Geospatial analysis of disease transmission and genetic diversity

Raj Rajeshwar Malinda[1,#]

[1]PGD in Data Science, Health and Climate Change for Social Impact, Indraprastha Institute of Information Technology Delhi, New Delhi 110020, India.

[#]Corresponding author: contact@rrmalinda.com

Abstract

Geospatial analysis provides insights into understanding tuberculosis disease transmission in Botswana, and highlights genetic diversity issues in available public database.

Keywords

Geospatial, genetics, genomics, disease transmission, public health

The term 'geospatial' refers to location-specific geographical information, and the data associated with these locations are considered geospatial data. Geospatial data are crucial to understand various phenomenon in fields, such as climate change in environmental sciences and health sciences. In the health domain, integrating geography-based information contributes to understanding disease etiology, and enhances the comprehension of disease prevention, diagnosis, transmission, and treatment (Richardson et al., 2013). Recent studies suggest, that leveraging geospatial data in disease settings can effectively advance health crisis responses (Shaweno et al., 2021; Shrestha et al., 2021).

Researchers utilized geospatial and genomic data to study disease transmission (Baker et al., 2023), and identified global imbalanced distribution patterns in genetic data (Peng et al., 2023). Genomic data, increasingly complex due to advancements in techniques, 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 significant burden on healthcare system, particularly in under-developing and developing economies (Azzopardi et al., 2023; Levin et al., 2022; Smith et al., 2022).

In a recent study, researchers focused on tuberculosis, one of the leading causes of mortality in lower-middle-income countries like Botswana, by integrating whole-genome sequencing and geospatial analysis to track potential disease transmission (Baker et al.,

2023). This approach identifies high burden transmission areas based on clustering method, and, therefore, suggests timely preventive measures during major outbreaks. Such studies propose combining genomic sequencing and geospatial data analysis to understand local and global disease transmission, aiding policymakers in implementing effective measures and interventions (Baker et al., 2023; Shaweno et al., 2021).

Geographical distribution patterns also reveal insights into genetic diversity (Peng et al., 2023). In a sense, it is argued that, biological diversity and distribution patterns can be well-understood through variations in available genetic data (Toczydlowski et al., 2021). 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 (Peng et al., 2023).

To address the current status and distribution of genetic data linked to geographical locations, researchers in this study, have highlighted taxonomical imbalance within genetic datasets. They further identified patterns of disproportion in data sharing by analyzing millions of nucleotide sequencing datasets across Animalia, Plantae, and Fungi kingdoms, and indicated unequal distribution patterns in shared genetic data (Peng et al., 2023). Such disparities highlight the need for comprehensive analyses to address local and global genetic data distribution issues.

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. Such patterns reflect gaps in historical dataset collection, further limiting our understanding of large-scale biodiversity and genetic variations. 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 (Peng et al., 2023), and facilitate timely interventions for outbreaks (Baker et al., 2023). Additionally, such studies hold particular significance for low-income countries, where genetic diseases are prevalent due to low literacy rates, and historical cultural perspectives (Thong et al., 2018). Effective collaboration from these countries is critical to advancing genetic studies, and can establish better infrastructure for genomic 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 sequence data, and applying suitable statistical approaches to understand genetic variations and geographic distances (Chen et al., 2024). Expanding these theories to integrate geospatial analysis with genomic data goes beyond disease transmission (Baker et al., 2023), testing, clinical management, and the study of genetic

diversity (Peng et al., 2023), such research studies can further accentuate its relevance in cancer and non-cancer rare diseases, spanning from their early assessment to survivorship (Best et al., 2022; Schootman et al., 2017).

In conclusion, harnessing geospatial analysis in genetic and genomic data identifies localized disease transmission, thereby preparing concerned authorities to take preemptive actions to protect residents of specific locations from health outbreaks (Anderson et al., 2023). Furthermore, advancements in the field of gene sequencing within the context of geospatial data can further enhance understanding of genomic datasets, aiding in early prediction of potential genetic diseases, and enables healthcare professionals to devise more effective strategies for clinical diagnosis, management and improving overall survival rates among individuals. This integration informs policymakers about effective measures to address unforeseen health crisis related to both communicable and non-communicable diseases.

Author contributions

RRM has conceptualized the idea, wrote, and edited the article.

Funding

No financial assistance was received for this study.

Ethical approval

Not applicable for this study.

Competing interests

The author declares no relevant competing interests.

References

Anderson, B., Ouyang, D., D'Agostino, A., Bonin, B., Smith, E., Kraushaar, V., Rudman, S. L., & Ho, D. E. (2023). Leveraging genomic sequencing data to evaluate disease surveillance strategies. *IScience*, *26*(12). https://doi.org/10.1016/j.isci.2023.108488

Azzopardi, P. S., Kerr, J. A., Francis, K. L., Sawyer, S. M., Kennedy, E. C., Steer, A. C., Graham, S. M., Viner, R. M., Ward, J. L., Hennegan, J., Pham, M. D., Habito, C. M. D., Kurji, J., Cini, K. I., Beeson, J. G., Brown, A., Murray, C. J. L., Abbasi-Kangevari, M., Abolhassani, H., … Patton, G. C. (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). 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., Tobias, J. L., Moonan, P. K., Click, E. S., Finlay, A., Oeltmann, J. E., Minin, V. N., Modongo, C., Zetola, N. M., Niemann, S., & Shin, S. S. (2023). Use of High-Resolution Geospatial and Genomic Data to Characterize Recent Tuberculosis Transmission, Botswana. *Emerging Infectious Diseases*, *29*(5). https://doi.org/10.3201/eid2905.220796

Best, S., Vidic, N., An, K., Collins, F., & White, S. M. (2022). A systematic review of geographical inequities for accessing clinical genomic and genetic services for non-cancer related rare disease. In *European Journal of Human Genetics* (Vol. 30, Issue 6). https://doi.org/10.1038/s41431-021-01022-5

Chen, Z., Lemey, P., & Yu, H. (2024). Approaches and challenges to inferring the geographical source of infectious disease outbreaks using genomic data. In *The Lancet Microbe* (Vol. 5, Issue 1). https://doi.org/10.1016/S2666-5247(23)00296-3

Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A. R., & Kuchenbaecker, K. (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

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

Levin, A. T., Owusu-Boaitey, N., Pugh, S., Fosdick, B. K., Zwi, A. B., Malani, A., Soman, S., Besançon, L., Kashnitsky, I., Ganesh, S., McLaughlin, A., Song, G., Uhm, R., Herrera-Esposito, D., De Los Campos, G., Peçanha Antonio, A. C. P., Tadese, E. B., & Meyerowitz-Katz, G. (2022). Assessing the burden of COVID-19 in developing countries: Systematic review, meta-Analysis and public policy implications. In *BMJ Global Health* (Vol. 7, Issue 5). https://doi.org/10.1136/bmjgh-2022-008477

Malinda, R. R. (2022). Clinical management approaches in Bardet–Biedl syndrome. In *European Journal of Human Genetics* (Vol. 30, Issue 11). https://doi.org/10.1038/s41431-022-01187-7

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*. https://doi.org/10.3389/fevo.2023.1112636

Richardson, D. B., Volkow, N. D., Kwan, M. P., Kaplan, R. M., Goodchild, M. F., & Croyle, R. T. (2013). Spatial turn in health research. In *Science* (Vol. 339, Issue 6126). https://doi.org/10.1126/science.1232257

Schootman, M., Gomez, S. L., Henry, K. A., Paskett, E. D., Ellison, G. L., Oh, A., Taplin, S. H., Tatalovich, Z., & Berrigan, D. A. (2017). Geospatial approaches to cancer control and population sciences. *Cancer Epidemiology Biomarkers and Prevention*, *26*(4). https://doi.org/10.1158/1055-9965.EPI-17-0104

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

Shrestha, S., Reja, M., Gomes, I., Baik, Y., Pennington, J., Islam, S., Faisel, A. J., Cordon, O., Roy, T., Suarez, P. G., Hussain, H., & Dowdy, D. W. (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*. https://doi.org/10.1017/S0950268821000832

Smith, J. P., Oeltmann, J. E., Hill, A. N., Tobias, J. L., Boyd, R., Click, E. S., Finlay, A., Mondongo, C., Zetola, N. M., & Moonan, P. K. (2022). Characterizing tuberculosis

transmission dynamics in high-burden urban and rural settings. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-10488-2

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

Toczydlowski, R. H., Liggins, L., Gaither, M. R., Anderson, T. J., Barton, R. L., Berg, J. T., Beskid, S. G., Davis, B., Delgado, A., Farrell, E., Ghoojaei, M., Himmelsbach, N., Holmes, A. E., Queeno, S. R., Trinh, T., Weyand, C. A., Bradburd, G. S., Riginos, C., Toonen, R. J., & Crandall, E. D. (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). https://doi.org/10.1073/pnas.2107934118