Published in final edited form as:
Curr Opin Crit Care. 2013 October ; 19(5): 432–439. doi:10.1097/MCC.0b013e3283636b8c.

Selected Insights from Application of Whole Genome Sequencing for Outbreak Investigations

Vien Thi Minh Le[1] and **Binh An Diep**[1,2,*]

[1]Division of Infectious Diseases, Department of Medicine, University of California, San Francisco, CA, USA

[2]Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, CA, USA

Abstract

Purpose of review—The advent of high-throughput whole genome sequencing has the potential to revolutionize the conduct of outbreak investigation. Because of its ultimate pathogen strain resolution, whole genome sequencing could augment traditional epidemiologic investigations of infectious disease outbreaks.

Recent findings—The combination of whole genome sequencing and intensive epidemiologic analysis provided new insights on the sources and transmission dynamics of large-scale epidemics caused by *Escherichia coli* and *Vibrio cholerae*, nosocomial outbreaks caused by methicillin-resistant *Staphylococcus aureus*, *Klebsiella pneumonia*, and *Mycobacterium abscessus*, community-centered outbreaks caused by *Mycobacterium tuberculosis*, and natural disaster-associated outbreak caused by environmentally acquired molds.

Summary—When combined with traditional epidemiologic investigation, whole genome sequencing has proven useful for elucidating sources and transmission dynamics of disease outbreaks. Development of a fully automated bioinformatics pipeline for analysis of whole genome sequence data is much needed to make this powerful tool more widely accessible.

Keywords

whole genome sequencing; outbreak investigation; infection control; nosocomial infection; foodborne outbreaks

Introduction

Integration of conventional molecular strain typing methods with epidemiologic investigation have greatly expanded our understanding of the source and mode of transmission of pathogens and enhanced our ability to control and prevent disease outbreaks. In recent years rapid technological advances made possible the use of high-throughput whole genome sequencing—which provides the ultimate in pathogen strain resolution—for

*Correspondence to: Binh An Diep, PhD Assistant Professor 1001 Potrero Ave, Building 30, Room 3318 San Francisco CA 94110 USA Phone: (408) 410-3897 Fax: (415) 648-8425 bdiep@medsfgh.ucsf.edu.

Authors declare no conflicts of interest.

outbreak investigations. This review provides selected insights learned from the application of whole genome sequencing for investigation of outbreaks caused by major human pathogens (Table).

Foodborne disease outbreaks

Foodborne disease outbreaks are caused by many different species of bacteria, virus and parasites. The goals of the food-borne outbreak investigation are to identify the pathogen that caused the outbreak, identify the source of the infection and, most importantly, control and prevent further spread of the infection. Whole genome sequencing has been applied recently for foodborne outbreak investigations [1-5, 15-17].

From May through July 2011 two major foodborne outbreaks caused by *E. coli* of serotype O104:H4, which had not previously been associated with large-scale outbreaks, occurred in Germany and France. The German outbreak involved about 4000 cases of bloody diarrhea, 850 case of hemolytic uremic syndrome and 50 deaths, whereas the smaller French outbreak involved 15 cases of bloody diarrhea, 9 cases of hemolytic uremic syndrome. Traditional epidemiologic investigation revealed that the source of this international outbreak was linked to a 15,000-kg contaminated sprout seed shipment from Egypt. Conventional strain typing methods, including antimicrobial susceptibility testing, serotyping, virulence gene content typing, rep-PCR, multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), confirmed that outbreak isolates from Germany and France are identical. What follows were the worldwide efforts to determine the whole genome sequence of representative isolates from German and French outbreaks [1-3*]. An open-source genomics approach was developed, which allowed for the high-throughput sequencing and immediate public release of the whole genome sequence of a German outbreak isolate within days after the DNA sample was received, and the rapid crowd-sourcing of bioinformatics analyses that provided within weeks a detailed understanding of the evolutionary origins and molecular basis for the emergence of the new pathogenic outbreak strain [1].

The genome sequence of the *E. coli* O104:H4 outbreak strain revealed the following features that may contribute to its ability to cause unusually severe disease outbreaks [1-3*]: (i) prophage encoding Shiga toxin, a toxin that inhibits protein synthesis and causes the severe hemolytic uremic syndrome [18], (ii) a rare combination of virulence factors, termed serine protease autotransporters of Enterobacteriaceae (SPATEs), that has been implicated in mucosal damage and colonization [19], (iii) an intermediate-size plasmid encoding several enteroaggregative *E. coli*-specific virulence factors, and (iv) a large plasmid encoding extended-spectrum beta-lactamase. Although *E. coli* O104:H4 outbreak strain produces Shiga toxin, which is commonly found in enterohemorrhagic *E. coli*, its whole genome sequence revealed unequivocally that it was an enteroaggregative *E. coli* strain. The outbreak strain formed a distinct clade with strains of Shiga toxin-negative enteroaggregative *E. coli* O104:H4 from African countries [20]. This suggests that horizontal acquisition of the Shiga toxin-encoding phage may be a relatively recent event, which then confers upon outbreak strain of enteroaggregative *E. coli* O104:H4 an enhanced ability to cause unusually severe epidemic disease.

Enterohemorrhagic *E. coli* O157:H7 is a major cause of foodborne outbreaks [15]. A recent enterohemorrhagic *E. coli* outbreak occurred in the United Kingdom, involving 93 infected persons, with 22% of these individuals developed hemolytic-uremic syndrome [4**]. Whole genome sequencing was performed on 8 isolates derived from human patients and eight from animals on a farm suspected to be the source of the outbreak. A total of 4 SNPs were identified from the whole genome sequence data, which allowed for development of a rapid SNP assay for characterization of the remaining 106 isolates from the outbreak. The combination of whole genome sequencing data, SNP typing data, and classic epidemiologic investigation showed that contamination with the outbreak strain occurred all over the farm, which was the likely source for human cases.

Vibrio cholerae causes large-scale disease epidemics through contaminated water [21]. Although cholera has not caused outbreaks in Haiti until recently, a severe epidemic of cholera occurred in 2010, involving all 10 Haitian provinces and a neighboring country. 93000 persons were infected and, of those, 2100 have died [5]. To determine the origins of the outbreak strain, whole genome sequencing was performed on 5 isolates, including 2 clinical *V. cholerae* isolates from the current outbreak, 1 strain caused cholera in Latin America in 1991, and 2 strains isolated in South Asia in 2002 and 2008. These 5 genome sequences were compared to one another and also with a set of previously obtained partial genomic sequences of 23 diverse strains of *V. cholera*. The outbreak *V. cholerae* strain was shown to be nearly identical to so-called variant seventh-pandemic El Tor O1 strains widespread in South Asia and unrelated to *V. cholera* strains circulating Latin America and East Africa. The Haitian cholera epidemics likely began with human introduction of a *V. cholerae* strain into Haiti from a distant geographic source.

Methicillin-resistant *Staphylococcus aureus* (MRSA)

Since its first emergence in 1961, MRSA has become endemic in hospitals worldwide. Among hundreds of pathogenic species of bacteria, viruses, fungi and parasites, MRSA has the distinction of being the most frequent cause of nosocomial outbreaks [22]. Although nosocomial outbreaks, defined as an increase in the occurrence of infection above the background rate, may account for <10% of all nosocomial infections [22], it is often the focus of most infection control efforts. Over the past 15 years, MRSA has also emerged within community settings, causing disease outbreaks among otherwise healthy persons with no antecedent healthcare exposure [23]. MRSA outbreak investigation is often triggered by a laboratory-based surveillance program that identifies an increase in prevalence of MRSA infections, or a cluster of MRSA-infected patients, or a new MRSA strain with an atypical susceptibility profile to a panel of antimicrobials (antibiogram). Molecular genotyping methods are used as an adjunct to traditional epidemiologic investigation to assess whether the outbreak was caused by a single clone, or multiple clones. Because MRSA has a clonal population structure in which a limited number of bacterial genotypes causes the vast majority of infection [24], conventional molecular genotyping methods, e.g. PFGE and MLST, may not have sufficient discriminatory power to differentiate between clones causing outbreak disease from those causing endemic disease. Several reports demonstrated the usefulness of data obtained from whole genome sequencing to assist in classic epidemiologic investigation of MRSA outbreaks [6**, 7**].

To validate the use whole genome sequencing for MRSA outbreak investigation, Koser et al. used a benchtop sequencer to characterize outbreak and non-outbreak isolates recovered during an epidemiological investigation of MRSA infections in a neonatal intensive care unit (NICU) at the Rosie Hospital in Cambridge, United Kingdom [6**]. Outbreak isolates were differentiated from non-outbreak isolates by having a distinctive antibiogram, which provided a basis for epidemiologic characterization of transmission events within NICU. Whole genome sequencing confirmed that epidemiologically-linked NICU outbreak isolates were clonally related to one another and that they are more distantly related to non-outbreak isolates prevalent at this hospital. Conventional molecular typing method, i.e. MLST, would have failed to differentiate NICU outbreak isolates and non-outbreak isolates because they belonged to the same sequence type 22, a strain type that is endemic in hospitals across the United Kingdom. Whole genome sequencing also identified a missed transmission event outside of the NICU between two patients in adjacent beds who had bacteremia caused by a sequence type 1 MRSA [6**]. It should be noted that whole genome sequencing did not have any impact on this outbreak investigation because it was done retrospectively.

In another MRSA outbreak investigation conducted at the Rosie Hospital, whole genome sequencing was used in real-time to assist in the epidemiologic investigation of three small clusters of infected babies in the special care baby unit (SCBU) over a 6-month period [7**]. Although the SCBU isolates shared a distinctive antibiogram, the three clusters were interspersed with gaps of 17 days and 33 days, during which no MRSA cases were detected. Because of the temporal gaps, the infection control team was not able to determine whether the three clusters were part of a single clonal outbreak. Retrospective application of whole genome sequencing showed that SBCU outbreak isolates with the distinctive antibiogram were indeed clonally related, belonging to a new sequence type 2371, which differed from the sequence type 22 MRSA strain at only a single MLST gene loci. Whole genome sequencing was then integrated into a prospective longitudinal surveillance and epidemiologic investigation program, which then allowed for identification of additional cases infected the same outbreak strain and a previously undetected transmission network involving infants, mothers, and a health care worker—the latter was likely the source of the protracted outbreak [7**]. In this instance, real-time application of whole genome sequencing had a major impact on control of this MRSA outbreak.

MRSA is known historically for its capacity to cause large-scale disease epidemics across international borders. Whole genome sequencing was recently used to illustrate the spread of the TW20 MRSA clone from Asia, where this clone is prevalent, to the United Kingdom where it caused a large 2-year outbreak in London [8]. TW20, a recently emerged subclone of the sequence type 239, was found to harbor a ϕSPβ-like prophage encoding a novel virulence determinant later termed SasX, for *Staphylococcus aureus* surface protein X [25]. Molecular pathogenesis studies showed that SasX plays a major role in enhancing biofilm formation as well as nasal colonization, lung disease and abscess formation in animal models [26]. This indicates that whole genome sequencing not only allowed for enhancement of outbreak investigations but the actual genomic data can be used to develop a better understanding of the molecular basis for the emergence of new pathogenic MRSA strain.

Carbapenem-resistant *Klebsiella pneumoniae*

A nosocomial outbreak of carbapenem-resistant *Klebsiella pneumoniae* occurred recently at the United States National Institutes of Health Clinical Center, involving 18 patients, with 6 deaths attributable to this gram-negative bacteria [13**]. Like MRSA, *K. pneumoniae* has a clonal population structure that makes it difficult to gain useful data from conventional genotyping methods, e.g. PFGE and MLST, for outbreak investigation. All carbapenem-resistant *K. pneumoniae* isolates belonged to sequence type 258 lineage, which is ubiquitous in U.S. hospitals [27]. Multiple isolates colonizing different body sites of the putative index patients were compared retrospectively to isolates from the other 17 patients by whole genome sequencing to help elucidate transmission events. Integrated bacterial SNPs analysis and epidemiologic data revealed three independent transmission events from isolates colonizing the different body sites of the index patient to the other cases, as well as the identification of an unexpected transmission route involving contaminated medical equipment.

Mycobacterium tuberculosis

A critical component of *M. tuberculosis* disease control and prevention is contact tracing, which allow for identification and diagnosis of persons who may have come into contact with an infected person. Two conventional molecular typing methods, restriction-fragment-length polymorphism analysis based on insertion sequence 6110 (RFLP-IS6110) and 24-loci mycobacterial interspersed repetitive unit -variable number tandem repeats (MIRU-VNTR), have been used to either refute or confirm transmission links between patients with tuberculosis. If the infecting isolates exhibit different RFLP-IS6110 or MIRUVNTR genotypes, then epidemiologic links between the infected patients can be refuted. However, if the infecting isolates are of the same RFLP-IS6110 or MIRU-VNTR genotypes, then epidemiologic data are still needed to confirm a transmission event. However, matched RFLP-IS6110 or MIRU-VNTR genotypes may trigger long and costly contact tracing efforts that may not result in identification of plausible transmission events because these methods may cluster distantly related isolates. Whole genome sequencing has the potential to clearly resolve outbreak-associated isolates from non-outbreak isolates.

A 3-year tuberculosis outbreak occurred in British Columbia, Canada, involving mostly crack cocaine users [9]. The outbreak isolates had identical MIRU-VNTR genotype, but intensive contact tracing and social-network analysis failed to identify a source. Whole genome sequencing was then used in a retrospective analysis that allowed for differentiation of the outbreak isolates into two distinct genetic lineages of *M. tuberculosis* that caused two concomitant outbreaks. Integration of whole genome sequencing data with the epidemiological data allowed for construction of a more accurate transmission network and identification of “superspreader” cases who were most likely responsible for infecting multiple other cases. It is of interest to note that outbreak isolates from the same genetic lineage, including those with direct epidemiological links, were separated by at least 18 single nucleotide polymorphisms (SNPs) as determined by whole genome sequencing [9], which seemed high considering the mutation rate for *M. tuberculosis* in a cynomolgus macaque model has been estimated at 0.5 SNPs per genome per year [28]. However, many

of these SNPs were found in repetitive DNA regions that are often difficult to map and analyze because of the large volume of short reads generated by next-generation sequencing technology.

Two recent whole genome sequencing studies of *M. tuberculosis* outbreaks aimed at determining the type of epidemiological inferences that can be drawn from SNPs data [10**, 11*]. A retrospective outbreak investigation from the United Kingdom Midlands found no more than 5 SNPs separating isolates from 69 epidemiologically linked patients [10**]. A rate of change of 0.5 SNPs per genome per year was estimated from longitudinal isolates from 30 individuals and 25 households [10**], which is consistent with the rate of SNPs accumulation in macque infection model [28]. Another retrospective outbreak investigation from Hamburg, Germany, revealed a maximum of 3 SNPs separating 31 patient isolates recovered from eight confirmed human-to-human transmission chains [11]. Thus, the level of SNPs variation in definite transmission chains is limited to no more than 3-5 SNPs. These analyses are focused on SNPs not identified in repetitive DNA regions, to minimize risk of making false positive SNP calls [11*]. Taken together, it was proposed that transmission links may be expected to exist between isolates different by five or fewer SNPs, and not to exist between isolates differing by >12 SNPs [10**].

Mycobacterium abscessus

The intrinsic and acquired resistance of the rapidly growing nontuberculous mycobacterium, *Mycobacterium abscessus*, to commonly used antibiotics limits treatment options to the chronic pulmonary infection caused by this species. In the United States and Europe, 3-10% of cystic fibrosis patients are infected with *M. abscessus*. Transmission of *M. abscessus* from person to person has not been definitely demonstrated, although there is concern that this species may spread among patients with cystic fibrosis. Combined use of whole genome sequencing and epidemiologic investigation of 11 cystic fibrosis patients revealed two clustered outbreaks of *M. abscessus* subspecies *massiliense*, with isolates from each cluster differing by less than 10 SNPs. Mutation conferring high-level amikacin resistance selected during treatment of the presumed index case was found in isolates infecting four other patients who did not have a history of treatment with aminoglycosides, supporting the conclusion that a resistant *M. abscessus* clone is transmitted between cystic fibrosis patients.

Fungal infection

Mucormycosis is a rare fungal infection caused by environmentally acquired molds. A cluster of 13 patients developed necrotizing cutaneous mucormycosis caused by *Apophysomyces trapeziformis* after sustaining tornado-related injuries, including fracture, blunt and penetrating trauma [14**]. Contaminated medical equipments, including bandages, wooden tongue depressors, and ostomy bags, are potential nosocomial sources for mucormycosis. However, this seems unlikely to be the source among these case patients because they were treated at six different hospitals. Whole genome sequencing based on 11 of the 13 isolates showed that four different clones infected the case patients, which effectively ruled out a common nosocomial source. Clonally related isolates did not cluster geographically; instead, they were scattered throughout the tornado path, with the two predominant clones causing infections in patients living closest to the touchdown site of the

tornado. One or more environmental sources of apophysomyces may exist along the tornado path and that aerosolized spores may be carried along by tornado and inoculated into wounds of case patients who sustained tissue-penetrating trauma.

Conclusion

Although whole genome sequencing has been used largely for retrospective analysis of pathogens suspected to be part of disease outbreaks (Table), its usefulness for outbreak investigation and identification of transmission dynamics is limited by a lack of real-time deployment of this new technology. The plummeting cost and increasing throughput of whole genome sequencing may pave the way for its integration into many laboratory-based disease surveillance program. However, the enormous volume of data generated by next-generation sequencing technology will be a major challenge to traditional infection control teams without dedicated bioinformatics support. Development of a fully automated bioinformatics pipeline for analysis of whole genome sequence data is much needed for its integration into infection control practice and outbreak investigation.

Acknowledgments

Dr. Diep is supported by a United States Public Health Service grant NIH NIAID R01 AI087674.

REFERENCES

1. Rohde H, Qin J, Cui Y, et al. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. The New England journal of medicine. 2011; 365(8):718–24. [PubMed: 21793736]

2. Rasko DA, Webster DR, Sahl JW, et al. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. The New England journal of medicine. 2011; 365(8):709–17. [PubMed: 21793740]

3*. Grad YH, Lipsitch M, Feldgarden M, et al. Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(8):3065–70. [PubMed: 22315421] [This article reported a molecular epidemiological analysis using multiplatform whole-genome sequencing and analysis of multiple isolates from the German and French outbreaks.]

4**. Underwood AP, Dallman T, Thomson NR, et al. Public health value of next-generation DNA sequencing of enterohemorrhagic Escherichia coli isolates from an outbreak. Journal of clinical microbiology. 2013; 51(1):232–7. [PubMed: 23135946] [Whole genome sequence was used to investigate an outbreak of enterohemorrhagic *E. coli* in the UK. Based on 16 genoes analysis, 4 SNPs was identified and used to design an assay to type 122 other outbreak isolates. The sequence data indicated that the outbreak strain was expanded and contaminated the environment and animals which led ultimately to human cases.]

5. Chin CS, Sorenson J, Harris JB, et al. The origin of the Haitian cholera outbreak strain. The New England journal of medicine. 2011; 364(1):33–42. [PubMed: 21142692]

6**. Koser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. The New England journal of medicine. 2012; 366(24):2267–75. [PubMed: 22693998] [The authors investigated a putative MRSA outbreak in a neonatal intensive care unit by using whole genome sequence with a clinically relevant turnaround time. Seven isolates associated with the outbreak and another seven MRSA isolates associated with carriage of MRSA or bacteremia in the same hospital were sequenced. This revealed a clear separation between outbreak and non-outbreak isolates. In addition, the authors detected a previously missed transmission event between 2 patients who were not part of the outbreak.]

7**. Harris SR, Cartwright EJ, Torok ME, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis. 2013; 13(2):

130–6. [PubMed: 23158674] [The authors studied a putative MRSA outbreak on a special care baby unit by using whole genome sequence to validate and expand findings from conventional analysis of epidemiological data and antibiogram profiles. MRSA isolates from all colonized patients in the ward and from patients in the hospital or community with the same antibiogram profile as the outbreak strain were sequenced. Analysis of the whole genome sequence identified 26 related cases of MRSA carriage, and showed transmission occurred within the ward, between mothers on a postnatal ward, and in the community.]

8. Harris SR, Feil EJ, Holden MT, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 2010; 327(5964):469–74. [PubMed: 20093474]
9. Gardy JL, Johnston JC, Ho Sui SJ, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med. 2011; 364(8):730–9. [PubMed: 21345102]
10**. Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. The Lancet infectious diseases. 2013; 13(2):137–46. [PubMed: 23158499] [In this retrospective observational study, 390 separate Mtb isolates, including representatives from all five major lineages, from 254 patients were sequenced. These isolates were characterized into four groups: cross-sectional, longitudinal, household, and community. The sequence data estimated rate of change in DNA sequences was 0.5 SNPs per genome per year (95% CI 0.3–0.7) in longitudinal isolates. Divergence is rarely higher than five SNPs in 3 years. Genetic trees and clinical and epidemiological data suggest that super-spreaders were present in two community clusters.]
11. Roetzer A, Diel R, Kohl TA, et al. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS medicine. 2013; 10(2):e1001387. [PubMed: 23424287] [A total of 2,301 patients were subjected in this long-term prospective population-based molecular epidemiological surveillance from 1997 to 2010. Whole genome sequence was analyzed from 86 isolates revealed 85 SNPs, subdividing the outbreak into seven genome clusters (two to 24 isolates each), plus 36 unique SNP profiles.]
12**. Bryant JM, Grogono DM, Greaves D, et al. Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study. Lancet. 2013 [The aim of this study is to identify the transmission of *M. abscessus* between patients with cystic fibrosis. Whole genome sequence and antimicrobial susceptibility testing were performed on 168 consecutive isolates of *M. abscessus* from 31 patients. Phylogenetic analysis revealed two clustered outbreaks of near-identical isolates of the *M abscessus* subspecies massiliense, differing by less than ten base pairs. This variation represents less diversity than that seen within isolates from a single individual, strongly indicating between-patient transmission.]
13**. Snitkin ES, Zelazny AM, Thomas PJ, et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med. 2012; 4(148):148ra116. [A noscomial outbreak of carbapenem-resistant Klebsiella pneumoniae occurred in the USA, involving 18 patients, with 6 deaths. Whole genome sequence was performed on *K. pneumoniae* isolates to gain insight of this outbreak. Integrated genomic and epidemiological analysis traced the outbreak to three independent transmissions from a single patient who was discharged 3 weeks before the next case became clinically apparent. Additional genomic comparisons provided evidence for unexpected transmission routes, with subsequent mining of epidemiological data pointing to possible explanations for these transmissions.]
14**. Neblett Fanfair R, Benedict K, Bos J, et al. Necrotizing cutaneous mucormycosis after a tornado in Joplin, Missouri, in 2011. The New England journal of medicine. 2012; 367(23):2214–25. [PubMed: 23215557] [This article reported a cluster of cases of cutaneous mucormycosis among tornado survivors that were associated with substantial morbidity and mortality. The authors conducted a case–control study by reviewing medical records and conducting interviews with case patients and hospitalized controls. Whole genome sequence was performed on clinical specimens which showed that the apophysomyces isolates were four separate strains.]
15. Eppinger M, Mammel MK, Leclerc JE, et al. Genomic anatomy of Escherichia coli O157:H7 outbreaks. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(50):20142–7. [PubMed: 22135463]

16. McDonnell J, Dallman T, Atkin S, et al. Retrospective analysis of whole genome sequencing compared to prospective typing data in further informing the epidemiological investigation of an outbreak of Shigella sonnei in the UK. Epidemiology and infection. 2013:1–8.

17. Lienau EK, Strain E, Wang C, et al. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med. 2011; 364(10):981–2. [PubMed: 21345093]

18. Kaper JB, Nataro JP, Mobley HL. Pathogenic Escherichia coli. Nat Rev Microbiol. 2004; 2(2): 123–40.

19. Dutta PR, Cappello R, Navarro-Garcia F, Nataro JP. Functional comparison of serine protease autotransporters of enterobacteriaceae. Infect Immun. 2002; 70(12):7105–13. [PubMed: 12438392]

20. Bernier C, Gounon P, Le Bouguenec C. Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative Escherichia coli as a sensitive probe for detecting the AAF-encoding operon family. Infect Immun. 2002; 70(8):4302–11. [PubMed: 12117939]

21. Harris JB, LaRocque RC, Qadri F, et al. Cholera. Lancet. 2012; 379(9835):2466–76. [PubMed: 22748592]

22. Vonberg RP, Weitzel-Kage D, Behnke M, Gastmeier P. Worldwide Outbreak Database: the largest collection of nosocomial outbreaks. Infection. 2011; 39(1):29–34. [PubMed: 21153042]

23. Miller LG, Diep BA. Clinical practice: colonization, fomites, and virulence: rethinking the pathogenesis of community-associated methicillin-resistant Staphylococcus aureus infection. Clin Infect Dis. 2008; 46(5):752–60. [PubMed: 18220477]

24. Enright MC, Robinson DA, Randle G, et al. The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). Proc Natl Acad Sci U S A. 2002; 99(11):7687–92. [PubMed: 12032344]

25. Holden MT, Lindsay JA, Corton C, et al. Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant Staphylococcus aureus, sequence type 239 (TW). J Bacteriol. 2010; 192(3):888–92. [PubMed: 19948800]

26. Li M, Du X, Villaruz AE, et al. MRSA epidemic linked to a quickly spreading colonization and virulence determinant. Nat Med. 2012; 18(5):816–9. [PubMed: 22522561]

27. Tzouvelekis LS, Markogiannakis A, Psichogiou M, et al. Carbapenemases in Klebsiella pneumoniae and other Enterobacteriaceae: an evolving crisis of global dimensions. Clin Microbiol Rev. 2012; 25(4):682–707. [PubMed: 23034326]

28. Ford CB, Lin PL, Chase MR, et al. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat Genet. 2011; 43(5):482–6. [PubMed: 21516081]

Key Points

- The combination of whole genome sequencing and intensive epidemiologic analysis provided new insights on the sources and transmission dynamics.
- Usefulness of whole genome sequencing for outbreak investigation is limited by a lack of real-time deployment of this new technology.
- Development of automated bioinformatics analysis of whole genome sequence data is much needed to make this powerful tool more widely accessible for real-time outbreak investigation.

Table 1

Use of whole genome sequencing in outbreak investigation

Time period	Location, Country	Pathogens	Application ofWGS	WGS isolates	Sequencer	Source of the outbreak	Comments	Ref.
2011	Germany	*E.coli* O104:H4	Real time	1	Illumina	contaminated imported seeds	The outbreak strain belonged to enteroaggregative *E. coli* lineage and had acquired Shiga toxin 2 and antibiotic resistance genes.	[1]
2011	Germany	*E.coli* O104:H4	Real time	12	PacBio RS	contaminated imported seeds	The outbreak strain was different from other O104:H4 clone in containing prophage encoding Shiga toxin 2, additional virulence and antibiotic resistance factors.	[2]
2011	Germany	*E. coli* O104:H4	Real time	15	Illumina	contaminated imported seeds	The outbreak strains from Germany and France were clonally related to E coli O104:H4 from African countries.	[3*]
2009	Surrey, UK	*E. coli* O517:H7	Real time	16	454 pyrosequencing and Illumina	animal in a farm	WGS showed gross contamination of the outbreak strain across a farm that had been epidemiologically linked to human cases.	[4**]
2010	Artibonite, Haiti	*Vibrio cholerae*	Real time	5	PacBio RS		The outbreak strain is closely related to pandemic El Tor O1 strain.	[5]
2009	Cambridge, UK	*Staphylococcus aureus*	Retrospective	14	Illumina	nosocomial transmission	WGS showed a clear distinction between isolates in the outbreak	[6**]

Time period	Location, Country	Pathogens	Application ofWGS	WGS isolates	Sequencer	Source of the outbreak	Comments	Ref.
							and non-outbreak group. and non-outbreak group.	
2011	Cambridge, UK	*Staphylococcus aureus*	Real time	15	Illumina	nosocomial transmission	WGS identified the transmission of bacteria among patients, which could not be confirmed with conventional method.	[7**]
1982-2003 2006-2007	ST239 worldwide	*Staphylococcus aureus*	Retrospective	46	Illumina		WGS data was used to estimate the time frame for the emergence of a bacterial pathogen clone and its evolution during epidemic spread.	[8]
2006-2008	Canada	*Mycobacterium tuberculosis*				transmission	WGS allowed for differentiation of the outbreak isolates into two distinct genetic lineages of *M. tuberculosis* that	[9]
1994-2011	Birmingham, Leicester, UK	*Mycobacterium tuberculosis*	Retrospective	390	Illumina	person-to-person tramission	WGS revealed no more than 5 SNPs separating epidemiologically-linked isolates.	[10**]
1997-2010	Hamburg, Germany	*Mycobacterium tuberculosis*	Retrospective	86	454 pyrosequencing	person-to-person transmission	WGS provides a measure of *M. tuberculosis* genome evolution over time in its natural host context.	[11]
2007-2011	Cambridge, UK	*Mycobacterium abscessus*	Retrospective	168	Illumina	person-to-person transmission	WGS showed person-person transmission of drug- resistant *M. abscessus* between cystic fibrosis patients.	[12**]
2011-2012	NewYork, US	*Klebsiella pneumonia*	Retrospective	18	454 pyrosequencing	nosocomial transmission	Three independent transmission events from isolates colonizing the different body	[13**]

Time period	Location, Country	Pathogens	Application ofWGS	WGS isolates	Sequencer	Source of the outbreak	Comments	Ref.
							sites of the index patient to the other cases w	
							sites of the index patient to the other cases w	
							sites of the index patient to the other cases w	
							sites of the index patient to the other cases w	
2011	Missouri, US	*Apophysomyces trapeziformis*	Real time	18	Illumina	environmental molds	WGS identified 4 different types of *Apophysomyces trapeziformis* that caused the mycormycosis outbreak.	[14**]