A Microbiological Revolution Meets an Ancient Disease: Improving the Management of Tuberculosis with Genomics

Marta Wlodarska, James C. Johnston, Jennifer L. Gardy, Patrick Tang

The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, USA; Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada; School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada; Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY

Tuberculosis (TB) is an ancient disease with an enormous global impact. Despite declining global incidence, the diagnosis, phenotyping, and epidemiological investigation of TB require significant clinical microbiology laboratory resources. Current methods for the detection and characterization of Mycobacterium tuberculosis consist of a series of laboratory tests varying in speed and performance, each of which yields incremental information about the disease. Since the sequencing of the first M. tuberculosis genome in 1998, genomic tools have aided in the diagnosis, treatment, and control of TB. Here we summarize genomics-based methods that are positioned to be introduced in the modern clinical TB laboratory, and we highlight how recent advances in genomics will improve the detection of antibiotic resistance-confering mutations and the understanding of M. tuberculosis transmission dynamics and epidemiology. We imagine the future TB clinic as one that relies heavily on genomic interrogation of the M. tuberculosis isolate, allowing for more rapid diagnosis of TB and real-time monitoring of outbreak emergence.

INTRODUCTION

Tuberculosis (TB) is a major global health problem and the second leading cause of death from an infectious disease, after HIV (1). It is an ancient disease that has been with humankind for thousands of years, and our relationship with its causative agent, Mycobacterium tuberculosis, is so intimate that pathogen and host have coevolved (2–4). M. tuberculosis spreads through the inhalation of infectious aerosols, and newly infected individuals progress to one of two states: a symptomatic and potentially infectious state, known as active TB, or an asymptomatic, noninfectious state called latent TB infection (LTBI) (Fig. 1). Approximately one-third of the global human population has LTBI (1), and although LTBI does not manifest with any clinical symptoms, it comes with the risk of developing into active TB disease. Active TB typically presents as pulmonary disease but can also present as extrapulmonary or disseminated disease (5, 6).

As with other airborne diseases, the spread of M. tuberculosis is facilitated by high population densities and crowded indoor environments that optimize the aerosol transmission of the pathogen. Host factors, such as immune suppression, smoking, poor nutrition, diabetes, and respiratory comorbidities, also play a role by increasing the risk of transitioning from latent to active TB (6). This is particularly an issue in the developing world, where overcrowding, malnutrition, and HIV infection all contribute to a high burden of disease. Worldwide, the majority of cases occur in Southeast Asia and the Western Pacific Regions (56%), Africa (25%), India (24%), and China (11%) (1). Globally, the mortality rate of TB has decreased in the last decade, and the incidence of active disease is steadily but slowly declining (approximately 1 to 3% per year) (7). However, the burden of disease remains substantial, especially in low-income countries— in 2013, an esti...
M. tuberculosis is a complex (MTBC). The predominantly human pathogen contains 9 million people developed TB, and 1.5 million died from TB threatens to reverse the gains we have achieved to date. The rapid and accurate clinical laboratory diagnosis and characterization of M. tuberculosis is pivotal to battling TB (7). This is complicated by the fact that M. tuberculosis belongs to a group of genetically related species collectively termed the Mycobacterium tuberculosis complex (MTBC). The predominantly human pathogens within the MTBC, in addition to M. tuberculosis, include M. africanum, M. canetti, M. bovis, and the related M. bovis bacillus Calmette-Guérin strain used in the TB vaccine. Recent genomic investigations suggest that M. canetti is substantially more recombinogenic than other mycobacteria and may not be grouped correctly within the MTBC (8, 9). The remaining members of the complex largely infect animals (3) but have been documented as having caused human infections—these include M. microti, M. caprae, M. orygis, and M. pinnipedii (10–13).

Many classical microbiological techniques for species identification and phenotyping are of limited utility for the MTBC due to the genetic homogeneity of its members (14), with most of the species in the MTBC sharing 99.9% nucleotide identity and identical 16S rRNA sequences (15). In contrast, genomics has been integral in revealing the population structure of the MTBC, as recently reviewed by Coscolla and Gagneux (16). Seven lineages of human-adapted MTBC organisms have been described and strongly correlate with geography. M. tuberculosis lineages 2 (East Asian, including Beijing strains) and 4 (Euro-American) are widely distributed globally, lineages 1 (Indo-Oceanic) and 3 (East African-Indian) are restricted to eastern Africa and parts of Asia, and lineage 7 (Ethiopian) is restricted to a small region in the Horn of Africa. Lineages 5 and 6 comprise M. africanum isolates designated West Africa 1 and 2, respectively. Molecular and genomic investigations have also added approximately 50 new members to the list of over 150 nontuberculous mycobacteria (NTMs) since 2006, including several that have been documented as causing human disease (17), and have pointed to seals as a potential vector for the introduction of TB to the New World in the Holocene era (4).

The genomic revolution offers new opportunities for diagnosis and characterization of the MTBC. With the advent of rapid and cost-effective platforms for whole-genome sequencing (WGS) (reviewed in reference 18), many of the multiple time-consuming steps in the clinical mycobacteriology laboratory, including detection, species identification, and antimicrobial resistance pheno- typing, can be replaced with a single WGS analysis. This review focuses on describing the genomic tools that have aided in the identification, diagnosis, treatment, and control of TB thus far and the potential future impact of WGS technologies on detection of antibiotic-resistance-conferring mutations and TB epidemiology in the context of the clinical microbiology laboratory. While WGS-based advances, to date, have largely been in laboratories in high-income, low-burden settings, we also discuss the role that these technologies may play now and in the future in low-income regions. The general principles behind WGS in the clinical microbiology laboratory and the differences between available platforms are not covered here; instead, we refer the reader to a recent comprehensive review of these topics by Didelot et al. (19).

**ROLE OF GENOMICS IN THE DIAGNOSIS OF M. TUBERCULOSIS**

The modern TB laboratory’s role is the detection and species identification of mycobacteria, as well as drug susceptibility testing to optimize each patient’s treatment plan. For pulmonary TB, the most common form of active TB, diagnosis is typically made with a sputum sample from the patient. In the clinical laboratory, the sputum sample is decontaminated to enhance the recovery of mycobacteria, including M. tuberculosis. This treated sputum sample is concentrated, fixed onto a slide, stained with acid-fast stains, and visualized microscopically for the presence of mycobacteria in what is known as the acid-fast bacillus (AFB) smear test. Unfortunately, this test has a high limit of detection, requiring at least 10^5 CFU/ml in the sputum sample to be able to consistently detect mycobacteria, and the test does not distinguish between different species of mycobacteria. In resource-limited settings, the AFB smear test may be the only test available, often without the added steps of decontamination and concentration. However, for modern TB laboratories with biosafety level 3 facilities, the treated sputum sample will also be cultured using selective media for mycobacteria. Although culture-based methods are the gold standard for the diagnosis of mycobacterial disease, mycobacteria grow slowly, meaning that culture can require 6 to 8 weeks. After growth of the mycobacteria, phenotypic methods have traditionally been used to categorize and differentiate some Mycobacterium species, but lateral flow immunoassays specifically for the MTBC (20), as well as molecular methods, such as nonamplified DNA probes, line probe assays, or DNA sequencing, are more com-

---

**FIG 1** Infectious spread of M. tuberculosis and resulting disease. An infectious individual with pulmonary TB can spread disease via aerosols—bacilli are expelled during coughing, talking, singing, and other activities, establishing infection in approximately 25 to 50% of close contacts. Most newly infected contacts develop LTBI (90 to 95%), though a small percentage (5 to 10%) immediately develop primary progressive active disease. Among LTBI cases, 5 to 10% of individuals will have reactivated disease over their lifetime, in what is termed postprimary disease. Risk factors associated with reactivation include immunodeficiency caused by HIV coinfection, use of immunosuppressive drugs, and diabetes mellitus; socioeconomic status; and smoking. Active disease, either primary or postprimary, may present as pulmonary disease or extrapulmonary disease, depending on age, the presence or absence of underlying disease, the genotype of the M. tuberculosis strain, and immune status. Colored circles represent contacts of an “index case.” Red represents active disease, blue represents latent infection, gray represents no disease, black solid lines represent person-to-person aerosol transmission, and the black dashed line represents reactivation of latent infection.
monly used to definitively identify *M. tuberculosis* and nontuberculosis mycobacteria to the species level (21–23). Additionally, nucleic acid amplification tests for *M. tuberculosis* can be applied directly to the sputum sample for the immediate diagnosis of TB independent of the culturing process.

**Advances in *M. tuberculosis* Detection**

In an effort to improve the rapid identification and characterization of *M. tuberculosis*, several molecular tools have been developed based on our knowledge of the genome of *M. tuberculosis*, which was first completely sequenced in 1998 (14). For rapid TB diagnosis, the WHO has endorsed the use of the Xpert MTB/RIF test (Cepheid), an automated real-time PCR-based test that simultaneously detects the presence of *M. tuberculosis* and common mutations for rifampin resistance, considered a marker of multidrug resistance. The GeneXpert platform can provide a result within hours directly from a sputum sample and has a higher sensitivity than the AFB smear test, being able to detect as few as 10^2 CFU/ml (1, 24, 25), though the sensitivity is lower than those of the gold standard of culture and of the microscopic observation drug susceptibility (MODS) assay (26). The platform has been used successfully in low- and middle-income countries; however, its utility is debated for remote and/or highly underresourced settings, where a lack of reliable electricity, limited technical support, and other environmental factors exclude all but the simplest diagnostic methods. Further methods are being developed that will allow for the rapid, sensitive, and specific detection of *M. tuberculosis* directly from the patient sample, be it sputum, cerebrospinal fluid, or another specimen type. Selective enrichment is one such method. Studies have shown the effectiveness of ligand-coated magnetic beads to capture and concentrate *M. tuberculosis* in sputum samples prior to microscopic examination (27–29). This is useful for TB diagnosis in low-income countries, where decontamination and centrifugation to concentrate *M. tuberculosis* may not be feasible due to a lack of resources. Elitas et al. showed a label-free approach based on the dielectric properties of bacterial cells (30). Enrichment uses dielectrophoresis on three-dimensional (3D) carbon electrode arrays to discriminate bacteria based on small changes in their intrinsic properties, such as membrane integrity (30). Further developments in nanotechnology and microfluidics offer the possibility of rapid and accurate detection of *M. tuberculosis* from sputum samples (reviewed by Wang et al. [31]). The development of next-generation sequencing technology also offers the potential to identify the presence of *M. tuberculosis* directly from a sputum sample. Recently, Pallen and colleagues used shotgun metagenomics to detect MTBC sequences directly from eight sputum samples, without an initial culture step (32). Unprocessed sputum was subjected to differential lysis to enrich microbial DNA, and the resulting nucleic acids were sequenced on an Illumina MiSeq instrument (32). By comparing the sequence reads to the H37Rv reference genome, Pallen and colleagues were able to assign seven of the eight isolates to an MTBC species and lineage (32). With the possibility of longer read lengths through emerging nanopore sequencing technology and single-cell genomics, a very low concentration of genomic DNA may be sufficient not only to diagnose the presence of *M. tuberculosis* directly from patient samples (32–34) but also to provide insight into genomic determinants of antibiotic susceptibility (20).

**Genomics for Predicting Antibiotic Susceptibility of *M. tuberculosis***

Following positive identification of *M. tuberculosis*, a specimen will undergo antibiotic susceptibility testing. The results are critical for establishing an appropriate treatment regimen for a given patient, particularly with the recent emergence of MDR and extensively drug-resistant (XDR) TB. Globally, 480,000 cases of MDR TB are estimated to have occurred in 2013, and rapid testing for MDR is a WHO priority action for TB control (1). A culture-based phenotypic assay can be used to determine the susceptibility to and/or MICs for first- and second-line drugs, including isoniazid, rifampin, pyrazinamide, and ethambutol, but may take up to 2 weeks to complete. Line probe assays, targeting antibiotic resistance mutations revealed by genomics, can accelerate this process. These are DNA strip-based tests that use nucleic acid amplification and reverse hybridization methods for the rapid detection of mutations associated with drug resistance and can be used directly on sputum samples or cultured isolates (35). Currently, two products targeting mutations relevant to resistance to first-line drugs are commercially available: the INNO-LiPA Rif.TB test (Innogenetics NV, Belgium) and the “Hain test,” i.e., the GenoType MTBDRplus assay (Hain LifeScience GmbH, Germany) (35–37). A notable drawback to these assays is that they are limited to detecting a specific set of mutations in the rpoB, katG, and inhA genes (e.g., the INNO-LiPA Rif.TB test detects only nine mutations in rpoB [36]) and cannot identify novel mutations conferring resistance (37), although certain assays do contain probes for wild-type alleles—observing a resistance phenotype and a wild-type allele would suggest a novel mutation but would not provide further details. Recently, Hain LifeScience introduced the GenoType MTBDRsl assay, which extends resistance typing to gyrA mutations conferring fluoroquinolone resistance, rrs mutations conferring aminoglycoside resistance, and embB mutations conferring ethambutol resistance. This new assay enables the identification of XDR TB strains but, again, is incapable of characterizing novel mutations.

WGS is emerging as a promising technology for resistance typing of *M. tuberculosis*. Early sequencing studies first targeted specific *M. tuberculosis* genes known to confer resistance to first- and second-line drugs. A study by Campbell et al. sequenced nine loci associated with antibiotic resistance in 314 clinical isolates, 52% of which were MDR and 3% of which were XDR, and compared the phenotypic data with the sequencing data (38). They found that the sequencing data were able to capture the observed resistance phenotype frequently in the case of MDR isolates, with a sensitivity and specificity of 90.8% and 94.7%, respectively. However, accurate sequence-based typing of XDR isolates was more challenging, with only 40.0% sensitivity and 99.3% specificity (38). Recent studies used WGS on hundreds of clinical isolates to identify known and novel antibiotic resistance-conferring mutations (39–41). The largest of these, a study of 1,000 Russian isolates, 48% of which were MDR, identified 238 single nucleotide polymorphisms (SNPs) associated with drug resistance (42). At the other end of the spectrum, a 2014 study by Eldholm et al. examined the emergence of resistance within a single XDR TB patient sampled serially over a 42-month period (43). They observed a total of 35 mutations in the nine clinical isolates sequenced, 15 of which became fixed in the population (43). Twelve of the mutations were determined to play a role in drug resistance, and the
laboratory phenotyping results were concordant with these mutations (43). For five of the seven antibiotics for which resistance arose, two independent resistance mutations were observed (43).

WGS has also been used to examine the potential for and mechanism of resistance to novel therapeutic agents. In 2005, Andries et al. sequenced the genomes of M. tuberculosis and M. smegmatis isolates before and after treatment with the investigational compound R207910 (now called bedaquiline) (44). They detected resistance-conferring mutations in the atpE gene, suggesting that the novel compound targeted the F$_r$, subunit of the ATP synthase protein (44). However, WGS is not the only step in such an analysis; candidate resistance-conferring mutations should be confirmed with Sanger sequencing or a PCR-based single nucleotide variant (SNV) assay and then must be followed up with in vitro functional assays to confirm the link between the observed mutation and phenotype.

WGS studies of drug-resistant isolates have expanded our knowledge of drug resistance mutations but have also shown the difficulty in data interpretation. In the case of ethambutol resistance, a study showed that a low-level resistance that would be undetected by current phenotypic antibiotic susceptibility testing may accumulate within the genome of M. tuberculosis (41). In contrast, another study reported that phenotypic testing indicated isoniazid resistance, but WGS failed to identify the mutation responsible (45). Mutations conferring drug resistance are known to come at the cost of a loss in an organism’s fitness; however, recent genomic studies revealed that compensatory mutations can occur which overcome the loss of fitness and create a “perfect storm” of resistant isolates without reduced fitness (42, 46–48). Furthermore, it has been shown that M. tuberculosis strains of different lineages can differ in their virulence, acquiring mutations conferring antibiotic resistance, a delayed host immune response, enhanced transmissibility, and more severe disease at different rates (49–51). The success of some of the more virulent Beijing and Euro-American strains, for example, has been attributed to the genomic diversity of these specific lineages (16). Understanding the emergence of compensatory mutations, their interactions with other mutations (epistasis), and the impact of these on treatment decisions is an emerging area into which WGS can provide important insights, as recently reviewed by Trauner et al. (52).

A recent case study by Köser et al. (53) showed retrospectively how WGS could be used in a clinical setting to examine drug resistance in a patient simultaneously infected with two strains of XDR M. tuberculosis. The WGS-based prediction of drug resistance was concordant with the phenotypic testing results done at a reference laboratory, but the WGS data were generated in a matter of days, versus the weeks required at the reference laboratory. Furthermore, the WGS analysis also identified resistance to five additional antibiotics not tested by the reference laboratory (53). This data suggest that WGS will allow faster identification of drug resistance, allowing physicians to more rapidly determine the most effective treatment plan for the patient and potentially limit the further spread of drug-resistant strains. However, before WGS can replace standard phenotypic testing, more data on the concordance between SNPs and resistance phenotypes, including the role of compensatory mutations, are needed. Furthermore, although Köser et al. were able to accurately predict the phenotype in a case of mixed infection, this is likely to not always be the case. In a mixed infection with susceptible and resistant bacteria, the observed in vivo phenotype may be difficult to predict from WGS data, especially if the sample being sequenced contains a large proportion of genomic DNA from the susceptible bacteria relative to the resistant bacteria. A sample collected early in the infection or one that has been cultured in the laboratory may have a ratio of susceptible to resistant bacteria that does not reflect the host infection, where the resistant mutants may eventually dominate if the treatment regimen is tailored only for the susceptible isolate. Still, WGS does offer a promising alternative to current growth-based susceptibility testing methods, whose performance is variable due to differences in the physiochemical environment in which the testing is performed and the protocols used for inoculum preparation and testing, and whose in vitro results may not reflect the in vivo clinical reality (54).

Currently, many mutations conferring antibiotic resistance are known for M. tuberculosis, and more continue to be identified (42, 55, 56); however, in order for WGS to be implemented as a routine clinical tool for resistance typing, a centralized and freely accessible database must be created to catalog resistance- and heightened virulence-conferring mutations, similar to the Stanford HIV Drug Resistance Database (57). A legacy database, TBDReaMDB (58), exists, as does the more recent MUBII-TB-DB (59). Both compile reported genetic polymorphisms conferring drug resistance in M. tuberculosis, with TBDRea MDB including mutations reported until 2010 and MUBII-TB-DB including data through 2014. The use of such databases in a clinical setting has yet to be formalized, but the recent announcement that the Gates Foundation has funded a software project to warehouse TB genomes to facilitate and formalize resistance typing suggests that clinical use is on the horizon (Genomeweb). The success of a centralized, standard reference database will depend on continued efforts to identify emerging mechanisms of resistance, regular comparisons of genotype with phenotype, agreement of the TB community to utilize this database globally, and dedicated maintenance of the resource.

**GENOMICS AS A TOOL FOR TB OUTBREAK INVESTIGATION**

In addition to rapid diagnosis and phenotyping, contact tracing and effective outbreak management are critical to controlling and preventing the spread of TB in settings where resources permit (60, 61). Contact tracing is a routine epidemiological practice in which persons who may have been exposed to an infectious case of TB are identified, screened for evidence of LTBI or active TB, and offered preventative or active TB therapy when appropriate (1). Outbreak management strategies are deployed when a TB outbreak has been declared, and these involve contact tracing coupled with enhanced community-wide screening activities to capture secondary cases in individuals who may not be in a patient’s immediate circle of named contacts (1). Contact investigations have been enhanced by molecular techniques in recent years, and WGS is revealing new insights into how TB spreads within a contact network (62, 63).

**Current Molecular Epidemiology Tools**

A suspected epidemiological link between two TB cases can be confirmed by use of molecular epidemiology techniques in which the two M. tuberculosis isolates are compared at the genetic level: identical or nearly identical genetic fingerprints suggest contact between individuals or exposure to a common source. These techniques can also be used to discover new epidemiological links, revealing clusters of related cases. Molecular epidemiological techniques take advantage of known genetic variation within the
M. tuberculosis genome. The majority of this genetic variation occurs as SNPs, insertions, deletions, genomic rearrangements, and transpositions, and the genome contains numerous repeat sequences, including transposable elements, variable-number tandem repeats (VNTRs), mycobacterial interspersed repetitive units (MIRUs), and direct repeat regions (64–67).

Classical DNA fingerprinting or genotyping methods exploiting variation in the M. tuberculosis genome include spoligotyping (64), IS6110 restriction fragment length polymorphism (RFLP) analysis (66), and MIRU-VNTR analysis (65) (Table 1). Of these, 24-locus MIRU-VNTR assay is considered to be the gold standard method, but it is often combined with spoligotyping and/or IS6110 RFLP analysis to further increase its discriminatory power (68). Even so, current fingerprinting methods can only identify linked cases—they cannot reveal the direction of transmission or discriminate between primary, secondary, and tertiary cases. Furthermore, cases with matching fingerprints generally trigger resource-intensive contact tracing efforts, and these techniques often cluster more distantly related isolates even where there is no direct epidemiological connection (67). WGS is an attractive alternative: by examining the whole genome, not just a small fraction of the available genetic information, it captures all the variation present, providing a high-resolution, accurate, and reproducible method for identifying fine-scale transmission events, while simultaneously providing information for use in other aspects of TB management, such as predicting antimicrobial resistance (Table 1) (53, 67).

### Advancing Molecular Epidemiology with WGS

The principle behind using WGS to identify transmission events is simple. Over the time scale of an outbreak, random mutations will accrue in the M. tuberculosis genome. A mutation arising in one host will be transmitted to all individuals infected by that host and will similarly be passed on to people those individuals go on to

### TABLE 1 Comparison of available molecular epidemiology tools for M. tuberculosis

<table>
<thead>
<tr>
<th>Method</th>
<th>Attributes</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoligotyping</td>
<td>May be used directly on clinical samples or early cultures</td>
<td>Low DNA concentration required</td>
<td>Less discriminatory than other genotyping methods</td>
</tr>
<tr>
<td></td>
<td>Rapid initial PCR amplification step</td>
<td>Short turnaround time</td>
<td>False clustering</td>
</tr>
<tr>
<td></td>
<td>Involves probe hybridization</td>
<td>Objective and reproducible results</td>
<td>Unable to resolve transmission events within outbreaks</td>
</tr>
<tr>
<td></td>
<td>Numerical results</td>
<td>Fingerprint defined with a numerical code, which facilitates comparisons</td>
<td></td>
</tr>
<tr>
<td>IS6110-based RFLP analysis</td>
<td>Cultured isolates are required to obtain sufficient DNA for RFLP analysis</td>
<td>High discriminatory power</td>
<td>High DNA concentration required</td>
</tr>
<tr>
<td></td>
<td>Restriction digestion followed by gel electrophoresis and blot transfer</td>
<td></td>
<td>Banding pattern may be difficult to interpret and standardize</td>
</tr>
<tr>
<td></td>
<td>Hybridization of IS6110-specific probe</td>
<td></td>
<td>Technically demanding and has a long turnaround time</td>
</tr>
<tr>
<td></td>
<td>Generates a band pattern fingerprint</td>
<td></td>
<td>Not suitable for strains with no or low-copy-number IS6110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>False clustering</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Unable to resolve transmission events within outbreaks</td>
</tr>
<tr>
<td>15- and 24-locus MIRU-VNTR analyses</td>
<td>Require cultured isolates</td>
<td>Low DNA concentration required</td>
<td>Interrogates only a small fraction of the entire genome</td>
</tr>
<tr>
<td></td>
<td>Rapid initial PCR amplification step</td>
<td>Short turnaround time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fragment lengths can be resolved using gel or capillary electrophoresis systems</td>
<td>Objective and reproducible results</td>
<td></td>
</tr>
<tr>
<td>WGS</td>
<td>Currently requires culture&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Superior tool for resolving transmission dynamics</td>
<td>Technically demanding for sample preparation and data analysis</td>
</tr>
<tr>
<td></td>
<td>Fragmentation of genomic DNA followed by addition of sequencing adapters</td>
<td></td>
<td>Not globally available due to higher costs</td>
</tr>
<tr>
<td></td>
<td>Desktop-sized sequencers perform sequencing by synthesis</td>
<td></td>
<td>No standardized methods for data analysis and data sharing</td>
</tr>
<tr>
<td></td>
<td>Data analysis requires bioinformatics expertise</td>
<td>Yields information about virulence factors, antibiotic resistance, and epidemiology</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Newer sequencing technologies may be able to generate whole-genome sequences without culture.
<table>
<thead>
<tr>
<th>Study focus and period</th>
<th>Location</th>
<th>No. of WGS isolates</th>
<th>Genome coverage (%)</th>
<th>No. of SNVs&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sequencer&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Major finding(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies on transmission of TB</td>
<td>Uzbekistan</td>
<td>2</td>
<td>95.9</td>
<td>1,209&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Illumina</td>
<td>Showed that isolates with identical IS6110 patterns differed by 130 SNVs</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Netherlands</td>
<td>3</td>
<td>95.4</td>
<td>8&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Roche/454</td>
<td>Discovered that SNVs can be used to identify transmission chains in RFLP clusters</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>36</td>
<td>99.2</td>
<td>204</td>
<td>Illumina</td>
<td>Integration of WGS with a social network questionnaire generated a more accurate transmission network; used WGS to identify &quot;superspreader&quot; cases</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>2</td>
<td>95</td>
<td>0</td>
<td>Illumina</td>
<td>Outbreak identification with WGS; showed the utility of WGS to identify drug resistance markers</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>390</td>
<td>88.5</td>
<td>1,096</td>
<td>Illumina</td>
<td>Rate of genetic changes was 0.5 SNV/genome/year; used WGS to identify &quot;superspreader&quot; cases</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>86</td>
<td>96.4</td>
<td>85</td>
<td>Roche/454</td>
<td>WGS was proven to be more effective at generating clustering patterns than classical genotyping methods; used WGS to estimate a rate of genetic change of 0.4 SNV/genome/year</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>9</td>
<td>95.7</td>
<td>7&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Illumina</td>
<td>WGS allowed the identification of new epidemiological links in the transmission chain; used WGS to characterize a mutation rate of 0–2 SNVs/transmission event resulting in a secondary case</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>256</td>
<td>92.6</td>
<td>1,715</td>
<td>Illumina</td>
<td>Patients born in low-incidence countries are more likely to have pulmonary disease and social risk factors, causing secondary cases in the UK</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Netherlands</td>
<td>199</td>
<td>95.6</td>
<td>11,879</td>
<td>Roche/454</td>
<td>Estimated an average mutation rate of 0.3 SNV/genome/year, with a large degree of variation, i.e., 0.4–17 SNVs/genome/year</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>China</td>
<td>32</td>
<td>ND</td>
<td>1,790</td>
<td>Illumina</td>
<td>Evaluated the usefulness of WGS in a high-incidence country</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Switzerland</td>
<td>69</td>
<td>98.3</td>
<td>133</td>
<td>Illumina</td>
<td>WGS suggested a single origin for an outbreak involving 68 patients over 21 years that could be divided into 3 subclusters with epidemiological links differing by 0–11 SNVs</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>33</td>
<td>ND</td>
<td>21</td>
<td>Illumina</td>
<td>WGS revealed within-host diversity in an index case; a computational method was developed to infer transmission from phylogenetic data, accounting for within-host diversity</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>36</td>
<td>ND</td>
<td>3,523</td>
<td>Illumina</td>
<td>WGS of &lt;i&gt;M. tuberculosis&lt;/i&gt; strains belonging to the Manila sublineage showed better resolution than that of MIRU-VNTR assay and decreased the frequency of clustered cases</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>61</td>
<td>ND</td>
<td>722</td>
<td>Illumina</td>
<td>WGS revealed 6 subclusters of the 17-year outbreak showing unique patterns of evolution</td>
<td>75</td>
</tr>
</tbody>
</table>
Studies on microevolution of *M. tuberculosis* 1994–2011

<table>
<thead>
<tr>
<th>Year</th>
<th>Country</th>
<th>N</th>
<th>SNVs</th>
<th>Illumina</th>
<th>Defined epidemiologically linked transmission events by using SNV cutoffs of &lt;5 SNVs for transmission and &gt;12 SNVs for no evidence of transmission; intrainfant variation was limited to &lt;5 SNVs; rate of genetic changes was 0.5 SNV/genome/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994–2011</td>
<td>UK</td>
<td>390</td>
<td>88.5</td>
<td>1,096</td>
<td></td>
</tr>
<tr>
<td>2003–2010</td>
<td>Spain</td>
<td>36</td>
<td>98</td>
<td>28</td>
<td>Examined inter- and intrainfant variations and determined them to be equivalent</td>
</tr>
<tr>
<td>2008</td>
<td>Malaysia, South Africa, Thailand</td>
<td>96</td>
<td>ND</td>
<td>1,419</td>
<td>Showed that WGS can be used to discriminate between relapse and reinfection</td>
</tr>
<tr>
<td>1990–2010</td>
<td>New Zealand</td>
<td>10</td>
<td>98</td>
<td>747</td>
<td>Used WGS to show that mutation rates during latent infection in humans are substantially lower than those during active disease</td>
</tr>
</tbody>
</table>

63. SNVs were verified by secondary Sanger sequencing.

86. ND, not defined in the report.

89. The Illumina, Roche/454, and SOLiD platforms all represent short-read sequencing-by-synthesis platforms and are reviewed in reference 18. Newer technologies, such as nanopore-based sequencing or the PacBio platform, have also been applied to microbial genomics but have yet to be used in a TB study.

93. SNPs were verified by secondary Sanger sequencing.

102. The utility of WGS and DNA fingerprinting methods to improve the resolution of transmission events than that of historical fingerprinting methods, even when both techniques are combined.

103. Regardless of the nomenclature, we might reasonably expect a zero- to five-SNV (or SNP) difference between the infectious source and an infected recipient.

104. Recent retrospective studies showed that WGS offers far superior resolution of transmission events than that of historical fingerprinting methods, even when both techniques are combined.

105. A 2009 study by Niemann et al. was the first to compare the utility of WGS and DNA fingerprinting methods within the same outbreak by using whole-genome sequencing, we have chosen the more conservative definition of these mutations as SNVs, as it is not yet possible to calculate the degree of fixation of these mutations within any given genotype of *M. tuberculosis*.
for typing of *M. tuberculosis* isolates to identify transmission events in outbreaks (70, 77).

The following year, Gardy et al. were the first to sequence a complete outbreak, sequencing the first 32 isolates of an ongoing outbreak in British Columbia, Canada, along with 4 preoutbreak isolates from the same region with the same MIRU-VNTR pattern (62). The study revealed substantial genetic variation within the outbreak and, combined with epidemiological data derived from a social network questionnaire, provided the first nearly complete transmission network of an outbreak, which revealed the contribution of previously unrecognized “superspreaders” responsible for seeding multiple secondary infections. A second study, published in 2013, demonstrated that WGS could also link previously unsuspected cases (78). In the United Kingdom, two active TB outbreaks in 2013, demonstrated that WGS could also link previously unsuspected cases (78). Walker et al. sequenced isolates from 247 cases and found that foreign-born individuals were less likely to cause onward transmission, confirming results from earlier studies using other genotyping methods (79). They hypothesized that among the United Kingdom-born individuals, social risk factors may have led to a failure to seek health care, and physicians may not have considered a TB diagnosis; together, these factors would contribute to delayed diagnosis and increased community transmission. These social risk factors, including homelessness, emerged as drivers of several outbreaks recently characterized using WGS, including sequencing of 61 *M. tuberculosis* isolates from a protracted outbreak among the homeless in Toronto, Canada (75), which revealed six independent transmission chains, and of 33 isolates from an outbreak originating in a Kelowna, Canada, homeless shelter (80). In both scenarios, WGS suggested the occurrence of transmission within the shelter environment, as well as transmission from shelter clients to shelter staff.

WGS of every outbreak isolate can be prohibitively expensive, particularly for long outbreaks with hundreds of cases. Recent work suggests that limited WGS may also have utility in the clinical laboratory in such scenarios. Stucki et al. carried out WGS on three isolates from a large Swiss TB outbreak occurring over 21 years and on four unrelated control isolates (76). The whole-genome data revealed 118 SNVs present in the outbreak strains but not the control strains; 2 of these were selected as the targets of a PCR-based SNV typing assay that was then applied to 1,642 archived isolates. The assay ruled in 68 archived isolates as belonging to the outbreak; these were then analyzed using the full WGS approach (76). Using this limited WGS/SNV typing method, the authors were able to rapidly characterize the outbreak at a substantially reduced cost ($4,900) compared to that for other genotyping methods: the SNV assay was 15 times less expensive than MIRU-VNTR typing, and the targeted WGS was 20 times less expensive than performing WGS on the entire strain collection (76).

In another limited-WGS approach, a core genome multilocus sequence typing (cgMLST) scheme was recently proposed as a method to convert WGS-identified SNVs in coding regions of the genome into an allele numbering system (81). Although this method is standardized, shareable across jurisdictions, portable, and not computationally intensive, it failed to resolve outbreak clusters as effectively as approaches that use the entirety of the genome (81). However, the advantages of universally implementing such an approach at a national or international level mean that cgMLST may provide a high-level view of large trends that are not apparent with traditional genotyping, with the resulting cgMLST clusters then targeted for follow-up with more detailed SNV-based phylogenies. Approaches such as cgMLST also raise the possibility of leveraging WGS data to “back-calculate” traditional molecular epidemiology fingerprints from genomic data. SpolPred calculates a spoligotyping pattern directly from short-read sequencing data (82), and as sequencing platforms improve and are able to handle longer repetitive regions, generating in silico MIRU-VNTR patterns should also be possible, permitting comparisons between WGS isolates and historically fingerprinted isolates.

Importantly, it should be noted that the above-described genomic epidemiology studies all took available epidemiological data into account in establishing the direction of transmission. Recent work has shown that in the absence of epidemiological information, WGS does provide finer subclustering of isolates than would be possible with traditional genotyping, but it cannot identify individual transmission events with any certainty (80). The direction of transmission can be difficult to resolve given genic data alone; to make plausible inferences requires knowledge of a patient’s symptomatic and infectious periods, their relative infectiousness, and their social contacts and places of social aggregation. Even with this knowledge, delayed diagnoses and long infectious periods—common in vulnerable populations—relinquent or unable to seek care—introduce significant differences that can complicate linking sequenced isolates to each other. A second caveat is that much of the potentially informative variation in MTBC genomes lies in repetitive regions, such as the PE/PPE genes, which constitute ~10% of the genome; however, current short-read sequencing technologies and bioinformatic algorithms struggle with these types of repetitive regions, meaning that studies to date have largely ignored variation in these regions, which may provide even deeper insight into the nature of transmission and of mutation rates.

The studies described above have largely been retrospective, examining outbreaks that have concluded or are already well under way, and given the prolonged nature of TB outbreaks, which typically span several years, it is likely that we will not see “real-time” genomic epidemiology unfold for TB the same way it has for other organisms, including hospital-acquired infections (83, 84). However, there are indications that while genomic data alone cannot identify individual transmission events, it can provide some insight into the overall dynamics of the outbreak, including whether or not the outbreak shows evidence of a superspreader (85) and which patients are most central to the outbreak network (80). These patterns can be detected within the first several isolates from an outbreak (85), suggesting that real-time WGS may be useful in understanding an outbreak and its key players, thereby informing the public health approach to managing the situation.

**UNDERSTANDING THE MUTATION RATE OF *M. TUBERCULOSIS***

In 2013, studies from Walker et al., Roetzer et al., and Kato-Maeda et al. further demonstrated that WGS outperforms traditional fingerprinting methodologies in resolving outbreaks and identifying direct transmission events (63, 71, 72), with their work providing insights into the *M. tuberculosis* mutation rate in an outbreak sce-
nario and into the expected genomic differences between related isolates (Table 2). Most notably, Walker et al. sequenced 390 isolates representing a range of epidemiological scenarios to determine whether the number of unique SNVs between isolates could be used to draw epidemiological inferences about cases (63). They determined that if two isolates differed by fewer than 5 SNVs, they likely belonged to the same transmission network, whereas if they differed by more than 12 SNVs, they could not represent a direct transmission event (63). Their study also examined the accrual of SNVs within a patient over time, finding that longitudinal isolates collected up to 8 years apart varied by fewer than five SNVs (63). Furthermore, they showed that cross-sectional diversity within a host, as revealed by isolates taken from various body sites (pulmonary versus nonpulmonary) at one time, was similarly limited to five or fewer SNVs (63). This five-SNV threshold was also shown to be useful in differentiating relapse of infection from reinfection, as cases of relapse did not differ by more than six SNVs (86). Roetzer et al. sequenced 86 isolates from a large outbreak, noting a maximum of three SNVs between epidemiologically linked patients, while Kato-Maeda et al. sequenced nine isolates from a small cluster of cases, observing zero to two SNVs between linked cases (71, 72).

These and other studies also shed light on the rate of genetic change in M. tuberculosis in a natural outbreak scenario. Walker et al. reported a mutation rate of 0.5 SNV genome/year, in agreement with a previous study done in cynomolagus macaques (63, 87). A subsequent study by Bryant et al. also reported a similar mutation rate of approximately 0.3 SNV genome/year (88), while Roetzer et al. calculated a rate of 0.4 SNV genome/year (71). This consistency appears to support the use of total accumulated SNVs to rule in or rule out transmission; however, Bryant et al. noted a large degree of variation in numbers of SNVs genome/year across their data set of 199 sequenced epidemiologically linked isolates, and they concluded that phylogenetic analysis would offer an advantage over SNV thresholds in transmission studies (88). The SNV cutoffs defined by Walker et al. were also challenged by another study, which showed that although inter- and intrapatient variations were similar, a constant rate of SNV acquisition may not be valid for all strains and situations (89). Recently, an examination of serial evolution within an XDR TB patient reported substantially higher rates, from 1.1 to as high as 7.0 SNVs genome/year (43). This increased rate was attributed to selective sweeps driven by drug treatment, and the authors cautioned against reliance on strict SNV count thresholds in ruling transmissions in or out, noting that two serial isolates from the patient were separated by 11 SNVs (43).

INSIGHTS INTO MIXED INFECTIONS AND WITHIN-HOST GENETIC DIVERSITY ACCRUED DURING LTBI

The earliest outbreak reconstructions also failed to consider the potential impact of mixed infection or within-host genetic diversity on our ability to make accurate inferences about transmission from genomic data (Fig. 2). This is a particularly important issue for TB, as LTBI and chronic, undiagnosed active disease both contribute to within-host diversity, whose magnitude depends on the time of infection and the host’s immune background. Recently, efforts to account for this diversity were proposed, including a method calibrated for TB infections. Utilizing a novel Bayesian approach, Didelot and colleagues combined the phylogenetic inference package BEAST with a Markov chain Monte Carlo inference of the suite of possible transmission events that could have given rise to an observed phylogenetic tree (80). This transmission network can be refined further with epidemiological data, such as time of infection, location, and smear-positive or -negative status (80). When applied to genomic data derived from a real outbreak, their TransPhylo method was able to identify the source case and the major transmission subclusters within the outbreak, and they reported a mutation rate of approximately 0.5 SNV genome/year, in agreement with earlier reports (63, 80, 87, 88).

The contribution of LTBI to the generation of within-host genetic diversity and its consequences for reconstruction of disease transmission are unclear. Upon initial infection of the lungs, M. tuberculosis is phagocytosed by innate immune cells, including macrophages, and initiates intracellular survival within these immune cells, modulating the immune response and leading to the formation of a granuloma (90, 91). M. tuberculosis can persist within the granuloma during LTBI, during which time its replication rate is thought to decline (92). Understanding mutation rates during LTBI is important, as LTBI treatment often consists of isoniazid treatment alone for 6 to 9 months, and this long-term mono-therapy may lead to the emergence of antibiotic resistance (6).

Two studies, one in macaques and the other in humans, estimated the mutation rate of M. tuberculosis during latency, with conflicting results (87, 93). Ford et al. showed that in M. tuberculosis infections of macaques, the mutation rates of M. tuberculosis were identical during active disease, latent infection, and in vitro growth (87). They speculated that either M. tuberculosis has similar growth and mutation rates during latent infection, accounting for the similarity to the rates for active disease, or the growth rate decreases during latent infection, with a concurrent increase in the mutation rate (87). They provided evidence that the mutations that arise during latency are a result of oxidative damage, which is consistent with the exposure of the bacterium to oxidative stress during adaptation to intracellular survival within macrophages (87). In order to further characterize the mutation rate of M. tuberculosis during latency in humans, Colangeli et al. used epidemiological and phylogenetic analyses of two recently acquired M. tuberculosis isolates and two cases of reactivated latent infection (20 years after initial exposure) from a multidecade outbreak in New Zealand (93). They reported that during latent infection, the mutation rate was approximately 30-fold lower than that during active disease, in contrast to the macaque study results, and they did not report any evidence to suggest that oxidative damage increases the mutation rate of M. tuberculosis during latency (87, 93). The discrepancy in the results between these two studies may be driven by the choice of host, duration of latent infection (which was shorter in the macaque model), host-driven alterations in the immune response, and the human study’s assumption that in the 2 years prior to reactivation, the replication and mutation rates gradually increased to match the rates observed in active disease (93).

UTILITY OF WGS IN COUNTRIES WITH HIGH TB BURDENS

Genomic studies of TB transmission have, to date, been performed largely in low-incidence, highly resourced settings where facilities for culture-based diagnosis are available and where the cost of WGS makes such analyses feasible. The utility of WGS in high-incidence countries has yet to be explored thoroughly, but a recent study by Luo et al. is a first step in this direction. They evaluated the effectiveness of WGS in China and suggested that WGS did allow for the resolution of some
transmission chains (73). However, due to the genetic homogeneity of *M. tuberculosis* and the high incidence of disease in the region, transmission events were more difficult to elucidate than the case for studies in low-incidence countries (73). It is unlikely that WGS would offer much direct benefit to the control of TB outbreaks in high-incidence countries, particularly given the resource limitations of these countries; however, the information gained from WGS studies in low-incidence countries is informative with respect to transmission dynamics and may suggest screening strategies or outbreak interventions that might be useful in higher-incidence settings, where case finding may be challenging.

In the developing world, gains in TB control are most likely to come from scale-up of accurate diagnostics. While we do not foresee WGS-based diagnostics in low-income settings in the near term, the public release of increasing amounts of TB genomic data is likely to result in novel molecular diagnostics that can be used at the point of care and in the absence of culture in resource-limited laboratories. Comparative genomic analyses will likely reveal novel molecular targets specific to particular MTBC members and lineages that may be used in GeneXpert-like rapid-testing diagnostic platforms, while the discovery of new resistance-linked markers may be coupled to these detection methods for enhanced resistance surveillance efforts.

Indeed, it is in the area of drug resistance that WGS is most likely to have an impact in high-incidence countries. Sequencing of 56 TB genomes from Lisbon, Portugal, a city with high MDR rates, demonstrated the potential of genomics for identifying MTBC lineages responsible for local trends in drug-resistant TB (94). WGS-based insights into the emergence of drug resistance and the order in which mutations arise might provide a barometer for evaluating a region’s resistance trends. Viewed through the molecular epidemiology lens, WGS data might also suggest the extent to which drug-resistant strains are transmitted from person

---

**FIG 2** Within-host genetic diversity can complicate transmission inferences from genomic data. (A) Hypothetical transmission scheme, starting with an index patient in whom three distinct bacterial populations are present as a result of a long period of latency, mixed infection, or undiagnosed active disease. Percentages indicate the abundance of each *M. tuberculosis* population. (B) Potential results of a WGS-based outbreak investigation. The index patient transmits different organisms to different secondary individuals and may potentially transmit a mixed infection, and only the most abundant genotypes are identified in each patient. Thus, in this case, WGS fails to link some secondary cases to the index patient and resolves this outbreak into two independent clusters.
to person within a region, providing data that may be used to guide local interventions.

LOOKING TO THE FUTURE: TB AS A MODEL FOR PUBLIC HEALTH GENOMICS

The studies reviewed here offer proof of principle that genomics can positively affect TB care from bench to bedside, from improved diagnostics to new outbreak management strategies. It is likely that TB will be the first pathogen for which a complete genomics approach can be implemented within the well-resourced public health environment, given these early successes. Indeed, the first pilot project using WGS for diagnosis, antibiotic susceptibility determination, and epidemiology was recently launched by Public Health England (https://publichealthmatters.blog.gov.uk/2014/06/16/the-genomics-of-tuberculosis/). One can imagine the TB clinic and laboratory of the future as a place where genomic analyses of the pathogen, and perhaps even the host, are combined to give a complete view of an individual’s infection (Fig. 3).

Advances in selective enrichment and in next-generation sample preparation methods, including single-cell genome sequencing, mean that we will be able to derive the complete genome sequence of a patient’s mycobacterial infection directly from their clinical specimen within days, versus the current standard of weeks. From this sample, we will be able to identify the M. tuberculosis species infecting a patient and any genomic factors potentially associated with virulence, such as belonging to the virulent and transmissible Beijing lineage, which may affect decisions about how a patient is managed at home or in the hospital. As the quality and scope of drug resistance mutation databases increase and they become validated for use in a clinical context, we will be able to use WGS data to rapidly determine the best combination of antibiotics for each patient, thus accelerating the delivery of appropriate TB therapy. Continuous real-time genomics will enable detection of the emergence of antimicrobial resistance over the course of a patient’s therapy, allowing real-time adjustments in therapy prior to the development of a more complicated drug-resistant infection requiring longer treatment. High-throughput targeted sequencing may also prove useful for large resistance typing efforts. Rather than sequencing entire genomes, next-generation sequencing platforms can be used instead for massively parallel sequencing of specific amplicons of interest—in this fashion, an instrument such as Illumina’s MiSeq platform can be used to sequence thousands of amplicons spanning known resistance markers in a single run, rather than only 10 to 50 whole genomes. This parallelized analysis may prove useful in national or international surveys of resistance markers in which WGS of all isolates would be too costly, particularly if advances in sample preparation make culture-independent targeted sequencing easier for nonexpert laboratories.

The pathogen genome data generated in the clinical laboratory will then flow directly to TB epidemiologists, who will interpret the genomic relationships between isolates to identify emerging clusters of disease, reconstruct disease transmission patterns within a cluster or an outbreak, and monitor large-scale trends in the movement of TB genotypes within their region. Understanding outbreaks on an individual level will inform screening strategies, while developing a larger knowledge base of general trends in TB transmission will be key to creating evidence-based policies around managing outbreaks and preventing future occurrences of disease. Cross-border sharing of WGS data should be encouraged not only for identifying transmissions occurring across borders but also to reveal trends in TB dynamics in high-incidence countries, whose emigrants may progress to active disease in a low-incidence country—data from one region alone may be insufficient to draw broad conclusions, but when these data are pooled with data from many regions, we may be able to make inferences about disease occurring in other parts of the world.

We also envision a more distant future in which much more than the pathogen’s DNA is analyzed. Human DNA, which can be recovered from the initial sputum or other specimen or collected using a cheek swab or saliva sample during a patient’s visit to the TB clinic, can provide important insight into a patient’s response to their infection and treatment plan. Antibiotic susceptibility testing data may be combined with a pharmacogenomic panel interrogating human SNPs and alleles linked to adverse drug reactions. For example, specific polymorphisms in the enzyme N-acetyltransferase 2 (NAT2) are known to increase the risk of hepatotoxicity during isoniazid treatment, as well as to affect the pharmacokinetics of the antibiotic (95–98). In addition, immunogenetic factors may also play a role in antibiotic-induced hepatotoxicity, with one study showing that the human leukocyte antigen (HLA) alleles can be predictive for isoniazid-induced hepatoxicity (99). The treating physician would consider both of these inputs (bacterial antibiotic susceptibility and patient genetics) in designing a regimen to provide the most effective treatment while minimizing the risk of adverse reactions, such as hepatotoxicity, by using the appropriate antibiotics at the appropriate doses. As our knowledge of the human microbiome advances, a metagenomic survey of a patient’s intestinal microbiota may also influence the choice of treatment if organisms known to affect the metabolism of antituberculosis drugs are detected. Although no microbe-antituberculosis drug interactions have been described to date, this new area of research is an important component of pharmacology and is rapidly advancing our understanding of how microbial xenobiotic metabolism affects treatment outcomes (100–103). These sorts of personalized treatment plans would require substantial trialing and refinement and would likely be used only in highly resourced settings, but the recent rise of personal-
ized medicine in cancer suggests that they may one day be a possibility for the control of communicable diseases.

Analysis of specific biomarkers from the host genome may detect genetic signatures predictive of a poor response to treatment, with such patients flagged for enhanced follow-up, and host biomarkers may also be leveraged to predict which patients are at highest risk of developing primary progressive or postprimary active TB (104, 105). For example, NRAMP1 polymorphisms have been implicated in the rapid progression to primary TB and have been found in various populations spanning the globe, suggesting that these markers may have some utility in stratifying an individual’s risk of developing active TB (104). This would allow for targeted preventative therapy in these individuals and, again, might be combined with pharmacogenomic data to select the optimal course of antibiotics. However, as with other polymorphism-based biomarkers, any biomarkers used in TB risk stratification must be found across a range of populations and confer a substantially increased risk to have any broad clinical decision-making utility. Azad and colleagues reviewed many innate immune polymorphisms linked to TB susceptibility, disease progression, and variable responses to tuberculosis skin test (TST) and gamma interferon release assay testing (104), noting that interpretation of these biomarkers is further complicated by gene-gene interactions, HIV coinfection and other comorbidities, epigenetic modification, and variations in the infecting MTBC strain (104).

FUTURE DIRECTIONS AND CONCLUSIONS

WGS technology promises a rapid, efficient, and accurate platform for TB diagnosis, treatment planning, and epidemiology and offers the clinical laboratory the tantalizing possibility of using a single platform, a DNA sequencer, for the detailed detection and typing of a range of bacterial and viral pathogens. However, progress toward the lab of the future is not without its barriers. As with any new technology, cost and implementation are significant hurdles. Although the per-base cost of sequencing is dropping at an astonishing rate, there are numerous ancillary costs associated with establishing WGS capability in a laboratory. The required infrastructure includes a sequencing platform, equipment for high-throughput DNA library preparation, and appropriate computational power and storage for handling the large data sets arising from sequencing projects and meeting requirements related to data privacy and security. Similarly, one must also budget for library preparation and sequencing reagents, recognizing that the more “hands-off” a reagent kit is, requiring less technician time, the more expensive it will be. Finally, even the smallest sequencing group must comprise at least two trained technicians—one to handle the wet-lab aspects of preparing and sequencing the DNA and one to handle the computational, dry-lab analyses. Developing bioinformatics expertise, which must include everything from maintaining the computer hardware to understanding the various bioinformatic tools available and their associated options and parameters, is nontrivial and is frequently the rate-limiting step in laboratories’ attempts to establish genomics cores.

A related personnel issue is the need for clinicians, epidemiologists, and other public health staff who have been trained in the interpretation of this new form of data and in understanding what it is and is not capable of providing. In other clinical disciplines in which genomic data are becoming increasingly common, efforts are being made to understand how users interact with online data reports in order to develop better user interfaces for services such as reporting of personal genomic data (106). These information visualization and user interface studies will inform the development of simplified reports that synthesize complex genomic information and associated risk factors into simple scores or visualizations that provide both actionable information and the necessary caveats to a clinician. Much like the GeneXpert output is simplified to clearly report diagnosis and resistance typing, we envision that WGS-based reporting will follow a similar model of a single-page report in lay language.

Sequencing platforms are also constantly changing—the current short-read platforms, which may not be able to read through long stretches of repetitive sequence, are being complemented by newer methods, such as nanopore sequencing, that can read several thousand bases of DNA in one contiguous read. As the platform landscape changes, labs must quickly adapt to assess the utility of these new methods and to integrate them into clinical practice. This will require validating new platforms for their utility in various diagnostic or characterization methods each time that the platform or the reagents change and will complicate the development of sustainable universal protocols. Similarly, updates to bioinformatic software would also merit recalibrating the system and potentially rerunning archived data sets to examine them through an updated lens.

Another critical barrier is standardization of data processing, analysis, and management in a way that is acceptable in the accredited clinical laboratory environment. The Global Microbial Identifier (GMI) project (http://www.globalmicrobialidentifier.org), an international consortium of microbial genomics researchers working within the public health sphere, was created with the vision of “develop[ing] a global system to aggregate, share, mine and use microbiological genomic data to address global public health and clinical challenges.” The GMI consortium has convened working groups to address these issues, examining the need for a minimum data and metadata format and addressing gaps in the analytical tool space, and is also working toward development of a quality assurance testing procedure that genomic clinical labs can use in their proficiency testing. Because it is organism agnostic, the work of the GMI project will have important impacts on the clinical laboratory landscape, and the policies and practices it helps to shape will be applicable across a range of pathogenic bacteria and viruses.

Despite this community-wide focus on establishing analytical standards, there will still be challenges to integrating WGS data with other clinical and epidemiological data sets and interpreting them through the appropriate lenses—these challenges must be dealt with on an organism by organism basis. In the TB community, where the consequences of incorrectly assigning a susceptibility phenotype or misinterpreting an outbreak’s patterns have very real public health implications, we must work particularly hard toward reproducible and accurate analytical pipelines for interpreting WGS data, a deeper understanding of how finely genomic variation separates populations, and data-sharing approaches that will permit discovery of national and international trends in TB population genomics. As more WGS studies of TB are published, a consensus is beginning to emerge around issues such as the depth and quality of sequencing required to make confident variant calls; however, these guidelines should be formalized and shared with the community. The GMI project may be a vehicle for this, as may community-driven sites dedicated to TB genomics.
Similarly, insights derived from WGS will be translated successfully into public health interventions only if the appropriate epidemiological and clinical infrastructure is in place. The lab of the future may be capable of generating new insights into personalized treatment plans or the identification and management of emerging outbreaks, but putting this new knowledge to work will require open lines of communication between the laboratory and public-facing TB control programs, TB clinicians informed enough and confident enough around WGS data to act on their recommendations, and enough public health nurses to appropriately investigate and manage newly discovered clusters of cases.

The GMI initiative is also tackling another barrier to uptake—political involvement. Cementing new platforms into public health practice requires overcoming evidentiary thresholds of multiple stakeholders and obtaining enough buy-in from public health decision-makers to be able to roll out a new platform across an entire health care system. In the resource-constrained public health environment, a convincing case must be made to introduce a technology such as genomics, which requires a significant initial investment in infrastructure and training.

Finally, genomic testing of samples beyond the pathogen itself, i.e., samples of either the human host or his or her microbiota, introduces ethical issues. These include classical issues encountered throughout human genomics, including whether a patient’s data might affect their chances of being insured or, when a family unit is being tested, issues related to paternity, but they also include issues only recently appearing on the community’s radar. For example, to what degree can we rely on predictive biomarkers, such as those that can identify patients at higher risk of progressing from LTBI to active disease, and how does this influence our decision to use preventative therapy in LTBI, especially when the antibiotics used in prophylaxis carry a risk of adverse events?

Challenges aside, the genomic revolution is set to dramatically alter the clinical microbiology landscape, and we expect to see the first inroads in this area in the diagnosis, treatment, and epidemiology of tuberculosis in well-resourced settings. In order to effectively manage what is truly a global TB epidemic, interventions aimed at reducing poverty and health inequalities will be most critical. However, technological advances that create new knowledge about the biology and epidemiology of this important organism can readily be translated into sustainable interventions appropriate for low-resource settings and will also play a role in TB control efforts. This is especially true for regions that transition from high to medium to low incidence and will require increasingly targeted approaches to reducing TB rates. Given recent successes in this area, such as with the GenExpert system, we anticipate a future in which genomics directly contributes to a reduction in the incidence of this important global disease.

ACKNOWLEDGMENTS

We thank the British Columbia Centre for Disease Control Foundation, Genome British Columbia, and the British Columbia Lung Association for their support of our TB genomics research.

REFERENCES


Brandis G, Hughes D. 2013. Genetic characterization of compensatory
Improving the Management of Tuberculosis with Genomics

61. Shrivastava, 50. Reiling, Gardy, Comas, Sandgren

57. 52.

49.

51.


Marta Wlodarska has 6 years of research experience in infectious disease and immunology and is currently a Postdoctoral Fellow at Novartis Institutes for Biomedical Research in Boston, MA. She obtained her B.Sc. with Honors at the University of British Columbia (UBC), Vancouver, Canada. She also completed her Ph.D. at the University of British Columbia, in the laboratory of B. Brett Finlay. Her research focused on discovering immune mechanisms required to maintain intestinal health and their relevance to enteric infection. After completing her Ph.D., she moved to the British Columbia Centre for Disease Control (BCCDC) in Vancouver to gain experience in public health and to understand the transmission dynamics of infectious disease, with a focus on the molecular epidemiology of tuberculosis.

James C. Johnston is a Clinical Assistant Professor in the Department of Medicine at UBC and a Michael Smith Foundation Scholar. He received his bachelor’s degree from McGill and his M.D. from Queen’s University. He returned to McGill for internal medicine training before moving to UBC for his Respiratory Medicine Fellowship. He completed postdoctoral training in tuberculosis research at the BC Centre for Disease Control and has a master’s degree in public health from the Harvard School of Public Health. Dr. Johnston joined the BCCDC TB Services as the TB Evaluation Lead in 2010. His interests include TB genomics, TB screening in migrants to countries with low TB incidence, and MDR TB epidemiology and treatment.

Jennifer L. Gardy is a Senior Scientist at the British Columbia Centre for Disease Control and an Assistant Professor in UBC’s School of Population and Public Health. She received a bachelor’s degree in cell biology and genetics from UBC in 2000, completed a Ph.D. in computational microbial genomics at Simon Fraser University in 2006, and did 3 years of postdoctoral training in network analysis and systems biology at UBC before joining the BCCDC in 2009. Her research interests are in the use of genomics as a tool for understanding the origins and spread of communicable disease, with a particular emphasis on TB. She is particularly interested in using genomics to reconstruct outbreaks of infectious disease and translating the knowledge gained from genomics studies into real-world public health interventions.

Patrick Tang is a Medical Microbiologist at the British Columbia Centre for Disease Control and a Clinical Associate Professor in Pathology and Laboratory Medicine at the University of British Columbia. He completed his M.D./Ph.D. at UBC, followed by a medical microbiology residency at the University of British Columbia. He completed his M.D./Ph.D. at UBC, followed by a medical microbiology residency at the University of British Columbia. He has led the British Columbia TB/Mycobacteriology Laboratory and the Core Molecular/Genomics Laboratory at the BCCDC. His research interests are in the application of genomics and metagenomics to public health and infectious diseases, as well as the development of new molecular diagnostics. He uses these tools to diagnose outbreaks and unusual clinical diseases at the BCCDC. His other research projects include searching for infectious agents in chronic diseases and outbreaks, using genomics to investigate the dynamics of tuberculosis transmission, studying the microbiomes of people and animals, and even looking for better ways to test for water pollution.