HHS Public Access

Author manuscript

Trends Microbiol. Author manuscript; available in PMC 2023 November 01.

Published in final edited form as:

Trends Microbiol. 2022 November ; 30(11): 1036–1044. doi:10.1016/j.tim.2022.04.005.

Connecting the dots: understanding how human mobility shapes TB epidemics

Tyler S. Brown[1,2], **D. Ashley Robinson**[3], **Caroline O. Buckee**[1], **Barun Mathema**[4,*]

[1]Center for Communicable Disease Dynamics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[2]Infectious Diseases Division, Massachusetts General Hospital, Boston, MA, USA

[3]Department of Microbiology and Immunology, University of Mississippi Medical Center, Jackson, MS, USA

[4]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA

Abstract

Tuberculosis (TB) remains a leading infectious cause of death worldwide. Reducing TB infections and TB-related deaths rests ultimately on stopping forward transmission from infectious to susceptible individuals. Critical to this effort is understanding how human host mobility shapes the transmission and dispersal of new or existing strains of *Mycobacterium tuberculosis* (*Mtb*). Important questions remain unanswered. What kinds of mobility, over what temporal and spatial scales, facilitate TB transmission? How do human mobility patterns influence the dispersal of novel *Mtb* strains, including emergent drug-resistant strains? This review summarizes the current state of knowledge on mobility and TB epidemic dynamics, using examples from three topic areas, including inference of genetic and spatial clustering of infections, delineating source–sink dynamics, and mapping the dispersal of novel TB strains, to examine scientific questions and methodological issues within this topic. We also review new data sources for measuring human mobility, including mobile phone-associated movement data, and discuss important limitations on their use in TB epidemiology.

Introduction

The future of global TB control rests ultimately on our ability to avert disease transmission from infectious to susceptible individuals [1,2]. This goal remains elusive, with progress toward it impeded by both limited implementation of existing interventions [3] and continued gaps in our scientific understanding of TB transmission [4-6]. Renewed investment in TB transmission science [7] has led to important insights into individual and environmental determinants of transmission (including *inter alia* HIV-mediated immune suppression, environmental exposure to inhaled pollutants, and exposure to congregate

*Correspondence: bm2055@cumc.columbia.edu (B. Mathema).

Declaration of interests

No interests are declared.

living environments) [4,5] and has expanded the evidence base on transmission-interrupting interventions for TB [8,9]. Concurrent with these advances, whole genome sequencing (WGS) has yielded unprecedented new insights into the transmission of *Mtb* in both highly endemic [10-12] and low-incidence contexts [13,14].

Progress on other key scientific questions relevant to TB transmission has been slow by comparison. Importantly, our understanding of how human host mobility, a central force in other infectious disease epidemics [15-17], influences transmission in TB is still limited [18]. An important body of literature has explored mobility as an individual risk factor for TB exposure [19-22], but work examining mobility as a determinant of population-level TB dynamics is still lacking [18]. Multiple important questions in this area remain largely unexplored. What kinds of mobility, over what geographic and temporal scales, facilitate contact between infectious and susceptible individuals? How does host mobility influence our ability to identify TB transmission clusters or delineate sources and sinks of new TB cases? How does mobility shape the geographic dispersal of new TB strains, including emergent drug-resistant strains [10]?

Understanding how human mobility shapes epidemic dynamics is a longstanding and ongoing scientific question [16]. For directly transmitted infections with short incubation periods and straightforward clinical diagnoses, host mobility patterns are reflected directly in the spatiotemporal distribution of observed cases. These epidemics (including measles and influenza) typically exhibit distinct wave-like dispersal across space [15,23]. However, for infections with more complicated transmission dynamics, including those with extended or variable latent periods or vector-mediated transmission, the distribution of incident cases over space and time is less directly linked to host mobility. Recent studies in malaria [17,24], chikungunya [25], and dengue [26] have demonstrated how empiric estimates of human mobility flows, combined with novel analytical approaches, can reveal important insights into epidemics with more complicated dynamics. For example, mobile phone-associated mobility data has helped to uncover how holiday travel during Eid drove dispersal of the 2017 chikungunya epidemic in Bangladesh [25].

Despite these advances, we still know very little about how human host mobility shapes TB transmission and dispersal. What explains this discrepancy? Like other pathogens that can cause chronic and asymptomatic infections [27], there are several features of *Mtb* infection that may confound attempts to understand mobility-related determinants of transmission. Preinfectious latency is highly variable in TB, lasting decades in some individuals and months or weeks in others [28,29], and many others never develop active disease. Subclinical TB, in which individuals have active transmissible infection with minimal or no symptoms, poses a related challenge to tracking TB transmission [28]. Individuals may relocate during latent infection, obfuscating spatial patterns in transmission [30,31]. Likewise, individual mobility patterns may change considerably in the time between exposure and the development of active (i.e., symptomatic), transmissible disease, particularly during extended periods of latent infection. In addition, infectiousness is highly variable between individuals with active disease [32] (with 10- to 20-fold variation observed in experimental studies [33]), confounding efforts to disambiguate individual mobility patterns versus disease physiology as drivers of increased transmission. Individuals

with active TB may be more infectious if they have a physiologically more transmissible infection [34] or if they have a greater number of effective contacts with susceptible individuals due to increased mobility and other behavioral factors. Importantly, the mutation rate of *Mtb* is markedly slower than it is for most other epidemic pathogens [35], limiting the utility of many genomic epidemiological approaches that have proven useful elsewhere [36]. Lastly, multiple technical obstacles, including lack of reliable mobility data in most highly endemic countries [37], and challenges inherent to mathematical modeling of TB epidemics [27], have impeded deeper understanding of mobility and TB transmission.

This review focuses on three specific topic areas with direct relevance to understanding and preventing TB transmission: identifying clustered infections, delineating source–sink dynamics, and mapping the dispersal of novel TB strains. For each topic area we review existing research, outline unanswered questions or limitations, and, by examining parallel work on other epidemics, identify opportunities for advancing our understanding in this important field. We use this framework to inform a wider discussion on the questions listed above and revisit the challenges inherent to studying mobility and TB epidemiology. Lastly, we discuss available data sources for measuring human mobility (including mobile phone-associated movement data [37]), their potential applications to TB epidemiology, and important limitations on their use.

Topic 1: Genetic and spatial clustering of incident infections

Detecting clusters of incident cases, and identifying groups of individuals and/or locations with ongoing transmission, can provide important information for TB control efforts. This information can inform geographically or demographically targeted interventions, including intensified case-finding and screening efforts. TB cases are considered to be genetically clustered if their respective *Mtb* isolates are genetically similar to one another, as compared to a prespecified threshold of genetic similarity that varies across genotyping methods. *Mtb* has a relatively slow mutation rate, and *Mtb* populations are in general genetically less diverse than other disease-relevant microbes [38,39]. In TB outbreaks driven by infections with short latency periods and relatively rapid progression to active disease (for example, those occurring in populations with high rates of HIV coinfection), outbreak-associated isolates typically exhibit highly restricted genetic diversity and often complete clonality [10]. Whole genome sequencing (WGS) has largely replaced earlier *Mtb* genotyping methods (reviewed in Mathema *et al.* [40]) that are limited in their power to discriminate between closely related *Mtb* isolates. WGS-based methods have uncovered previously unknown diversity within *Mtb* populations [41] and revealed important aspects of TB transmission in both high transmission [11,12,42] and low transmission [13,14,43] contexts. Multiple methods exist for inferring genetic clustering of infections from WGS data, including simple single-nucleotide polymorphism (SNP) difference thresholds (reviewed in Hatherell *et al.* [44]) and novel probabilistic methods that combine SNP difference, timing of infections, and an *Mtb* molecular clock rate [45]. There remain important challenges for all of these methods, including those related to sampling bias, polyclonal infections, and variability of within-host evolution between individuals [46,47]. Recent studies have found significant within-host diversity in some *Mtb* infections [48,49], even within a single pulmonary cavity [50], which may have important implications for sampling *Mtb* isolates

from TB cases. Specifically, clinical samples from TB patients (including expectorated sputa, bronchoalveolar lavage, or samples from other anatomical sites), may capture only a limited subset of the multiple diverse *Mtb* subpopulations present in a given infection. Likewise, different TB subpopulations within individual patients may be more or less able to grow in standard laboratory cultures [51], introducing another potential source of bias that may influence our ability to infer transmission between individuals. These features may have important implications for TB surveillance programs that use SNP difference thresholds to infer linked transmission between TB cases [52].

'Spatial' clustering refers to infections that occur in close physical proximity to another, resulting in a geographically localized excess number of incident TB cases in a given area (i.e., a spatial cluster of infections). Methods for inferring spatial clusters of TB infections are similar to those used for other infectious diseases, and their use and limitations have been reviewed in detail [53]. Sampling bias, introduced via geographic heterogeneity in TB case detection, and confounding, introduced via concurrent reactivation TB in colocated individuals with shared risk factors, are important issues for these methods [53].

Prior to the introduction of WGS, a number of studies identified limited spatial aggregation of genetically clustered TB cases [54-56]. There are likely multiple factors that explain this phenomenon. These include the lower discriminatory power of pre-WGS genotyping methods, leading to misclassification of clustered versus nonclustered cases [57]. Using WGS data, which can distinguish isolates at the limit of heritable strain differences, largely obviates this problem and has enabled improved identification of infection clusters that would not be detected via spatial clustering methods (for example, closely related infections propagated via spatially dispersed or otherwise cryptic contact networks). Movement of individual hosts, during the extended period over which TB infections and disease unfold, may obfuscate spatial cohesion between cases with shared transmission links [30]. Transmission outside the home, an important contributor in highly endemic contexts [58], may link cases separated by larger geographic distances; understanding what these locations are, and how they contribute to longer-distance linkages between cases, will likely require more detailed data on daily and long-term mobility behaviors [19]. Importantly, many other epidemics exhibit spatially incoherent dispersal patterns, in which geographic distance is poorly correlated with linked transmission [59]. For example, epidemic arrival times for influenza and severe acute respiratory syndrome (SARS) dispersal are poorly correlated with geographic distance, but highly correlated with effective connectivity, as measured by airline traffic volume between locations [59]. Importantly, infections with longer incubation periods (like Ebola) typically exhibit less spatially coherent dispersal than those with short incubation periods (like cholera), which exhibit more spatially organized, wave-like dispersal [31]. Thus, spatial clustering methods that rely on geographic distance alone can fail to detect clustering that may be captured if alternative measures of 'effective connectivity' are considered (for example, host movement between locations, Figure 1) [60]. Revealing these hidden patterns of connectivity, and resolving how TB cases are linked across complex spatial networks, would provide important, actionable guidance for TB control interventions, particularly in highly endemic countries. Specifically, identifying these more complex or cryptic networks may guide the development and implementation

of interventions targeting networks of individuals or locations that are highly linked by mobility.

Topic 2: Source–sink dynamics and spatially-targeted interventions

Source–sink dynamics, in which highly endemic locations ('hotspots') serve as sources of infections exported to lower-incidence sink locations, are important for propagating and sustaining epidemics. Important examples of this phenomenon include malaria [17,61,62], Ebola virus disease [63], and chikungunya [25]. TB is typically heterogeneous across populations, and TB hotspots in both high- and low-burden countries often colocalize with areas where higher proportions of residents experience relative social disadvantage or marginalization (based on poverty, class, or racialization) [64-66]. In addition, there is evidence from mathematical modeling studies indicating that reducing transmission in hotspots can reduce overall transmission in both source and sink locations [67]. Thus, reducing transmission in hotspots may be a high value intervention for TB control programs. For these reasons, there is growing interest in spatially targeting TB control interventions in higher-incidence source locations [68], but evidence supporting these approaches is mixed [8] and additional evaluation of their effectiveness is still needed [9].

Improved understanding of host mobility may be helpful for the design and implementation of these spatially targeted interventions. For example, if hotspots are identified via spatial clustering methods they may (as described above) fail to capture spatially incoherent populations that, although not geographically contiguous with one another, may be strongly linked via mobility. Thus, spatial clustering may fail to detect these 'hidden' or 'dispersed' hotspots (or 'hot networks'), or fail to identify their full extent, resulting in improperly targeted spatial interventions. In addition, studies in other epidemic pathogens [24] have highlighted the importance of delineating connectivity, and thus mobility, between sources and potential sinks, given that this information can help to prioritize non-source areas where importation of infections from source locations are most likely. In the context of TB, this information could potentially guide surveillance and case detection activities, and may be particularly useful for understanding the dispersal of drug-resistant TB strains (as described below). Implementing disease control interventions in transmission 'hubs', that is, TB hotspots that are also highly connected nodes in transmission-relevant mobility networks, may yield outsized TB control benefits compared to interventions in isolated or poorly connected locations.

Topic 3: Emergence and dispersal of novel TB strains

Emergent *Mtb* strains, including those with novel drug-resistance phenotypes, are often identified only after they have achieved sustained transmission and dispersed into larger geographic areas away from their origin location [10,12,42]. Epidemic transmission of specific drug-resistant strains [12], including widespread community transmission via household and casual contact [42], is the primary driver of new drug-resistant TB cases in most highly endemic settings. Improved understanding of how mobility shapes transmission and broader geographic dispersal of these strains could have important implications for the early detection of these threats to global health. This work may also guide early control

interventions to stop the transmission of new drug-resistant *Mtb* strains at the critical juncture between local outbreak and widespread epidemic.

Novel methods examining the population genetic signatures of geographic dispersal have yielded important insights into the geographic origins of drug-resistant pathogens, including methicillin-resistant *Staphylococcus aureus* in the USA [69] and extensively drug-resistant *Mtb* in South Africa [10]. Briefly, these methods, initially developed for studying wildlife and other nonmicrobial populations, use geolocated bacterial genome sequences to characterize genetic changes that are expected to arise as a population expands across a larger geographic range. These signatures are generated during serial founder events as a given strain spreads to new locations away from its origin (Figure 2) [70]. Relevant population genetic signatures include decreasing genetic diversity [71] and increasing derived allele frequency within populations [72], and increasing genetic differentiation between populations (typically measured using estimators of the fixation index, F_{ST}) with increasing distance away from the origin location. The serial founder events are also expected to reduce effective population size and may reduce the efficacy of natural selection – a phenomenon referred to as 'expansion load' [73].

These methods are still early in their development and there are important technical issues that are relevant for applying these methods to TB epidemiology. For example, the time required for detectable signatures of range expansion to accrue is expected to vary across different microbial populations undergoing range expansion, due to both intrinsic biological factors (including mutation rate) and extrinsic epidemiological factors that may shape the magnitude and frequency of bottlenecking during serial founder events. Very short time intervals between strain origin and sampling, which may be expected in the case of new drug-resistant TB strains, may require very sensitive measures of genetic diversity and differentiation to detect differences within and between populations. Data size requirements for evaluating population genetic signatures of range expansion, including requirements for the number of genetic markers that need to be typed and the number of isolates that need to be sampled, are still not clear, although existing and ongoing work may provide some guidance [74].

These methods are also subject to many of the same mobility-related issues that challenge spatial-clustering methods. Thus far, studies have relied on geographic distances, including Euclidean and shortest road distances, to represent spatial changes in genetic signatures [10,69], but efforts to infer strain origin locations may be unsuccessful if geographic distance does not correspond with true connectivity between individuals in different locations. Likewise, underlying human mobility networks, which may prove to be informative for predicting where new drug-resistant TB strains will disperse, are often spatially incoherent and effective connectivity is often poorly correlated with geographic distance [75]. Recent studies on chikungunya and dengue have highlighted how incorporating empiric estimates of human mobility can markedly improve forecasting and prediction of epidemic dispersal (for example, from urban Dhaka and Bangkok into outlying areas of Bangladesh [25] and Thailand [76], respectively). Using such data to understand and detect early dispersal of novel TB strains could potentially yield important improvements to existing approaches.

Finally, existing models of geographic range expansion that were developed with empirical data from nonmicrobes may require modifications to be more appropriate for studying emergent pathogens. The population genetic signatures that are expected under a range expansion models assume a series of short-distance dispersals. Long-distance dispersals, via founder events that 'leapfrog' over more proximate locations, can strongly influence observed spatial patterns of genetic diversity [77,78]. Even if relatively rare compared to short-distance dispersals, these long-distance dispersals can form irregular and spatially incoherent patterns in genetic diversity [79]. Incorporating empiric data on connectivity between locations may help to account for long-distance dispersals, but this approach may still not completely resolve this issue. More work is needed in order to better understand the application of these methods to real world data from emerging epidemics.

Data sources for studying human mobility and TB epidemic dynamics

Identifying informative data sources on human mobility, including those that can delineate movement patterns relevant to TB transmission, is an important research goal. Desirable features of such a data source would include: (i) capture of movement patterns for both infectious and susceptible individuals, with minimal and measurable selection bias across demographic, socio-economic, or other important population subgroups; (ii) spatial and temporal resolution adequate for identifying nonhousehold locations where individuals may interact; and (iii) adequate capture of different epidemiologically relevant types of mobility, including daily recurrent movements (commuting), seasonal migration, and long-term or permanent relocation. Few if any data sources will have all of these features, and many analyses may involve multiple kinds of data, including both conventional and newer (digital) data sources. Conventional data sources include travel surveys collected via random household sampling, which can measure the location, distance, duration, and purpose of different kinds of human movement in a given population, and individual-level data on specific groups of interest (for example, new TB cases [19]). Recent work characterizing the 'activity space' of individual TB cases has provided early insights into new types of movement and location data that can be used for understanding TB transmission [80-82]. These approaches account for the inherent 'spatial polygamy' of human movement [83], mapping the multiple locations an individual visits over time rather than anchoring on a single location (e.g., home address at the time of TB diagnosis), and thus provide much richer information on the spatiotemporal space in which TB transmission may occur. Convergent approaches, in which novel data sources (for example, mobile device-associated movement data) are used to measure individuals' activity spaces, may provide highly useful information for studying TB epidemic dynamics. Lastly, extending data collection efforts to include not only TB-infected individuals but also noninfected individuals at risk for transmission, may be important for more fully characterizing the mobility-related determinants specific to TB transmission.

The use of mobile phone-associated movement data, including data records from mobile network operators and GPS-enabled mobile applications, requires robust safeguards to protect end-user privacy [84]. Anonymization of data records alone is insufficient in this regard, given that individual users can be readily identified by the locations they visit frequently [84,85]. In fact, individual users can be identified in large mobility datasets

using as few as four time-location data points [85]. Precomputed, aggregated mobility indicators (for example, matrices of total trips between locations) are an important tool for protecting privacy [86], and even coarsely aggregated indicators can be informative for some applications (specifically, modeling influenza dispersal [87]). Developing aggregation methods and research protocols that both protect privacy and meet the criteria for usefulness described above is an important research objective.

Concluding remarks

Understanding how human movement contributes to TB transmission and dispersal is an important front in global efforts to curb new TB infections. As described here, achieving this goal will require researchers to solve multiple important challenges, some attributable to inherent features of TB infection and disease, some attributable to the difficulties in measuring relevant human mobility patterns, and some due to the confluence of these factors. Overcoming these challenges will likely require a wide-reaching, multidisciplinary scientific effort drawing on expertise from infectious disease epidemiologists, modelers, population geneticists, and specialists in the use of human mobility data (see Outstanding questions).

References

1. Migliori GB et al. (2019) Reducing tuberculosis transmission: a consensus document from the World Health Organization Regional Office for Europe. Eur. Respir. J 53, 1900391 [PubMed: 31023852]
2. Sekkides O (2019) Understanding tuberculosis transmission might be the gamechanger we need. Lancet Infect. Dis 19, e63 [PubMed: 30554994]
3. Reid MJA et al. (2019) Building a tuberculosis-free world: The Lancet Commission on tuberculosis. Lancet 393, 1331–1384 [PubMed: 30904263]
4. Churchyard G et al. (2017) What we know about tuberculosis transmission: an overview. J. infect. Dis 216, S629–S635 [PubMed: 29112747]
5. Khan PY et al. (2019) Transmission of drug-resistant tuberculosis in HIV-endemic settings. Lancet infect. Dis 19, e77–e88 [PubMed: 30554996]
6. Mathema B et al. (2017) Drivers of tuberculosis transmission. J. infect. Dis 216, S644–S653 [PubMed: 29112745]
7. Auld SC et al. (2017) Research roadmap for tuberculosis transmission science: where do we go from here and how will we know when we're there? J. infect. Dis 216, S662–S668 [PubMed: 29112744]
8. Cudahy PGT et al. (2019) Spatially targeted screening to reduce tuberculosis transmission in high-incidence settings. Lancet Infect. Dis 19, e89–e95 [PubMed: 30554997]
9. Dowdy DW et al. (2017) Designing and evaluating interventions to halt the transmission of tuberculosis. J. Infect. Dis 216, S654–S661 [PubMed: 29112743]
10. Brown TS et al. (2019) Pre-detection history of extensively drug-resistant tuberculosis in KwaZulu-Natal, South Africa. Proc. Nati. Acad. Sci. U. S. A 116, 23284–23291
11. Nelson KN et al. (2018) Spatial patterns of extensively drug-resistant tuberculosis (XDR-tuberculosis) transmission in KwaZulu-Natal, South Africa. J. Infect. Dis 218, 1964–1973 [PubMed: 29961879]
12. Shah NS et al. (2017) Transmission of extensively drug-resistant tuberculosis in South Africa. N. Engi. J. Med 376, 243–253
13. Walker TM et al. (2013) Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect. Dis 13, 137–146 [PubMed: 23158499]

14. Walker TM et al. (2014) Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. Lancet Respir. Med 2, 285–292 [PubMed: 24717625]
15. Grenfell BT et al. (2001) Travelling waves and spatial hierarchies in measles epidemics. Nature 414, 716–723 [PubMed: 11742391]
16. Prothero RM (1977) Disease and mobility: a neglected factor in epidemiology. Int. J. Epidemio 6, 259–267
17. Wesolowski A et al. (2012) Quantifying the impact of human mobility on malaria. Science 338, 267–270 [PubMed: 23066082]
18. Okano JT and Blower S (2019) New conceptual framework for tuberculosis transmission. Lancet Infect. Dis 19, 578
19. Robsky KO et al. (2021) Characterization of geographic mobility among participants in facility- and community-based tuberculosis case finding in urban Uganda. PLoS One 16, e0251806 [PubMed: 33989343]
20. Long Q et al. (2008) Barriers to accessing TB diagnosis for rural-to-urban migrants with chronic cough in Chongqing, China: a mixed methods study. BMC Health Serv. Res 8, 202 [PubMed: 18828929]
21. Stuckler D et al. (2011) Mining and risk of tuberculosis in sub-Saharan Africa. Am. J. Public Heaith 101, 524–530
22. Chen KS et al. (2016) Tuberculosis transmission and risk factors in a Chinese antimony mining community. Int J. Tuberc. Lung Dis 20, 57–62 [PubMed: 26688529]
23. Viboud C et al. (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 312, 447–451 [PubMed: 16574822]
24. Chang HH et al. (2019) Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. eLife 8, e43481 [PubMed: 30938289]
25. Mahmud AS et al. (2021) Megacities as drivers of national outbreaks: The 2017 chikungunya outbreak in Dhaka, Bangladesh. PLoS Negi. Trop. Dis 15, e0009106
26. Salje H et al. (2021) Reconstructing unseen transmission events to infer dengue dynamics from viral sequences. Nat. Commun 12, 1810 [PubMed: 33753725]
27. Childs LM et al. (2015) Modelling challenges in context: lessons from malaria, HIV, and tuberculosis. Epidemics 10, 102–107 [PubMed: 25843394]
28. Kendall EA et al. (2021) The epidemiological importance of subclinical tuberculosis. a critical reappraisal. Am. J. Respir. Crit. Care Med 203, 168–174 [PubMed: 33197210]
29. Pai M et al. (2016) Tuberculosis. Nat. Rev. Dis. Primers 2, 16076 [PubMed: 27784885]
30. Bishai WR et al. (1998) Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA 280, 1679–1684 [PubMed: 9831999]
31. Kahn R et al. (2020) Incubation periods impact the spatial predictability of cholera and Ebola outbreaks in Sierra Leone. Proc. Natl. Acad. Sci. U. S. A 117, 5067–5073 [PubMed: 32054785]
32. Cadena AM et al. (2017) Heterogeneity in tuberculosis. Nat. Rev. Immuno 17, 691–702
33. Wood R et al. (2016) Real-time investigation of tuberculosis transmission: developing the respiratory aerosol sampling chamber (RASC). PLoS One 11, e0146658 [PubMed: 26807816]
34. Turner RD et al. (2017) Tuberculosis infectiousness and host susceptibility. J. Infect. Dis 216, S636–S643 [PubMed: 29112746]
35. Ford CB et al. (2013) *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat. Genet 45, 784–790 [PubMed: 23749189]
36. Drummond AJ et al. (2003) Measurably evolving populations. Trends Ecoi. Evoi18, 481–488
37. Wesolowski A et al. (2016) Connecting mobility to infectious diseases: the promise and limits of mobile phone data. J. Infect Dis 214, S414–S420 [PubMed: 28830104]
38. Comas I et al. (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. PLoS One 4, e7815 [PubMed: 19915672]

39. Sreevatsan S et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Nat . Acad. Sci. U. S. A 94, 9869–9874
40. Mathema B et al. (2006) Molecular epidemiology of tuberculosis: current insights. Clin. Microbiol. Rev 19, 658–685 [PubMed: 17041139]
41. Brown TS et al. (2016) Genomic epidemiology of Lineage 4 *Mycobacterium tuberculosis* subpopulations in New York City and New Jersey, 1999–2009. BMC Genom. 17, 947
42. Auld SC et al. (2018) XDR tuberculosis in South Africa: genomic evidence supporting transmission in communities. Eur. Respir. J 52, 1800246 [PubMed: 30115614]
43. Gardy JL et al. (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N. Eng . J. Med 364, 730–739
44. Hatherell HA et al. (2016) Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC Med. 14, 21 [PubMed: 27005433]
45. Stimson J et al. (2019) Beyond the SNP threshold: identifying outbreak clusters using inferred transmissions. Mol. Biol. Evol 36, 587–603 [PubMed: 30690464]
46. Dheda K et al. (2017) The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. Lancet Respir. Med 5 (4), 291–360
47. Worby CJ et al. (2014) Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. PLoS Comput Bioi. 10, e1003549
48. O'Donnell MR et al. (2019) Early detection of emergent extensively drug-resistant tuberculosis by flow cytometry-based phenotyping and whole-genome sequencing. Antimicrob. Agents Chemother 63, e01834–18 [PubMed: 30670422]
49. Brown TS et al. (2021) Evolution and emergence of multidrug-resistant Mycobacterium tuberculosis in Chisinau, Moldova. Microb. Genom 7, 000620 [PubMed: 34431762]
50. Dheda K et al. (2018) Drug-penetration gradients associated with acquired drug resistance in patients with tuberculosis. Am. J. Respir. Crit. Care Med 198, 1208–1219 [PubMed: 29877726]
51. Chengalroyen MD et al. (2016) Detection and quantification of differentially culturable tubercle bacteria in sputum from patients with tuberculosis. Am. J. Respir. Crit. Care Med 194, 1532–1540 [PubMed: 27387272]
52. Witney AA et al. (2016) Clinical use of whole genome sequencing for *Mycobacterium tuberculosis*. BMC Med. 14, 46 [PubMed: 27004841]
53. Shaweno D et al. (2018) Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med. 16, 193 [PubMed: 30333043]
54. Richardson M et al. (2002) Historic and recent events contribute to the disease dynamics of Beijing-like *Mycobacterium tuberculosis* isolates in a high incidence region. Int J. Tuberc. Lung Dis 6, 1001–1011 [PubMed: 12475147]
55. Feske ML et al. (2011) Including the third dimension: a spatial analysis of TB cases in Houston Harris County. Tuberculosis (Edinb) 91, S24–S33 [PubMed: 22094150]
56. Tuite AR et al. (2013) Epidemiological evaluation of spatiotemporal and genotypic clustering of *Mycobacterium tuberculosis* in Ontario, Canada. Int. J. Tuberc. Lung Dis 17, 1322–1327 [PubMed: 24025385]
57. Shaweno D et al. (2017) A novel Bayesian geospatial method for estimating tuberculosis incidence reveals many missed TB cases in Ethiopia. BMC Infect. Dis 17, 662 [PubMed: 28969585]
58. Wang W et al. (2014) Role of casual contacts in the recent transmission of tuberculosis in settings with high disease burden. Clin. Microbiol. Infect 20, 1140–1145 [PubMed: 24941878]
59. Brockmann D and Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. Science 342, 1337–1342 [PubMed: 24337289]
60. Brown TS et al. (2021) The impact of mobility network properties on predicted epidemic dynamics in Dhaka and Bangkok. *Epidemics* 35, 100441 [PubMed: 33667878]
61. Ihantamalala FA et al. (2018) Estimating sources and sinks of malaria parasites in Madagascar. Nat. Commun 9, 3897 [PubMed: 30254280]

62. Martens P and Hall L (2000) Malaria on the move: human population movement and malaria transmission. Emerg. Infect. Dis 6, 103–109 [PubMed: 10756143]
63. Yang W et al. (2015) Transmission network of the 2014–2015 Ebola epidemic in Sierra Leone. J. R. Soc. Interface 12, 20150536 [PubMed: 26559683]
64. Peterson ML et al. (2019) Extensively drug-resistant tuberculosis 'hotspots' and sociodemographic associations in Durban, South Africa. Int. J. Tuberc. Lung Dis 23, 720–727 [PubMed: 31315705]
65. Alene KA et al. (2017) Spatial patterns of multidrug resistant tuberculosis and relationships to socio-economic, demographic and household factors in northwest Ethiopia. PLoS One 12, e0171800 [PubMed: 28182726]
66. Dememew ZG et al. (2020) The yield of community-based tuberculosis and HIV among key populations in hotspot settings of Ethiopia: A cross-sectional implementation study. PLoS One 15, e0233730 [PubMed: 32469997]
67. Dowdy DW et al. (2012) Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proc. Natl. Acad. Sci. U. S. A 109, 9557–9562 [PubMed: 22645356]
68. Cegielski JP et al. (2014) Eliminating tuberculosis one neighborhood at a time. Am. J. Public Health 104, S214–S233 [PubMed: 24899457]
69. Challagundla L et al. (2018) Range expansion and the origin of USA300 North American epidemic methicillin-resistant *Staphylococcus aureus*. mBio 9, e02016–17
70. Excoffier L et al. (2008) Genetic consequences of range expansions. Annu. Rev. Ecol . Evol . Syst 40, 481–501
71. Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Mol. Ecol 13, 853–864 [PubMed: 15012760]
72. Peter BM and Slatkin M (2013) Detecting range expansions from genetic data. Evolution 67, 3274–3289 [PubMed: 24152007]
73. Peischl S and Excoffier L (2015) Expansion load: recessive mutations and the role of standing genetic variation. Mol. Ecol 24, 2084–2094 [PubMed: 25786336]
74. Brown TS et al. (2021) Distinguishing gene flow between malaria parasite populations. PLoS Genet. 17, e1009335 [PubMed: 34928954]
75. Simini F et al. (2012) A universal model for mobility and migration patterns. Nature 484, 96–100 [PubMed: 22367540]
76. Kiang MV et al. (2021) Incorporating human mobility data improves forecasts of Dengue fever in Thailand. Sci. Rep 11, 923 [PubMed: 33441598]
77. Bialozyt R et al. (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. J. Evol. Biol 19, 12–20 [PubMed: 16405572]
78. Becheler R et al. (2016) Genetic signatures of a range expansion in natura: when clones play leapfrog. Ecol. Evol 6, 6625–6632 [PubMed: 27777735]
79. Paulose J and Hallatschek O (2020) The impact of long-range dispersal on gene surfing. Proc. Natl. Acad. Sci. U. S. A 117, 7584–7593 [PubMed: 32198207]
80. Bui DP et al. (2021) Community transmission of multidrug-resistant tuberculosis is associated with activity space overlap in Lima, Peru. BMC Infect. Dis 21, 275 [PubMed: 33736597]
81. Worrell MC et al. (2017) Use of activity space in a tuberculosis outbreak: bringing homeless persons into spatial analyses. Open Forum Infect. Dis 4, ofw280 [PubMed: 28480272]
82. Izumi K et al. (2015) Detection of tuberculosis infection hotspots using activity spaces based spatial approach in an urban Tokyo, from 2003 to 2011. PLoS One 10, e0138831 [PubMed: 26382251]
83. Matthews SA and Yang TC (2013) Spatial polygamy and contextual exposures (SPACEs): promoting activity space approaches in research on place and health. Am. Behav. Sci 57, 1057–1081 [PubMed: 24707055]
84. de Montjoye YA et al. (2018) On the privacy-conscientious use of mobile phone data. Sci. Data 5, 180286 [PubMed: 30532052]
85. de Montjoye YA et al. (2013) Unique in the crowd: the privacy bounds of human mobility. Sci. Rep 3,1376 [PubMed: 23524645]

86. Kishore N et al. (2020) Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. Lancet Digit. Heath 2, e622–e628

87. Engebretsen S et al. (2020) Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh. J. R. Soc. Interface 17, 20190809 [PubMed: 32546112]

Highlights

Mobility-related determinants of tuberculosis (TB) transmission and dispersal remain poorly understood, lagging behind recent advances for other epidemic pathogens. Multiple inherent features of TB infection contribute to this knowledge gap, including variable and often prolonged latency of infection, the slow mutation rate of *Mycobacterium tuberculosis* (*Mtb*), and high within-host diversity of *Mtb* populations in many patients.

Methods that use geographic distance between infections (for example, those that attempt to detect spatial clustering of infections) may fail to detect meaningful epidemiological patterns if connectivity and transmission linkages between locations are not consistently correlated with distance. Empiric measurement of human mobility patterns may improve the use of these methods in this situation.

Geolocated pathogen whole genome sequence data (i.e., sequence data that is linked to the home or clinic location for incident infections), and spatial patterns of genetic diversity and divergence in these data, have yielded important insights into the geographic origins and dispersal of drug-resistant TB and other epidemics. Models of geographic range expansion, which detect genetic signatures of a population expanding away from its origin, may provide an important tool for understanding the origin of novel TB strains.

Outstanding questions

The mutation rate of *Mtb* is slow, compared to many other epidemic pathogens, and TB outbreaks are often characterized by collections of isolates with very low genetic diversity, posing difficulties for genomic epidemiology approaches that may require a larger amount of genetic variation to make reliable inferences. Do novel approaches exist that can address this difficulty and provide meaningful information on TB transmission and dispersal given limited genetic variation?

Euclidean or other geographic measures of distance may correlate poorly with true connectivity between locations, with important implications for understanding the spatial epidemiology of TB. Can empiric measurements of human mobility patterns (for example, those derived from mobile phone-associated movement data) improve inference of spatial clustering and dispersal patterns in TB?

Population genetic models of range expansion are a promising approach to inferring the origin location and dispersal networks of emerging TB strains, but applying these models to real-world epidemics involves multiple challenges. What improvements are needed to existing models of range expansion, and how can these models benefit from using empiric measurements of human mobility?

Using digital data sources on human mobility involves multiple critical considerations around individual privacy and identifiability. Are there data protocols that can ensure these essential protections and also provide useable, informative data for understanding TB epidemiology?

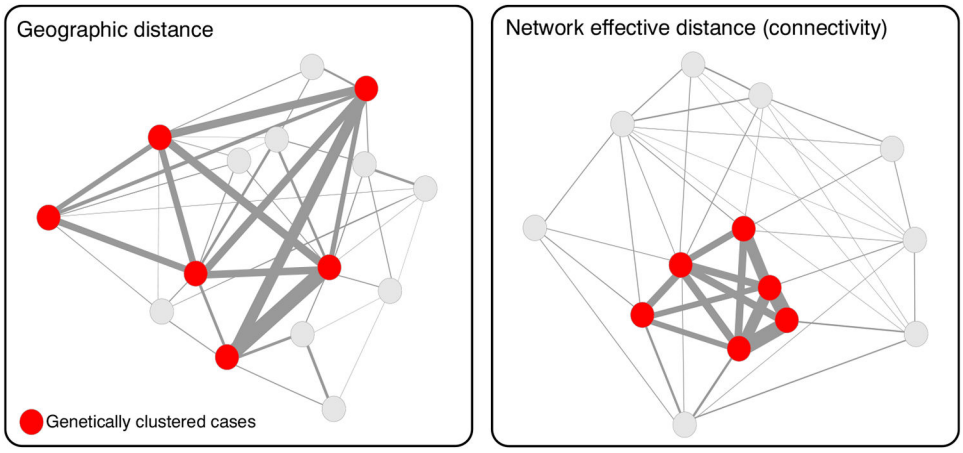

Figure 1. Geographic distance, effective connectivity, and genetic clustering of *Mycobacterium tuberculosis* cases.

Genetically clustered cases may not exhibit spatial clustering if geographic distance is not consistently correlated with effective connectivity between locations (left panel). Considering effective network distances (for example, connectivity via human mobility between locations) may identify clustered infections (right panel).

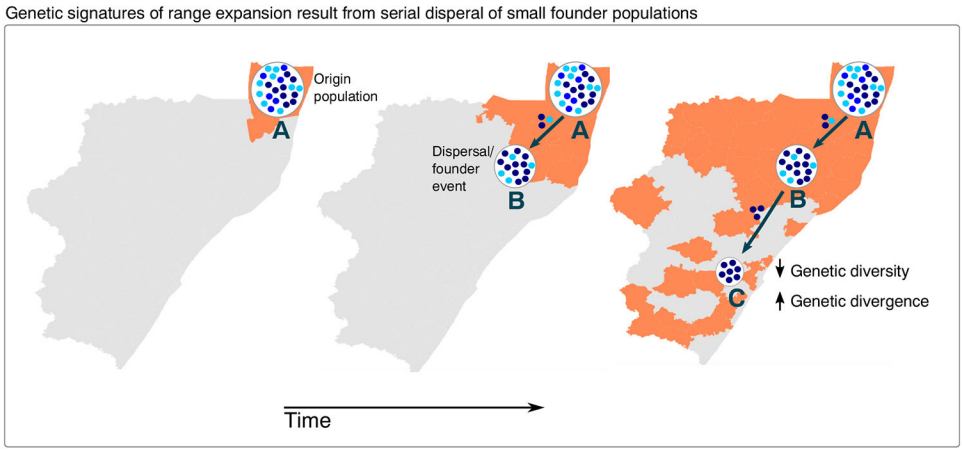

Figure 2. Range expansion of a *Mycobacterium tuberculosis* strain.

During geographic range expansion of a strain, a population spreads via serial founder events that result in distinct population genetic signatures, including decreasing genetic diversity and increasing genetic divergence when compared with samples from the origin population. A ➔ B: short-distance founder events between adjacent territories. B ➔ C: long-distance founder event between noncontiguous territories. Orange territories indicate geographic areas in which a new TB strain is present at a given time.