Medicine®

Visualization and analysis of mapping knowledge domains for coronavirus research

Weizhi Yao[a],[*] [ID], Ling Tan[b], Liang Liu[a]

Abstract

Background: In recent years, many countries around the world have been threatened by COVs. The aim of this study was to better grasp developments and trends in research on coronavirus around the world and to promote theoretical research into their prevention and control.

Methods: Research on coronavirus was reviewed and analyzed using bibliometrics based on a total of 4860 publications collected from the Web of Science Core Collection database. Yearly quantitative distribution of literature, country/region distribution, organization distribution, main source journal distribution, subject category distribution, research knowledge bases, and research hotspots and frontiers were all analyzed, and CiteSpace and VOSviewer were used to plot knowledge domain maps, Excel was used to plot keyword strategy diagram.

Results: Coronavirus research could be roughly divided into 4 stages: preliminary development stage (before 2000), rapid growth stage (2000–2005), slow decline stage (2006–2011) and sustained growth stage (since 2012). America had taken the leading position in this field. The study of COVs involves many subject categories, mainly includes virology, veterinary sciences, biology, and immunology. At present, the key words in the field of coronavirus research were mainly divided into 6 major hot clusters, namely, the introduction and structure analysis of coronavirus, the research on the outbreak source and transmission of coronavirus, the research on the infection pathway of coronavirus in human body, the research on the pathogenesis of coronavirus, the research on the diagnosis and symptoms of coronavirus infection, and the research on the treatment of coronavirus.

Conclusion: Coronavirus, which occurs all over the world, often causes huge casualties and economic losses, and poses a serious threat to the safe and stable operation of the social and economic system. Objective literature review and analysis can help scholars in related fields to deepen their overall understanding. And, there are several key issues that should be further explored in future research.

Abbreviations: 2019-nCoV = 2019 new coronavirus, COVs = coronavirus, NIAID = National Institute of Allergy and Infectious Diseases, SARS = severe acute respiratory syndrome, SARS-CoV = sever acute respiratory syndrome coronavirus, WoS = web of science.

Keywords: bibliometric analysis, coronavirus, mapping knowledge domains, visualization

1. Introduction

Coronavirus (COVs) is a type of enveloped RNA virus with non-segmented genomes of about 30,000 nucleotides. It can infect mammals and birds and mainly causes respiratory and gastrointestinal diseases Gonzalez et al.[1] In recent years, many countries around the world have been attacked by COVs, such as the Sever acute respiratory syndrome (SARS-CoV) that originated in Guangdong, China in 2002, and the Middle East respiratory syndrome (SRS) that broke out in Saudi Arabia in 2012 (Middle East Respiratory Syndrome, MERS-CoV) and the

2019 New Coronavirus (2019-nCoV), have attracted widespread attention from the government and academia. It is important that the current outbreak of 2019-nCoV, caused by the severe acute syndrome coronavirus-2, has been considered as a major anxiety of the twenty first century Dey et al.[2] The common symptoms of people infected with 2019-nCoV include respiratory symptoms, fever, cough, shortness of breath and dyspnea Cw et al.[3] In more serious cases, infection can lead to pneumonia, severe acute respiratory syndrome, renal failure and even death (D. Thomas-Rüddel et al.[4] The main routes of transmission of 2019-nCoV are respiratory droplet transmission

and contact transmission. Aerosol and fecal-mouth transmission pathways need to be further clarified. People of all ages may be infected, mainly adults, among whom the elderly and the weak and sick seem to be more likely to be infected Yang et al,[5] and hence researching on 2019-nCoV seems to be a necessary strategy to control the current pandemic of 2019-nCoV in this critical time.

With the continuous deepening of the process of global trade liberalization, the mutual exchanges between countries around the world have gradually deepened, and COVs originating from 1 region can spread to other countries or regions in various ways. In addition, the convenience of transportation, such as the increasing popularity of high-speed rail and aircraft, also provides a feasible way for the spread of coronavirus Putri et al.[6] As a result, these coronaviruses that have spread out all over the world often cause huge casualties and economic losses, increasing the risk of the safe and stable operation of socioeconomic systems Al-Tawfiq et al.[7] Due to the complexity of COVs and its serious social impact, scholars have conducted a lot of research on COVs, including COVs genomic RNA packaging Masters,[8] COVs-induced ER stress response Fung et al,[9] treatments of the COVs, effective control of outbreaks Al-Tawfiq et al.[10]

However, bibliometrics have rarely been used to analyze the development of research on COVs by mapping knowledge domains. Bibliometrics is a research method for processing a large number of references. It has knowledge-oriented functions and belongs to scientific metrology Shiffrin et al.[11] We can understand the status of a research field and the development trends for related research by using visualization methods and mapped knowledge domain analysis to study the distribution laws, quantitative relationships, and internal relationships within the literature. Therefore, this paper attempts to sort out the research status, hotspots, deficiencies and trends of COVs through objective bibliometric analysis methods. The contribution to the research of COVs lies in the analysis of existing COVs in the form of bibliometric analysis, analysis and discussion of key issues, and pointing out the shortcomings of the current research, providing a useful reference for subsequent research.

The rest of the paper is organized as follows: Section 2 introduces the data sources and research methods; Section 3 uses a bibliometric analysis method to objectively analyze the related publications, including the yearly quantitative distribution of COVs research, country/region distribution, organization distribution, main source journal distribution, subject category distribution, research knowledge bases, and research hotspots and frontiers; Section 4 discusses the results of the above studies; Section 5 makes a conclusion.

2. Data sources and research methods

2.1. Data source

The data used in this study come from the Web of Science (WoS) core collection database. The database contains more than 10,000 multi-disciplinary, high impact, international, authoritative and comprehensive academic journals, and is the most academically authoritative source of citation information in the world Zhao and Liu et al.[12–13] Drawing on previous research, the following search code was used in our analysis: TI = Coronaviru*. The language was "English," the literature type was "Article," and the time span of articles was set as "1990–2020." "TI" stands for the title of the

Table 1

The process and results of literature identification and selection.

Process and results	
Data source	Web of science core collection
Search format	TI = Coronavirus*
Literature language	English
Literature type	Article
Time span	1900-2020
Retrieval date	2020–02-06
Search results	4860 literature

publication, and in this research, only journal articles were used as valid data, because through peer review, the quality of the articles is higher, and covers more extensive and authoritative information. The data was last updated on February 6, 2020, and WoS Document Information Management Online System was used to output the literature information data (full records and cited references). At the initial stage of the study, downloaded data were first deduplicated using CiteSpace, and the final sample size was 4860 articles. These papers were authored in 79 countries by 2601 organizations; they spanned 80 subject categories, comprised 680 source journals, and contained 14475 authors. In this paper, we included all the organizations, subject categories, journals and authors in the paper into the analysis to describe the relationship and strength among them better. We set the maximum look back years to "-1" in the analysis so that all the documents in the sample data would be included. When analyzing the data, years per slice were set to "1 year." The process and results of literature identification and selection are shown in Table 1.

2.2. Research methods

The knowledge map is a graphic showing the relationship between the development process and the structure of scientific knowledge. It uses a certain method to map abstract data into 2D or 3D graphics, and reveals the development of a field and discipline from the macro, meso and micro levels. The overview enables people to comprehensively review the structure of a discipline and research hotspots from all angles Chen, Chaomei.[14]

The data (i.e., the annual number of publications, country, organization, journal, subject, author, h-index, citation and so on) were collected using the analysis function from the WOS core collection and were analyzed after drawing and tabulation. Moreover, knowledge domain maps for country/region co-authorship, organization co-authorship, author co-authorship, journal co-citation and keyword co-occurrence were plotted using VOSviewer Van Eck and Waltman and Van Eck and Waltman[15–16]; keyword time zone were plotted using Cite-Space Chen Chaomei[17]; yearly quantitative distribution of the literature and keyword strategy diagram were plotted using Excel.

3. Results

3.1. Yearly quantitative distribution of the literature

The spatiotemporal change of the volume of literature is an important indicator for evaluating the development of a field,
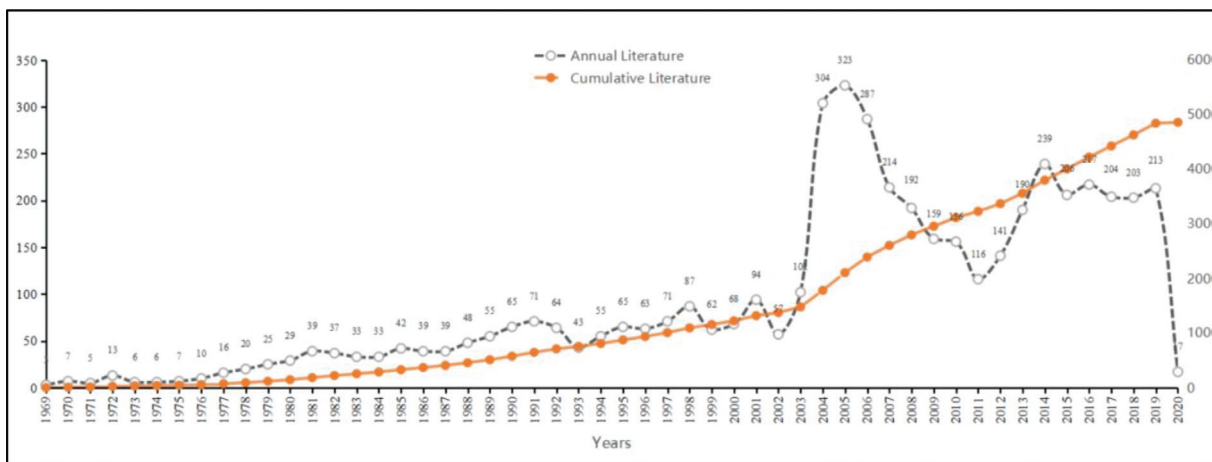
Figure 1. Publication of COVs related literature from 1969 to 2020.

which is of great significance to evaluate the development trend and dynamic evolution of the field. This paper first describes the number of literature and trends since 1969 (see Fig. 1). As a whole, the literature on coronavirus research is increasing. From the perspective of specific growth trends, research in this area can be roughly divided into 4 stages. The first stage is the preliminary development stage (before 2000). The number of literature published during this stage increased slowly year by year, from 3 in 1969 to 68 in 2000, with an average of 38 published articles per year. The second stage is the rapid growth stage (2000–2005). Due to the outbreak of SARS-CoV and other coronaviruses, its impact has attracted great attention from scholars. The research in this stage increased rapidly and the number of literature peaked at 323 in 2005. The third stage is the slow decline stage (2006–2011). Coronavirus research holds the characteristics of periodicity and timeliness. When the large-scale epidemic is over, its research fever will gradually decline. The fourth stage is the sustained growth stage (since 2012). During this period, the average annual volume of literature has accelerated, from 116 in 2012 to 213 in 2019, with an average annual literature of 87.3. The reason for the small number of papers issued in 2020 is that the search period is until February 2020, so the literature for that year is incomplete.

3.2. Quantitative analysis of productive countries/regions

To identify the main countries/regions involved in research on Coronavirus and the cooperation among them, country/region distribution was analyzed. Countries/region distribution and the mapped knowledge domains for the co-authoring countries/regions are shown in Table 2 and Figure 2, respectively.

There are as many as 79 countries in the world that are involved in research on COVs. As shown in the Table 2, the United States clearly dominates the number of publications in this field (1525, 31.38%), followed by China (755, 15.53%) and Germany (294, 6.05%). A network map was created using VOSviewer to visualize the geographical distribution of countries or regions contributing to the field of COVs, which is presented in Figure 2, matching the general ranking trends in Table 2. Although China ranked second in number of literature, the overall performance in terms of centrality was inferior to that of Germany and Netherlands. Thus, nearly one third of the top

20 institutes contributing to this field are located in the USA, suggesting that American scientists have taken the leading position in this field. Indeed, this has been accompanied by huge benefits to the population.

The knowledge domain map of co-authoring countries/regions is shown in Figure 2. The nodes represent different countries, and the size of the nodes represents activity and article number. Links between 2 nodes indicate that they have a cooperative relationship. The thicker the connection, the stronger the cooperation between the 2 countries. In Figure 2, it can be seen that cooperation between the USA and Germany is the most frequent, followed by USA-England, USA-Netherlands and USA-China. The total link strength for the USA and Germany is similar (the total link strength is 8.70 for the USA and 8.38 for Germany), but international cooperation with China is mainly concentrated in the USA, Germany, Australia, Netherlands, England and France.

3.3. Quantitative analysis of the main research organizations

Information on the most productive organizations that specialize in a certain theme can be identified via analysis of organizational cooperation Hong et al.[18] To discover the main research organizations, we listed the top 10 organizations with the most publications (see Table 3) and plotted the knowledge domain

Table 2

Top 10 countries/regions with the most literature in COVs studies.

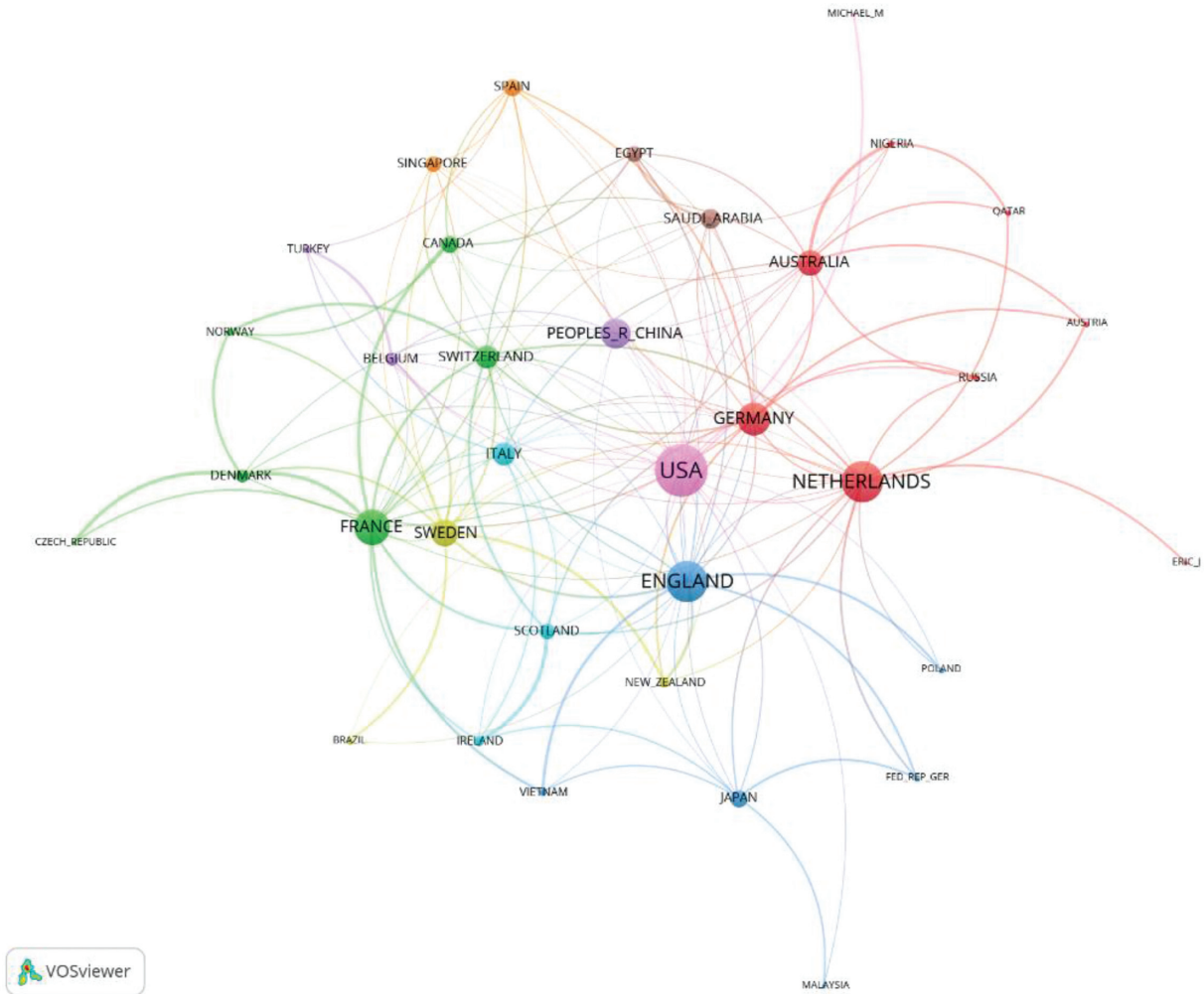| Rank	Country/region	Number of literature	Percentage of total	Centrality
1	USA	1525	31.38%	0.59
2	PEOPLES R CHINA	755	15.53%	0.10
3	GERMANY	294	6.05%	0.30
4	NETHERLANDS	259	5.33%	0.18
5	JAPAN	239	4.92%	0.09
6	CANADA	215	4.42%	0.07
7	ENGLAND	191	3.93%	0.15
8	TAIWAN	155	3.19%	0.00
9	FRANCE	147	3.02%	0.13
10	SAUDI ARABIA	137	2.82%	0.06

Figure 2. Mapping knowledge domains of co-authoring countries/regions in COVs studies.

map for co-authoring organizations (Fig. 3) using VOSviewer. Among these organizations in Table 3, organizations come from USA and China, and 1 organization each is from Holland and Spain. The University of Hong Kong has the most publications at 202. Utrecht University ranks second with 127 publications, and National Institute of Allergy and Infectious Diseases (NIAID) ranks third with 111 publications. In terms of Centrality, Utrecht University ranks first with 0.12, followed by NIAID with 0.08. The Centrality for the Superior Council of Scientific Investigations in Spain, whose publications rank 9th, is 0.05, and it ranks third.

Mapped knowledge domains for co-authoring organizations are plotted in Figure 3. Each node represents an organization, and its size is related to its publication number. The thickness of

Table 3

Top 10 organizations with the most publications in COVs studies.

| Rank	Organization	Country	Literature	Proportion	Centrality
1	The University of Hong Kong	China	202	4.16%	0.04
2	Utrecht University	Holland	127	2.61%	0.12
3	National Institute of Allergy and Infectious Diseases	USA	111	2.28%	0.08
4	Chinese Academy of Sciences	China	107	2.20%	0.04
5	University of Texas	USA	86	1.77%	0.03
6	The University of North Carolina	USA	83	1.71%	0.04
7	The University of Iowa	USA	76	1.56%	0.04
8	University of Southern California	USA	64	1.32%	0.03
9	Superior Council of Scientific Investigations	Spain	61	1.26%	0.05
10	University of Pennsylvania	USA	59	1.21%	0.02
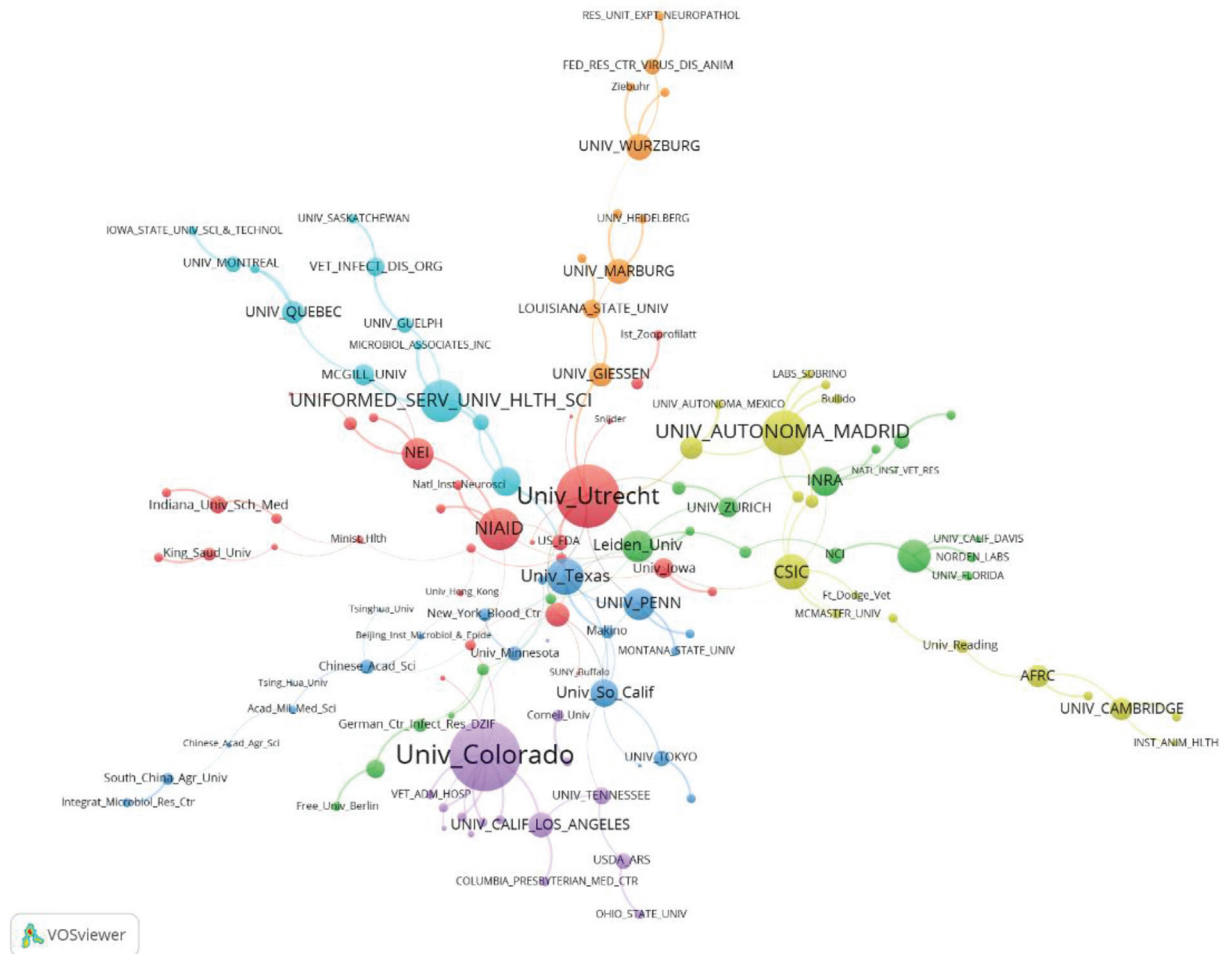
Figure 3. Mapping knowledge domain of co-authoring organizations in COVs studies.

the connecting lines indicates the strength of cooperation between 2 organizations. The 2 sets of research organizations with the closest cooperation are (NIAID, USA) and National Eye Institute (USA) and Utrecht University (Holland) and University of Giessen (Germany). The total link strength for the University of Colorado and Utrecht University is similar (the total link strength is 8.46 for University of Colorado and 7.97 for Utrecht University), but the total link strength of the University of Hong Kong is just 0.64 whose publications rank first, suggesting that COVs studies of the University of Hong Kong is relatively independent.

3.4. Quantitative analysis of main source journals

Journals are the most important sources for academic communication and the dissemination of scientific achievements. The core journals for a field can be identified by analyzing journal distribution. Based on the retrieved results, the top 10 journals with the most literature are listed in Table 4, and they are all included in SCIE. In Table 4, there are 5 journals from the USA, 3 journals from Netherlands, and 1 journal each from the England and Germany. In terms of literature, *Journal of Virology* ranks first with 772 literature, followed by *Virology* with 282 literature and *Journal of General Virology* with 181 literature. Although *Emerging Infectious Diseases* rank 8th in number of literature,

the SJR 2018 rank first. Obviously, these journals focus on the theme of virology and diseases.

3.5. Subject category distribution for the literature

Given the complexity of COVs and the breadth of its serious impact, the study of COVs involves many subject categories, including but not limited to Virology, Veterinary Sciences, Microbiology, Biochemistry & Molecular Biology, and Infectious Diseases. Virology has the most literature with 2044. The highest centrality belongs to Immunology, whose literature ranks 6th. In general, research on COVs mainly focuses on virology, veterinary sciences, biology and immunology (see Table 5).

3.6. Knowledge bases for COVs studies

The concept of co-citation was proposed by Henry Small, who is an American intelligence scientist. Co-citation means that 2 publications were cited together in the references of the third publication Small.[19] The research knowledge base is made up of co-cited articles. Moreover, co-cited articles are published by journals that report on the state of the knowledge base. High co-citation journals also represent core journals at the forefront of research. To identify core journals and the knowledge base for

Top 10 journals with the most literature in COVs studies.

Rank	Journal title	Country	Literature	Proportion	H index	SJR 2018	Citations index
1	*Journal of Virology*	USA	772	15.88%	271	2.59	SCIE
2	*Virology*	USA	282	5.80%	162	1.64	SCIE
3	*Journal of General Virology*	England	181	3.72%	152	1.32	SCIE
4	*Advances In Experimental Medicine And Biology*	USA	128	2.63%	107	0.65	SCIE
5	*Virus Research*	Netherlands	124	2.55%	104	1.09	SCIE
6	*Archives of Virology*	Germany	109	2.24%	102	0.91	SCIE
7	*Veterinary Microbiology*	Netherlands	92	1.89%	114	1.17	SCIE
8	*Emerging Infectious Diseases*	USA	91	1.87%	202	3.14	SCIE
9	*Journal of Virological Methods*	Netherlands	77	1.58%	91	0.78	SCIE
10	*Plos One*	USA	76	1.56%	268	1.10	SCIE

the field of COVs, the VOSviewer tool was used to analyze co-citation publications and their source journals.

3.6.1. Journal co-citation analysis. Co-citation of journals is described by mapping the knowledge domain of co-citation journals, as shown in Figure 4. A line connecting 2 journals means that these 2 journals were cited in the same literature, and the thickness of the connecting line represents the strength of co-citation between these 2 journals.

The biggest node is for the *Journal of Virology*, meaning that the *Journal of Virology* is the journal most frequently cited with other journals. In terms of co-citation strength, the line connecting *Journal of Virology* and *Journal of General Virology* is the thickest, indicating that *Journal of Virology* and *Journal of General Virology* were co-cited most frequently. The connecting line between *Research in Veterinary Science* and *Avian Diseases* is the second thickest, followed by *Nature* and *Virology*.

3.6.2. Literature co-citation analysis. To determine the distribution of the most influential literature in the field of COVs research, we collected the top ten articles with the most co-citations, as shown in Table 6. Eight articles were published in 2003, one was published in 2004, and one was published in 2012. Additionally, all the ten articles were co-authored.

As for subject, 8 of the 10 literature were studied the characteristic and transmission of SARS-Cov, 1 article was related to a fourth human coronavirus, HCoV-NL63, and one paper was about reporting a novel betacoronavirus species whose closest known relatives are bat coronaviruses HKU4 and HKU5.

The paper with the most citations is "Characterization of a novel coronavirus associated with severe acute respiratory syndrome," which was written by Rota et al[20] and published in Science. The authors presented the initial characterization of the viral genome of SARS-CoV. The article with the second most citations was written by Ksiazek et al.[21] They attempted to identify potential pathogens of SARS-CoV. The third-ranked paper searched for unknown viruses with the use of cell cultures and molecular techniques of SARS-CoV Drosten et al.[22]

A paper discussing a fourth human coronavirus (HCoV-NL63) entitled "Identification of a new human coronavirus" was published in Nature Medicine in 2004 van der Hoek et al.[23] This paper reported the identification of a fourth human coronavirus, HCoV-NL63, using a new method of virus discovery, which was isolated from a 7-month-old child suffering from bronchiolitis and conjunctivitis. The complete genomic sequence indicates that the virus is not a recombinant virus, but a new group 1 coronavirus, and the virus has spread widely with the human population. An article studying a novel betacoronavirus species whose closest known relatives are bat coronaviruses HKU4 and HKU5 was authored by Zaki et al[24] and published in New England Journal of Medicine in 2012. They introduced a previously unknown coronavirus (called HCoV-EMC) which was isolated from the sputum of a 60-year-old man, and the results showed that it was remarkably similar to that of the severe acute respiratory syndrome (SARS) outbreak in 2003 and reminded us that animal coronaviruses can cause severe disease in humans.

Top 10 disciplines with the most literature in COVs studies.

Rank	Subject category	Literature	Proportion	Centrality
1	Virology	2044	42.06%	0.11
2	Veterinary Sciences	698	14.36%	0.09
3	Microbiology	607	12.49%	0.13
4	Biochemistry & Molecular Biology	557	11.46%	0.20
5	Infectious Diseases	554	11.40%	0.06
6	Immunology	547	11.26%	0.32
7	Biotechnology & Applied Microbiology	332	6.83%	0.10
8	Medicine	309	6.36%	0.09
9	Science & Technology - Other Topics	234	4.81%	0.04
10	Research & Experimental Medicine	228	4.69%	0.07

Figure 4. Mapped knowledge domains for journal co-citation in COVs studies.

Top 10 literature with the most citations in COVs studies.

Rank	Title	Journal	Authors	Year	Co-citations
1	Characterization of a novel coronavirus associated with severe acute respiratory syndrome	Science	Rota P A, Oberste M S, Monroe S S, et al[16]	2003	615
2	A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome	New England Journal of Medicine	Ksiazek T G, Erdman D, Goldsmith C S, et al.[17]	2003	612
3	Identification of a novel coronavirus in patients with severe acute respiratory syndrome	New England Journal of Medicine	Drosten C, Günther S, Preiser W, et al.[18]	2003	553
4	Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia	New England Journal of Medicine	Zaki A M, Van Boheemen S, Bestebroer T M, et al.[19]	2012	544
5	The genome sequence of the SARS-associated coronavirus	Science	Marra M A, Jones S J M, Astell C R, et al.[20]	2003	519
6	Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study	Lancet	Peiris J S M, Chu C M, Cheng V C C, et al.[21]	2003	465
7	Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus	Nature	Li W, Moore M J, Vasilieva N, et al.[22]	2003	295
8	Unique and Conserved Features of Genome and Proteome of SARS-coronavirus, an Early Split-off From the Coronavirus Group 2 Lineage	Journal of Molecular Biology	Snijder E J, Bredenbeek P J, Dobbe J C, et al.[23]	2003	275
9	Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China	Science	Guan Y, Zheng B J, He Y Q, et al.[24]	2003	239
10	Identification of a new human coronavirus	Nature Medicine	van der Hoek L, Pyrc K, Jebbink M F, et al.[25]	2004	214

Top 10 authors with the most literature in COVs studies.

Rank	Authors	Country	Organization	Literature	Proportion	H-index	Total Link Strength
1	Yuen KY	China	The University of Hong Kong	104	2.14%	92	18.41
2	Enjuanes L	Spain	Superior Council of Scientific Investigations	102	2.10%	54	33.88
3	Perlman S	USA	University of Iowa	99	2.04%	49	14.72
4	Drosten C	Germany	Humboldt University of Berlin	87	1.79%	67	29.62
5	Rottier PJM	Netherlands	Utrecht University	86	1.77%	59	19.23
6	Baric RS	USA	University of North Carolina	82	1.69%	67	25.20
7	Weiss SR	USA	NIH National Institute on Drug Abuse (NIDA)	77	1.58%	49	17.58
8	Makino S	USA	University of Texas Medical Branch Galveston	72	1.48%	49	13.94
9	Chan KH	China	University of Hong Kong	65	1.34%	71	14.44
10	Liu DX	China	South China Agricultural University	64	1.32%	36	6.10

3.7. Research hotspots and frontier analysis

3.7.1. Authors of hotspot literature.
Because many subject categories are involved in the study of COVs, 14475 relevant authors were obtained from the WOS core database. We list the top 10 authors with the most literature and include their countries and organizations, literature, proportions, h-index and total link strength in Table 7.

Total link strength, which combines a number of co-authors and frequency of co-authorship, is derived from VOSviewer. Among these authors, 4 are in USA, 3 are in China and the rest are in Spain, Germany and Netherlands. The author who published the largest number of articles,104 in total, is Yuen KY from the University of Hong Kong (his h-index is also ranked first). The second largest node is Enjuanes L from the Superior Council of Scientific Investigations. Perlman S from the University of Iowa ranked third with 99 literature. In terms of co-author intensity, Enjuanes L from Superior Council of Scientific Investigations (Enjuanes_L in Fig. 5) ranked first with a total link strength of 33.88, indicating a high number of co-authored papers and a high number of co-authors, followed by Drosten C and Baric RS.

Creating and analyzing knowledge maps for the co-authorship network can provide valuable information that can help research organizations develop cooperative groups, help individual researchers find partners, and help publishers look for editorial teams. A knowledge domain map for co-authorship was plotted using the VOSviewer software, as shown in Figure 5. Each node represent an author, and the node sizes indicate the number of co-authored published articles. Links between 2 nodes represent collaborations, with greater link width representing closer collaboration between the authors. The colors represent different author cooperation clusters. There are some closely collaborating author clusters, such as Yuen KY (University of Hong Kong), Baric RS (University of North Carolina), Lai MMC (University of Southern California) and Makino S (University of Texas Medical Branch Galveston). Core clusters also show connections between different subclusters, such as the red and green clusters with Yuen KY (University of Hong Kong) and Drosten C (Humboldt University of Berlin) at their core. In general, most studies on COVs have been cooperative research projects within the same institution, supplemented by a small number of interagency cooperations.

3.7.2. Keywords co-occurrence analysis.
Keywords co-occurrence analysis which studies the co-occurrence of keywords in a large number of articles has been used to describe the core content and structure for certain academic domains and to reveal research frontiers for that subject category. The number of keywords used for COVs studies was extracted from the Web of Science core database and is as high as 328. Using the VOSviewer software for cluster analysis a knowledge domain map for keyword co-occurrence was generated and a total of 6 clusters were obtained as shown in Figure 6. Nodes represent keywords and node size indicates the occurrence frequency of the keyword. Node connection thickness indicates the strength of co-occurrence between keywords. According to the clustering diagram, it can be found that the keywords in the field of coronavirus research are mainly divided into 6 main hotspot clusters, namely, the introduction and structure analysis of coronavirus, the research on the outbreak source and transmission of coronavirus, the research on the infection pathway of coronavirus in human body, the research on the pathogenesis of coronavirus, the research on the diagnosis and symptoms of coronavirus infection, and the research on the treatment of coronavirus.

Cluster 1 (Blue): The blue cluster is mainly centered around the keywords "coronavirus," "crystal structure," "sars coronavirus," "murine coronavirus," "2019-nCoV" and so on, which contains 58 keywords and it is closely linked to core keywords of other clusters. In general, keywords in cluster 1 mainly focus on the introduction and structure analysis of coronavirus. Rota et al[20] thought that the coronaviruses are a diverse group of large, enveloped, positive-stranded RNA viruses that cause respiratory and enteric diseases in humans and other animals. Yang et al[25] compared the virus surface spikes of MERS-CoV and a related bat coronavirus, HKU4. The results showed that although HKU4 spike cannot mediate viral entry into human cells, 2 mutations enabled it to do so by allowing it to be activated by human proteases. Aleanizy et al[26] described the demographic characteristics, mortality rate, clinical manifestations and comorbidities with confirmed cases of MERS-CoV using non parametric binomial test and Chi-Squared test. Huang et al[27] reported clinical features of patients infected with 2019 novel coronavirus in Wuhan and found that the 2019-nCoV infection could cause clusters of severe respiratory illness similar to severe acute respiratory syndrome coronavirus and was associated with ICU admission and high mortality. Moreover, Chan et al[28] studied a familial cluster of pneumonia associated with the 2019 novel coronavirus and found that 2019-nCoV can spread from person to person.

Cluster 2 (Yellow). The main keywords in the yellow cluster are "antibody," "outbreak," "epidemiology" and so on, which
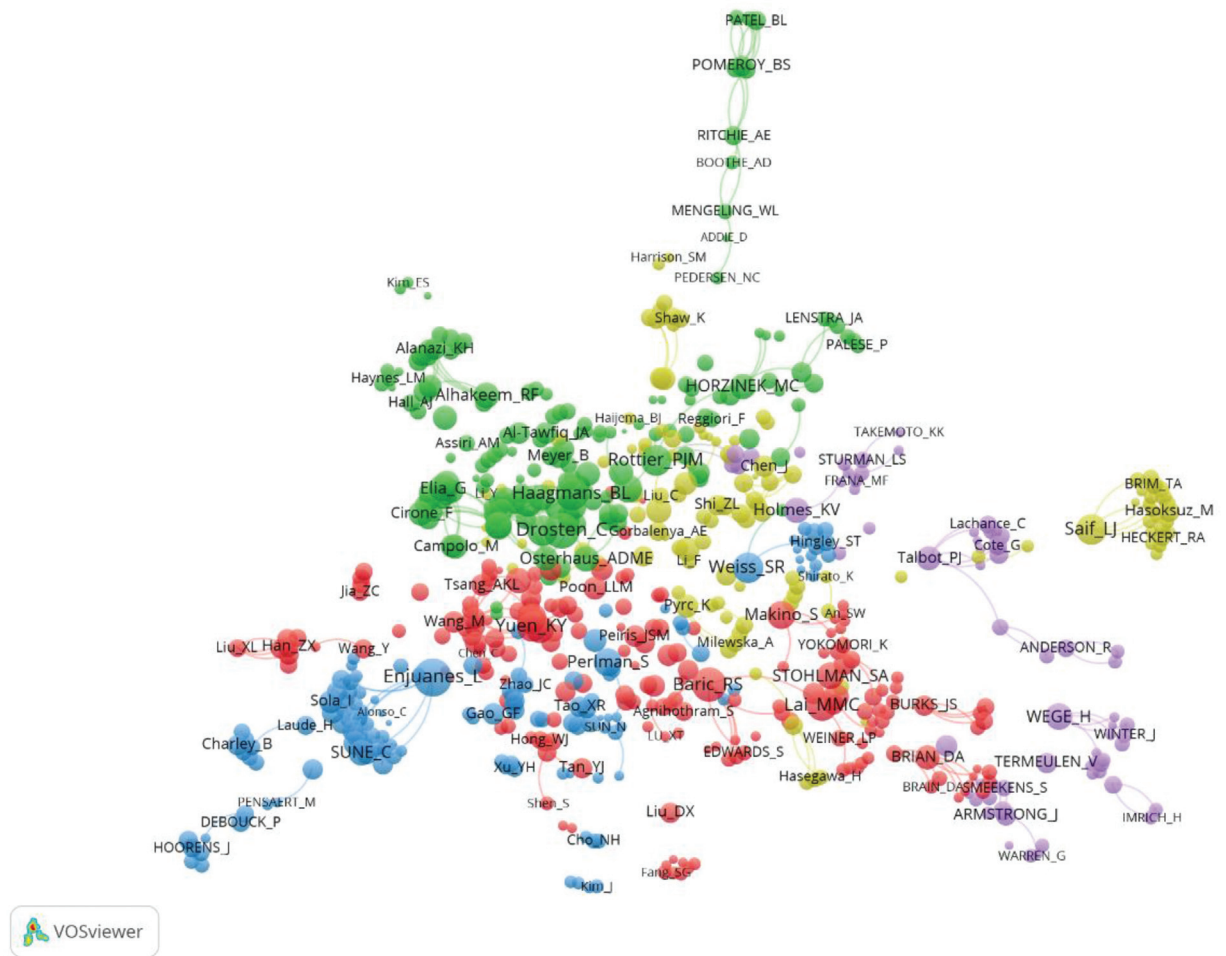
Figure 5. Knowledge domain map for the co-authorship network in COVs studies.

contains 36 keywords. It can be seen that the keywords in cluster 2 mainly focus on the research on the outbreak source and transmission of coronavirus. Reusken et al[29] investigated possible animal reservoirs of MERS-CoV by assessing specific serum antibodies in livestock. Alraddadi et al[30] conducted a retrospective cohort study among healthcare personnel in hospital units that treated MERS-CoV patients. Participants were interviewed about exposures to MERS-CoV patients, use of personal protective equipment, and signs and symptoms of illness after exposure. Infection status was determined by the presence of antibodies against MERS-CoV. JL Harcourt et al[31] attempted to determine the prevalence of MERS-CoV seropositivity in dromedary camels in Israel.

Cluster 3 (Red). The red cluster mainly focuses on the research on the infection pathway of coronavirus in human body represented by keywords such as "messenger rna," "infectious bronchitis virus," "gene," "sequence analysis," and "rna polymerase," which contains 87 keywords. PJM Rottier et al[32] established that the feline infectious peritonitis virus spike protein is the determinant for efficient macrophage infection by showing that infection by both viruses was equally blocked by antibodies directed against the feline aminopeptidase N receptor. Ma et al[33] revealed a new strategy used by transmissible gastroenteritis virus to escape the type I interferons

response by engaging the inositol-requiring enzyme 1α-miR-30a-5p/SOCS1/3 axis, thus improving our understanding of how transmissible gastroenteritis virus escapes host innate immune defenses.

Cluster 4 (Green). Among the 78 keywords, "strain," "pathogenesis," "antigen," "acolony," and "cytokine" are the main keywords. In the green cluster, the main topic is the research on the pathogenesis of coronavirus. Sevajol et al[34] summarized current understanding of SARS-CoV enzymes involved in RNA biochemistry, such as the in vitro characterization of a highly active and processive RNA polymerase complex. Terada et al[35] succeeded in obtaining infectious cDNA clones derived from type I feline coronavirus (F-CoV) that retained its virulence, which are powerful tools for increasing understanding of the viral life cycle and pathogenesis of feline infectious peritonitis-inducing type I F-CoV.

Cluster 5 (Light Blue). The light blue cluster mainly reflects the topic of the research on the diagnosis and symptoms of coronavirus infection, which contains 31keywords such as "infection," "assay," "diagnosis," and so on. Bermingham et al[36] described the clinical and virological features of a novel coronavirus infection causing severe respiratory illness in a patient transferred to London, United Kingdom, from the Gulf region of the Middle East. Ajlan et al[37] described the chest CT
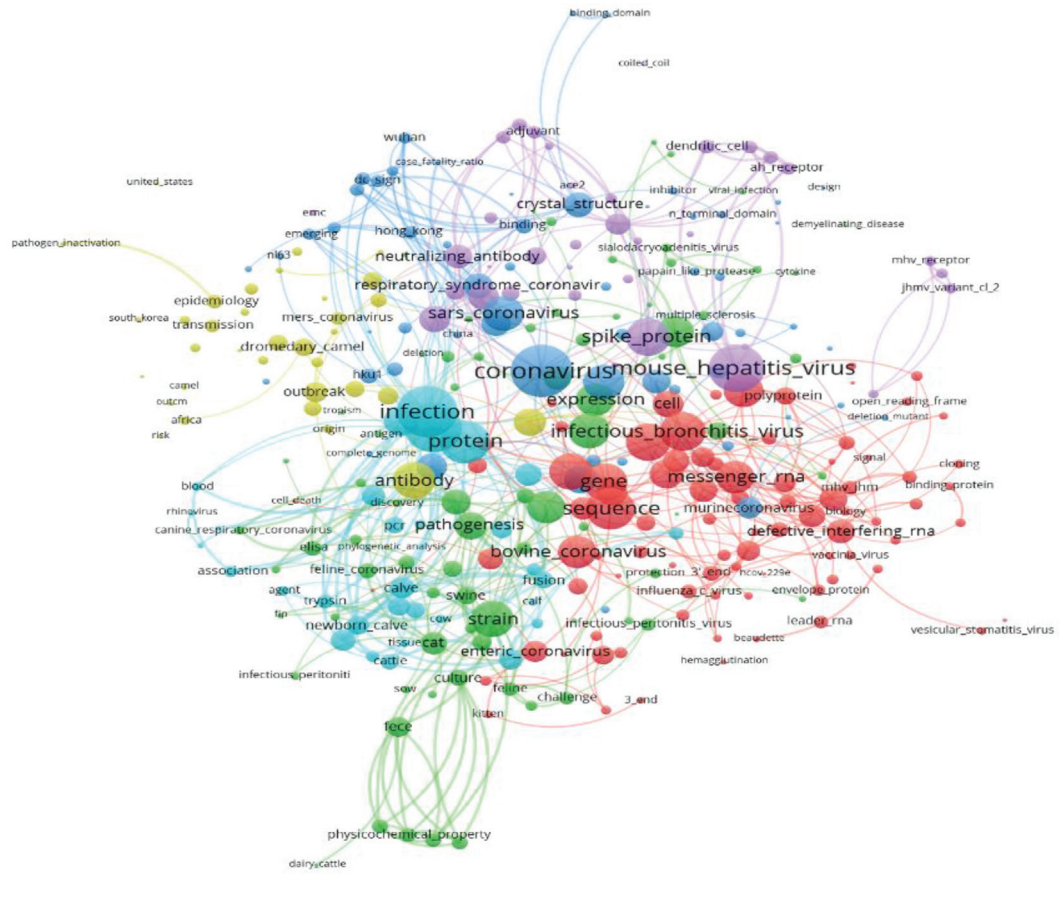
Figure 6. Knowledge domain map for a keyword co-occurrence network in COVs studies.

findings in 7 patients with Middle East respiratory syndrome coronavirus (MERS-CoV) infection. Trivedi et al[38] described the development and evaluation of a multiplexed magnetic microsphere immunoassay to simultaneously detect immuno-globulin G antibodies specific for recombinant nucleocapsid proteins from h-CoVs 229E, NL63, OC43, HKU1, SARS-CoV, and MERS-CoV.

Cluster 6 (Purple). The purple cluster contains 36 keywords, including "spike protein," "neutralizing antibody," "adjuvant," and so on. The main topic of purple cluster is the research on the treatment of coronavirus. Ying et al[39] reported the identification of human monoclonal antibodies from a large nonimmune antibody library that target MERS-CoV. Three human monoclo-nal antibodies and especially m336, neutralized the virus with exceptional potency. It therefore may have great potential as a candidate therapeutic and as a reagent to facilitate the develop-ment of vaccines against MERS-CoV. Menachery et al[40] explored whether manipulation of CoV NSP16, a conserved 2'O methyltransferase, could provide a broad attenuation platform against future emergent strains with an ongoing threat posed by circulating zoonotic strains. Coupled with increased safety and reduced pathogenesis, the study highlights the potential for 2'O methyltransferase attenuation as a major component of future live attenuated coronavirus vaccines.

3.7.3. Research frontier identification. The time zone view can show the research progress across time for each hotspot and can

show the evolution of research themes. By detecting burst keywords, a time zone view is generated, as shown in Figure 7. The darker the node color, the stronger its burst, and a link indicates that there is a connection between 2 nodes.

At the same time, the keyword strategy matrix map can also directly reflect the research hotspots and trends. We used the frequency of keywords as the horizontal axis, the degree of center as the vertical axis, and the average value of the frequency of keywords and the degree of center as the basis for division. We selected (256.9302, 0.0465) as the intersection and made the coordinates of the keyword strategy map System, as shown in Figure 8. Combining time zone view and keyword strategy matrix map can more objectively analyze the hotspots, frontiers, trends and challenges of COVs research.

In Figure 7, the whole time zone for COVs researches is shown. From this we can see that the research on coronavirus have a certain periodicity, such as SARS-CoV, mouse hepatitis virus, or MERS-CoV and other viruses. Every large virus outbreak will trigger a series of research booms. Judging from the current specific research situation, the coronavirus researches can be divided into 4 parts in combination with the keyword strategy matrix diagram.

Firstly, keywords such as "sars coronavirus," "virus," "infection," "identification," "protein," and "sequence" are located in the first quadrant and belong to the "mainstream issues" which have high frequency and high centrality character-istics. Research topics represented by these keywords are not
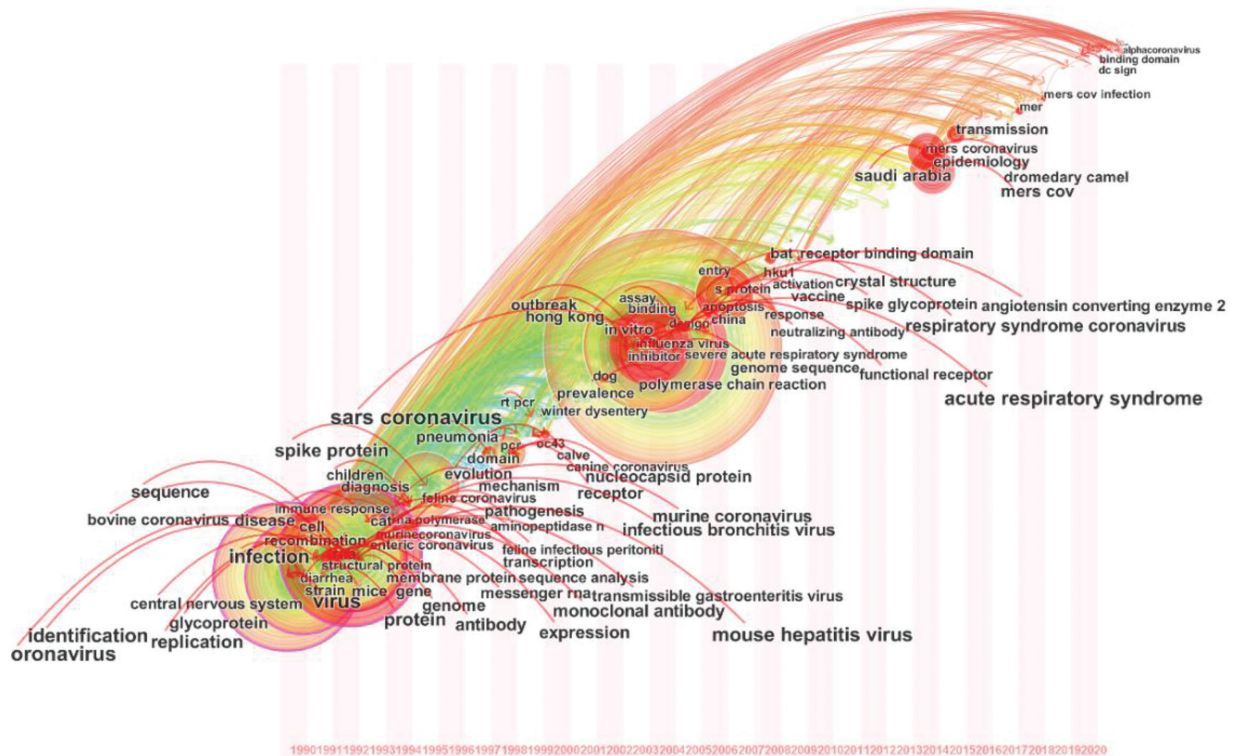
Figure 7. Keyword time zone view of COVs studies.

only mature in their own right, but also closely related to other research topics, which are the focus of current research. Such as Marra et al,[41] sequenced the 29,751-base genome of the severe acute respiratory syndrome (SARS)–associated coronavirus known as the Tor2 isolate, and discovered that the genome sequence could aid in the diagnosis of SARS virus infection in humans and potential animal hosts, in the development of antivirals, and in the identification of putative epitopes for vaccine development.

Secondly, keywords such as "pathogensis," "genome," and "receptor" are located in the second quadrant and belong to "high latent hotspots," which have the characteristics of low frequency and high centrality. The research represented by these keywords is still in the early stages of development, but it is closely related to other topics and has the potential to become a new hot spot. Sampath et al[42] described a new approach for infectious disease surveillance that facilitates rapid identification of known and emerging pathogens, which applicabled to the surveillance of bacterial, viral, fungal, or protozoal pathogens.

Thirdly, keywords such as "messenger rna," "evolution," "transmissible," "outbreak," and "pneumonia" are located in the third quadrant, which belongs to the "island area" of the current research. The research represented by them has the characteristics of low frequency and low centrality, which belongs to the subject that has not yet developed well, and has poor connection with other subject research, but it is also an area that needs to be focused on in the future, which requires more research we will continue to strive for breakthroughs. Chen et al[43] investigated the epidemiology and genetic diversity of transmissible gastroenteritis virus in the United States by testing clinical cases for TGEV by real time RT-PCR.

Fourthly, the keywords represented by "respiratory syndrome coronavirus," "acute respiratory syndrome," "murine coronavirus," and "expression" are located in the fourth quadrant and belong to the "marginal zone," which has the characteristics of high frequency and low centrality, indicating that although these themes are mature, they have poor connection with other themes. But it may also be combined with other hot topics or potential topics to continue to be focused on. Lau et al[44] described the identification and molecular characterization of a SARS-CoV-related virus from Chinese horseshoe bats in Hong Kong which provided evidence for interspecies transmission in the genesis of the SARS epidemic.

4. Discussion

In this paper, a bibliometric analysis was performed on COVs literature from the WOS core database and knowledge domain maps were obtained via information visualization technology. To date, COVs research shows the following characteristics:

1. Coronavirus research can be roughly divided into 4 stages: preliminary development stage (before 2000), rapid growth stage (2000–2005), slow decline stage (2006–2011) and sustained growth stage (since 2012). There are as many as 79 countries in the world that are involved in research on COVs. The United States clearly dominates the number of literature in this field, followed by China and Germany. And, nearly one third of the top 20 institutes contributing to this field are located in the USA, suggesting that American scientists have taken the leading position in this field. A relatively complete system of cooperation has been formed between organiza-
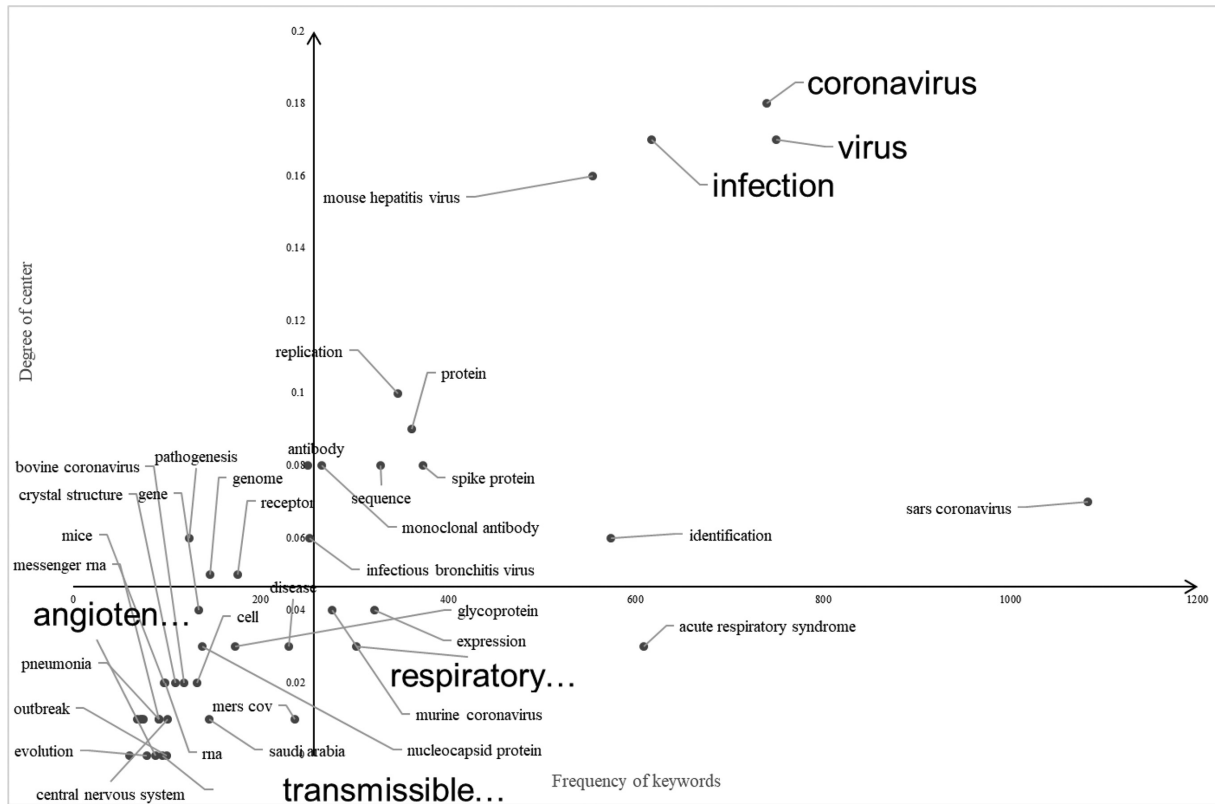
Figure 8. Keywords strategy map of COVs studies.

tions and authors. Relevant organizations and authors from the USA, China, Spain and Holland are the main driving force of the research.

2. The study of COVs involves many subject categories, mainly includes virology, veterinary sciences, biology and immunology. Among the top ten articles cited most frequently, there are 8 articles about SARS-CoV, which mainly focus on the medical research such as virus genome. In terms of main source journals, Journal of Virology ranks first, followed by Virology and Journal of General Virology. Journal of Virology is the journal most frequently cited with other journals. Journal of Virology and Journal of General Virology were co-cited most frequently, followed by Veterinary Science and Avian Diseases and Nature and Virology.

3. At present, the key words in the field of coronavirus research are mainly divided into 6 major hot clusters, namely, the introduction and structure analysis of coronavirus, the research on the outbreak source and transmission of coronavirus, the research on the infection pathway of coronavirus in human body, the research on the pathogenesis of coronavirus, the research on the diagnosis and symptoms of coronavirus infection, and the research on the treatment of coronavirus. The research represented by such keywords as "sars coronavirus," "virus," "infection," "identification," "protein," "sequence," "pathogensis," "genome," and "receptor" are relatively mature, which belongs to the current research hotspot or high latent hotspot. The subject research represented by "messenger rna," "evolution," "transmissible," "outbreak," "pneumonia," "respiratory syndrome coronavirus," "acute respiratory syndrome," "murine coronavirus," and "expression" are relatively

insufficient, and belongs to the field that needs to be enriched in future research.

5. Conclusion

Coronavirus, which occurs all over the world, often causes huge casualties and economic losses, and poses a serious threat to the safe and stable operation of the social and economic system. Objective literature review and analysis can help scholars in related fields to deepen their overall understanding.

However, there are several key issues that should be further explored in future research. For instance, the research data is based on the WoS core collection database, which may be affected by the coverage of WoS. In addition, the retrieval style and keyword structure were designed through many trials, however, there may still be some related articles that have not been captured. Future research can address these limitations by using multiple databases or focus on the development of methods to collect data more widely.

References

[1] Gonzalez JM, Gomez-Puertas P, Cavanagh D, et al. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Arch Virol 2003;148:2207–35.
[2] Dey SK, Rahman MM, Siddiqi UR, et al. Analyzing the epidemiological outbreak of COVID-19: a visual exploratory data analysis approach. J Med Virol 2020;92:632–8.
[3] Cw A, Yang LB, Yy B, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods-science direct. Acta Pharmaceutica Sinica B 2020;10:766–88.

[4] Thomas-Rüddel D, Winning J, Dickmann P, et al. Coronavirus disease 2019 (COVID-19): update für Ansthesisten und Intensivmediziner Mrz 2020. Der Anaesthesist 2020;69:225–35.

[5] Yang C, Ma QY, Zheng YH, et al. Transmission routes of 2019-novel coronavirus (2019-nCoV). Chin J Prevent Med 2020;54:374–7.

[6] Putri WCWS, Muscatello DJ, Stockwell MS, et al. Economic burden of seasonal influenza in the United States. Vaccine 2018;36:3960–6.

[7] Al-Tawfiq JA, Gautret P. Asymptomatic Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: extent and implications for infection control: a systematic review. Travel Med Infectious Dis 2019;27:27–31.

[8] Masters PS. Coronavirus genomic RNA packaging. Virology 2019;537:198–207.

[9] Fung TS, Huang M, Liu DX. Coronavirus-induced ER stress response and its involvement in regulation of coronavirus–host interactions. Virus Res 2014;194:110–23.

[10] Al-Tawfiq JA, Auwaerter PG. Healthcare-associated infections: the hallmark of Middle East respiratory syndrome coronavirus with review of the literature. J Hospital Infection 2019;101:20–9.

[11] Shiffrin RM, Börner K. Mapping knowledge domains. National Acad Sciences 2004;101:5183–5.

[12] Zhao X. A scientometric review of global BIM research: analysis and visualization. Automation Construction 2017;80:37–47.

[13] Liu H, Hong R, Xiang C, et al. Visualization and analysis of mapping knowledge domains for spontaneous combustion studies. Fuel 2020;262:116598.

[14] Chen Chaomei. Expert review. Science mapping: a systematic review of the literature. J Data Inform Sci 2017;2:1–40.

[15] Van Eck NJ, Waltman L. VOS: a New Method for Visualizing Similarities between Objects[M]//Advances in Data Analysis. Berlin, Heidelberg: Springer; 2007. 299–306.

[16] Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84:523–38.

[17] Chen Chaomei. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inform Sci Technol 2006;57:359–77.

[18] Hong R, Xiang C, Liu H, et al. Visualizing the knowledge structure and research evolution of infrared detection technology studies. Information 2019;10:227.

[19] Small H. Co-citation in the scientific literature: a new measure of the relationship between two documents. J Am Society Inform Sci 1973;24:265–9.

[20] Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003;300:1394–9.

[21] Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Eng J Med 2003;348:1953–66.

[22] Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Eng J Med 2003;348:1967–76.

[23] van der Hoek L, Pyrc K, Jebbink MF, et al. Identification of a new human coronavirus. Nat Med 2004;10:368–73.

[24] Zaki AM, Van Boheemen S, Bestebroer TM, et al. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Eng J Med 2012;367:1814–20.

[25] Yang Y, Liu C, Du L, Jiang S, et al. Two mutations were critical for bat-to-human transmission of Middle East respiratory syndrome coronavirus. J Virol 2015;89:9119–23.

[26] Aleanizy FS, Mohmed N, Alqahtani FY, et al. Outbreak of Middle East respiratory syndrome coronavirus in Saudi Arabia: a retrospective study. BMC Infectious Dis 2017;17:1–7.

[27] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506.

[28] Chan JFW, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020;395:514–23.

[29] Reusken CB, Haagmans BL, Marcel A Müller, et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infectious Dis 2013;13:859–66.

[30] Alraddadi BM, Al-Salmi HS, Jacobs-Slifka K, et al. Risk factors for Middle East respiratory syndrome coronavirus infection among healthcare personnel. Emerg Infectious Dis 2016;22:1915.

[31] Harcourt JL, Rudoler N, Tamin A, et al. The prevalence of Middle East respiratory syndrome coronavirus (MERS-CoV) antibodies in dromedary camels in Israel. Zoonoses Public Health 2018;65:749–54.

[32] Rottier PJM, Nakamura K, Schellen P, et al. Acquisition of macrophage tropism during the pathogenesis of feline infectious peritonitis is determined by mutations in the feline coronavirus spike protein. J Virol 2005;79:14122–30.

[33] Ma Y, Wang C, Xue M, et al. The coronavirus transmissible gastroenteritis virus evades the type I interferon response through IRE1 alpha-mediated manipulation of the MicroRNA miR-30a-5p/SOCS1/3 Axis. J Virol 2018;92:e00728–18.

[34] Sevajol M, Subissi L, Decroly E, et al. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. Virus Res 2014;194:90–9.

[35] Terada Y, Kuroda Y, Morikawa S, et al. Establishment of a virulent full-length cDNA clone for type I feline coronavirus strain C3663. J Virol 2019;93:e01208–19.

[36] Bermingham A, Chand M, Brown C, et al. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. Euro Surveill 2012;17:20290.

[37] Ajlan AM, Ahyad RA, Jamjoom LG, et al. Middle East respiratory syndrome coronavirus (MERS-CoV) infection: chest CT findings. Am J Roentgenol 2014;203:782–7.

[38] Trivedi SU, Miao C, Sanchez JE, et al. Development and evaluation of a multiplexed immunoassay for simultaneous detection of serum IgG antibodies to six human coronaviruses. Sci Rep 2019;9:1–8.

[39] Ying T, Du L, Ju TW, et al. Exceptionally potent neutralization of Middle East respiratory syndrome coronavirus by human monoclonal antibodies. J Virol 2014;88:7796–805.

[40] Menachery VD, Gralinski LE, Mitchell HD, et al. Combination attenuation offers strategy for live attenuated coronavirus vaccines. J Virol 2018;92:e00710–18.

[41] Marra MA, Jones SJM, Astell CR, et al. The genome sequence of the SARS-associated coronavirus. Science 2003;300:1399–404.

[42] Sampath R, Hofstadler SA, Blyn LB, et al. Rapid identification of emerging pathogens: coronavirus. Emerg Infectious Dis 2005;11:373.

[43] Chen F, Knutson TP, Rossow S, et al. Decline of transmissible gastroenteritis virus and its complex evolutionary relationship with porcine respiratory coronavirus in the United States. Sci Rep 2019;9:1–11.

[44] Lau SKP, Woo PCY, Li KSM, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proceedings National Academy Sci 2005;102:14040–5.