Molecular epidemiology

Vineeta Singh

National Institute of Malaria Research (ICMR), Sector-8, Dwarka, New Delhi-110077, India

Molecular epidemiology is majorly defined as the combined effect of genetic and environmental risk factors at the molecular level, to the etiology, geographic distribution and prevention of disease within and across populations. The expression- "molecular epidemiology" is loosely used for the articles that use molecular profiling/fingerprinting techniques—regardless of whether any epidemiologic relevance exists in the article. However, molecular epidemiology is meant to include both the terms; *viz* "molecular," for the application of molecular biology techniques, and the "epidemiology," for studying the distribution and determinants of disease occurrence in human populations. The growth of this discipline has gained momentum by the decoding of the human genome and major advances in molecular biology. The knowledge about molecular epidemiology not only facilitates the ability of scientists to conduct etiologic research but also increases their insight about the determinants of disease alongwith contributing to disease prevention measures to improve public health.

Characterizing the pathogen leads to evaluating the impact of pathogens genetic diversity on their relevant medical properties. Major advances in molecular epidemiological methodologies have helped in eluciadting the role of host and pathogen genetic factors with respect to the infectious diseases. Also the exponential development of molecular technologies like genome sequencing, postgenomic studies, and bioinformatics, continue to spark enthusiasm about future prospects. This chapter also attempts to provide a comprehensive overview of the various aspects involved which collectively makeup the discipline of molecular epidemiology.

Keywords: molecular epidemiology; molecular techniques; infectious diseases

1. Introduction

The discipline of molecular epidemiology has evolved by merging the boundaries of traditional epidemiology, human and molecular genetics and molecular biology. The genetic epidemiology, falling within the molecular epidemiology, focuses on the use of genetic factors specifically to elucidate epidemiologic relationships. The advantage of molecular epidemiology lies in the collective knowhow of the rapidly evolving cellular and molecular biology and genetic research, along with classic and advanced epidemiologic methods for better understanding of the disease, the characterization of risk factors for disease and the development of more effective and targeted medical treatments [1,2] (Figure 1). This discipline is multifaceted which encompasses the environmental and genetic risk factors in the etiology of chronic diseases to understand the underlying disease mechanisms thus, establishing a better rationale for disease prevention, by developing biomarkers with diagnostic or prognostic utility in the clinical setting [3]. It is the involvement of molecular biomarkers in to population-based studies for understanding the molecular basis of diseases to facilitate the scientists in their disease etiological research, increases the existing knowledge about disease determinants, hence contributing to the development of disease prevention approaches resulting in public health improvement [4,5].

Fig. 1 The integrated four disciplines which make up molecular epidemiology.

2. Infectious diseases

Infectious diseases account for most part of the global health problem, both in terms of morbidity and mortality with major burden falling in developing countries [6]. The molecular epidemiological studies of infectious diseases include evaluation of the factors leading to the cause of disease, host genetics and environment all put together. The knowledge of all these characteristics is important for control, treatment and prevention of the diseases as they kill more people worldwide caused by germs found everywhere - in air, soil and water than any other single disease. The host can get infected by germs *via* several modes of transmission eg: physical/sexual contact, contaminated air/water/food or

through infected animal and/or insect bites [7]. The spread of infectious diseases can be contained by practicing hygienic healthy habits, with proper medication during sickness and preventive measures like vaccines, etc. [8].

2.1 Classification of Microorganisms

a) Bacteria - one-celled germs (Gram-positive/gram-negative, antibiotic sensitivity differs, diagnostic and therapeutic uses of gram-negative capsule) that multiply quickly releasing chemicals causing illness like strep throat, urinary tract infections, tuberculosis etc. [9].
b) Viruses – have capsules that contain genetic material (RNA/DNA, RNA viruses genetically unstable, lipid enveloped/nonenveloped, solvent-detergent treatments virucidal only for enveloped viruses) are smaller than bacteria and cause diseases ranging from common flu to AIDS [10].
c) Fungi - primitive plants (Disseminated vs. Superficial, mold vs. biphasic), like mushrooms or mildew and infect skin, lung or nervous system [11].
d) Protozoa - one-celled animals that live on other living organisms for food and survival [10,11]. The common infectious diseases are listed in the table 1.

Table 1 List of Common Infectious Diseases.

S. No	**Disease**	**Mode of infection**	**Microbial pathogens**			
			Virus	**Bacteria**	**Fungi**	**Protozoa**
1.	Meningitis	Spread through the exchange of respiratory and throat secretions . spread of a fungus through blood to the spinal cord.	*Non-polio enteroviruses*	*H.influenzae*, *S.pneumoniae* , group B *Streptococcus*, *Listeria monocytogenes*, and *Neisseria meningitidis*.	*Cryptococcus*	*N.fowleri.*
2.	Salmonella infections	Contaminted water and food		*Salmonella* spp.		
3.	Mumps	Spreads by saliva or mucus ejected out by mouth or nose	*Mumps virus*			
4.	Sexually transmitted diseases	Sexual contact, contaminated body fluids	*Human immunodeficiency virus, Human papilloma virus*	*C. trachomatis, N.gonorrhoeae, T.pallidum*		*T.vaginalis*
5.	Poliomyelitis	It spreads through fecal-oral and the oropharyngeal routes	*Polio virus*			
6.	Severe acute respiratory syndrome (SARS)	Air borne, contaminated objects	*SARS corona virus*			
7.	Whooping cough	Air-borne		*B.pertussis*		
8.	Yellow fever	Spread by bite of female *Aedes aegypti*	*Yellow fever virus*			
9.	Typhoid	Spread by contaminated food or water and occasionally through direct contact.		*S. typhi*		
10.	Tuberculosis	Air-borne		*M.tuberculosis*		
11.	Strep throat	Air-borne		*Streptococcus group A beta-haemolytic*		
12.	Shigellosis	Contaminated water, oral-faecal route, directly person-to-person hand-to-mouth.		*Shigella family*		

13.	Schisotsomiasis	Contaminated fresh water				*S.mansoni, S. haematobium*, or *S.japonicum*
14.	Rota virus	Faeco-oral route	*Rota virus A,B,C,D,E,F,G,H*			
15.	Pneumonia	Air-borne	*Rhino corona, influenza, respiratory syncytial virus (RSV),aden o and para influenza virus*	*S.pneumonia, M. pneumonia, H.influenzae, C.pneumonia*		
16.	Onchocerciasis (river blindness)	Spread by bite of black fly *Simulium* type				*O. volvulus*
17.	Measles	Air-borne	*Measles virus*			
18.	Malaria	Bite of infected female *Anophele s*				*P.falciparum, P.vivax, P.malariae, P.ovale*
19.	Influenza	Direct transmission; the airborne route or through contact.	*Influenza virus*			
20.	Chicken-pox	*Via* sneeze/cough, or by contact with the clothing, bed linens, or oozing blisters of an infected person	*Varicella zoster virus*			
21.	Hepatitis A, B, C	Faeco-oral route or by ingestion of contaminated water or food	*Hepatitis A , B, C virus*			
22.	Dengue	Through the bite of *Aedes aegypti* mosquito.	*Dengue virus*			
23.	Cryptosporiodos is	Spread by contaminated water				*Cryptosporidiu m*
24.	Cholera	Contaminated drinking water and unsanitary conditions		*V. cholera*		

3. Host genetics

The host genetic factor, the genomic diversity of host and pathogen are the major determinants in deciding disease susceptibility. The exisiting genomic diversity in host and pathogen both; are influenced by various factors such as evolutionary pressures, migration, culture, regional and various other man made variables [12,13]. Humans are exposed to variety of microorganisms and infectious agents in their day to day life but, not all infected people develop diseases. This variability in the people in contracting a disease could be attributed to the difference in the host genes responsible for their immune response [14].

To identify the genes involved in host susceptibility to infectious diseases several studies on Major Histocompatilibity Complex (MHC) and non-MHC genes have been undertaken [15]. Several studies on racial differences in susceptibility, family segregation analyses, association studies, candidate gene studies and genome scan studies have observed that host genetics play a crucial role in determining the susceptibility or resistance to many infectious diseases [16-17]. It is also reported that genetic factors play a major role in susceptibility to infectious diseases [18]. From the available information it is clear now that in majority of infectious diseases host susceptibility is most likely to be highly polygenic [19]. Currently, the major advances in genetic epidemiological methodologies have helped in understanding the role of host genetic factors in susceptibility/resistance to infectious diseases [20,21]. Recent

human genome mapping information and the identification of a large number of candidate genes provide the tools for studying the role of host genetic factors in disease susceptibility/resistance in populations to give an insight on pathogenesis of disease for the preventive and therapeutic strategies [22,23]. Better understanding of these factors can be obtained by employing high throughput genomic approaches [24]. The few major genes that have been identified in several genomewide linkage scans for bacterial, parasitic and viral infectious diseases demonstrate or reveal the genetic susceptibility widely distributed among numerous polygenes [25].The available knowledge about the host's genetic profile would be considerable in identifying targets for new therapies, in improved vaccination strategies and eventually in disease elimination.

4. Epidemiological factors

The epidemiological factors encompass a wide range of areas like the identification of environment and genetic risk factors, the etiology of the diseases, the disease mechanisms, disease prevention strategies including both, the primary and secondary strategies etc. Analyzing the epidemiological factors in detail while conducting descriptive and analytical studies allow the epidemiologists to evaluate gene / environment interactions in disease etiology and be able to provide risk factor-specific morbidity rates for purposes of education and intervention [26]. The epidemiologists today make use of the molecular techniques to aid in the diagnosis of the infections. The laboratory identification and typing of infectious microorganisms facilitate the clinicians and epidemiologists to assess the molecular clinical and epidemiological relevance accurately [27]. This knowledge also helps to reduce heterogeneity in disease classification and allow examination of susceptibility or molecular markers with increased accuracy.

Epidemiological understanding includes definitions and nomenclature, outbreak investigations, disease surveillance, case-control studies, cohort studies, laboratory diagnosis, molecular epidemiology, dynamics of transmission, etc. The insights from this increasingly-important discipline are now involving policy-makers thus playing a major role in research. The need to understand the transmission patterns of infectious diseases, to be able to interpret and critically-evaluate the epidemiological data is an important aspect of the molecular epidemiological studies (figure 2).

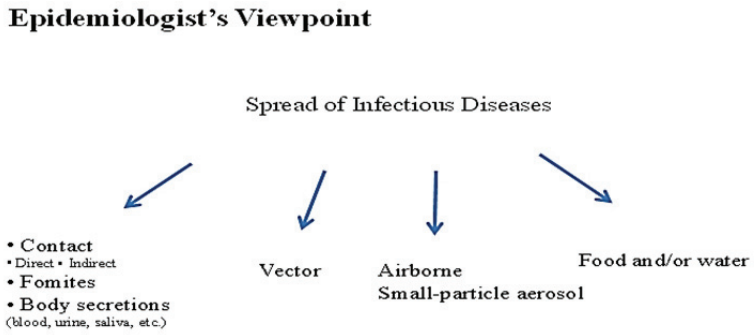

Fig. 2 Epidemiologists viewpoint.

5. Moelcular technologies

The molecular technlogical progress for the detection and characterisation of infectious organisms today, has made a wide range of molecular assays available, faciltiating the study of the distribution and determinants of disease occurrence in human populations thus, presenting the greatest opportunities in disease prevention. Molecular methods today have progressed from identification of antimicrobial resistance genes and provide public health information such as strain characterisation by genotyping [28]. The advantages of these assays are many, ranging from rapid turn-around time to high sensitivity and specificity to giving accurate results and user-friendly with cost-effective. The molecular diagnosis of disease microorganisms are mostly done by PCR assay, either using single step PCR or by nested PCR followed by gel electrophoresis detection [29]. Now with the advent of various stages of DNA or RNA extraction, amplification and product detection with real-time PCR, makes molecular assays more efficient. Recently several new molecular tools are developed, for the direct examination, diagnosis of the microorganisms from clinical/ non-clinical samples thus, providing access to relatively fast, cheap and advanced molecular assays [30].The DNA chip- the microarray technology further increases the utility of molecular diagnosis of microorganisms in the laboratory [31,32]. The molecular typing methods for microorganisms are summarized in the table 2.

Table 2 Typing methods used in the laboratory for diagnosis of microorganisms.

	Type	Methods	Examples	Advantages	Disadvantages
Phenotypic typing Conventional, Non-molecular Techniques	Biotyping	Strains are referred to as "biotypes". Biotyping may be performed manually or using automated systems	Tolerance to pH, chemicals and dyes, Hydrolysis of compounds, Haemagglutination, Hemolysis	Reproducible technique. Most strains are typeable	Discriminatory power poor,
	Phage Typing	Strains characterised by susceptibility or resistance pattern of to standard bacteriophages	Presence/absence of specific receptors	This test can be reproduced, has discriminatory power with ease of interpretation.	Requires maintenance of biologically active phages, a major limitation. Most strains are non-typeable.
	Serotyping	Strains differentiated by antigenic differences	Bacterial/latex agglutination, co-agglutination, fluorescent and enzyme labelling assays	Good reproducibility and ease of interpretation	Poor discriminatory power of serotypes, crossreaction and untypeable nature of some strains.
	Antibiogram	Typing to a set of antibiotics	identification of new or unusual pattern of antibiotic resistance among isolates	Ease of performance and fair reproducibility	poor discriminating power,
	Protein Typing	Typing the differences in the proteins made by different strains, immunoblotting	SDS page, enzymelabelled anti-immunoglobulins	Good reproducibility and ease of interpretation	Expensive, technically demanding, multiple strains difficult to type.
	Multilocus Enzyme Electrophoresis	Analysed on differences in the eletrophoretic mobilities of a set of metablolic enzymes	Cell extracts containing soluble enzymes are electrophoresed in starch gels	Excellent reproducibility with ease of interpretation	moderately discriminatory for the epidemiological analysis of clinical isolates, needs specific instruments for the test.
Genotypic typing Molecular Techniques	Plasmid analysis	PCR and gel electrophoresis.	Plasmid extraction	Good ease of interpretation.	isolates lacking plasmids are untypeable.
	RFLP	The DNA is sheared at a specific nucleotide sequence by restriction endonuclease	DNA digestion with endonuclease enzymes.	Good reproducibility for all the typed strains.	complex profile reduces the interpretation.
	PFGE	Gel electrophoresis	Orientation of the electric field across the gel is changed periodically	Restriction profiles are easily read and interpreted	Process long, costly and specific instruments needed.
	Southern blot analysis	only the particular restriction fragment is detected.	specific sequences are detected by labelled DNA probes.	Reproducible, easy to interpret.	Costly reagents, equipment and labour intensive
	Sequence Analysis	DNA (or RNA) nucleotide-base sequences	By amplifying a known DNA segment and automated techniques to sequence the amplified product	Reproducible results	Costly reagents, equipment and labour intensive
	Next-Generation Sequencing	Whole genome sequencing	High-resolution pathogen typing	Highly accurate, replacing multiple tests to identify the organism	Costly reagents, equipment and labour intensive
	Nucleic acid arrays	100s/1000s of sequence specific nucleotide probes	Microarray technique	Entire genome sequence can be determined on single chip	Expensive techniques, specific techniques, special expertise required.

6. Future prospects

Today, the human population globally faces continuous threat with respect to infectious diseases. Molecular data today are widely used in epidemiological studies to evaluate the transmission dynamics, to characterize the biology and diversity of the pathogens. The emerging and remerging infectious diseases can be dealt with accurate surveillance, evolving newer control and preventive strategies to contain the global spread of infectious microorganisms. The success stories of recent advances in molecular epidemiology of population studies gives hope to all the clinicians and researchers to keep their efforts united for better disease free tomorrow [33,34].

7. Conclusions

The field of molecular epidemiology is best understood as multi-disciplinary sub-specialty of epidemiology that makes use of the molecular, cellular and other biologic measurements into epidemiologic evaluations collectively.The major challenge in the discipline of molecular epidemiology is the immense number of studies being undertaken this broad umbrella but unable to weave the whole story due to the lack of the involvement of multidisciplinary team of clinicians, epidemiologists, biostatisticians, toxicologists and environmental scientists. A combined effort of all these necessary expertise is needed to address the broad spectrum of molecular epidemiology of infectious diseases.

References

[1] Foxman B, Riley L Molecular epidemiology: focus on infection. Am J Epidemiol. 2001 Jun 15;153(12):1135-41.

[2] Struelens MJ, Brisse S. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Euro Surveill. 2013;18(4):pii=20386.

[3] Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, et al. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerg Infect Dis. 2012;18(11):e1

[4] Nigel Field, Ted Cohen, Prof Marc J Struelens, Daniel Palm, Barry Cookson, et al. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. The Lancet Infect Dis. 2014; 14(4):341-352.

[5] Laupland KB, Lyytikainen O, Sogaard M, Kennedy KJ, Knudsen JD, Ostergaard C, Galbraith JC, Valiquette L, Jacobsson G, Collignon P, Schonheyder HC, for the International Bacteremia Surveillance Collaborative: The changing epidemiology of *Staphylococcus aureus*bloodstream infection: a multinational population-based surveillance study. Clin Microbiol Infect 2013, 19:465-471

[6] Christopher J L Murray, Katrina F Ortblad, Caterina Guinovart, Stephen S Lim, Timothy M Wolock, D Allen Roberts et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet, 2014;Vol. 384, No. 9947, p1005–1070.

[7] D Palm, A Johansson, A Ozin, AW Friedrich, JT Larsson, MJ Struelens. Molecular epidemiology of human pathogens: how to translate breakthroughs into public health practice, Stockholm, November 2011, Euro Surveill, 17; 2012, p. 2.

[8] Kramer A, Bekeschus S, Bröker BM, Schleibinger H, Razavi B, Assadian OJ. Maintaining health by balancing microbial exposure and prevention of infection: the hygiene hypothesis versus the hypothesis of early immune challenge. J. Hosp Infect. 2013 Feb;83 Suppl 1:S29-34. doi: 10.1016/S0195-6701(13) 60007-9.

[9] Schaumburg F, Köck R, Friedrich AW, Soulanoudjingar S, Ngoa UA, et al. Population structure of *Staphylococcus aureus* from remote African Babongo Pygmies. PLoS Neglect Trop Dis 2011;5: e1150. doi: 10.1371/journal.pntd.0001150

[10] National Institutes of Health (US); Biological Sciences Curriculum Study. NIH Curriculum Supplement Series [Internet]. Bethesda (MD): National Institutes of Health (US); 2007-. Understanding Emerging and Re-emerging Infectious Diseases.Available from: http://www.ncbi.nlm.nih.gov/books/NBK20370/

[11] Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Introduction to Pathogens. Available from: http://www.ncbi.nlm.nih.gov/books/NBK26917/

[12] Ambrose CB, Kadlubar FF. Toward an integrated approach to molecular epidemiology. Am J Epidemiol 1997;146:912–18.

[13] Kamangar F. Confounding variables in epidemiologic studies: basics and beyond. Arch Iran Med. 2012 Aug;15(8):508-16. doi: 012158/AIM.0014.

[14] LW Riley. Molecular epidemiology of infectious disease: principles and practicesASM Press, Washington, DC (2004)

[15] Ramagopalan SV, Knight JC, Ebers GC. Multiple sclerosis and the major histocompatibility complex. Curr Opin Neurol. 2009 Jun;22(3):219-25. doi: 10.1097/WCO.0b013e32832b5417.

[16] Melanie Newport and Chris Finan. Genome-wide association studies and susceptibility to infectious diseases. Briefings in Functional Genomics (2011) 10 (2): 98-107 doi:10.1093/bfgp/elq037.

[17] Doeschl-Wilson, Andrea B. et al. “Implications of Host Genetic Variation on the Risk and Prevalence of Infectious Diseases Transmitted Through the Environment.” Ed. D. W. Threadgill. Genetics 188.3 (2011): 683–693. *PMC*. Web. 15 Sept. 2015.

[18] Bishop S. C., Woolliams J. A. On the genetic interpretation of disease data. PLoS ONE 2010;5(1): e8940.

[19] Segal S., Hill A. V. S. Genetic susceptibility to infectious disease. Trends Microbiol., 2003;11(9): 445–448.

[20] Daetwyler H. D., Villanueva B., Woolliams J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 2008;3(10): e3395

[21] Woolhouse M. E., Etard J. F., Dietz K., Ndhlovu P. D., Chandiwana S. K. Heterogeneities in *schistome* transmission dynamics and control. Parasitology 1998;117: 475–482.

[22] Bernard Kaevney. Genetic epidemiological studies of coronary heart disease Int. J. Epidemiol. 2002; 31(4): 730-736 doi:10.1093/ije/31.4.730
[23] Hill, Adrian V. S. "Evolution, Revolution and Heresy in the Genetics of Infectious Disease Susceptibility." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1590 (2012): 840–849. *PMC*. Web. 15 Sept. 2015.
[24] Hirschhorn J. N., Daly M. J. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 2005;6: 95–10
[25] McClellan J., King M. C. Genetic heterogeneity in human disease. Cell 2010;141, 210–217
[26] Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl Jm, Laurent F, Grundmann H, Friedrich AW. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. ESCMID Study Group of Epidemiological Markers (ESGEM).Euro Surveill. 2013 Jan 24; 18(4):20380. *Epub 2013 Jan 24*
[27] A van Belkum, PT Tassios, L Dijkshoorn, the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group on Epidemiological Markers (ESGEM), *et al.*Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect, 13 (suppl 3) 2007; pp. 1–46
[28] Sabat A, Krzyszton-Russjan J, Strzalka W, Filipek R, Kosowska K, Hryniewicz W, et al. New method for typing Staphylococcus aureus strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. J Clin Microbiol. 2003;41(4):1801-4.
[29] Sabat A, Malachowa N, Miedzobrodzki J, Hryniewicz W. Comparison of PCR-based methods for typing Staphylococcus aureus isolates. J Clin Microbiol. 2006;44(10):3804-7.
[30] Burd EM. Validation of Laboratory-Developed Molecular Assays for Infectious Diseases. Clinical Microbiology Reviews. 2010;23(3):550-576. doi:10.1128/CMR.00074-09.
[31] Friedrich AW, Witte W, Harmsen D, de Lencastre H, Hryniewicz W, Scheres J, et al. SeqNet.org: a European laboratory network for sequence-based typing of microbial pathogens. Euro Surveill. 2006;11(2):pii=2874. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2874
[32] Urwin R, Maiden MC. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. 2003;11(10):479-87.
[33] Braun SD, Ziegler A, Methner U, Slickers P, Keiling S, Monecke S, et al. Fast DNA Serotyping and Antimicrobial Resistance Gene Determination of Salmonella enterica with an Oligonucleotide Microarray-Based Assay. PloS one. 2012;7(10):e46489.
[34] Jackson SA, Kotewicz ML, Patel IR, Lacher DW, Gangiredla J, Elkins CA. Rapid genomic-scale analysis of Escherichia coli O104:H4 by using high-resolution alternative methods to next-generation sequencing. Appl Environ Microbiol. 2012;78(5):1601-5.

