VOLUME 4, ISSUE 4

FOCUS on Field Epidemiology

Laboratory Diagnosis: Molecular Techniques

CONTRIBUTORS

Authors:

Amy Nelson, PhD, MPH

Lauren N. Bradley, MHS

FOCUS Workgroup*

Reviewers:

FOCUS Workgroup*

Production Editors:

Tara P. Rybka, MPH

Lorraine Alexander, DrPH

Rachel A. Wilfert, MD, MPH

Editor in chief:

Pia D.M. MacDonald, PhD, MPH

*** All members of the FOCUS Workgroup are named on the last page of this issue.**

UNC SCHOOL OF PUBLIC HEALTH

NORTH CAROLINA CENTER FOR PUBLIC HEALTH PREPAREDNESS

The North Carolina Center for Public Health Preparedness is funded by Grant/Cooperative Agreement Number U90/CCU424255 from the Centers for Disease Control and Prevention. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the views of the CDC.

Have you ever sat down with microbiologist colleagues for what you thought would be pleasant chit-chat, only to find yourself completely lost? You start to comment on the weather when a co-worker asks, "So, did you get the PFGE done on the MRSA?" Another coworker says, "No, I'm having problems with the DNA. Has anyone done RFLP on MRSA?" The first coworker replies, "Hmm, not that I know of, but maybe PCR would work…" And on they continue, until you don't know whether your RFLP is in the PCR or your PFGE got lost on the way to the MRSA.

When you start thinking nonsense like this, you definitely have the molecular biology blues. Luckily FOCUS can lift your spirits by providing a review of molecular diagnostic tests. Then the next time you sit next to a microbiologist on your coffee break, you might even have a suggestion or two.

As noted in the last issue of FOCUS, laboratory techniques such as culture, microscopy, serology, phagetyping, and molecular methods can be used both to verify the presence of an organism and to identify that organism. Molecular techniques involving DNA or RNA are particularly powerful tools for the laboratory professional.

In this issue of FOCUS, we look at these molecular techniques in detail and discuss those most commonly used in public health: PCR, PFGE, and ribotyping.

What is DNA?

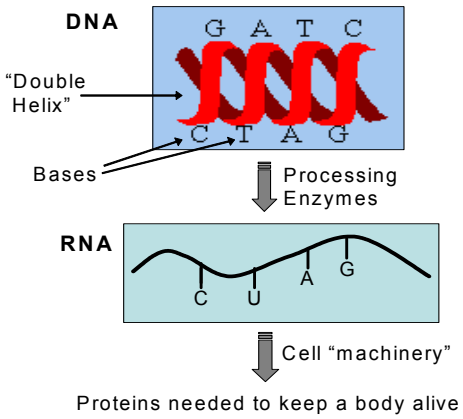
You've heard it before: *the double helix*. DNA (deoxyribonucleic acid) is the twisty, ladder-like molecule that is the genetic material present in every bacterium, plant, and animal. DNA is the code used to build all the molecules that make up a living being. Some viruses also have DNA, though others have RNA as their genetic material.

DNA is composed of a long string of four special molecular units—adenine (A), thymine (T), cytosine (C), and guanine (G)—called **bases.** Each base is linked with a partner (A with T, C with G), like two sides of a ladder linked by the rung. Together, they are known as base-pairs. The bases are arranged in an exact order called a **sequence**, such as AATTCGCG or CATAGCGTA.

This pattern of A's, T's, C's, and G's is like a recipe for the protein that will be created by that particular piece of DNA. DNA also codes for RNA, but in RNA, thymine (T) is replaced by uracil (U). To replicate DNA or create proteins, the two sides of the DNA ladder separate from each other, and new bases pair up with the existing sequence. RNA is used in living cells as the copy messenger to DNA. From the DNA template, the cell makes a copy of RNA. RNA then runs all over the cell, carrying the code to create and maintain the living being to the cell's building machinery (see Figure 1 on page 2).

More in-depth information on DNA is available from the additional resources listed on page 4.

Figure 1. DNA being converted to RNA in a living cell



Why is DNA Useful in Epidemiology?

From an epidemiologist's point of view, the useful thing about DNA sequences is that it can be used to identify an organism causing a disease outbreak. Certain DNA sequences are unique to each organism.

In a case of gastrointestinal illness, a stool sample might be tested for the presence of DNA from several different organisms. If the DNA of a particular organism is detected, that organism may be the cause of the illness.

Also , examination of the right sections of DNA can allow two different strains of the same species of microbe to be distinguished from each other. Some sequences will be exactly the same among strains of the same species, while other sequences will have areas of variation that can be used to distinguish one strain from another.

This property of DNA sequences is useful in determining whether different cases of the same disease are actually part of an outbreak.

For example, if Norovirus is identified in two cases of gastrointestinal illness, they may or may not be part of the same outbreak. If you determine that these are different strains of Norovirus, you know the cases are not related.

If the cases have the same strain, they might have acquired the infection from the same source, or one case might have transmitted the infection to the other.

Using molecular techniques such as polymerase chain reaction (PCR) to examine DNA sequences can help to identify what strain of a pathogen is present in a specimen.

Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a laboratory technique that makes multiple copies of a piece of DNA or RNA in a process called amplification. Amplification makes it easier to detect the tiny strands of an organism's DNA by generating numerous copies of the DNA to work with during analysis. Also, PCR can start with very small amounts of DNA, which is an advantage if the specimen contains a limited number of organisms from which to obtain DNA or if the organism cannot be cultured. PCR can be used on either viruses or bacteria.

PCR starts with a sample of the DNA of interest, such as that from a clinical specimen suspected to contain a pathogen. Then a primer is added to the sample. A primer is a very short sequence of DNA that will seek out and bind to a specific sequence of the target DNA. The primer is the key element: it can be designed to be very specific, for example to "match" echovirus 30, or it can be designed to be more general, for example to match any echovirus.

Other materials added to the mixture include a polymerase enzyme that will "read" a DNA sequence and create copies, and DNA base "building blocks" that can used as raw material to make the copies. The polymerase enzyme will make copies only of the DNA that matches the primer. Then, if the DNA has been amplified, we know that the DNA in the specimen matched the primer used. If the DNA fails to amplify, the particular bacterium that the primer was designed to match was not present in the specimen. Further tests must then be conducted to determine whether different bacteria were present.

So if you believe *Salmonella* is causing an outbreak of diarrheal illness, you would amplify a gene that is unique to *Salmonella.* After the PCR reaction, you would use the genes amplified by PCR to confirm that the organism is indeed *Salmonella.* As with all detection and identification techniques that require a sample of the organism, proper sample collection, shipment, and storage are essential (see *FOCUS* Volume 4, Issue 2 for more information on specimen collection). If the organism does not survive the trip from the patient or outbreak setting to the laboratory, it may be difficult or impossible to identify.

DNA Fingerprinting

If you are still unsure what the infecting organism might be after performing PCR on the specimen, you probably ran a non-specific PCR reaction (that is, you amplified whatever genetic material was present). With the supply of genetic material obtained after amplification, your next step is to sequence the DNA.

Figure 2. Comparison of the DNA sequences of a nucleoprotein gene in infections of two patients with different strains of rabies

A. Gene sequence AY138566; rabies virus isolate 1360, India
B. Gene sequence AY138567; rabies virus isolate 945, Kenya

```
Line 1a gaaaaagaac ttcaagaata tgagacggca
Line 1b gaggaagaac ttcaagaata cgagacggct

Line 2a gaattgacaa agactgacgt agcgctggca
Line 2b gaactgacaa agactgacgt ggcattggca

Line 3a gatgatggaa ctgtcaattc ggatgacgag
Line 3b gatgatggaa ctgtcaactc tgacgatgag
```

Differences between the sequences are shown in boxes. The exact sequence of genes in a microorganism can be used to identify that organism or strain.

Full sequence available from query at:bhttp://www.ncbi.nlm.nih.gov/entrez/query.fcgi

You can determine the specific order of the bases in the DNA strand(s) that you amplified. This particular sequence of A's, T's, G's and C's can then be compared to known sequences of an organism or strain to determine which organism's DNA matches the sequence you obtained. For an example of comparison of DNA sequences, see Figure 2.

It is also possible that the DNA sequence amplified will be that of a known gene from a specific organism. If the laboratory suspects *Salmonella* and runs the experiment to amplify the DNA of a *Salmonella* gene, this gene will be amplified if *Salmonella* was the infecting organism. The gene will not amplify if *Salmonella* was NOT the infecting organism. After PCR amplification, the laboratory technician will run the PCR product on a special gel that helps to visualize the DNA (more on gels below). Since the gene we are interested in is known, we know how many base pairs it is supposed to have (i.e. how big the sequence is). Once we see our sample of DNA on a gel, we can determine whether the gene is present and whether it has the correct length segment. If so, the organism is the expected organism (see Figure 3).

The DNA obtained through PCR can also be further processed to identify its DNA fingerprint, a pattern on a gel that will identify the organism (more on this below). DNA fingerprinting is generally done when a specific organism is suspected, in order to determine which strain of the organism is present.

- For example, tuberculosis (TB) has clear symptoms, but DNA fingerprinting might be used to determine whether different cases of TB are infected with the same strain, possibly due to an outbreak or to a common exposure.

How Do Gels Work?

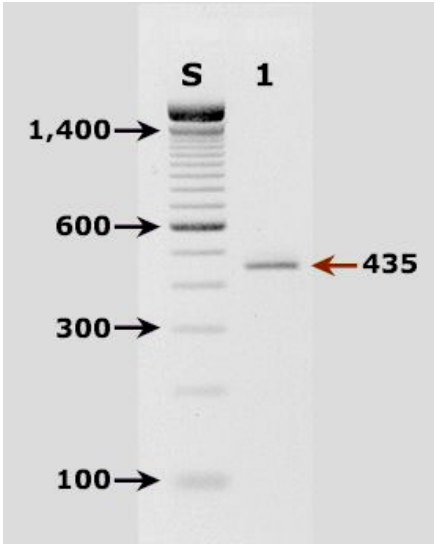
The PCR product is placed in a lane at one end of the gel, much like a runner getting ready for his race around the track. If the organism is unknown, the PCR product might first be divided into pieces by a special enzyme that cuts the DNA wherever there is a certain sequence. A small electric field is then applied to the gel, which causes the DNA to migrate through the gel from one end of the gel to the other.

The distance traveled by the DNA depends on the sequence of the DNA and the length of the piece(s) of DNA. DNA bases A, T, C and G have natural electrical charges that determine the speed and direction at which a DNA fragment can move when an electrical field is applied to the gel. Negatively charged particles (A and T) tend to move more quickly, while positively charged particles (C and G) move more slowly through the gel. Additionally, each piece of DNA is unique in size, depending on where the DNA-cutting enzyme divided it into pieces. Smaller pieces of DNA will move through the gel at a faster rate than larger pieces.

After a defined time period, the electric field is turned off, freezing the DNA race so that the scientist can examine the pattern of DNA in the gel. Special techniques are

Figure 3. Picture of a PCR gel for diagnosing *Cryptosporidium parvum* from a fecal sample

Each dark band represents many strands of DNA that are the same length. The lane marked "S" is a DNA ladder; like a mile-marker for DNA, each band shows DNA strands with a specific number of base pairs (marked on the side) that can be used to measure the length of DNA amplified in the PCR reaction. In this case, the 435 base pair band from *C. parvum* is a positive identification. (1)



used to look at the clusters of DNA, which appear as solid bands in the gel (see Figure 3). When this process is conducted on two different organisms, it will result in very different DNA patterns, just as fingerprints from two people look very different to the trained eye. If samples of an organism taken from two patients have exactly the same DNA pattern, these people were both infected with the same organism, which lends support to the possibility of an outbreak.

Pulsed Field Gel Electrophoresis (PFGE)

DNA can also be detected by **pulsed field gel electrophoresis (PFGE)**, which is used for the analysis of large DNA fragments. PFGE is advantageous because it requires less processing and sample preparation of the DNA. To perform PFGE, special enzymes can be used to cut the DNA into a few rather long pieces. Instead of applying an electrical field such that the DNA fragments race straight to the end, after the electrical field is applied, the direction is changed, and then changed back, and then changed again.

This is like a race composed entirely of large, slow-moving runners. At the start, they take up so much room and are so slow that it can be difficult to tell them apart; they appear to be just a mass of runners. Once they start running, the finish line is moved from directly ahead of them to a spot 400 yards to the left. Once all the runners manage to turn themselves around and head toward the new goal, the finish line is switched back to its original spot. All the runners turn and head toward the finish line again.

Switching directions separates the runners (the DNA pieces) into two planes and spreads out the DNA more distinctly. The electric field is often applied in a hexagonal pattern, so the field is alternated in six different directions.

PFGE is used to identify bacteria, but not viruses. The DNA used for PFGE analyses can be extracted from a microorganism in culture, from a clinical specimen, or from an environmental specimen. Like regular gels, PFGE can be used to identify an organism or to distinguish between strains of the same organism to determine whether several cases of disease are related to each other (identical strains) or not (different strains). However, the turnaround time for PFGE is longer than the time needed for running a regular gel.

- For example: Let's go back to an outbreak mentioned in the last issue of *FOCUS* involving *Escherichia coli* O157:H7 infections among Colorado residents in June 2002. (2) In this outbreak the case definition required that *E. coli* be cultured from the patient**.** In addition, the case definition required that all cultures exhibit the same PFGE pattern. Molecular techniques were used to fine-tune a case definition. The Colorado investigators did not want to include in the outbreak every case of *E. coli* found in the city, only the cases that were the exact strain of interest.

PFGE patterns are often used in this manner to link cases together in an outbreak. While PFGE is not successful in fingerprinting every bacterial organism, it can be used to fingerprint a wide variety of pathogens, such as *E. coli, Heliobacter pylori, Staphylococcus aureus, Legionella, Psuedomonas*, and others.

PulseNet

In partnership with state health departments and the Association of Public Health Laboratories (APHL), the CDC is creating a molecular subtyping network called PulseNet, which is based on DNA fingerprinting of bacteria that cause foodborne diseases.

The project is being implemented first for *Escherichia coli* O157:H7 and will then be expanded to include *Salmonella* typhimurium and other foodborne pathogens. All participants will use standardized equipment and protocols, and a centralized database of DNA patterns ("DNA fingerprints") will be stored on a computer server at the CDC.

Through the participation of the U.S. Department of Agriculture and the Food and Drug Administration, the database will include fingerprints derived from contaminated foods as well as from clinical isolates.

Website: http://www.cdc.gov/pulsenet/

Resources:

The Columbia Electronic Encyclopedia: suggested searches on nucleic acid and DNA fingerprinting.
Available at: http://education.yahoo.com/reference/encyclopedia/

Source Molecular Company: an interesting application of molecular diagnostic techniques to water quality.
Available at: http://www.sourcemolecular.com/ribotyping.htm

Ribotyping

Ribotyping is another molecular diagnostic technique. Its name is derived from the **ribosome**, which is part of the cellular machinery that creates proteins. Ribosomes are found only in cells, so ribotyping is a method of identifying bacteria, not viruses. (Viruses are molecules with genetic material and protein only. They do not have a cellular structure.)

A ribosome is composed of RNA that is folded up on itself in a particular way. This RNA is referred to as "rRNA" for *ribosomal* RNA. We noted at the beginning of this issue that DNA codes for RNA. Since living cells in everything from lizards to people create proteins, the DNA genes that code for rRNA have much in common, even across vastly different species. However, some parts of the genes that code for rRNA are highly variable. That is, certain sequences are quite different from one species to the next, or even from one strain of bacteria to the next. These variable regions can be used to identify a particular strain type of bacteria.

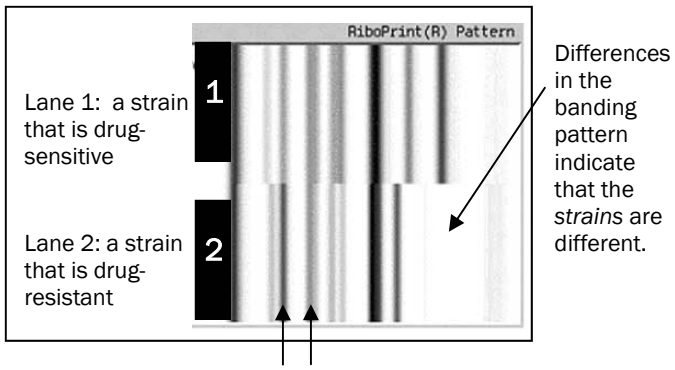
How are these variable regions determined? As in other methods of DNA analysis, DNA-cutter enzymes are used. These enzymes divide the RNA only when a specific sequence occurs. Thus, if a strain of bacteria has that sequence in its rRNA, the rRNA will be cut at that location. If another strain of bacteria has a few different bases in the same spot, the rRNA will not be cut. The rRNA is then run on a gel so the number and size of the pieces can be seen (see Figure 4). rRNA that has been cut in the expected locations will appear different from rRNA that was not cut.

The advantage of ribotyping as a method of identifying infectious microorganisms is that the procedure is fully automated. Not only is less labor involved in performing the procedure, but the procedure is standardized. However, because of the equipment needed, ribotyping is rather expensive, and is usually only performed in reference laboratories. Ribotyping is most commonly used for typing strains of *Staphylococcus aureus*, but it can also be used for typing other species of *Staphylococcus* and for *E. coli*.

In this issue of *FOCUS*, we have discussed molecular techniques, that is, laboratory analyses that use DNA or RNA. These techniques can be used to identify pathogens in a sample, or to determine what strain of a particular pathogen is causing infection.

A future issue of *FOCUS* will walk you through the uses of laboratory diagnostics in an outbreak setting and provide examples obtained from real investigations.

Figure 4. A ribotype image showing two strains of *Salmonella* Newport (3)



Glossary:

Base—a molecular unit that forms the backbone of a DNA molecule; DNA has four bases: adenine (A), thymine (T), cytosine (C), and guanine (G)

DNA—any of various nucleic acids that are usually the molecular basis of heredity; constructed of a double helix held together by hydrogen bonds

DNA fingerprinting—a method of identification by determining the sequence of base pairs in the DNA of a person (or other creature)

Nucleic acids—biochemical macromolecules composed of nucleotide chains that convey genetic information

Ribosome—RNA-rich structure in the cell that is the site of protein creation

RNA—any of various nucleic acids that contain ribose and uracil as components; involved with the control of cellular chemical activities

Virus—any of a large group of sub-microscopic infective agents composed of a protein coat that surrounds an RNA or DNA core; capable of growth and multiplication only in living cells

CONTACT US:

The North Carolina Center for Public Health Preparedness
The University of North Carolina at Chapel Hill
Campus Box 8165
Chapel Hill, NC 27599-8165

Phone: 919-843-5561
Fax: 919-843-5563
Email: nccphp@unc.edu

FOCUS Workgroup:

- Lorraine Alexander, DrPH
- Meredith Anderson, MPH
- David Bergmire-Sweat, MPH
- Lauren N. Bradley, MHS
- Anjum Hajat, MPH
- Pia D.M. MacDonald, PhD, MPH
- Gloria C. Mejia, DDS, MPH
- Amy Nelson, PhD, MPH
- Tara P. Rybka, MPH
- Rachel A. Wilfert, MD, MPH

REFERENCES:

1. Johnson DW, Pieniazek NJ, Griffin DW, Misener L, Rose JB. Development of a PCR protocol for sensitive detection of *Cryptosporidium* oocysts in water samples. *Appl Environ Microbiol*. 1995;61:3849-3855.
2. Centers for Disease Control and Prevention. Multistate outbreak of *Escherichia coli* O157:H7 infections associated with eating ground beef --- United States, June–July 2002. *MMWR Morb Mort Wkly Rep*. 2002;51:637-639. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5129a1.htm. Accessed November 30, 2006.
3. Fontana J, Stout A, Bolstorff B, Timperi R. Automated ribotyping and pulsed field electrophoresis for rapid identification of multidrug-resistant *Salmonellas* Serotype Newport. *Emerg Infect Dis* [serial online]. 2003;9:496-499. Available at: http://www.cdc.gov/ncidod/EID/vol9no4/02-0423.htm. Accessed December 14, 2006.

If you would like to receive electronic copies of FOCUS on Field Epidemiology, please fill out the form below:

- NAME: _____
- DEGREE (S): _____
- AFFILIATION: _____
- E-MAIL ADDRESS: _____
- May we e-mail any of your colleagues? If so, please include their e-mail addresses here:

  _____

  _____

  _____

Please fax to: (919) 919-843-5563

or mail to: North Carolina Center for Public Health Preparedness
The University of North Carolina at Chapel Hill
Campus Box 8165
Chapel Hill, NC 27599-8165

Or go online: http://www.sph.unc.edu/nccphp/focus/

UPCOMING TOPICS!

- **Laboratory Diagnosis in Outbreak Investigations**
- **Contact Tracing**
- **Biosafety Levels**

We are on the web!
http://www.sph.unc.edu/nccphp