Analysis of the characteristics of Global synthetic biology research and reference to technical layout

Sujuan Zhang, Xu Lu, Tianyao Sun, Rubing Yuan*

Beijing Academy of Science and Technology, 140 Xizhimen Wai Street, Beijing, 100044, China

Abstract: Background. In order to understand the research status and hotspots in the field of global synthetic biology, this paper analyzes the gaps and differences between China's synthetic biology research and foreign counterparts.
Methods. Using the scientmetric tool CiteSpace, 12040 academic papers with the theme of synthetic biology, which were included in the Web of Science core set database in the past 15 years from 2008 to 2023, were visually analyzed.
Results. The study found that: at the country level, the United States continues to lead the research in the field of global synthetic biology, and had obvious advantages and dominant positions. Through the bibliometric analysis of the literature, we found that, Synthetic biologye research focused mostly on the keywords "genome", "models", "proteins", "complex", "cells", and "gene cluster". The fragmentation of China's synthetic biology research network is prominent. **Conclusions**. On the whole, if China can fully absorband learn from the strategically oriented scientific and technological management experience in the field of synthetic biology from major competing countries such as the United States, give full play to the advantages of the national system, transform and utilize the existing achievements, and at the same time attach importance to emphasis on basic R&D investment and talent base training, expanding orderly participation, will have the opportunity to become a high ground for global synthetic biology and innovation practice in the technology industry in the future.

1. Introduction

As a frontier interdisciplinary discipline, synthetic biology, gathering and integrating life science, engineering, genomics, informatics, mathematics, chemistry, computer science and many other disciplines, had extremely broad application prospects in medicine, energy, materials, chemical industry, agriculture and other fields.[1-2] As a quantitative statistical analysis method, bibliometry can realize multi-dimensional measurement analysis of paper data.[3] Compared with the qualitative method of expert consultation, it can significantly improve the analysis efficiency. Scientific and technological papers are an important part of scientific and technological achievements, which can reflect the academic research level of scientific research subjects to a certain extent. At present the global widely used knowledge map drawing software for Cite Space, UCINET, SPSS and Bibexcel, etc., Yang Siluo, Han Ruizhen and zewen hu, Ji'an jun sun figure experts of knowledge map drawing tool after the function and the advantages and disadvantages of comparison, Dreel (Drexel) university professor Chen Chaomei Cite Space and his team developed in the literature clustering, co-occurrence and center of visualization has obvious advantages.[4] Therefore, this study adopts the basic principles and methods of bibliometry, and uses the literature analysis platform of Web of Science database, InCites database and Excel and other data processing tools to analyze the timing number of literature publications, literature producing countries, scientific research institutions, authors, journals, etc. Tool such as Cite space and Vosviewer for theme mining and visual display to interpret the main content of highly cited literature in this field. Build a social network map based on the literature reference relationship to determine the evolutionary characteristics of key technology topics. It is hoped to provide certain information support for domestic synthetic biology researchers to carry out scientific research and for managers of government departments to make relevant decisions.

2. Study protocol design and study methods

2.1. Data Collection

According to domestic databases for instance CNKI, it can be seen that the total number of scientific research papers themed on "synthetic biology" in Chinese literature is small and scattered. The research frontiers of

* Corresponding author: yuanrb@bjstinfo.ac.cn

domestic scholars and research institutions are mostly published in foreign SCI and SSCI journals. Therefore, this study focused on collecting literature data from American Science (Web of Science). The Web of Science database was selected as the data source of scientific papers, and the comprehensiveness, authority and popularity of the database were also recognized by the researchers.A total of 12113 paper data were obtained. After removing duplicate records, incomplete literature information or low content correlation through machine learning and manual reading, 12,040 articles were obtained through repeated refined search, and the search date was up to October 12, 2023. Data was loaded into the Web of Science "download_1-12040" data format, and the Web of Science 12040 bar collection underwent date "input-output" information unit separation. Make the subject layer required for establishing the network measurement model of unit nodes, namely, researcher, research unit, region, carrier, etc., namely hot words, keywords, theme and research field.

2.2. Methods

Bibliometrics.The bibliometric data was also analyzed and biblio graphic maps were constructed using Visualization of Simi larities Viewer software (VOSviewer v.1.6.15, Centre for Science and Technology Studies, Leiden University, The Netherlands). The co-authorship network, co-authorship network of countries, and keywords co-occurrence network maps were employed to present bibliographic networks. The minimum published article number by an author has been set to 12, to increase the clarity and reduce the density of co authorship network maps. Similarly, the minimum published article number by a country has been set to three, and the minimum number of occurrences of a keyword has been set to six, as inclusion criteria in the co-authorship network of countries and the keyword co-occurrence maps.

Analytic software.As mentioned above, the preprocessing of SCI paper data mainly includes extraction and co-occurrence calculation of disciplines agencies, authors and keywords, which is completed using MS Excel Vba. The co-occurrence results will be input into the clustering and visualization software Gephi to complete the clustering of disciplines and keywords and then generate clustering images. The analysis of results (e. g., the classification of SCI paper, calculation of the Crown Index and other statistic work) is largely conducted using MS Excel Vba. Furthermore, Tool such as Cite space and Vosviewer for theme mining and visual display to interpret the main content of highly cited literature in this field. Build a social network map based on the literature reference relationship to determine the evolutionary characteristics of key technology topics.[5-6]

3. Results

3.1. Global trend of publications and citations

A total of 12040 papers were collected from WoSCC database inception according to our data searching strategy. Research on synthetic biologye started in 2008 and has increased every year (Figure 1). As of the search date, all papers have been cited 35317 times. The citation rate started to speedy drop in 2018, however.

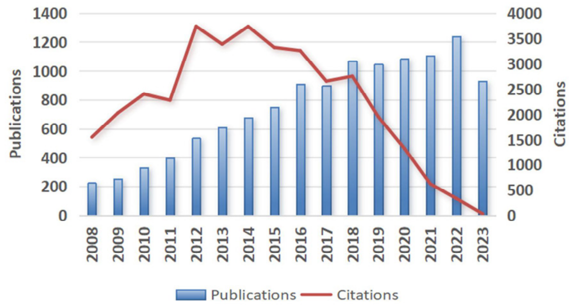

Figure 1. Global trend of publications and citations on synthetic biologye research from 2008 to 2022.

3.2. Analysis of productive countries/regions

A total of 105 countries/regions had published related articles in this field, of which the top 10 in terms of publication count are USA (4656), China (2245), UK (1296), Germany (1146), Japan (555), France (554), Spain (521), Canada (445), Switzerland (441), and Netherlands (353), accounting for 38.13%, 18.38%, 10.61%, 9.38%, 4.54%, 4.54%, 4.27%, 3.64%, 3.61%, and 2.89% of total publications, respectively.

The USA ranked first in total citations, with 200438 citations, respectively, which were both much higher than that of China in second place (44741 citations, respectively). Moreover, the USA was first in terms of average citations per paper, followed by Canada and Switzerland. Figure 3 shows the degree of cooperation between countries when the minimum number of publications was set to at least 9. The lines between nodes indicate co-authorships between countries, where a thicker line indicates stronger cooperation (total link strength [TLS]). The top 5 TLSs were associated with the USA, UK, China, Germany, and France (Figure 2).

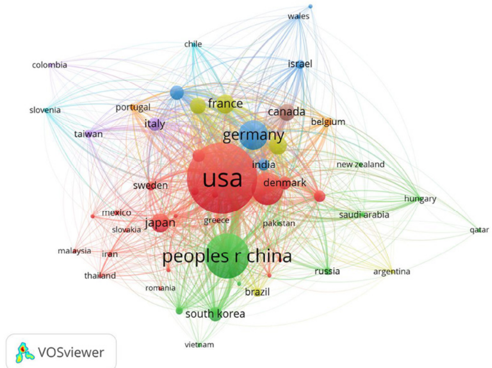

Figure 2. Citation network visualization map of countries/regions. The thickness of the lines reflects the citation strength.

3.3. Institutional and collaborative analysis

The network of national and institutional collaborations of synthetic biologye research is calculated by associating the Cite Space operational with network model "PF-NET", setting Nodes (TopN, e) = {n(i), TopN = 50, TopN% = 10%, and the literature monolith with "g2 ≤ kΣi ≤ gci, k∈z+ , k = 25" as the interval value.

Figure 3 shows the timeline visualizations of co-authorship networks of institutes constructed in the same way as for the authors. This process showed that a number of institutes are interconnected with each other. We set the threshold for the minimum number of publications by an institute at five. The program data shows that the network: N=187 and E=224. In the past 15 years, a total of 947 research institutions have carried out synthetic biologye research, mainly led by renowned universities. A quantitative disparity is evident in the collaborative network. Universities are the main force in global synthetic biologye research. From the most critical institutional cooperation map, University of California System, Chinese Academy of Sciences, UDICE-French Research Universities, United States Department of Energy (DOE) and Massachusetts Institute of Technology (MIT) have the highest prominence in synthetic biologye research. The University of California System was the leading institute with the most collaborative relationships and has a centrality of 0.0.16. It was worth noting, additionally, that the Chinese Academy of Sciences maintains a close cooperative relationship with most institutes. However, the four research institutions, Chinese Academy of Sciences in China, UDICE-French Research Universities in the France, United States Department of Energy (DOE) in the USA and Massachusetts Institute of Technology (MIT) in China, although they have a high publication output, their centrality is low. This data indicates that their research collaboration with the outside world may still need further strengthening in order to better integrate into the global synthetic biologye research network.

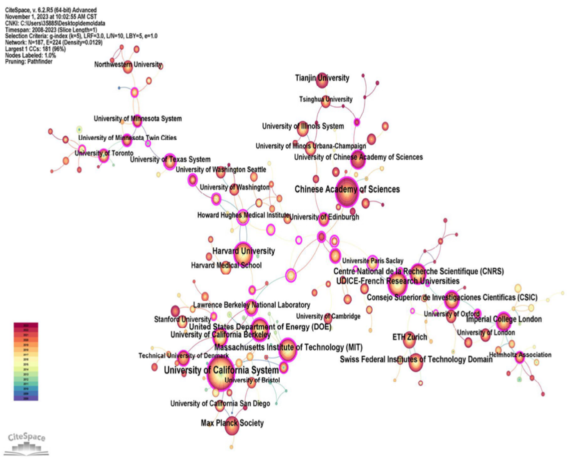

Figure 3. Institutional cooperation map.

3.4. Research content subject evolution analysis

3.4.1 Co-Occurrence of Keywords Analysis

From the analysis results of keyword co-occurrence map (Figure 4), it is found that the hot words in 2016 are mainly concentrated in the systems biology field. A large number of biological researchers had started to do systems biology, because the basic design of synthetic biology is based on a systematic understanding of the research process (from the genome to expression, to the final metabolism and product release). Therefore, well-defined genomic and metabolome studies and their bioinformatic analysis must be performed to ensure the design of engineering, and to apply it to systems biology. In 2017, "self-assembly" became a research hotspot. As synthetic technology continues to mature and iterate, creating "stable" semisythetic organisms and storing digital information in DNA, the application of "self-assembly" technology continues to expand in the field. In 2018, "metabolic engineering" became a research hotspot, and the first single chromosomal yeast strain was synthesized. The research hotspots in the past two years are "crispr (cas 9)", and "membranes". In 2020, crispr successfully cured two inherited blood diseases for the first time, and AI accurately predicted the three-dimensional structure of protein for the first time. In 2021, the artificial synthesis of carbon dioxide starch technology emerged. As a new biotechnology, it can convert carbon dioxide into starch, so as to realize the effective use of carbon dioxide and reduce environmental pollution. The research and application of this technology is of great significance to solving the global energy and environmental problems. It can be seen that the application fields of synthetic biology are mainly low-carbon biosynthesis, synthetic bioenergy, artificial synthesis and innovative application of bioactive molecules, health and medicine, agriculture and food, nano and materials, and environment. However, how to design and build intelligent and automated customized cells and gene lines to meet the needs of different practical application scenarios is a bottleneck problem to be solved urgently.

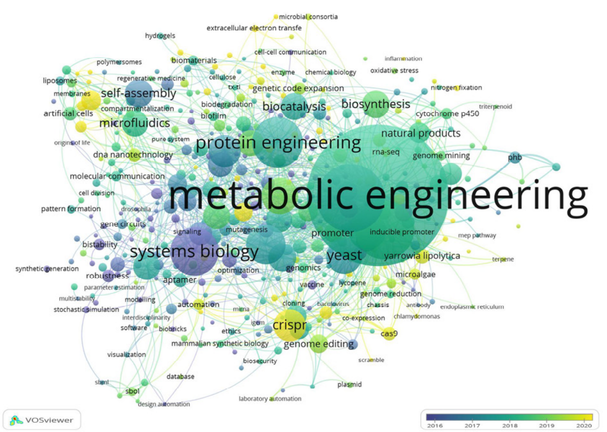

Figure 4. Co-occurrence map of keywords.

https://doi.org/10.1051/bioconf/202411103012

3.4.2 Cluster Analysis of the Keywords

On the basis of keyword co-occurrence, the acquaintance measurement of LLR in the correlation operation rules is selected for clustering, and the theme clustering time line is developed by the Timeline model under the cluster group, which can effectively grasp the dynamic change of the research in this field to explore its trend evolution and hot spot development. We demonstrate the current situation in synthetic biologye research with a timeline pattern of existing clusters. This timeline view helps to discover better the trends in the hot topics in synthetic biologye research at different periods. The modularityQ is 0.8671 and the weighted mean silhouette S is 0.9663. Nine cocitation clusters are presented with their keywords (Figure 5), They were respectively Cluster #0"DNA", Cluster #0"DNA", #1"natural products", #2" transcription", #3"system", #4"cloning", #5"identification", #6"binding", #7"evolution", #8"synthetic biology", and #9"saccharomyces cerevisiae". In the field of synthetic biology, it involved a wide range of fields, and there was some overlap between different cluster units, indicating that the co-citation phenomenon between cluster units is common.

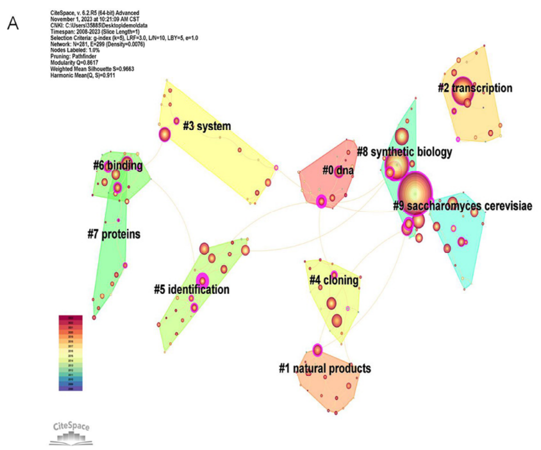

Figure 5. Visualization of the keywords Cluster analysis.

3.4.3Citation Burst Detection of Keywords

Keyword co-occurrence analysis is based on the detection of keywords with high frequency and fast growth within a certain period of time. The number of citations for some keywords has increased sharply in a short period, allowing us to detect some exciting trends. A total of 20 emergent words were detected (Figure 6). "systems biology " had the highest burst intensity, reaching 27.47, and occurred between 2008 and 2016. "noise" was the second highest, occurring between 2008 and 2014. Both are important research issues and application areas in the field of systems biology. The early explosive growth of keyword citations mainly focused on enzyme engineering, protein engineering, and genetic engineering. With further research, the hotspots of research have become more diversified. Since 2017, "synthetic organisms" began their applications in protein engineering. In addition, "protein engineering" was the latest keyword to have emerged in the last five years.The noteworthy points are the explosive references of keywords such as "mechanisms" (2020-2023, 22.8), "acid" (2019-2023, 22.08), "transformation" (2021-2023, 20.29), "in vitro" (2021-2023, 18.43), and so on, which will continue to be the future research hotspots, as indicated by their sustained prevalence until 2023.

Top 20 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2008 - 2023
systems biology	2008	27.47	2008	2016	
noise	2008	26	2008	2014	
toggle switch	2008	25.31	2008	2012	
enzymes	2009	24.29	2017	2020	
network	2008	24.02	2008	2014	
mechanisms	2008	22.8	2020	2023	
acid	2015	22.08	2019	2023	
networks	2008	21.97	2008	2015	
synthetic gene	2008	20.35	2008	2012	
transformation	2017	20.29	2021	2023	
specificity	2009	19.43	2009	2016	
in vitro	2008	18.43	2021	2023	
saccharomyces-cerevisiae	2008	16.35	2012	2018	
replication	2009	15.06	2014	2015	
pathways	2016	14.64	2016	2017	
resistance	2008	14.47	2019	2020	
protein engineering	2017	13.96	2020	2021	
in-vivo	2010	13.48	2016	2017	
gene-expression	2008	13.38	2016	2018	
transcriptional regulation	2008	12.91	2008	2013	

Figure 6. The top 20 keywords with the strongest citation bursts.

3.5. Analysis of the funding effect of the scientific research fund

Science fund is an important strategic support for the state to support basic research and cultivate the source innovation ability. Previous studies showed that the improvement of scientific research output is comprehensive: at the micro level, the fund will increase the published number of scholars[7] ; in terms of the influence of papers, the fund will increase the reference rate of published papers[8]; at the macro level, the government financial support for scientific research promotes the overall scientific research output of the countrys[9-10] .

Figure 7 showed the total number of publications and the number of funded papers in the field of synthetic biology by country. It can be seen that the top five countries in the total number of papers published in the field of synthetic biology are the same as the number of fund papers, namely the US, China, UK, Germany and Japan. The total number of US publications and the number of fund papers published was significantly higher than other countries, playing a "leader" role in the field, followed by China as a "catch-up", followed by the UK, Germany and Japan, but compared to the United States and China, the gap between the three countries was not obvious. Finally, this paper calculates the ratio between the number of papers issued by national funds and the total number of papers issued by the top five countries in the world. The proportion of articles funded by the Discovery Fund is more than 70%, which is higher than the global average, indicating that the development of synthetic biology is closely related to the fund, and its development has been highly valued by global public policy makers.

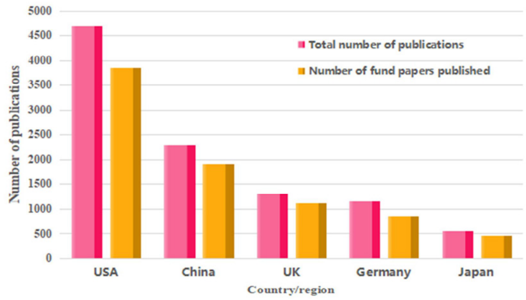

Figure 7. The total number of publications and the number of fund papers published by the top five countries in the field of synthetic biology from 2008 to 2023.

As can be seen from Table 1) From the average value of the global share of the top five countries in synthetic biology, the US accounted for 40.71% of the global share, China 15.51%, the UK 10.30%, Germany 9.23% and Japan 4.86%; From the perspective of annual growth trend, the proportion of the US in the global share had slightly decreased since 2015, China's global share had risen rapidly since 2014 (from 9.48% to 34.55%), the UK had been stable at 10% both in terms of average and annual growth, and Germany has been stable at more than 9% since 2015. Japan's global share hadslowed slightly since 2016, falling from 6.94% in 2015 to 4.30%; 2) From the proportion of the synthetic biology Global Fund papers, In 2008, the proportion of fund papers was 43.50%, and since then it has been above 60%, indicating that scientific research funding has strongly promoted the development of global synthetic biology to a large extent. From the perspective of the proportion of funded papers in the top five countries in synthetic biology (2008-2023), the US had always maintained a level of about 80% - although the private sector is also actively involved in the research and development of synthetic biology, but the US in this field is mainly the government invested a lot of money, and the funding is increasing year by year, of which the US government funded a large part of the funding for biofuel research. In the past decade, China and the UK had been stable at more than 80%. It was worth mentioning that from the trend of the proportion of fund papers, China and the UK showed a very similar trend, which is closely related to the funding of synthetic biology and the policy of the national strategic support in recent years. The proportion of funded papers in Germany and Japan has declined slightly in the past three years, which may be related to the weakening of funding for scientific research in this field in both countries.

Table 1. The publication volume of the top five countries accounted for the global share and trend unit from 2008 to 2023, (Unit:%)

Years	US's global share	The proportion of US's fund papers	China's global share	The proportion of Chinese fund papers	UK's global share	The proportion of UK's fund papers	Germany's global share	The proportion of German fund papers	Japan's global share	The proportion of Japanese fund papers
2023	28.74	76.78	34.55	77.57	7.75	80.56	8.18	72.37	3.77	71.43
2022	28.40	81.13	30.32	81.27	10.64	81.95	8.16	76.47	3.20	77.50
2021	34.91	76.69	25.20	80.21	11.29	86.05	10.06	85.22	3.85	88.64
2020	35.92	80.96	23.06	86.56	10.67	88.03	10.39	79.82	4.47	91.84
2019	38.09	85.64	21.50	89.22	13.25	90.91	11.31	76.23	3.52	84.21
2018	39.44	81.75	17.57	88.83	11.78	93.65	11.78	79.37	4.86	80.77
2017	44.21	83.12	16.70	88.67	11.25	88.12	9.69	81.61	4.90	93.18
2016	39.47	85.20	14.44	89.31	12.02	90.83	9.92	61.11	4.30	79.49
2015	41.39	81.94	13.62	90.20	10.01	89.33	9.75	75.34	6.94	80.77
2014	45.93	86.77	9.48	81.25	11.70	86.08	7.70	63.46	6.52	88.64
2013	43.68	79.70	8.70	86.79	10.67	75.38	9.85	70.00	6.24	76.32
2012	49.34	82.51	9.94	84.91	8.26	86.36	7.13	71.05	3.75	75.00
2011	45.98	85.79	5.53	72.73	10.55	78.57	9.55	60.53	4.77	89.47
2010	46.67	81.82	6.67	81.82	9.39	83.87	8.18	74.07	6.06	90.00
2009	45.20	86.73	6.80	82.35	8.40	85.71	5.20	61.54	5.60	71.43
2008	43.95	65.31	4.04	44.44	7.17	81.25	10.76	37.50	4.93	45.45
mean value	40.71	81.37	15.51	81.63	10.3	85.42	9.23	70.36	4.86	80.26

4. Conclusions

Through the bibliometric analysis of the literature, we found that, Synthetic biologye research focused mostly on the keywords "genome", "models", "proteins", "complex", "cells", and "gene cluster". It can be seen that the application fields of synthetic biology are mainly low-carbon biosynthesis, synthetic bioenergy, artificial synthesis and innovative application of bioactive molecules, health and medicine, agriculture and food, nano and materials, and environment. Synthetic biology has great development prospects in the fields of exploring the mysteries of life, preparing clean energy, environmental governance, drug research and development, vaccine development, medical and health care and so on.

Over the past 15 years, scientific research funding investment in synthetic biology has maintained rapid growth around the world, and government funding is playing an increasingly important role in scientific research activities. The scientific research fund in this field has a larger scale and stronger guidance than before, and the science fund is facing a key process from quantitative change to qualitative change. Cooperation is, however, gradually improving. All five continents are involved in Synthetic biologye research, forming a complex network of collaborations, with the United States and China leading the way in synthetic biologye research. The US has the most complex and diverse collaborations, although not as many as China in terms of the volume of published research emerging from these collaborations.

Compared with the United States and the European Union, although China's synthetic biology started a little later, it is now in a state of following and running. China's biological manufacturing has thus entered the 5G era of "high-throughput screening, high-level breeding, high-activity products, high-precision regulation, and efficient transformation in various fields and scenarios." In the past five years, China has gradually strengthened top-level strategic planning, attaching importance to basic research and macro deployment of the technology industry. Shanghai, Guangdong, Tianjin, Chongqing and other local governments have also included synthetic biology in the 14th Five-Year Plan for scientific and technological innovation. On May 10, 2022, the National Development and Reform Commission issued the "Fourteenth Five-Year Plan" for Bioeconomy Development, which made it clear that the bioeconomy, including synthetic biology, is a new driving force for China's economic transformation in the future. In that year, the Ministry of Science and Technology invested 186 million yuan to support the national key research and development plan "Synthetic Biology". Focusing on the deployment of three tasks: "Artificial genome synthesis and high-version chassis cells", "artificial components and gene circuits", and "functional synthetic biological systems", the company will build practical major artificial biological systems and innovate cutting-edge synthetic biological technologies to promote the innovative development of the biological industry and the green growth of the economy.

Oligonucleotide synthesis, eukaryotic cell and prokaryotic cell technology were currently in the outbreak stage, and the technology was gradually mature, becoming the emerging technology point with the highest market attention. Enterprises should be prepared and timely adjust the development strategy. In a word, novel synthetic biology including nucleic acid, genome, epigenome, and protein synthesis and engineering, including design tools is the future research direction and technology competition highland.

Acknowledgement

Funded by the Youth Cultivation Project of Beijing Academy of Science and Technology: Research on the path of Accelerating the cultivation of world-class enterprises in Beijing -- A case study in the field of synthetic biology

References

1. Knuuttila T, Loettgers A. (Un) Easily Possible Synthetic Biology [J]. Philosophy of Science. 2022, 89(5):908-917.
2. Zhang XE, Liu CL, Dai JB, et al. Enabling technology and core theory of synthetic biology [J]. Science China Life Sciences. 2023, 66(8):1742-1785.
3. Zhou L, Zhang L, Zhao Y, et al. A Scientometric Review of Blockchain Research. Inf. Syst. E-Bus. Manag. 2021, 19, 757–787.
4. Nielsen AA, Der BS, Shin J, et al. Genetic circuit design automation [J]. Science. 2016, 352(6281)
5. Verma MK, Khan D, Yuvaraj M. Scientometric assessment of funded scientometrics and bibliometrics research (2011-2021) [J]. Scientometrics. 2023, 128(8):4305-4320.
6. Grinev, AV. The Scientometric Portrait of a Scientist as a Tool to Assess His or Her Achievements [J]. Herald of the Russian Academy of Sciences. 2022, 92(2):141-150.
7. Gush J, Jaffe A, Larsen V, et al The effect of public Funding on research output: The New Zealand Marsden Fund [J]. New Zealand Economic. 2018, 52(2):227-248
8. Gonzalez-Brambila C N, Veloso F M, Krackhardt D. The impact of network embeddedness on research output [J]. Research Policy, 2013, 42(9):1555-1567.
9. Gyorffy B, Nagy A M, Herman P, et al. Factors influencing the scientific performance of Momentum grant holders: An evaluation of the first 117 research groups [J]. Scientometrics, 2018, 117(1):409-426.
10. Baruffaldi S H, Marino M, Visentin F. 2020. Money to move: The effect on researchers of an international mobility grant [J]. Research Policy, 49(8):1-18.