Fighting Outbreaks with Bacterial Genomics: Case Review and Workflow Proposal

M.K. Cheung H.S. Kwan

School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong

Key Words

Bioinformatics • Epidemiology • Next-generation • NGS • Outbreak genomics • Sequencing

Abstract

Background: Disease outbreak investigation is a key aspect of public health. Whole-genome sequencing of bacterial pathogen based on new generation high-throughput sequencing technologies has facilitated outbreak investigations recently. Whilst the approach has become more affordable and accessible to research and clinical laboratories, a system for adequate and efficient analyses of genome data in the context of bacterial outbreak investigations is missing.

Methods: We performed a literature review of timely genomic investigations performed during the course of bacterial outbreaks that are based on new generation sequencing technologies. Currently available bioinformatics tools for genomic analyses are also reviewed here. **Results:** Genomic investigations in early stages of bacterial outbreaks have shown to provide timely information on evolutionary origin, transmission route, pathogenic potential, and resistance information of the outbreak strains and allow development of strain-specific typing methods. A systematic genomic analytical workflow is proposed here for the first time to facili-

tate efficient extraction of epidemiologically useful information from genome data of bacterial pathogens in future bacterial outbreak investigations. **Conclusion:** With the continuous reduction of genome sequencing cost and development of user-friendly analytical tools, it is expected that high-throughput genome sequencing will be applied routinely for timely genomic analysis in bacterial outbreaks in the near future.

Copyright © 2012 S. Karger AG, Basel

Introduction

Disease outbreaks represent one of the major public health threats worldwide. In 2010, there were 857 reported foodborne outbreaks in the US (<http://wwwn.cdc.gov/foodborneoutbreaks>). Timely identification and characterization of the pathogen during onset of a deadly outbreak, such as understanding its pathogenic potential, antimicrobial resistance and route of transmission, could save lives. Pulse field gel electrophoresis (PFGE) is the current gold standard for molecular typing of bacterial pathogens. However, inter-laboratory pattern comparisons could be difficult and subjective [1]. More importantly, the approach lacks resolution for closely related

isolates and is not well-suited for detection of novel strains. On the other hand, conventional strain characterization methods such as virulence assays and antimicrobial susceptibility tests are time-consuming and labor-demanding [2].

Whole-genome sequencing (WGS) takes into account all the genetic information within a genome and is able to provide the ultimate resolution possible and allows the discovery of the 'unknown unknowns' [3]. The recent advent of next- and third-generation high-throughput sequencing technologies has greatly improved the speed of WGS; draft genome of a bacterium can now be generated within days [4]. In addition, the introduction of affordable benchtop new-generation sequencers has made WGS of bacterial pathogens feasible for small and medium-sized research and clinical laboratories [5]. The state-of-the-art sequencing technologies have recently been employed to study bacterial outbreaks, both in a retrospective manner [e.g. 6, 7] and during the course of outbreaks [3, 8–13].

However, while sequencing a bacterial genome nowadays is not limiting, how to translate effectively and efficiently the complex genome data into information that could benefit population health has been one of the main topics of public health genomics. Unlike basic research studies in which a genome can be analyzed for months with various approaches before the final conclusions are drawn, understanding the route of transmission and the antibiotic resistance profile of the bacterial pathogen, for instance, could be the instant concerns in deadly outbreak investigations. Also, although crowdsourcing efforts could generate a massive amount of analytical outputs within short periods of time [3], redundant analyses may occur, and results that are not directly comparable may make interpretation of data even more vague. It is believed that a system for the analysis of genome data has to be established for the public health system [14].

In this review, we revisit real life examples of genomic investigations based on WGS and the use of new-generation sequencers during early stages of bacterial outbreaks that provided timely information on evolutionary relatedness, pathogenic potential and antimicrobial resistance of the bacterial pathogens. We then propose a systematic genomic analytical workflow that aims to facilitate efficient extraction of epidemiologically useful information from genome data of bacterial pathogens in the context of outbreak investigations.

Timely Genomic Investigations during Bacterial Outbreaks

2008 Canadian Listeriosis Outbreak

Listeriosis is an infection caused by the Gram-positive bacterium *Listeria monocytogenes*. In the summer of 2008, an outbreak of listeriosis associated with ready-to-eat meat products occurred in Canada. The outbreak was caused by a *L. monocytogenes* strain of serotype 1/2a and resulted in more than 20 deaths and over 50 illnesses (http://www.phac-aspc.gc.ca/alert-alerte/listeria_200808-eng.php). Using the next-generation 454 pyrosequencer [15], draft genome sequences of 2 *L. monocytogenes* isolates were generated within 3 days [10] (table 1). Comparative genomic analysis revealed a novel plasmid in the primary outbreak isolate, which harbors cadmium resistance genes, *cadA* and *cadC*, that are associated with resistance to sanitizers used in food-processing facilities [16]. A novel *Listeria* phage and a genetic island which is unique among all other sequenced *L. monocytogenes* isolates and encodes putative translocation, resistance, and regulatory factors were also found in the primary outbreak isolate. Virulence factor investigations revealed the presence of an intact internalin-encoding *inlA* locus, which plays a role in the promotion of mammalian host cell invasion [17], and some additional internalin-like loci which may partly account for the pathogenicity of the outbreak strain. Phylogenetic analysis based on whole-genome alignment indicated that the outbreak isolates belong to clonal complex 8 within lineage II and are most closely related to strain EGDe isolated in 1924 [18]. The analysis demonstrated that lineage II strains of *L. monocytogenes* can also cause a large outbreak of severe invasive disease, despite the fact that listeriosis outbreaks are usually caused by strains belonging to serotype 4b in lineage I [19]. The study represents one of the first attempts to use next-generation DNA sequencing technology in an ongoing bacterial outbreak investigation and provides a proof-of-concept that the approach could offer real-time responses to bacterial outbreaks.

Multidrug-Resistant *Acinetobacter baumannii*

Military patients from Iraq and Afghanistan are often colonized with multidrug-resistant *Acinetobacter baumannii* (MDRAB) strains, which can subsequently cause nosocomial infections in civilian patients and healthcare workers [20]. In 2008, a hospital outbreak of MDRAB occurred in the UK, in which isolates of the pathogen were recovered from 2 civilian patients following admission of 4 military patients colonized with MDRAB in the same

Table 1. Examples of timely bacterial genomics during outbreaks

Outbreak	Causative agent	Sequencing platform	Genome size, Mb	Draft genome	Virulence profile	Resistance profile	Phylogenetic relatedness	New typing method	Source tracking	Transmission route	Reference
Listeriosis	<i>Listeria monocytogenes</i>	454 GS FLX Standard	3.03	<3 days	X	X	X				[10]
MDRAB	<i>Acinetobacter baumannii</i>	454 GS FLX Titanium	4.12	NA			X			X	[11]
Cholera	<i>Vibrio cholerae</i>	PacBio RS Sequencer	NA	NA			X		X		[9]
HUS	<i>Escherichia coli</i>	454 GS FLX Titanium	5.31	NA	X	X	X				[8]
		Ion Torrent PGM	5.45	<3 days			X				[12]
		PacBio RS Sequencer	5.51	<3 days	X		X				[13]
		Ion Torrent PGM, Illumina HiSeq	5.49	3 days ¹	X	X	X	X			[3]

¹ Using Ion Torrent PGM. NA = Not available.

unit [11]. PFGE and variable number tandem repeat analyses grouped the outbreak isolates into European clone 1 [21] but generated indistinguishable profiles among them. Identical antimicrobial resistance profiles were also obtained for all the 6 outbreak isolates. The transmission events thus remained unclear. By identifying 3 well-validated single nucleotide polymorphisms (SNPs) in draft genomes of the outbreak isolates using 454 pyrosequencing and subsequent mapping of the data to the complete genome sequence of a reference strain also in European clone 1, one of the military isolates was identified as bearing the ancestral genotype at all the 3 SNP loci, suggesting a transmission route from the wound of that military patient to the respiratory track of the civilian patient in the adjacent bed. The study highlights the potential of using genome sequencing to examine transmission events in bacterial outbreaks.

2010 Haitian Cholera Outbreak

Cholera, an acutely dehydrating diarrheal disease that could be deadly, is caused by the Gram-negative bacterium *Vibrio cholerae* [22]. In late 2010, a large outbreak of cholera started in Haiti, causing over 6,600 deaths and 0.47 million cases (http://new.paho.org/disasters/index.php?option=com_content&task=view&id=1423&Itemid=1). Cholera had not been epidemic in Haiti for at least 100 years, and the origin of the outbreak was controversial [23]. Using the third-generation PacBio RS sequencing system [24], genome sequences of 2 clinical outbreak isolates of *V. cholerae* and 3 historical isolates from other

regions of the world were determined [9]. Phylogenetic analyses based on core SNPs placed the outbreak isolates in group V of the seventh-pandemic group [25] and revealed a close relationship to the South Asian isolates from Bangladesh. Investigations on 20 previously described hyper-recombinant chromosomal elements [26] in the 5 *V. cholerae* genomes revealed structural variations in 3 regions: superintegron, VSP-2 and SXT, which in turn suggested a closer relationship of the 2 outbreak isolates to the Bangladesh strain CIRS101, isolated in 2002, than to the other Bangladesh strain M4, isolated in 2008. Detailed comparative genomic analysis of the 2 outbreak isolates with 3 additional outbreak isolates from the Centers for Disease Control and Prevention (CDC) indicated that the outbreak is clonal. In addition, the distant phylogenetic relationship between the Haitian outbreak isolates and those circulating in Latin America and the US Gulf Coast showed that the cholera epidemic in Haiti is not associated with climatic events, unlike some other cases of cholera epidemic [27]. Instead, the close relationship of the Haitian outbreak isolates with historical South Asian isolates from Bangladesh suggested that the Haitian epidemic is probably due to human activity that brought the *V. cholerae* strain from a distant geographic source to Haiti. The study represents the first application of third-generation sequencing in an ongoing bacterial outbreak and provides policy implications for public health officials on consideration of measures for controlling cholera [28].

2011 German *Escherichia coli* O104:H4 Outbreak

In mid 2011, a large outbreak of diarrhea with associated hemolytic-uremic syndrome (HUS) started in Germany, causing nearly 4,000 reported cases and over 40 deaths (http://www.ecdc.europa.eu/en/healthtopics/escherichia_coli/Pages/index.aspx). Diarrhea associated with HUS is usually caused by enterohemorrhagic *E. coli* (EHEC) of serotype O157:H7 [29]. However, the outbreak strain was serotyped to be O104:H4, a rare serotype of Shiga toxin-producing *E. coli* that had only been linked to sporadic cases of HUS [30]. The outbreak was also characterized by a higher incidence in adults, a higher incidence of HUS and a predominance of female patients among HUS cases, which are all unusual [31]. Using various next- and third-generation sequencers, draft genomes of 10 outbreak isolates, as well as some related and historical isolates, were made available within days [3, 8, 12, 13] (table 1). A crowd-sourcing effort, in which analyses of the publicly released genome data were outsourced to bioinformaticians worldwide, was also set in motion in the early stage of the outbreak to gather analytical outputs rapidly (<https://github.com/ehc-outbreak-crowd-sourcing/BGI-data-analysis/wiki>).

Genomic comparisons of the outbreak strain with all previously sequenced complete genomes of *E. coli* revealed the enteroaggregative *E. coli* (EAEC) strain 55989, isolated in the late 1990s [32], to be the closest relative of the outbreak strain [3]. The result was confirmed by multi-locus sequence analysis (MLSA) [8] and whole-genome phylogenetic analysis [12]. Genes encoding virulence factors that are typical of EAEC were also found in the genomes of the outbreak strain [3]. However, a Shiga toxin-encoding phage, highly similar to a phage from EHEC O157:H7, was identified in the outbreak strain [8, 13], although the locus of enterocyte effacement pathogenicity island (PAI) which is typical in EHEC [29] was found missing [3, 8, 13]. Comparison of genome sequences from the outbreak isolates derived from different patients suggested a stable genome of the outbreak strain during its infection in different hosts [8] as well as a clonal nature of the outbreak [13]. Two large plasmids were revealed in the outbreak strain; the larger plasmid is highly similar to the pEC_Bactec plasmid that harbors extended-spectrum beta-lactamase genes of the TEM-1 and CTX-M-15 classes, and the smaller one is similar to the pAA plasmid found in EAEC 55989 but contains a rare type of aggregative adherence fimbria, AAF/I, instead of the more common AAF/III type [3, 8, 13]. Based on the characteristic presence of the AAF/I gene cluster, strain-specific diagnostic kits were designed and released

for outbreak isolate identification 5 days after the release of the genome sequence data [3]. Genes involved in mercury resistance, tellurium resistance and antimicrobial resistance were also identified in the outbreak strain [8]. The scenario represents the largest sequencing effort on a bacterial pathogen using different high-throughput platforms in an outbreak investigation at the moment.

Genomic Analytical Workflow for Bacterial Outbreaks

As illustrated in the case examples reviewed above, timely genomic investigations in early stages of bacterial outbreaks could rapidly provide information on the evolutionary position, transmission route, pathogenic potential, and resistance information of the outbreak strains and allow development of quick strain-specific typing methods. In order to facilitate future investigations of bacterial outbreaks, here we propose a systematic genomic analytical workflow (fig. 1), with suggestions of some ready-to-use tools and pipelines (table 2), that is specifically designed to facilitate efficient extraction of epidemiologically useful information from genome data of bacterial pathogens in the context of outbreak investigations. It should be noted that since the analysis of new-generation sequencing data is a fast-evolving field in which analytical tools are constantly being improved and developed [33], tools and pipelines listed here are not aimed to be exhaustive.

Genome Sequencing, Assembly and Annotation

When a bacterial outbreak occurs, isolation of the causative pathogen in pure culture and extraction of genomic DNA first take place. High-throughput sequencing of the whole bacterial genome is then feasible using any of the state-of-the-art new-generation sequencing platforms. Currently available next-generation sequencing platforms include 454 [15], Illumina [34] and SOLiD [35]; while third-generation sequencing platforms include Ion Torrent [36] and PacBio [24]. Various platforms employ different sequencing chemistry, have different sequencing throughput, generate sequence reads of different lengths, and are subject to different intrinsic errors. A comparison of these platforms has been reviewed recently [4].

Raw sequencing reads have to be assembled after genome sequencing. The purpose of genome assembly is to group the fragments of a DNA sequence into contigs, and then contigs into scaffolds, to reconstruct the original

Fig. 1. Proposed genomic analytical workflow for bacterial outbreaks using new-generation sequencing technologies. When a bacterial outbreak occurs, whole genome of the bacterial pathogen is first sequenced on next-generation sequencing (NGS) or third-generation sequencing (TGS) platforms. After genome assembly and genome annotation, comparative genomic analyses could be carried out to investigate the pathogenic potential and antimicrobial resistance of the outbreak strain. Targets of study include horizontally acquired elements such as plasmid, prophage and genomic island; SNPs and insertions and deletions (indels), and synteny; as well as searching against online databases including the virulence factor database (VFDB) and the antibiotic resistance

genes database (ARDB). Virulence and resistance profiles of the outbreak strain obtained would assist treatment decision making and design of preventive measures, and any genetic elements unique to the outbreak strain could allow establishment of rapid strain-specific typing methods. On the other hand, phylogenetic relatedness of the outbreak strain with other related strains could be elucidated using core SNPs or core genome alignment data. Phylogenetic identity of the outbreak strain, whether novel or known, would affect treatment decision and preventive measure implementation, and its evolutionary relationship with other strains would allow source tracking and understanding of the transmission route of outbreak.

DNA sequence. Genomes could be assembled using either the de novo or mapping approach [37]. The de novo approach is more mathematically complex and computationally demanding and is usually employed on reconstructing genomes that have never been sequenced before, while the mapping approach allows quicker assembly but is only feasible when a closely related reference sequence is available. Many genome assemblers are currently available, examples for de novo assembly include Velvet [38], MIRA 3 [39] and Allora; and those for mapping assembly include BWA [40], SOAP2 [41] and Bowtie [42] (table 2). Bao et al. [43] recently compared the performance of various genome assemblers and suggested guidelines for tool selection under varying conditions. Finishing of genome assembly often requires a time-consuming gap-closure process. However, it has been shown that unfinished draft genomes of pathogens are informa-

tive enough in the context of emerging bacterial outbreaks [44]. The utilization of draft genomes for rapid outbreak investigations is therefore generally recommended.

After the genome is assembled, whether in draft or completed form, genome annotation follows. Genome annotation is a process of adding biological interpretations to DNA sequences and involves gene prediction and functional annotation. In gene prediction, a gene finder is applied to the genome sequence, producing a set of predicted protein-coding genes. Subsequent functional annotation attaches biological information to the set of predictions via sequence similarity searches against available databases. Various tools and pipelines have been developed for automatic genome annotation, including RAST [45], Gent [46] and DIYA [47]. However, none of them is capable of generating a functional annotation

Table 2. List of selected ready-to-use tools and servers for genomic analyses in bacterial outbreaks

Process	Tool/server	Notes	Reference
De novo genome assembly	Velvet	Free stand-alone software; for NGS and Ion torrent	[38]
	MIRA 3	Free stand-alone software; for NGS and Ion torrent	[39]
	Allora	Free stand-alone software; for PacBio RS sequencer	NA
Mapping assembly	BWA	Free stand-alone software; for NGS and Ion torrent	[40]
	SOAP2	Free stand-alone software; for NGS and Ion torrent	[41]
	Bowtie	Free stand-alone software; for NGS and Ion torrent	[42]
Genome annotation	RAST	Free online server	[45]
	GenDB 2.2	Free stand-alone software	[46]
	DIYA	Free stand-alone software	[47]
Prophage detection	PHAST	Free online server	[53]
	Prophage Finder	Free online server	[54]
	Prophinder	Free online server	[55]
Genomic island detection	IslandViewer	Free online server	[61]
	MobilomeFINDER	Free stand-alone software	[62]
	Alien_Hunter	Free stand-alone software	[63]
SNP/indel calling	Samtools	Free stand-alone software	[65]
	GATK	Free stand-alone software	[66]
	SOAPSnp	Free stand-alone software	[41]
Whole-genome alignment	Mauve	Free stand-alone software	[67]
	MUMmer 3.0	Free stand-alone software; also for SNP / indel calling	[68]
	ACT	Free stand-alone software	[69]
Circular genome visualization	BRIG	Free stand-alone software	[70]
	DNAPlotter	Free stand-alone software; also for linear visualization	[71]
	CGView	Free stand-alone software	[72]
VF and AR gene searching	VFDB	Free online database of VFs and some PAIs	[73]
	ARDB	Free online database of AR genes	[74]
Phylogeny reconstruction	MEGA 5	Free stand-alone software	[80]
	RAXML	Free stand-alone software	[81]
	PhyML 3.0	Free stand-alone software	[82]
Others	GenePRIMP	Free online server for manual genome curation	[48]
	Panseq	Free online server for core genome/SNP alignment construction	[79]

VF = Virulence factor; AR = antimicrobial resistance.

without any error and thus manual curation, in which experts are deployed to re-examine the prediction set, is always required. This could be assisted with GenePRIMP, a web-based post-processing pipeline that identifies erroneously predicted genes and which has been used by the US Department of Energy Joint Genome Institute on over 300 genomes [48].

Analyses of Pathogenicity and Antimicrobial Resistance

Rapid identification and characterization of pathogenicity-related and antimicrobial resistance genes is crucial in order to quickly get information on what the emerging pathogen is capable of and to assure susceptibility to drugs of choice. These kinds of genes are usually

harbored on plasmids, prophages and genomic islands (GIs), which are acquired via horizontal gene transfer.

Plasmids are self-replicating pieces of extrachromosomal DNA that usually carry virulence-related and antimicrobial resistance genes. For instance, acquisition of the virulence plasmid pINV makes enteroinvasive *E. coli* invasive [49], and the presence of plasmid-encoded Qnr protein confers quinolone resistance in various bacterial genera [50]. Prophages are bacteriophages that have physically integrated into genomes of their preferred bacterial host [51]. The presence of prophage sequences may allow some bacteria to become pathogenic or to acquire antimicrobial resistance. SPC-P1, for example, is a pathogenicity-associated prophage of *Salmonella enterica* serovar Paratyphi C [52]. Available tools that allow prophage

identification include PHAST [53], Prophage Finder [54] and Prophinder [55]. Performance of these tools has been compared in a recent review [53].

GIs refer to horizontally transferred gene clusters that are typically 10–200 kb in size [56]. Several classes of GIs are recognized according to their gene content, including PAIs, resistance islands, secretion islands, and metabolic islands [57]. PAIs carry genes coding for virulence factors such as toxins and adhesins that confer pathogenicity to bacteria and resistance islands harbor genes related to antimicrobial resistance and metal resistance [58]. For instance, virulence-related genes are found within the *Francisella* PAI of *F. tularensis* LVS [59], and the presence of *Salmonella* GI confers multidrug resistance to *S. Typhimurium* DT104 [60]. Ready-to-use tools for GI prediction include IslandViewer [61], MobilomeFinder [62] and Alien_Hunter [63].

Besides acquisition of virulence- and resistance-related gene elements via horizontal gene transfer, point mutations and DNA rearrangements might also contribute to pathogenicity and antimicrobial resistance of pathogens. An example includes SNPs in *gyrA* that could confer bacterial resistance against quinolones and fluoroquinolones [64]. Commonly used tools for point mutation detection include Samtools [65], GATK [66] and SOAPsnp [41]. On the other hand, conservation of synteny among genomes could be identified and analyzed using whole-genome aligners such as Mauve [67], MUMmer [68] and ACT [69] or circular genome viewers such as BRIG [70], DNAPlotter [71] and CGView [72].

Several online databases contain collections of pathogenicity-related and antimicrobial resistance genes that could also facilitate rapid identification of such elements in the genomes of bacterial pathogens. For instance, the virulence factor database contains sequences of 418 experimentally demonstrated virulence factors and 2,353 virulence-factor-related genes from 24 genera of medically important bacterial pathogens [73], and the antibiotic resistance genes database contains sequences of 380 types of antimicrobial resistance genes that encode resistance to 249 antibiotics [74].

Elucidation of Phylogenetic Relatedness

Apart from knowing the pathogenic potential and antimicrobial resistance profile of outbreak strains, timely information on their phylogenetic relatedness to other strains is equally important to facilitate source tracking and understand their evolutionary positions and routes of transmission. As illustrated in the German *E. coli* scenario, various approaches could be employed to address

the issue, these include the average nucleotide identity method [3], core SNPs (<http://bacpathgenomics.wordpress.com/2011/06/15/snp-base-phylogeny-confirms-similarity-of-e-coli-outbreak-to-eaec-ec55989/>), MLSA [8], core genome open reading frames [12], core genome alignment [13], and alignment-free approach [75]. While MLSA is based on information from only 7 housekeeping genes, accuracy of alignment-based methods relies heavily on the sequence alignment, and alignment-free methods are sometimes opposed due to the lack of biological background [76]. Nevertheless, among the approaches, those based on the entire core genome and concatenated SNP sets seem to be more common and well developed [e.g. 77, 78]. Core genome and core SNP data could be extracted from the input genomes using the online tool, Panseq [79], which could also automatically create input files for phylogenetic tree building programs such as MEGA [80], RAxML [81] and PhyML [82]. The approach is thus more accessible, especially to non-bioinformaticians.

Conclusion and Perspectives

The genome of a bacterium contains too much information one could extract from and make sense of, and which could fit for various different purposes [83]. In the context of disease outbreak investigations, understanding the pathogenic potential and drug susceptibility of the pathogen, developing rapid strain-specific typing methods, knowing the route of transmission, and tracking the source of outbreak should be of top priorities for outbreak control. In order to facilitate efficient and targeted extraction of such epidemiologically useful information from genome data of bacterial pathogens in future bacterial outbreak investigations, a genomic analytical workflow is developed here. In case of a bacterial outbreak, studying elements such as plasmids, prophages and GIs on which pathogenicity-related and antimicrobial resistance genes are usually located allows the investigator to quickly get information on what the emerging bacterial pathogen is capable of, thus allows design of preventive measures, and to assist antibiotic treatment decision making so that susceptibility to drugs of choice could be assured. Strands of DNA unique to the outbreak strain may also be found in studying these elements, allowing establishment of rapid strain-specific typing methods that could in turn control and prevent further spread of the disease. On the other hand, by studying core SNPs or core genome alignment, phylogenetic related-

ness of the outbreak strain could be uncovered. Revealing the phylogenetic identity of an outbreak strain could affect treatment decision and preventive measure implementation: when the outbreak strain is identified to be a known one, effective antibiotics and preventive measures previously employed could be simply re-adopted. And, an evolutionary profile of the outbreak strain with other related strains would allow source tracking and understanding of the transmission route of outbreak.

Although WGS with new-generation high-throughput sequencers has revolutionized genomic investigations during bacterial outbreaks, there is no simple path from genome sequence to understanding of the virulence or transmissibility [84]. For instance, although the most parsimonious transmission route was suggested in the case of the MDRAB outbreak, the exact time and mode of transmission remained undetermined [11]. For the 2010 Haitian cholera outbreak, additional epidemiologic investigations are required to understand how exactly the South Asian *V. cholerae* strain was introduced to Haiti [9]. And, it remained unclear why the incidence rates of HUS in adult and female are unusually higher in the 2011 German *E. coli* outbreak. Also, as the current WGS approach relies on isolation of pure cultures, it is infeasible to be directly applied on clinical samples in which a mixture of pathogen(s) and the normal microbiota is present [85]. Possible culture-independent approaches that may be used to tackle the problem include metagenomic sequencing [86] and single cell genome sequencing [87]. However, these have to be tested in the clinical setting in the future before put into routine use.

The size and composition of the genomic database of bacterial strains is also crucial for subsequent biological interpretations, and a lack of representative members could result in biased or even wrong interpretation of data. Taking the 2011 German *E. coli* case as an example, EAEC 55989 had been identified as the closest relative of the outbreak strain before the genome sequence of the 2001 isolate from the HUS-associated *E. coli* collection [88] was made available. However, later phylogenetic analysis revealed a closer relationship of the outbreak strain to the 2001 isolate than to EAEC 55989 [75]. This example demonstrated how the availability of the *E. coli* collection facilitated phylogenetic grouping and also revealed the need of additional genome sequences from related strains. Indeed, apart from the benefits to ongoing outbreaks, an expanded genomic database of clinical bacterial isolates might also allow detection of outbreaks in advance [5].

Although the reduction in cost of WGS has made genomic analysis of bacterial pathogens more affordable to small and medium-sized clinical laboratories, many of these laboratories are currently still using conventional methods such as PFGE and instrumentations for genome sequencing are lacking. Also, although our genomic analytical workflow provides a clear and more focused direction of data analysis that is specifically designed for the purpose of outbreak investigations, a certain level of knowledge on bioinformatics is still required with the present set of available tools. As it is currently unrealistic to have dedicated bioinformatics specialists in every diagnostic laboratory, we expect an initial introduction of the WGS approach into country-level or regional core sequencing centers in which specialized technical expertise is present. After public health officials are more informed about genomics and more user-friendly bioinformatics tools become available, decentralization with genome sequencers in local public health laboratories and then diagnostic laboratories in hospitals across countries is expected.

For WGS to be used in routine disease outbreak investigations, the CDC and the European Centre for Disease Prevention and Control (ECDC) have to play important leading roles. For instance, they should coordinate regional and national meetings that bring together scientists, public health practitioners and the academia for discussions. Besides, it is important for the CDC and ECDC to coordinate and provide adequate training activities to public health professional to enrich their knowledge on genomics so that correct decisions could be made from the genomic information obtained. The CDC and ECDC should also establish and maintain an international database for central storage and sharing of genome data of bacterial pathogens for the purpose of disease outbreak investigations.

We are now in a new era of high-throughput, genome-based epidemiology. WGS would soon provide a cost-effective alternative to the conventional methods [10] or even replace them [11, 89]. It is therefore inevitable for public health laboratories to prepare for appropriate analysis and interpretation of genome data in the context of molecular epidemiology [10]. We move one step forward here by proposing a genomic analytical workflow that provides a clear and focused direction for efficient extraction of epidemiologically useful information from the complex genome data of bacterial pathogens for the purpose of outbreak investigations, and which should be able to facilitate and accelerate the revolution of outbreak genomics. In the near future, perhaps with the help of a

further reduction of genome sequencing cost and development of user-friendly and powerful analytical pipelines and tools, we can perform routine genomic investigations to fight against bacterial outbreaks.

Acknowledgements

This work is supported by RFCID CHP-PH-06 from Food and Health Bureau of Hong Kong SAR, China.

References

- Wattiau P, Boland C, Bertrand S: Methodologies for *Salmonella enterica* subsp. *enterica* subtyping: gold standards and alternatives. *Appl Environ Microbiol* 2011;77:7877–7885.
- Jorgensen JH, Ferraro MJ: Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin Infect Dis* 2009;49:1749–1755.
- Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, Chen W, Pu F, Peng Y, Li J, Xi F, Li S, Li Y, Zhang Z, Yang X, Zhao M, Wang P, Guan Y, Cen Z, Zhao X, Christner M, Kobbe R, Loos S, Oh J, Yang L, Danchin A, Gao GF, Song Y, Li Y, Yang H, Wang J, Xu J, Pallen MJ, Wang J, Aepfelbacher M, Yang R, *E. coli* O104:H4 Genome Analysis Crowd-Sourcing Consortium: open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med* 2011;365:718–724.
- Glenn TC: Field guide to next-generation DNA sequencers. *Mol Ecol Resour* 2011;11:759–769.
- Kupferschmidt K: Epidemiology: outbreak detectives embrace the genome era. *Science* 2011;333:1818–1819.
- Gardly JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P: Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011;364:730–739.
- Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao G, Meng J, Stones R: Identification of a salmonellosis outbreak by means of molecular sequencing. *N Engl J Med* 2011;364:981–982.
- Brzuszkiewicz E, Thürmer A, Schuldes J, Leimbach A, Liesegang H, Meyer FD, Boelter J, Petersen H, Gottschalk G, Daniel R: Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Enterotoxigenic *Escherichia coli* (EAHEC). *Arch Microbiol* 2011;193:883–891.
- Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK: The origin of the Haitian cholera outbreak strain. *N Engl J Med* 2011;364:33–42.
- Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel KM, Larios O, Allen V, Lee B, Nadon C: High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. *BMC Genomics* 2010;11:120.
- Lewis T, Loman NJ, Bingle L, Juma P, Weinstock GM, Mortiboy D, Pallen MJ: High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. *J Hosp Infect* 2010;75:37–41.
- Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H: Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* 2011;6:e22751.
- Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK: Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 2011;365:709–717.
- Zimmern RL, Khoury MJ: The impact of genomics on public health practice: the case for change. *Public Health Genomics* 2012;15:118–124.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgeson S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhiyani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomas A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376–380.
- Mullapudi S, Siletsky RM, Kathariou S: Heavy-metal and benzalkonium chloride resistance of *Listeria monocytogenes* isolates from the environment of turkey-processing plants. *Appl Environ Microbiol* 2008;74:1464–1468.
- Vázquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Domínguez-Bernal G, Goebel W, González-Zorn B, Wehland J, Kreft J: *Listeria* pathogenesis and molecular virulence determinants. *Clin Microbiol Rev* 2001;14:584–640.
- Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couvé E, de Daruvar A, Dehoux P, Domann E, Domínguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, García-del Portillo F, Garrido P, Gautier L, Goebel W, Gómez-López N, Hain T, Hauf J, Jackson D, Jones LM, Kaerst U, Kreft J, Kuhn M, Kunst F, Kurapkat G, Madueno E, Maitournam A, Vicente JM, Ng E, Nedjari H, Nordsiek G, Novella S, de Pablos B, Pérez-Díaz JC, Purcell R, Rammel B, Rose M, Schlueter T, Simoes N, Tierrez A, Vázquez-Boland JA, Voss H, Wehland J, Cossart P: Comparative genomics of *Listeria* species. *Science* 2001;294:849–852.
- Swaminathan B, Gerner-Smidt P: The epidemiology of human listeriosis. *Microbes Infect* 2007;9:1236–1243.
- Peleg AY, Seifert H, Paterson DL: *Acinetobacter baumannii*: emergence of a successful pathogen. *Clin Microbiol Rev* 2008;21:538–582.
- Dijkshoorn L, Auckan H, Gerner-Smidt P, Janssen P, Kaufmann ME, Garaizar J, Ursing J, Pitt TL: Comparison of outbreak and non-outbreak *Acinetobacter baumannii* strains by genotypic and phenotypic methods. *J Clin Microbiol* 1996;34:1519–1525.
- Sack DA, Sack RB, Nair GB, Siddique AK: Cholera. *Lancet* 2004;363:223–233.
- Enserink M: Haiti's outbreak is latest in cholera's new global assault. *Science* 2010;330:738–739.

- 24 Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S: Real-time DNA sequencing from single polymerase molecules. *Science* 2009; 323:133–138.
- 25 Lam C, Octavia S, Reeves P, Wang L, Lan R: Evolution of seventh cholera pandemic and origin of 1991 epidemic, Latin America. *Emerg Infect Dis* 2010;16:1130–1132.
- 26 Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, Taviani E, Jeon YS, Kim DW, Lee JH, Brettin TS, Bruce DC, Challacombe JF, Detter JC, Han CS, Munk AC, Chertkov O, Meincke L, Saunders E, Walters RA, Huq A, Nair GB, Colwell RR: Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci USA* 2009;106:15442–15447.
- 27 Constantin de Magny G, Murtugudde R, Sapiiano MR, Nizam A, Brown CW, Busalacchi AJ, Yunus M, Nair GB, Gil AI, Lanata CF, Calkins J, Manna B, Rajendran K, Bhat-tacharya MK, Huq A, Sack RB, Colwell RR: Environmental signatures associated with cholera epidemics. *Proc Natl Acad Sci USA* 2008;105:17676–17681.
- 28 Enserink M: No vaccines in the time of cholera. *Science* 2010;329:1462–1463.
- 29 Kaper JB, Nataro JP, Mobley HL: Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2004;2:123–140.
- 30 Bae WK, Lee YK, Cho MS, Ma SK, Kim SW, Kim NH, Choi KC: A case of hemolytic uremic syndrome caused by *Escherichia coli* O104:H4. *Yonsei Med J* 2006;47:437–439.
- 31 Scheut F, Møller Nielsen E, Frimodt-Møller J, Boisen N, Morabito S, Tozzoli R, Nataro JP, Caprioli A: Characteristics of the enteroaggregative Shiga toxin/verotoxin-producing *Escherichia coli* O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011. *Euro Surveill* 2011;16:pii 19889.
- 32 Mossoro C, Glaziou P, Yassibanda S, Lan NT, Bekondi C, Minssart P, Bernier C, Le Bouguénec C, Germani Y: Chronic diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome associated with HEP-2 adherent *Escherichia coli* in adults infected with human immunodeficiency virus in Bangui, Central African Republic. *J Clin Microbiol* 2002;40:3086–3088.
- 33 Nielsen R, Paul JS, Albrechtsen A, Song YS: Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet* 2011; 12:443–451.
- 34 Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al: Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456:53–59.
- 35 Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM: A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* 2008;18:1051–1063.
- 36 Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinkas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J: An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011;475:348–352.
- 37 Pop M: Genome assembly reborn: recent computational challenges. *Brief Bioinform* 2009;10:354–366.
- 38 Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;18:821–829.
- 39 Chevreaux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S: Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 2004;14:1147–1159.
- 40 Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–1760.
- 41 Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J: SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25:1966–1967.
- 42 Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.
- 43 Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ: Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet* 2011;56:406–414.
- 44 La Scola B, Elkarkouri K, Li W, Wahab T, Fournous G, Rolain JM, Biswas S, Drancourt M, Robert C, Audic S, Löfdahl S, Raoult D: Rapid comparative genomic analysis for clinical microbiology: the *Francisella tularensis* paradigm. *Genome Res* 2008;18:742–750.
- 45 Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O: The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008;9:75.
- 46 Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Pühler A: GenDB – an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* 2003;31:2187–2195.
- 47 Stewart AC, Osborne B, Read TD: DIYA: a bacterial annotation pipeline for any genomics lab. *Bioinformatics* 2009;25:962–963.
- 48 Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC: GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010;7:455–457.
- 49 Maurelli AT: Black holes, antiviral genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiol Lett* 2007;267:1–8.
- 50 Robicsek A, Jacoby GA, Hooper DC: The worldwide emergence of plasmid-mediated quinolone resistance. *Lancet Infect Dis* 2006;6:629–640.
- 51 Casjens S: Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 2003;49:277–300.
- 52 Zou QH, Li QH, Zhu HY, Feng Y, Li YG, Johnston RN, Liu GR, Liu SL: SPC-P1: a pathogenicity-associated prophage of *Salmonella* Paratyphi C. *BMC Genomics* 2010; 11:729.
- 53 Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS: PHAST: a fast phage search tool. *Nucleic Acids Res* 2011;39:W347–W352.
- 54 Bose M, Barber RD: Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol* 2006;6:223–227.
- 55 Lima-Mendez G, Van Helden J, Toussaint A, Leprieux R: Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* 2008;24:863–865.
- 56 Hacker J, Kaper JB: Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 2000;54:641–679.
- 57 Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H: Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 1997;23: 1089–1097.
- 58 Dobrindt U, Hochhut B, Hentschel U, Hacker J: Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2004;2:414–424.
- 59 Bröms JE, Lavander M, Meyer L, Sjöstedt A: IglG and IglI of the *Francisella tularensis* LVS. *Infect Immun* 2011;79:3683–3696.

- 60 Boyd DA, Peters GA, Ng L, Mulvey MR: Partial characterization of a genomic island associated with the multidrug resistance region of *Salmonella enterica* Typhimurium DT104. *FEMS Microbiol Lett* 2000;189:285–291.
- 61 Langille MG, Brinkman FS: IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 2009;25:664–665.
- 62 Ou HY, He X, Harrison EM, Kulasekara BR, Thani AB, Kadioglu A, Lory S, Hinton JC, Barer MR, Deng Z, Rajakumar K: MobilomeFINDER: web-based tools for in silico and experimental discovery of bacterial genomic islands. *Nucleic Acids Res* 2007;35:W97–W104.
- 63 Vernikos GS, Parkhill J: Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 2006;22:2196–2203.
- 64 Levy DD, Sharma B, Cebula TA: Single-nucleotide polymorphism mutation spectra and resistance to quinolones in *Salmonella enterica* serovar Enteritidis with a mutator phenotype. *Antimicrob Agents Chemother* 2004;48:2355–2363.
- 65 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
- 66 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The Genome Analysis Toolkit: a Map-Reduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
- 67 Darling AC, Mau B, Blattner FR, Perna NT: Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–1403.
- 68 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: Versatile and open software for comparing large genomes. *Genome Biol* 2004;5:R12.
- 69 Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: ACT: the Artemis Comparison Tool. *Bioinformatics* 2005;21:3422–3423.
- 70 Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA: BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.
- 71 Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J: DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009;25:119–120.
- 72 Stothard P, Wishart DS: Circular genome visualization and exploration using CGView. *Bioinformatics* 2005;21:537–539.
- 73 Yang J, Chen L, Sun L, Yu J, Jin Q: VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. *Nucleic Acids Res* 2008;36:D539–D542.
- 74 Liu B, M Pop: ARDB – Antibiotic Resistance Genes Database. *Nucleic Acids Res* 2009;37:D443–D447.
- 75 Cheung MK, Li L, Nong W, Kwan HS: 2011 German *Escherichia coli* O104:H4 outbreak: whole-genome phylogeny without alignment. *BMC Res Notes* 2011;4:533.
- 76 Schwarz RF, Fletcher W, Förster F, Merget B, Wolf M, Schultz J, Markowitz F: Evolutionary distances in the twilight zone – a rational kernel approach. *PLoS One* 2010;5:e15788.
- 77 DeLeo FR, Kennedy AD, Chen L, Bubeck Wardenburg J, Kobayashi SD, Mathema B, Braughton KR, Whitney AR, Villaruz AE, Martens CA, Porcella SF, McGavin MJ, Otto M, Musser JM, Kreiswirth BN: Molecular differentiation of historic phage-type 80/81 and contemporary epidemic *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 2011;108:18091–18096.
- 78 Kuroda M, Serizawa M, Okutani A, Sekizuka T, Banno S, Inoue S: Genome-wide single nucleotide polymorphism typing method for identification of *Bacillus anthracis* species and strains among *B. cereus* group species. *J Clin Microbiol* 2010;48:2821–2829.
- 79 Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, Thomas JE, Gannon VP: Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics* 2010;11:461.
- 80 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–2739.
- 81 Stamatakis A, Hoover P, Rougemont J: A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 2008;57:758–771.
- 82 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
- 83 Wren BW: Microbial genome analysis: insights into virulence, host adaptation and evolution. *Nat Rev Genet* 2000;1:30–39.
- 84 Anonymous: Outbreak genomics. *Nat Biotech* 2011;29:769.
- 85 Pallen MJ, Loman NJ: Are diagnostic and public health bacteriology ready to become branches of genomic medicine? *Genome Med* 2011;3:53.
- 86 Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J: Metagenomic pyrosequencing and microbial identification. *Clin Chem* 2009;55:856–866.
- 87 Yilmaz S, Singh AK: Single cell genome sequencing. *Curr Opin Biotechnol* 2012;23:437–443.
- 88 Mellmann A, Bielaszewska M, Köck R, Friedrich AW, Fruth A, Middendorf B, Harmsen D, Schmidt MA, Karch H: Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg Infect Dis* 2008;14:1287–1290.
- 89 Schürch AC, Siezen RJ: Genomic tracing of epidemics and disease outbreaks. *Microb Biotechnol* 2010;3:628–633.