.6 Language-wise Distribution

Table 6 Language-wise Distribution

S. No.	Languages	Total Publications (%)
1	English	35466 (99.59%)
2	Chinese	52 (0.15%)
3	German	23 (0.06%)
4	Spanish	22 (0.06%)
5	French	20 (0.06%)
6	Portuguese	8 (99.59%)
7	Russian	5 (99.59%)
8	Japanese	4 (99.59%)
9	Hungarian	3 (99.59%)
10	Polish	3 (99.59%)
11	Korean	2 (99.59%)
12	Czech	1 (99.59%)
13	Danish	1 (99.59%)
14	Turkish	1 (99.59%)

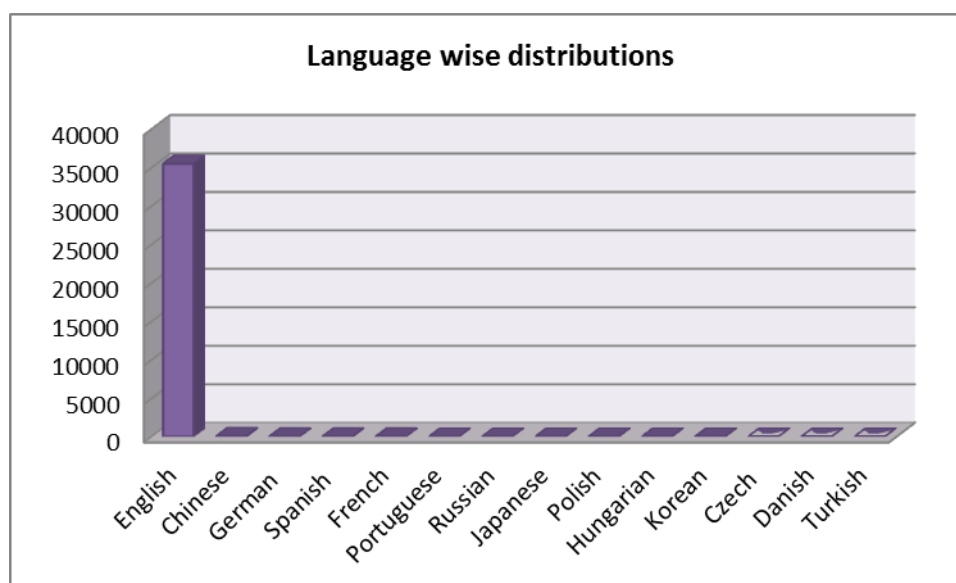


Figure 5 Language wise distributions

Publications on bioinformatics are spread over 14 languages. The study reveals that the maximum number of publications have been published in the English language with 35466 (99.59%) publications, followed by Chinese language with 52 (0.15%) publications, German language ranks third position with 23 (0.06%) publications, Spanish language with 22 (0.06%) publications, French language with 20 (0.06%) publications. And the remaining languages such as Portuguese, Russian, Japanese and other languages are constituted in negligible percentage. The English language superiority was found in every year in total productivity on the subject during the study period.

4.7 Major Source Title of Publications

Table 7 Source Title of Publications

Rank	Source Title	No. of Publications	Percentage
1	PLOS one	1074	3.02
2	BMC bioinformatics	1030	2.89
3	Bioinformatics	773	2.17
4	Scientific Reports	561	1.58
5	BMC Genomics	551	1.55
6	Nuclei Aids Research	459	1.29
7	Molecular medicine Reports	436	1.22
8	Proteomics	415	1.17
9	Oncotarget	336	0.94
10	Gene	317	0.89
11	Journal of Proteome Research	317	0.89

The publication share of most productive source titles (≥ 300 publications) on bioinformatics is given in Table 7. The scientific literature on bioinformatics is spread over 3733 different source journals and conference publications. It reveals that PLOS one the list with the highest number of publications 1074 (3.02%) followed by BMC bioinformatics with a share of 1030 (2.89%) publications. Bioinformatics occupies the third position with 773 (2.17%) publications. The fourth highest source title is Scientific Reports with 561 (1.58%) publications, BMC genomics with 551 (1.55%) publications and Nucleic acids research with 459 (1.29%) publications.

4.8 High Productivity Subject Areas

Table 8 High Productivity Subject Areas

Rank	Subject	No. of Articles	Percentage
1	Biochemistry molecular biology	10050	28.22
2	Biotechnology applied microbiology	4327	12.15
3	Oncology	3830	10.76
4	Genetics heredity	3488	9.79
5	Mathematical computational biology	3455	9.70
6	Computer science	2928	8.22
7	Science technology	2889	8.11
8	Cell biology	2783	7.82
9	Research Experimental medicine	2373	6.66
10	Pharmacology pharmacy	2373	6.66

Table 8 shows high productivity subjects which are contributing more than 2000 articles. It is found that Biochemistry molecular biology has the highest number of articles with 10050 (28.22%) followed by Biotechnology applied microbiology contributing 4327 (12.15%) articles. Oncology occupies the third position with 3830 (10.76%) articles. The fourth highest articles belonged to the subject Genetics heredity with 3488 (9.79%), Mathematical computational biology with 3455 (9.70%) and Computer science with 2928 (8.22%) articles respectively.

CONCLUSION

The study reveals that there is a quantum of information literature output available in the subject of agriculture, medicine, microbial genetics, and plant science. A number of research works are being carried out all over the world in this field. A total of 35611 publications were published on bioinformatics research during a 2010-2019 and average number of publications per year was 3561. The density of the research output during the year 2015 with 3389 publications. Chinese Academy of Science, China with 818 (2.30%) publications is the most productive institutions in the field of bioinformatics research followed by Shanghai Jiao Tong University, China with 640 (1.80%) publications. Chinese Academy of Science, China with 379 (1.84%) publications, is the most productive institutions in the field of bioinformatics research followed by Harvard University with 344 (1.67%) publications. Among 148 countries, China topped the list with highest share (36.71%) of publications followed by USA with 27.42% share of publications and England 6.14% share of publications. Among source titles, PLOS one the list with the highest number of publications 1074 (3.02%) followed by BMC bioinformatics with a share of 1030 (2.89%) publications. Publications on bioinformatics are spread over 14 languages. The maximum number of publications has been published in the English language with 35466 (99.59%) publications, followed by Chinese language with 52 (0.15%) publications.

REFERENCES

- [1] Gupta, B. M., Kaur, H., & Kshitig, A. (2012). Dementia research in India: A scientometric analysis of research output during 2002-11.
- [2] Santha Kumar R. (2016). Publications Trends in Nuclear Physics: A Global Perspective. *Library Philosophy and Practice (e-journal)*. Paper 1361. <http://digitalcommons.unl.edu/libphilprac/1361>
- [3] Santha Kumar R. (2016). Research Trends in Medical Physics: A Global Perspective. *Library Philosophy and Practice (e-journal)*. Paper 1362. <http://digitalcommons.unl.edu/libphilprac/1362>
- [4] Santha Kumar R & Kaliyaperumal K. (2015). A Scientometric Analysis of Mobile Technology Publications. *Scientometrics*, 105 (2): 921-939.
- [5] Santha Kumar R & Kaliyaperumal K. (2014). Mapping of Mobile Technology Publications: A Scientometric Approach. *DESIDOC Journal of Library and Information Technology*, 34(4): 298-303
- [6] <http://www.biotecharticles.com>
- [7] http://abc.cbi.pku.edu.cn/2can/genomes/eukaryotes/oryza_sativa.html