Journal of Cambodian Health
Volume. 4, No. 1, 2026, v4103
https://doi.org/10.70243/jch2026.4103

OPEN ACCESS

Geospatial Analysis of Disease Transmission and Genetic Diversity

Commentary

Raj Rajeshwar Malinda[1,2*]

[1] Indraprastha Institute of Information Technology Delhi (IIIT-Delhi), New Delhi 110020, India.
[2] Graduate School of Information Science, University of Hyogo, Kobe, Hyogo 651-2197, Japan.

Article history
Received: 6 February 2026
Revised: 26 June 2026
Accepted: 15 July 2026
Published: 5 August 2026

Keywords
Geospatial Data
Disease Transmission
Data Disparities
Public Health

Abstract: This commentary discusses key studies to highlight the importance of geospatial analysis with genome sequencing data to provide insights into the transmission of infectious disease such as tuberculosis and covid, particularly in low-income Southeast Asian countries, including Cambodia. This geospatial genomic framework further explores how geographical-based analysis can help to identify issues of imbalance in historical data collection, processing and sharing on public databases, connecting genetic diversity to geospatial data analysis. In summary, it highlights the significance of integrating geospatial analysis with data-driven genetic and genomic studies, to better understand disease transmission and data disparities across health systems, with a focus on developing nations, for improved healthcare infrastructure and outcomes.

Geospatial Genomic Analysis Importance in Lower-Income Nations

Geospatial data often refers to location-specific geographical information, and thereafter any associated data. Geospatial data have become crucial to understand various phenomena in several fields, such as climate change in environmental sciences and biomedical sciences. Particularly in the public health domain, integrating geography-based information contributes to understanding disease etiology, and further enhances the comprehension of disease prevention, diagnosis, transmission, and treatment (Richardson et al., 2013; Srivastava et al., 2024). Low- and middle-income countries often struggle with adequate healthcare infrastructure to combat disease transmission, especially in the transmission of infectious diseases (Prem et al., 2019; Spies et al., 2024; Su et al., 2024). Previous studies suggest that, in low-income nations, leveraging geospatial data in disease settings can effectively advance health crisis responses (Kattenberg et al., 2024; Malinda, 2024; Shaweno et al., 2021; Shrestha et al., 2021).

In low-income and developing nations including the Southeast Asian block, researchers utilized geospatial and genomic data to study disease transmission (Baker et al., 2023; Malinda & Mishra, 2025; Prem et al., 2019; Spies et al., 2024). To extend, genomic data, increasingly complex due to advancements in technologies, such as next-generation sequencing (NGS), and genome-wide association studies (GWAS), continue to expand daily (Goodwin et al., 2016). Infectious diseases such as tuberculosis, pneumonia, flu, cholera, and the recent addition to the list, coronavirus disease, impose a significant burden on the healthcare system, particularly in underdeveloped and developing economies, including the Southeast Asian nation Cambodia (Azzopardi et al., 2023; Levin et al., 2022; Prem et al., 2019; Spies et al., 2024; Su et al., 2024).

Disease transmission is spatially heterogeneous, and therefore requires multi-scale approaches for effective understanding, thereby genomic data becoming one of the most informative when interpreted within their geographical context (Lin & Wen, 2022; Malinda, 2024). The value of geospatial genomics

***Corresponding author**: Raj Rajeshwar Malinda, rrmalinda@gsis.u-hyogo.ac.jp

lies not merely in locating outbreaks but in revealing transmission patterns that conventional epidemiological surveillance may overlook (Grad & Lipsitch, 2014; Lin & Wen, 2022). Recent studies suggest that the principal contribution of geospatial genomic integration resides not simply in improved disease mapping, for example, in the COVID-19 epidemic, but rather a more informative understanding of how disease transmission processes are structured across space (Lin & Wen, 2022; Su et al., 2024). Likely to understand the growing interests and importance, traditional surveillance systems often quantify disease burden at administrative scales, whereas the combination of genomic sequencing and spatial analysis can reveal localized transmission networks, emerging hotspots, and heterogeneous patterns of spread that remain obscured in aggregated epidemiological data in traditional settings (Belman et al., 2022; Grad & Lipsitch, 2014; Lin & Wen, 2022; Zeghbib et al., 2023).

Evidence from the Southeast Asia illustrates such points. In Cambodia, to understand the importance of geospatial integrated data-driven genomic analysis in disease transmission, a few case studies highlighted the public health concerns of the COVID-19 and tuberculosis transmission, as shown in Table 1 (Prem et al., 2019; Su et al., 2024). Cambodia experienced a high emergence of the COVID-19 variants alpha and delta, translating into a high public healthcare burden. At that time, researchers utilized spatiotemporal evaluation analysis to track the disease dynamics during specific waves of the pandemic, highlighting the value of integrating efforts of genomic data sequencing with geographic surveillance to understand epidemic progression (Su et al., 2024). In another study, using geospatial Bayesian statistics, researchers pointed out active tuberculosis burden prevalence in the Cambodian population. It is further demonstrated that, location-specific factors influence transmission dynamics and intervention priorities. Such studies concluded that transmission is not uniformly distributed across the population but instead exhibits distinct spatial signatures that can inform targeted public health responses. Furthermore, it highlights the importance of geographical data in the mapping of disease prevalence and control efforts of disease transmission impact in Cambodia (Prem et al., 2019).

The importance of spatial heterogeneity becomes even more apparent when comparing findings across different settings, and geographical locations. To highlight such cases, another study of drug-susceptible and multidrug-resistant of tuberculosis using spatial distribution analysis in Ho Chi Minh, a densely populated city in the Southeast Asian region was conducted (see Table 1) (Spies et al., 2024). Interestingly, spatial autocorrelation was further analysed using global Moran's I and Getis-Ord Gi* statistics, and findings revealed substantial variation in the distribution of drug-susceptible and multidrug-resistant tuberculosis cases between urban wards. Using global Moran's I statistic, a weak positive correlation was revealed among parameters, while the Getis-Ord Gi* statistic identified various hot locations of infections. Such strategic demographics that were spatially targeted were able to conclude various insights of disease transmission (Spies et al., 2024). It demonstrates that disease transmission operates through localized ecological and social contexts rather than across homogeneous urban populations.

In the African continent, a similar yet profound case of tuberculosis transmission was observed. Tuberculosis is one of the leading causes of mortality in lower-middle-income countries. In this study from Botswana, researchers integrated whole-genome sequencing and geospatial analysis to track potential disease transmission pathways, summarized in Table 1 (Baker et al., 2023). Their approach identified high burden transmission areas based on clustering methods, and, therefore, suggested timely preventive measures during major outbreaks. Despite considerable geographical differences between Southeast Asia and Africa, such studies propose a similar observation, that integrating genomic sequencing and geospatial data improves the identification of local and global disease transmission

clusters, and provide actionable evidence for policymakers in implementing effective measures and interventions, especially in low and middle-income countries (Malinda, 2024).

Table 1. Summary of case studies discussed

Reference study	Region	Key findings
(Prem et al., 2019)	Southeast Asia	Applied geospatial Bayesian statistics to identify active tuberculosis burden and prevalence
(Su et al., 2024)	Southeast Asia	Investigated COVID-19 Alpha and Delta variant transmission using spatiotemporal evaluation analysis
(Spies et al., 2024)	Southeast Asia	Examined spatial distribution of drug-susceptible and multidrug-resistant tuberculosis cases
(Baker et al., 2023)	African continent	Integrated whole-genome sequencing with geospatial analysis to track tuberculosis transmission

Collectively, these studies suggest that geographical genomic data analysis should be viewed as more than a methodological combination (see Table 1). Such an approach provides a conceptual framework for understanding the spatial organization of disease transmission mapping (see Figure 1). Linking with disease transmission with its geographical context, the integration of geospatial genomic data generates insights that are critical for outbreak detection, resource allocation, and precision public health. These benefits are particularly important in resource-constrained settings, and in lower-income settings, where the efficient targeting of interventions is essential (Baker et al., 2023; Getchell et al., 2024; Grad & Lipsitch, 2014; Lin & Wen, 2022; Prem et al., 2019; Spies et al., 2024; Su et al., 2024).

Furthermore, beyond understanding the disease transmission, the highlighted case studies expose a broader context, that both disease transmission and genomic evidence are spatially structured (Brito et al., 2022; Corpas et al., 2025). The ability to identify transmission hotspots in Cambodia, Vietnam, and Botswana (see Table 1) depends fundamentally on where genomic surveillance is conducted and where spatial data are available. Consequently, regions with limited sequencing capacity or inadequate geospatial surveillance risk becoming epidemiological unexplored areas (Corpas et al., 2025; Peng et al., 2023). To understand this perspective, geospatial genomics serves not only as a tool for understanding disease spread but also as a means of identifying inequalities in data distribution within public health (Brito et al., 2022; Corpas et al., 2025; Grad & Lipsitch, 2014; Lin & Wen, 2022). Therefore, such approaches can be applied to reveal geographical gaps in genomic surveillance and further connect to genetic data representation, discussed in the following section.

Geographical Distribution Connecting to Genetic Diversity Patterns

Geographical distribution patterns also reveal insights into genetic diversity (Brito et al., 2022; Corpas et al., 2025; Peng et al., 2023). In a sense, it is argued that biological diversity and distribution patterns can be understood through variations in available genetic data (Toczydlowski et al., 2021), such as sampling bias in phylogeography (Gámbaro et al., 2025; Lemey et al., 2020; Liu et al., 2022), geospatial genomic transmission (Baker et al., 2023), and database-level spatial imbalance (Corpas et al., 2025; Peng et al., 2023). It is, now, a well-established phenomenon, that advancements in generating genomic data have tremendously increased since the human genome project began. However, challenges persist in production, and sharing geographically specific genetic data, as noted by the scientific community (Chen et al., 2024; Corpas et al., 2025; Peng et al., 2023), because data sparsity itself is considered spatially patterned (Corpas et al., 2025). In order to understand the geographical distribution, it is also important

to emphasize how genetic data are geographically concentrated and create spatial data imbalance in available public databases, further understanding the concerns in the public health, as shown in Figure 1.

To address the current status and distribution of genetic data linked to geographical locations, researchers highlighted taxonomical imbalance within genetic datasets. They further identified patterns of disproportionate representation in data sharing by analyzing millions of nucleotide sequencing datasets (Corpas et al., 2025; Peng et al., 2023). Such disparities highlight the need for comprehensive analyses to address local and global genetic data distribution issues. However, it is further emphasized that genetic data are concentrated in a few geographic areas, and noted that many sequences do not have location-specific metadata, further limiting our overall understanding of how geospatial data are distributed across the scales (Corpas et al., 2025; Liu et al., 2022).

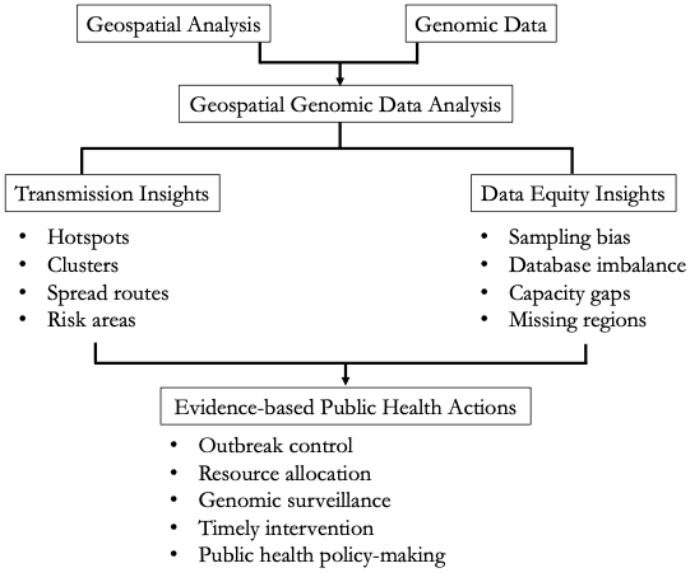

Figure 1. Illustrative representation of geospatial genomic data analysis for public health actions.

Interestingly, the global distribution of shared sequence data is mostly concentrated at a few geographical coordinates, depending on the public databases used in the study (Peng et al., 2023). Such patterns reflect gaps in historical dataset collection, further limiting our understanding of large-scale biodiversity and genetic variations. For example, spatiotemporal sampling bias in genomic data collection has raised concerns regarding the data-sharing strategies, and furthermore genomic surveillance for practical implications in geography-based approaches (Corpas et al., 2025; Gámbaro et al., 2025; Lemey et al., 2020; Liu et al., 2022; Peng et al., 2023). In general, unexplored data locations can bring more information, in addition to known clusters in genomic data collection (Brito et al., 2022). Such arguments shift the overall uneven representation of genomic datasets, and therefore perpetuates bias and health disparities. Nevertheless, these gaps in pre-existing genetic databases present remarkable opportunities for future researchers to address, cite and explore (Fatumo et al., 2022; Peng et al., 2023). It further encourages genomic researchers to actively adopt inclusive practices in data collection, processing, and sharing within the research community.

Combining geographical understanding with gene-associated data provides insights into local genetic diversity (Corpas et al., 2025; Peng et al., 2023), and facilitate understanding of disease emergence, tracking and timely interventions for diseases transmissions (see Figure 1) (Baker et al., 2023; Malinda, 2024; Prem et al., 2019; Spies et al., 2024; Su et al., 2024). Additionally, such studies are particularly significant for low-income countries, where levels of genetic literacy are limited and their associated health

impacts are not well understood (Thong et al., 2018). Effective collaboration from these countries is critical to advancing genetic data distribution and diversity, and can help to establish better infrastructure for healthcare services (Thong et al., 2018), clinical counselling and disease management (Malinda, 2022).

Several challenges are still prominent in utilizing geospatial and genomic data, including managing the large size of sequenced data, and applying suitable statistical approaches to understand genetic variations and geographical distances (Chen et al., 2024; Corpas et al., 2025). Expanding these theories to integrate geospatial analysis with genomic data goes beyond disease transmission, testing, clinical management (Malinda, 2022), and the study of genetic diversity (Peng et al., 2023). Such research studies can further accentuate their relevance in cancer and non-cancer rare diseases, spanning from their early assessment to overall survivorship (Malinda, 2016; Malinda et al., 2020; Nielsen et al., 2015). Therefore, the key connection between genetic data collection strategies and geospatial analysis not only helps to locate disease transmission patterns, but it further reveals where the genomic evidence is more clustered, unexplored, and systematically missing, especially in low-income nations.

Conclusion

In conclusion, geospatial genomic integrated analysis, especially in low-income Southeast Asian settings, is essential because it converts sequencing from a descriptive surveillance tool into an equity-sensitive analytical framework that can identify localized disease transmission pathways, and expose unequal data distribution. It thereby helps prepare concerned authorities to take preemptive actions to protect residents of specific locations from health outbreaks. Furthermore, advancements in genomic data sequencing within the context of geospatial become far more informative, and aid in the early spread of potential diseases transmission understanding, and enable healthcare professionals to devise more effective strategies for clinical management and improvement of overall healthcare facilities. Unexplored geographic information, combined with the context of genomic analysis, could provide future research directions for understanding data disparities. Overall, geospatial genomic analysis offers a valuable framework for strengthening health health surveillance, improving data equity, and supporting more resilient health systems of both communicable and non-communicable diseases, and promotes a more equitable health knowledge infrastructure.

Acknowledgment

The author sincerely thanks all the anonymous reviewers for their critical feedback, which helped improve the overall content of the manuscript. A preprint version of this article is available at Jxiv https://doi.org/10.51094/jxiv.779.

Conflicts of Interest

The author declares that there are no relevant personal or financial conflicts of interest regarding this study.

Funding

The author declares that no financial assistance in any form was received for this study.

References

Azzopardi, P. S., Kerr, J. A., Francis, K. L., et al. (2023). The unfinished agenda of communicable diseases among children and adolescents before the COVID-19 pandemic, 1990–2019: a systematic analysis of the Global Burden of Disease Study 2019. *The Lancet*, *402*(10398), 313-333. https://doi.org/10.1016/S0140-6736(23)00860-7

Baker, C. R., Barilar, I., de Araujo, L. S., Rimoin, A. W., Parker, D. M., Boyd, R....Shin, S. S. (2023). Use of High-Resolution Geospatial and Genomic Data to Characterize Recent Tuberculosis Transmission, Botswana. *Emerging Infectious Diseases, 29*(5), 977-987. https://doi.org/10.3201/eid2905.220796.

Belman, S., Saha, S., & Beale, M. A. (2022). SARS-CoV-2 genomics as a springboard for future disease mitigation in LMICs. *Nature Reviews Microbiology*, *20*, 3. https://doi.org/10.1038/s41579-021-00664-y

Brito, A. F., Semenova, E., Dudas, G., et al. (2022). Global disparities in SARS-CoV-2 genomic surveillance. *Nature Communications*, *13*, 7003. https://doi.org/10.1038/s41467-022-33713-y

Chen, Z., Lemey, P., & Yu, H. (2024). Approaches and challenges to inferring the geographical source of infectious disease outbreaks using genomic data. *The Lancet Microbe, 5*, e81-e92. https://doi.org/10.1016/S2666-5247(23)00296-3

Corpas, M., Pius, M., Poburennaya, M., et al. (2025). Bridging genomics' greatest challenge: The diversity gap. *Cell Genomics*, *5*(1), 100724. https://doi.org/10.1016/j.xgen.2024.100724

Fatumo, S., Chikowore, T., Choudhury, A., et al. (2022). Diversity in Genomic Studies: A Roadmap to Address the Imbalance. *Nature Medicine*, *28*, 243–250. https://doi.org/10.1038/s41591-021-01672-4

Gámbaro, F., Layan, M., Baele, G., et al. (2025). Navigating Sampling Bias in Discrete Phylogeographic Analysis: Assessing the Performance of an Adjusted Bayes Factor. *Molecular Biology and Evolution*, *42*(11), msaf253. https://doi.org/10.1093/molbev/msaf253

Getchell, M., Wulandari, S., De Alwis, R., et al. (2024). Pathogen genomic surveillance status among lower resource settings in Asia. *Nature Microbiology*, *9*(10), 2738–2747. https://doi.org/10.1038/s41564-024-01809-4

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics 17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Grad, Y. H., & Lipsitch, M. (2014). Epidemiologic data and pathogen genome sequences: A powerful synergy for public health. *Genome Biology 15*(11), 538. https://doi.org/10.1186/s13059-014-0538-4

Kattenberg, J. H., Mutsaers, M., Nguyen, V. H., et al. (2024). Genetic surveillance shows spread of ACT resistance during period of malaria decline in Vietnam (2018-2020). *Frontiers in Genetics*, *15,* 1478706. https://doi.org/10.3389/fgene.2024.1478706

Lemey, P., Hong, S. L., Hill, V., et al. (2020). Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nature Communications*, *11*, 5110. https://doi.org/10.1038/s41467-020-18877-9

Levin, A. T., Owusu-Boaitey, N., Pugh, S., et al. (2022). Assessing the burden of COVID-19 in developing countries: Systematic review, meta-Analysis and public policy implications. *BMJ Global Health*, 7(5), e008477. https://doi.org/10.1136/bmjgh-2022-008477

Lin, C. H., & Wen, T. H. (2022). How Spatial Epidemiology Helps Understand Infectious Human Disease Transmission. *Tropical Medicine and Infectious Disease*, 7(8), 164. https://doi.org/10.3390/tropicalmed7080164

Liu, P., Song, Y., Colijn, C., & MacPherson, A. (2022). The impact of sampling bias on viral phylogeographic reconstruction. *PLOS Global Public Health*, *2*(9), e0000577. https://doi.org/10.1371/journal.pgph.0000577

Malinda, R. R. (2016). Cellular Mechanisms Regulating Ciliary Disassembly and EMT: Roles of Ion Transport and Implications for Cancer [*Ph.D. Thesis*]. University of Copenhagen.

Malinda, R. R. (2022). Clinical management approaches in Bardet–Biedl syndrome. *European Journal of Human Genetics*, *30*, 1203–1204. https://doi.org/10.1038/s41431-022-01187-7

Malinda, R. R. (2024). Geospatial analysis of disease transmission and genetic diversity. *Jxiv preprint*. https://doi.org/10.51094/jxiv.779

Malinda, R. R., & Mishra, D. (2025). Genomic data, effective communication through visualization. *Acta Biologica Szegediensis*, *68*(2), 111–114. https://doi.org/10.14232/abs.2024.2.111-114

Malinda, R. R., Zeeberg, K., Sharku, P. C., et al. (2020). TGFβ Signaling Increases Net Acid Extrusion, Proliferation and Invasion in Panc-1 Pancreatic Cancer Cells: SMAD4 Dependence and Link to Merlin/NF2 Signaling. *Frontiers in Oncology*, *10,* 687. https://doi.org/10.3389/fonc.2020.00687

Nielsen, B. S., Malinda, R. R., Schmid, F. M., et al. (2015). PDGFRβ and oncogenic mutant PDGFRα D842V promote disassembly of primary cilia through a PLCγ- and AURKA-dependent mechanism. *Journal of Cell Science*, *128*(19), 3543–3549. https://doi.org/10.1242/jcs.173559

Peng, X., Li, Q., Cheng, Z., & Huang, X. (2023). The geography of genetic data: Current status and future perspectives. *Frontiers in Ecology and Evolution*, *11*, 1112636. https://doi.org/10.3389/fevo.2023.1112636

Prem, K., Pheng, S. H., Teo, A. K. J., et al. (2019). Spatial and temporal projections of the prevalence of active tuberculosis in Cambodia. *BMJ Global Health*, *4*(1), e001083. https://doi.org/10.1136/bmjgh-2018-001083

Richardson, D. B., Volkow, N. D., Kwan, M. P., et al. (2013). Spatial turn in health research. In *Science, 339*(6126), 1390-1392. https://doi.org/10.1126/science.1232257

Shaweno, D., Trauer, J. M., Doan, T. N., et al. (2021). Geospatial clustering and modelling provide policy guidance to distribute funding for active TB case finding in Ethiopia. *Epidemics*, *36,* 100470. https://doi.org/10.1016/j.epidem.2021.100470

Shrestha, S., Reja, M., Gomes, I., et al. (2021). Quantifying geographic heterogeneity in TB incidence and the potential impact of geographically targeted interventions in South and North City Corporations of Dhaka, Bangladesh: a model-based study. *Epidemiology and Infection*, *149,* e106. https://doi.org/10.1017/S0950268821000832

Spies, R., Hong, H. N., Trieu, P. P., Lan, L. K., Lan, K., Hue, N....Walker, T. M. (2024). Spatial Analysis of Drug-Susceptible and Multidrug-Resistant Cases of Tuberculosis, Ho Chi Minh City, Vietnam, 2020–2023. *Emerging Infectious Diseases*, *30*(3), 499-509. https://doi.org/10.3201/eid3003.231309.

Srivastava, A., Mahilkar, S., Upadhyaya, C. P., et al. (2024). Alkhumra Hemorrhagic Fever Virus (AHFV): A Concise Overview. *Yale Journal of Biology and Medicine*, *97*(4), 505-514. https://doi.org/10.59249/QSPC8835

Su, Y. C. F., Zeller, M. A., Ou, T. P., et al. (2024). Spatiotemporal evolution and transmission dynamics of Alpha and Delta SARS-CoV-2 variants contributing to sequential outbreaks in Cambodia during 2021. *Communications Medicine*, *4*(1), 252. https://doi.org/10.1038/s43856-024-00685-7

Thong, M. K., See-Toh, Y., Hassan, J., & Ali, J. (2018). Medical genetics in developing countries in the Asia-Pacific region: challenges and opportunities. *Genetics in Medicine*, *20*(10), 1114-1121. https://doi.org/10.1038/s41436-018-0135-0

Toczydlowski, R. H., Liggins, L., Gaither, M. R., et al. (2021). Poor data stewardship will hinder global genetic diversity surveillance. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(34), e2107934118. https://doi.org/10.1073/pnas.2107934118

Zeghbib, S., Kemenesi, G., & Jakab, F. (2023). The importance of equally accessible genomic surveillance in the age of pandemics. *Biologia Futura*, 74, 81–89. https://doi.org/10.1007/s42977-023-00164-5