Editorial

Progress in pathogen detection by whole-genome sequencing

Volume 3 Issue 2 - 2016

Chung F Wong
Department of Chemistry and Biochemistry, University of Missouri-St. Louis, USA

Correspondence: Chung F Wong, Department of Chemistry and Biochemistry, University of Missouri-St. Louis, Tel 3145165318, Email wongch@umsl.edu

Received: January 28, 2016 | **Published:** February 01, 2016

Editorial

Methods such as staining/microscopic examination and culturing have served us well for a long time and are still commonly used today for pathogen detection. However, these methods mostly focus on looking for one pathogen at a time and are not effective when a large number of potential pathogens need to be considered. They fall short when the proper diagnosis of a seriously ill patient needs to be achieved quickly to decide treatment options. They are also ineffective when it is necessary to identify the causative agent of a disease outbreak quickly to suggest strategies to control the outbreak. Identifying the right causative bacterial pathogen early also allows the proper antibiotics to be administered at the outset, reducing the chance of developing antibiotic resistance, an important consideration in modern antibiotic stewardship. Molecular methods have opened up new opportunities. They can detect pathogens that are not cultivatable and many can be performed quickly. Although many methods still focus on the analysis of single pathogens, encouraging progresses have been made in the development of techniques that can detect a large number of pathogens in a single assay.

Microarrays provide one example. Arrays such as the Virochip[1] uses tens of thousands of short sequences of DNA to detect a large number of viruses. A drawback of this approach is that a microarray misses pathogens that it is not designed to look for. It also uses only part of the genome of a microbe for detection - using the whole-genome sequence of a microbe could improve sensitivity. To this end, whole-genome sequencing provides an attractive solution. With high-throughput next-generation sequencing technology, this technique can potentially detect many pathogens from a minimally processed metagenomic sample in a single assay. Although this approach has not yet been widely examined, encouraging results are emerging. In 2008, Nakamura and colleagues[2] demonstrated the feasibility of using this technique to detect Campylobacter in the stool sample of a patient. This bacterium was missed by several other traditional techniques that they used before. This approach did not assume what pathogens might be present in a sample. Instead, it unbiasedly searched a large database containing many pathogens for sequences that matched with the sample.

Later, Loman et al.,[3] applied next-generation sequencing to analyze the stool samples of patients during the outbreak of Shiga-toxigenic E. coli O104:H4 in Germany in 2011. With the benchtop-sequencing platform Illumina MiSeq, they were able to obtain good coverage of the genome of this outbreak strain from the metagenomic stool samples. More recently, Chiu and coworkers have sped up the use of metagenomics to detect pathogens by developing the SURPI program.[4] This program brings this approach closer to practical uses by significantly reducing the computational time required for diagnosis, which requires mapping many short DNA fragments from next-generation sequencing experiments to large databases of microbes. Chiu and co-workers have demonstrated the utility of this technique in identifying the causative agents of several illnesses, including neuroinvasive astrovirus infection,[5] Balamuthia mandrillaris encephalitis,[6] and neuroleptospirosis.[7]

The impact of this technique should increase further with the introduction of faster and portable sequencers such as the MinIon being developed by Oxford Nanopore TechnologiesTM. Greninger et al.,[8] have already demonstrated that this platform can produce useful diagnostic results in even shorter time. Because this sequencer is small and portable that can be plugged into a laptop computer, it can potentially be used in remote areas without sending the data to remote high-performance computing facilities for processing if fast computer programs for diagnosis can be run directly on the laptop computer. Gontarz & Wong[9] made a step in this direction by using several strategies:

I. Focus on checking microbes that are pathogenic such as those in the PATRIC database.[10]

II. Develop compact genome representations that can fit into the smaller RAMs of portable computers.

III. Develop short-read aligners such as SRmapper[11] that requires less computer memory.

Thoughtful developments of compact genome representations have not only made it easier to check many pathogens in computers with smaller memories but can also improve the sensitivity of detection. Although it is easier to pick out a pathogen from sequencing experiments of metagenomic samples when its genome is covered to a large extent by sequencing reads, it is harder when the microbial load or the sequencing depth is low. Using marker genes, such as from MetaRef,[12] is not effective as these genes are not necessarily covered when only a small number of reads from a microbial genome is produced in a next-generation-sequencing experiment. Using the whole-genome sequence of a microbe improves the odds that reads originating from the microbe can be detected. However, some parts of the genomes are more useful for detection than the others. In particular, the parts that do not overlap with the genomes of other organisms known to be present in a specific type of samples-e.g., saliva - are most useful as the presence of a microbe can be deduced even when only a small number of reads are aligned to these regions. Using only the unique regions of the genomes of microbes could improve signal/noise (S/N) ratio significantly. Consider an example when 50% of a reference genome is known to overlap with other known species, one may estimate the S/N ratio by 100%/50%=2. However, removing the overlapping regions gives a S/N ratio of 50%/ (A number close to zero), which could be several orders of magnitudes larger than two.

Gontarz and Wong[9] showed that the presence of Mycobacterium tuberculosis and bacteria in the Human Oral Microbiome Database from metagenomic samples could be deduced when only 0.2% of their unique genomes were covered. It will be interesting to see whether such a compact representation of genome can be developed for most pathogens in the PATRIC database. Although practical tools for medical diagnostics require stringent validation for approval by the US Food and Drug Administration, the outlook is bright for using whole-genome sequencing of metagenomic samples for this purpose.

Acknowledgements

Support by a Research Award from the University of Missouri-St. Louis is appreciated. The author also thanks Dr. Paul Gontarz for performing part of the work described in this Editorial when he was a graduate student.

Conflict of interest

The author declares no conflict of interest.

References

1. Chen EC, Miller SA, DeRisi JL, et al. Using a pan–viral microarray assay (Virochip) to screen clinical samples for viral pathogens. *J Vis Exp*. 2011;50:e2536.
2. Nakamura S, Maeda N, Miron IM, et al. Metagenomic diagnosis of bacterial infections. *Emerg Infect Dis*. 2008;14(11):1784–1786.
3. Loman NJ, Constantinidou C, Christner M, et al. A culture–independent sequence–based metagenomics approach to the investigation of an outbreak of shiga–toxigenic Escherichia coli O104:H4. *JAMA*. 2013;309(14):1502–1510.
4. Naccache SN, Federman S, Veeraraghavan N, et al. A cloud–compatible bioinformatics pipeline for ultrarapid pathogen identification from next–generation sequencing of clinical samples. *Genome Res*. 2014;24(7):1180–1192.
5. Naccache SN, Peggs KS, Mattes FM, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next–generation sequencing. *Clin Infect Dis*. 2015;60(6):919–923.
6. Greninger AL, Messacar K, Dunnebacke T, et al. Clinical metagenomic identification of Balamuthia mandrillaris encephalitis and assembly of the draft genome:The continuing case for reference genome sequencing. *Genome Med*. 2015;7(1):113.
7. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next–generation sequencing. *N Engl J Med*. 2014;370(25):2408–2417.
8. Greninger AL, Naccache SN, Federman S, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real–time nanopore sequencing analysis. *Genome Med*. 2015;7(1):99.
9. Gontarz PM. *Fast and Sensitive Genome–Hashing Software and its Applicaiton in Using NGS as a Detection Agent for Bacterial Presence in Oral Metagenomic Samples*. St. Louis: Department of Chemistry and Biochemistry, University of Missouri; 2015. 198 p.
10. Wattam AR, Abraham D, Dalay O, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*. 2014;42:D581–D591.
11. Gontarz PM, Berger J, Wong CF. SRmapper: A fast and sensitive genome–hashing alignment tool. *Bioinformatics*. 2013;29(3):316–321.
12. Huang K, Brady A, Mahurkar A, et al. MetaRef: A pan–genomic database for comparative and community microbial genomics. *Nucleic Acids Res*. 2014;42:D617–D624.