



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Monitoring infectious diseases in the big data era

Yuanqiang Zou · Yousong Peng · Lizong Deng ·
Taijiao Jiang

Published online: 24 December 2014
© Science China Press and Springer-Verlag Berlin Heidelberg 2014

Infectious diseases have caused and will continue to cause a significant impact on public health. Following the unprecedented H7N9 outbreak in China in early 2013, the recent spreading of the Ebola virus in West Africa has once again placed the danger of infectious diseases in the public eye and has caused significant public alarm all over the world. Given the frequent recurrence of rapidly evolving pathogens, such as seasonal influenza viruses, and the sporadic introduction of novel pathogens, such as the 2003 SARS coronavirus in China, the prevention and control of infectious diseases has become a major global public health issue, which relies heavily on an effective surveillance strategy.

In China, the surveillance of infectious diseases is mainly conducted by a nationwide monitoring network consisting of numerous hospitals and health departments. Under the guidance of the Chinese Center for Disease Control and Prevention (CDC), these monitors will watch, report, collect and analyze suspicious samples. In addition to the traditional methods of collecting immunological and biological data for emerging pathogens, large-scale gene sequencing has now been widely used in the surveillance of infectious diseases. The large amount of genomic data on infectious diseases have facilitated a more accurate and rapid identification of pathogens, thereby assisting in the

prediction of their potential pathological and epidemical characteristics. Taking the unprecedented 2013 H7N9 virus as an example, the considerable genomic data enabled Chinese researchers to rapidly identify the novel pathogens and infer their origins and evolutionary pathways [1, 2]; this assisted the formulation of proper measures for the control of infections of the viruses to humans.

Given the ever increasing genomic data, how to effectively model the big genomic data to infer the characteristics of infectious diseases has challenged the traditional molecular evolutionary analysis approaches. Previous studies have shown that, compared with the traditional phylogenetic analysis, the extraction of co-evolutionary signals and advanced features from the big gene data relating to seasonal influenza viruses could capture the characteristics of their antigenic changes; this would lead to a more accurate and timely recommendation of seasonal vaccine strains [3, 4]. There is no doubt that the coupling of large-scale gene sequencing with advanced computational modeling has led to new opportunities for effectively fighting infectious diseases.

Knowing how the infectious diseases emerge and spread is the key to control them. Unfortunately, it is very difficult and even infeasible to collect important information, such as the infection and transmission rate at the beginning of the epidemic. Due to the development of Internet technology, large-scale public participation will become an important means of gathering outbreak information. Nowadays, more and more people are seeking medical help and thus sharing their health information on the Internet. Therefore, the proper mining of this type of Internet big data could help monitor the dynamics of infectious diseases. The most famous example is Google Flu Trends, which can rapidly predict influenza activity by aggregating Google search queries [5].

Y. Zou · Y. Peng · T. Jiang (✉)
College of Computer Science and Electronic Engineering,
Hunan University, Changsha 410082, China
e-mail: taijiao@moon.ibp.ac.cn

L. Deng
Suzhou Institute of System Medicine, Suzhou 215123, China

T. Jiang
National Laboratory of Biomacromolecules, Institute of
Biophysics, Chinese Academy of Sciences, Beijing 100101,
China

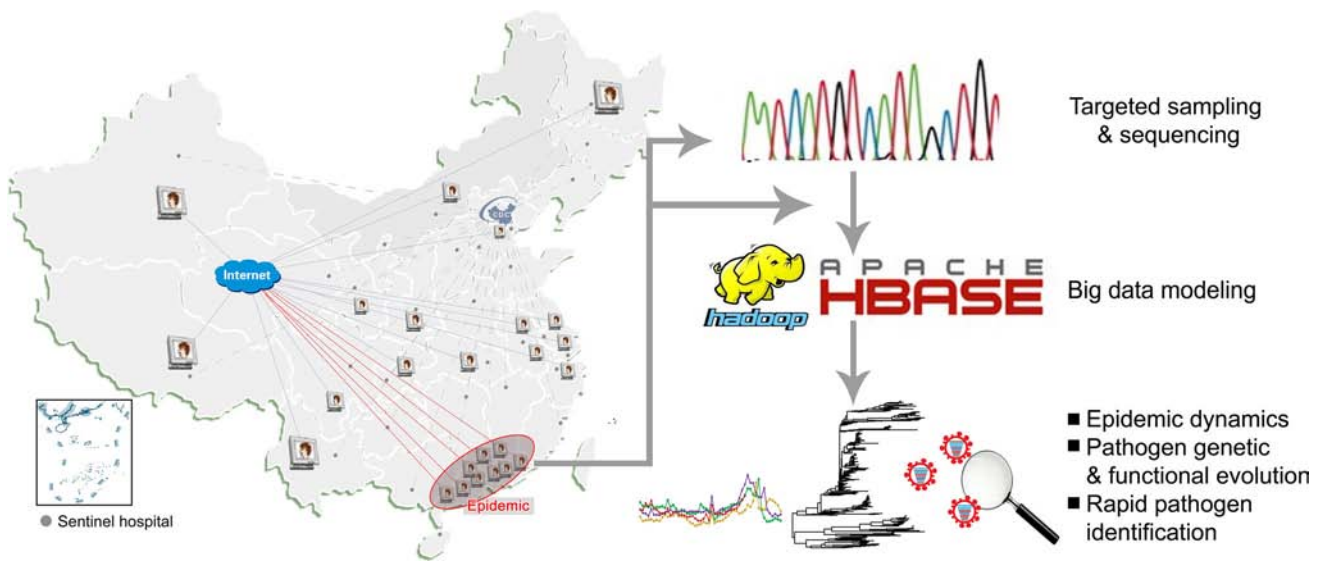


Fig. 1 Formulation of an effective surveillance strategy for infectious diseases through the coupling of Internet monitoring and focused gene sequencing

The big data aggregated from the Internet could contribute to a real-time monitoring of the activity of infectious diseases; however, they lack specific and critical information about the genetic, immunological and drug-resistance features of the newly emerging pathogens. On the other hand, the large-scale genomic and immunological characterization of emerging pathogens would not be an inconsequential investment for developing countries. Moreover, these genomic data can usually be generated on a weekly basis, typically with a reporting lag of up to 1–2 weeks. Is there a more efficient and economic way to monitor infectious diseases? Here we propose a strategy based on the coupling of Internet monitoring and focused genomic sequencing approaches. As illustrated in Fig. 1, the aggregation of big Internet data could monitor closely the activity of any suspicious infectious diseases on a national scale. Once an unusual activity has emerged, the China CDC should react immediately and collect the samples, based on Internet reporting, and further determine their genomic sequences. It is believed that with the assistance of powerful computational modeling, the

integration of big data from both the Internet and genomics will alter the paradigm of the prevention and control of infectious diseases.

References

1. Gao R, Cao B, Hu Y et al (2013) Human infection with a novel avian-origin influenza a (H7N9) virus. *N Engl J Med* 368: 1888–1897
2. Liu D, Shi W, Shi Y et al (2013) Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *Lancet* 381:1926–1932
3. Du X, Wang Z, Wu A et al (2008) Networks of genomic co-occurrence capture characteristics of human influenza a (H3N2) evolution. *Genome Res* 18:178–187
4. Du X, Dong L, Lan Y et al (2012) Mapping of H3N2 influenza antigenic evolution in china reveals a strategy for vaccine strain recommendation. *Nat Commun* 3:709
5. Ginsberg J, Mohebbi MH, Patel RS et al (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: U1012–U1014