REVIEW ARTICLE

Genetic polymorphism: evolution with technological advances and future direction

Randrita Pal, Ankita Samaddar, Sanchari Chakraborty, Barnali R. Basu*

Abstract

Genetic polymorphisms are considered as one of the major contributing factors for the variability in disease development and pathogenesis as well as drug response in individuals. Fortunately, in the last few decades a range of technological advancements eased the way for polymorphic studies to reveal the association between genetic polymorphism like single nucleotide polymorphisms (SNPs) or Variable number tandem repeats (VNTRs) and human diseases. Starting from Mendelian inheritance to recent Next Generation Sequencing (NGS) technologies not only helped to understand human disease biology better, paved the way towards personalised therapy by studying individual drug/therapy responses based on genetic makeup (mutation/variant) of individuals. Literature mining from PubMed, Google Scholar and Medline databases using keywords likes 'polymorphism', 'genetic polymorphism', 'SNP', 'VNTR', 'CNV'. The massive parallel sequencing capability of the NGS facilitates clinicians towards therapeutic decisions and aids follow up of patients by identifying minimal residual disease. However, this is the beginning of the era of targeted and personalised therapy and scientific world which will be able to touch the tip of the iceberg. However, much focus is needed to develop more user-friendly and cost-effective technologies to reach more patients along with development of much simpler and robust statistical methodologies to handle or interpret big data.

Keywords: Diseases, evolution, genetics, polymorphism, technological advancement.

Indian Journal of Physiology and Allied Sciences (2022); **ISSN:** 0367-8350 (Print)

INTRODUCTION

Variations in DNA is a very common phenomenon but when multiple forms of a gene are expressed in at least 1% of a population it is termed as genetic polymorphism. Genetic polymorphism alters among species and between genomes which justifies its impact on evolution as well as conservation of species. Evolution induces genetic variation and genetic adaptiveness towards *cis* or *trans* as a response to environment.[1] Variation in effective population size governs genetic diversity and polymorphism plays a vital role in this.[2] It is known to have several effects on the plant and animal world, especially disease prognosis in humans. Not only diseases and its early predictions, it is also known to affect animal behaviour.[3] The advancement of genotyping and sequencing technologies along with sophisticated statistical tools have resulted in easy detection of sequence variations and its relation with development of human diseases by high-density mapping of single nucleotide polymorphisms (SNPs) and haplotypes.[4] The evolution of genetic polymorphism and relation between technological advancement and genetic studies is the main focus of this review particularly in human diseases.

The human genome consists of 23 pairs of chromosomes, each pair inherited from respective parents, contains about 6 billion nucleotides of DNA.[5] Due to this vast size, the probability of having polymorphic DNA is very high. This polymorphism within the human genome varies from single nucleotide change to several base pair changes and presence of repetitive sequences.[6-7]Among all different types of polymorphisms, single nucleotide polymorphism (SNPs) is the most abundant.[8] SNPs are well scattered throughout the genome with variation in numbers mostly present in non-coding, regulatory regions of the genome and serve as excellent genetic markers to map several human diseases, evolutionary studies and population genetics.[9] Depending on the location within the genome, SNPs act differently at phenotypic levels.[8] Another major progress in the field of genomics occurs with the discovery of highly polymorphic repetitive DNA that comprises about 30-90% of the whole genome.[10] The length of variable number tandem repeats (VNTRs) that remains adjacent to each other varies from 2 to many thousand copies and on the basis of this they are classified into macro, mini and microsatellites.[8] According to the previous studies, about 700000-1000000 microsatellite loci is about 2-6 base pairs long whereas di and tetra nucleotide constitutes 75% of the human reference genome. Although VNTRs can be present widely at un-translated

Clinicogenomics Lab, Department of Physiology, Surendranath College, University of Calcutta, Kolkata, India.

***Corresponding author:** Barnali R. Basu, Clinicogenomics Lab, Department of Physiology, Surendranath College, University of Calcutta, Kolkata, India., Email: braybasu@gmail.com

How to cite this article: Pal R, Samaddar A, Chakraborty S, Basu BR. Genetic Polymorphism: Evolution with Technological Advances and Future Direction. Indian Journal of Physiology and Allied Sciences. 2022;74(4):12-15.

Conflict of interest: None

Submitted: 05/03/2022 **Accepted:** 15/11/2022 **Published:** 05/12/2022

regions, introns and protein coding regions, the dinucleotide repeats specifically abundant in regulatory regions than any other genomic regions.[11] These polymorphic tandem repeats were initially utilised in DNA fingerprinting but their association with wide range of pathological conditions are now well established.[12] Another important polymorphic DNA sequence involves structural and copy number variations (CNVs), discovery of which overruled the concept of having strictly two copies on genes within the genome. Several studies have illustrated that a substantial portion of DNA which ranges from thousands to million base pair of DNA can vary in copy numbers and they become one of the most important sources of genomic diversity.[13]

With the advancement of time, it becomes evident that polymorphisms are well associated with several diseases or may regulate a number of disease outcomes. SNPs, being changes in one nucleotide, can cause several genetic conditions that involve single or multiple genes.[14] SNPs present in the protein coding region of the genome can be either synonymous or non-synonymous resulting in either silent mutation or missense/nonsense mutation. One such disease that occurs due to SNPs located in coding sequence is Sickle cell anaemia. Disease causing SNPs can also present within the regulatory or promoter region of a gene that largely influence the expression and activity of that particular gene. Another different class of SNPs are responsible for alteration of the primary structure of proteins associated with drug metabolism which become potent target for the pharma-cogenetics studies. All the SNPs are not directly harmful or causative but they, being in close association with a disease-causing sequence, segregate with that particular gene. Therefore, the existence of that SNP can be a potent marker of the presence or the increased risk of development of that disease. These SNPs are widely studied for disease diagnosis, prediction and other applications.[8] Along with SNPs, STRs also correlates with a number of diseases and phenotypic changes depending on the increase or decrease in repetitive sequence length. One such example is Huntington's disease in which the causative mutation is due to the increase in number of CAG repeats from normal range of about 11-14 copies to an abnormal range of at least 38 copies that produces the excess glutamine.[15] Not only that there is other 40 neurological diseases including spinocerebellar ataxia with polyglutamine tracts also have significant correlation with the changes in length of trinucleotide repeats.[16]

With the recent advancement of technologies and introduction of the human genome project and next generation sequencing it is now a virgin area of genetic association studies that had revolutionised the field of scientific research.

In the present study we have conducted a detailed analysis of the genetic polymorphism studies specifically focused on humans by integrating the recent technological advancement over the past decades (i) to find out the technological advancement in genetic studies since inception, and (ii) to highlight time-framed polymorphism and genetic polymorphism studies and its implications on the human system.

MATERIALS AND METHODS

For this review, PubMed, Google Scholar and Medline data bases were extensively searched using keywords like 'polymorphism', 'genetic polymorphism', 'SNP', 'VNTR', 'CNV'.

RESULTS

Genetic studies are not possible without the knowledge of DNA and the various advancements in technology. Here we had categorised our search into two; one is the advancement of technology and simultaneous hallmark discoveries in genetic studies; second is the polymorphism study which although being multiple (expressive or non-expressive) forms of a particular gene, might have other biological implications.

Genetic Discoveries and Technological Advancement

Gregory Mendel had introduced fundamentals of inheritance which led to discovery of DNA and Chromosomes. DNA was then discovered as the hereditary carrier and later its double helical structure was identified which fetched Watson and Crick a Nobel prize.[17] In 1977, Sanger sequencing technique (dideoxy method) was discovered which was the stepping stone for all genetic studies till the current time which eluted complete genome sequences for the first time.[18] Post the completion of the Human Genome Project in 2003, further development of sequencing techniques like Next generation sequencing (NGS), the second-generation sequencing was comparatively faster and cheaper but they were short reads.[19] The third generation or long-read sequencing methods can produce novel genome assemblies.[20] SMRT has helped to resolve high-GC content regions and also highly repetitive regions, commonly known as short term Tandem Repeats (STR) which are known to be disease initiators.[21] All of this has resulted in a huge amount of generated data which calls for big data management and thereby the use of data management and developed sophisticated statistical tools has been ever increasing. Figure 1 gives an insight into detailed technological advances along with hallmarks of genetic studies.

Polymorphism Studies and their Implications Over Time

Trailing the path of several scientific discoveries related to polymorphism revealed that the prior polymorphic investigations solely depended on phenotypic visible coexistence of several characteristics within the organism. The most ancient study on polymorphism done on butterflies termed as mimetic polymorphism to illustrate the visible colour differences found among butterflies.[22] Immediately afterwards there are polymorphism studies on colour profile of drosophila and origin and evolution of polymorphism

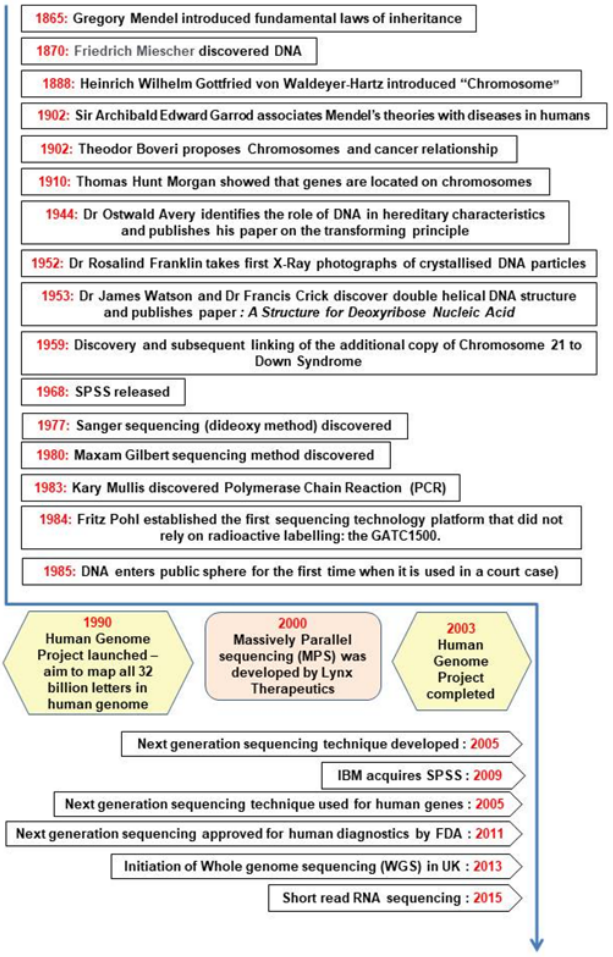

Figure 1: Time-lapse representation of hallmarks of Genetic discoveries and Technological advancements contributing to evolution of genetic studies.

on ants.[23,24] Another striking coexistence of different characteristics in animals are blood groups and with the help of the technological advancements' genetic polymorphism study on ABO blood groups has been done.[25] After interpreting these visible components of polymorphism, the area of the research moved further into the cellular levels analysing the polymorphic genes in complement system, Gc globulin and collagen.[26-28] With the advancement of technology, the loci of research move further from cellular towards genetic/DNA levels and therefore different varieties of polymorphisms have surfaced that are responsible for different sets of characteristics or diseases among human population. Single nucleotide polymorphism study was earliest in this genre and subsequently all other varieties such as Variable number tandem repeats, short tandem repeats and copy number variations polymorphism has also enter the arena of research.[29-31] Polymorphism detection techniques and invention of Sanger DNA sequencing has further correlated different diseases with polymorphisms. In very recent times, Human genome project and advanced sequencing techniques enables the researchers to analyse

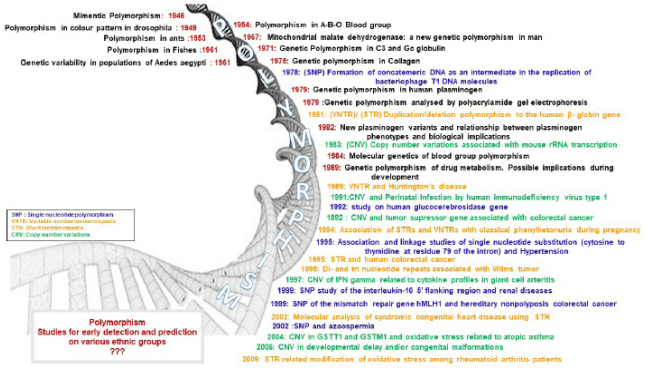

Figure 2: Polymorphism studies from initial to recent times associated with different types of polymorphism and related diseases.

the relationship between different diseases and presence of a specific polymorphism which further facilitates disease prediction at very early stage. Intricate DNA sequencing and analysis help the researchers to even design individualised drugs depending on their polymorphism profile to minimise the disease outcome.[32-33] Figure 2 ropes in the studies of polymorphism and its various forms and how 90s decade saw a sudden change of general polymorphism study being metamorphosed into disease based polymorphism studies and also its various forms came into limelight.

DISCUSSION

Genetic studies enrichment is impossible without various technological advances which mainly targets at achieving maximum yield of data with minimum resources, be it DNA sample or/and low cost of reagents and machines. Machines have made detection easier but data analysis has to be ultimately done by humans not by machines. Nowadays much of genetic studies are targeted with in-silico approaches before laboratory work but machines are only tools so the data for machine learning comes from experiments conducted in laboratories.

Our study highlighted various technological advances and how genetic polymorphism studies have evolved and developed over the years. One significant observation is the vividness and minute up gradation of DNA studies with each evolving technology.

REFERENCES

1. Signor SA, Nuzhdin SV. The evolution of gene expression in cis and trans. Trends in Genetics. 2018;34(7):532-44. DOI: 10.1016/j.tig.2018.03.007
2. Ellegren H, Galtier N. Determinants of genetic diversity. Nature Reviews Genetics. 2016;17(7):422-33. DOI: 10.1038/nrg.201658.
3. Inou-Murayama M. Genetic polymorphism as a background of animal behavior. Animal Science Journal. 2009;80(2):113- 20. DOI: 10.1111/j.1740-0929.2008.00623.x
4. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. Ann Rev Med. 2005;56:303. DOI: 10.1146/annurev.med.56.082103.104540.
5. Brieba LG. Template dependent human DNA polymerases. Current Topics in Medicinal Chemistry. 2008;8(15):1312- 26.

Available from: www.ingentaconnect.com/content/ben/ctmc/2008/00000008/00000015/art00002
6. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Nat Acad Sci. 1989;86(8):2766-70. DOI: 10.1073/pnas.86.8.2766
7. Teama S. DNA polymorphisms: DNA-based molecular markers and their application in medicine. Genetic Diversity and Disease Susceptibility. 2018;25. DOI: 10.5772/intechopen.79517.
8. Ismail S, Essawi M. Genetic polymorphism studies in humans. Middle East J Med Genet. 2012;1(2):57-63. DOI: 10.1097/01.MXE.0000415225.85003.47
9. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348-52. DOI: 10.1038/nature10242
10. Rao SR, Trivedi, S, Emmanuel, D, Merita K, Hynniewta M. DNA repetitive sequences-types, distribution and function: a review. J Cell Mol Biol. 7 and 8(2), 1-11. Available from: http://bspdf.freeweb.pk/od/bba/cell%20 and%20molecular%20biology.pdf#page=7
11. Willems T, Gymrek M, Highnam G, Mittelman D, Erlich Y, 1000 Genomes Project Consortium. The landscape of human STR variation. Genome Res. 2014;24(11):1894-904. DOI: 10.1101/gr.177774.114.
12. Bruce HA, Sachs NA, Rudnicki DD, et al. Long tandem repeats as a form of genomic copy number variation: structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations. Psychiat Genet. 2009;19(2):64. DOI: 10.1097/ YPG.0b013e3283207ff6
13. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54. DOI: 10.1038/ nature05329
14. Sobrino B, Brión M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int. 2005;154(2-3):181-94. DOI: 10.1016/j.forsciint.2004.10.020
15. La Spada AR. Glutamine repeats and neurodegenerative diseases: molecular aspects. Am J Hum Genetics. 2002;70(3):820. PMC384963.
16. Pearson JC, Lemons D, McGinnis W. Modulating Hox gene functions during animal body patterning. Nature Rev Genetics. 2005;6(12):893-904. DOI: 10.1038/nrg1726.
17. Watson JD, Crick FHC. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171:737-8. DOI: 10.1038/171737a0
18. Sanger F. DNA sequencing with chain termination inhibitors. Proc Nat Acad Sci. 1977;74:5463-7. DOI: 10.1073/pnas.74.12.546
19. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science. 2003;11;300(5617):286-90. DOI: 10.1126/science.1084564
20. Van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The third revolution in sequencing technology. Trends Genet. 2018;34(9):666-81. DOI: 10.1016/j. tig.2018.05.008
21. Schmidt MH, Pearson CE. Disease-associated repeat instability and mismatch repair. DNA repair. 2016;38:117-26. DOI: 10.1016/j.dnarep.2015.11.008
22. Hale Carpenter GD. Mimetic polymorphism. Nature. 1946;158(4008):277-9. DOI: 10.1038/158277a0.
23. Da Cunha AB. Genetic analysis of the polymorphism of color pattern in Drosophila polymorpha. Evolution. 1949:239- 51. DOI: 10.1111/j.1558-5646.1949.tb00024.x.
24. Wilson EO. The origin and evolution of polymorphism in ants. Q Rev Biol. 1953:136-56. DOI: 10.1086/399512
25. Brues AM. Selection and polymorphism in the A-B-O blood groups. Am J Phys Anthropol. 1954;12(4):559-98. DOI: 10.1002/ajpa.1330120408
26. Alper CA, Robin NI, Refetoff S. Genetic Polymorphism in Rhesus C3 and Ge Globulin. J Immunol. 1971;107(1):96-8. DOI: 10.4049/jimmunol.107.1.96
27. Bailey AJ, Sims TJ, Le Lous M, Bazin S. Collagen polymorphism in experimental granulation tissue. Biochem Biophys Res Commun. 1975;66(4):1160-5. DOI: 10.1016/0006-291X(75)90480-5
28. Bazin S, Bailey A, Le Lous M, Nicolétis C, Delaunay A. Collagen polymorphism in human scars. C R Acad Hebd Seances Acad Sci D. 1975;281(19):1447-9. PMID: 815035.
29. Spritz RA. Duplication/deletion polymorphism 5'-to the human β globin gene. Nucleic Acids Res. 1981;9(19):5037-48. DOI: 10.1093/nar/9.19.5037
30. Ritchie DA, Joicey DH. Formation of concatemeric DNA as an intermediate in the replication of bacteriophage T1 DNA molecules. J Gen Virol. 1978;41(3):609-22. DOI: 10.1099/0022-1317-41-3-609
31. Kuehn M, Arnheim N. Nucleotide sequence of the genetically labile repeated elements 5' to the origin of mouse rRNA transcription. Nucleic Acids Res. 1983;11(1):211-24. DOI: 10.1093/nar/11.1.211
32. Elsayed AH, Cao X, Mitra AK, et al. Polygenic Ara-C response score identifies pediatric patients with acute myeloid leukemia in need of chemotherapy augmentation. J Clin Oncol. 2022;40(7):772-83. DOI: 10.1200/JCO.21.01422
33. Ozberk D, Haywood A, Sutherland HG, et al. Association of polymorphisms in ARRB2 and clinical response to methadone for pain in advanced cancer. Pharmacogenomics. 2022;23(5):281-9. DOI: 10.2217/pgs-2021-0139