Molecular epidemiology of infectious diseases

Sana Eybpoosh[1], Ali Akbar Haghdoost[1], Ehsan Mostafavi[2], Abbas Bahrampour[3], Kayhan Azadmanesh[4], Farzaneh Zolala[3]

[1] HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[2] Department of Epidemiology and Biostatistics, Research Centre for Emerging and Reemerging infectious diseases, Pasteur Institute of Iran, Tehran, Iran
[3] Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[4] Department of Virology, Pasteur Institute of Iran, Tehran, Iran

Type of article: Review

Abstract

Molecular epidemiology (ME) is a branch of epidemiology developed by merging molecular biology into epidemiological studies. In this paper, the authors try to discuss the ways that molecular epidemiology studies identify infectious diseases' causation and pathogenesis, and unravel infectious agents' sources, reservoirs, circulation pattern, transmission pattern, transmission probability, and transmission order. They bring real-world examples of research works in each area to make each study design more understandable. They also address some research areas and study design aspects that need further attention in future. They close with some thoughts about future directions in this field and emphasize on the need for training competent molecular epidemiology specialists that are capable of dealing with rapid advances in the field.

Keywords: Epidemiology, Molecular; Infectious disease

1. Introduction

Epidemiology aims to study the distribution, determinants, and causes of health-related problems in order to prevent and control them effectively. The far-reaching influence of epidemiology in prevention of human diseases can be traced back to several large-scale epidemiological studies initiated in the 1940s, such as community trials of fluoride supplementation in 1940 which led to primary prevention of dental carries (1), Framingham Heart Study initiated in 1947 which led to identification of many risk factors and determinants of heart diseases (2), or polio vaccine trials initiated in 1954 which provided the basis for prevention and eradication of poliomyelitis (3). Despite beneficial effects of traditional epidemiological studies on the prevention and control of many diseases, there still remain important questions about the biological mechanisms underlying either infectious or chronic diseases. In cases of infectious diseases, scientists have always been curious about the answers to the following questions: 1) What is the cause and pathogenesis of an infectious disease? 2) Why do people respond differently to a specific pathogen, therapy, or vaccine (e.g., why do some people die from influenza while others stay asymptomatic)? 3) Why are some HIV subtypes transmitted more frequently through sexual contact rather than other modes of transmission? Or, 4) how can the transmission of an infectious disease be prevented more effectively? Gaining in-depth insight into these questions is necessary for developing more effective prevention and control strategies. However, using traditional epidemiological studies alone, it is difficult to gain deeper understanding of the above questions (4), while, molecular epidemiological (ME) studies have a great potential to address these issues. Molecular epidemiology is a branch of epidemiology developed by merging molecular biology into epidemiological studies. It was the development of high throughput laboratory techniques, applicable to large sample sizes that enabled epidemiologists to conduct large-scale epidemiological studies at molecular levels (5). Although it is claimed that

the term "molecular epidemiology" was first introduced by Lower et al. in 1979 (6), it seems that Kilbourne had used the term before, in 1973 to describe Influenza subtype distribution around the world (7). Since then, a dramatic rise in the number of molecular epidemiological studies have been observed in various research areas (Figure 1). In the field of infectious disease, ME studies have provided valuable information about infectious disease causation, pathogenesis, circulation, transmission, prevention, and therapy (Figure 2). Some methods that ME studies have used to cover knowledge gaps in these areas are discussed in this paper. Table 1 summarizes these methods as well.

Table 1. Methodological overview and rationale for some ME methods used in the field of infectious disease

| Domain	Study design or analysis method	
Causation of infectious disease	Case-control	Cases: patients with clinical manifestations of the disease; Controls: patients without clinical manifestations of the disease; Exposure: genetic material of pathogens
	Nested case-control/cohort	Exposed: individuals infected with the pathogen; Unexposed: individuals not infected with the pathogen; Outcome: disease onset
Pathogenesis of infectious disease	Case-control	Cases: 1) Severe, fatal, drug-resistant patients, 2) Patients with a disease or a especial clinical symptom; Controls: 1) Mild, non-fatal, drug-sensitive patients, 2) Patients without a disease or a especial clinical symptom Exposure: 1) Pathogen's genetic sequence, 2) Metagenome in a specific body site
	Genome Wide Association studies	Cases: patients with the disease or disease susceptibility Controls: patients without the disease or disease susceptibility Exposure: millions of SNPs along human genome
Sources/reservoirs	Phylogenetic analysis	Sampling from suspected sources and infected individual
Circulation pattern	Molecular surveillance	Routine, population-based genotyping of circulating pathogens
Transmission probability	Cross-sectional studies using phylogenetic analysis	Pathogen's genetic sequence isolated from temporally and epidemiologically related individuals
Transmission patterns	Cross-sectional studies using phylogenetic analysis in combination with behavioral data	Pathogenic genetic material is isolated from infected individuals within or between communities.
Transmission order	Bayesian phylogeographic studies	The order of virus spread across large geographic distances is estimated using information accumulated in the virus genome. Viral genomes are used to reconstruct a family tree showing the relationships between these viral samples
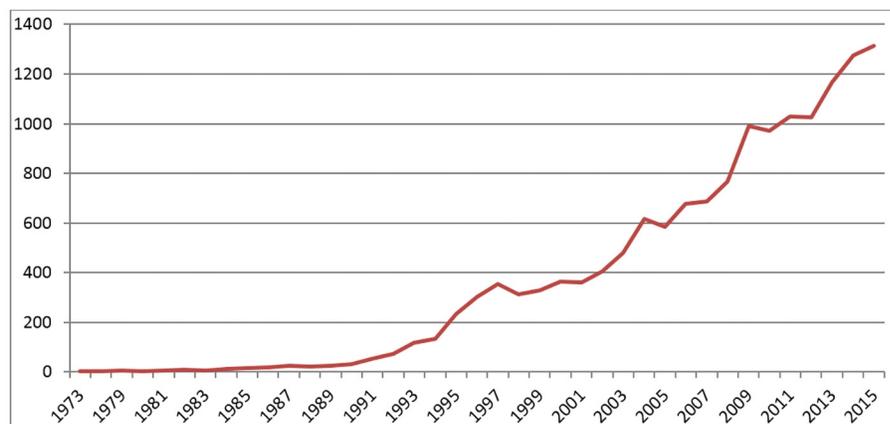
Figure 1. The rising trend in the number of scientific papers per year cited in PubMed including the term "molecular epidemiology" in the title, abstract or key word (by the end of December 2015)
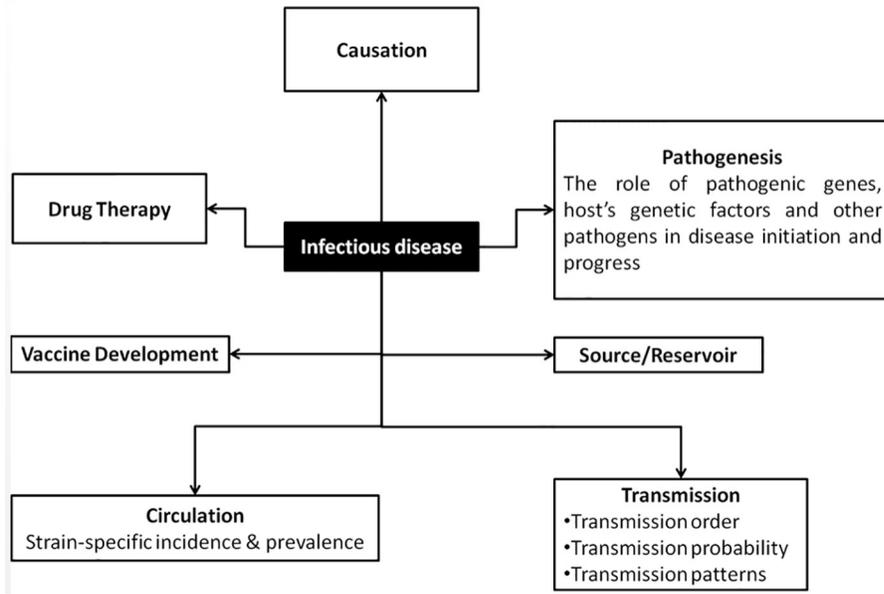
Figure 2. Main domains of infectious diseases addressed by molecular epidemiological studies

2. Discussion

2.1. Identification of the cause of infectious diseases

Molecular epidemiological studies are needed to confirm the causative role of a candidate pathogen in a given disease. In this regard, usually a case-control study is conducted at first, including cases that have and controls that do not have the disease symptoms. The aim is to see whether a given pathogen is present in cases, and absent in controls or not (8). It has been noted that the vast majority of pathogens are not detectable by traditional culture-based techniques, therefore, culture-based research in this area might not detect the causative agent in cases (patients), and might provide misleading information about the exact cause of infectious disease (9); however, advanced molecular techniques have provided the opportunity to detect trace amounts of genetic materials of a pathogen in various specimens with sensitivity that is far beyond culture-based methods (10). Here, pathogens whose genetic material is detected in specimens of cases but not in specimens of controls become candidate causative agent(s) of the disease of interest. After this step, we need to confirm if the pathogen(s) do cause the disease or not. To achieve this goal, we need to see if infection with the pathogen precedes the disease onset. This hypothesis can be checked efficiently in a cohort or a nested case-control study. A real-world case study might better illustrate the process. Kaposi's Sarcoma (KS) was hypothesized to be caused by an infectious agent, but culture-based techniques could not detect any infectious agent in specimens of KS patients. In 1994, Chang et al., found a unique DNA sequence in more than 90 percent of KS patients who had also Acquired Immunodeficiency Syndrome (AIDS). The sequence was also detected in tissue of non-AIDS patients with KS but was absent in tissue of non-AIDS and non-KS patients. Comparison of the DNA sequence with known pathogenic sequences revealed that the sequence is similar to herpes viruses, but is distinct from them in a small number of genes. Therefore, they concluded that the identified sequence probably belongs to a virus which probably belongs to the human herpes virus family (11). Another study was then needed to see if the infection with the virus precedes the onset of KS; based on Hill's criteria of causation, this condition is known as temporality and is an important prerequisite for establishing causality between the virus and the disease (12). For this reason, Moore and colleagues (1996) conducted a nested case-control study on ongoing cohorts of HIV positive individuals. They compared the presence of KS-associated herpes virus in blood samples of incident cases of KS+HIV+ and controls of KS-HIV+ individuals and concluded that KS+HIV+ patients are significantly more likely to show evidence of infection with the virus prior to the onset of KS (13). Therefore, the role of this virus in development of KS was established.

2.2. Identification of pathogenesis of infectious diseases

Combination of molecular techniques with epidemiological studies can reveal more secrets about the pathogenesis process of infectious diseases. For example, by comparing the genome sequence of a particular pathogen between severe, fatal, or drug-resistant cases and mild, non-fatal or drug-sensitive controls, we can find out if there is any difference between genome sequences of microbes in both groups of patients. Genetic regions that are not identical

in two groups of patients might be responsible for the observed clinical symptoms in cases, here, the virulence, case fatality, or drug-resistance, respectively (Figure 3). For example, a new type of methicillin-resistant Staphylococcus aureus (MRSA) was identified to cause community-acquired MRSA with high fatality rate. Therefore, using sequencing techniques, Baba et al., (2002) determined whole genome sequence of this species, isolated from an infant with fatal septicemia and septic arthritis due to this species. They compared the genome sequence of this species with those obtained from two cases of hospital-acquired MRSA presenting milder clinical manifestations. They found that the pathogen responsible for community-acquired MRSA carries a range of genes that is distinct from extant hospital-acquired MRSA strains which could be attributed to medically important phenotypes of the new species, i.e. virulence and case fatality (14). Such findings would be applicable for developing appropriate tools to detect if there is a potential for severe or fatal conditions and prevent them appropriately. Different levels of disease severity caused by a pathogen can also be due to differences in human host response to the infection. Such a difference might be due to differences in individuals' genetic susceptibility to or inherent immunity against the pathogen. One epidemiological method that has a great ability to assess this possibility is Genome Wide Association (GWA) study. These types of studies can be considered as a derivative of case-control studies; they compare the DNA of two groups of individuals: those who have the disease (cases) and those who are similar to cases but do not have the disease (controls).
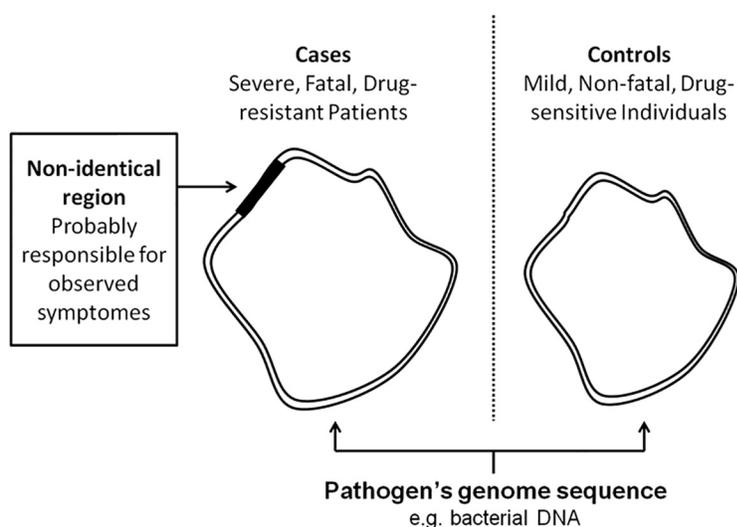

Figure 3. Pathogen's genome sequence in cases and controls. Non-identical genetic region might code for medical phenotype seen in cases, such as disease severity, fatality, drug-resistance, etc.

The exposure variables here are millions of common genetic variants (mutations) in human genome, known as single nucleotide polymorphisms (SNPs). Those SNPs which are seen more/less frequently in cases rather than controls are considered to be associated with the disease of interest. Since 2005, there have been an increasing number of published GWA studies, some of which have found SNPs in humans that are associated with susceptibility to and progress of infectious disease like HIV (15-17), AIDS (18), tuberculosis (19, 20), and malaria (21, 22). For example, Timmann et al. (2012) conducted a GWA study to unravel relevant SNPs that associate with malaria resistance. In this regard, they recruited 1,325 severe malaria cases and 828 unaffected controls. They found two novel SNPs associated with malaria resistance; one of them was identified within a gene that encodes the main calcium pump of erythrocytes (target cells for malaria parasite) and the other was an SNP linked to a protein that had a role in vascular endothelial cell damage during parasitemia (21). Such findings provide insight about mechanisms underlying pathogenesis of infectious diseases and the role of the human host's genetic characteristics in disease severity and progression. There have been attempts to use such genetic markers (SNPs) as a means to distinguish more genetically vulnerable populations in an epidemic, and improve the accuracy of predictions about disease prognosis. Today, it has been realized that infectious agents often do not act independently; rather, they interact with each other in a way that their infectivity can be mediated by other pathogens in a significant manner (23). The Human Microbiome Project (HMP) has determined the metagenome of species living in the human body. Metagenome is the whole genetic material recovered directly from an environment or human body such as the gut, nasal cavity, mouth, vagina, etc. The findings of HMP has brought the idea of Pathogen-Pathogen Interaction (PPI), and persuaded investigators to investigate the role of PPI in health and disease (24). For example, Samb-Ba, et al.,

(2014) conducted a case-control study on 162 cases of diarrhea and 185 healthy controls in Senegal. Combining various laboratory techniques (phenotypic methods, mass spectrometry, and 16S rRNA sequencing), they determined gut microbiota in cases and controls. They realized that the number of bacterial species were significantly lower among patients with diarrhea than those without diarrhea in a way that the numbers of commensal bacterial species were significantly decreased in patients with diarrhea. On the other hand, several Bacillus spp. (some of which were not detected or reported before) were significantly more frequent in these patients rather than the control group (25).

2.3. Prevention of infectious disease transmission
Another important challenge to combat infectious diseases is the effective prevention of their transmission. To achieve this goal, epidemiologists have to obtain exact knowledge about the origin, sources, reservoirs, circulation pattern, transmission probability, and transmission patterns of pathogens (10). Molecular epidemiological studies can enrich our knowledge in these aspects, as well.

2.3.1. Identification of infectious diseases' sources/reservoirs
Molecular tools have enhanced the ability to detect trace amounts of genetic materials in the environment (such as water, soil, human, etc.) which was not traceable before by traditional culture techniques. This opportunity has helped scientists to find new reservoirs/sources of infection and design more effective preventive strategies accordingly (9). In this regard, they need to identify if the pathogen isolated from the infected individual is genetically similar to the pathogen obtained from suspected sources or not. If they are genetically similar in a significant manner, then the suspicious source is considered as a new source of infection for that pathogen. To check for genetic similarity between two pathogens, epidemiologists have turned to grouping/clustering methods used in evolutionary biology. Phylogenetics is one of these methods that attempts to study the similarity and evolutionary relationships among various species. An example in this area is the identification of chicken as a new reservoir for Extraintestinal Pathogenic Escherichia coli (ExPEC) causing Urinary Tract Infection (UTI) in Humans. Previously, the host's intestinal tract was considered to be the only source of ExPEC-induced UTI. Usually such endogenous diseases are not expected to produce outbreaks. However, by 2010, a small number of ExPEC outbreaks were detected in different parts of the world (26). This observation lead scientists to hypothesize that a common source or vehicle - other than the host's intestinal tract - might be the source of these outbreaks. To search for additional sources, Bergeron et al. conducted a series of studies; they collected samples of temporally and geographically matched E. coli isolates from animals and from humans with UTIs. With the aid of phylogenetic methods, they found that E. coli isolates from chicken are genetically similar to isolates from humans with UTIs in a significant manner (27, 28). So a new reservoir was found for the pathogen which brought appropriate preventive strategies to the community accordingly.

2.3.2. Circulation of microbial strains
Routine, population-based genotyping of circulating pathogens is now implemented in many settings as a tool for surveillance of existing microbial strains and drug resistance mutations in the community (29). For pathogens, in which different strains have different propensities to cause the diseases, knowledge about circulating strains is essential for developing strain-specific vaccines. For this reason, the influenza vaccine is being prepared and distributed based on circulating strains each year (30). Furthermore, monitoring local, national and regional levels of drug resistance inform treatment guidelines and provide feedback on the success of infectious disease treatment and prevention programs (29).

2.3.3. Identification of infectious diseases' transmission probability
Molecular epidemiological studies can also provide valuable information about transmission probability of infectious diseases (an important component in computing basic reproductive number-R0 (31)) which cannot be provided correctly using questionnaire data alone (10). A typical example in this regard is the accurate and efficient estimation of transmission probability of sexually transmitted infections (STIs) using ME studies. Previously, transmission probability of STIs could be estimated by conducting longitudinal studies and following couples where one was infected and the other was susceptible (32); however, besides logistic and ethical issues of such study designs, it was difficult to verify if the transmission event (if happened) came from within the partnership. Now, epidemiologists can conduct cross-sectional studies on infected couples and use phylogenetic analysis to verify if bacterial type isolated from both partners are similar to each other or not. Phylogenetic methods can also reveal the order of transmission between infected couples; in other words, these methods can identify which partner has transmitted the infection to the other. When combined with contact data (i.e. use of condom and number of contacts), this information also helps epidemiologists to estimate the probability of infection per unprotected contact (33). Additionally, as molecular tools can also determine genome of the pathogen in asymptomatic partners,

transmission events are ascertained more exactly this way, therefore, estimation of transmission probabilities are more accurate in ME studies rather than traditional methods (10).

2.3.4. Identification of between-community transmission patterns

Phylogenetic methods, when combined with sociodemographic and behavioral data, can provide valuable information about transmission of the pathogen between communities (33). For example, Jahanbakhsh et al., (2014) found that the same HIV-1 subtype (the CRF35_AD clade) was circulating in Afghanistan and Iran. Therefore, they hypothesized that transmission events might have happened between these countries. To test this hypothesis, they applied phylogenetic analysis on DNA sequence of HIV-1 isolates obtained from these countries. They realized that the DNA sequence of HIV-1 obtained from Afghanistan and Iran were significantly similar together (i.e. they made significant clusters together). Behavioral data showed that all patients within these clusters were injecting drug users (IDUs). This finding indicated that there were probably between-country transmission events, occurring among IDU populations. Considering sociodemographic characteristics of these communities, authors concluded that the transmission might have happened through migration of Afghan refugees to Iran or through drug trafficking activities between these countries (34). The order of transmission between two or more communities can also be identified by the aid of phylogeographic methods. Phylogeography is the study of the historical processes that probably shaped current geographic distributions of living species, such as viruses. Today, these studies are done in light of the information gained from gene genealogies. Genealogies are family trees that represent the ancestor-descendent relationships between individuals in a population, e.g., viruses in an epidemic (35). A number of methods have been proposed to trace dispersal patterns of pathogens. Within these methods, the family tree is first reconstructed, without including any spatial information, and the phylogeographic inference is then made from this tree (36-39). These methods, however, do not address all the uncertainties associated with spatial dispersion of a pathogen. Therefore, more rigorous and sophisticated methods have been recently developed that use the Bayes theorem to integrate and simultaneously estimate all the parameters associated with spatio-temporal dispersion of a pathogen. In Bayesian phylogeographic methods, geographic locations from which the pathogens are isolated are considered as discrete variables. This information is then integrated into a phylogenetic tree - that shows the ancestor-descendent relationships of the isolated pathogens - to estimate the location of ancestral strains using the well-known method of "ancestral state reconstruction" (37-40). The information obtained in this method, can be used for inference about dispersion pathways of the epidemics and the order of virus transmission across countries. Using Bayesian phylogeographic methods, Eybpoosh, et al. (2016) investigated if the aforementioned CRF35_AD epidemic observed among IDUs had been transmitted from Afghanistan to Iran or vice versa. In this regard, they applied a Bayesian phylogenetic model on geo-referenced and time-stamped CRF35_AD sequence data from Afghanistan and Iran. They observed that despite the general perception about the unidirectional transmission of CRF35_AD, this strain continuously transmitted between Afghanistan and Iran in a bi-directional manner. The authors supposed that mass migration of Afghan refugees to Iran and drug trafficking and injection networks might play an important role in this regard (41). This example illuminates how integration of molecular data with epidemiological concepts and statistical techniques can improve our understanding of the transmission dynamics within an epidemic. Another example in this area is the identification of HIV-1 subtype B origin in Iran. For a long period of time, it was assumed that the subtype B epidemic in Iran originated from contaminated blood products imported from France (42-43). However, in another study conducted by Eybpoosh, et al. (2016), on all HIV-1 subtype B strains available from Iran revealed that multiple parallel subtype B epidemics existed in Iran with independent origins. They noted that the subtype B epidemic observed among heterosexuals, has an independent evolutionary history from the epidemic observed among hemophiliac patients and is probably imported to Iran from USA, Germany, and Hong Kong, but not from France (44).

3. Conclusions

The application of molecular techniques to epidemiologic studies have enhanced our understanding of infectious diseases. Molecular tools have improved our ability to sub-classify the organisms into meaningful groups, detect even trace amounts of living organisms in the human body or in the environment, and early detect diseases' symptoms. Conduction of epidemiologic studies on these data, can assist molecular biologists to identify previously-unknown aspects of infectious diseases, such as infectious diseases' etiology and pathophysiology and infectious agents' sources, reservoirs, circulation pattern, transmission probability, and transmission patterns. The information obtained in each of the above aspects helps practitioners and decision makers in the development of more effective prevention and control strategies at individual and population levels. These achievements, can also be used to further characterize the discipline of epidemiology.

4. Future Directions

ME studies have provided novel information about infectious diseases. However, compared to the dramatic progress of ME studies in the field of chronic disease; studies in the field of infectious disease are in their infancy. In the future, greater attention should be paid to the design of ME studies, especially to incorporate appropriate cases and controls and to account for potential confounders. In addition, application of longitudinal studies should be considered as a great opportunity for ME studies especially when working with metagenomic and metatranscriptomic data. Since metagenomic and metatranscriptomic (study of whole transcripts present in a specific environment) studies provide a snapshot of whole genetic materials present at each time in an environment, having chronological information of changes in genes or transcript composition in a body site can provide great knowledge about the pathogenesis of infectious diseases and the way hosts and other pathogens interact with the infecting agent during the course of the disease. In the case of GWA studies, further attempts are needed to establish the clinical utility of genomic information and to issue guidelines for clinical practice (45). In this regard, identified SNPs should be established as a means to diagnose genetically vulnerable populations, develop proper drugs, and provide individualized therapies. At the same time, attempts should be made to improve the accuracy of predictions about disease prognosis based on identified genetic markers. In the field of chronic disease, a small number of genetic risk scores have been introduced that attempt to predict prostate cancer (46, 47), survival after cardiovascular events (48), coronary artery bypass graft surgery (49), etc. For management of long-period infectious diseases such as AIDS, malaria and tuberculosis, providing such risk scores also seems beneficial. In general, as the time goes by, the "Omics" area (the field of study in biology ending in -omics, such as genomics, transcriptomics, proteomics, etc.) continues to progress rapidly by development of more sensitive techniques, databases that store omics data, and bioinformatics tools that can efficiently handle such data. Moreover, the cost of molecular techniques continues to fall day by day. Therefore, it becomes more convenient to conduct large-scale and multicenter investigations even in low and middle income countries. All these opportunities make it possible to answer more complex questions about the nature of infectious diseases in different populations and geographic areas, which were not investigable questions before; It should be emphasized that to take the best advantage of these opportunities, it is necessary to systematically train individuals who are familiar with methods and theories in the field of epidemiology and molecular biology, and are capable of dealing with rapid advances in these areas. Finally, we believe that the potential benefits of molecular epidemiology in prevention of infectious diseases justify a dedication to the further development and use of this approach to infectious disease prevention. We are hopeful that this new discipline will continue to provide new insights into the causation, pathogenesis and transmission of infectious diseases, and will help us arrive at more effective strategies for prevention of infectious diseases.

Conflict of Interest:

There is no conflict of interest to be declared.

Authors' contributions:

All authors contributed to this project and article equally. All authors read and approved the final manuscript.

References:

1) Arnold FA Jr, Dean HT, Jay P, Knutson JW. Effects of fluoridated water supplies on dental caries incidence: Results of the seventh year of study at Grand Rapids and Muskegon, Mich. Public Health Rep. 1953; 68: 141-8.
2) Oppenheimer GM. Becoming the Framingham Study 1947-1950. Am J Public Health. 2005; 95(4): 602-10. doi: 10.2105/AJPH.2003.026419. PMID: 15798116, PMCID: PMC1449227.
3) Francis T, Korns R, Voight R, Boisen M, Hemphill F, Napier J. An evaluation of the 1954 poliomyelitis vaccine trials. Am J Public Health Nations Health. 1955; 45(5 Pt 2): 1-63. PMID: 14361811, PMCID: PMC1622939.
4) Susser M. Does risk factor epidemiology put epidemiology at risk? Peering into the future. J Epidemiol Community Health. 1998; 52: 608-11. PMID: 10023453, PMCID: PMC1756623.

5) Haghdoost AA. Molecular Epidemiology, Concepts and Domains. J Kerman University of Medical Sciences. 2008; 15(1): 97-104.

6) Vineis P. Commentary: First steps in molecular epidemiology: Lower et al. 1979. Int J Epidemiol. 2007; 36(1): 20-2. doi: 10.1093/ije/dyl291.

7) Kilbourne ED. The molecular epidemiology of influenza. J Infect Dis. 1973; 127(4): 478-87. PMID: 4121053.

8) Giesecke I. Modern Infectious Disease Epidemiology, Second Edition. Boca Raton: CRC Press; 2001.

9) Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol. 1998; 180(18): 4765-74. PMID: 9733676, PMCID: PMC107498.

10) Foxman B. Molecular Tools and Infectious Disease Epidemiology. Burlington, MA: Elsevier; 2012.

11) Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science. 1994; 266(5192): 1865-9. PMID: 7997879.

12) Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med. 1965; 58(5): 295-300. PMID: 14283879, PMCID: PMC1898525.

13) Moore PS, Kingsley LA, Holmberg SD, Spira T, Gupta P, Hoover DR, et al. Kaposi's sarcoma-associated herpesvirus infection prior to onset of Kaposi's sarcoma. AIDS. 1996: 10(2): 175-80. PMID: 8838705.

14) Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, Oguchi A, et al. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet. 2002; 359(9320): 1819-27. PMID: 12044378.

15) McLaren PJ, Coulonges C, Ripke S, van den Berg L, Buchbinder S, Carrington M, et al. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. PLoS Pathog. 2013; 9(7): e1003515. doi: 10.1371/journal.ppat.1003515. PMID: 23935489, PMCID: PMC3723635.

16) Lingappa JR, Petrovski S, Kahle E, Fellay J, Shianna K, McElrath MJ, et al. Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. PLoS One. 2011; 6(12): e28632. doi: 10.1371/journal.pone.0028632. PMID: 22174851, PMCID: PMC3236203.

17) Petrovski S, Fellay J, Shianna KV, Carpenetti N, Kumwenda J, Kamanga G, et al. Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. AIDS. 2011; 25(4): 513-8. doi: 10.1097/QAD.0b013e328343817b. PMID: 21160409, PMCID: PMC3150594.

18) Herbeck JT, Gottlieb GS, Winkler CA, Nelson GW, An P, Maust BS, et al. Multistage genomewide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. J Infect Dis. 2010; 201(4): 618-26. doi: 10.1086/649842. PMID: 20064070, PMCID: PMC2928718.

19) Chimusa ER, Zaitlen N, Daya M, Möller M, van Helden PD, Mulder NJ, et al. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet. 2014; 23(3): 796-809. doi: 10.1093/hmg/ddt462. PMID: 24057671, PMCID: PMC3888262.

20) Mahasirimongkol S, Yanai H, Mushiroda T, Promphittayarat W, Wattanapokayakit S, Phromjai J, Yuliwulandari R, et al. Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. J Hum Genet. 2012; 57(6): 363-7. doi: 10.1038/jhg.2012.35. PMID: 22551897.

21) Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature. 2012; 489(7416): 443-6. doi: 10.1038/nature11334. PMID: 22895189.

22) Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet. 2009; 41(6): 657-65. doi: 10.1038/ng.388. PMID: 19465909, PMCID: PMC2889040.

23) Singer M. Pathogen-pathogen interaction: a syndemic model of complex biosocial processes in disease. Virulence. 2010; 1(1): 10-8. doi: 10.4161/viru.1.1.9933. PMID: 21178409, PMCID: PMC3080196.

24) Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett C, Knight R, Gordon JI. The human microbiome project. Nature. 2007; 449(7164): 804–10. doi: 10.1038/nature06244. PMID: 17943116, PMCID: PMC3709439.

25) Samb-Ba B, Mazenot C, Gassama-Sow A, Dubourg G, Richet H, Hugon P, et al. MALDI-TOF Identification of the Human Gut Microbiome in People with and without Diarrhea in Senegal. PLoS One. 2014; 9(5): e87419. doi: 10.1371/journal.pone.0087419. PMID: 24784934, PMCID: PMC4006720.

26) George DB, Manges AR. A systematic review of outbreak and non-outbreak studies of extraintestinal pathogenic Escherichia coli causing community-acquired infections. Epidemiol Infect. 2010; 138: 1679-90. doi: 10.1017/S0950268810001639. PMID: 20642873.

27) Cortés P, Blanc V, Mora A, Dahbi G, Blanco JE, Blanco M, et al. Isolation and characterization of potentially pathogenic antimicrobial-resistant Escherichia coli strains from chicken and pig farms in Spain. Appl Environ Microbiol. 2010; 76(9): 2799-805. doi: 10.1128/AEM.02421-09. PMID: 20228098, PMCID: PMC2863447.

28) Bergeron CR, Prussing C, Boerlin P, Daignault D, Dutil L, Reid-Smith RJ, et al. Chicken as reservoir for extraintestinal pathogenic Escherichia coli in humans, Canada. Emerg Infect Dis. 2012; 18(3): 415-21. doi: 10.3201/eid1803.111099. PMID: 22377351, PMCID: PMC3309577.

29) Shafer RW, Rhee SY, Bennett DE. Consensus drug resistance mutations for epidemiological surveillance: basic principles and potential controversies. Antivir Ther. 2008; 13(Suppl 2): 59-68. PMID: 18575192, PMCID: PMC4388302.

30) Wang R, Taubenberger JK. Methods for molecular surveillance of influenza. Expert Rev Anti Infect Ther. 2010; 8(5): 517-27. doi: 10.1586/eri.10.24. PMID: 20455681, PMCID: PMC2882197.

31) Haghdoost A, Baneshi MR, Zolala F, Farvahari S, Safizadeh H. Estimation of Basic Reproductive Number of Flu-like Syndrome in a Primary School in Iran. Int J Prev Med. 2012; 3(6): 408-13. PMID: 22783467, PMCID: PMC3389438.

32) Hooper RR, Reynolds GH, Jones OG, Zaidi A, Wiesner PJ, Latimer KP, et al. Cohort study of venereal disease. I: the risk of gonorrhea transmission from infected women to men. Am J Epidemiol. 1978; 108(2): 136-44. PMID: 707474.

33) Lemey P, Salemi M, Vandamme A. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. 2nd edition. Cambridge: Cambridge University Press; 2009.

34) Jahanbakhsh F, Eybpoosh S, Mostafavi E, Haghdoost AA, Azadmanesh K. Molecular Epidemiology of HIV-1 in Afghanistan, Iran, and Pakistan. J Med Microbiol Infec Dis, 2015, 3 (3-4): 44-47

35) Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009; 5: e1000520. doi: 10.1371/journal.pcbi.1000520. PMID: 19779555, PMCID: PMC2740835.

36) Holmes E. The phylogeography of human viruses. Mol Ecol. 2004; 13: 745-56. doi: 10.1046/j.1365-294X.2003.02051.x. PMID: 15012753.

37) Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics. 1989; 123: 603-13. PMID: 2599370, PMCID: PMC1203833.

38) Templeton A. Statistical phylogeography: methods for evaluating and minimizing inference errors. Mol Ecol. 2004; 13: 789-809. doi: 10.1046/j.1365-294X.2003.02041.x. PMID: 15012756.

39) Lanciotti R, Gubler D, Trent D. Molecular evolution and phylogeny of dengue-4 viruses. J Gen Virol. 1997; 78: 2279-86. doi: 10.1099/0022-1317-78-9-2279. PMID: 9292015.

40) Swofford WM. Systematics, Historical Ecology, and North American Freshwater Fishes. Parsimony, character-state reconstructions and evolutionary inferences. Stanford University Press; 1992.

41) Eybpoosh S, Bahrampour A, Karamouzian M, Azadmanesh K, Jahanbakhsh F, Mostafavi E, et al. Spatio-Temporal History of HIV-1 CRF35_AD in Afghanistan and Iran. PLoS One. 2016; 11(6): e0156499. doi: 10.1371/journal.pone.0156499. PMID: 27280293, PMCID: PMC4900578.

42) Sarrami-Forooshani R, Das SR, Sabahi F, Adeli A, Esmaeili R, Wahren B, et al. Molecular analysis and phylogenetic characterization of HIV in Iran. J Med Virol. 2006; 78: 853-63. doi: 10.1002/jmv.20634. PMID: 16721846.

43) Soheilli ZS, Ataiee Z, Tootian S, Zadsar M, Amini S, Abadi K, et al. Presence of HIV-1 CRF35_AD in Iran. AIDS Res Hum Retroviruses. 2009; 25: 123-4. doi: 10.1089/aid.2008.0199. PMID: 19182925.

44) Eybpoosh S, Bahrampour A, Azadmanesh K, Mostafavi E, Haghdoost AA, Zolala F. Spatio-temporal dynamics of HIV-I subtype B and Circulating Recombinant Form 01_AE clades in Iran: a phylogeographic approach. Iran Red Crescent Med J. 2017; 19(2): e34609

45) Khoury MJ, Feero WG, Reyes M, Citrin T, Freedman A, Leonard D, et al. The genomic applications in practice and prevention network. Genet Med. 2009; 11(7): 488-94. doi: 10.1097/GIM.0b013e3181a551cc. PMID: 19471162, PMCID: PMC2743616.

46) Samzadeh M, Hasanzad M, Jamaldini SH, Haghdoost AA, Afshari M, Ziaee SA. Association of G/A polymorphism, rs266882, in AREI region of the prostate-specific antigen gene with prostate cancer risk and clinicopathological features. Urol J. 2012; 9(4): 691-9. PMID: 23235975.

47) Ziaei SA, Samzadeh M, Jamaldini SH, Afshari M, Haghdoost AA, Hasanzad M. Endothelial nitric oxide synthase Glu298Asp polymorphism as a risk factor for prostate cancer. Int J Biol Markers. 2013; 28(1): 43-8. doi: 10.5301/JBM.2012.9585. PMID: 23015399.

48) Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, et al. Association between a Literature-Based Genetic Risk Score and Cardiovascular Events in 19,313 Women. JAMA. 2010; 303 (7): 631-7. doi: 10.1001/jama.2010.119PMID: 20159871, PMCID: PMC2845522.

49) Muehlschlegel JD, Liu KY, Perry TE, Fox AA, Collard CD, Shernan SK, et al. Chromosome 9p21 Variant Predicts Mortality after CABG Surgery. Circulation. 2010; 122 (11 Suppl): S60-5. doi: 10.1161/CIRCULATIONAHA.109.924233. PMID: 20837927, PMCID: PMC2943860.