



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Bio-informatics Research Progress in the Post-genome Era Based on the Quantitative Analysis of SCIE

Yongqin ZHAN, Min YU

Library of Sichuan Agricultural University, Chengdu 611130, China

Abstract SCIE paper output can reflect the status quo and trend of discipline research and 7 038 scientific articles concerning bioinformatics are retrieved in SCIE database during the years between 2008 and 2012. Quantitative analysis of paper output and citation frequency are conducted according to nations, institutions, publications, research direction as well as hot articles, which provides assistance for bioinformatics researchers to understand the present situation of this subject, carry out cooperative studies and display scientific research achievements.

Key words Bioinformatics, Status quo, SCIE database, Quantitative analysis

1 Introduction

Bioinformatics is a newly emerging subject after the proliferation of computer technology in the 1980s. In 1995, American Human Genome offered bioinformatics a relatively comprehensive definition that bioinformatics is an interdisciplinary which includes the obtaining, processing, storing, analysis as well as explanation of biological information and uses various methods in maths, computer science and biology to explain and understand the biological meanings involved in the large data. Bioinformatics is the organic combination of three main scientific issues, namely "genome", "information structure" and "complexity", in nature science and technology. Its primary task is to discover "the complexity of genome information structure as well as fundamental rules of genetic languages"^[1-2].

The emergence of bioinformatics stems from the application of computer in biology at the end of the 1950s. Its development is closely connected with Human Genome Project and has experienced three stages^[2]. The initial stage is the pre-genome era which means the period from the end of the 1950s to the end of the 1980s with the main research direction focusing on the establishment of biology database, development and application of retrieval tools as well as comparison and analysis of DNA and protein sequence. The second stage is called the genome era with the main task focusing on determination and analysis of nucleotide sequence, discovery of new genes, exploration and application of large database based on the network and interface as well as extraction analysis of gene sequence information. Since the drawing of human genome mapping was completed in 2003, development of bioinformatics has stepped into the post-genome era. The main tasks in this stage involve researches on proteomics, epigenomics, disease phenomics human genome annotation. During this stage, how to obtain the genetic information of organism promptly and accurately still remains a significant issue in scientific research.

With the arrival of the post-genome information, researches on bioinformatics become deeper and achievements richer. Analysis on bioinformatics research progress is indispensable in further development of bioinformatics study and is beneficial to bioinformatics researchers' understanding the research situation, frontier and direction. Achievements of bioinformatics research progress are relatively abundant^[1-6], covering several special subjects, however, research methods are mostly limited in the qualitative analysis and quantitative analysis has not been widely adopted. Pritchard^[7] first put forward the concept of bibliometrical study which uses quantitative statistical method to analyze research achievements and predict the research tendency. Bibliometrical study is adopted mostly in fields like information science and medical science^[8,9]. Meanwhile, there are not enough research achievements on bioinformatics research progress with the method of bibliometrical study. Periodical theses are characterized with short publishing period, large information content, wide covering, high transmission speed and convenient retrieval, therefore studying periodical thesis promotes the understanding of research tendency of related subjects.

2 Materials and methods

2.1 Database and relevant tools Mainly adopting SCIE citation database as the retrieval object, all retrieved documents are classified with the statistical analysis function provided by the database. Then, evaluation database are established with Excel and data are recorded and analyzed.

2.2 Retrieval strategy In database of Web of Science, "bioinformatics" is inputted as the subject word with time span from January 1, 2008 to December 31, 2012 and retrieval time as December 31, 2012. Then, all periodical theses of bioinformatics collected in the SCIE database between January, 2008 and December, 2012 are retrieved.

3 Result analysis

3.1 Analysis of paper distribution by nation 7 038 periodi-

cal theses with "bioinformatics" as the subject word have been collected in SCIE during the five years between 2008 and 2012 and are ranked in amount of papers from different countries (Table 1). The top 10 countries possess 6 796 SCIE theses which account for 96.56% of total amount. These top 10 countries are the main power of bioinformatics and represent the research level of the current bioinformatics in the world. The remaining 242 SCIE theses are from the other 71 countries, each of which has at least one collected document and account for 3.44% of total amount. Therefore, SCIE papers of bioinformatics concentrate on some countries obviously.

Table 1 Distribution of SCIE Article Amount in the Top Ten Countries

Nation	Article Amount	Percentage %	Average Citation Frequency
America	2 531	35.962	12.23
China	1 298	18.443	4.26
Britain	629	8.937	13.21
Germany	609	8.653	11.98
Canada	342	4.859	9.43
France	322	4.575	8.25
Italy	295	4.192	7.25
Japan	276	3.922	16.11
Australia	249	3.538	9.30
India	245	3.481	3.25

Citation frequency can reflect not only the research level of scientific paper research achievements but also how the achievements are concerned worldwide. The more the papers are cited, the wider influence they have in the research fields. And average quotation frequency is a significant indicator to measure the research level. The total cites of 7 038 SCIE papers amount to 56 146 and average cites (hereinafter referred to as international average cited times) reach 7.98. Japan, Britain and America are the top three countries with average cited times as 16.11, 13.21 and 12.23 respectively and display mighty power in bioinformatics research. Besides, average cited times in Germany (11.98), Canada (9.43), Australia (9.30) and France (8.25) also overtake the average.

3.2 Analysis of paper distribution by institution Bioinformatics displays a trend of both centralization and decentralization in terms of scientific research institutions. There are relatively a larger number of research institutions involved in bioinformatics and bioinformatics articles from altogether 1 868 institutions are collected by SCIE. It can be found after analysis that the top ten countries ranked by amount enjoy relatively high academic prestige as well as grade (Table 2). 854 bioinformatics papers published by these ten institutions are collected by SCIE, accounting for 12.13% in total number. However, papers from the other 1 858 institutions represent 87.87% and institutes which have only 1 bioinformatics paper being collected by SCIE during the five years amount to 16. Among the ten institutions, four institutions which belong to America have altogether 352 SCIE papers, accounting for 41.22% of total SCIE bioinformatics papers in the top ten institutions, which highlights America's dominant position in bioinformatics research. Three Chinese institutions enter the top ten

ranking and have 300 SCIE papers, representing 35.19% of total bioinformatics papers in the ten institutions.

Table 2 Top Ten Institutions by SCIE Article Amount

Institution	Nation	Article Amount	Percentage %	Average Citation Frequency
Chinese Acad Sci	China	148	2.103	6.59
Harvard University	America	137	1.947	11.92
Shanghai Jiao Tong University	China	93	1.321	8.52
Stanford University	America	78	1.108	10.99
University of California San Diego	America	76	1.080	17.78
University of Manchester	Britain	71	1.009	8.14
Cambridge University	Britain	66	0.938	12.32
European Bioinformat Inst	EU	65	0.924	25.14
University of Washington	America	61	0.867	10.23
Fudan University	China	59	0.838	5.03

In terms of average citation frequency, European Bioinformat Inst from the European Union has the highest cited times (25.14), which is more than three times as much as international cited times (7.98). Average citation frequency of 8 institutions among the top ten is higher than the international level, which suggests that these ten institutions have might research power and play a leading role in bioinformatics research although SCIE article amounts in the ten institutions do not have absolute advantage. Therefore, bioinformatics papers published by these institutions are worth being concerned by bioinformatics researchers.

3.3 Analysis of article source journals 7 038 bioinformatics SCIE papers published during the five years between 2008 and 2012 are distributed in 810 kinds of SCIE source journals, among which the top ten journals have the most journal articles, reaching 1 878 and representing 26.68% of total amount. Average citation frequency of 6 kinds of journals is higher than that of total bioinformatics papers during the five years and Nucleic Acids Research has the highest average citation frequency, arriving at 22.06, which is almost three times of total average citation frequency. And Nucleic Acids Research has the highest 2012 impact factor among the ten journals. Comprehensive analysis of the impact factor in 2012 and average citation frequency of these ten journals (Table 3) reflects that compared with the other 800 kinds of journals, these ten journals can be regarded as authoritative core journals in bioinformatics research and the excellent platform for bioinformatics scientific researchers to exhibit their achievements to the world.

3.4 Analysis of paper research direction According to the classification of research direction in ISI database, 7 038 research bioinformatics papers collected in SCIE between 2008 and 2012 are distributed in 98 research directions, among which research papers in biochemistry and molecular biology are the most, reaching 2 811 and representing nearly 40% of total amount, followed closely by those in biotechnology and applied biology with the paper amount reaching 1 148 and representing 16.311%. The amount of papers in mathematics and computational biology ranks

third, amounting to 1 049 and accounting for 14.905%. The other research directions with paper amount ranking the top ten involve computer science, gene inheritance, mathematics, cell biology, chemistry, biophysics and so forth. However, bioinformatics research papers on cell biology have the highest average citation frequency, reaching 13.18, which suggests that bioinformatics research achievements on this direction have captured relatively higher attention. While, articles on chemistry have the least average citation frequency with only 4.4. Bioinformatics is an interdisciplinary involving biology, mathematics, computer science and so forth. It can be seen from Table 6 that average citation frequency of bioinformatics papers on mathematics and computer science is 6.47 and 4.98 respectively, higher than those of pure mathematics as well as computer science, which proves that achievements of bioinformatics researches with the application of mathematics and computer science can attract higher attention. Average citation frequency of bioinformatics papers on maths and computational science amounts to 7.10, which further justifies the point.

Table 3 Top Ten Journals by SCIE Article Amount

Journal	Article Amount	Percentage %	Average Citation Frequency	2012 Impact Factor
Plos One	380	5.399	4.57	4.092
BMC Bioinformatics	362	5.144	8.66	2.751
Bioinformatics	246	3.495	9.81	5.468
Nucleic Acids Research	231	3.282	22.06	8.026
Proteomics	165	2.344	7.48	4.505
BMC Genomics	141	2.003	9.18	4.073
Journal of Proteome Research	125	1.776	10.63	5.113
Briefings in Bioinformatics	77	1.094	7.71	5.202
Journal of Biological Chemistry	76	1.080	8.45	4.773
International Journal of Data Mining and Bioinformatics	75	1.066	1.63	0.427

3.5 Analysis of heated theses Based on the analysis of Essential Science Indicators, heated theses of bioinformatics researches can be classified into two categories. One is about the basic theo-

ries and methods of bioinformatics, developing more efficient calculations, exploration of supporting softwares as well as data integrating and resource processing. Due to the innovativeness and extensive application of these research achievements, these articles have captured wide attention and have relatively high citation frequency. Over the latest five years, 3 bioinformatics research papers which have the highest citation frequencies among the top ten belong to this category with the citation frequencies surpassing 1200 (Table 5). A case in point is that the article accomplished by four scientists from American University of Arizona and University of Pennsylvania^[10] reports their newly developed Molecular Evolutionary Genetics Analysis version 5 used for molecular evolutionary genetic analysis as well as bioinformatics data mining. The improved software is more accessible for beginners and superior in analyzing gene sequences, therefore it has captured high attention of bioinformatics researchers. The citation frequency of this article amounted to 2 154 within only 15 months from October, 2010 to December, 2012, which was the highest. Another kind of articles aims at some specific substance or molecular to understand its structure and function as well as business development and research. Theses in this category are the subject of bioinformatics research articles in SCIE database and their subjects are the main directions of bioinformatics researchers. Among the top ten by citation frequency, 6 articles belong to this category (Table 5).

Table 4 Top Ten Research Directions Funding Agencies by SCIE Article Amount

Research Directions	Article Amount	Percentage %	Average Citation Frequency
Biochemistry Molecular Biology	2 811	39.940	10.74
Biotechnology Applied Microbiology	1 148	16.311	8.18
Mathematical Computational Biology	1 049	14.905	7.10
Computer Science	930	13.214	4.98
Genetics Heredity	701	9.960	11.87
Science Technology Other Topics	536	7.616	7.67
Mathematics	449	6.380	6.47
Cell Biology	319	4.533	13.18
Chemistry	270	3.836	4.40
Biophysics	248	3.524	7.85

Table 5 Top Ten Papers Cited Frequency in SCIE

Ranking	Title	First Author	Institution	Journal	Publication Date	Citation Frequency
1	Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods	Tamura K	Arizona State Univ	Molecular Biology and Evolution	2011.10	2 154
2	Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources	Huang DW	NCI, Lab Immunopathogenesis & Bioinformat, Clin Serv Program, SAIC Frederick Inc	Nature Protocols	2008.01	1 736
3	A Rapid Bootstrap Algorithm for the RAxML Web Servers	Stamatakis A	Univ Munich	Systematic Biology	2008.01	1 286
4	AMPK phosphorylation of raptor mediates a metabolic checkpoint	Gwinn Dana M.	Salk Inst Biol Studies, Dulbecco Ctr Canc Res, Mol & Cell Biol Lab	Molecular Cell	2008.04	537

(Table 5)

Ranking	Title	First Author	Institution	Journal	Publication Date	Citation Frequency
5	Micro RNA – 21 (miR – 21) post – transcriptionally downregulates tumor suppressor Pdc4 and stimulates invasion, intravasation and metastasis in colorectal cancer	Asangani I. A	Univ Heidelberg	Oncogene	2008.04	431
6	I – TASSER server for protein 3D structure prediction	Zhang Y	Univ Kansas	BMC Bioinformatics	2008.01	385
7	The University Protein Resource (UniProt)	Bairoch A	European Bioinformat Inst	Nucleic Acids Re- search	2010.01	333
8	Protein secondary structure analyses from circulardichroism spectroscopy: Methods and reference databases	Whitmore, Lee	Univ London	Biopolymers	2008.05	307
9	The University Protein Resource (UniProt) 2009	Bairoch A	Univ Geneva	Nucleic Acids Re- search	2009.01	302
10	Rapid transcriptome characterization for anonmodel organism using 454 pyro – sequencing	Vera, J. Cristobal	Penn State Univ	Molecular Ecology	2008.04	266

4 Conclusion

Bioinformatics research is widely concerned throughout the world. Although bioinformatics research has stepped into the post-genome era, bioinformatics remains a new engineering and technology discipline and the post-genome era provides a wider stage. There are urgent issues to be solved in both theory and method and bioinformatics researchers are still confronted with opportunities as well as challenges.

References

[1] ZHAO Q, GU RS, DU SM. Research status and development trend of bioinformatics[J]. Journal of Medical Information, 2012, 3(5): 2 – 6. (in Chinese).

[2] CHEN M. Bioinformatics in post genome era[J]. China Journal of Bioinformatics, 2004, 2(2): 29 – 34. (in Chinese).

[3] NIU QW, CHEN JP. Review of common method of bioinformatics and application software[J]. Journal of Tropical Medicine, 2012, 12(7): 908 – 910. (in Chinese).

[4] CHEN YH, SUN K. Bioinformatics research progress and application[J]. Journal of Qinghai Normal University: Natural Science Edition, 2011(3): 69 – 72. (in Chinese).

[5] LI W, MAO M. New progress of bioinformatics—Review of the sixth national bioinformatics and genome research annual meeting[J]. Foreign Medical Sciences: Genetics, 1999, 22(2): 104 – 108. (in Chinese).

[6] TANG XQ, ZHU P. Developing trend of bioinformatics in post genome era [J]. China Journal of Bioinformatics, 2007(3): 142 – 145. (in Chinese).

[7] PRITCHARD A. Statistical bibliography or bibliometrics [J]. Journal of Documentation, 1969, 25(4): 348 – 349.

[8] TIAN YG, WEN C, HONG S. Global scientific production on GIS research by bibliometric analysis from 1997 to 2006 [J]. Journal of Informetrics, 2008, 2(1): 65 – 74.

[9] SWAPAN KUMAR PATRAA, PRAKASH CHAND. HIVAIDS research in India: A bibliometric study[J]. Library & Information Science Research, 2007, 29(1): 124 – 134.

[10] TAMURA K, DUDLEY J, NEI M, *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. Molecular Biology and Evolution, 2011, 28(10): 2731 – 2739.

(From page 148)

resort to administrative mediation first before the judicial proceedings. In the long run, judicial protection capacity should be gradually improved, experience of hearing such cases should be summarized timely, and professional trainings organized for judges to improve their law enforcement quality and protect benefits of right owners effectively.

References

[1] China's Ministry of Commerce. The analysis report of China agricultural products export in 2009[R]. (in Chinese).

[2] Japan's investigation of carnation from China[N]. Reference News, 2006 – 05 – 12. (in Chinese).

[3] SI YN. Guard against the Japanese new "species barrier[J]. Northwest Horticulture, 2004(2): 53 – 55. (in Chinese).

[4] GU HJ. Analysis on economy effect of Japan's amendment[J]. World Agriculture, 2005(3): 21 – 24. (in Chinese).

[5] Chinese Oil Network. China's soybean imports helped Monsanto rewarding [EB/OL]. (2001 – 04 – 07) www.cnyouzh.com/html/Articles/200

90407/39762.htm.

[6] China's Ministry of Commerce. The analysis report of China agricultural products export in 2009[R]. (in Chinese).

[7] HU RF, HUANG J, PRAY C, *et al.* Impacts of plant breeders rights on rice seed prices and variety adoption in China[J]. China Soft Science, 2006(1): 49 – 56. (in Chinese).

[8] ZHU FC, YIN X. Institution reflection on the reform of traditional agricultural technology extension system and its innovation[J]. Forum on Science and Technology in China, 2009(2): 117 – 121. (in Chinese).

[9] LIAO WH. National Intellectual Property Bureau: 99% of the enterprises did not apply for a patent in China[EB/OL]. (2010 – 02 – 23) <http://finance.people.com.cn/GB/1037/4335236.html>.

[10] HE XY. Investigation of intellectual property rights protection in Yangling demonstration zone of agricultural enterprises[J]. Shaanxi Journal of Agricultural Sciences, 2006(6): 134 – 136. (in Chinese).

[11] LIU LF, GUO XW. The protection status, problems and countermeasures of agriculture intellectual property in China[J]. China Science and Technology Achievements, 2007(15): 15 – 17. (in Chinese).

[12] QU SQ. On Chinese traditional culture and intellectual property[M]. Beijing: Intellectual Property Press, 2006. (in Chinese).

[13] LUO ZL. Research on intellectual property rights of agricultural crop variety[D]. Wuhan: Central China Agricultural University, 2006. (in Chinese).