

## Editorial Article

# Big Data and Microbiology

Andreas Erich Zautner

Institute of Medical Microbiology, University of Uzbekistan, Kreuzberggring, Germany

**\*Corresponding author:** Andreas E Zautner, Institute of Medical Microbiology, University of Uzbekistan Göttingen, Kreuzberggring 57, D-37075 Göttingen, Germany, Phone: +49-551-395927, E-mail: azautne@gwdg.de

**Citation:** Andreas EZ (2017) Big Data and Microbiology. J Microbiol Genet 2017: J113

**Received Date:** 16 January, 2017; **Accepted Date:** 20 January, 2017; **Published Date:** 27 January, 2017

## Editorial

Computer, internet and telecommunication have significantly changed our daily life. Big Data - starting with the shift from analog data storage to digital data storage - will be the next big change that likewise will influence the work in the field of microbiology significantly. While the founders of modern microbiology about 100 years ago have cultivated pathogens and infected experimental animals to meet the 'Henle Koch postulates', today we are confronted with techniques that produce a vast amount of data. In particular, two specific techniques enter the playing field. The first is Next Generation whole genome Sequencing (NGS), the second quantitative mass spectrometry.

After the technique of pyrosequencing as a pioneer approach has made the break through for NGS in microbiology, many new sequencing methods have been developed. In particular sequencing by synthesis (Illumina), which is relatively inexpensive, and single-molecule real-time sequencing (SMRT-Sequencing; Pacific Biosciences), which mostly delivers completely closed genomes, are now well-established in microbiology. Both techniques provide Gigabytes of data and with these raw data a huge amount of information, which is to extract and visualize by the microbiologist - that turns him or her more or less into a bioinformatician.

New NGS-based phylogenetic typing procedures like core genome or whole genome multi-locus sequence typing (cg/wg MLST) provide much higher discriminatory power and there with well-established typing methods like Pulsed-Field Gel Electrophoresis (PFGE), ribotyping, Multiple Loci VNTR Analysis (MLVA), serotyping, seven - gene Multi - Locus Sequence Typing (MLST) and some other methods loose their role as gold standard in their particular field. Only a few years ago an out break situation was confirmed by presenting e.g. identical PFGE pattern, today 'Call SNPs & Infer Phylogeny' analysis often give an all-clear where it has not been technically possible recently. However, still there is no agreement on how many single nucleotide polymorphisms in one genome are likely sequence errors or real mutations that are a

result of micro evolution. Additionally NGS data provide information about virulence factors, resistance genes, plasmids, phages, restriction modification systems and toxins. Thus it is feasible to predict essential aspects of the phenotype. There are already approaches to determine the ribotype from the genome sequence. Additionally SMRT sequencing (Pacific Biosciences) provides excellent information about the microbial methylome, which opens new doors to study epigenetic regulation of genes and DNA uptake.

While whole genome sequencing gives an encompassing overview of the potential of a microorganism; quantitative mass spectrometric techniques provide comprehensive data on the proteomic changes in response to specific stressors. The most important techniques in this field are 'Stable Isotope Labeling by Amino acids in Cell culture' (SILAC) based on the detection of labelled proteins using non-radioactive isotopes, and label-free 'Sequential Window Acquisition of all Theoretical Mass Spectra' (SWATH-MS) that combines the shotgun proteomics strategy, in which proteolytic peptides are analyzed, with data independent acquisition connected with peptide spectral library match. Both, techniques provide qualitative and quantitative information about several hundreds of protein species that can be detected in a series of experimental runs. Furthermore, the data of different experimental approaches e.g. varying intensities or kinds of stressors can be compared. This very complex information provides a deeper understanding not only of the regulatory networks of the particular micro-organism but also of the host cells optionally included in an experimental setting.

Transcriptome analysis is biologically in-between whole genome sequencing and quantitative mass spectrometry and could have been mentioned as a third big data providing technology, but I think this technique has lost much of its importance due to the quantitative mass spectrometric techniques. Otherwise, RNA sequencing should be seen as a sub-discipline of NGS. The sum of all messenger RNA molecules gives a better conception of the biological processes than the genome sequencing alone especially if in addition to the molecular identities each particular RNA is

quantified.

As a clinical microbiologist, I would like to add one further big data providing system here, the clinical information obtained from the electronic health record. It completes the picture of host tissue tropism, antibiotic resistance, virulence and microevolution with data about therapeutic efficiency, pathogenesis, pharmacokinetics and final outcome.

All together all these techniques are significantly changing

the day-to-day routine of the microbiologist. To obtain the raw data, to educe the crucial information out of it and sometimes to visualize them to present them to colleagues from a different subject area requires some more computer skills than years before. Previously, the microbiologist had to enlarge small things to make them understandable - today he has to reduce large amounts of data to a small comprehensible measure.