Editorial

iMedPub Journals
http://www.imedpub.com/

Archives of Clinical Microbiology
ISSN 1989-8436

2016
Vol.7 No.2:e6

# Connecting the Dots between Microbial Epidemiology and Genomics

**Xuehua Wan***

Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii, USA

***Corresponding author**: Xuehua Wan, Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii, USA, Tel: +808-956-8039; E-mail: xuehua@hawaii.edu

**Received date:** March 18, 2016; **Accepted date:** March 21, 2016; **Published date:** March 31, 2016



**Keywords:** Bacteria; Epidemiology; Pathogens; Microbiota; Molecular biology

## Microbial Epidemiology and Genomics

On the Earth, bacteria may be traced up to at least 3.5 billion years ago in Archean eon, which evolved thousands of million years earlier than ancient land plants (450 million years ago, Ordovician) and land animals (Cambrian)? Besides the time axis, these tiny unicellular organisms also occupy a broad range in the space axis, living almost everywhere on the earth including air, sea water, fresh water, soil, plants, animals, humans, and extreme environments. Although some of them are our friends, some of them are our foes which can cause diseases in plants, animals, and humans. Many species and genus are unknown yet.

Compared to billion years of evolution of bacteria, the germ theory of disease was proposed only several hundred years ago and convinced by solid evidence around 1860s. As one of the fathers of modern epidemiology, Doctor John Snow investigated on the local dot map of the cholera cases near Broad Street of London and correlated the high frequency of cholera cases with the pump handle to identify the origin of the disease. His discovery not only leads to the removal of the pump handle in Soho, but also to the initiation of statistical mapping method.

Nowadays, this pioneered methodology is applied in global wide. Disease data have been collected worldwide and shared publically by the efforts of World Health Organization (WHO). Moreover, the recent groundbreaking technologies, high-throughput Whole Genome Sequencing (WGS) and comparative genome analyses, have been carried out to investigate on the historical and geographical origins of pathogenic bacteria causing diseases, e.g. Tuberculosis (TB) [1]. TB is caused by Mycobacterium tuberculosis Complex (MTBC) including various Mycobacterium species and strains. At the end of last century, it was thought that there was no big genetic variation in M. tuberculosis strains. However, using WGS method, Single Nucleotide Polymorphisms (SNPs) and large/small-scale genome arrangements were identified in MTBC, indicating the genetic diversity and evolutionary differences of the strains [2,3]. In addition, WGS technique has been applied to correlating bacterial antibiotic resistance with bacterial DNA mutations [4]. These data further provide information for strain typing and new medicine development.

With rapid progress of computational and biological sciences, we are standing at the starting point of the big data era. Numerous nucleotide data have been and will be generated and stored in international databases including National Center for Biotechnology Information (NCBI), European Nucleotide Archive (ENA), and DNA Data Bank of Japan (DDBJ). Moreover, several public databases have been built for microbial pathogens. National Microbial Pathogen Data Resource (NMPDR) provides on-line annotation service, Rapid Annotation using Subsystem Technology (RAST), and database for a series of bacterial pathogens [5]. Human Pan-Microbe Communities Database (HPMCD) provides metagenomic data to help studies in human gastrointestinal microbiota [6]. These shared databases will help international collaborations and efforts on deep learning of diversity, ecology and disease association of human microbiomes. Furthermore, with developments of text mining and image mining techniques, microbial physiology databases will also be helpful to data-driven research and biological knowledge discovery.

Genomics can be integrated into epidemiology not only as a research method but also as a monitoring method for clinical surveillance in future. Traditional surveillance protocols include culturing, microscopic examination, molecular biology diagnostics, x-ray imaging, etc. With decreasing of WGS costs and improvements of WGS protocols, WGC techniques may also be an efficient method to detect both culturable and non-culturable pathogens in animals and humans.

## References

1. Gagneux S (2012) Host-pathogen coevolution in human tuberculosis. Philos Trans R Soc B 367: 850-859.
2. Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, et al. (2014) A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun 5: 4812.
3. Ioerger TR, Koo S, No EG, Chen X, Larsen MH, et al. (2009) Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. PLoS One 4: e7778.
4. Zhang H, Li D, Zhao L, Fleming J, Lin N, et al. (2013) Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nature Genetics 45: 1255-1260.

5. McNeil LK, Reich C, Aziz RK, Bartels D, Cohoon M, et al. (2007) The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation. Nucleic Acids Res 35: D347-53.

6. Forster SC, Browne HP, Kumar N, Hunt M, Denise H, et al. (2015) HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes. Nucleic Acids Res.