

Visualization analysis of CRISPR/Cas9 gene editing technology studies^{*}

Quan-sheng DU^{†‡1}, Jie CUI², Chun-jie ZHANG³, Ke HE⁴

⁽¹⁾Department of Life Sciences, National Natural Science Foundation of China, Beijing 100085, China)

⁽²⁾Guangdong Techno-economy Research and Development Center, Guangzhou 510070, China)

⁽³⁾School of Science, Hangzhou Dianzi University, Hangzhou 310018, China)

⁽⁴⁾College of Animal Science and Technology, Zhejiang A&F University, Lin'an 311300, China)

[†]E-mail: duqs@nsfc.gov.cn

Received Aug. 12, 2016; Revision accepted Aug. 28, 2016; Crosschecked Sept. 11, 2016

Abstract: The clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system is an adaptive immune defense system that resists the invasion of viruses and plasmids heterologous genetic material in bacteria and archaea. Taking the literature related to gene editing technology of CRISPR/Cas9 from the Web of Science database from 2002 to 2015, we use the software CiteSpaceV to analyze co-cited literature in order to establish the research hotspots and fronts recently in this field by knowledge mapping.

Key words: CRISPR/Cas9, CiteSpaceV, Visualization analysis

<http://dx.doi.org/10.1631/jzus.B1601985>

CLC number: Q78; G35


1 Introduction

Previous work has shown that there are many clustered regularly interspaced short palindromic repeats (CRISPRs) in bacteria and archaea genomes. CRISPR loci are usually composed of short and highly conserved repeats of 21–48 bp, which are separated by 26–72-bp non-repetitive spacer sequences. CRISPR identifies the target gene through these spacer sequences and CRISPR-associated (Cas) protein families promote cleavage of the target genes. This forms an immune defense system of prokaryote cell and provides protection against the invasion of foreign genetic material to maintain genome stability and integrity. In recent years, the rapid progress in developing CRISPR/Cas9 system into a set of tools

for cell and molecular biology research has been remarkable, likely due to the simplicity, high efficiency and versatility. The system is by far the most user-friendly for gene editing. Therefore, it is widely used in the direct genetic transformation of human cells, as well as mouse, zebra fish, rice, etc. Furthermore, this system shows a broad prospect of clinical application in the treatment of acute lymphocytic leukemia, hereditary vision loss disease, acquired immune deficiency syndrome, etc., which has made it the most popular technology in the field of gene editing. CRISPR has emerged in *Science's* Breakthrough of the year thrice, in 2012, 2013, and 2015, the first two times as a second place together with other gene editing techniques. However, it broke away from the others, manifesting its spectacular achievements in 2015. In order to understand the path of knowledge evolution, the dynamic process and characteristics in CRISPR/Cas9 research, this paper uses the information visualization software CiteSpaceV to analyze the recently published literature on CRISPR/Cas9 and draw knowledge maps.

[‡] Corresponding author

^{*} Project supported by the Zhejiang Provincial Natural Science Foundation of China (No. LY16A010013)

 ORCID: Quan-sheng DU, <http://orcid.org/0000-0003-4300-4522>

© Zhejiang University and Springer-Verlag Berlin Heidelberg 2016

2 Methods

2.1 Data retrieval

We retrieved research papers in the field of gene editing from the Web of Science (WOS) database by keywords, and the search formula is TS (Topic)=CRISPR/Cas9 or CRISPR or CAS9. The retrieval time is from 2002 to 2015 (since CRISPR was officially named in 2002), the publication type only includes Article and Review, and the language is limited to English. We obtained 1908 papers using CiteSpaceV built-in deduplication function. Each paper contains 17 fields: PT (publication type), AU (authors), TI (document title), SO (publication name), DE (author keywords), ID (keywords plus[®]), AB (abstract), C1 (author address), CR (cited references), NR (cited reference count), TC (times cited), PY (year published), VL (volume), PG (page count), SC (subject category), UT (unique article identifier), and ER (end of record), which serves as a formatted file.

2.2 Analysis methods

CiteSpace is visual information analysis software based on literature co-citation analysis theory and pathfinder network scaling, which was developed by Prof. Chao-mei CHEN built on Java platform (CiteSpaceV, Version 4.4. R1 (32-bit); <http://cluster.cis.drexel.edu/~cchen/citespace>) (Chen, 2004; 2006).

CiteSpaceV presents the citation history of a paper as a tree ring of the node. The color of a citation ring denotes the time of corresponding citations. The radius of the node indicates the number of total citations. The area of circles represents the number of publications. The thin lines show the cooperation relationships between countries or institutions. Cluster maps and time-zone views can reveal the development trend of a discipline or field of knowledge in a specific period, and this can be used to present the evolution process of the research fronts and hotspots of related fields (Chen, 2012; Chen and Leydesdorff, 2014).

CiteSpace software adopts several structural metrics of co-citation networks, which include betweenness centrality, modularity, and silhouette (Chen *et al.*, 2010). The betweenness centrality metric measures the extent of one node in the middle of a path that connects other nodes in the network (Brandes, 2001; Freeman, 1977). High betweenness

centrality values recognize potentially supreme scientific literature (Chen, 2005). The modularity (Q) measures the strength of a network divided into independent communities or modules. The modularity score ranges from 0 to 1, and usually $Q > 0.3$ means that the division of the community structure is conspicuous (Newman, 2006). Mean Silhouette (S) metric shows which nodes are well within their clusters, and which ones are only somewhere in between clusters, and generally $S > 0.5$ means that the cluster is considered logical (Rousseeuw, 1987).

Cluster labels are chose from noun phrases and index terms of citing articles of each cluster. These terms are sorted by term frequency-inverse document frequency (TF-IDF) algorithm, which is used to evaluate the importance of the word in an article. Labels selected by TF-IDF weightiness prefer to represent the most salient aspect of a cluster and emphasize the research mainstream (Chen *et al.*, 2010).

3 Results and analysis

3.1 Brief history of CRISPR research

The story about CRISPR/Cas9 system began with the discovery of a highly homologous sequence clustered in the *Escherichia coli* genome (Ishino *et al.*, 1987; Nakata *et al.*, 1989). In 2002, Jansen and colleagues from Utrecht University (Utrecht, the Netherlands) officially named it Clustered Regularly Interspaced Short Palindromic Repeats and created the CRISPR acronym. Furthermore, CRISPR-associated (*cas*) genes (*cas1*, *cas2*, *cas3*, *cas4*) were identified to be well conserved and usually adjacent to the repeat elements. However, the biological function of CRISPR was still unclear (Mojica *et al.*, 2000; Jansen *et al.*, 2002).

The incubation period is from 2002 to 2006 when our understanding of CRISPR and associated genes was refined. Even though the annual number of research papers was few, some remarkable research results appeared in 2005. Bioinformatics analysis revealed attractive picture, wherein the CRISPR loci of three different types of CRISPR systems (types I–III) might specifically defense against invading foreign genetic elements, such as bacteriophages and plasmids, with the constitutive

spacers. In addition, some families of *cas* genes are extremely conserved. It suggests that CRISPR sequences may be associated with bacterial immunity (Bolotin *et al.*, 2005; Haft *et al.*, 2005; Mojica *et al.*, 2005).

The germination period is from 2007 to 2011 when a series of artful experimentation has provided fresh viewpoints into CRISPR functional mechanism. Among of them the significant findings include: Barrangou *et al.* (2007) from the food ingredients company Danisco in Copenhagen, Denmark, found that bacteria integrated new sequences originated from phage genomic sequences. CRISPR, together with associated *cas* genes, provided immunity against phages, and resistance specificity was controlled by spacer-phage sequence similarity. After the acquisition of Danisco, DuPont used this technique to breed a more resistant strain for industrial production. Marraffini and Sontheimer (2008) found that horizontal gene transfer (HGT) in bacteria and archaea happens through phage transduction or conjugation. CRISPR loci can confine the spread of antibiotic resistance by counteracting multiple channel of HGT. Research on CRISPR/Cas function aroused further interest in the study of its mechanism, and total cites of this paper were more than 300. Brouns *et al.* (2008) first proved that the guide RNA in the CRISPR system is the critical factor for bacterial immunity to viruses, and the CRISPR-associated complex is the key role in the antiviral defense.

Year 2012–2013 was a milestone in the progress of CRISPR/Cas9 research. Representative research results are: Jinek *et al.* (2012) published a highlight article with total cites up to 661 and a betweenness centrality of 0.45 on *Science* journal. The paper reveals a family of endonucleases that uses dual RNAs for site-specific DNA cleavage and the potential to develop an RNA-dependent gene editing system. Mali *et al.* (2013) reported that they successfully engineered the type II bacterial CRISPR system to function with custom guide RNA (gRNA) in human cells. The results demonstrated the possibility of CRISPR-mediated gene targeting for RNA guided, powerful, and multipurpose mammalian genome editing. The total cites of this paper is 669, ranking the second highest. Cong *et al.* (2013) concluded that short RNAs might guide Cas9 nucleases to trigger precise cleavage at endogenous genomic loci in mammalian

cells. Multiple guide sequences permute into single CRISPR loci to enable simultaneous editing of several target sites within the mammalian genome, demonstrating easy programmability and wide applicability of the technology. The paper has the highest total cites (714) in the field.

So far, scientists have successfully applied the technology on distinct species, including human cells, animals, and plants, indicating the simplicity and efficiency of the CRISPR/Cas9 system. The power of CRISPR systems will undoubtedly transform life science research and spur the development of novel cell genetics and molecular biology. Fig. 1 shows the number changes of research papers in the field of CRISPR/Cas9 collected by the WOS database from 2002 to 2015 (up to December 2015).

3.2 Analysis of countries' (districts') and institutions' cooperation

By drawing the hybrid network map of cooperative relationships, we analyze cooperative relationships between countries (districts) and institutions, research strength, and other information. We set the node type as Country and Institution. We cluster it and obtain a mixed network map consisting of 147 nodes and 197 edges (a density of 0.0184), as shown in Fig. 2. The United States has the highest number of publications, which is far more than those of other countries and accounts for more than half of the total with total cites up to 3098. This is followed by China (691) ranking the second, then Germany (434), Japan (381), and France (346). The citation number of the University of California, Berkeley, is 250, ranking first among research institutions. Harvard University has 246, Massachusetts Institute of Technology (MIT) 181, and the Chinese Academy of Sciences 128.

Most countries have cooperative relationships with the United States, which is at the core of this field. The centrality score of United States is 0.64, much higher than those of other countries, France (0.28), and then Canada (0.21) and England (0.24) follow. The centrality of China is 0.09, indicating that the international influence in this field of China is relatively weak. As a research powerhouse of gene editing technology, the United States statistics are distributed among various universities. There are six American universities in the Top 10 research institutions, including Harvard University, University of

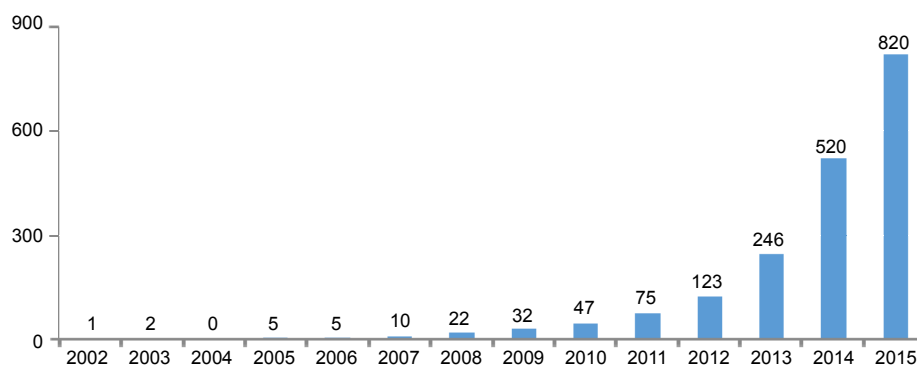


Fig. 1 Number of CRISPR/Cas9 research papers (Data source: Web of Science, up to December 2015)

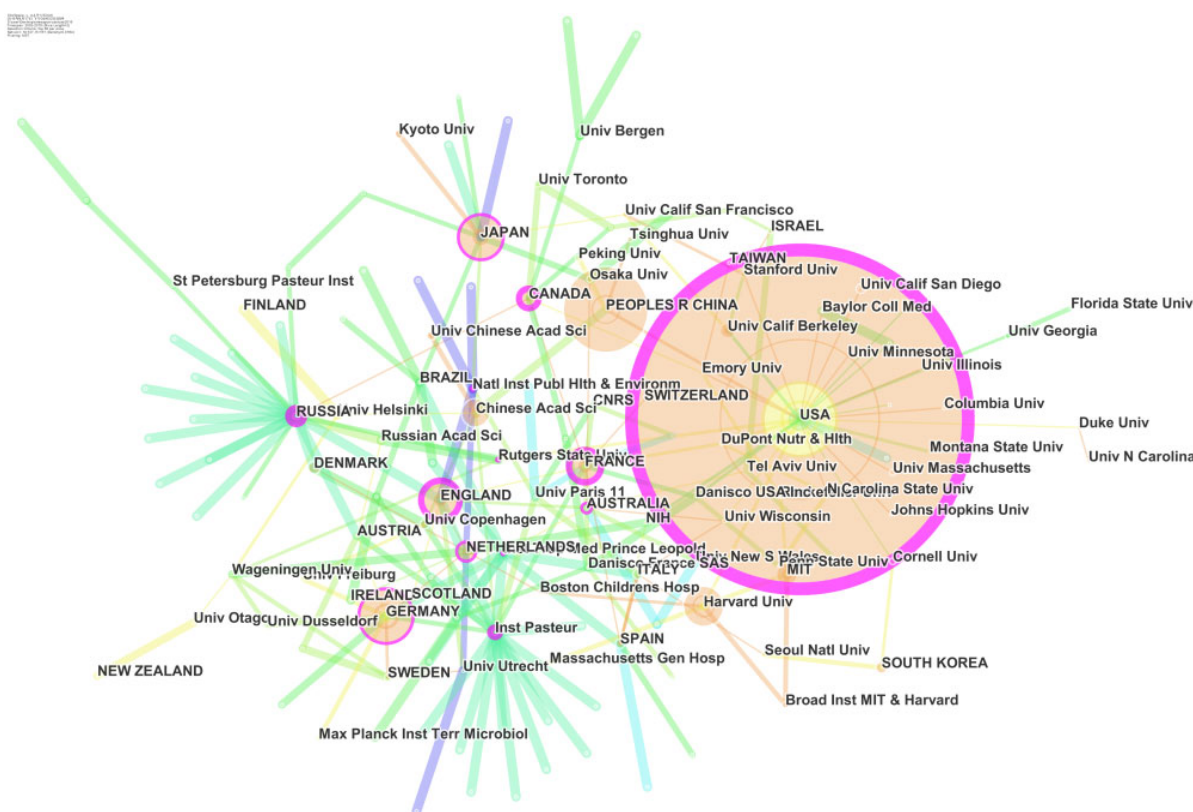


Fig. 2 National and institutional hybrid network map

California (UC) Berkeley, MIT, and University of Georgia. Chinese research efforts are mainly in the Chinese Academy of Sciences, which has the highest centrality in the institutions as shown in Table 1.

3.3 Statistical analysis of main authors

For this the process is: retrieve main authors in the field of gene editing from WOS database, and arrange Top 20 authors by the number of research papers, as shown in Table 2. K.S. Makarova from

Biotechnology Information Center of National Institutes of Health (NIH) and R.M. Terns from University of Georgia both have the highest betweenness centrality of 0.17. Makarova *et al.* (2006) found that CRISPR sequences might be associated with immune protection mechanisms of bacteria through comparative genomics study, which is the important turning point of gene editing research. Hale *et al.* (2009) mainly studied a CRISPR/Cas effector complex that is comprised of small invader-targeting RNAs from

the CRISPR loci and the related Cas proteins, which cleave complementary target RNAs at a fixed distance. The results indicated that prokaryotes possess a unique RNA silencing system that functions by homology-dependent cleavage of heterogenous RNAs. R. Barrangou from North Carolina State University has the most papers at 54, with total cites of 4978, field H index of 25, and a betweenness centrality of 0.06. His main contribution is to verify that the adaptive immunity of bacteria related to spacer sequence, which lays the foundation for further study (Barrangou *et al.*, 2007). F. Zhang from Broad Institute of MIT and Harvard published 31 papers, whose total cites, highest cites of a single paper, and average cites

all are the highest, 6041, 1629, and 194.87, respectively, and has the field H index of 21. F. Zhang first applied CRISPR technology to the specific removal of two genes in human cells (Cong *et al.*, 2013) and opened paradigm application of gene editing technology in eukaryotic cells.

4 Research frontier analysis

Document co-citation analysis (DCA) studies are among the most commonly used methods in quantitative studies of science (Chen, 2004; 2006). We set network nodes as Cited References and select

Table 1 A partial list of the countries and institutions in the hybrid network map (Top 10 citations)

Citation	Betweenness centrality	Country	Publish year	Citation	Betweenness centrality	Institution	Publish year
3098	0.64	USA	2005	250	0.03	UC Berkeley	2008
691	0.09	China	2010	246	0.08	Harvard University	2012
434	0.17	Germany	2006	181	0.00	MIT	2013
381	0.16	Japan	2003	128	0.10	Chinese Academy of Sciences	2010
346	0.28	France	2005	89	0.03	University of Georgia	2008
171	0.24	England	2003	85	0.05	UC San Francisco	2013
165	0.12	The Netherlands	2002	71	0.03	Duke University	2013
151	0.00	Korea	2012	69	0.01	Massachusetts General Hospital	2013
150	0.21	Canada	2007	63	0.02	University of Copenhagen	2008
100	0.16	Australia	2005	61	0.06	Stanford University	2011

Table 2 Statistical analysis of main authors in gene editing research field

Author	Author institution	Paper number	Total cites	Highest cites of a single paper	Average cites	Field H index	Betweenness centrality
R. Barrangou	N. Carolina State Univ.	54	4978	1046	92.19	25	0.06
J.A. Doudna	Univ. Calif. Berkeley	40	4921	1191	123.03	27	0.02
E.V. Koonin	NIH, Natl. Ctr. Biotechnol. Informat.	33	2720	595	82.42	21	0.00
F. Zhang	Broad Inst. MIT & Harvard	31	6041	1629	194.87	21	0.03
J. van der Oost	Wageningen Univ.	28	2821	595	100.75	20	0.09
L.A. Marraffini	Rockefeller Univ.	26	4139	1629	159.19	19	0.04
J.S. Kim	Seoul Natl. Univ.	25	1189	399	47.56	12	0.01
S.J.J. Brouns	Wageningen Univ.	24	2712	595	113.00	18	0.00
P. Horvath	DuPont Nutr. & Hlth.	22	4216	1046	191.64	20	0.03
R.A. Garrett	Univ. Copenhagen	22	802	115	36.45	16	0.04
B. Wiedenheft	Montana State Univ.	21	1763	452	83.95	13	0.01
K.S. Makarova	NIH, Natl. Ctr. Biotechnol. Informat.	21	2480	599	118.10	18	0.17
H. Li	Florida State Univ.	21	567	224	27.00	10	0.16
R.M. Terns	Univ. Georgia	20	1385	368	69.25	15	0.17
M.P. Terns	Univ. Georgia	20	1385	368	69.25	15	0.00
Y. Zhang	Gladstone Inst. Cardiovasc. Dis.	19	825	453	43.42	7	0.01
Y.I. Wolf	NIH, Natl. Ctr. Biotechnol. Informat.	19	1569	595	82.58	12	0.01
T. Sakuma	Hiroshima Univ.	19	198	43	10.42	8	0.00
L. Randau	Max Planck Inst. Terr. Microbiol.	19	234	48	12.32	10	0.00
E.R. Westra	Wageningen Univ.	18	1847	562	102.61	14	0.06

Only research article and review papers are contained

the Top 35 most cited or occurring items from each slice, with other parameters remaining the same with above, and then obtain an automatic cluster map. Extract labels in citing literature of citation clusters, thereby characterizing the research fronts in the field of gene editing technology. The co-cited literature in the hybrid clustering map of this paper, 11 co-citation clusters are composed of 162 nodes and 205 connection lines, with a Q value of 0.7506 and S value of 0.5644, indicating that the clustering structure is significant and efficiency and reliability are within reasonable limits, as shown in Fig. 3.

The network divided into 11 co-citation clusters and they are labeled by index terms from their own citers. The largest four clusters are summarized in the co-cited literature cluster map, as shown in Table 3.

The largest cluster (#0) has 26 members and a silhouette value of 0.796. It is labeled as experimental definition | new spacers by TF-IDF. CRISPR system is a bacterial acquired immune system. After the first phage infection, bacteria obtain the phage DNA fragment and integrate it into its own genome as a formation of memory. Stern *et al.* (2010) published a research paper in *Trends in Genetics*, in which citing literature accounts for 69% of the cited literature in the cluster. Studies show that the CRISPR is the key adaptive defense system of bacterial and archaeal genomes to resist phage infections.

The second largest cluster (#1) has 21 members and a silhouette value of 0.772. It is labeled with RNA-directed adaptive immunity by TF-IDF. Karginov and Hannon (2010) published a review paper in

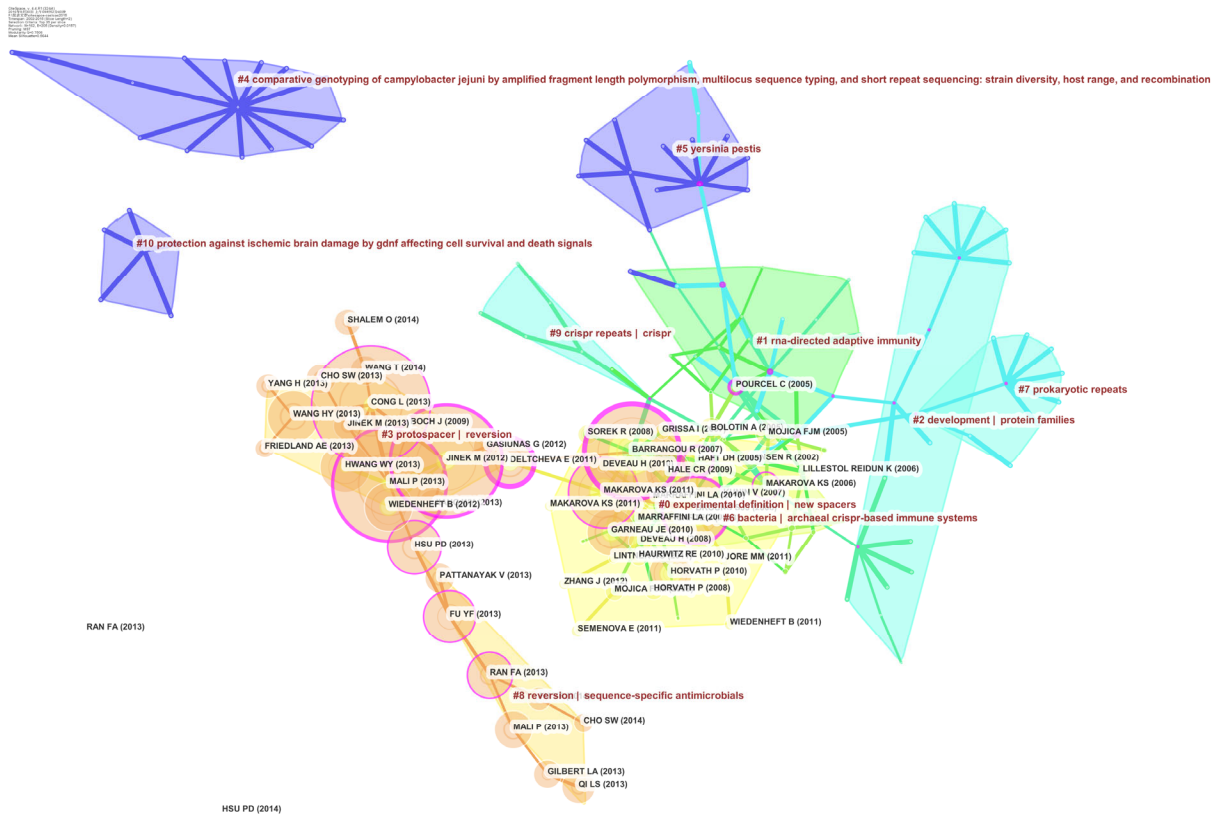


Fig. 3 An overview of the co-citation networks

Table 3 Main clusters and label cluster information of the co-cited literature map

Cluster ID	Size	Silhouette	Label (TF-IDF)	Average year
0	26	0.796	Experimental definition new spacers	2008
1	21	0.772	RNA-directed adaptive immunity	2001
2	18	0.984	Development protein families	2000
3	16	0.937	Protospacer reversion	2012

Molecular Cell, accounts for 67% of cited literature in the cluster. This paper summarized the CRISPR/Cas system and its mechanism found in bacteria and archaea, and compared it with the gene silencing mechanism mediated by small RNA molecules in eukaryotic cells.

The third largest cluster (#2) has 18 members and a silhouette value of 0.984. It is labeled as development | protein families by TF-IDF. The paper with the most citing literature in this cluster is about the proteomics structure and function research of psychrophilic methanogens (Saunders *et al.*, 2005), and pointed out that the genome of psychrophilic methanogens presents a similar CRISPR locus and it may be related to the gene bidirectional metastasis of bacteria and archaea.

The fourth largest cluster (#3) has 16 members and a silhouette value of 0.937. It is labeled as protospacer | reversion by TF-IDF. The paper published in *Science* in 2012 is an important breakthrough in this field, which clarified the mechanism of the CRISPR/Cas9 system (Jinek *et al.*, 2012). Cas9 is a 160-kD protein that can be utilized for RNA recognition and cleavage of target DNA. The paper published in *Genes & Development* by Malina *et al.* (2013), refers to 56% of the literature in the cluster. This paper showed that researchers used CRISPR/Cas9 to target cut p53 gene in order to prove its potential use in the construction of mouse cancer models.

5 Summary and perspectives

The research history of the CRISPR/Cas9 system amply shows the importance of interdisciplinarity. Interdisciplinary points are usually new scientific growth points and new scientific fronts, which are the most likely to lead to significant scientific breakthroughs. CRISPR's research is composed of the disciplines of microbiology, food science, bioinformatics, comparative genomics, clinical medicine, etc., and especially bioinformatics studies show that CRISPR sequences are complementary to phage DNA sequences, which suggests that it may relate to bacterial immune protection and points out the direction for subsequent function studies. Subsequently, researchers will carry out extensive studies in a variety of model organisms, especially clinical research re-

lated to human cells and major diseases, making CRISPR be one of the most popular methods of biotechnology and medicine.

The gene editing technology of CRISPR/Cas9 needs to be studied in greater depth. Though in the past few years an explosion has been seen in our understanding of the CRISPR/Cas9 system, many issues remain unresolved. Using co-cited literature analysis, we demonstrated the research fronts in the field, and the largest cluster is the study on the CRISPR/Cas9's mechanism. There are scientific issues worthy of further study. Why the CRISPR locus only exists in some bacteria and how bacteria and bacteriophage co-evolve? Since the presence of more serious off-target effects in the mammalian cell's gene editing process, what is the mechanism by which CRISPR/Cas9 accurately identifies and cuts the heterologous genetic material in bacterial cells? What is the mechanism to precisely regulate Cas9 protein's expression and jointly assemble them with single-guide RNA (sgRNA) into a functional complex?

Proposing significantly original scientific issues remains the dilemma we face. China obtains the second place in number of papers published in the field of gene editing, with betweenness centrality of 0.09, far behind in the number of personal papers and highly cited papers, indicating that Chinese research forces in this field are dispersed and have little influence. Many scientists are good at tracking international hot spots, following scientific ideas of others and doing some expansion of work in accordance with existing theories. However, because of a lack of original thought, they rarely propose significant scientific issues and thus it is difficult to make influential achievements. Physicist Albert Einstein quotes, "The mere formulation of a problem is far more often essential than its solution, which may be merely a matter of mathematical or experimental skill."

Acknowledgements

We thank especially Dr. Xin-zhang GAO (Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences) for assistance with the analysis software.

Compliance with ethics guidelines

Quan-sheng DU, Jie CUI, Chun-jie ZHANG, and Ke HE declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Barrangou, R., Fremaux, C., Deveau, H., *et al.*, 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**(5819):1709-1712.
<http://dx.doi.org/10.1126/science.1138140>
- Bolotin, A., Oquinis, B., Sorokin, A., *et al.*, 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**(8):2551-2561.
<http://dx.doi.org/10.1099/mic.0.28048-0>
- Brandes, U., 2001. A faster algorithm for betweenness centrality. *J. Math. Sociol.*, **25**(2):163-177.
<http://dx.doi.org/10.1080/0022250X.2001.9990249>
- Brouns, S.J., Jore, M.M., Lundgren, M., *et al.*, 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**(5891):960-964.
<http://dx.doi.org/10.1126/science.1159689>
- Chen, C.M., 2004. Searching for intellectual turning points: progressive knowledge domain visualization. *PNAS*, **101**(Suppl. 1):5303-5310.
<http://dx.doi.org/10.1073/pnas.0307513100>
- Chen, C.M., 2005. The centrality of pivotal points in the evolution of scientific networks. IUI '05 Proceedings of the 10th International Conference on Intelligent User Interfaces, San Diego, California, USA. ACM, New York, USA, p.98-105.
<http://dx.doi.org/10.1145/1040830.1040859>
- Chen, C.M., 2006. Citespace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inform. Sci. Technol.*, **57**(3):359-377.
<http://dx.doi.org/10.1002/asi.20317>
- Chen, C.M., 2012. Predictive effects of structural variation on citation counts. *J. Am. Soc. Inform. Sci. Technol.*, **63**(3):431-449.
<http://dx.doi.org/10.1002/asi.21694>
- Chen, C.M., Leydesdorff, L., 2014. Patterns of connections and movements in dual-map overlays: a new method of publication portfolio analysis. *J. Assoc. Inform. Sci. Technol.*, **65**(2):334-351.
<http://dx.doi.org/10.1002/asi.22968>
- Chen, C.M., Ibekwe-sanjuan, F., Hou, J.H., 2010. The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis. *J. Am. Soc. Inform. Sci. Technol.*, **61**(7):1386-1409.
<http://dx.doi.org/10.1002/asi.21309>
- Cong, L., Ran, F.A., Cox, D., *et al.*, 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**(6121):819-823.
<http://dx.doi.org/10.1126/science.1231143>
- Freeman, L.C., 1977. A set of measuring centrality based on betweenness. *Sociometry*, **40**(1):35-41.
<http://dx.doi.org/10.2307/3033543>
- Haft, D.H., Selengut, J., Mongodin, E.F., *et al.*, 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLOS Comput. Biol.*, **1**(6):e60.
<http://dx.doi.org/10.1371/journal.pcbi.0010060>
- Hale, C.R., Zhao, P., Olson, S., *et al.*, 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, **139**(5):945-956.
<http://dx.doi.org/10.1016/j.cell.2009.07.040>
- Ishino, Y., Shinagawa, H., Makino, K., *et al.*, 1987. Nucleotide-sequence of the *iap* gene, responsible for alkaline-phosphatase isozyme conversion in *Escherichia-coli*, and identification of the gene product. *J. Bacteriol.*, **169**(12):5429-5433.
- Jansen, R., van Embden, J.D.A., Gastra, W., *et al.*, 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**(6):1565-1575.
<http://dx.doi.org/10.1046/j.1365-2958.2002.02839.x>
- Jinek, M., Chylinski, K., Fonfara, I., *et al.*, 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**(6096):816-821.
<http://dx.doi.org/10.1126/science.1225829>
- Karginov, F.V., Hannon, G.J., 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell*, **37**(1):7-19.
<http://dx.doi.org/10.1016/j.molcel.2009.12.033>
- Makarova, K.S., Grishin, N.V., Shabalina, S.A., *et al.*, 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.*, **1**(7):1-26.
<http://dx.doi.org/10.1186/1745-6150-1-7>
- Mali, P., Yang, L.H., Esvelt, K.M., *et al.*, 2013. RNA-guided human genome engineering via Cas9. *Science*, **339**(6121):823-826.
<http://dx.doi.org/10.1126/science.1232033>
- Malina, A., Mills, J.R., Cencic, R., *et al.*, 2013. Repurposing CRISPR/Cas9 for in situ functional assays. *Gene Dev.*, **27**(23):2602-2614.
<http://dx.doi.org/10.1101/gad.227132.113>
- Marraffini, L.A., Sontheimer, E.J., 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**(5909):1843-1845.
<http://dx.doi.org/10.1126/science.1165771>
- Mojica, F.J., Diez-Villasenor, C., Soria, E., *et al.*, 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.*, **36**(1):244-246.
<http://dx.doi.org/10.1046/j.1365-2958.2000.01838.x>
- Mojica, F.J.M., Diez-Villaseñor, C., García-Martínez, J., *et al.*, 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**(2):174-182.
<http://dx.doi.org/10.1007/s00239-004-0046-3>
- Nakata, A., Amemura, M., Makino, K., 1989. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.*, **171**(6):3553-3556.
- Newman, M.E., 2006. Modularity and community structure in networks. *PNAS*, **103**(23):8577-8582.
<http://dx.doi.org/10.1073/pnas.0601602103>

- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **20**:53-65.
[http://dx.doi.org/10.1016/0377-0427\(87\)90125-7](http://dx.doi.org/10.1016/0377-0427(87)90125-7)
- Saunders, N.F.W., Goodchild, A., Raftery, M., et al., 2005. Predicted roles for hypothetical proteins in the low-temperature expressed proteome of the Antarctic archaeon *Methanococoides burtonii*. *J. Proteome Res.*, **4**(2): 464-472.
<http://dx.doi.org/10.1021/pr049797+>
- Stern, A., Keren, L., Wurtzel, O., et al., 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.*, **26**(8):335-340.
<http://dx.doi.org/10.1016/j.tig.2010.05.008>

中文概要

题 目: 基因编辑技术 CRISPR/Cas9 研究的可视化分析

目 的: 理解 CRISPR/Cas9 研究的知识进化路径, 以及动态过程和特点。

创新点: CRISPR/Cas9 系统是细菌和古细菌在长期演化过程中形成的一种抵御异源遗传物质入侵的适应性免疫防御机制。它被广泛应用于人类细胞、小鼠、斑马鱼、大米等的基因定向改造。本研究利用信息可视化软件 CiteSpaceV 分析了近年发表的 CRISPR/Cas9 有关文献, 并绘制了知识图谱。

方 法: 基于 Web of Science 数据库中 2002 年至 2015 年有关基因编辑技术 CRISPR/Cas9 的文献资料为研究对象。利用信息可视化软件 CiteSpaceV 进行共被引文献分析, 以知识图谱方式展现该领域近年来的研究热点和前沿。

结 论: CRISPR/Cas9 系统的历史研究充分表明跨学科的重要性。跨学科点通常是新的科学的增长点和新的科学前沿, 这是最有可能产生显著科学突破的地方。

关键词: CRISPR/Cas9; CitespaceV; 可视化分析