American Journal of Infection Control 45 (2017) 170-9

Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

State of the Science Review

Infection control in the new age of genomic epidemiology

Patrick Tang MD, PhD [a,*], Matthew A. Croxen PhD [b], Mohammad R. Hasan PhD [a,c], William W.L. Hsiao PhD [b,d], Linda M. Hoang MD [b,d]

[a] Department of Pathology, Sidra Medical and Research Center, Doha, Qatar
[b] British Columbia Centre for Disease Control Public Health Laboratory, Vancouver, BC, Canada
[c] Department of Pathology and Laboratory Medicine, Weill Cornell Medical College in Qatar, Doha, Qatar
[d] Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

Key Words:
Whole-genome sequencing
Next-generation sequencing
Genomic epidemiology
Outbreak
Genotyping
Molecular epidemiology

With the growing importance of infectious diseases in health care and communicable disease outbreaks garnering increasing attention, new technologies are playing a greater role in helping us prevent health care–associated infections and provide optimal public health. The microbiology laboratory has always played a large role in infection control by providing tools to identify, characterize, and track pathogens. Recently, advances in DNA sequencing technology have ushered in a new era of genomic epidemiology, where traditional molecular diagnostics and genotyping methods are being enhanced and even replaced by genomics-based methods to aid epidemiologic investigations of communicable diseases. The ability to analyze and compare entire pathogen genomes has allowed for unprecedented resolution into how and why infectious diseases spread. As these genomics-based methods continue to improve in speed, cost, and accuracy, they will be increasingly used to inform and guide infection control and public health practices.

Infectious diseases continue to be one of the leading causes of death worldwide.[1] In contrast with noncommunicable diseases, infectious pathogens can often evolve and spread rapidly, leading to the emergence of novel human pathogens, more virulent forms of existing pathogens, and antibiotic resistance organisms.[2,3] Our infection control practices must continue to evolve and improve in order to combat the ever changing threat of infectious diseases in the health care and population settings.[4,5] Over the last few decades, the new methods in the microbiology laboratory, such as molecular diagnostics, have become central in allowing for the rapid detection and characterization of pathogens.[6,7] The ability to genotype pathogens using various molecular techniques has led to the rise of molecular epidemiology, which has enhanced our ability to detect outbreaks and uncover the events and factors involved in transmission.[8]

Through genotyping, a molecular fingerprint of each pathogen isolate can be generated and compared with others in the same suspected cluster. Isolates with the same or highly similar genotypes, and linked by epidemiologic data, are likely to represent related cases of the infectious disease within a cluster or outbreak. The laboratory methods for genotyping range from determining single nucleotide variants (SNVs) to evaluating repetitive regions of the genome to examining the length of various fragments of the genome. The earlier techniques for fingerprinting infectious agents mostly involved the electrophoretic separation of DNA or RNA fragments, such as examining plasmids, electrophoresis of ribosomal RNA, or restriction fragment length polymorphism (RFLP).[7] Polymerase chain reaction (PCR) can also be used to create fragments of the pathogen genome for analysis through techniques such as randomly amplified polymorphic DNA, arbitrarily primed PCR, or amplified fragment length polymorphism. Other methods amplify repetitive regions of the genome to determine the number of repeats within each region, such as multiple-locus variable number of tandem repeats analysis. Finally, the sequence composition of portions of the genome can be examined by hybridization methods (eg, with DNA probes or microarrays), SNV typing, or DNA sequencing of specific genes (multilocus sequence typing [MLST]).

All of these laboratory genotyping methods support epidemiologic investigations by determining whether any 2 isolates belong to the same strain and are therefore potentially epidemiologically related. The resolution at which these methods can distinguish different strains of a microorganism depend on the species being

* Address correspondence to Patrick Tang, MD, PhD, Department of Pathology, Sidra Medical and Research Center, PO Box 26999, Doha, Qatar.
E-mail address: ptang@sidra.org (P. Tang).
Conflicts of Interest: None to report.

0196-6553/
http://dx.doi.org/10.1016/j.ajic.2016.05.015

examined, the percentage of the genome being interrogated, and the variability of those regions. These molecular fingerprinting techniques are useful for confirming clusters or outbreaks of an infectious disease. Although the identification of a cluster of epidemiologically and genetically related cases can be enough to intervene or implement measures to prevent future occurrences, additional details about the outbreak, such as the phylogenetic origins and pathogenic potential of the etiologic agent, and delineation of individual transmission events can improve our understanding of the outbreak and therefore better focus interventions and infection prevention measures.

With the recent introduction of massively parallel DNA sequencing technology or next-generation sequencing (NGS), the real-time sequencing of entire pathogen genomes is now possible.[9,10] In contrast with genotyping, where only a small fraction of the pathogen genome is used to infer phylogenetic relationships, the whole genome of the pathogen can be used to resolve the transmission dynamics of an outbreak in much greater detail.[11] Current second-generation sequencing platforms involve fragmentation of the genomic DNA, amplification and tagging of the fragments, followed by massively parallel sequencing by synthesis during which millions to billions of DNA fragments are sequenced in parallel.[12] The many sequence reads are then assembled back into a genome using various bioinformatics algorithms.[13] The multiple pathogen genomes within a cluster of related cases can then be examined for their virulence potential and compared with one another to determine their relatedness at the genomic level. Because the genomes of isolates within a cluster are expected to be highly related, the most common way to compare these genomes is to examine the differences in the SNVs between isolates (when SNVs become fixed within a population they are called SNPs [single nucleotide polymorphisms]).[14] These new sequencing technologies along with the associated bioinformatics algorithms have given rise to the field of genomic epidemiology where whole-genome analysis methods are integrated with epidemiologic investigations to yield the ultimate resolution into communicable disease outbreaks.[15] Armed with knowledge of the entire genome, we can not only better identify and characterize the pathogen responsible for an outbreak so we can better estimate its risk and select the most appropriate interventions, but we can also begin to understand the origins and dynamics of the outbreak.

Genomic epidemiology has been applied to many outbreaks in the last few years and is becoming a widely accepted method to investigate outbreaks.[16] Most outbreaks have been examined retrospectively because the sequencing costs have been relatively high, the time required to prepare the samples and generate the sequences has been lengthy, and the complexity of the bioinformatics analysis poses a challenging problem. However, with improvements in sequencing technology and continuing optimization and standardization of bioinformatics algorithms, genomic epidemiology investigations can now be conducted during the course of an ongoing outbreak to provide real-time guidance for infection control and prevention interventions.[9,17] Genomic epidemiology has played a large role in furthering our understanding of outbreaks in health care settings (health care–associated infections [HAIs]) and larger communicable disease outbreaks at the community, national, and international scales. In this article, we review the evolution from molecular to genomic epidemiology and the technologic and computational advances which have enabled this, and we also discuss the remaining challenges that need to be addressed for the wider adoption of genomics in infection control.

MOLECULAR EPIDEMIOLOGY

Molecular epidemiology is a relatively new discipline that refers to the use of molecular methods in the microbiology laboratory, coupled with conventional epidemiologic tools, to identify potentially linked cases and aid in the investigation of outbreaks. Prior to the 1970s, the laboratory tools for detecting clusters of infectious disease cases were limited to the identification of the causative agent at the genus and species level along with common antigenic features or antimicrobial resistance patterns. Although these traditional methods still play a role in identifying clusters, they have since been replaced by nucleic acid–based methods, which have greatly improved our ability to identify epidemiologically related cases of infectious diseases. Microevolution occurs within all microorganisms through point mutations, genetic rearrangements, or horizontal gene transfer, leading to various degrees of genetic diversity within each species. By finding and measuring those genetic changes that occur at a fixed rate, we are then able to follow the microevolution of an organism over time. Because the timescale of an outbreak is exceedingly short compared with a pathogen's evolutionary timescale, one fundamental assumption is that the degree of genetic diversity of a given pathogen within an outbreak will be less than the genetic diversity within the overall pathogen population. Molecular genotyping methods that are able to reveal these differences in genetic diversity have since become the foundation of molecular epidemiology.

Molecular methods are used to genotype or generate molecular fingerprints of the individual pathogens in a suspected cluster of cases. The choice of methods is dependent on many factors, including the species of the pathogen, the availability of comparative genotypes, and the epidemiologic context within which the method can be applied. Isolates of different pathogen species can be highly related or highly diverse at the genomic level, and the genotyping method must be able to examine the relevant genomic regions to detect the diversity within the given species. Species, which tend to be clonal within the context of an outbreak, will require genotyping methods that target regions of the genome that are more variable but yet exhibit rate of change that is relevant to the timescale of the outbreak. Difficulties can arise when an organism's genome is too monomorphic, in which case, it may be difficult or even impossible to find a genotyping method that is able to show differences between epidemiologically unrelated isolates of that species. Conversely, when a pathogen species is panmictic, with mutations arising through genetic rearrangements or horizontal gene transfer, it may be difficult to interpret genotyping results because epidemiologically related isolates may show different genotypes and even isolates from the same patient may not have identical genotypes. In this situation, it will be necessary to identify genotyping methods which are less affected by these types of mutations or else combine multiple methods to obtain a more complete understanding of this diversity.

The availability of reference genotypes and reference databases, generated with the same genotyping method, is essential for the interpretation of the genotypes generated from a suspected cluster, favoring the use of methods which are already more widely adopted for a given pathogen. The genotyping method must also consider the type of outbreak and the epidemiologic questions being asked: different methods may be needed to characterize an outbreak of a more ubiquitous pathogen versus a rare pathogen, and different methods may be needed to confirm an outbreak versus surveillance for additional related cases. In any case, the tool must be able to clearly demonstrate relatedness between epidemiologically related cases while differentiating unrelated cases. Ultimately, for a genotyping method to be widely accepted and adopted, it needs to be relatively inexpensive, have acceptable turn-around time, be highly reproducible even in different laboratories, and be easy to perform and interpret.[18,19]

There are numerous genotyping methods for molecular epidemiology, and no single method is universally applicable for all