Molecular biology techniques for research development. A literature review

Técnicas de Biología Molecular en el desarrollo de la investigación. Revisión de la literatura

Maritza Angarita Merchán, María Inés Tórres Caycedo, Andrea Katherine Díaz Torres

Universidad de Boyacá. Tunja, Colombia.

How to cite this paper:

Received: December 2, 2016. Approved: September 20, 2017.

This paper is a translation to English of its original version, available on: http://www.revhabanera.sld.cu/index.php/rhab/article/view/1651/1867

ABSTRACT

Introduction: Epidemiology, from the etymological point of view, means "science that studies diseases that affect the communities". It has been developing through centuries, describing and explaining the dynamics of population health; it has integrated new branches such as molecular epidemiology, which is defined as a discipline in which molecular techniques are implemented for clinical, research, and scientific contributions.

Objective: To present techniques with basis in molecular biology, which have contributed to the research development.

Material and methods: A review of scientific articles was made during the months of August-October 2016, and July-September, 2017 in English, Portuguese, French, and Spanish languages. A search for information in scientific journals such as Pubmed, Scielo, Biomed Central, Free Medical Journals, LILACS, Redalyc, Inbiomed, and Dialnet was also made using DeCs term descriptors of Health Sciences, and the MeSH descriptor; articles published during the time period from 2012-2017 were used; and publications of previous years were also taken into consideration as a contribution to the
history of this topic.

**Results:** Five techniques of molecular biology which have contributed to research development were presented: PCR, sequencing, hybridization with DNA probes, RAPD, and RFLP.

**Conclusions:** At present, the use of molecular techniques allows the complete genome study or short and long DNA sequencings with the aim of detecting and analyzing sequences of interest for research in agronomy and forensic sciences; clinical diagnosis and basic, translational, and applied research; each of them characterized by reliability and quickness in obtaining result, strength, specificity, sensitivity, and flexibility, compared to phenotypic methods.

**Keywords:** Epidemiology, polymerase chain reaction, Hybridization, Sequencing, RAPD, RFLP.

**INTRODUCTION**

The optimization of specificity, sensitivity and rapidity of traditional diagnostic techniques has been necessary in the fight against infectious diseases; however, with the rise of research and
the need of efficient and opportune diagnoses, laboratory techniques have emerged with a basis on molecular biology applied to prevention, control, and treatment programs. Among the diagnostic alternatives proposed to face these challenges we describe some techniques such as polymerase chain reaction; DNA hybridization proves; genes sequencing; parallel sequencing, also known as next-generation sequencing (NGS); pyrosequencing; Random Amplification of Polymorphic DNA (RAPD); and Restriction Fragment Length Polymorphism (RFLP), whose introduction in laboratories is intended to support the obtaining of highly reliable results.

Polymerase chain reaction (PCR) has been the main diagnosis tool which has taken advantage of the goodness of molecular biology, getting the point of reaching a high versatility as an analysis technique. The specificity, efficiency, and accuracy of PCR is directly influenced by the different components which integrate it as the mixture of reaction, cycling regime, and DNA polymerase; the technique permits the selective amplification of any DNA segment, knowing the sequences that flank it, and obtaining a concrete DNA sequence without turning to cloning in a host organism. Its applications are variable and unlimited, for example, it gives the possibility of making genetic expressions, direct sequencing of amplified sequences, mutation detection, follow-up of the effectiveness in the treatment of diseases, diagnoses of genetic and infectious diseases; and in forensic sciences, it is used in the identification of biological remainders, paternity determination, and expert evidences in criminalistics.

DNA sequencing consists on determining the order of A, C, G and T bases in a DNA fragment; this method was described by Sanger in 1977, and permits to obtain the sequence of a determined DNA fragment, a gen or a part of this one, to be used at present. This method has evolved through the time; nowadays different kinds of sequencings have been implemented, standing out the New-Generation Sequencing (NGS) that permits the exploration of complete genomes in humans and other species; and Pyrosequencing, in which it is possible to determine the sequencing of a DNA molecule, identifying individual bases or short nucleic acid sequences in certain positions. Hybridization is a method that is based on the assembling of two nucleic acid chains that produce a stranded structure, which are DNA, RNA-RNA or DNA- RNA hybrids. Hybridization is based on the development of two nucleic acid molecules: a homogenous one with a distinguished sequence as a sounding line and a heterogeneous one with an unknown sequence, which contains the target sequence to analyze. Single-stranded nucleic acids come from cloned DNA that is fragmented by restriction enzymes, or synthetic oligonucleotides. There are other molecular techniques that have contributed significantly to research such as RAPD markers (Random Amplification of Polymorphic DNA), which are based on PCR; due to their existence, the detection of existent polymorphisms in the DNA sequence under study and RFLP (Restriction fragment length polymorphisms) is possible, which express the differences among individuals in specific DNA sequences that are recognized by different enzymes that cut those sequences and give rise
to little fragments which can be analyzed through electrophoresis.\textsuperscript{11}

**OBJECTIVE**
The objective of this review article is to present techniques based on molecular biology, which have contributed to research development in different fields of application.

**MATERIAL AND METHODS**
A review of scientific articles was made during the months of August-October 2016, and July-September, 2017 in English, Portuguese, French, and Spanish languages. A search for information in Ibero-American scientific journals indexed on Pubmed, Scielo, Biomed Central, Free Medical Journals, LILACS, Redalyc, Inbiomed, and Dialnet, was made using DeCs term descriptors of Health Sciences, and MeSH descriptor for the validation of key words.
The articles published during the time period from 2012-2017 were the most used, although it was necessary to use publications of previous years as a contribution to the history of this topic.
A total of 80 articles were obtained, to which inclusion and exclusion criteria consistent in validity and contribution to the topic under analysis were applied; 56 of them were finally selected. A database was made from these data, from which a bibliometric analysis was developed for their classification according to subjects of interest, authors and publication date. It is important to highlight that the main limitation of this review was that the access to some of these issues do not allow a free consultation of their texts.

**CASE PRESENTATION**

**Molecular Epidemiology**
Molecular epidemiology is a branch of the discipline applied to the study of infectious diseases in which molecular techniques used for the identification of pathogenic agents in epidemiological studies are implemented. Its objective is to describe the proliferation of the disease and its risk factors, in order to intervene in the course of its natural development.\textsuperscript{12} It is based on statistical analyses through geographic methods, which permit to evaluate the development of the affectedness,\textsuperscript{13} detect and quantify specific genetic material coming from biological samples, outbreak studies, characterization of microorganisms, relations existing among genotypes and virulence factors studies.\textsuperscript{14}
Molecular diagnosis is a dynamic area in constant development that has revolutionized the clinical diagnosis, demonstrating an impact in health areas,\textsuperscript{15} and obliging to the implementation of key tools for clinical equipment, which generate a direct benefit for the patient.\textsuperscript{16}
The principle of molecular epidemiology is based on the study of infectious diseases through the use of molecular techniques that allow the study of the genome of bacteria, viruses, viroids, fungi
and parasites, that are etiological agents of those diseases.17

**Applications**

Molecular epidemiology is used as a diagnostic method for different pathologies; its main application is found in:

1. **Molecular method for typification:**
   Typification is known as the identification and characterization of pathogenic microorganisms that permit to establish the identity of microorganisms which cause infectious outbreaks, determining the infectious source and its possible patterns of dissemination; likewise, it establishes the infectious agent prevalence in a population.18

   The typification technique to use will depend on the requirements and characteristics of the system analyzed; however, regardless the typification method, it must be previously evaluated according to its capacity to generate the required epidemiological information.

   Typification can be evaluated taking into account the following criteria:
   - Detection, identification and typification of the whole analyzed isolates.
   - Repeatability and reproducibility of the method.
   - Genetic stability of the marker, neutral for evolutionary forces.
   - Exclusion of the different groups of individuals with high probabilities.
   - Capacity of the method to generate similar results to the obtained through other techniques.
   - Effectiveness among economical costs generated by the application of the method and the profits obtained to reach the prevention and control of the disease.

2. **Phenotypic and genotypic molecular methods:**
   Phenotypic methods are based on the determination of biochemical and/or physiological characteristics; they constitute the first tool for the comparison of microorganisms that includes the determination of enzymatic activities, metabolic capacity and susceptibility to antigenic determinants or bactericidal agents; however, with these kind of methods; genes, polymorphism, or mutations that determine the expression of the visible characteristics of culture mediums, and biochemical and susceptibility tests cannot be identified.20,22

   Genotypic methods study the genome of the microorganism that causes the disease and make possible the analysis of the characteristics of genetic polymorphism presented in etiological agents.23 They are based on the location of the genetic material, which allows to generate new changes in the genetic expression pattern and offers more stable and reproducible alternatives.24

   **Among the employed techniques in genotyping, the following ones are described:**
   1. Polymerase chain reaction (PCR).
   2. Genome sequencing.
      a. NGS sequencing.
      b. Pyrosequencing.
   3. DNA hybridization proves.
   4. RAPD.
   5. RFLP.

   Each technique has offered an alternative for epidemiological research; however, they also have limited applications.25

   **1. Polymerase chain reaction (PCR)**

   In recent years, new molecular techniques based
on PCR have been developed, which have brought a great advance in the evolution of research on infectious diseases through molecular epidemiologic studies, which have the objective to determine the existent clonal relation among several isolates of a same species, through typification techniques that involve genes amplification or polymorphic DNA sequences. Genotypic methods amplify specific DNA in vitro areas by employing sequences that delimit the amplification area; from a copy of the area to the one to amplify, millions of copies that make their detection possible and reflect the presence of the DNA area in the sample to analyze are acquired; in this transformation, there are several proteins that cooperate with the synthesis of new DNA strands, from another one which acts as a mold. This technique has had different advances and applications, which are presented in the chart.

**Chart.** Characteristics and applications of the different types of PCR

<table>
<thead>
<tr>
<th>TYPE DE PCR</th>
<th>CHARACTERISTICS</th>
<th>APPLICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard PCR</td>
<td>Amplification of a DNA segment by using two timings. The detection of the amplification is throughout agarose gel or polyacrylamide used for observations of small regions in number of pair bases.</td>
<td>Qualitative detection of a DNA segment.</td>
</tr>
<tr>
<td>Nested PCR</td>
<td>In this variant the product of amplification is used as a mold for making a second amplification with starters that are going to be placed in the first amplified sequence.</td>
<td>Qualitative detection of a DNA segment. Highly sensitive and specific</td>
</tr>
<tr>
<td>In situ PCR</td>
<td>The products generated from biological samples as secretions and tissues, can be visualized in the place of amplification, allowing the detection of small quantities of genetic material.</td>
<td></td>
</tr>
<tr>
<td>Multiple PCR</td>
<td>Amplification of two or more DNA segments using several primers in just a reaction of amplification. The detection of the amplification is visualized throughout agarose gel.</td>
<td>Qualitative detection of several DNA segments in just a PCR reaction.</td>
</tr>
<tr>
<td>RFLP- PCR (restriction fragment length polymorphism)</td>
<td>Standard PCR with subsequent step with restriction enzymes digestion.</td>
<td>Detection of RFLPs polymorphisms</td>
</tr>
<tr>
<td>Reverse transcription PCR (RT- PCR)</td>
<td>This kind of PCR use RNA as a main mold and a reverse transcriptase is required for the conversion of RNA to a</td>
<td>Genes expression. Detection of RNA virus</td>
</tr>
</tbody>
</table>

26 Genotypic methods amplify specific DNA in vitro areas by employing sequences that delimit the amplification area; from a copy of the area to the one to amplify, millions of copies that make their detection possible and reflect the presence of the DNA area in the sample to analyze are acquired; in this transformation, there are several proteins that cooperate with the synthesis of new DNA strands, from another one which acts as a mold.27,28 This technique has had different advances and applications, which are presented in the chart.29,30
type of DNA called cDNA (complementary DNA).  

<table>
<thead>
<tr>
<th>RT-PCR (Real time) or qRCP</th>
<th>Standard PCR where stains or proves with fluorophores are used for the detection of the amplified fragments; they can present type multiple.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Qualitative detection of one or several DNA segments. DNA quantification in the sample (charges) or genes expression.</td>
</tr>
</tbody>
</table>

**Source**: Adapted from [Biología molecular aplicada al diagnóstico clínico - BO. Mauricio J. Farfán, PhD](37)

### 2. Genome sequencing

It is a technique that determines the whole DNA sequence in the genome of a person; it consists on determining the order of the Adenine, Guanine, Cytosine and Thymine bases in a DNA fragment. With this technique, approximately 500 bases can be obtained; they are assembled to a genome of reference that sequences a whole genome. This method has changed the way to understand Genetics basing on the identification of the real causes of heredity, focusing on genetic studies of individuals with a defined phenotype and mendelian inheritance diseases produced by known genes; they evaluate the phenotype and sequencing of the gen that can be affected and present a very high sensitivity for detecting mutation.  

One of the most famous projects in the history of molecular biology was the Human Genome Project (HGP), which proposed to determine the complete sequencing (more than 3 000x10⁶ base pairs) of the human genome; the project locates exactly the 100 000 DNA genes approximately and the rest of the heredity material of human beings, responsible for genetic instructions of everything that forms a human being from the biological point of view.

Through time, sequencing has experienced several modifications to the method described at first by Sanger, and generated other kinds of sequencing as NGS and pyrosequencing, constantly employed in clinical research and epidemiological studies nowadays; the methods previously mentioned are presented below.

a. Next-generation Sequencing (NGS)

Nucleic acid sequencing permits to establish the order of nucleotides presented in DNA or RNA molecules to study; that is why, its use has increased exponentially in research and clinical laboratories around the world in recent years; NGS has been implemented because of the possibility that it offers when making massive and parallel sequencing of millions of DNA and RNA fragments presented in the sample, with the use of leading technology, at a low cost and a very high efficiency, which made the amplification of a complete genome possible in only one day.

NGS has a high application in epidemiological studies due to the advantages offered by this kind of method as the use of complete genomes for establishing phylogenetic relations among species, identification of possible epidemiological combinations and markers that contribute to the identification of possible mutations in a population; therefore, it is considered as a revolutionary technology in epidemiological studies applied to basic science, translational research, clinical diagnosis,
agronomy, forensic science, and applied science. This kind of sequence is also known as Non-Sanger, and it is available in different formats that permit the generation of data with advantages and disadvantages inherent to each matrix; among the advantages, the quality of the data obtained from the sequences, the robustness, and low noise present in the chromatogram are highlighted; as disadvantages, the availability of a laboratory with bioinformatics capacity that guarantees the quality in obtaining and interpretation of the data, as well as the need of making a control over the random and unspecified sequences that can interfere with the sequencing have been reported.

b. Pyrosequencing
It is characterized by the DNA sequencing synthesis with a detection in real time; this technique is used for the identification of individual bases or short nucleic acids sequences in certain positions, through the usage of phosphate during the incorporation of nucleotides to the DNA chain, followed by a series of enzymatic reactions. This is the only sequencing method that was implemented as an alternative to the classic DNA sequencing; if it is compared with other molecular techniques, pyrosequencing is simple, robust, rapid, sensitive, highly quantitative and accurate, flexible, effective and has the capacity of automation of the sample; it has been employed in studies of genetic variations analysis, agronomical studies that permit the implementation of priming and specific proves that contribute to certification programs of food quality, changes in microbial communities of different environments, resolution in forensic sciences cases, microbiology, and the detection of mutations in pathologies of clinical interest.

3. DNA hybridization proves
It is known as the analysis in samples to detect the presence of nucleic acids (DNA or RNA), by making an antiparallel combination of them with a double-chain molecule. Its techniques are used to detect a target molecule starting by a complementary probe. Many molecular techniques are based on hybridization as PCR; they are used in the diagnosis of diseases, the identification of pathogenic microorganisms, studies of genetic expression profiles, localization of genes in chromosomes or ARNm in tissues, and the comparison of pathogenic species.

d. RAPD (Random amplification of polymorphic DNA)
This is a technique which employs molecular markers for PCR amplification of short polymorphic DNA sequencings by using a primer of short sequence (10 to 12 base pairs –pb). As it is a technique based on PCR, it needs control over certain factors that can affect directly the technique performance as dNTPs, TaqDNA polymerase, hybridization temperature, extension, cycles and mold chain integrity. On research, this kind of technique, is applied in genetic analyses that permit the establishment of similarities among communities of the same species (example: bacteria and plants); an example of this is the study of the relation between the resistance to arsenic in bacterial flora coming from ground samples, which was made in India and was published in Molecular Phylogenetics and Evolution in 2016.
e. RFLP (restriction fragment length polymorphism)
This is the result of a variation of a DNA sequencing recognized by restriction enzymes used for cutting DNA sequences in unknown places; they are used principally as markers in genetic maps. This is a due to its rapidity in obtaining the results, low price, and specificity; this needs certain conditions for its functioning, consistent on the use of appropriate restriction enzymes, conditions of amplification and optimization, and analyses of the amplified products (restriction fragments) through electrophoresis, mainly in agarose gel.\textsuperscript{52-54}
Among the advantages described, it is found that, it needs minimum items of laboratory for its performance; it has been applied in diverse studies that have allowed to establish or identify bacterial species inherent to human being and animals (example: melitensis biovare),\textsuperscript{55} the discrimination between pathogenic species of different microorganisms that cause infections in human beings or which are presented in some human consumption products and metagenomic studies.\textsuperscript{56}

CONCLUSIONS
The techniques employed in research with molecular foundation, have allowed a significant progress on research, contributing mainly to the molecular epidemiology development as a science applied for the knowledge of genotypic of bacterial communities in different environmental, veterinarian and human areas; generating knowledge about epidemiological performance and changes that populations, specially bacterial, have developed as defense and adaptation mechanisms to their habitat conditions. Nowadays, the use of molecular techniques as NGS, pyrosequence, RAPD and RFLP, allow the study of a complete genome or specific long or short DNA sequencings in order to detect and analyze sequences of interests for research in agronomic and forensic sciences, clinical diagnosis, and basic, translational and applied research.
Each method presented in this review is characterized for the reliability and rapidity in the obtaining of the results, strength, specificity, sensitivity and flexibility, when compared to phenotypic methods; being this, a direct and accessible contribution to the development of molecular epidemiology.

ACKNOWLEDGEMENTS
To the Bacteriology and Clinical Laboratory Program at the University of Boyacá.

REFERENCES
2. Ranjbar R, et al. "Typing methods used in the
11. Rasmussen HB. Restriction fragment length polymorphism analysis of RCP-amplified fragments (RCP-RFLP) and gel electrophoresis-valuable tool for genotyping and genetic fingerprinting. Chapter from the book Gel Electrophoresis - Principles and Basic. 2012.
23. Tamay de Dios L, Ibarra C, and Velasquillo C. "Fundamentos de la reacción en cadena de la
polimerasa (RCP) y de la RCP en tiempo real”. Investigación en discapacidad .2013; 2.2: 70-78.

Maritza Angarita Merchán. Universidad de Bogotá. Tunja, Colombia.
E-mail: mangarita@uniboyaca.edu.co