and are key to characterizing the events that define *Y. pestis* success. Notably, these questions are all related to an even greater question: Why is the immune system defenseless overall when encountering *Y. pestis*? Any efforts towards answering this question will not only contribute to the understanding of this fascinating bacterium but also of the biology of the immune system.

[1]Department of Microbiology and Immunology, University of North Carolina, 125 Mason Farm Road, CB# 7290, University of North Carolina, Chapel Hill, NC 27599, USA
[2]Department of Genetics, University of North Carolina, 120 Mason Farm Road, Genetic Medicine Building, CB# 7264, University of North Carolina, Chapel Hill, NC 27599, USA
[‡]Current address: Department of Microbiology and Immunobiology, Harvard Medical School, 77 Avenue Louis Pasteur, NRB, Room 836, Boston, MA 02115, USA.

*Correspondence: vlmiller@med.unc.edu (V.L. Miller).

http://dx.doi.org/10.1016/j.tim.2016.01.010

References

1. Sebbane, F. *et al.* (2005) Kinetics of disease progression and host response in a rat model of bubonic plague. *Am. J. Pathol.* 166, 1427–1439
2. Gonzalez, R.J. *et al.* (2012) Bioluminescence imaging to track bacterial dissemination of *Yersinia pestis* using different routes of infection in mice. *BMC Microbiol.* 12, 147
3. Gonzalez, R.J. *et al.* (2015) Comparison of models for bubonic plague reveals unique pathogen adaptations to the dermis. *Infect. Immun.* 83, 2855–2861
4. Gonzalez, R.J. *et al.* (2015) Dissemination of a highly virulent pathogen: tracking the early events that define infection. *PLoS Pathog.* 11, e1004587
5. Shannon, J.G. *et al.* (2013) *Yersinia pestis* subverts the dermal neutrophil response in a mouse model of bubonic plague. *MBio* 4, e00170–e213
6. Bosio, C.F. *et al.* (2012) Kinetics of innate immune response to *Yersinia pestis* after intradermal infection in a mouse model. *Infect. Immun.* 80, 4034–4045
7. Pujol, C. and Bliska, J.B. (2003) The ability to replicate in macrophages is conserved between *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Infect. Immun.* 71, 5892–5899
8. Titball, R.W. *et al.* (2003) *Yersinia pestis* and plague. *Biochem. Soc. Trans.* 31, 104–107
9. Guinet, F. *et al.* (2015) Dissociation of tissue destruction and bacterial expansion during bubonic plague. *PLoS Pathog.* 11, e1005222
10. Lorange, E.A. *et al.* (2005) Poor vector competence of fleas and the evolution of hypervirulence in *Yersinia pestis*. *J. Infect. Dis.* 191, 1907–1912
11. Brubaker, R. (2006) *Yersinia pestis* and bubonic plague. *Prokaryotes* 6, 399–442
12. Shannon, J.G. *et al.* (2015) Dermal neutrophil, macrophage and dendritic cell responses to *Yersinia pestis* transmitted by fleas. *PLoS Pathog.* 11, e1004734

Science & Society

Sharing Data for Global Infectious Disease Surveillance and Outbreak Detection

Frank M. Aarestrup[1,*] and Marion G. Koopmans[2]

Rapid global sharing and comparison of epidemiological and genomic data on infectious diseases would enable more rapid and efficient global outbreak control and tracking of diseases. Several barriers for global sharing exist but, in our opinion, the presumed magnitude of the problems appears larger than they are, and solutions can be found.

Global Disease Surveillance

Globally, infectious diseases are the cause of about 22% of all human deaths [1]. In addition to the direct consequences for human health, infectious diseases also cause an increased financial burden on health systems and may imply restrictions on travel and trade. The longer it takes before the causative agents are detected, the greater the consequences for the individual patient, and – in the case of transmissible pathogens – entire populations. The recent Ebola outbreak serves as a harsh warning: a 4-month delay in diagnosis likely triggered the largest Ebola outbreak on record [2], with devastating impact on an already weak health-care system. Because many infectious diseases are international or even global, rapid global surveillance systems for exchange and comparison of information on the worldwide spread of zoonotic and human pathogens are greatly needed [3,4].

This has been recognized for many years, and combating the global burden of infectious diseases has been one of the main responsibilities for the World Health Organization (WHO); however, despite improvements in the past decades, global disease detection and surveillance remain patchy [5], and the time to diagnosis of outbreaks can be several months [6]. One practical barrier is that surveillance is organized at national or regional levels, often in an unsystematic and unstandardized way. The obvious need to look globally has resulted in several, often successful, attempts to implement a common vocabulary for specific diseases or pathogens. However, this has also led to the establishment of closed groups, isolated around a single or a few pathogens, or small groups of national reference laboratories.

Even if disease detection and surveillance capacity exists, a major hurdle is the timeliness of data sharing. This was one of the drivers for the revision of the International Health Regulations (IHR) in 2005 that specified the need for agnostic, event-based surveillance and international sharing of information to combat emerging infectious diseases [7].

Thus, despite the obvious increasing needs and benefits, as well as the global recognition, why has a real-time global surveillance and exchange of information to all not been implemented several decades ago? Here, we review some of the potential reasons and offer suggestions for the future against the background of new technologies that have tremendous potential for global surveillance.

Next Generation Sequencing (NGS), the Common Language Enabling Exchange and Comparison of Information

The decreasing costs and fast development of NGS has the potential to

completely change disease detection and surveillance capacity and organization. A huge advantage is the generation of the genomic information independently of the type of pathogen (viruses, bacteria, and parasites) which could change the focus from working in pathogen-specific silos to focusing on the disease symptoms and studying them across pathogens.

In clinical and public health microbiology, NGS-based diagnostics has started to gain ground [8,9], and global cost-efficiency of NGS is within reach for many countries, including some of those in the developing world, even though the current cost might still be a barrier. NGS can also be applied directly in metagenomic analyses, potentially advancing diagnostics even more and generating in-depth molecular surveillance data in the same run [10–12]. Currently, the most important limiting factor in many countries is the lack of access to bioinformatic expertise, especially when used as part of frontline diagnostics.

Why Are We Not Sharing Freely?

We seek to support the sharing of data prepublication, in line with the statement from the recent WHO meeting on global data sharing (http://www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/en/). However, a number of real obstacles are given in Table 1, and sharing data freely in real-time could certainly cause some problems. However, it is also our opinion that the presumed magnitude of the problems appears larger than they are in real life, and that solutions can be found. For global surveillance we do not need detailed or person-sensitive epidemiological data. To detect emerging threats and outbreaks, as well as performing valuable research studies, data on country, year, origin, and whether it is from an infection, are often sufficient. This is also the minimum set of meta-data now encouraged by the National Center for Biotechnology Information (NCBI) and the European Molecular Biology Laboratory (EMBL) European Bioinformatics Institute's European Nucleotide Archive (ENA) and recommended by international organizations such as the Global Microbial Identifier (GMI) (www.globalmicrobialidentifier.org). It should, however, also be accepted that there can be special circumstances where sharing even minimum meta-data could lead to identification of, for example, a specific food producer or a single hospital. In such special cases it may be considered whether the information might be shared without all minimum meta-data.

Special notice should be given to the many uncertainties about legal frameworks that often lead to a conservative non-sharing or an untimely sharing reaction. This is not only the case among scientists, but is also found among governments and health authorities. Clarification of international treaties such as IHR, the Nagoya Protocol on genomic resources, and EU regulations on data protection and cross-border health threats, is of utmost importance for a realistic political and professional attitude and support towards data sharing for the improvement of infectious disease control.

A specific issue is the competition for scientific output. Many people with access to NGS data are employed in positions where publications are essential for employment, research funding, and career development. The scientific competition is tough, and scientists have experienced examples where other scientists have used data generated by them without their consent. Though this is a fact, it is also our opinion that this must not hamper the exchange of information for the benefit of global health. Rather, the scientific community, with the assistance of the scientific journals and research funds, should nurture standards for responsible practice with respect to appropriate attribution and respect for formally and informally requested embargoes on non-urgent analyses of shared data.

Benefits of Sharing in Real Time

Besides the benefits associated with timely detection of emerging threats and outbreaks, as well as improved possibilities to identify vaccine targets and performing better research, there is also the benefit that the information collected will be saved for the future. When researchers retire, institutions undergo changes, or novel analytic techniques are being implemented, and the historical data are often discarded. With NGS data, the available sequence information remains useful, even if novel sequencing technologies emerge, and sharing the raw data in global repositories ensures that a backup is made in the International Nucleotide Sequence Database Collaboration (INSDC) [13].

Models for Common Sharing and Analysis of Data

It is important to realize that, with NGS data, we are talking about sharing, analyzing, and comparing very large amounts of data. Today, a typical bacterial genome yields 2–500 MB of data, while meta-genomic datasets often in GB. Research studies comparing a few hundred strains can easily be performed and handled locally, but in the future we will need to compare in real-time tens or even hundreds of thousands of genomes and between single strains and meta-genomic datasets. This will be several TB of data, and considering the current internet connections would often require several days/weeks of transfer time if the data were not immediately available. Another important debate is whether the intention is to share the complete genomic information (raw data) or pre-analyzed data. We strongly recommend sharing raw data since this will maximize reusability as new computational methodologies and reference and comparitor data become available.

The Traditional Research Approach

Genomic data have until now mainly been generated and analyzed in the research community. Here, the individual research groups have sufficient time and funding to allow for local storage and analysis of the data, and after publication the data are typically released into the global repositories. The major problem with this is the time delay and the obvious need to