Korean Journal of Microbiology (2018) Vol. 54, No. 3, pp. 188-199
DOI https://doi.org/10.7845/kjm.2018.8045

pISSN 0440-2413
eISSN 2383-9902

Review

Genomic epidemiology for microbial evolutionary studies and the use of Oxford Nanopore sequencing technology

Sang Chul Choi*

Department of Biotechnology, Sungshin Women's University, Seoul 01133, Republic of Korea

미생물 진화 연구를 위한 유전체 역학과 옥스포드 나노포어 염기서열분석 기술의 활용

최상철*

성신여자대학교 지식서비스공과대학 바이오생명공학과

(Received July 9, 2018; Revised August 25, 2018; Accepted August 26, 2018)

Genomic epidemiology exploits various basic microbial research areas. High-throughput sequencing technologies dramatically have been expanding the number of microbial genome sequences available. Abundant genomic data provide an opportunity to perform strain typing more effectively, helping identify microbial species and strains at a higher resolution than ever before. Genomic epidemiology needs to find antimicrobial resistance genes in addition to standard genome annotations. Strain typing and antimicrobial resistance gene finding are static aspects of genomic epidemiology. Finding which hosts infected which other hosts requires the inference of transient transmission routes among infected hosts. The strain typing, antimicrobial resistance gene finding, and transmission tree inference would allow for better surveillance of microbial infectious diseases, which is one of the ultimate goals of genomic epidemiology. Among several high-throughput sequencing technologies, genomic epidemiology will benefit from the more portability and shorter sequencing time of the Oxford Nanopore Technologies's MinION, the third-generation sequencing technology. Here, this study reviewed computational methods for quantifying antimicrobial resistance genes and inferring disease transmission trees. In addition, the MinION's applications to genomic epidemiology were discussed.

***For correspondence.** E-mail: sangchulchoi@sungshin.ac.kr; Tel.: +82-2-920-2751; Fax: +82-2-920-2047

Keywords: antimicrobial resistance, infectious disease outbreaks, public health, surveillance, transmission tree inference

Microbial genomics is ushering in an era of genomic epidemiology (Traynor, 2009), employing the whole-genome sequencing (Metzker, 2010) in traditional molecular epidemiology (Eybpoosh *et al.*, 2017) studies. Fleischmann *et al.* (1995) triggered the start of microbial genomics by sequencing the complete bacterial genome for *Haemophilus influenzae*. Thousands of complete microbial genomes have been available since then (Mukherjee *et al.*, 2017). Wu *et al.* (2009) made efforts to expand taxon sampling coverage in an attempt of inferring a complete microbial phylogeny. These efforts even increased the number of microbial genome sequences. Earlier studies in microbial genomics documented genomes using gene prediction tools on a single genome basis (Alm *et al.*, 1999). Comparative genomics changed the direction of research into examining a collection of genomes (Tettelin *et al.*, 2008). Studies on some microbial genomes led to the analysis of microbial pan-genome, consisting of core and accessory genes. Core genes under strong negative selective pressure are often passed on vertically, not changing much over evolutionary time. Accessory genes are more likely

to be horizontally transferable, allowing microbial organisms to adapt to changing environments (Lapierre and Gogarten, 2009). Microbial genomes consist of operons, each holding a static set of functionally related genes as a unit of transcription (Jacob and Monod, 1961). Microbial transcriptome, later, exhibited dynamic RNA expression, varying in different culture conditions. For instance, Sharma *et al.* (2010) employed a massively parallel differential RNA sequencing, showing that *Helicobacter pylori* displayed diverse repertoires of the microbial transcriptome. The microbial transcriptome explained the complexity of alternative splicing in eukaryotes. Thus far, high-throughput sequencing technologies helped document microbial genomes and transcriptomes in many species, mostly aiming at better understanding microbial pathogenesis (Klemm and Dougan, 2016). The past several years has seen that the earlier basic research on the catalog of microbial genomes and transcriptomes are transitioning to a more practical application to molecular epidemiology (Guthrie and Gardy, 2017).

Molecular epidemiology: strain identification

Traditional molecular epidemiology researchers apply molecular biology techniques to epidemiology studies for disease prevention

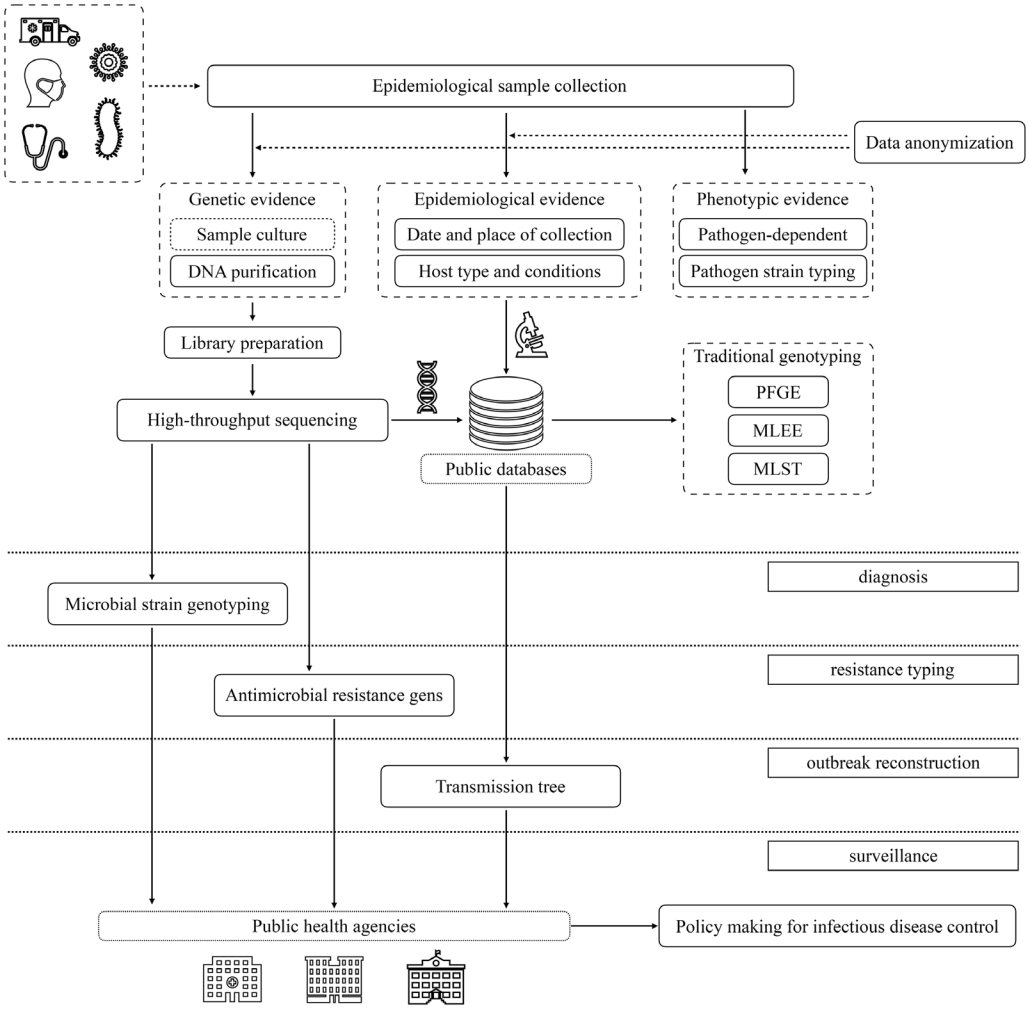

Fig. 1. The diagrammatic flow of genomic epidemiology studies. The workflow starts in the top left corner with epidemiological sampling. Anonymization of sampling data, if necessary, would protect the privacy of individuals under surveillance. The sampled specimen may undergo experimental procedures for phenotypic, genotypic, or epidemiological evidence. Genetic and some of the epidemiological evidence should be stored at the central public databases for further analyses. Sequencing techniques allow for microbial strain typing and antimicrobial resistance gene finding. Transmission tree inference for outbreak reconstruction would be suggested from genetic and epidemiological data. Public health agencies would utilize diagnosis, resistance typing, and outbreak reconstruction for making a better policy of controlling and limiting the dispersal of infectious diseases.

and management (Kilbourne, 1973). They track the development of infectious diseases and identify the causes of such diseases to control and oversee the spread of the diseases (Fig. 1). Finding out the causes of microbial infectious diseases amounts to identifying the species or strains of such microorganisms genotypically or phenotypically. First, phenotypic identification of microorganisms often depends on examining blood serum, the method of which varies among different microbial species or strains. Phenotypic identification methods are often instrumental in diagnoses, and not easily yet transferrable for comparisons among public health agencies because phenotypic measurements are not easy to compare one with another. Second, traditional genotyping techniques include pulsed-field gel electrophoresis (PFGE), multi-locus enzyme electrophoresis (MLEE), and multi-locus sequence typing (MLST), to name a few (Maiden, 1998). They all use the variation of microbial sequences to discriminate differences in genotypes of microorganisms. While PFGE often was valuable for discriminating between minute variations and identifying strains occurring in a local area, MLEE was better at clustering closely related strains (Maiden, 1998). MLST technique uses only several genes that exist universally in bacteria to genotype strains, having gained popularity among population genetics studies to understand the distribution of infectious diseases caused by pathogens. The last decade has seen the development of MLST's derivatives, resolving the unfortunate situation of the use of a fraction of a total genome. The development includes ribosomal MLST (Jolley *et al.*, 2012), core-genome MLST, and whole-genome MLST (Maiden *et al.*, 2013). The further development of MLST expanded genotyping widths, allowing for higher resolution of microbial species delimitation. Applying a statistical approach to genotype data, one can learn of the population structure of individual organisms (Pritchard *et al.*, 2000). Similarly, MLST genotype data with a clustering method such as eBURST grouping were instrumental in relating bacterial isolates for evolutionary investigation (Feil *et al.*, 2004). Although MLST and its derivatives suggested better resolution in typing bacterial strains, they would use still a gene-by-gene comparison. Thus, they often failed to make full use of the totality of sequence data by the whole genome sequencing.

High-throughput DNA sequencing

DNA sequencing technologies have undergone a few generations (Heather and Chain, 2016), each of which evolved microbial genome sequencing. After the publication of the chain termination sequencing (Sanger *et al.*, 1977), researchers made efforts to mechanically parallelize the improved Sanger sequencing (Green, 2001). The genome sequencing of the free-living organism (Fleischmann *et al.*, 1995) realized the earlier idea of shotgun sequencing (Staden, 1979). Pyrosequencing technology by Roche 454 heralded a major change in sequencing technologies, popularizing the term of next-generation sequencing (NGS) (Margulies *et al.*, 2005). The second-generation sequencing technologies included a sequencing-by-ligation method called SOLiD (Shendure *et al.*, 2005) and a sequencing-on-bead-via-emulsion PCR method called Ion Torrent sequencing (Nakano *et al.*, 2003). Illumina sequencing has been dominating the DNA sequencing market share using the so-called sequencing-by-synthesis method (Bentley *et al.*, 2008). Before Illumina sequencing, hundreds of bacterial genomes existed (Land *et al.*, 2015). Since the drop in sequencing cost in 2007, the second-generation sequencing has increased the number of bacterial genome sequences up to tens of thousands. Notably, it was because of a surge of metagenomic projects in the last decade (Land *et al.*, 2015). The third-generation sequencing technologies read DNA nucleotides at the single-molecule level. Two technologies of third-generation sequencing included single molecule real-time sequencing by Pacific Bioscience (Eid *et al.*, 2009), and the Oxford Nanopore Technologies's (ONT) MinION (Deamer *et al.*, 2016). These third-generation sequencers would produce longer read sequences, facilitating finishing microbial genomes (Chin *et al.*, 2013; Wick *et al.*, 2017). The development of high-throughput sequencing allowed molecular epidemiology to gain access readily to microbial genomes, ushering in the era of genomic epidemiology.

Genomic epidemiology: antimicrobial resistance genes

Genomic epidemiology studies call for genome annotations that feature epidemiologically informative microbial types, including strain typing and antimicrobial resistant gene finding (Fig. 1). Genomic data analysis begins with annotating genome sequences with the locations and descriptions of genes. Many

efforts have been made to develop prokaryotic genome annotation methods despite the much simpler genomes of prokaryotic organisms compared with eukaryotic ones. For instance, NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) developed in 2005 has allowed users to simultaneously deposit and annotate a prokaryotic genome sequence (Tatusova *et al.*, 2016). The Rapid Annotations using Subsystems Technology (RAST) having started as a web-based service in 2007 annotates functional genes based on manually curated protein family libraries of subsystems (Aziz *et al.*, 2008). As these general-purpose genome annotation tools hinge on experimental biological evidence, molecular epidemiological features prediction tools need datasets serving as experimentally corroborated evidence. McArthur and Tsang (2017) reported a detailed review of molecular epidemiological databases for antimicrobial resistance surveillance. This part discusses key features of computational tools for quantifying antimicrobial resistance genes. Before that, one NCBI's database worth mentioning is Bacterial Antimicrobial Resistance Reference Gene Database (NCBI BioProject Accession PRJNA313047. Retrieved 2018, June 10. from https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047), allowing access to DNA and protein sequences involved in antimicrobial resistance genes. Other databases, e.g., Resfinder (Zankari *et al.*, 2012) and the Comprehensive Antimicrobial Resistance Database (CARD) (Jia *et al.*, 2017), contributed to some of the NCBI's antimicrobial resistance sequence datasets (see more contributors' references in the BioProject PRJNA313047). The different approaches used by computational tools for predicting antimicrobial resistance genes are as follow. NCBI's AMRFinder, employed by the NCBI Pathogen Detection pipeline (NCBI's Pathogen Detection. Retrieved 2018, June 10. from https://www.ncbi.nlm.nih.gov/pathogens/), uses the Bacterial Antimicrobial Resistance Reference Gene Database to predict antimicrobial resistant genes (NCBI's AMRFinder. Retrieved 2018, June 10. from https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/). ARGs-OAP along with Structured Antibiotic Resistance Genes (SARG) database uses hidden Markov models to quantify antibiotic resistance genes in metagenomic data (Yin *et al.*, 2018). ResFinder (Zankari *et al.*, 2012) and Sequence Search Tool for Antimicrobial Resistance (SSTAR) (de Man and Limbago, 2016) use NCBI's BLAST search for antimicrobial resistance genes. KvarQ, provided as a graphical and command-line user interface, mines short read sequence data for drug resistance mutations without reference-based genome mapping or genome assembly (Steiner *et al.*, 2014). Similarly, Short Read Sequence Typing (SRST2) uses a read mapping technique to determine the types of short read sequences in the MLST fashion (Inouye *et al.*, 2014). Rather than devising a new computational technique, Search Engine for Antimicrobial Resistance (SEAR), provided as a command-line tool and a website service, combines off-the-shelf computational tools to quantify antimicrobial resistance genes (Rowe *et al.*, 2015). Rowe and Winn (Retrieved 2018, July 3. from https://github.com/will-rowe/groot) implemented a graph representation of genes for antibiotic resistance gene typing by taking advantage of an assembly graphs visualization tool (Wick *et al.*, 2015), which helped visualize the variation graphs of short read alignments. Some of the antimicrobial resistance prediction tools are specialized for *Mycobacterium tuberculosis*; e.g., Comprehensive Analysis Server for the *Mycobacterium tuberculosis* complex (CASTB) (Iwai *et al.*, 2015), TB Profiler (Coll *et al.*, 2015), and Mykrobe predictor (Bradley *et al.*, 2015). Microbial genome sequencing and antimicrobial resistance gene finding are static aspects of genomic epidemiology. Infectious disease surveillance needs to track infected hosts, which was described in the next subsection.

Genomic epidemiology: transmission tree inference

Tracing epidemiological footprints left in the pathogen genomes after or during infectious disease outbreaks considers relations among multiple individual hosts that are infected by pathogens while genotyping microbial pathogens is a single individual endeavor. Epidemiological data such as times and places of infected individuals suggest possible outbreak investigators of possible dispersal routes of infectious diseases. Because such efforts are labor-intensive to reach decisive transmission mappings, increasingly available genotype data have been used with computational tools employing mathematical models (Table 1). The SIR model states that hosts of disease-causing microbial parasites experience infection stages of susceptible, infectious, and removed (Didelot *et al.*, 2014). The microbial parasites in infected hosts would undergo microevolution in different speed. Microbial genomes in an infected individual may change over time during the infectious period. This simplistic

model imposes two different evolutions on tracing epidemiological footprints. First and foremost, potential and infected hosts are related by the unknown network that connects those hosts as vertices with directed edges denoting pathogen transmission events between hosts. Second, microbial pathogens may undergo molecular evolution within each of their hosts. The two distinct evolutions of pathogens and hosts led to the pioneering statistical treatment of both epidemiological and genetic data for inferring transmission trees (Cottam *et al.*, 2008). While Cottam *et al.* (2008) developed a maximum-likelihood approach, most of the following notable implementations of transmission tree inference have been Bayesian approaches. Note that, contrary to most of the other approaches, Numminen *et al.* (2014) used an importance sampling approach, which may take advantage of transmissions known *a priori*. Morelli *et al.* (2012) are one of the Bayesian approaches that combine epidemiological and genetic data to rebuild transmission trees.

The transmission tree inference procedures need simplifying assumptions for mathematical and computational conveniences. Each of such inference procedures tried to relax some of those assumptions. Some of the assumptions include a single transmission of pathogens to a susceptible host, all observed infected hosts in an epidemic, constant host infection potential, and transmission tree interpretation of a pathogen's phylogenetic tree. Examination of the four assumptions is in order. First, the single transmission event assumption is flawed; after susceptible hosts became exposed and infected, other pathogens could be carried to the hosts more than once. Second, the assumption that we observe all the infected hosts is rarely valid because observing all the infected hosts would be impossible. Epidemiologists could collect infected host records retrospectively after the end of a disease outbreak. However, not all of the infected hosts could be reported to public health agencies because some of the infected hosts might be recovered before they were reported to such agencies. Infectious disease control agencies would want to design policies that could limit the spread of the disease during infectious disease epidemic, suggesting that our observations of infected hosts would be inevitably partial. Third, the assumption of constant host infection potential is not easily justified because pathogens within an infected host may change infectivity depending on the corresponding host's state in times and places. Lastly, increasingly available genetic data have triggered the use of phylogenetic trees of pathogens as a surrogate of transmission trees of hosts. Many efforts have been made to relate evolutionary trees of pathogens to the epidemiological trajectories of the hosts (Grenfell *et al.*, 2004). However, one should be cautious to treat phylogenetic trees as epidemiological interpretation because pathogen's evolutionary trajectories and host transmission

Table 1. Comparison of transmission tree inference methods

Software package (language)[a]	Statistical method[b]	Partial observation[c]	Reference
n/a	ML	no	Cottam *et al*. (2008)
n/a	MCMC	no	Morelli *et al*. (2012)
n/a	MCMC	no	Ypma *et al*. (2013)
outbreaker (R)	MCMC	yes	Jombart *et al*. (2014)
n/a	MCMC	yes	Mollentze *et al*. (2014)
n/a	IS	yes	Numminen *et al*. (2014)
TransPhyloMatlab (Matlab)	MCMC	no	Didelot *et al*. (2014)
n/a	MCMC	yes	Lau *et al*. (2015)
BEAST package BEASTLIER (Java)	MCMC	no	Hall *et al*. (2015)
BEAST package SCOTTI (Java)	MCMC	yes	De Maio *et al*. (2016)
phybreak (R)	MCMC	no	Klinkenberg *et al*. (2017)
TransPhylo (R)	MCMC	yes	Didelot *et al*. (2017)

[a]Software package names mean that some computational tools are available. The note of n/a implies that no such software packages are available. The word in the parentheses is a computer programming language used for implementing the package.
[b]Statistical methods are one of three: ML for maximum likelihood, MCMC for Markov chain Monte Carlo, and IS for importance sampling.
[c]Methods assume that we observe either all or some of the infected hosts.

networks may not coincide (Romero-Severson *et al.*, 2014).

A phylogenetic tree of pathogens could help understand the infection trajectories of hosts especially if pathogens like RNA viruses underwent a rapid evolution. However, pathogens may not much evolve within hosts, which may result in multiple lineages inside hosts. Multiple infections of pathogens from different host origins into a host may occur. Therefore, within-host evolution complicates the transmission tree inference (Worby *et al.*, 2014). Ypma *et al.* (2013), still assuming that they observed all the infected hosts, was the first to develop a method for jointly inferring the transmission tree and phylogenetic tree with consideration of the within-host evolution. However, they failed to provide a hands-on implementation for use. Jombart *et al.* (2014) improved the statistical methods developed by Ypma *et al.* (2013) and Morelli *et al.* (2012) via relaxing the assumptions that they observed all infected hosts and a single infection led to an infected host. Fortunately, Jombart *et al.* (2014) were the first that provided the computational tool for inferring transmission trees, *outbreaker*, as a statistical software R package. De Maio *et al.* (2016) applied their structured coalescent method (De Maio *et al.*, 2015) to the transmission tree inference problem where populations and migration events represented hosts and transmission events, respectively. They considered transmission cases that result in the conflict between the phylogenetic and transmission trees; i.e., within-host evolution, non-sampled host, multiple infections to a host, and incomplete bottleneck.

Didelot *et al.* (2014), still assuming that they observed all infected hosts, employed a two-step procedure of inferring first phylogenetic trees using DNA sequence data. Afterward, they would estimate the transmission tree based on the phylogenetic trees as data to the Bayesian approach. The two-step procedure might fail to account for the uncertainty in pathogen's phylogenetic trees although Didelot *et al.* (2017) defended their preference over the combined approach. More elaborate approaches jointly estimated the pathogen's phylogenetic tree and the host's transmission tree in a combined Markov chain Monte Carlo (MCMC) (Hall *et al.*, 2015; Klinkenberg *et al.*, 2017). There were methods for inferring transmission trees to deal with partially observed epidemiological data that were often collected during infectious disease epidemic (Mollentze *et al.*, 2014; Didelot *et al.*, 2017).

Genomic epidemiological research with the MinION sequencer

Genomic epidemiology studies are increasingly using genomic sequence data to control and oversee the spread of infectious diseases effectively. One of the crucial features of epidemiological approaches is time. It takes time to sequence microbial genomes from pathogen sampling with epidemiological documentation. Currently, most high-throughput DNA sequencing would need the transfer of biological samples back to the laboratories for further purification or amplification of DNA molecules from a biological specimen. Sequencing library preparation itself requires laboratory work. Besides, microbial pathogens often need culture steps to obtain enough DNA molecules although DNA library preparation may be possible directly from clinical samples in a culture-independent manner. One DNA sequencing technology that could reduce time to sequence from samples is one of the third-generation sequencing technologies, the Oxford Nanopore Technologies's MinION sequencer.

The third-generation sequencing platforms are based on single-molecule sequencing technologies commercialized by Pacific Biosciences (PacBio) (Eid *et al.*, 2009) and the Oxford Nanopore Technologies (ONT) (Deamer *et al.*, 2016). Both of the technologies tend to produce read sequences longer than those generated by the second-generation sequencing platforms. The third-generation sequencers, unfortunately, would suffer from higher rates of sequencing errors compared with those of the second-generation sequencers. The sequencing error rate of the ONT's sequencing platform is known to be even larger than that of the PacBio's sequencing platform. The molecule-level model of the ONT's sequencing technology consists of three biological components: motor proteins, biological membranes, and pore proteins (Deamer *et al.*, 2016). A motor protein binds to a double-stranded DNA and guides the DNA to a transmembrane pore protein that is bound in a membrane. Then, the motor protein is attached to the pore protein and unzips the accompanied double-stranded DNA, one strand of which passes through the pore protein. Electrical current flows across the membrane and fluctuates as bases of the unzipped single-stranded DNA pass through the pore. The current fluctuation produces wiggling analog signals. Decoding the analog signals into digital values is a base-calling step in the ONT's nanopore

sequencing. The ONT's MinION with a single flowcell weighed less than a typical smartphone, which would enable it highly portable. The ONT's other benchtop sequencers would provide higher throughput than the MinION because of their accommodation of multiple flowcells.

The genomic epidemiology with sequencing data has been mostly retrospective; i.e., collecting sequence data after the end of infectious disease outbreaks. The second-generation sequencing platforms and the PacBio's sequencing would be more cost-effective for analyzing sequencing data collected from retrospective epidemiology studies because of the lower cost of their sequencing per base compared with that of the ONT's MinION. However, collecting sequencing data during outbreaks to execute infectious disease controlling measures would provide more effective ways of monitoring the spread of infectious diseases. The in-field sequencing capability of the ONT's MinION is currently unique in the sequencing market, potentially reducing costs in some of the downstream work of surveillance of infectious diseases. Thus, monitoring and controlling the spread of infectious diseases could greatly benefit from the in-field sequencing feature of the ONT's MinION. For example, transferring samples back to the laboratory for sequencing libraries could be difficult because of the physical distance or local regulatory limitation. Sequencing facilities might be too far away from the region of disease outbreaks. Biological samples might be prohibited from being moved into the area where sequencing facilities were available. Although sequencing facilities could be planned to be established near the region of a disease outbreak, sequencer manufacturers might be reluctant to send their employees to the region of disease outbreaks. Therefore, the portability of the MinION and the availability of the sequencing kits could allow epidemiologists or their collaborators to perform in-field sequencing for more effective surveillance of infectious diseases.

The ONT's nanopore sequencing is still evolving and genomic epidemiology with such in-field sequencing data has yet to be developed. Here, I survey some of the use of the MinION's sequencing in the areas related to genomic epidemiology. Votintseva *et al.* (2017) took advantage of the MinION's fast sequencing to produce a diagnostic data for tuberculosis. The MinION's sequencing starts just after loading DNA libraries, which allowed researchers to identify pathogenic viruses in three hours from sample receipt (Kilianski *et al.*, 2016) and pathogens from urine in 4 h (Schmidt *et al.*, 2017). Field epidemiologists may carry one of the USB-powered portable handheld machines to sequence pathogen DNAs after some library preparation steps. The MinION's portability allowed a metagenomic sequencing study in the extreme environment of arctic permafrost ice wedge (Goordial *et al.*, 2017). The ONT's portable sequencers have been used in bacteria-causing hospital outbreak (Quick *et al.*, 2015), Ebola surveillance (Quick *et al.*, 2016), sequencing *Plasmodium falciparum* for diagnostic purpose (Imai *et al.*, 2018), and enterovirus genotyping (Rames and Macdonald, 2018).

Additionally, the ONT's MinION sequencers have been applied to various microbial genetic and epidemiological studies; genome sequencing, strain identification, antimicrobial resistance gene finding (Fig. 1). First, the Oxford Nanopore Technologies mobile portable handheld sequencer, the MinION, has been used in genome assembly of *Escherichia coli* K-12 MG1655 (Loman *et al.*, 2015), *Bacteroides fragilis* BE1 (Risse *et al.*, 2015), *Agrobacterium tumefaciens* LBA4404 (Deschamps *et al.*, 2016), multidrug-resistant *Enterobacter kobei* (Judge *et al.*, 2016), methicillin-resistant *Staphylococcus aureus* USA300 (Bayliss *et al.*, 2017), plasmids from *Enterobacteriaceae* isolates (George *et al.*, 2017), *Pseudomonas baetica* a390T (Beaton *et al.*, 2018), and *Streptococcus pyogenes* Serotype M12 (You *et al.*, 2018). While many of these genomes were assembled with a combination of the long-read technology, the MinION, and other short-read technologies, some of them were completed using the nanopore technology only (Loman *et al.*, 2015). The MinION was applied to sequencing directly from clinical samples without culture for studying Zika virus related infectious diseases (Quick *et al.*, 2017). Second, the ONT's single-molecule sequencing capacity renders the MinION a biological molecule detection tool for strain identification rather than a DNA sequencing machine. It has been applied to identify *Salmonella enterica* from food samples (Hyeon *et al.*, 2017), multi-resistant *Escherichia coli* (Schmidt *et al.*, 2017), *Klebsiella pneumoniae* clinical strains from liver samples (Gong *et al.*, 2018), and arbovirus in mosquitos (Russell *et al.*, 2018). Also, the MinION sequencing was used as species identification based on 16S rRNA sequencing (Benítez-Páez *et al.*, 2016), and as a cost-effective bacterial communities profiling method (Kerkhof *et*

al., 2017). Third, the MinION has also been used in the studies of antimicrobial resistance genes in *Salmonella typhi* Haplotype 58 (Ashton *et al.*, 2015), *Staphylococcus aureus* and *Mycobacterium tuberculosis* (Bradley *et al.*, 2015), multiple antibiotic resistant coliform bacteria (Xia *et al.*, 2017), *Klebsiella pneumoniae* (Gorrie *et al.*, 2018; Simner *et al.*, 2018), and *Acinetobacter baumannii* (Hawkey *et al.*, 2018).

Furthermore, genomic epidemiology applications could benefit from the ONT's accessories that are worth mentioning. The ONT's VolTRAX version 2 may allow on-site sequencing library preparation without the need of a back-to laboratory. This still portable machine removes the laboratory needs for sequencing, which would allow for the DNA sequencing out in the field. Another ONT's product called SmidgION that was designed to work with a smartphone would boost the portability of the already small MinION sequencer. The MinION's sequencing raw data could be sent for data analysis to the cloud computing serviced by its sister company of Metrichor. The SmidgION could allow for place-independent sequencing as long as wireless signals are available. One of the drawbacks of the MinION's product is somewhat expensive flow cells compared with other sequencing technologies' ones. Recently, the ONT's Flongle that was for early access only at the time of writing might be more cost-effective in pathogen identification because it would cost much less than the MinION's flow cells. Another drawback of the ONT's sequencing product is the high computer hardware requirements and not-so-easy software installation. The ONT's MinIT would be an excellent alternative to a laptop computer for managing the MinION.

Concluding Remarks

We are entering an exciting era of microbial applications. In near future, researchers will carry portable devices for identifying pathogens and infected hosts *in situ*. Just as we consider the weather forecast as a norm, our descendants could watch infectious disease forecast in the nightly news. To achieve such an ambitious goal, we need to see methodological developments and societal readiness for the new era. Genotypes of microbial species and strains need to be stored and accessible publicly as soon as microbial sequencing is performed. Microbial sequencing needs quality control for data integrity. Pathogen's epidemiological information about infected hosts needs to be anonymized to avoid any privacy concerns. Microbial identifying methods are evolving significantly fast. Therefore, public health agencies will possibly use real-time surveillance information to limit the spread of infectious diseases.

적 요

다양한 미생물학 연구 분야의 발전에 힘입어 유전체역학은 발전되어 왔다. 예를 들어, 대용량서열화 기술의 발전으로 미생물 유전체의 수는 급속도로 증가해 오고 있다. 이러한 풍부한 유전체 데이터는 전에는 보지 못한 보다 더 정확한 미생물 종의 동정에 도움을 주는 균주종 타이핑에 새로운 기회를 제공한다. 유전체역학은 유전체에 일반적인 유전자를 찾고 표기하는 것 뿐만 아니라 항균 저항성 유전자를 찾을 수 있다. 균주종 타이핑과 항균 저항성 유전자 찾기는 각각 종을 구분하고 유전체내의 유전자 위치를 결정하는 유전체 역학의 방법들로 시간에 따른 변화가 없는 측면이다. 이에 반하여, 하나의 숙주가 어떤 숙주를 감염시켰는지 알아내기 위해서는 감염된 숙주들 사이의 시간에 따른 동적인 전염 경로를 추론해야 한다. 이렇게, 균주종 타이핑, 항균 저항성 유전자 찾기, 전염 계통수 추론을 통하여 유전체역학의 궁극적인 목표 중 하나인 미생물성 전염병을 보다 효율적으로 감시할 수 있을 것으로 기대된다. 그리고, 대용량서열화 기술 중, 3세대 서열화기술 중 하나인 옥스포드 나노포어 MinION의 보다 나은 휴대성과 빠른 서열화의 성능 덕분에 유전체역학은 더 많은 발전을 거듭할 것으로 보인다. 이에, 본 연구는 항균 저항성 유전자를 찾고 전염병 경로를 추론하는 계산적인 방법에 대하여 살펴보고, 미생물 유전체역학에서 MinION이 응용된 예들에 대하여 논하였다.

Conflict of Interest

The author has no financial conflict of interest.

Acknowledgements

I thank the reviewers for valuable comments and suggestions that helped improve the manuscript. This work was supported

by the Sungshin Women's University Research Grant of 2016-2-28-001/1.

References

Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, de Jonge BL, *et al.* 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**, 176–180.

Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J, and O'Grady J. 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.* **33**, 296–300.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, *et al.* 2008. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75.

Bayliss SC, Hunt VL, Yokoyama M, Thorpe HA, and Feil EJ. 2017. The use of Oxford Nanopore native barcoding for complete genome assembly. *GigaScience* **6**, 1–6.

Beaton A, Lood C, Cunningham-Oakes E, MacFadyen A, Mullins AJ, Bestawy WE, Botelho J, Chevalier S, Coleman S, Dalzell C, *et al.* 2018. Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas baetica* a390t sequenced by ion semiconductor and nanopore technologies. *FEMS Microbiol. Lett.* **365**, 103.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, *et al.* 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59.

Benítez-Páez A, Portune KJ, and Sanz Y. 2016. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. *GigaScience* **5**, 4.

Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, *et al.* 2015. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat. Commun.* **6**, 1761.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, *et al.* 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569.

Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, *et al.* 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med.* **7**, 321.

Cottam EM, Thébaud G, Wadsworth J, Gloster J, Mansley L, Paton DJ, King DP, and Haydon DT. 2008. Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proc. R. Soc. B* **275**, 887–895.

Deamer D, Akeson M, and Branton D. 2016. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524.

De Maio N, Wu CH, O'Reilly KM, and Wilson D. 2015. New routes to phylogeography: A Bayesian structured coalescent approximation. *PLoS Genet.* **11**, e1005421.

De Maio N, Wu CH, and Wilson DJ. 2016. SCOTTI: Efficient reconstruction of transmission within outbreaks with the structured coalescent. *PLoS Comput. Biol.* **12**, e1005130.

de Man TJB and Limbago BM. 2016. SSTAR, a stand-alone easy-to-use antimicrobial resistance gene predictor. *mSphere* **1**, e00050-15.

Deschamps S, Mudge J, Cameron C, Ramaraj T, Anand A, Fengler K, Hayes K, Llaca V, Jones TJ, and May G. 2016. Characterization, correction and *de novo* assembly of an Oxford Nanopore genomic dataset from *Agrobacterium tumefaciens*. *Sci. Rep.* **6**, 28625.

Didelot X, Fraser C, Gardy J, and Colijn C. 2017. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. *Mol. Biol. Evol.* **34**, 997–1007.

Didelot X, Gardy J, and Colijn C. 2014. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol. Biol. Evol.* **31**, 1869–1879.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, *et al.* 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138.

Eybpoosh S, Haghdoost AA, Mostafavi E, Bahrampour A, Azadmanesh K, and Zolala F. 2017. Molecular epidemiology of infectious diseases. *Electron. Physician* **9**, 5149–5158.

Feil EJ, Li BC, Aanensen DM, Hanage WP, and Spratt BG. 2004. eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* **186**, 1518–1530.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, *et al.* 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, Mathers A, Norris R, Navickaite I, Eaton C, *et al.* 2017. Resolving plasmid structures in *Enterobacteriaceae* using the MinION nanopore sequencer: Assessment of MinION and MinION/Illumina hybrid data assembly approaches. *Microb. Genom.* **3**, e000118.

Gong L, Huang YT, Wong CH, Chao WC, Wu ZY, Wei CL, and Liu PY. 2018. Culture-independent analysis of liver abscess using nanopore sequencing. *PLoS One* **13**, e0190853.

Goordial J, Altshuler I, Hindson K, Chan-Yam K, Marcolefas E, and Whyte LG. 2017. *In situ* field sequencing and life detection in remote (79°26'N) Canadian high arctic permafrost ice wedge microbial communities. *Front. Microbiol.* **8**, 2594.

Gorrie CL, Mirceta M, Wick RR, Judd LM, Wyres KL, Thomson NR, Strugnell RA, Pratt NF, Garlick JS, Watson KM, *et al.* 2018. Antimicrobial resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition

in the referring hospital. *Clin. Infect. Dis.* **11**, 589.

Green ED. 2001. Strategies for the systematic sequencing of complex genomes. *Nat. Rev. Genet.* **2**, 573–583.

Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, and Holmes EC. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332.

Guthrie JL and Gardy JL. 2017. A brief primer on genomic epidemiology: Lessons learned from *Mycobacterium tuberculosis*. *Ann. N. Y. Acad. Sci.* **1388**, 59–77.

Hall M, Woolhouse M, and Rambaut A. 2015. Epidemic reconstruction in a phylogenetics framework: Transmission trees as partitions of the node set. *PLoS Comput. Biol.* **11**, e1004613.

Hawkey J, Ascher DB, Judd LM, Wick RR, Kostoulias X, Cleland H, Spelman DW, Padiglione A, Peleg AY, and Holt KE. 2018. Evolution of carbapenem resistance in *Acinetobacter baumannii* during a prolonged infection. *Microb. Genom.* **4**, 148.

Heather JM and Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8.

Hyeon JY, Li S, Mann DA, Zhang S, Li Z, Chen Y, and Deng X. 2017. Quasi-metagenomics and realtime sequencing aided detection and subtyping of *Salmonella enterica* from food samples. *Appl. Environ. Microbiol.* **84**, e02340-17.

Imai K, Tarumoto N, Runtuwene LR, Sakai J, Hayashida K, Eshita Y, Maeda R, Tuda J, Ohno H, Murakami T, *et al.* 2018. An innovative diagnostic technology for the codon mutation C580Y in *kelch13* of *Plasmodium falciparum* with MinION nanopore sequencer. *Malar. J.* **17**, 217.

Inouye M, Dashnow H, Schultz MB, Zobel J, and Holt KE. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6**, 90.

Iwai H, Kato-Miyazawa M, Kirikae T, and Miyoshi-Akiyama T. 2015. CASTB (the comprehensive analysis server for the *Mycobacterium tuberculosis* complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. *Tuberculosis* **95**, 843–844.

Jacob F and Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.

Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, *et al.* 2017. CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573.

Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, *et al.* 2012. Ribosomal multilocus sequence typing: Universal characterization of bacteria from domain to strain. *Microbiology* **158**, 1005–1015.

Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, and Ferguson N. 2014. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput. Biol.* **10**, e1003457.

Judge K, Hunt M, Reuter S, Tracey A, Quail MA, Parkhill J, and Peacock SJ. 2016. Comparison of bacterial genome assembly software for MinION data and their applicability to medical microbiology. *Microb. Genom.* **2**, e000085.

Kerkhof LJ, Dillon KP, Häggblom MM, and McGuinness LR. 2017. Profiling bacterial communities by MinION sequencing of ribosomal operons. *Microbiome* **5**, 116.

Kilbourne ED. 1973. The molecular epidemiology of influenza. *J. Infect. Dis.* **127**, 478–487.

Kilianski A, Roth PA, Liem AT, Hill JM, Willis KL, Rossmaier RD, Marinich AV, Maughan MN, Karavis MA, Kuhn JH, *et al.* 2016. Use of unamplified RNA/cDNA-hybrid nanopore sequencing for rapid detection and characterization of RNA viruses. *Emerg. Infect. Dis.* **22**, 1448–1451.

Klemm E and Dougan G. 2016. Advances in understanding bacterial pathogenesis gained from whole-genome sequencing and phylogenetics. *Cell Host Microbe* **19**, 599–610.

Klinkenberg D, Backer JA, Didelot X, Colijn C, and Wallinga J. 2017. Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks. *PLoS Comput. Biol.* **13**, e1005495.

Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, *et al.* 2015. Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* **15**, 141–161.

Lapierre P and Gogarten JP. 2009. Estimating the size of the bacterial pan-genome. *Trends Genet.* **25**, 107–110.

Lau MSY, Marion G, Streftaris G, and Gibson G. 2015. A systematic Bayesian integration of epidemiological and genetic data. *PLoS Comput. Biol.* **11**, e1004633.

Loman NJ, Quick J, and Simpson JT. 2015. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat. Methods* **12**, 733–735.

Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, *et al.* 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**, 3140–3145.

Maiden MCJ, van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, and McCarthy ND. 2013. MLST revisited: The gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* **11**, 728–736.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

McArthur AG and Tsang KK. 2017. Antimicrobial resistance surveillance in the genomic age. *Ann. N. Y. Acad. Sci.* **1388**, 78–91.

Metzker ML. 2010. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**, 31–46.

Mollentze N, Nel LH, Townsend S, le Roux K, Hampson K, Haydon DT, and Soubeyrand S. 2014. A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases

from space-time-genetic data. *Proc. Biol. Sci.* **281**, 20133251–20133251.

Morelli MJ, Thébaud G, Chadœuf J, King DP, Haydon DT, and Soubeyrand S. 2012. A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. *PLoS Comput. Biol.* **8**, e1002768.

Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemska O, Isbandi M, Thomas AD, Ali R, Sharma K, Kyrpides NC, *et al.* 2017. Genomes online database (GOLD) v.6: Data updates and feature enhancements. *Nucl. Acids Res.* **45**, D446–D456.

Nakano M, Komatsu J, Matsuura SI, Takashima K, Katsura S, and Mizuno A. 2003. Single-molecule PCR using water-in-oil emulsion. *J. Biotechnol.* **102**, 117–124.

Numminen E, Chewapreecha C, Sirén J, Turner C, Turner P, Bentley SD, and Corander J. 2014. Two-phase importance sampling for inference about transmission trees. *Proc. Biol. Sci.* **281**, 20141324–20141324.

Pritchard JK, Stephens M, and Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, Nair S, Neal K, Nye K, Peters T, *et al.* 2015. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16**, 7.

Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, *et al.* 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276.

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, *et al.* 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232.

Rames E and Macdonald J. 2018. Evaluation of MinION nanopore sequencing for rapid enterovirus genotyping. *Virus Res.* **252**, 8–12.

Risse J, Thomson M, Patrick S, Blakely G, Koutsovoulos G, Blaxter M, and Watson M. 2015. A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. *GigaScience* **4**, 60.

Romero-Severson E, Skar H, Bulla I, Albert J, and Leitner T. 2014. Timing and order of transmission events is not directly reflected in a pathogen phylogeny. *Mol. Biol. Evol.* **31**, 2472–2482.

Rowe W, Baker KS, Verner-Jeffreys D, Baker-Austin C, Ryan JJ, Maskell D, and Pearce G. 2015. Search engine for antimicrobial resistance: A cloud compatible pipeline and web interface for rapidly detecting antimicrobial resistance genes directly from sequence data. *PLoS One* **10**, e0133492.

Russell JA, Campos B, Stone J, Blosser EM, Burkett-Cadena N, and Jacobs JL. 2018. Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: A field-forward biosurveillance protocol. *Sci. Rep.* **8**, 5417.

Sanger F, Nicklen S, and Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, Khan AM, Woodford N, Saunders NJ, Wain J, *et al.* 2017. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* **72**, 104–114.

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, *et al.* 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**, 250–255.

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, and Church GM. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732.

Simner PJ, Antar AAR, Hao S, Gurtowski J, Tamma PD, Rock C, Opene BNA, Tekle T, Carroll KC, Schatz MC, *et al.* 2018. Antibiotic pressure on the acquisition and loss of antibiotic resistance genes in *Klebsiella pneumoniae*. *J. Antimicrob. Chemother.* **37**, 1288.

Staden R. 1979. A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res.* **6**, 2601–2610.

Steiner A, Stucki D, Coscolla M, Borrell S, and Gagneux S. 2014. KvarQ: Targeted and direct variant calling from fastq reads of bacterial genomes. *BMC Genomics* **15**, 881.

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, and Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624.

Tettelin H, Riley D, Cattuto C, and Medini D. 2008. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477.

Traynor BJ. 2009. The era of genomic epidemiology. *Neuroepidemiology* **33**, 276–279.

Votintseva AA, Bradley P, Pankhurst L, del Ojo Elias C, Loose M, Nilgiriwala K, Chatterjee A, Smith EG, Sanderson N, Walker TM, *et al.* 2017. Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *J. Clin. Microbiol.* **55**, 1285–1298.

Wick RR, Judd LM, Gorrie CL, and Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* **3**, e000132.

Wick RR, Schultz MB, Zobel J, and Holt KE. 2015. Bandage: Interactive visualization of *de novo* genome assemblies. *Bioinformatics* **31**, 3350–3352.

Worby CJ, Lipsitch M, and Hanage WP. 2014. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Comput. Biol.* **10**, e1003549.

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al.* 2009. A

phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056–1060.

Xia Y, Li AD, Deng Y, Jiang XT, Li LG, and Zhang T. 2017. MinION nanopore sequencing enables correlation between resistome phenotype and genotype of coliform bacteria in municipal sewage. *Front. Microbiol.* **8**, 2105.

Yin X, Jiang XT, Chai B, Li L, Yang Y, Cole JR, Tiedje JM, and Zhang T. 2018. ARGs-OAP v2.0 with an expanded SARG database and hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics* **535**, 435.

You Y, Kou Y, Niu L, Jia Q, Liu Y, Davies MR, Walker MJ, Zhu J, and Zhang J. 2018. Complete genome sequence of a *Streptococcus pyogenes* serotype M12 scarlet fever outbreak isolate from China, compiled using Oxford Nanopore and Illumina sequencing. *Genome Announc.* **6**, e00389-18.

Ypma RJF, van Ballegooijen WM, and Wallinga J. 2013. Relating phylogenetic trees to transmission trees of infectious disease outbreaks. *Genetics* **195**, 1055–1062.

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, and Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644.