Pak. J. Biochem. Mol. Biol., 2025, 58(1):1-3

Perspective

Genomics of the Pakistani Population:

Perspectives on Long-Read Next-Generation Sequencing

M. Kamran Azim (mkazim@uok.edu.pk)

Department of Biochemistry, University of Karachi, Karachi, Pakistan.

Abstract

Pakistan represents one of the most genetically diverse regions in South Asia, yet remains underrepresented in global genomic databases. The advent of long-read next-generation sequencing (NGS) technologies offers unprecedented opportunities to explore this diversity, resolve complex genomic structures, and address pressing health and agricultural challenges. This perspective highlights the potential of long-read sequencing in advancing Pakistani genomics, with implications for medicine, population genetics, and sustainable development.

Introduction

Pakistan's population of over 240 million is characterized by extensive ethnic diversity, shaped by centuries of migration and admixture across South Asia, Central Asia, and the Middle East (1,2). Despite this diversity, genomic studies of Pakistani populations remain limited compared to other regions. Long-read sequencing technologies, such as Pacific Biosciences (PacBio) HiFi and Oxford Nanopore platforms, provide new avenues to capture structural variants, repetitive regions, and haplotype phasing with greater accuracy than short-read sequencing.

Genetic Diversity and Clinical Relevance

Whole-genome sequencing of Pakistani individuals has revealed millions of variants, many of which are rare or absent in global reference datasets (1). Sequencing of an ethnic Pathan genome demonstrated unique variants linked to migration across Asia, underscoring the population's evolutionary significance (3).

Clinically, Pakistan faces a high burden of recessive genetic disorders due to elevated rates of consanguinity (4). Long-read sequencing can improve variant detection in complex genomic regions, enabling more accurate diagnosis of inherited diseases. Moreover, pharmacogenomic insights are critical for optimizing treatments for tuberculosis, cardiovascular disease, and diabetes, which are prevalent in the region.

Advantages of Long-Read Sequencing

Long-read sequencing offers several advantages for Pakistani genomics:

Structural variant detection: Captures large insertions, deletions, and translocations often missed by short reads.

Improved genome assembly: Resolves repetitive and GC-rich regions.

Haplotype phasing: Enables direct phasing of maternal and paternal alleles in diverse populations.

Epigenetic insights: Nanopore sequencing can detect DNA methylation, revealing population-specific regulatory mechanisms.

Applications in Health and Medicine

Rare genetic disorders: Identification of pathogenic variants in families affected by recessive conditions.

Cancer genomics: Detection of gene fusions and structural variants for precision oncology.

Pharmacogenomics: Understanding drug metabolism variations to guide personalized medicine.

Public health genomics: Informing national screening and preventive strategies.

Population Genetics and Evolutionary Insights

Long-read sequencing can reconstruct population histories by identifying ancient haplotypes, tracing admixture events, and exploring adaptations to diverse environments such as high-altitude regions of Gilgit-Baltistan. These insights enrich global understanding of human diversity while situating Pakistan within broader evolutionary narratives (3).

Agricultural and Environmental Genomics

Pakistan's reliance on crops such as wheat, rice, and cotton highlights the importance of agricultural genomics. Long-read sequencing has already been applied to pathogens affecting basmati rice, revealing structural features of effector regions in *Xanthomonas oryzae* isolates (5). Expanding such efforts can improve crop resilience to drought, salinity, and disease, supporting food security.

Challenges and Opportunities

Cost and infrastructure: Long-read sequencing remains expensive; national centers are needed.

Bioinformatics capacity: Training programs and collaborations are essential for handling large datasets.

Ethical considerations: Policies on data sharing, privacy, and informed consent must be prioritized.

Global representation: Inclusion of Pakistani genomes in international databases will reduce biases in medical research.

Future Directions

1. Establish a "National Genomics Initiative" leveraging long-read sequencing.

2. Integrate genomic data into "clinical practice" for personalized medicine.

3. Build "bioinformatics capacity" through training and international collaboration.

4. Engage the public to foster trust and participation in genomic research.

5. Expand research to "agricultural and environmental genomics" for sustainable development.

Conclusion

The genomics of the Pakistani population represents a frontier of discovery. Long-read sequencing technologies provide the tools to unlock this diversity, offering insights into health, evolution, and agriculture. By investing in infrastructure, training, and ethical frameworks, Pakistan can position itself as a leader in global genomics research.

References

1- Khan, S. Y., Ali, M., Lee, M. W., Ma, Z., Biswas, P., Khan, A. A., Naeem, M. A., Riazuddin, S., Riazuddin, S. A., & Hejtmancik, J. F. (2020). Whole genome sequencing data of multiple individuals of Pakistani descent. Scientific Data, 7(1), 1–8.

2- Azim, M. K., Yang, C., Yan, Z., Choudhary, M.I., Khan, A.I., Sun, X., Li, R., Asif, H., Sharif, S., & Zhang, Y. (2013). Complete genome sequencing and variant analysis of a Pakistani individual. Journal of Human Genetics, 2013, 58(9), 622-626.

3- Mehmood, R., et al. (2015). Whole genome sequencing of an ethnic Pathan (Pakhtun) from the north-west province of Pakistan. BMC Genomics, 16, 172.
4- Ghias, K., Rehmani, S.S., Razzak, S.A., Madhani, S., Azim, M.K., Ahmed, R., & Khan, M.J. Mutational landscape of head and neck squamous cell carcinomas in a South Asian population. (2019). Genetics and Molecular Biology, 4242(3):526-542.
5- Ejaz, K., Zakria, M., Zhang, P., Tapia, J. H., Arif, M., White, F., & Yasmin, S. (2025). Comparative genomics using long-read sequencing identifies nearly identical TAL effector regions in Xanthomonas oryzae isolates from Pakistan. Frontiers in Microbiology, 16, 1560969.