Genomics and Applied Biology 2017, Vol.8, No.5, 30-48
http://gab.biopublisher.ca

Research Article **Open Access**

Visualization Map Analysis of Literature on Genomics Research Based on Bibliometrics of CiteSpace

Chunhe Zhao, Weixi Cui, Junsheng Meng, Xudong Zhang, Xiaoyu Zhao, Lei Wang, Zhanjun Dong
Hebei General Hospital, Hebei, 050051, China
Corresponding author email: tcmanalcwx@163.com, useforpubc@163.com

Genomics and Applied Biology, 2017, Vol.8, No.5 doi: 10.5376/gab.2017.08.0005
Received: 13 Nov., 2017
Accepted: 19 Dec., 2017
Published: 29 Dec., 2017

Preferred citation for this article:
Zhao C.H., Cui W.X., Meng J.S., Zhang X.D., Zhao X.Y., Wang L., and Dong Z.J., 2017, Visualization map analysis of literature on genomics research based on bibliometrics of CiteSpace, Genomics and Applied Biology, 8(5): 30-48 (doi: 10.5376/gab.2017.08.0005)

Abstract In this paper, we compared the domestic and foreign research situation in the field of genomics from the perspective of bibliometrics, aiming to provide some references for further study. We retrieved genomics related literature through the Web of Science and CNKI from 1985 to 2016, and then analyzed the number of published articles, the networks of collaboration among authors, research institutes and countries, the co-occurrence of keywords and co-citations of the retrieved literature through the software CiteSpace 4.0.R5 SE. The number of published periodical articles develops in fluctuation and shows a rising trend in general. The USA holds the largest share of world publications in genomics, the following is the UK and China in order. A core institution with considerable scale has formed in America and good cooperative relationships have been established between universities in the institution. Chinese Academy of Sciences is the main research institution in China. The cooperation between the international institutions is more often and closer than that at home. The researchers with the most articles published in Chinese and foreign journals are Liu J.Q. and Zhang Y respectively. The main co-occurrence and burst keywords are Pharmacogenomics, Functional genomics, Comparative genomics, Proteomics, etc. The articles with the highest centrality, highest citation frequency and highest burst strength are (Yang, 2007), (Lander, 2001) and (Altschul, 1997), respectively. Interagency cooperation should be enhanced so as to push forward the development of genomics research in China.

Keywords Genomics; CiteSpace; Bibliometrics; Visualization analysis

Background

Genomics is a discipline which studies on biological genomes to solve the practical problems. It was firstly put forward by Thomas Roderick in 1986. It covers genome mapping, sequencing, gene tagging and functional analyzing of the whole genomes (Li and Yin, 2000). Genomics provides information on biological genomes and systematic utilization of related data in order to work out the significant issues in the fields of biology, industry and medicine. We retrieved genomics related literature through CNKI and the Web of Science that have been published since 1985, and then analyzed the authors, institutions, collaboration among institutes and that among countries, keywords burst and co-occurrence, and co-citations of the retrieved literature through the visualization analysis software CiteSpace 4.0.R5 SE. We aims to explore the research hot spots and fronts to provide some references for further study on genomics.

CiteSpace is a citation visualization analysis software developed by Java language, focusing on analyzing the potential knowledge contained in scientific analysis. Using pathfinder network scaling and co-citation analysis theory, CiteSpace measures the literature (collection) of a specific domain and explores the key documents that will or may play an important role on the evolution of a certain knowledge domain or scientific field, which will be the intellectual turning point of the domain or the creative path that may affect the development of the domain. Through the knowledge structure and its transformation displayed by visualization network map of citing articles and its references or keywords, CiteSpace reflects the dynamic evolution of research hot spots in the certain field. So far, CiteSpace has been widely used in mapping and file management, management science and engineering, education, public administration, sociology, sport science, basic medicine, theoretical economics, philosophy, business administration, biology, applied economics, history of science and technology, psychology and other fields. CiteSpace serves as a valuable scientific research tool for the studies of the above fields (Chen, 2004; Chen, 2006; Chen et al., 2010; Chen, 2012; Chen and Leydesdorff, 2013; Chen et al., 2014).

1 Results and Analysis

1.1 Publication amount analysis

According to the retrieving results in the Web of Science and CNKI, we analyzed the number of published genomics related articles (from 1985 to 2016) and made a list of them (Table 1).

Table 1 The number of published periodical articles (piece)

Year	Chinese literature	Foreign literature
1985	0	0
1986	0	0
1987	0	0
1988	0	1
1989	0	1
1990	0	2
1991	0	7
1992	0	7
1993	0	10
1994	0	9
1995	0	23
1996	0	45
1997	6	101
1998	19	199
1999	49	348
2000	135	655
2001	292	931
2002	325	1280
2003	434	1494
2004	389	1886
2005	331	2139
2006	459	2216
2007	490	2343
2008	575	2570
2009	571	2633
2010	352	2905
2011	483	3051
2012	349	3343
2013	956	3498
2014	1132	3655
2015	1037	3982
2016	128	1185
Total	8512	40519

The number of published genomics related articles is increasing year by year (Figure 1). It reaches almost as high as 4000 pieces in 2015, which indicates that more and more researchers and researching institutes abroad have attached great importance to genomic studies and have obtained abundant research achievement. Compared with that, the number of published articles in China shows an upward trend in fluctuation since 1997. Moreover, there is sudden increase from 2013 to 2015, which illustrates that our country has increased the investment in genomics related fields in recent years, and has made great progress.

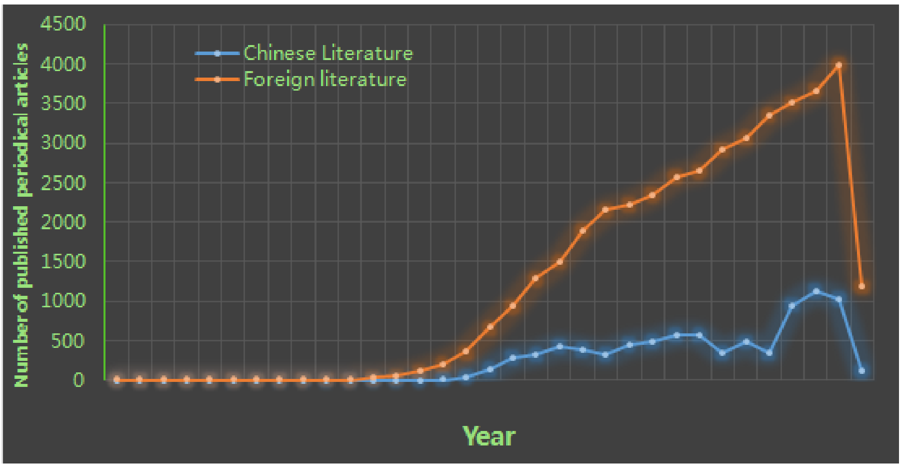
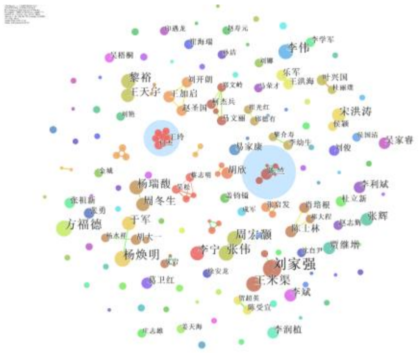


Figure 1 The number of published periodical articles (piece)

1.2 Visualized collaboration map analysis

1.2.1 Co-authorship network analysis

Table 2 shows the main authors that have published more than 10 Chinese-language articles in the field of genomics from 1985 to 2016. Jiaqiang Liu ranks the first with 29 published papers. The visualized co-authorship network map (Figure 2) shows that there are totally 168 authors and 78 items of collaboration among them, which has formed co-authorship networks represented by Jiaqiang Liu and Miqu Wang, Wei Zhang and Honghao Zhou, Huanming Yang and Jun Yu, Dongsheng Zhou and Ruifu Yang, Hongtao Song and Ying Hou, Yu Li and Tianyu Wang, Shilin Chen, Dacheng Hao and Peigen Xiao, Jiaqi Wang, Shengguo Zhao and Kailang Liu, Wenli Ma, Jiebing Ke and Wenling Zheng, Chaoying He and Shouyi Chen, Song Wu and Zhiming Cai, Guanghong Cui and De Qiu, Jieshou Li and Yousheng Li, Jun Yue and Honghai Wang, Xingguo Ye and Lipu Du, Ling Wang and Jie Fu, and Zhu Chen, etc.



Freq	Author	Author	Freq	Author	Author	Freq	Author	Author
29	刘家强	Jiaqiang Liu	7	乐军	Jun Le	5	吴梧桐	Wutong Wu
19	方福德	Fude Fang	7	易家康	Jiakang Yi	4	蔡志明	Zhiming Cai
18	杨焕明	Huanming Yang	7	陈竺	Zhu Chen	4	刘娜	Na Liu
16	王米渠	Miqu Wang	7	葛卫红	Weihong Ge	4	金城	Cheng Jin
16	周宏灏	Honghao Zhou	6	杜立新	Lixin Du	4	姜天海	Tianhai Jiang
16	张伟	Wei Zhang	6	叶兴国	Xingguo Ye	4	沈岩	Yan Shen
15	李伟	Wei Li	6	张祖新	Zuxin Zhang	4	庄志雄	Zhixiong Zhuang
14	黎裕	Yu Li	6	刘开朗	Kailang Liu	4	赵寿元	Shouyuan Zhao
13	杨瑞馥	Ruifu Yang	6	赵圣国	Shengguo Zhao	4	郝大程	Dacheng Hao
13	李宁	Ning Li	6	肖培根	Peigen Xiao	4	刘艳	Yan Liu
12	宋洪涛	Hongtao Song	6	张勇	Yong Zhang	4	成军	Jun Cheng
12	于军	Jun Yu	6	胡大一	Dayi Hu	4	杨水祥	Shuixiang Yang
11	周冬生	Dongsheng Zhou	6	刘俊	Jun Liu	4	邱德有	Deyou Qiu
11	王天宇	Tianyu Wang	6	王洪海	Honghai Wang	4	付杰	Jie Fu
10	张辉	Hui Zhang	5	盖钧镒	Junyi Ge	4	侯国清	Guoqing Hou
9	贾继增	Jizeng Jia	5	李学军	Xuejun Li	4	沈自尹	Ziyin Shen
9	李斌	Bin Li	5	马文丽	Wenli Ma	4	徐安龙	Anlong Xu
8	李润植	Runzhi Li	5	李幼生	Yousheng Li	4	郑文岭	Wenling Zheng
8	陈士林	Shilin Chen	5	王玲	Ling Wang	4	黎介寿	Jieshou Li
8	胡欣	Xin Hu	5	柯杰兵	Jiebing Ke	4	吴松	Song Wu
8	李利斌	Libin Li	5	崔海瑞	Hairui Cui	4	孙洁	Jie Sun
8	王加启	Jiaqi Wang	5	侯颖	Ying Hou	4	杜丽璞	Lipu Du
8	吴家睿	Jiarui Wu	5	张启发	Qifa Zhang	4	贺超英	Chaoying He
			5	陈受宜	Shouyi Chen	4	马荣才	Rongcai Ma
			5	赵志辉	Zhihui Zhao	4	印遇龙	Yulong Yin
						4	崔光红	Guanghong Cui

Figure 2 The visualized co-authorship networks of Chinese-language articles

Table 2 The main authors of the Chinese-language articles

Number of published articles	Researchers
29	Jiaqiang Liu
19	Fude Fang
18	Huanming Yang
16	Miqu Wang
16	Honghao Zhou
16	Wei Zhang
15	Wei Li
14	Yu Li
13	Ruifu Yang
13	Ning Li
12	Hongtao Song
12	Jun Yu
11	Dongsheng Zhou
11	Tianyu Wang
10	Hui Zhang

Table 3 shows the main authors that have published more than 50 foreign-language articles in the field of genomics from 1985 to 2016. Zhang Y. ranks the first with 149 published papers. The visualized co-authorship network map (Figure 3) shows that there are totally 190 authors and 130 items of collaboration among them, which indicates that the cooperation between the international researchers is stronger and closer than that in China. The scale of co-authorship networks is relatively large which are formed by Zhang Y., Wang J., Liu L., Li H., Wang L., Wang Y., Lee S.H., Varshney R.K., Li J., Wang W., Zhang J., Li X., Liu Y., Li Y., Wang X., Wang Q., Katze M.G., and Kumar S., etc, and the output of articles is the largest.

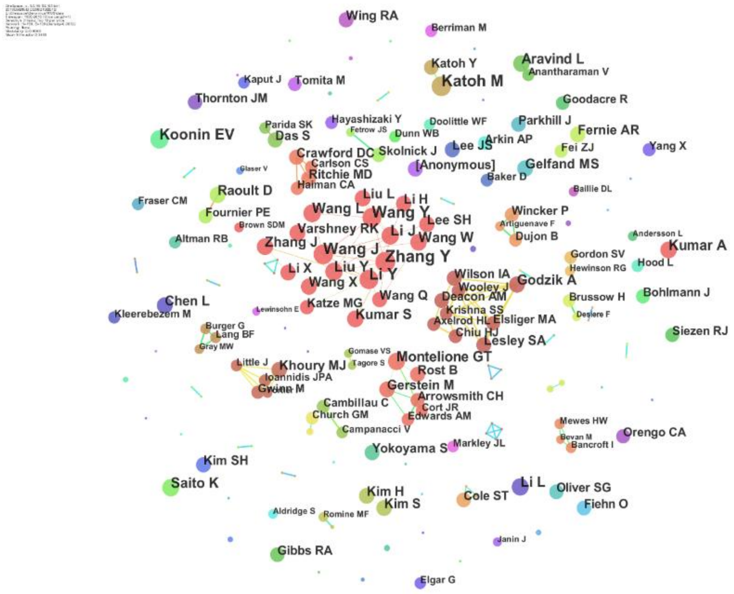


Figure 3 The visualized co-authorship networks of foreign-language articles

Table 3 The main authors of foreign-language articles

Number of published articles	Researchers
149	Zhang Y.
136	Wang J.
131	Katoh M.
122	Wang Y.
122	Li Y.
97	Koonin EV.
85	Li J.

Continued Table 3

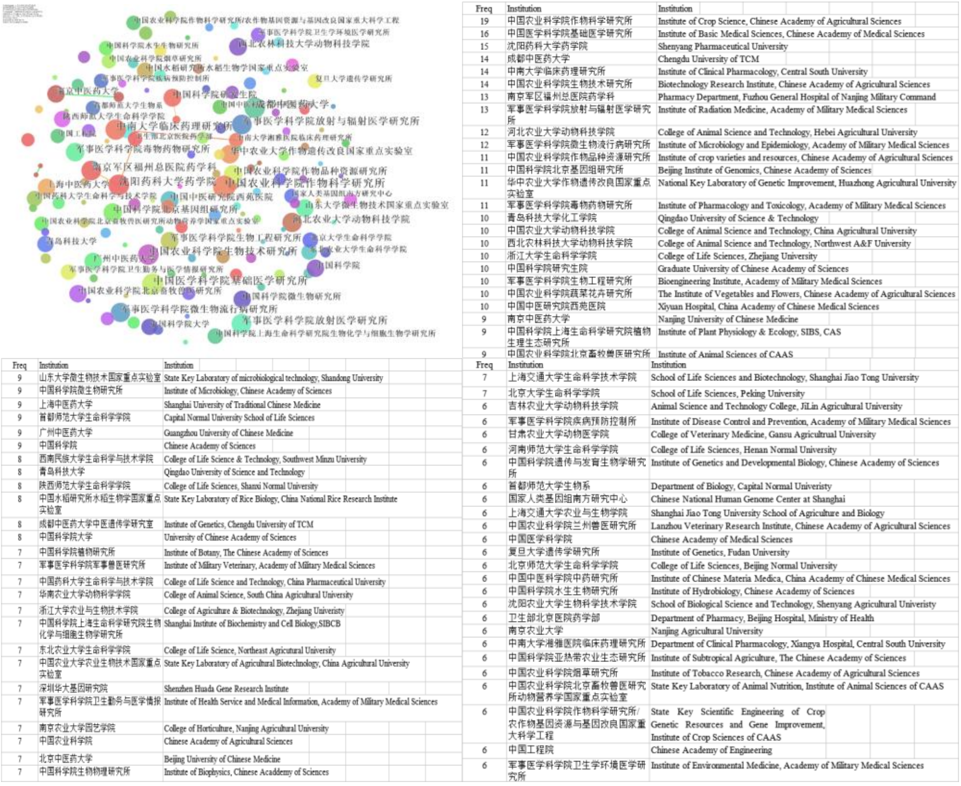
Number of published articles	Researchers
82	Wang L.
80	Liu Y.
77	Li L.
76	Montelione G.T.
75	Saito K.
67	Wang W.
67	Kumar A.
66	Aravind L.
62	Zhang J.
61	Kumar S.
60	Godzik A.
59	Liu L.
58	Chen L.
58	Raoult D.
55	Gelfand M.S.
55	Wang X.
54	Varshney R.K.
53	Khoury M.J.
53	Li X.
52	Li H.
51	Lee S.H.
50	Kim S.

1.2.2 Co-institution network analysis

The paper analyzed the research institutes of the first authors in studied articles in China and here are the top 10 research institutes (Table 4). Institute of Crop Science of Chinese Academy of Agricultural Sciences published the most articles in China. The visualized institute collaboration network map (Figure 4) shows that there are totally 153 institutes and 9 items of collaboration among them (1 Year Per Slice, Select top 10 from each slice), compared with 106 institutes and 59 items (Figure 5) respectively at abroad (1 Year Per Slice, Select top 10 from each slice), which indicates that the cooperation at abroad is much stronger and the number of programs is far more than that at home. There are 6 American institutes in the top 10 list of foreign institutes (Table 5) and all of them are universities, which shows that a core institution with considerable scale has formed in America. Chinese Academy of Sciences, the main research institution of genomics in China, ranks the second on the list.

Table 4 The top 10 domestic institutions in the number of published articles

Rank	Institutions	Number of published articles	First published year
1	Institute of Crop Science, Chinese Academy of Agricultural Sciences	19	2007
2	Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences	16	1997
3	Shenyang Pharmaceutical University	15	2009
4	Chengdu University of TCM	14	2001
5	Institute of Clinical Pharmacology, Central South University	14	2005
6	Biotechnology Research Institute, Chinese Academy of Agricultural Sciences	14	2003
7	Pharmacy Department, Fuzhou General Hospital of Nanjing Military Command	13	2009
8	Institute of Radiation Medicine, Academy of Military Medical Sciences	13	2006
9	College of Animal Science and Technology, Hebei Agricultural University	12	2007
10	Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences	12	2002



Freq	Institution	Institution
19	中国农业科学院作物科学研究所	Institute of Crop Science, Chinese Academy of Agricultural Sciences
16	中国医学科学院基础医学研究所	Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences
15	沈阳药科大学药学院	Shenyang Pharmaceutical University
14	成都中医药大学	Chengdu University of TCM
14	中南大学临床药理研究所	Institute of Clinical Pharmacology, Central South University
14	中国农业科学院生物技术研究所	Biotechnology Research Institute, Chinese Academy of Agricultural Sciences
13	南京军区福州总医院药学科	Pharmacy Department, Fuzhou General Hospital of Nanjing Military Command
13	军事医学科学院放射与辐射医学研究所	Institute of Radiation Medicine, Academy of Military Medical Sciences
12	河北农业大学动物科技学院	College of Animal Science and Technology, Hebei Agricultural University
12	军事医学科学院微生物流行病研究所	Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences
11	中国农业科学院作物品种资源研究所	Institute of crop varieties and resources, Chinese Academy of Agricultural Sciences
11	中国科学院北京基因组研究所	Beijing Institute of Genomics, Chinese Academy of Sciences
11	华中农业大学作物遗传改良国家重点实验室	National Key Laboratory of Genetic Improvement, Huazhong Agricultural University
11	军事医学科学院毒物药物研究所	Institute of Pharmacology and Toxicology, Academy of Military Medical Sciences
10	青岛科技大学化工学院	Qingdao University of Science & Technology
10	中国农业大学动物科技学院	College of Animal Science and Technology, China Agricultural University
10	西北农林科技大学动物科技学院	College of Animal Science and Technology, Northwest A&F University
10	浙江大学生命科学学院	College of Life Sciences, Zhejiang University
10	中国科学院研究生院	Graduate University of Chinese Academy of Sciences
10	军事医学科学院生物工程研究所	Bioengineering Institute, Academy of Military Medical Sciences
10	中国农业科学院蔬菜花卉研究所	The Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences
10	中国中医研究院西苑医院	Xiyuan Hospital, China Academy of Chinese Medical Sciences
9	南京中医药大学	Nanjing University of Chinese Medicine
9	中国科学院上海生命科学研究院植物生理生态研究所	Institute of Plant Physiology & Ecology, SIBS, CAS
9	中国农业科学院北京畜牧兽医研究所	Institute of Animal Sciences of CAAS

Freq	Institution	Institution
9	山东大学微生物技术国家重点实验室	State Key Laboratory of microbiological technology, Shandong University
9	中国科学院微生物研究所	Institute of Microbiology, Chinese Academy of Sciences
9	上海中医药大学	Shanghai University of Traditional Chinese Medicine
9	首都师范大学生命科学学院	Capital Normal University School of Life Sciences
9	广州中医药大学	Guangzhou University of Chinese Medicine
9	中国科学院	Chinese Academy of Sciences
8	西南民族大学生命科学与技术学院	College of Life Science & Technology, Southwest Minzu University
8	青岛科技大学	Qingdao University of Science and Technology
8	陕西师范大学生命科学学院	College of Life Sciences, Shanxi Normal University
8	中国水稻研究所水稻生物学国家重点实验室	State Key Laboratory of Rice Biology, China National Rice Research Institute
8	成都中医药大学中医遗传学研究室	Institute of Genetics, Chengdu University of TCM
8	中国科学院大学	University of Chinese Academy of Sciences
7	中国科学院植物研究所	Institute of Botany, The Chinese Academy of Sciences
7	军事医学科学院军事兽医研究所	Institute of Military Veterinary, Academy of Military Medical Sciences
7	中国药科大学生命科学与技术学院	College of Life Science and Technology, China Pharmaceutical University
7	华南农业大学动物科学学院	College of Animal Science, South China Agricultural University
7	浙江大学农业与生物技术学院	College of Agriculture & Biotechnology, Zhejiang Univeristy
7	中国科学院上海生命科学研究院生物化学与细胞生物学研究所	Shanghai Institute of Biochemistry and Cell Biology,SIBCB
7	东北农业大学生命科学学院	College of Life Science, Northeast Agricutural University
7	中国农业大学农业生物技术国家重点实验室	State Key Laboratory of Agricultural Biotechnology, China Agricultural University
7	深圳华大基因研究院	Shenzhen Huada Gene Research Institute
7	军事医学科学院卫生勤务与医学情报研究所	Institute of Health Service and Medical Information, Academy of Military Medical Sciences
7	南京农业大学园艺学院	College of Horticulture, Nanjing Agricultural University
7	中国农业科学院	Chinese Academy of Agricultural Sciences
7	北京中医药大学	Beijing University of Chinese Medicine
7	中国科学院生物物理研究所	Institute of Biophysics, Chinese Acaddemy of Sciences

Freq	Institution	Institution
7	上海交通大学生命科学技术学院	School of Life Sciences and Biotechnology, Shanghai Jiao Tong University
7	北京大学生命科学学院	School of Life Sciences, Peking University
6	吉林农业大学动物科技学院	Animal Science and Technology College, JiLin Agricultural University
6	军事医学科学院疾病预防控制所	Institute of Disease Control and Prevention, Academy of Military Medical Sciences
6	甘肃农业大学动物医学院	College of Veterinary Medicine, Gansu Agricultrual University
6	河南师范大学生命科学学院	College of Life Sciences, Henan Normal University
6	中国科学院遗传与发育生物学研究所	Institute of Genetics and Developmental Biology, Chinese Academy of Sciences
6	首都师范大学生物系	Department of Biology, Capital Normal Univeristy
6	国家人类基因组南方研究中心	Chinese National Human Genome Center at Shanghai
6	上海交通大学农业与生物学院	Shanghai Jiao Tong University School of Agriculture and Biology
6	中国农业科学院兰州兽医研究所	Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences
6	中国医学科学院	Chinese Academy of Medical Sciences
6	复旦大学遗传学研究所	Institute of Genetics, Fudan University
6	北京师范大学生命科学学院	College of Life Sciences, Beijing Normal University
6	中国中医科学院中药研究所	Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences
6	中国科学院水生生物研究所	Institute of Hydrobiology, Chinese Academy of Sciences
6	沈阳农业大学生物科学技术学院	School of Biological Science and Technology, Shenyang Agricultural Univeristy
6	卫生部北京医院药学部	Department of Pharmacy, Beijing Hospital, Ministry of Health
6	南京农业大学	Nanjing Agricultural University
6	中南大学湘雅医院临床药理研究所	Department of Clinical Pharmacology, Xiangya Hospital, Central South University
6	中国科学院亚热带农业生态研究所	Institute of Subtropical Agriculture, The Chinese Academy of Sciences
6	中国农业科学院烟草研究所	Institute of Tobacco Research, Chinese Academy of Agricultural Sciences
6	中国农业科学院北京畜牧兽医研究所动物营养学国家重点实验室	State Key Laboratory of Animal Nutrition, Institute of Animal Sciences of CAAS
6	中国农业科学院作物科学研究所/农作物基因资源与基因改良国家重大科学工程	State Key Scientific Engineering of Crop Genetic Resources and Gene Improvement, Institute of Crop Sciences of CAAS
6	中国工程院	Chinese Academy of Engineering
6	军事医学科学院卫生学环境医学研究所	Institute of Environmental Medicine, Academy of Military Medical Sciences

Figure 4 The visualized co-institution networks of Chinese-language articles

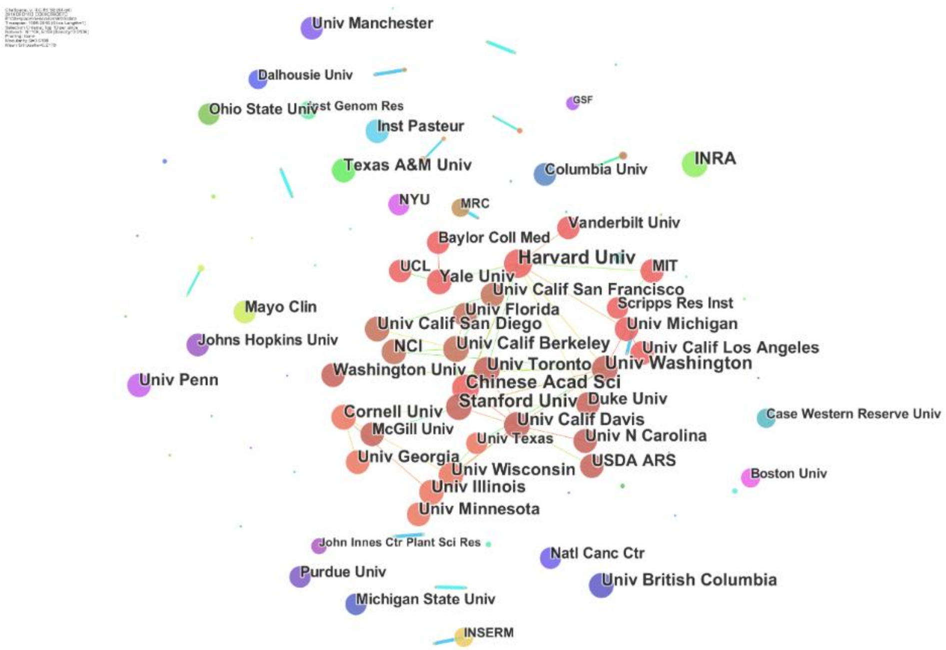


Figure 5 The visualized co-institution networks of foreign-language articles

1.2.3 Analysis on networks of co-operation among countries

The visualized networks of co-operation among countries (Figure 6) was generated by setting Time Slice Length=1, Node Type=Country, Selection Criteria: Top 50 per slice. The international cooperation networks are rather complicated among those 83 countries showing on the map with totally 606 items of collaboration among them. Table 6 shows the Top 10 countries in the number of published articles. USA ranks the first with 17276 published papers in the field of genomics which indicates that its scientific research level is at the forefront of the world. England ranks the second with 2915 published articles. China is the third on the list with 2662 published articles which is of the leading level in the world, indicating that the investment of China in genomics related fields is relatively large but it is also far behind compared with that of USA.

Table 5 The top 10 foreign institutions in the number of published articles

Rank	Institution	Number of published articles	First published year
1	Harvard University	602	1998
2	Chinese Academy of Sciences	486	2007
3	University of Washington	385	1999
4	Stanford University	380	1998
5	University of Toronto	350	2000
6	University of California, Davis	342	2001
7	University of California, Berkeley	330	1998
8	Institut National dela Recherche Agronomique, INRA	322	2008
9	Cornell University	297	2002
10	University of British Columbia	270	2006

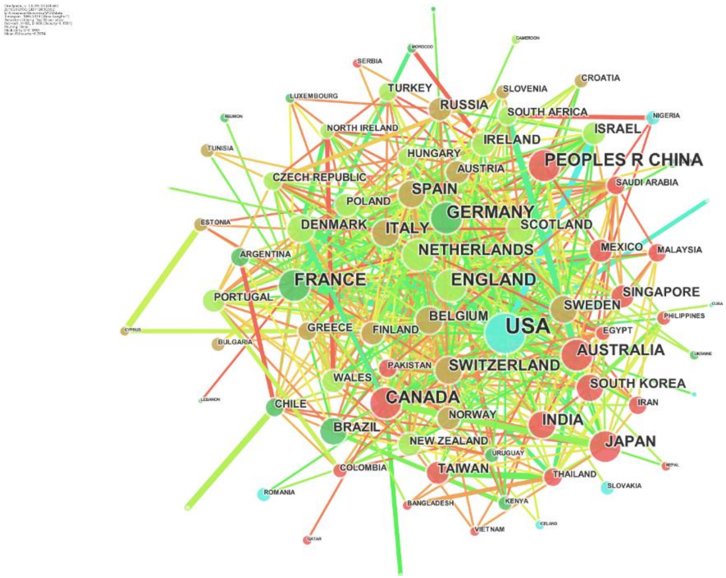


Figure 6 The visualized networks of co-operation among countries

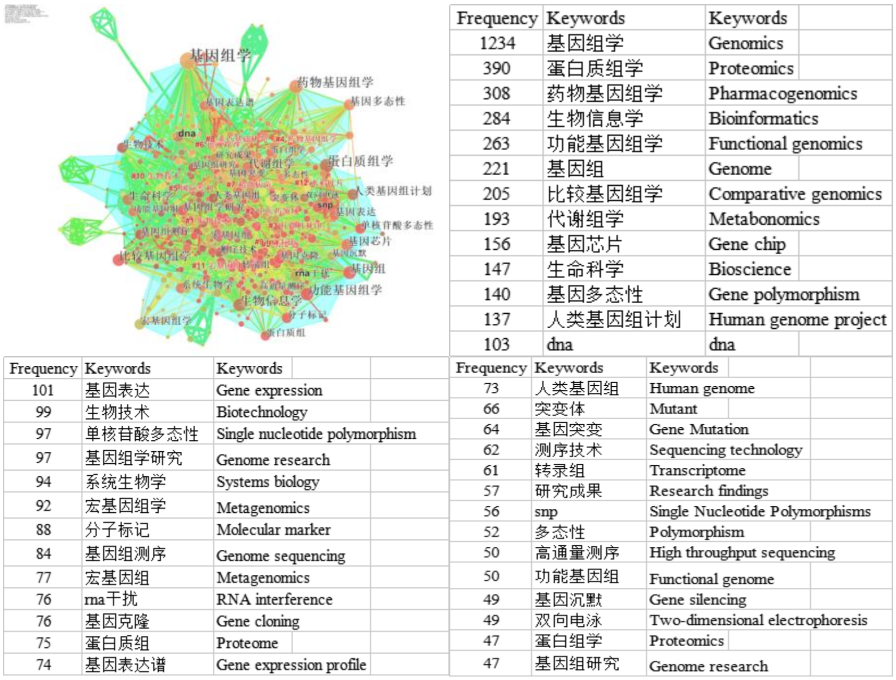
Table 6 The top 10 countries in the number of published articles

Rank	Country	Number of published articles	Betweenness centrality	First published year
1	USA	17276	0.38	1991
2	England	2915	0.07	1991
3	China	2662	0.03	1998
4	Germany	2467	0.19	1992
5	France	2167	0.24	1991
6	Canada	2124	0.12	1992
7	Japan	1590	0.05	1993
8	Australia	1204	0.09	1997
9	Netherlands	1118	0.08	1995
10	Spain	1086	0.12	1999

1.3 Visualized map analysis of co-keywords

The keywords with highest frequency are always the indicator of hot topics in a certain researching field. Thus, this paper analyzed the co-occurrence networks of keywords of the articles in the genomics related fields in order to explore the research hot spots and their evolution.

The paper analyzed the co-keywords networks in Chinese-language articles by setting Time Slice Length=1, Node Type=Keywords, Selection Criteria: Top 80 per slice, and we got the visualized networks of co-keywords consisting of 700 nodes and 2664 links (Figure 7). Table 7 shows the Top 20 keywords. Moreover, cluster analysis was also made which generated 18 relatively obvious clusters (words in red in Figure 7 are the label words of clusters). The module value (Q value)= 0.5709>0.3 and the mean silhouette value (S value)= 0.5112>0.5 illustrate that the generated cluster structure is apparent and the clusters are reasonable. And they are also the bases of the settling of Selection Criteria. The main burst keywords in those articles are listed in Figure 8. Burst keywords are the nodes where the frequency increases or decreases abruptly. Those nodes are usually considered as the turning points of a certain research hot spot.



Frequency	Keywords	Keywords
1234	基因组学	Genomics
390	蛋白质组学	Proteomics
308	药物基因组学	Pharmacogenomics
284	生物信息学	Bioinformatics
263	功能基因组学	Functional genomics
221	基因组	Genome
205	比较基因组学	Comparative genomics
193	代谢组学	Metabonomics
156	基因芯片	Gene chip
147	生命科学	Bioscience
140	基因多态性	Gene polymorphism
137	人类基因组计划	Human genome project
103	dna	dna

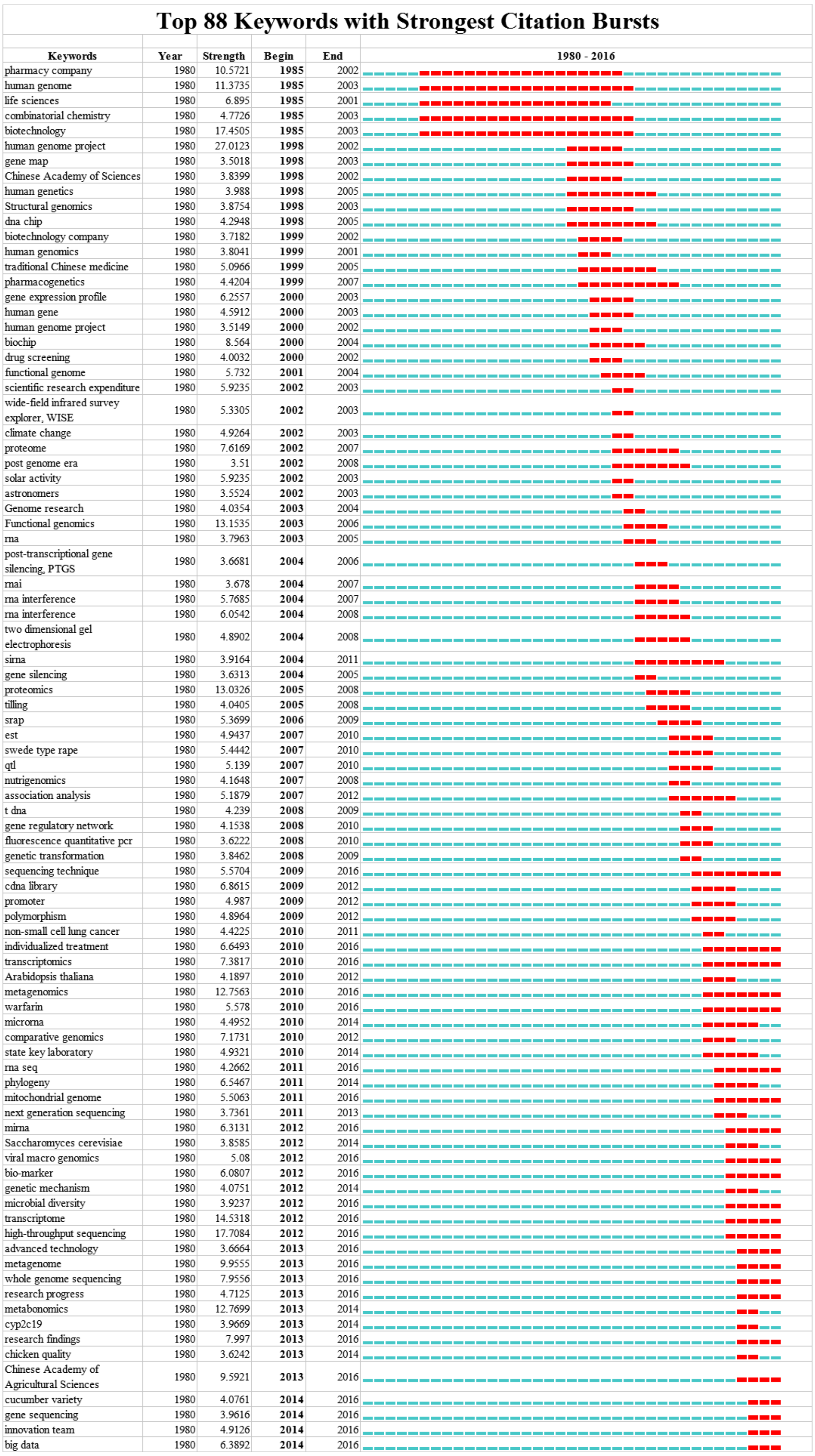
Frequency	Keywords	Keywords
101	基因表达	Gene expression
99	生物技术	Biotechnology
97	单核苷酸多态性	Single nucleotide polymorphism
97	基因组学研究	Genome research
94	系统生物学	Systems biology
92	宏基因组学	Metagenomics
88	分子标记	Molecular marker
84	基因组测序	Genome sequencing
77	宏基因组	Metagenomics
76	rna干扰	RNA interference
76	基因克隆	Gene cloning
75	蛋白质组	Proteome
74	基因表达谱	Gene expression profile

Frequency	Keywords	Keywords
73	人类基因组	Human genome
66	突变体	Mutant
64	基因突变	Gene Mutation
62	测序技术	Sequencing technology
61	转录组	Transcriptome
57	研究成果	Research findings
56	snp	Single Nucleotide Polymorphisms
52	多态性	Polymorphism
50	高通量测序	High throughput sequencing
50	功能基因组	Functional genome
49	基因沉默	Gene silencing
49	双向电泳	Two-dimensional electrophoresis
47	蛋白组学	Proteomics
47	基因组研究	Genome research

Figure 7 The visualized networks of co-keywords in Chinese-language articles

Table 7 The top 20 keywords with highest occurrence frequency in Chinese-language articles

Rank	Keywords	Frequency
1	Genomics	1234
2	Proteomics	390
3	Pharmacogenomics	308
4	Bioinformatics	284
5	Functional genomics	263
6	Genome	221
7	Comparative genomics	205
8	Metabonomics	193
9	Gene chip	156
10	Bioscience	147
11	Gene polymorphism	140
12	Human genome project	137
13	DNA	103
14	Gene expression	101
15	Biotechnology	99
16	Single nucleotide polymorphism	97
17	Genome research	97
18	Systems biology	94
19	Metagenomics	92
20	Molecular marker	88



Top 88 Keywords with Strongest Citation Bursts

Keywords	Year	Strength	Begin	End
pharmacy company	1980	10.5721	**1985**	2002
human genome	1980	11.3735	**1985**	2003
life sciences	1980	6.895	**1985**	2001
combinatorial chemistry	1980	4.7726	**1985**	2003
biotechnology	1980	17.4505	**1985**	2003
human genome project	1980	27.0123	**1998**	2002
gene map	1980	3.5018	**1998**	2003
Chinese Academy of Sciences	1980	3.8399	**1998**	2002
human genetics	1980	3.988	**1998**	2005
Structural genomics	1980	3.8754	**1998**	2003
dna chip	1980	4.2948	**1998**	2005
biotechnology company	1980	3.7182	**1999**	2002
human genomics	1980	3.8041	**1999**	2001
traditional Chinese medicine	1980	5.0966	**1999**	2005
pharmacogenetics	1980	4.4204	**1999**	2007
gene expression profile	1980	6.2557	**2000**	2003
human gene	1980	4.5912	**2000**	2003
human genome project	1980	3.5149	**2000**	2002
biochip	1980	8.564	**2000**	2004
drug screening	1980	4.0032	**2000**	2002
functional genome	1980	5.732	**2001**	2004
scientific research expenditure	1980	5.9235	**2002**	2003
wide-field infrared survey explorer, WISE	1980	5.3305	**2002**	2003
climate change	1980	4.9264	**2002**	2003
proteome	1980	7.6169	**2002**	2007
post genome era	1980	3.51	**2002**	2008
solar activity	1980	5.9235	**2002**	2003
astronomers	1980	3.5524	**2002**	2003
Genome research	1980	4.0354	**2003**	2004
Functional genomics	1980	13.1535	**2003**	2006
rna	1980	3.7963	**2003**	2005
post-transcriptional gene silencing, PTGS	1980	3.6681	**2004**	2006
rnai	1980	3.678	**2004**	2007
rna interference	1980	5.7685	**2004**	2007
rna interference	1980	6.0542	**2004**	2008
two dimensional gel electrophoresis	1980	4.8902	**2004**	2008
sirna	1980	3.9164	**2004**	2011
gene silencing	1980	3.6313	**2004**	2005
proteomics	1980	13.0326	**2005**	2008
tilling	1980	4.0405	**2005**	2008
srap	1980	5.3699	**2006**	2009
est	1980	4.9437	**2007**	2010
swede type rape	1980	5.4442	**2007**	2010
qtl	1980	5.139	**2007**	2010
nutrigenomics	1980	4.1648	**2007**	2008
association analysis	1980	5.1879	**2007**	2012
t dna	1980	4.239	**2008**	2009
gene regulatory network	1980	4.1538	**2008**	2010
fluorescence quantitative pcr	1980	3.6222	**2008**	2010
genetic transformation	1980	3.8462	**2008**	2009
sequencing technique	1980	5.5704	**2009**	2016
cdna library	1980	6.8615	**2009**	2012
promoter	1980	4.987	**2009**	2012
polymorphism	1980	4.8964	**2009**	2012
non-small cell lung cancer	1980	4.4225	**2010**	2011
individualized treatment	1980	6.6493	**2010**	2016
transcriptomics	1980	7.3817	**2010**	2016
Arabidopsis thaliana	1980	4.1897	**2010**	2012
metagenomics	1980	12.7563	**2010**	2016
warfarin	1980	5.578	**2010**	2016
microrna	1980	4.4952	**2010**	2014
comparative genomics	1980	7.1731	**2010**	2012
state key laboratory	1980	4.9321	**2010**	2014
rna seq	1980	4.2662	**2011**	2016
phylogeny	1980	6.5467	**2011**	2014
mitochondrial genome	1980	5.5063	**2011**	2016
next generation sequencing	1980	3.7361	**2011**	2013
mirna	1980	6.3131	**2012**	2016
Saccharomyces cerevisiae	1980	3.8585	**2012**	2014
viral macro genomics	1980	5.08	**2012**	2016
bio-marker	1980	6.0807	**2012**	2016
genetic mechanism	1980	4.0751	**2012**	2014
microbial diversity	1980	3.9237	**2012**	2016
transcriptome	1980	14.5318	**2012**	2016
high-throughput sequencing	1980	17.7084	**2012**	2016
advanced technology	1980	3.6664	**2013**	2016
metagenome	1980	9.9555	**2013**	2016
whole genome sequencing	1980	7.9556	**2013**	2016
research progress	1980	4.7125	**2013**	2016
metabonomics	1980	12.7699	**2013**	2014
cyp2c19	1980	3.9669	**2013**	2014
research findings	1980	7.997	**2013**	2016
chicken quality	1980	3.6242	**2013**	2014
Chinese Academy of Agricultural Sciences	1980	9.5921	**2013**	2016
cucumber variety	1980	4.0761	**2014**	2016
gene sequencing	1980	3.9616	**2014**	2016
innovation team	1980	4.9126	**2014**	2016
big data	1980	6.3892	**2014**	2016

Figure 8 The main burst keywords in Chinese-language articles

The paper analyzed the co-keywords networks in English-language articles by setting Time Slice Length=1, Node Type=Keywords, Selection Criteria: Top 20 per slice, and we got the visualized networks consisting of 138 nodes and 466 links (Figure 9). Table 8 lists the Top 20 keywords. Moreover, cluster analysis was also made which generated 10 relatively obvious clusters. The module value (Q value)= 0.5452>0.3 and the mean silhouette value (S value)= 0.7669>0.7 illustrate that the generated cluster structure is apparent and the clusters are highly efficient and reasonable. The main burst keywords in those articles are listed in Figure 10.

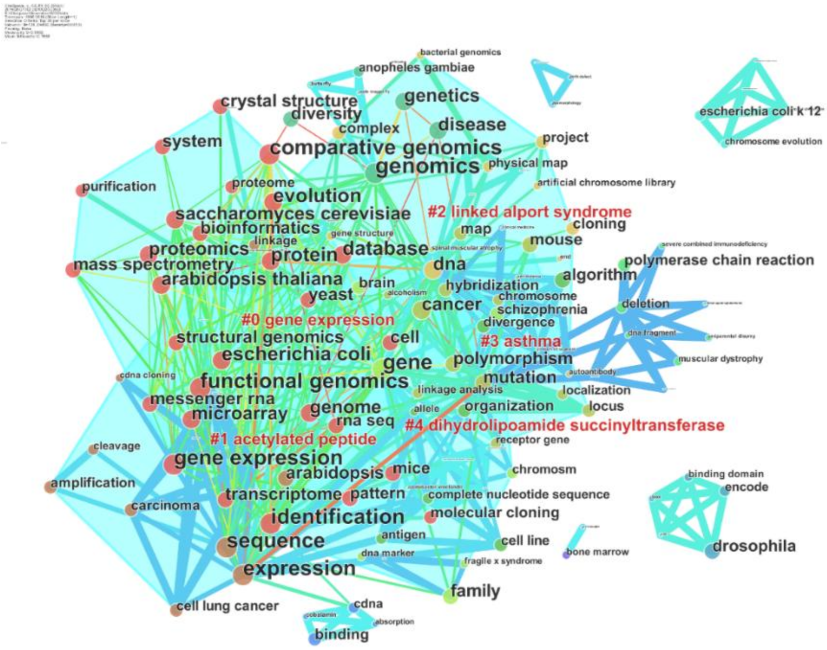
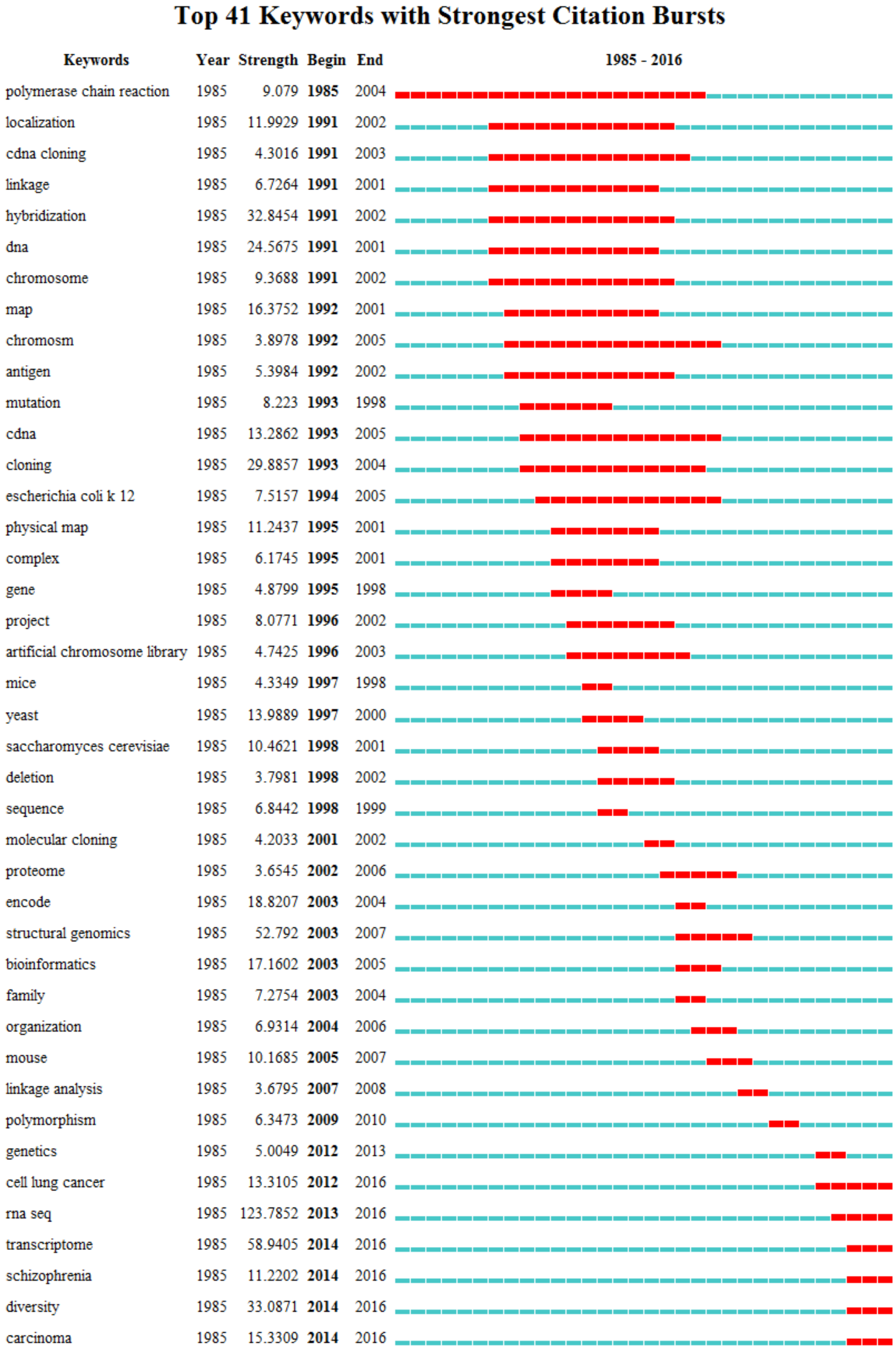


Figure 9 The visualized networks of co-keywords in foreign-language articles

Table 8 The top 20 keywords with highest occurrence frequency in foreign-language articles

Rank	Keyword	Frequency
1	Genomics	8085
2	Gene expression	3938
3	Identification	3602
4	Expression	3312
5	Gene	3296
6	Comparative genomics	3215
7	Functional genomics	3132
8	Sequence	3056
9	Evolution	2916
10	Protein	2351
11	Escherichia coli	2158
12	Database	1942
13	Arabidopsis thaliana	1720
14	Microarray	1519
15	Genome	1482
16	DNA	1462
17	Proteomics	1302
18	Saccharomyces cerevisiae	1246
19	Genetics	1176
20	Cancer	1165



Top 41 Keywords with Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	1985 - 2016
polymerase chain reaction	1985	9.079	**1985**	2004	
localization	1985	11.9929	**1991**	2002	
cdna cloning	1985	4.3016	**1991**	2003	
linkage	1985	6.7264	**1991**	2001	
hybridization	1985	32.8454	**1991**	2002	
dna	1985	24.5675	**1991**	2001	
chromosome	1985	9.3688	**1991**	2002	
map	1985	16.3752	**1992**	2001	
chromosm	1985	3.8978	**1992**	2005	
antigen	1985	5.3984	**1992**	2002	
mutation	1985	8.223	**1993**	1998	
cdna	1985	13.2862	**1993**	2005	
cloning	1985	29.8857	**1993**	2004	
escherichia coli k 12	1985	7.5157	**1994**	2005	
physical map	1985	11.2437	**1995**	2001	
complex	1985	6.1745	**1995**	2001	
gene	1985	4.8799	**1995**	1998	
project	1985	8.0771	**1996**	2002	
artificial chromosome library	1985	4.7425	**1996**	2003	
mice	1985	4.3349	**1997**	1998	
yeast	1985	13.9889	**1997**	2000	
saccharomyces cerevisiae	1985	10.4621	**1998**	2001	
deletion	1985	3.7981	**1998**	2002	
sequence	1985	6.8442	**1998**	1999	
molecular cloning	1985	4.2033	**2001**	2002	
proteome	1985	3.6545	**2002**	2006	
encode	1985	18.8207	**2003**	2004	
structural genomics	1985	52.792	**2003**	2007	
bioinformatics	1985	17.1602	**2003**	2005	
family	1985	7.2754	**2003**	2004	
organization	1985	6.9314	**2004**	2006	
mouse	1985	10.1685	**2005**	2007	
linkage analysis	1985	3.6795	**2007**	2008	
polymorphism	1985	6.3473	**2009**	2010	
genetics	1985	5.0049	**2012**	2013	
cell lung cancer	1985	13.3105	**2012**	2016	
rna seq	1985	123.7852	**2013**	2016	
transcriptome	1985	58.9405	**2014**	2016	
schizophrenia	1985	11.2202	**2014**	2016	
diversity	1985	33.0871	**2014**	2016	
carcinoma	1985	15.3309	**2014**	2016	

Figure 10 The main burst keywords in foreign-language articles

From 1985 to the beginning of 21st century, there were burst keywords like pharmacy company, biotechnology, combinatorial chemistry, human genome, and life sciences (Figure 8; Figure 10). New technologies like genomics, combinatorial chemistry and the following highly efficient screening of drugs have transformed biotechnological research and pharmacy industries. At the same time, Human Genome Project was launched by National Human Genome Research Institute of USA in 1990, aiming to uncover the secrets of 3 billion base pairs that constitute all the genes of human body. In 1998, Human Genome Center in Institute of Genetics and Developmental Biology was established in Chinese Academy of Sciences. In the same year, National Human Genome Center, Beijing and National Human Genome Center at Shanghai were established. The next year, Human Genome Center was registered internationally and completed the task of sequencing about 30Mb region from human chromosome break No.8 which accounts for 1% of the whole human genome. China is the 6th country following USA, UK, France, German and Japan to have participated in Human Genome Project (Wu, 2009). During the period, DNA chip technology was developed at the right moment to high-efficiently and quickly test and analyse a large sum of genetic information. Post-genome era came around when human gene map was portrayed completely in 2000. Functional genomics became the focus of genomic research. New technologies like two dimensional gel electrophoresis and DNA chips were developed and applied. It concentrated on cognizing and analyzing the genetic and non-genetic sequences and their functions of the whole genome in attempt to make out the encyclopedia to help interpret the profound DNA language. In 2010, there were burst keywords like metagenomics and comparative genomics. Metagenomics takes microorganism DNA extracted from environmental samples as study objects to construct the metagenomic library. It screens and searches for new physiological activators to acquire information about micro-organic genetic diversity and molecular ecology from the environment (Huang et al., 2009). Comparative genomics mainly compares the whole genome of different species and comprehends the functional and developmental correlation of the whole genome. In addition, post-genome era has made proteomics more popular. Proteomics is mainly about a comprehensive study of the properties of protein, providing theoretical basis and solution for the clarification and attack of many disease mechanisms at the protein level. It has also brought revolution to the medical field. Keywords like pharmacogenomics and individualized treatment have been extensively mentioned in references. In the course of clinical treatment, it is often found that different patients have different therapeutic effects and side effects on the same drug. Pharmacogenomics, based on gene theory, studies the relationship between the gene itself and its mutants and its drug effects. Through the gene detection of the patients, the individualized treatment plan is provided according to its genotype, so as to improve the efficacy and reduce the occurrence of adverse drug reactions. In recent years, the development of the treatment on non-small cell lung cancer depends largely on the research and application of pharmacogenomics. The direct motivation of Human Genome Project is to solve the basic genetic problems of human diseases including cancer, and to achieve early prevention and treatment, so as to reduce the risk of disease. In addition, in recent years, traditional Chinese medicine genomics which takes Chinese medicine raw species as study objects has also developed rapidly. With the help of techniques like gene chips and bio-informatics, the targets of traditional Chinese medicine and their mechanism of action can be explored to determine the effective parts of traditional Chinese medicine, identify Chinese herbs, distinguish genuine herbs, screen new drugs and shorten the cycle and finally open up a new path for the modernization of traditional Chinese medicine (Xing et al., 2007) (Figure 7). In 2014, the concept of big data burst. In the same year, Google marched towards genomics and joined Global Alliance for Genomics and Health in order to gather the resources in this field and build databases based on big data to solve the problem of interpreting the results of complex gene detection.

1.4 Visualized map analysis of co-citation

Visualized co-citation map was portrayed by CiteSpace to study the key references used in the field of genomics. At present, CiteSpace is still unable to carry out the literature co-citation analysis of the literature in CNKI, so this study only analyzes the English literature. Because of the huge amount of data cited in English literature, this study is divided into three time periods to analyze the cited literature in the English literature-the first 5 years (1985-2009), the second 5 years (2010-2014) and recent two years (2015 and 2016). According to the document quantity and cluster analysis Q value, Selection Criteria is set as Top 10, Top 30, Top 100, respectively (Figure 11; Figure 12; Figure 13). Q value is 0.7171, 0.5164 and 0.6984, respectively, which indicates the cluster structure is apparent. The main clusters and top terms of reference co-citation network in the three periods are listed below (Table 9; Table 10; Table 11).

Table 12 lists the top 10 literature with highest centrality. Literature with highest centrality means that papers occupy an important position in structure, that is, they play an important role in connecting other nodes or several different clusters. These documents can be regarded as a landmark in the field of genomics (Chen et al., 2014).

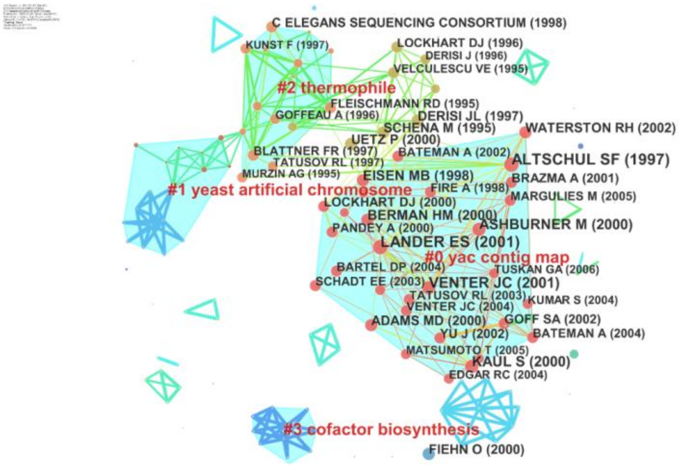


Figure 11 The visualized network of reference co-citation (1985-2009)

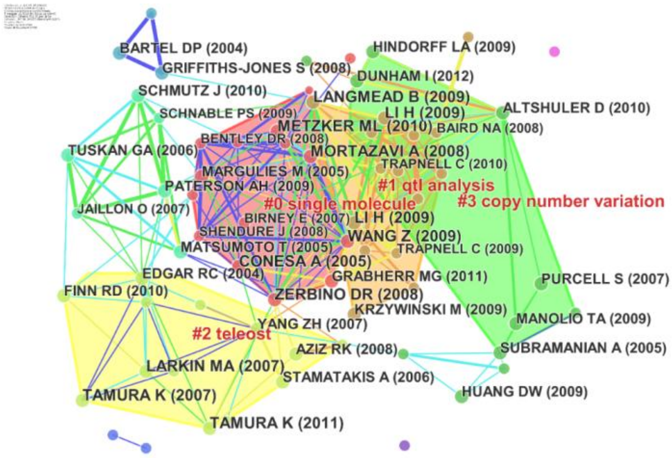


Figure 12 The visualized network of reference co-citation (2010-2014)

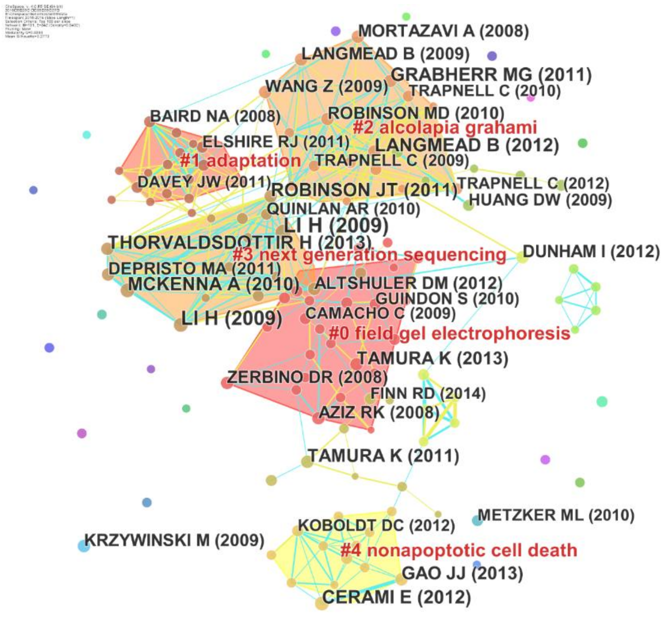


Figure 13 The visualized network of reference co-citation (2015-2016)

Table 9 The main clusters and top terms in the reference co-citation network (1985-2009)

Cluster	Size	Silhouette	Top terms (log-likelihood ratio, p-level)	Average year of publication
0	30	0.793	YAC contig map	2002
1	17	0.936	Yeast artificial chromosome(YAC)	1988
2	14	0.748	Thermophile	1995
3	10	1	Cofactor biosynthesis	1985
4	9	1	Severe combined immunodeficiency	1988
5	8	0.899	Gene expression pattern	1996

Table 10 The main clusters and top terms in the reference co-citation network (2010-2014)

Cluster	Size	Silhouette	Top terms (log-likelihood ratio, p-level)	Average year of publication
0	16	0.658	Single molecule	2006
1	15	0.851	Quantitative Trait Locus (QTL) analysis	2010
2	13	0.732	Teleost	2006
3	11	0.912	Copy number variation	2007
4	6	0.961	Genome duplication	2007

Table 11 The main clusters and top terms in the reference co-citation network (2015-2016)

Cluster	Size	Silhouette	Top terms (log-likelihood ratio, p-level)	Average year of publication
0	24	0.790	Field gel electrophoresis	2010
1	17	0.937	Adaptation	2011
2	16	0.761	Alcolapia grahami	2011
3	16	0.829	Next generation sequencing	2010
4	13	0.927	Nonapoptotic cell death	2012
5	8	0.972	Crangon crangon	2011
6	6	0.918	Zinc finger nuclease	2013
7	6	0.910	Venous thrombosis	2012
8	5	1	Draft genome	2009

Table 12 The top 10 literature with highest centrality

Rank	Centrality	Author	Year	Source	Volume	Page	Cluster
1	0.38	Yang ZH	2007	Molecular Biology and Evolution	V24	P1586	2b、0c
2	0.37	Dunham I	2012	Nature	V489	P57	3b、6c
3	0.36	Cong L	2013	Science	V339	P819	6c
4	0.33	Li H	2009	Bioinformatics	V25	P2078	1b、3c
5	0.29	Wang T	2014	Science	V343	P80	6c
6	0.28	Altshuler D	2010	Nature	V467	P1061	3b
7	0.27	Barretina J	2012	Nature	V483	P603	4c
8	0.26	Langmead B	2009	Genome Biology	V10	R25.1	1b、2c
9	0.24	Zerbino DR	2008	Genome research	V18	P821	0b、0c
10	0.2	Overbeek R	2014	Nucleic acids research	V42	D206	0c
10	0.2	Altschul SF	1990	Journal of Molecular Biology	V215	P403	2a

Note: a refers to the cluster in the reference co-citation network (1985-2009); b refers to the cluster in the reference co-citation network (2010-2014); c refers to the cluster in the reference co-citation network (2015-2016)

The literature “Yang ZH-2007” ranks the first with centrality of 0.38. The title is “PAML 4: Phylogenetic analysis by maximum likelihood”. PAML is a software package for phylogenetic analysis on DNA or protein sequences using maximum likelihood methods. It was developed by Ziheng Yang and provided for academic use for free. He is the author of the paper, and also he is a famous ethnic Chinese scientist, academician of the Royal Academy of Sciences and professor in statistical genetics of University of London. The use of multi-core PAML parallel algorithm has obvious acceleration effect on data set analysis of DNA and protein sequences. The literature “Dunham I-2012” ranks the second with centrality of 0.37. The title is “An integrated encyclopedia of DNA elements in the human genome”. The integrated encyclopedia of DNA elements can systematically map the areas of transcription, transcription factor association, chromosome structure, and histone modification. These data enable the biochemical function of 80% genes to be assigned. Many of the candidate regulatory elements are found to be associated with other regulatory elements and expression genes, providing new insights into the regulation mechanism of genes. Some newly discovered gene elements and sequence variations associated with human diseases show statistical correlation with sequence variation, which helps to explain variation. The literature “Cong L-2013” ranks the third with the centrality of 0.36. The title is “Multiplex Genome Engineering Using CRISPR/Cas Systems”. Gene editing technology CRISPR/Cas9 is listed by SCIENCE as one of the ten major advances in science and technology in the year of 2013. In this paper, two kinds of II CRISPR (regularly spaced short palindrome repeats) / Cas (CRISPR related protein) system were designed and proved that Cas9 nuclease can accurately split endogenous genomes directly induced by short RNA in human and mouse cells. Cas9 can also be converted into an incisional enzyme that promotes homologous directional repair with minimal mutagenic activity. Multiple boot sequences can be encoded into a single CRISPR array, enabling simultaneous editing of several parts of the mammalian genome. It has been proved that RNA guided nuclease technology is convenient, programmable and widely applicable. Besides, the most recently published high-centrality literature in 2014 is also about CRISPR/Cas9 system entitled “Genetic Screens in Human Cells Using the CRISPR-Cas9 System”. In the CRISPR/Cas9 system, the enzyme Cas9 cuts at the DNA target site. The target of DNA is determined in the following way: the RNA molecule called CRISPR RNA (crRNA) uses some of its sequences to bind to another RNA molecule called tracrRNA by base pairing. The chimeric RNA (tracrRNA/crRNA) is formed and then pairing with the target DNA site with another portion of the crRNA sequence. In this way, this chimeric RNA can guide Cas9 to the target site and cut it. In practical application, tracrRNA and crRNA can be used as the two guiding RNA (gRNA) or can be fused together to form a one-way guide RNA (single guide RNA, sgRNA), and it’s used to guide the enzyme Cas9 binding to the target DNA sequence and cutting. CAS9 together with sgRNA is called the Cas9-sgRNA system. Therefore, in order to construct nuclear gene modified animal models by using CRISPR/Cas9, Cas9/sgRNA software package was complied to assist in the rapid design and screening of highly active and specific sgRNA, constructing sgRNA expression library and large-scale construction of genetically modified animal models. Another high-centrality literature published in 2014 was “The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)”. Genome annotation belongs to the category of functional genomics. RAST is a fast annotation tool using Subsystem technology. It is a genome annotation tool for complete or nearly complete bacteria and archaea. The accuracy, consistency and completeness of RAST are based on two databases: the Subsystem Library of artificial rectification and the FIGfams Library of protein, which can be used to predict ORF, Rrna, Trna and corresponding functional genes, and these information can be used to build metabolic network.

By integrating the co-cited literature of three periods, the top 10 most cited references are listed in Table 13. The highly cited literature is generally an important document with a fundamental role. The top is “Lander ES-2001” with the 851 times of citation. The title is “Initial sequencing and analysis of the human genome”. This paper reports the results of an international collaboration dedicated to providing free drafts of human genome sequencing, and makes a preliminary analysis of the sequencing data. The second-place literature is “Altschul SF-1997” with 797 times of citation. The title is “Gapped BLAST and PSI-BLAST：a new generation of protein database search programs”. Gapped BLAST is Gapped Basic Local Alignment Search Tool while PSI-BLAST is Position Specific Iterated BLAST. BLAST is a search tool widely used to retrieve protein or DNA sequences similar to the current research sequences in protein or DNA databases. The improved BLAST allows the insertion of the vacancy, that is, the vacancy BLAST, and its running speed is up to three times that of the original. PSI-BLAST introduces the location specific scoring matrix into BLAST, and retrieves protein or DNA database by using this matrix. It finds out the best retrieval results through many iterations. The single iteration of PSI-BLAST is similar to that of vacancy BLAST, but its sensitivity to bio-correlation sequences with weak similarity is stronger than that of vacancy BLAST. It has thus been used to explore and discover the new and interesting BRCT protein super family members. The third-place literature is “Li H-2009” with 659times of citation. The title is “The Sequence Alignment/Map format and SAMtools”. It’s also the literature of high centrality. SAM format is a text format for storing and reading gene sequence data. It can support short sequence reading and long sequence reading (up to 128Mbp) generated by different sequencing platforms. In addition, the most recently published high-centrality literature in 2011 is “Molecular Evolutionary Genetics Analysis Using Maximum Likelihood，Evolutionary Distance，and Maximum Parsimony Methods”. The comparative analysis of molecular sequence data plays a vital role in reproducing the evolutionary history of species and inferring the impacts of natural selection on the creation of genes and species evolution. This document has released the latest version of MEGA (Molecular Evolutionary Genetics Analysis) software MEGA5, which introduces the maximum likelihood rate algorithm to deduce the evolutionary tree, select the best alternative model (nucleotide or amino acid), deduce the ancestral sequence and state (and probability), and estimate the rate of evolution. In computer simulation and analysis, the maximum likelihood rate algorithm adopted by MEGA5 is better than other softwares in inferring phylogenetic trees and replacing parameters. This version supports Windows, Mac OS X and Linux systems, available at http://www.megasoftware.net free of charge.

Table 13 The top 10 most cited references (1985-2016)

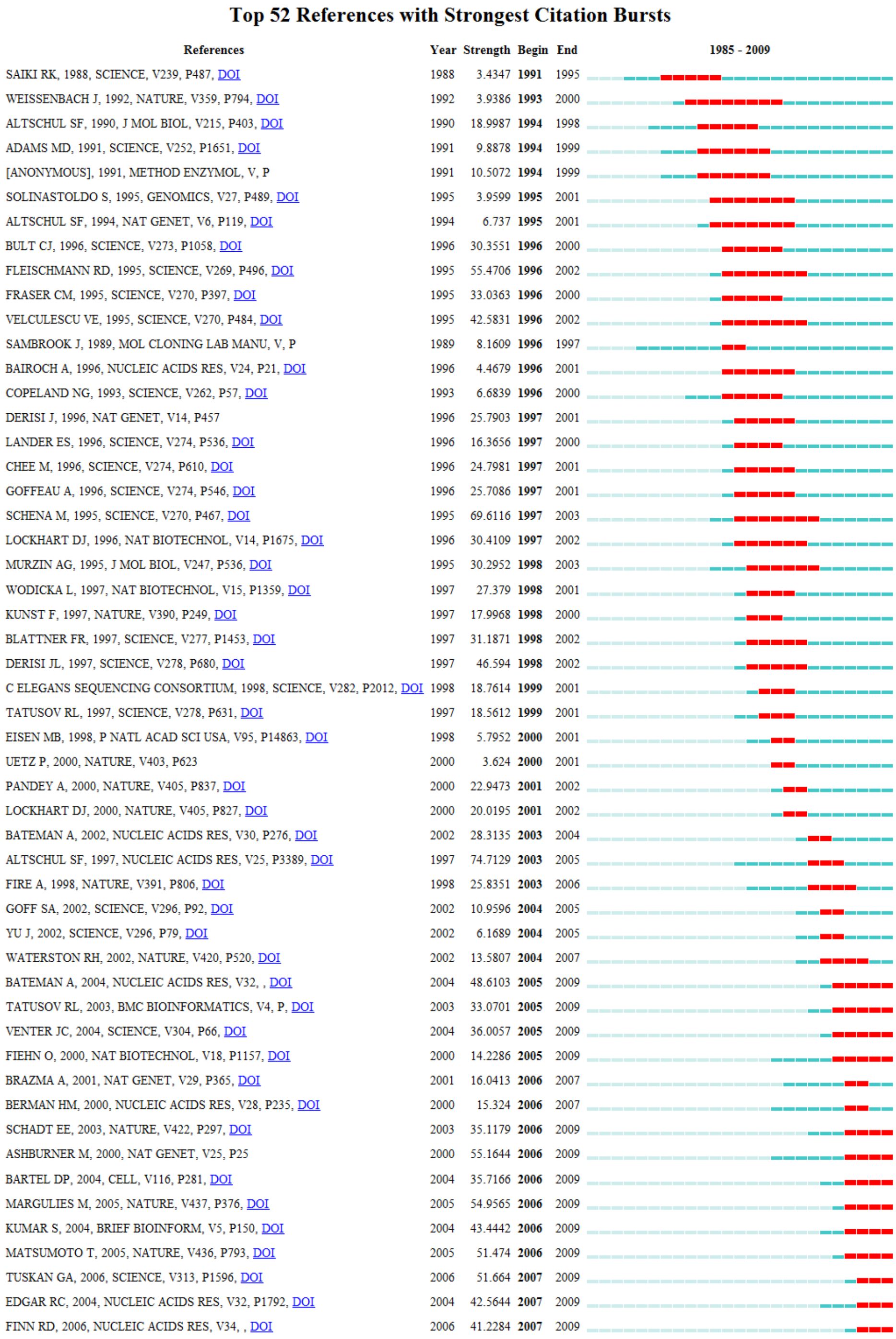
Rank	Frequency	Author	Year	Source	Volume	Page	Cluster
1	851	Lander ES	2001	Nature	V409	P860	0a
2	797	Altschul SF	1997	Nucleic Acids Research	V25	P3389	0a
3	659	Li H	2009	Bioinformatics	V25	P2078	1b、3c
4	629	Venter JC	2001	Science	V291	P1304	0a
5	523	Li H	2009	Bio-informatics	V25	P1754	1b、3c
6	455	Tamura K	2011	Molecular Biology and Evolution	V28	P2731	2b、5c
7	435	Ashburner M	2000	Nature Genetics	V25	P25	0a
8	432	Wang Z	2009	Nature Reviews Genetics	V10	P57	0b、2c
9	412	Kaul S	2000	Nature	V408	P796	0a
10	380	Berman HM	2000	Nucleic Acids Research	V28	P235	0a

Note: a refers to the cluster in the reference co-citation network (1985-2009); b refers to the cluster in the reference co-citation network (2010-2014); c refers to the cluster in the reference co-citation network (2015-2016)

The references with strong citation bursts of 1985-2009 and 2010-2014 are shown below (Figure 14; Figure 15). There is no citation burst in 2015 or 2016. The reference with strong citation burst refers to the sudden increase in the cited frequency of the reference at a time point or time period, so it contains two dimensions: the burst strength and the bursting time. The reference with the highest burst strength is “Altschul SF-1997” with the value of 74.7129. The title is “Gapped BLAST and PSI-BLAST：a new generation of protein database search programs”. It’s also the reference of high citation. It indicates that this paper has received great attention in genome research, especially from 2003 to 2005, and has played an important role in the research of this field. It is the research hot spot during this period. In addition, the literature of most recent bursting time is “Finn RD-2010”. The bursting time was 2011-2012. It is the research frontier in recent years. The title is “The Pfam protein families database”. Pfam is a protein motif database, which is widely used in proteomics research. It is based on hidden Markov model and provides multiple sequence alignment services. It is a large set of protein family. This paper introduces its latest version, Pfam24.0, which applies the latest version of Hidden Markov Model package, HMMER3. HMMER3 runs 100 times faster than HMMER2. The sensitivity is greatly improved by the application of forward algorithm. Pfam 24 contains 11912 protein families. Pfam application website: http://pfam.sanger.ac.uk/ (UK), http://pfam.janelia.org/ (USA), http://pfam.sbc.su.se/ (Sweden).

2 Discussion

Through the bibliometric analysis and visualization analysis in the field of genomics from 1985 to 2016, it is found that there are more papers in genomics in China, but there is less cooperation among researchers and among research institutions. Therefore, while strengthening the research of genomics and encouraging domestic research institutions to strengthen the research investment in this discipline, at the same time, we should also encourage the exchanges and cooperation of various scientific research institutions in various regions of China, as well as cross regional and transnational cooperation between China and other countries or regions in the field of genomics.

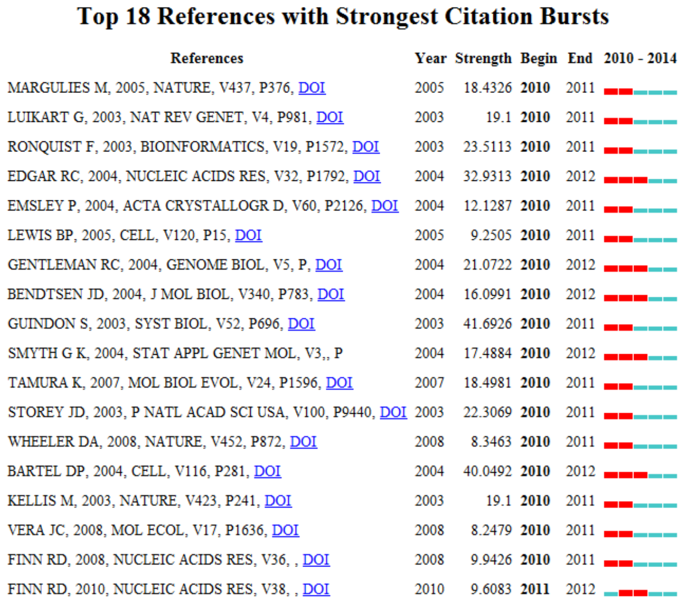


Top 52 References with Strongest Citation Bursts

References	Year	Strength	Begin	End	1985 - 2009
SAIKI RK, 1988, SCIENCE, V239, P487, DOI	1988	3.4347	**1991**	1995	
WEISSENBACH J, 1992, NATURE, V359, P794, DOI	1992	3.9386	**1993**	2000	
ALTSCHUL SF, 1990, J MOL BIOL, V215, P403, DOI	1990	18.9987	**1994**	1998	
ADAMS MD, 1991, SCIENCE, V252, P1651, DOI	1991	9.8878	**1994**	1999	
[ANONYMOUS], 1991, METHOD ENZYMOL, V, P	1991	10.5072	**1994**	1999	
SOLINASTOLDO S, 1995, GENOMICS, V27, P489, DOI	1995	3.9599	**1995**	2001	
ALTSCHUL SF, 1994, NAT GENET, V6, P119, DOI	1994	6.737	**1995**	2001	
BULT CJ, 1996, SCIENCE, V273, P1058, DOI	1996	30.3551	**1996**	2000	
FLEISCHMANN RD, 1995, SCIENCE, V269, P496, DOI	1995	55.4706	**1996**	2002	
FRASER CM, 1995, SCIENCE, V270, P397, DOI	1995	33.0363	**1996**	2000	
VELCULESCU VE, 1995, SCIENCE, V270, P484, DOI	1995	42.5831	**1996**	2002	
SAMBROOK J, 1989, MOL CLONING LAB MANU, V, P	1989	8.1609	**1996**	1997	
BAIROCH A, 1996, NUCLEIC ACIDS RES, V24, P21, DOI	1996	4.4679	**1996**	2001	
COPELAND NG, 1993, SCIENCE, V262, P57, DOI	1993	6.6839	**1996**	2000	
DERISI J, 1996, NAT GENET, V14, P457	1996	25.7903	**1997**	2001	
LANDER ES, 1996, SCIENCE, V274, P536, DOI	1996	16.3656	**1997**	2000	
CHEE M, 1996, SCIENCE, V274, P610, DOI	1996	24.7981	**1997**	2001	
GOFFEAU A, 1996, SCIENCE, V274, P546, DOI	1996	25.7086	**1997**	2001	
SCHENA M, 1995, SCIENCE, V270, P467, DOI	1995	69.6116	**1997**	2003	
LOCKHART DJ, 1996, NAT BIOTECHNOL, V14, P1675, DOI	1996	30.4109	**1997**	2002	
MURZIN AG, 1995, J MOL BIOL, V247, P536, DOI	1995	30.2952	**1998**	2003	
WODICKA L, 1997, NAT BIOTECHNOL, V15, P1359, DOI	1997	27.379	**1998**	2001	
KUNST F, 1997, NATURE, V390, P249, DOI	1997	17.9968	**1998**	2000	
BLATTNER FR, 1997, SCIENCE, V277, P1453, DOI	1997	31.1871	**1998**	2002	
DERISI JL, 1997, SCIENCE, V278, P680, DOI	1997	46.594	**1998**	2002	
C ELEGANS SEQUENCING CONSORTIUM, 1998, SCIENCE, V282, P2012, DOI	1998	18.7614	**1999**	2001	
TATUSOV RL, 1997, SCIENCE, V278, P631, DOI	1997	18.5612	**1999**	2001	
EISEN MB, 1998, P NATL ACAD SCI USA, V95, P14863, DOI	1998	5.7952	**2000**	2001	
UETZ P, 2000, NATURE, V403, P623	2000	3.624	**2000**	2001	
PANDEY A, 2000, NATURE, V405, P837, DOI	2000	22.9473	**2001**	2002	
LOCKHART DJ, 2000, NATURE, V405, P827, DOI	2000	20.0195	**2001**	2002	
BATEMAN A, 2002, NUCLEIC ACIDS RES, V30, P276, DOI	2002	28.3135	**2003**	2004	
ALTSCHUL SF, 1997, NUCLEIC ACIDS RES, V25, P3389, DOI	1997	74.7129	**2003**	2005	
FIRE A, 1998, NATURE, V391, P806, DOI	1998	25.8351	**2003**	2006	
GOFF SA, 2002, SCIENCE, V296, P92, DOI	2002	10.9596	**2004**	2005	
YU J, 2002, SCIENCE, V296, P79, DOI	2002	6.1689	**2004**	2005	
WATERSTON RH, 2002, NATURE, V420, P520, DOI	2002	13.5807	**2004**	2007	
BATEMAN A, 2004, NUCLEIC ACIDS RES, V32, , DOI	2004	48.6103	**2005**	2009	
TATUSOV RL, 2003, BMC BIOINFORMATICS, V4, P, DOI	2003	33.0701	**2005**	2009	
VENTER JC, 2004, SCIENCE, V304, P66, DOI	2004	36.0057	**2005**	2009	
FIEHN O, 2000, NAT BIOTECHNOL, V18, P1157, DOI	2000	14.2286	**2005**	2009	
BRAZMA A, 2001, NAT GENET, V29, P365, DOI	2001	16.0413	**2006**	2007	
BERMAN HM, 2000, NUCLEIC ACIDS RES, V28, P235, DOI	2000	15.324	**2006**	2007	
SCHADT EE, 2003, NATURE, V422, P297, DOI	2003	35.1179	**2006**	2009	
ASHBURNER M, 2000, NAT GENET, V25, P25	2000	55.1644	**2006**	2009	
BARTEL DP, 2004, CELL, V116, P281, DOI	2004	35.7166	**2006**	2009	
MARGULIES M, 2005, NATURE, V437, P376, DOI	2005	54.9565	**2006**	2009	
KUMAR S, 2004, BRIEF BIOINFORM, V5, P150, DOI	2004	43.4442	**2006**	2009	
MATSUMOTO T, 2005, NATURE, V436, P793, DOI	2005	51.474	**2006**	2009	
TUSKAN GA, 2006, SCIENCE, V313, P1596, DOI	2006	51.664	**2007**	2009	
EDGAR RC, 2004, NUCLEIC ACIDS RES, V32, P1792, DOI	2004	42.5644	**2007**	2009	
FINN RD, 2006, NUCLEIC ACIDS RES, V34, , DOI	2006	41.2284	**2007**	2009	

Figure 14 The references with strongest citation bursts (1985-2009)

The Human Genome Project was put forward in 1985 and officially launched in 1990. With the completion of the draft of the human genome, the post genome era has come around. Through the analysis of co-keywords, burst keywords and co-citation, we can also find that the current research focus of genomics has been transferred to post genomics. Functional genomics has become the focus of research, and proteomics is the core of it. Proteomics research can help people learn a comprehensive understanding about the occurrence and development of disease from the protein level, and it can provide theoretical basis and solution for the early diagnosis of disease, discovery of bio-markers of the disease and search of the target molecule for corresponding drugs. It has great application prospect in disease prevention and individualized treatment. What's more, pharmacogenomics is also a keyword with relatively high co-occurrence frequency in the study of genomics, and it has developed new ideas for traditional pharmaceutical research. Because of the different reactions of different patients to the same drugs, individualized medication guidance services are provided for patients by genotype detection to improve the efficacy and ensure the safety of drug use.



Top 18 References with Strongest Citation Bursts

References	Year	Strength	Begin	End	2010 - 2014
MARGULIES M, 2005, NATURE, V437, P376, DOI	2005	18.4326	**2010**	2011	🟥🟥🟦🟦🟦
LUIKART G, 2003, NAT REV GENET, V4, P981, DOI	2003	19.1	**2010**	2011	🟥🟥🟦🟦🟦
RONQUIST F, 2003, BIOINFORMATICS, V19, P1572, DOI	2003	23.5113	**2010**	2011	🟥🟥🟦🟦🟦
EDGAR RC, 2004, NUCLEIC ACIDS RES, V32, P1792, DOI	2004	32.9313	**2010**	2012	🟥🟥🟥🟦🟦
EMSLEY P, 2004, ACTA CRYSTALLOGR D, V60, P2126, DOI	2004	12.1287	**2010**	2011	🟥🟥🟦🟦🟦
LEWIS BP, 2005, CELL, V120, P15, DOI	2005	9.2505	**2010**	2011	🟥🟥🟦🟦🟦
GENTLEMAN RC, 2004, GENOME BIOL, V5, P, DOI	2004	21.0722	**2010**	2012	🟥🟥🟥🟦🟦
BENDTSEN JD, 2004, J MOL BIOL, V340, P783, DOI	2004	16.0991	**2010**	2012	🟥🟥🟥🟦🟦
GUINDON S, 2003, SYST BIOL, V52, P696, DOI	2003	41.6926	**2010**	2011	🟥🟥🟦🟦🟦
SMYTH G K, 2004, STAT APPL GENET MOL, V3,, P	2004	17.4884	**2010**	2012	🟥🟥🟥🟦🟦
TAMURA K, 2007, MOL BIOL EVOL, V24, P1596, DOI	2007	18.4981	**2010**	2011	🟥🟥🟦🟦🟦
STOREY JD, 2003, P NATL ACAD SCI USA, V100, P9440, DOI	2003	22.3069	**2010**	2011	🟥🟥🟦🟦🟦
WHEELER DA, 2008, NATURE, V452, P872, DOI	2008	8.3463	**2010**	2011	🟥🟥🟦🟦🟦
BARTEL DP, 2004, CELL, V116, P281, DOI	2004	40.0492	**2010**	2012	🟥🟥🟥🟦🟦
KELLIS M, 2003, NATURE, V423, P241, DOI	2003	19.1	**2010**	2011	🟥🟥🟦🟦🟦
VERA JC, 2008, MOL ECOL, V17, P1636, DOI	2008	8.2479	**2010**	2011	🟥🟥🟦🟦🟦
FINN RD, 2008, NUCLEIC ACIDS RES, V36, , DOI	2008	9.9426	**2010**	2011	🟥🟥🟦🟦🟦
FINN RD, 2010, NUCLEIC ACIDS RES, V38, , DOI	2010	9.6083	**2011**	2012	🟦🟥🟥🟦🟦

Figure 15 The references with strongest citation bursts (2010-2014)

Researchers in China should firmly grasp the frontier and hot spot of genomics research, and carry out in-depth research in advantageous fields. Researchers in the field of medicine should also increase their enthusiasm for genomics research, and apply the research results to clinical practice, making this strategy of individualized treatment a reality.

3 Materials and Methods

3.1 Domestic database retrieval

This study retrieves the genomics related studies in the Chinese Journal Full-text Database (CNKI), which is used as a data source in the field of genomics research in China. The subject word was "Genomics" and the time span was limited from 1985-01-01 to 2016-4-30. The data sources were China Academic Journal Network Publishing Database, China Doctoral Dissertations Full-text Database, China Master's Theses Full-text Database, Characteristic Journal and China Mono-graphic Series Full-text Database. 10164 articles were retrieved. A total of 8512 articles excluding conference notice, agency introduction, task interview, forum essay and publicity of science popularization were used as a sample of genomics research in China.

3.2 Foreign database retrieval

This study retrieves the relevant studies on genomics in the Web of Science ™ core collection database as a foreign source of data in this field. The data sources were databases like Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index - Science (CPCI-S), Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH) and Emerging Sources Citation Index (ESCI). The subject word was "Genomics" and the time span was limited from 1985 to 2016. The literature type was ARTICLE or REVIEW. A total of 40519 articles were retrieved and used as an international research sample in the field of genomics.

3.3 Analysis methods

This research adopts CiteSpace series of information visualization software CiteSpace 4.0.R5 SE version developed by Chaomei Chen, the professor of School of Computing and Information Science in Drexel University, USA. The networks of collaboration among authors, research institutes and countries, the co-occurrence of keywords and co-citations of the retrieved literature are all analyzed. It uses visualization map to explore the research frontiers and hot spots in the field of genomics at home and abroad, and also contracts and analyzes the domestic and foreign research status of genomics.

Authors' contributions

Zhao Chunhe and Cui Weixi were the executors of design and research. They were responsible for paper conception, database retrieval, data analysis, map making, document reading and sorting, first draft writing and revision. Meng Junsheng, Zhang Xudong, Zhao Xiaoyu, and Wang Lei participated in data retrieval, data sorting and analysis. Dong Zhanjun directed research design, data analysis, paper writing and revision. All the authors have read and agreed with the final manuscript.

Acknowledgments

This study was funded by Pharmacy Department of Hebei General Hospital. Thank Meng Junsheng, Zhang Xudong, Zhao Xiaoyu, Lei Wang, Cui Weixi and Dong Zhanjun for participating in this research design, data retrieval and analysis, guidance on paper writing and revision.

References

Chen C.M., 2004, Searching for intellectual turning points: Progressive knowledge domain visualization, Proceedings of the National Academy of Sciences, 101(l): 5303-5310
https://doi.org/10.1073/pnas.0307513100
PMid:14724295 PMCid:PMC387312

Chen C.M., 2006, CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, Journal of the American Society for Information Science and Technology, 57(3): 359-377
https://doi.org/10.1002/asi.20317

Chen C.M., 2012, Predictive effects of structural variation on citation counts, Journal of the American Society for Information Science and Technology, 63(3): 431-449
https://doi.org/10.1002/asi.21694

Chen C.M., and Leydesdorff L., 2013, Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis, Journal of the Association for Information Science and Technology, 65(2): 334-351
https://doi.org/10.1002/asi.22968

Chen C.M., Ibekwe-SanJuan F., and Hou J.H., 2010, The structure and dynamics of co-citation clusters: A multiple-perspective cocitation analysis, Journal of the American Society for Information Science and Technology, 61(7): 1386-1409
https://doi.org/10.1002/asi.21309

Chen Y., Chen C.M., Hu Z.G., and Wang X., eds., 2014, Principles and applications of analyzing a citation space, Science Press, Beijing, China, pp.12, 134-144

Huang X.L., Huang S.J., Guo L.Q., and Lin J.F., 2009, Advances of metagenomics, Microbiology China, 36(7): 1058-1066

Xing Z.W., Wang Z., Gao S.H., and Wang Y.Y., 2007, Microarray and research of Chinese medications: pharmacogenomics of Chinese medications, China Journal of Chinese Materia Medica, 32(4): 289-292