ORIGINAL RESEARCH ARTICLE

Visualizing a Field of Research for the Coronavirus Replication in Humans with Knowledge Mapping: Evidence from *Web of Science*

Hanyuan Liang[1] · Weikun Zhang[2] · Zhe Chen[1] · Xuexue Chen[3]

Abstract

Background The outbreak of COVID-19 sweeping the globe in 2020 has caused widespread fear and threatened global health security. Compared to SARS and MERS, COVID-19 also causes severe respiratory diseases and even fatal diseases but have many differences, such as the unidentified gene sequence and replication mechanism. From SARS to MERS, and then to COVID-19, coronaviruses have significant variations in host adaptation, virus evolution, infectivity, spread, and pathogenicity due to its unique replication mechanism.

Methods A field of research for the coronavirus replication in humans was visualized with a database covering 9177 kinds of literature in Web of Science from 2002 through October 2021 to provide cognitive direction for the epidemic situation of virus infection. Knowledge Mapping by CiteSpace and Bibliometrix Package in R Software was drawn to depict the underlying features of viral replication and changing trends of studies, with these analyses including co-citation, density visualization, keyword clustering, and time zone.

Results The keyword frequencies of "replication," "infection," and "spike protein" repeatedly appeared in published papers. Coronavirus can promote or inhibit apoptosis, depending on the balance between viral protein and apoptotic factors. When the living environment of cells is irreversibly damaged by the virus, cells have to start the apoptosis mechanism to prevent the replication, transmission, and spread of the virus. The replication, assembly and transmission of coronavirus can inhibit cells from entering the apoptosis prematurely with the fusion of spike protein and cell receptor in human.

Conclusion Our results indicated that "viral infection," spike protein," and "mutation" might be future research hotspots on coronavirus replication in humans. The attention should be paid to the mutations of S protein and these mutants carrying mutations.

Hanyuan Liang and Xuexue Chen contributed equally to this paper.

✉ Weikun Zhang
  weikunz@sids.org.cn

1   School of Economics and Finance, Zhanjiang University of Science and Technology, Zhanjiang 524088, Guangdong, People's Republic of China

2   School of Social and Public Administration, Lingnan Normal Univeristy, Zhanjiang 524088, Guangdong, People's Republic of China

3   Central People's Hospital of Zhanjiang, Zhanjiang 524037, Guangdong, People's Republic of China
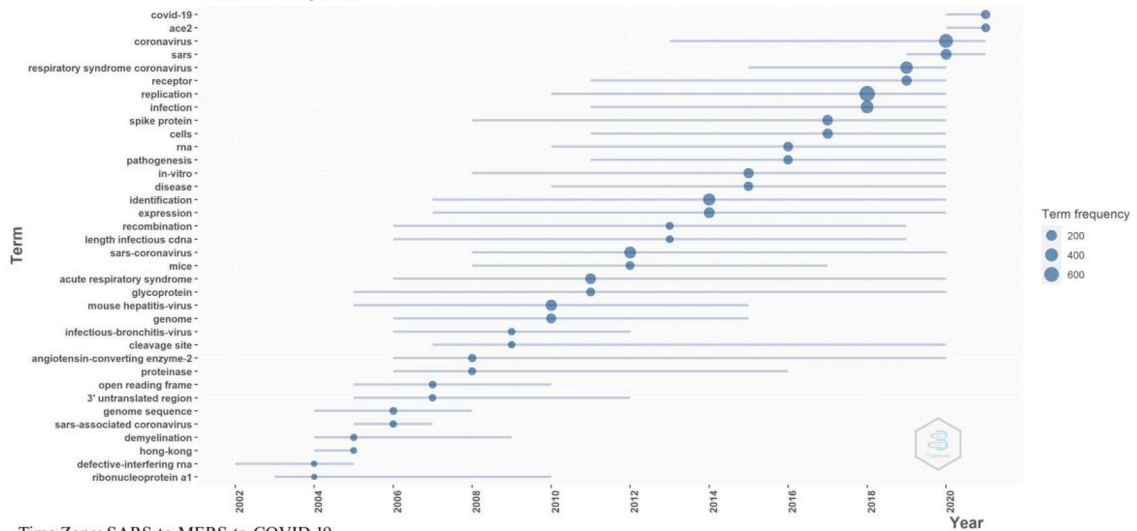
Graphical abstract
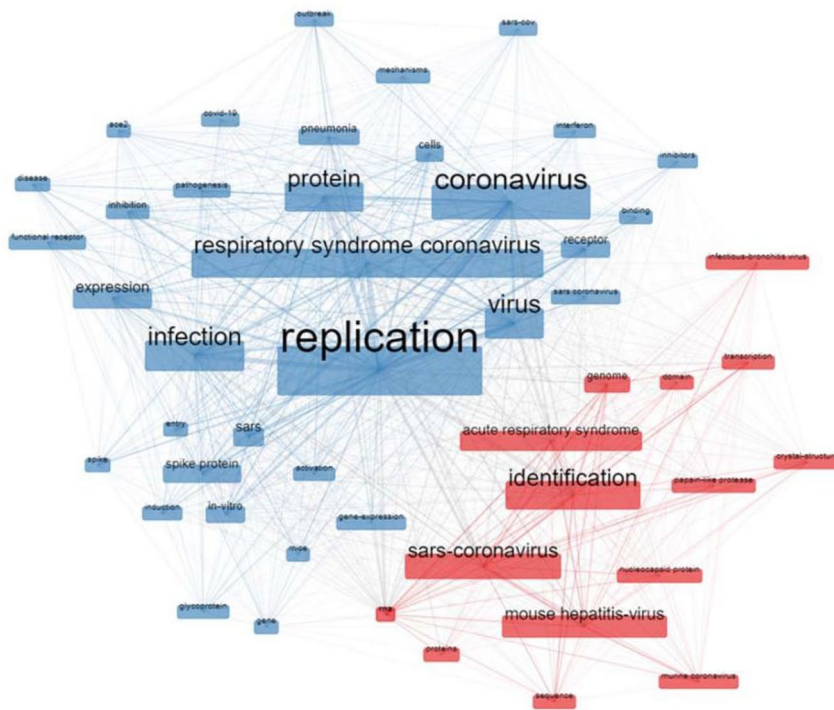
Core Words: "replication," "infection" and "protein"

High-frequency Words: "identification" and "acute respiratory syndrome"

Conclusion: The keyword frequencies of "replication," "infection," and "spike protein" repeatedly appeared in published papers.

Time Zone: SARS-to-MERS-to-COVID-19

Conclusion: ACE2, deemed to be a functional receptor for SARS-CoV, can aggravate the severity of the disease in the blend of cellular protein after the COVID-19 outbreak in China.

1 Background

Coronavirus is the pleomorphic enveloped viruses causing the diseases of the respiratory tract, gastrointestinal tract, and central nervous system, which contains the largest single-stranded positive-sense RNA genome enveloped virus ranging in length from 27 to 31 kb in nature [1, 2]. The Coronavirus family members consist of four genera based on their genetic characteristics, including Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus. Of all RNA viruses, coronavirus has the most abundant RNA genome ranging from 26 to 32 kb. Apart from humans, coronaviruses also can infect many different animal species, including swine, cattle, horses, camels, cats, dogs, rodents, birds, bats, rabbits, ferrets, martens, snakes, and other wildlife. Coronaviruses causing severe respiratory diseases in humans, such as SARS-CoV in 2003 and MERS-CoV in 2012, are mainly zoonotic pathogens that originate from severe respiratory diseases and sometimes related to other fatal diseases in humans [3].

The novel coronavirus COVID-19 that broke out in China in January 2020 is a newly discovered coronavirus in humans. This COVID-19, a novel coronavirus similar to certain Betacoronaviruses identified in bats, has different coronavirus-specific nucleic acid sequences compared with the known coronavirus in humans and is temporarily named by the World Health Organization, which enveloped RNA virus belong to an evolutionary branch of another cluster of Subgenus Sarbecovirus of Orthocoronavirinae subfamily. The novel coronaviruses in China has caused severe respiratory syndromes similar to SARS-CoV and MERS-CoV, and become one of the coronaviruses in humans with high mortality and the seventh member of the coronavirus family that infects humans [4].

COVID-19 as a novel RNA virus, namely SARS-CoV-2, mutates through the dependence on RNA polymerase and results in high RNA recombination rate, thus adapting to new hosts. RNA-dependent RNA polymerase is one of the important reasons for the high mutation rate of RNA virus. A variety of active substances in host cells may promote the mutation of virus gene. Out of its unique viral replication mechanism, coronavirus has a high frequency of recombination, which can infect a variety of mammals and birds through the ability of the spike entry protein to bind to cell surface receptors. Studies over the past 17 years have shown that bats are natural hosts for coronaviruses. Genetic diversity among zoonotic viruses in bats increases the likelihood of mutations crossing species barriers and causes human disease outbreaks [5]. Bat-CoV cannot replicate within the potential zoonotic virus range and requires intermediate hosts such as livestock, which

may be in close contact with humans to better bind to angiotensin-converting enzyme 2 (ACE2) in human cells [6]. Coronaviruses have the potential to spread through the respiratory or fecal pathways, and the ability to cause significant epidemics after intermediate host transformation. However, coronavirus detected in humans accounts for a large proportion with bats, maybe because of the transmission through bats to humans across species such as mammals [7]. In this vein, the pathological transmission mechanism of coronavirus replication in humans remains complex and difficult to recognize a comprehensive and clear understanding for the research of coronavirus especially for SARS-CoV-2.

Coronaviruses have a high frequency of recombination and mutation rate that can adapt to new hosts and niche fleetly because of the unique mechanism of virus replication. For instance, in the early stage of viral infection, SARS-CoV-2 can attack targeted organs with ACE2 through the fusion of human cell and virus membrane, triggering diseases such as lung, heart, and kidney; in the late stage of infection, the endothelial cells of alveolar capillaries can be infused with inflammatory reaction aggravated and neutrophil infiltration induced due to the increase of the replication rate of virus. Instead, our perceptions of viral infection in humans remain still incomplete due to the elusive nature of the illness. The exploration of the causes of respiratory diseases with similar symptoms is a particularly difficult task because many acute and chronic diseases of unknown etiology start from unrecognized viruses. The coronavirus replication in humans (CRH) is so quite diverse that many researchers cannot grasp critical points in an instant with the lack of system integration on the gene-expressed replication. Accordingly, the kinds of literature on such research domain published in *Science Citation Index Expanded* (SCIE) and *Social Science Citation Index* (SSCI) in *Web of Science* (WOS) from 2002 to October 2021 were employed to be visualized with the code of statistically Knowledge Mapping to depict the new development of underlying specialties on CRH.

The contributions of this article are threefold. First, a more comprehensive understanding of CRH with literature co-citation analysis, relatively poorly explored, is advanced through the visualization of the literature database on CRH. Second, the literature visualization of reviews for CRH, different from narrative reviews, is performed with main keywords frequency analysis and evolution time zone analysis. Finally, in spite of the use of the quantitative method widely, some quantitative literature analyses only rest on descriptions of statistical data rather than an in-depth visualization analysis in the specific field. Our study attempts to provide in-depth reviews for the findings on CRH through literature review visualization instead of the description of statistical data.

The remaining part of this article proceeds as follows. The next section introduces the literature database and method. Section 3 describes the visualization result. The discussion on CRH is presented in Sect. 4 and the conclusion in Sect. 5.

2 Main Text

2.1 Definition of COVID-19

Coronavirus mainly infects older children and cause common cold and pharyngolaryngitis, among which some strains can bring about diarrhea. A novel coronavirus called COVID-19 was isolated from the lower respiratory tract of patients with pneumonia of unknown cause in Wuhan China in January 2020. Like SARS-CoV and MERS-CoV, SARS-CoV-2, namely COVID-19, belongs to the same family of β coronavirus. In terms of gene sequence homology, SARS-CoV-2 has a resemblance but not the terrible characteristics with SARS-CoV. SARS-CoV-2 has already been lethal without the determined fatality rate, which high risks may be lower than that of SARS-CoV. Compared with SARS-CoV and MERS-CoV, SARS-CoV-2 spreads faster among humans.

The spread of the virus variants observed in the Republic of Korea in 2015 has confirmed the virus spike protein contains mutations reducing the sensitivity of antibody receptor binding but not uncovered what plays a promoting role [8]. The detection from the faecal samples for up to 30% of patients with MERS-CoV in the Middle East has revealed the existence of coronavirus replication, indicating that MERS-CoV can survive in gastrointestinal fluid with the ability to infect intestinal organs [9]. Of 62 faecal samples, some scholars has detected only 6.5% test is positive for SARS-CoV-2 [10]. Generally speaking, the spread of toxic gas may play a role in the rapid spread of SARS-CoV-2, along with respiratory and digestive system, and be a potential route for infections. The patient may be infected with novel coronavirus in the digestive system and then spread to lung infection of respiratory tract. Moreover, SARS-CoV-2 broke out during the winter-spring transition period in China, followed the peak flow of people during the Spring Festival in China, and spread rapidly around the world. In the clinical symptoms of novel coronavirus, many asymptomatic patients are diagnosed with the positive of RT-PCT Testing. Due to the lack of serious illnesses in healthy people, emerging undiscovered viruses may pose a significant risk to vulnerable people in underlying diseases such as Diabetes and Heart disease, which will affect our recognition to contain the spread of the virus. Note that many infected people are asymptomatic or have mild symptoms but can spread the virus, which brings the complexity in the chain of transmission and subsequent contract tracing. Furthermore, S protein located on the surface of SARS-CoV-2 (diameter 60–140 nm), an upgrade of SARS-CoV, is characterized by high mutation rate with strong adaptability through gene recombination.

2.2 Data

Before SARS in 2003, most known coronaviruses did not kill humans but could induce fatal diseases in livestock species [11]. SARS, emerging across Guangzhou, China in 2003, is a member of the coronavirus-derived family but has a long genetic distance from known coronaviruses [12]. To reflect the researching differences on the coronavirus replication before and after the outbreak of high-risk coronavirus in humans, we viewed the starting point of the scientific literature as the last year of the emergence of SARS in 2003.

Commonly the queries of scientific literature about the domain of coronaviruses can be constructed in the used sources, including the WOS, Scopus, Google Scholar, and PubMed. WOS, including thousands of journals, especially the most influential ones, is widely deemed to be an appropriate data source of scientific literature and contains abundant cited information [13]. To ensure the accuracy and reliability of the research content, a query-based search of "replicat* + coronavir*" or "unknown pneumonia + virus + replicat*" can be constructed in SCI (SCIE) and SSCI in WOS from 2002 through October 2021, in which the published papers have a very high degree of authority and recognition, and represent higher academic levels and eminently influential value in academia. We excluded the duplicate literature, only included the articles and reviews, and thus collected 9177 ones from the query-searched findings, which displayed such information covering "journal," "title," "abstract," "keyword," "author," "institution," "country," and "term."

The peak point of the publication for coronavirus research mainly was often the outbreak period of COVID-19 during the period from January 2020 through October 2021, as reported in Fig. 1. With the emergence of SARS-CoV in 2003, the research field on CRH has captured interests from scholars, which attention heat decreased with the control of the SARS epidemic after 2006. Due to the MERS-CoV in 2012, the coronavirus-based studies have become hotspots from 2013 to 2019. A new unknown coronavirus outbreak in Wuhan China in January 2020 makes CRH focus attractive and seductive to scholars all over the world widely as detected in the histogram of 2020 and 2021 in Fig. 1.

2.3 Methods

Knowledge mapping (KM) is a universal process for domain analysis and visualization with so many general tools such
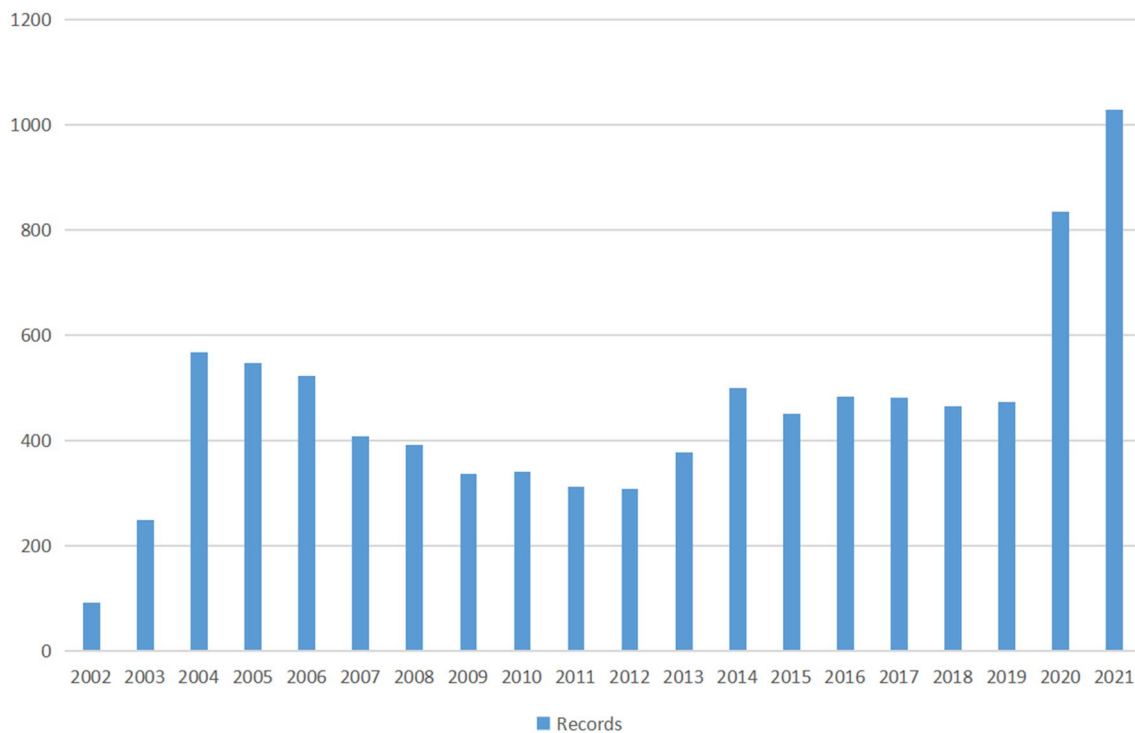
Fig. 1 Coronavirus research posts from 2002 through October 2021

as CiteSpace, VOSviewer, and HistCite. However, few systems are easy to access and to meet the needs of systematic evaluation of fast-changing and complex areas, especially the explanations of emerging trends and transition patterns from non-domain experts. CiteSpace as a scientific literature data mining and visual analysis tool provides co-citation analysis of literature, knowledge clustering distribution of mining citation space, and cooperative co-occurrence analysis with information consisting of author, institutions and countries or regions. It is well known that CiteSpace is an effective literature visualization tool to explore the research evolution path, hot topics, and knowledge boundary. Among these tools, CiteSpace, visualizing the structural patterns of scientific literature, has been used for the application of regenerative medicine [14]. CiteSpace based on the algorithm covering the co-citation analysis and pathfinder network scaling can be used to descript the knowledge mapping containing the structure of a research field, evolution path, and the exploration of scientific fronts on your studies. The co-reference network in CiteSpace contains more information than the classic ones, including various indicators and topic labels derived from the cited articles to the jointly cited clusters, conceptual tags tracking the evolution of clusters year by year, and hierarchical representations of concept terms extracted from the titles and abstracts of the cited articles. Furthermore, CiteSpace supports several types of visualization, including the network clustering, the

co-citation analysis for author and document, and the view for a timeline and a history.

To reflect network visualization, feature clustering, literature co-citation, and text statistics, we used KM with CiteSpace and Bibliometrix Package in R Software to draw the intellectual structure and thus interpret the nature of underlying specialties of the literature-based discovery by using analysis ways for citation, bibliometrics, and scientometrics. Due to the aggregations of co-cited individual items, we only showed what clusters and what information we get from the literature-based discovery.

3 Results

3.1 Country

We framed the node type as "country" and set the threshold to (200, 2, 2) to draw a KM of national research on CRH, with these results shown in Fig. 2. The countries with the most significant number of articles are the United States, China, and Germany. The total number of articles published by American scholars reached 2,665 rankings first in all countries. Among them, the U.S. research started early with abundant results, and its average annual volume of papers has been more than 100 since 2004, especially nearly 200 ones in 2006 and 2014. In contrast, Chinese in this domain

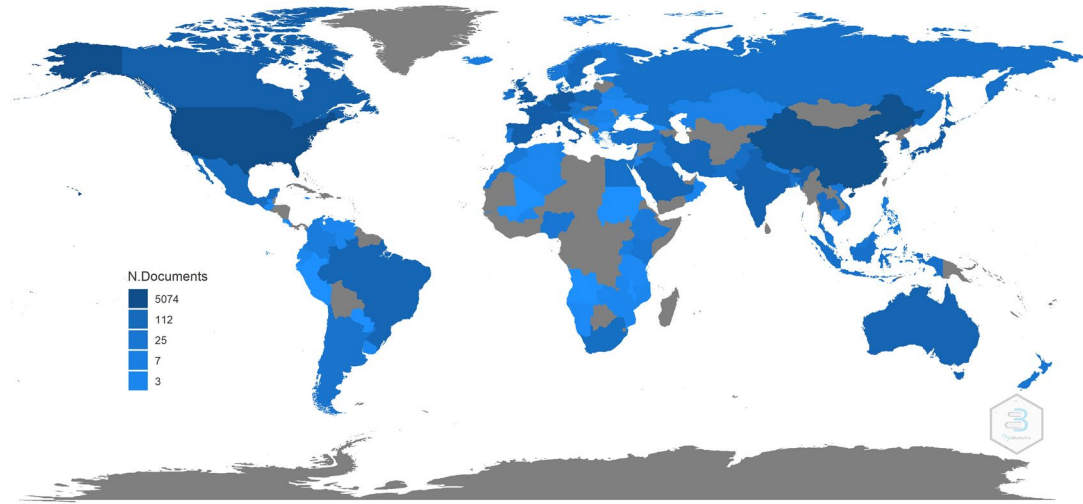## Country Scientific Production

Fig. 2 KM of countries with more than 200 published papers on CRH from 2002 to October 2021

began during the SARS outbreak with a total of 2049 published articles and had the most literature on SARS-CoV with more than 200 published articles in 2004 and 2005, but its annual number of articles published was less than 100 after 2006. Besides, Saudi Arabia's ones with few published articles on CRH, notably, have been a spurt of growth after MERS.

3.2 Journal Source

To explore the impact of literature source journals, we applied H Index meaning high citations to depict the local impacts of journals as shown in Fig. 3. The most influential source journal is "Journal of Virology"(H Index 94) which mainly focus on the research on virology. The next is
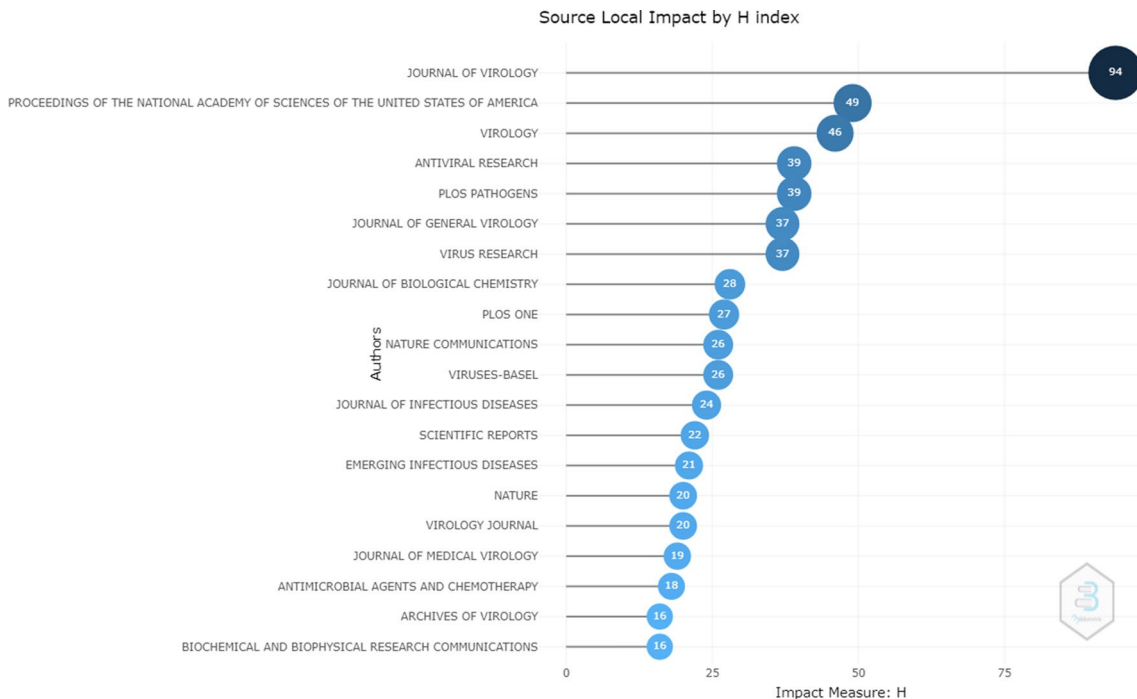

Fig. 3 The Impact of Journals with H Index

"Proceedings of National Academy of Sciences of the United States of America" and the third "Virology" (H Index 49 and 46), indicating the papers on CRH is mostly published in journals about pathology hosted by the United States. The readers can obtain the information on CRH from these journals if interested in virology. So, the local cited or relevant three sources are "Journal of Virology", "Proceedings of National Academy of Sciences of the United States of America", and " Virology", from which scholars get the references of further studies on CRH to detect the intellectual base of a key field.

3.3 Co-citation Analysis

Citation rate—the frequency of papers cited in the research—is a crucial indicator to evaluate the quality of papers. The higher the citation rate, the more authoritative the paper is in the field. Nine of the top ten cited published papers in 2003 with topics of "SARS-coronavirus" or "Severe Acute Respiratory Syndrome" (Fig. 4) were the knowledge base in the field of CRH. A possible explanation is that few works on the lethal new coronavirus in humans are studies in the early stage of an outbreak with directive significance for exploring CRH. The study of Ksiazek et al. [15] published in "*New England Journal of Medicine*" as

a widely cited document in the field on CRH opened a black box of the research on coronavirus infecting human cells and causing human diseases, which identified a novel coronavirus from severe acute respiratory syndrome [15]. On the basis of that the unique and conserved features of genome and proteome of SARS-CoV were demonstrated by later scholars such as Snijder [16]. These highly cited literature studies have confirmed that SARS is closely related to the coronavirus, but has an approximate distance with the known coronavirus because of 50–60% of the nucleotide similarity [17, 18]. Considerable effort has gone into SARS-CoV-2 with the outbreak of COVID-19. A consensus with SARS-CoV is that ACE2 is confirmed to be the receptor of SARS-CoV-2 infecting human cells [19].Suffice it to say that SARS-CoV-2 enters human cells using serine protease TMPRSS2 to activate the S protein through the receptor ACE2 [20].

Turning attention to the density visualization, the found research hotspots on CRH with statistical keyword frequency—if the frequency were more significant and more—were highly more popular in this field. We exploited "Keyword" for the density node with the threshold of 200 from 2002 through October 2021 in an attempt to generate a keyword knowledge network. According to the red area in Fig. 5, high-frequency keywords, on the whole,
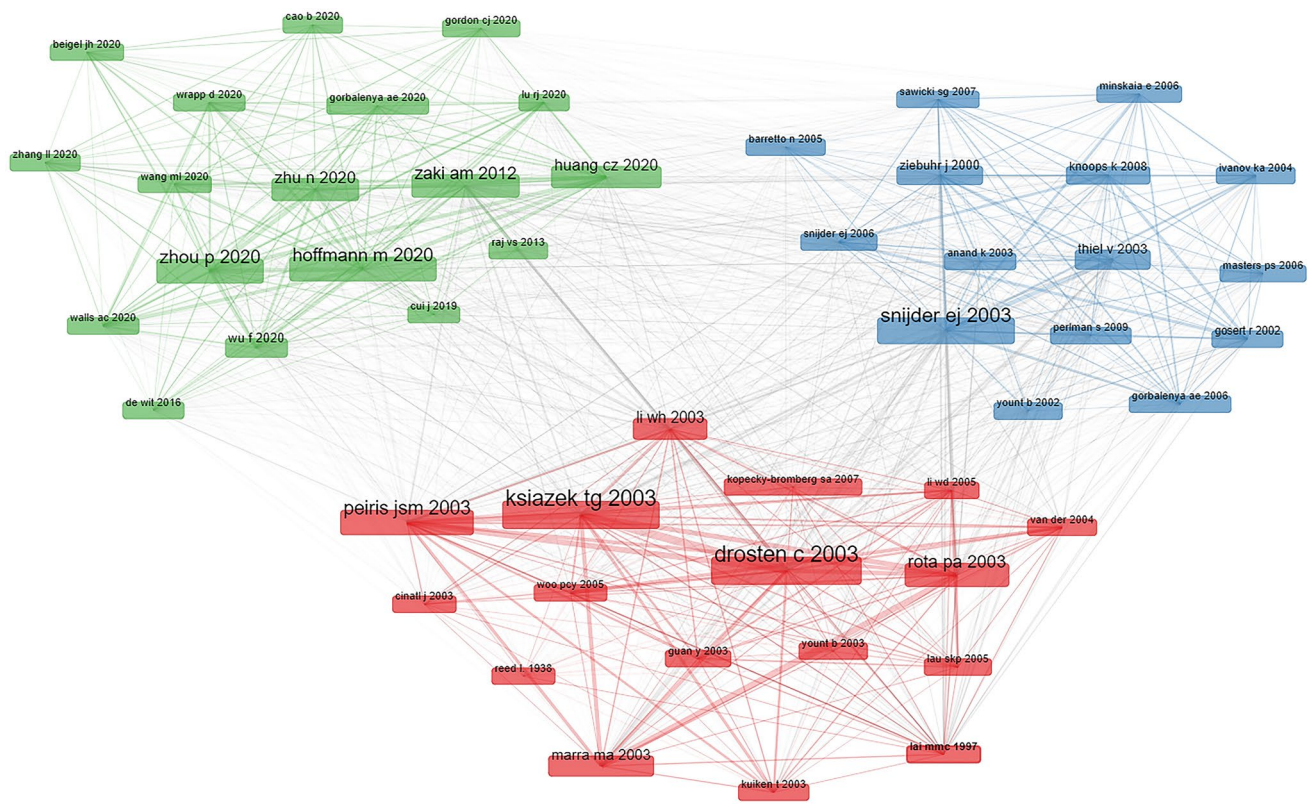

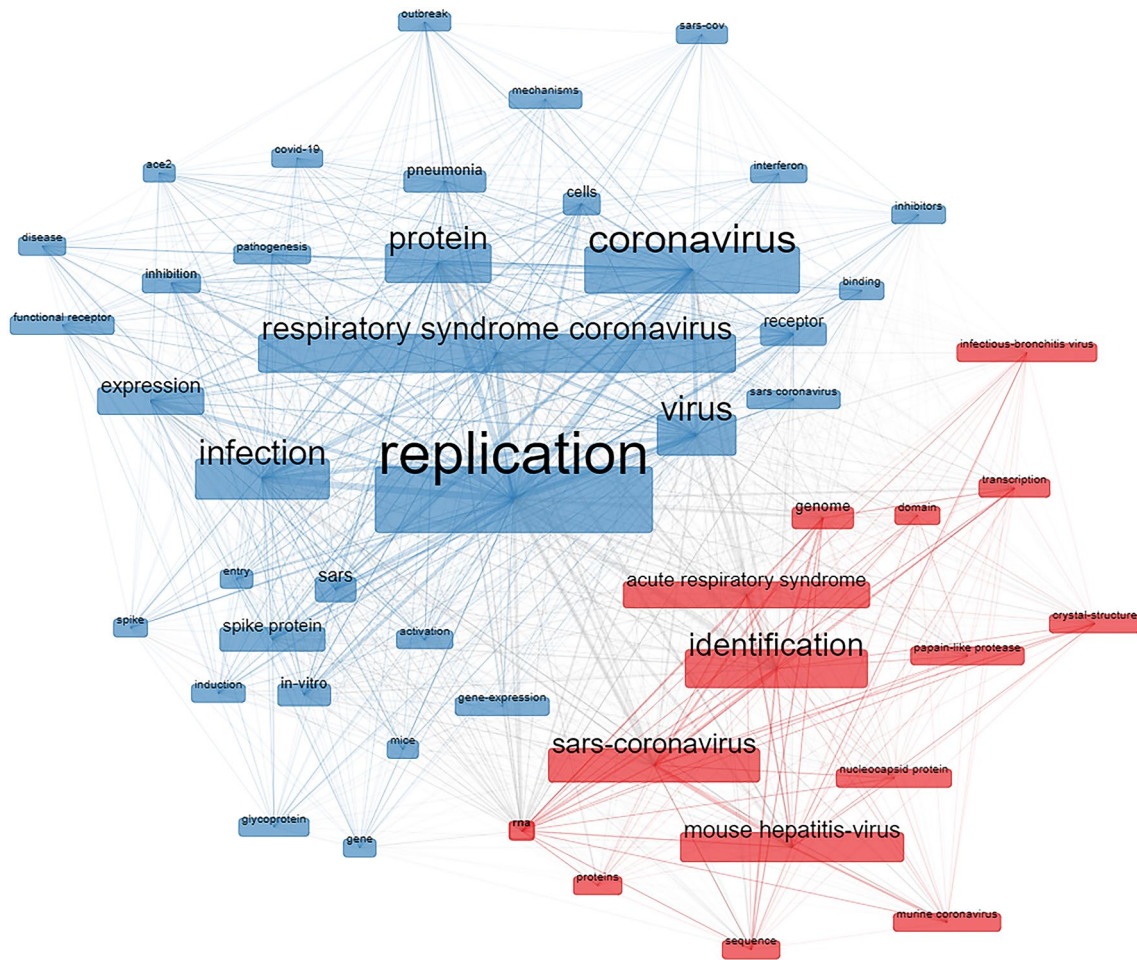**Fig. 4** KM of co-cited literature on CRH

Fig. 5 Density visualization of keywords on CRH ($n > 200$)

are composed of two categories, one for core words such as "replication," "infection" and "protein" in Coronavirus Domain, another for high-frequency words such as "SARS-CoV," "identification," and "acute respiratory syndrome".

To uncover what hotspots scholars studied intuitively, we reported the frequencies of keywords on CRH in Fig. 6. Coronavirus causes severe respiratory diseases in humans, associated with ICU entry and high mortality. Commonly the focuses on Coronavirus in humans are "replication," "infection," "identification," and "protein." Coronaviruses, a type of enveloped virus with a spinous protein similar to a crown under the microscope, consists of single-stranded RNA. When propagated in the host, viral RNA is "wrapped" into the coated core and finally "wrapped" in an envelope through ribosomal frameshifting. Specifically, spike protein on a cell space can cause infected cells to fuse with adjacent uninfected cells, and is transported into the endoplasmic reticulum–Golgi intermediate compartment through subgenomic mRNAs translated into structural and accessory proteins—thereby forming a sizeable multinucleate

syncytium infection. The current diagnostic tests for coronavirus have included RT-PCR, real-time reverse transcription, reverse transcription loop-mediated isothermal amplification, and real-time RT-LAMP in terms of "identification."

3.4 Keywords of Cluster Analysis

We used the Logarithmic likelihood algorithm to divide the keyword network into 8 cluster categories. The quality evaluation index of the clustering results is module value Q and the average contour value S. The obtained network community structure is definite with $Q > 0.3$, and the clustering results reasonable with $S > 0.5$. Q is 0.401 and S 0.5437, as shown in Fig. 7, implying that clustering works well. The clustering labels are sorted from 0 to 7, inferring that the smaller the number, the more keywords are.

To depict the feature on CRH intuitively, we reported the detail list for the size of clustering in Table 1.The most significant gathering point is #0 representing the mouse hepatitis virus (MHV) in terms of size, which is

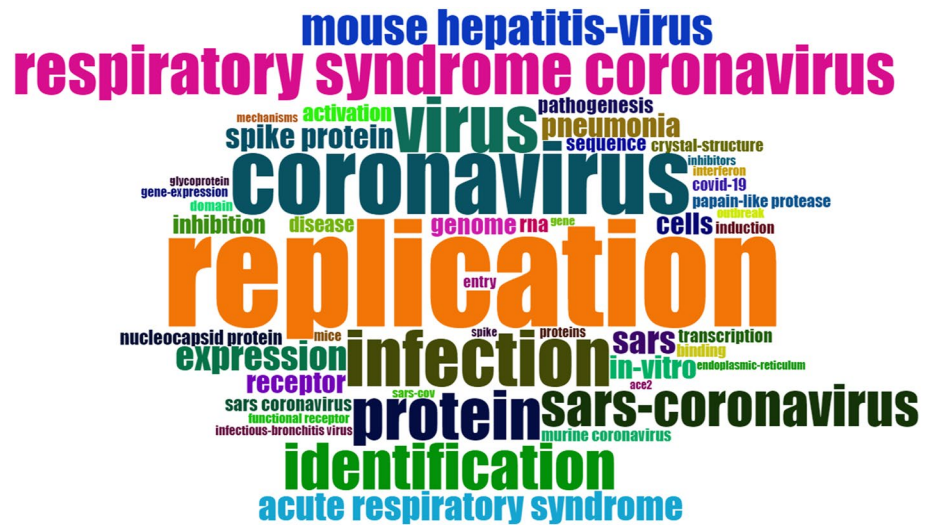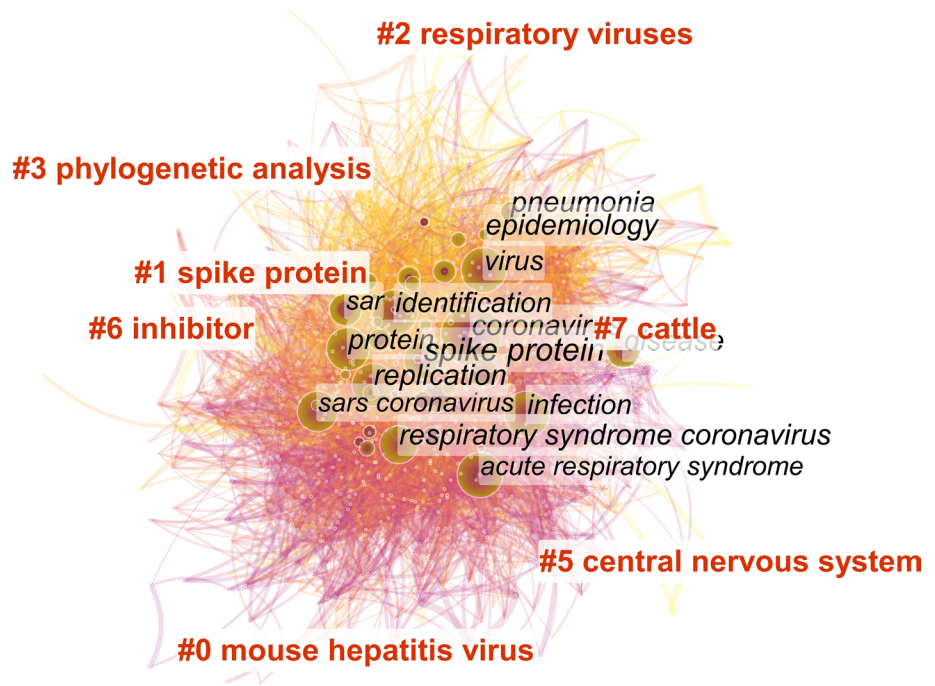**Fig. 6** The frequencies of keywords on CRH

Fig. 7 KM of thematic Atlas of CRH

characterized by "messenger RNA," "defective interfering RNA," and "viral RNA." It is an apparent fact that positive-strand RNA viral replication complexes of eukaryotes are always associated with cell intima. Specific RNA virus is modified into characteristics related to viral RNA synthesis with the membrane of specific organelles as targets. Morphology of these complexes, namely "virus factories," can provide a membrane-based structural framework for RNA synthesis. One potential benefit is that the induction of host defense mechanisms triggered by double-stranded RNA intermediates attributed to prion RNA viruses can be halted through the coordination in the viral life cycle.

The secondary topic on CRH centers on "spike protein"(#1) as shown in Table 1. Coronavirus invades spinous process proteins of host cells, of which ACE2 serves as a functional receptor [21]. There are essential viral-neutralizing epitopes that may be sufficient to elicit a protective antibody response across the receptor-binding domain in infected humans [6].

The topics of the "central nervous system"(#5), "inhibitor"(#6), and "cattle"(#7) are concerned in terms of the silhouette as given in Table 1. Regarding the "central nervous system," the diversity of coronaviruses, as we have known, is accelerated by the high frequency of

Table 1 CRH keyword clustering

ID	Size	Silhouette	Mean year	Clustering keywords
0	195	0.599	2004	Mouse hepatitis virus; murine coronavirus; messenger RNA; transcription; genome; defective interfering RNA; viral RNA; molecular biology; gene expression; infected cell; equine arteritis virus; viral replication
1	158	0.635	2005	Spike protein; s protein; envelope glycoprotein; functional receptor; angiotensin converting enzyme 2; membrane fusion potent inhibitor; proteolytic activation
2	145	0.792	2007	Respiratory viruses; children; human bocavirus; respiratory syncytial virus; young children; polymerase chain reaction; respiratory virus; respiratory infections; infant
3	141	0.694	2008	Phylogenetic analysis; recombination; infectious bronchitis virus; genetic diversity; metagenomic analysis; porcine epidemic diarrhea virus; spike protein; diversity; protection; molecular characterization; sequence
4	100	0.774	2010	Saudi Arabia; MERS-CoV; middle east respiratory syndrome coronavirus; middle east respiratory syndrome; transmission; MERS; outbreak
5	68	0.838	2004	Central nervous system; multiple sclerosis; neurotropic coronavirus; experimental autoimmune encephalomyelitis; cytokine; IFN gamma; encephalomyelitis; memory; chemokines; b cell; cd8(+) t cell; demyelination; demyelinating disease
6	64	0.827	2007	Inhibitor; main protease; design; 3c-like protease; m pro; cov main proteinase; active site; dimer interface; molecular dynamics; dimerization; crystal structure; docking; 3 cl protease; substrate specificity; coronavirus; small molecule inhibitor; dissociation
7	44	0.823	2004	Cattle; feline infectious peritonitis; canine coronavirus; feline coronavirus; 3c gene; cow; dairy calves; infectious peritonitis virus; tropism; bovine coronavirus

RNA recombination and the abnormally large RNA virus genome. That not only gives rise to the emergence of viruses with new characteristics but enables organisms to adapt to new hosts and niche, sometimes characterized by zoonotic events [22]. Previous researches on "inhibitor" have claimed coronaviruses target general host gene expression mechanisms to hurdle innate immune function through body fluid rather than T-cell-dependent immune mechanisms [23, 24]. Findings on "cattle" trace SARS back to the animal market [25], then to palm fruit [26], and later identify bats as animal cisterns.

3.5 Timeline of the Research on CRH

In the thematic evolution KM, the topics had evolved by the time sequence of "SARS-to-MERS-to-COVID-19" (Fig. 8). The topics such as "virus," "infection," and "spike protein" were core knowledge in this field in the early stage of the SARS outbreak. Coronavirus opts for spike protein to determine host-range trimers to accelerate cell-attached into host membranes fused with the virus [27].

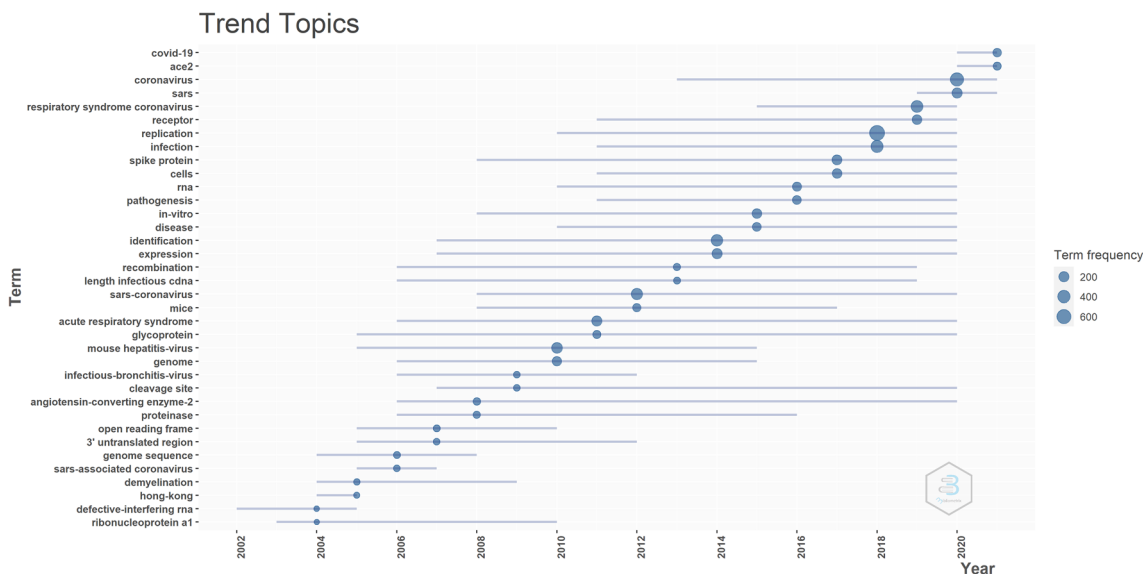The occurrence of SARS across Guangdong, China, in 2002, caused unprecedented effects and set off a wave

Fig. 8 Time zone for theme evolution

of research in this domain. The number of publications reached a peak during the outbreak with high-frequency words of "acute respiratory syndrome" and "SARS-CoV." ACE2, deemed to be a functional receptor for SARS-CoV, can aggravate the severity of the disease in the blend of cellular protein, as demonstrated by Imail et al. [28]. As is known to all, the replication from positive-strand RNA virus such as coronavirus can avoid antiviral defense mechanisms by the recombination of a host cell. The coronavirus recombination with a high mutation rate can adapt to new hosts and niche rapidly due to the unique mechanism of CRH [29], as shown in an outbreak of MERS across Saudi Arabia in 2012. Spike glycoprotein of MERS-CoV can be activated into fusion form by host cell transmembrane serine, endosome cathepsin, and furin protease, which is a way for viruses to enter cells and assemble virosomes. The RNA-dependent RNA polymerase as a vital factor of coronavirus especially for SARS-CoV-2 replication machinery can be supposed to be a primary target for the antiviral drug such as remdesivir and chloroquine [30].

SARS-CoV, MERS-CoV, and SARS-CoV-2 are quite different in terms of symptoms caused by coronavirus as shown in Table 2. SARS-CoV can cause severe acute respiratory syndrome (severe acute respiratory syndrome), mainly accompanied by fever, cough, headache, muscle pain, and respiratory infection symptoms. MERS broke out in Asia in 2015 and 2016, and had too many similarities with SARS in 2003. MERS compared to SARS is not easy to spread from person to person. MERS is less infectious than SARS, but which fatality rate is much higher (e.g. MERS close to 40%, SARS less than 10%). SARS-CoV-2 caused by antigenic variation of coronavirus—because of poverty of immunity of the population to the variant virus strain—spreads prevalently across the world in 2020, accompanied by mild dry cough, fatigue, breathless, diarrhea, and other symptoms but without runny nose and sputum. Compared with SARS-CoV and MERS-CoV, the speed and the spread of SARS-CoV-2 are really fast, but which attack performs weakly and leads to fatal complications rarely.

Coronavirus enters target cells through the binding of trimeric transmembrane spike glycoproteins to host cell receptors fused of virus [31]. Correctly, the prerequisite of coronavirus infection is the identification of host cells with receptors recognized by spike glycoproteins and the fusion of viral and cell membranes [32]. Although similar to SARS-CoV, SARS-CoV-2 can enter cells expressing ACE2 rather than ones without ACE2 or with other coronavirus receptors such as Aminopeptidase N and dipeptidyl peptidase, indicating ACE2 plays a crucial role in SARS-CoV-2 infection.

4 Discussion

Results binding Figs. 1 and 8 illustrate research on CRH, which originated from SARS, consist of three stages. In the first stage from 2003 through 2012, the number of literature on RCH increased sharply with the outbreak of SARS and then decreased slightly, which could be the first golden development period of studies for CRH. In the second phase from 2013 through 2019, those were on the rise rapidly with the emergence of MERS, viewed as the second valuable cycle. The emergence of Corona Virus Disease 2019 in Wuhan from China has aroused scholarly attention to CRH and will be a breakthrough point for the new type of CRH in the future. Scholars on CRH mainly come from the United States, China, and Germany, as given by Fig. 2.

Researches on CRH have focused on the domain of "characteristics of coronaviruses," "infection path," and "replication mechanism" in the light of "emergent keyword," according to Figs. 5, 6 and 7. Turning attention to "characteristics of coronaviruses" and "infection path," RNA compared with DNA virus mutates more frequently, as long as entering a host with weak immunity, and has enough space to generate something that can improve the infection efficiency of a new host. While replicating, RNA virus becomes the focal point of an epidemic through trans- or intra-species transmission. Coronaviruses have the giant genomes of any RNA virus, including segmented genomes that provide a large number of known gene expression strategies through this expanded coding capability. In short, a large number of unidentified human viruses exist in the global biosphere.

Commonly, coronavirus entry into the cell is mediated by transmembrane spike glycoprotein, forming a trimer with receptor binding and membrane fusion. For coronaviruses like SARS, MERS, and SARS-CoV-2, entering cells need to insert ACE2 protein exposed in mucosal cells, including the respiratory tract, digestive tract, and even ocular conjunctiva, which can spread and erupt rapidly in the crowd with a contract, droplets, and other means. Viewing SARS, SARS-CoV can inhibit the apoptosis of infected cells, promote cell survival, and give the virus time to complete the transcription, translation and assembly the virus in the host cell genome. The S protein in SARS-CoV is one of key proteins affecting receptor binding and host specificity. The receptor recognition and membrane fusion is mediated with an important role of virus infection of target cells. SARS-CoV-2 similar with SARS-CoV induces cell apoptosis with the internal replication machinery through the fusion of the mediation for S protein and ACE 2. Tissues with high expression of ACE2, such as lung, heart, kidney and intestine, are vulnerable to SARS-CoV-2. Hence, the interaction between S protein and cell membrane death receptor induces cells to initiate external apoptosis pathway.

Table 2 The comparison among CRH

Virus	SARS-CoV	MERS-CoV	SARS-CoV-2
Time	2002.11–2003.7	2012-	2019.12-
Spread	32 countries	27 countries	205 countries
Patients	Infection: 8422 Death: 916	Infection: 2454 Death: 858	Infection: 1,085,874+ Death: 59,112+
Mortality	10.88%	34.96%	5.44%
Transmission	Close respiratory tract flying, contact transmission, and fecal oral transmission; person-to-person transmission	Animal, limited human-to-human, and droplet transmission	Spray, contact(including self-contact caused by hand pollution), and short-range transmission in different respiratory aerosols
Incubation	General: 2–14 days Mostly: 3–5 days	General: 2–14 days Mostly: 3–5 days	General: 1–14 days Mostly: 3–7 days
Symptom	Fever, cough, headache, muscle pain, and respiratory tract infection	Typical: fever, cough, and shortness of breath; Majority: pneumonia; Others: gastrointestinal symptoms such as diarrhea	Typical: fever, fatigue, and dry cough; Majority: accompanied by nasal obstruction, runny nose, diarrhea, and other symptoms; Some cases: low fever or even no obvious symptoms
Interpersonal Communication	3	0.6	3.8
Gene similarity (on the basis of COVID-19)	0.793	0.528	1
Receptor	ACE2	DPP4	ACE2

Before talking about the "replication mechanism," as is known widely, viruses exhibit higher mutation rates due to loss of genetic fidelity function. RNA evolution success depends on phenotypic selection, which in turn depends on available genetic variability. Genetic variation derives from mutagenesis and can change the previous mutations followed by recombination. Concerning the "replication mechanism," the virus begins to proliferate when infecting original hosts. The more generations it proliferates, the more the cumulative variation will increase. Those with a higher chance of mutation, once evolving to have the potential to infect other species, can quickly adapt to new hosts and begin to spread widely. Out of mutation spontaneously and randomly, each generation of the virus inherits the existing mutations, on the basis, and continues to generate new mutations. Apart from genetic mutations, viruses can gain the ability to infect new species through genetic recombination between different strains. When two or more strains infect the same host, these can be genetically exchanged to produce a completely new one. When infecting the host cell, coronavirus can be continuously reproduced in the light of the genetic information of viral RNA. As proliferation and replication spread, some viruses undergo genetic mutation or genetic recombination and have acquired the ability to recognize human cell surface protein receptors to be enhanced to entering the human body through human–animal contact. They have a low infection capacity at the initial stage of entering a new host. With a period of adaptation, these viruses mutate and replicate again, begin to be transmitted from person to person forcefully, and thus result in a large-scale outbreak.

The database of research for CRH was retrieved and collected from the WOS, covering more information including country, institution, journal, and author to be analyzed more comprehensively and objectively. However, there are inherent limitations of our reviews of researches on CRH for lack of non-English literature and the consideration for other databases such as PubMed, Scopus, and Google Scholar. Among which, one shortcoming is the collection of core documents such as SCI and SSCI in WOS; another limitation does not contain the quasi-published or informally working papers on CRH. Further studies are needed to enrich the data sources and predict emerging trends on CRH using machine learning.

5 Conclusion

This article visualized a field of research for CRH using 9177 kinds of literature in SCIE and SSCI in WOS from 2002 through October 2021. KM with diverse ways, including country, co-citation, keywords clustering, and time zone, were employed to identify potential evolutionary mechanisms and research changing trends of CRH.

These results demonstrated that ACE2 might be a potential therapeutic target that inhibits virus-cell binding, suppresses inflammatory storms, and reduces lung and heart damage. These keywords of "infection," "spike protein," and "mutation with recombination" surveyed by literature-based visualization might be the vital entry hotspots for the researches on coronavirus replication in the future.

References

1. Cavanagh D, Horzinek MC (1993) GenusTorovirus assigned to the Coronaviridae. Arch Virol 128:395–397. https://doi.org/10.1007/BF01309450

2. Snijder EJ, Horzinek MC (1993) Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily. J Gen Virol 74:2305–2316. https://doi.org/10.1099/0022-1317-74-11-2305

3. Zhu N, Zhang D, Wang W et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017

4. Huang P, Wang H, Cao Z et al (2018) A rapid and specific assay for the detection of MERS-CoV. Front Microbiol 9:1101. https://doi.org/10.3389/fmicb.2018.01101

5. Huang C et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:15–21. https://doi.org/10.1016/S0140-6736(20)30183-5

6. Li W, Shi Z, Yu M et al (2005) Bats are natural reservoirs of SARS-like coronaviruses. Science 310:676–679. https://doi.org/10.1126/science.1118391

7. Reusken C et al (2013) Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infect Dis 13:859–866. https://doi.org/10.1016/S1473-3099(13)70164-6

8. Kleine-Weber H, Elzayat MT, Wang L et al (2018) Mutations in the spike protein of MERS-CoV transmitted in Korea increase resistance towards antibody-mediated neutralization. J Virol 93(2):e01381-e1418. https://doi.org/10.1128/JVI.01381-18

9. Zhou J, Li C, Zhao G et al (2017) Human intestinal tract serves as an alternative infection route for Middle East respiratory syndrome coronavirus. Sci Adv 3(11):eaao4966. https://doi.org/10.1126/sciadv.aao4966

10. Guan W, Ni Z, Hu Y et al (2020) Clinical characteristics of 2019 novel coronavirus infection in China. MedRxiv. https://doi.org/10.1101/2020.02.06.20020974

11. Park S, Sestak K, Hodgins DC et al (1998) Immune response of sows vaccinated with attenuated transmissible gastroenteritis virus (TGEV) and recombinant TGEV spike protein vaccines and protection of their suckling pigs against virulent TGEV challenge exposure. Am J Vet Res 59(8):1002–1008. https://doi.org/10.1016/S0167-5877(98)00069-5

12. Rota PA, Oberste MS, Monroe SS et al (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300(5624):1394–1399. https://doi.org/10.1126/science.1085952

13. Dahlander L, Gann DM (2010) How open is innovation? Res Policy 39(6):699–709. https://doi.org/10.1016/j.respol.2010.01.013

14. Chen C, Dubin R, Kim MC (2014) Emerging trends and new developments in regenerative medicine: a scientometric update (2000–2014). Expert Opin Biol TH 14(9):1295–1317. https://doi.org/10.1517/14712598.2014.920813

15. Ksiazek TG, Erdman D, Goldsmith CS et al (2003) A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 348(20):1953–1966. https://doi.org/10.1056/NEJMoa030781

16. Snijder E, Bredenbeek P, Dobbe J et al (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol 331(5):991–1004. https://doi.org/10.1016/S0022-2836(03)00865-9

17. Drosten C, Günther S, Preiser W et al (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 348:1967–1976. https://doi.org/10.1056/NEJMoa030747

18. Peiris JSM et al (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 361:1319–1325. https://doi.org/10.5222/terh.2003.26734

19. Zhou P et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–290. https://doi.org/10.1038/s41586-020-2012-7

20. Hoffmann M, Kleine-Weber H, Schroeder S et al (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181(2):271–280. https://doi.org/10.1016/j.cell.2020.02.052

21. Li W, Moore MJ, Vasilieva N et al (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426:450–454. https://doi.org/10.1038/nature02145

22. Zaki AM, Van Boheemen S, Bestebroer TM et al (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367:1814–1820. https://doi.org/10.1056/NEJMoa1211721

23. Yang Z et al (2004) A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. Nature 428:561–564. https://doi.org/10.1038/nature02463

24. Subbarao K et al (2004) Prior infection and passive transfer of neutralizing antibody prevent replication of severe acute respiratory syndrome coronavirus in the respiratory tract of mice. J Virol 78:3572–3577. https://doi.org/10.1128/JVI.78.7.3572-3577.2004

25. Guan Y, Zheng BJ, He YQ et al (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302:276–278. https://doi.org/10.1126/science.1087139

26. Kan B et al (2005) Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. J Virol 79(18):11892–11900. https://doi.org/10.1128/JVI.01072-06

27. Jackwood MW et al (2001) Spike glycoprotein cleavage recognition site analysis of infectious bronchitis virus. Avian Dis 45(2):366–372. https://doi.org/10.2307/1592976

28. Imai Y, Kuba K, Rao S et al (2005) Angiotensin-converting enzyme 2 protects from severe acute lung failure. Nature 436:112–116. https://doi.org/10.1038/nature03712

29. Woo PCY, Lau SKP, Yip CCY et al (2006) Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. J Virol 80(14):7136–7145. https://doi.org/10.1128/JVI.00509-06

30. Gao Y, Yan L, Huang Y et al (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science 368:779–782. https://doi.org/10.1126/science.abb7498

31. Shang J, Zheng Y, Yang Y et al (2018) Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. PLoS Pathog 14(4):e1007009. https://doi.org/10.1371/journal.ppat.1007009

32. Li F (2016) Structure, function, and evolution of coronavirus spike proteins. Ann Rev Virol 3(1):237–261. https://doi.org/10.1146/annurev-virology-110615-042301