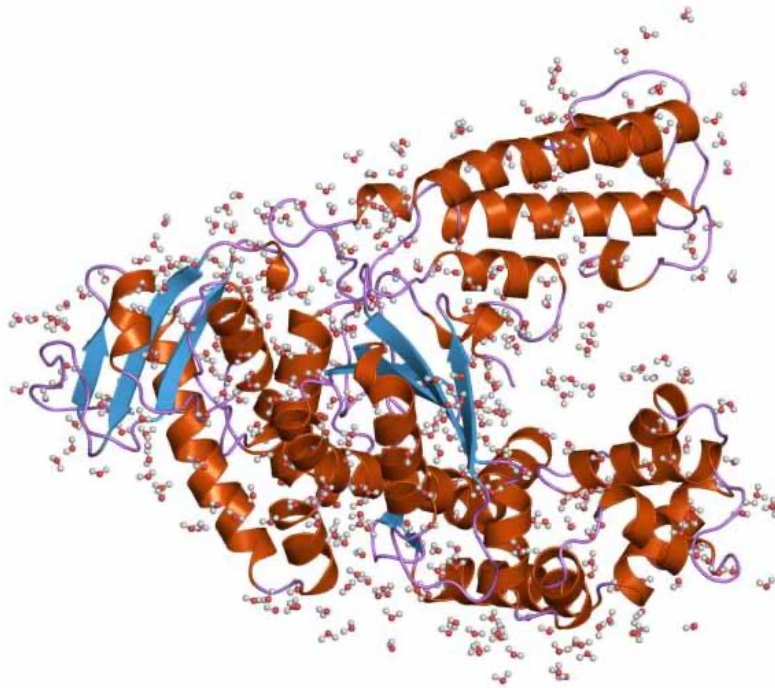


Portrait Gallery of Humanity-Changing Benefits of Microbes

Inspired by Nature – the Polymerase Chain Reaction

My friend at school has COVID. Do I have it, too?



Molecular scale structure of the DNA polymerase enzyme from the thermophilic bacterium *Thermus aquaticus* (Taq), isolated from Mushroom Pool, Yellowstone National Park, United States of America. Taq polymerase, as it is now termed, forms the heart of the polymerase chain reaction that has saved millions of lives and tremendously benefitted humanity. Credit Jawahar Swaminathan and staff at the European Bioinformatics Institute.

Eric S. Boyd and Rachel L. Spietz

Department of Microbiology and Cell Biology, Montana State University, Bozeman, Montana, USA

A child-centric microbiology education framework

Storyline

Imagine going on your next family vacation and making a discovery that would forever change our world. That is what happened when Drs. Thomas and Louise Brock stopped by Yellowstone National Park, U.S.A., and observed the hot springs and their colorful microbial inhabitants during a vacation in summer 1964. Little did the Brocks and their colleagues know at the time, but their successful cultivation of *Thermus aquaticus* from Mushroom Pool in Yellowstone was about to change humanity. Not only was *T. aquaticus* one of the first thermophilic, or heat-loving, organisms grown in the laboratory, it was also the source of the final ingredient necessary for the successful development of the polymerase chain reaction (PCR), which is now a mainstay method used in the scientific laboratory. As discussed below, the PCR has a variety of uses that continue to drive life-changing scientific discoveries, advance forensic science, and promote the development of vaccines, antibiotics, and other medicines.



Mushroom Pool, in the Lower Geyser Basin of Yellowstone National Park, as it looked on June 23, 1967. Thomas Brock, pictured near the edge of the pool, collected a sample from this spring that was the source of *Thermus aquaticus*, a heat-loving bacterium that contained the essential ingredient for the polymerase chain reaction, or PCR. Image from Thomas Brock's self-published "A Scientist in Yellowstone National Park". Importantly, approaching hot springs in Yellowstone National Park without a government-issued research permit is not only extremely dangerous it is also illegal. All visitors to the thermal features in Yellowstone must stay on a marked trail or a boardwalk or risk significant fines, injury, or even death.

Before we discuss the fundamentals of the PCR and how it changed our world, let's take a moment and review a few key biological principles that inspired and allowed for the

A child-centric microbiology education framework

development of the PCR. All living cells contain DNA, referred to as a cell's genome. A genome is essentially an instruction manual that tells a cell what to do and how to do it. This genomic instructional manual is made up of thousands of smaller units, called genes. Each gene holds information for different traits or characteristics, such as hair color. The genome from one type of organism is different from the genome of another type of organism, and these differences are responsible for variations in how organisms look and function. Each species of organism has their own unique genomic DNA sequence that can be used to differentiate it from another species. Even within the same species of organisms, the sequence of DNA can differ slightly. For example, each of us has genomic DNA with a sequence that varies slightly from others, and this is the reason that we each have a unique set of characteristics, including our height and the colors of our eyes. Thus, genomic DNA is like a fingerprint – a unique characteristic that defines an organism.

The Problem

Despite the advantage of using DNA to identify different species or even individuals within the same species, there is an inherent problem with this approach: the amount of DNA in a cell is really, really small. An average microbial cell only contains several femtograms of DNA and a human cell only contains several picograms of DNA. For perspective, a small paperclip weighs close to a gram. Thus, we would need to take the DNA from about a quadrillion (1,000,000,000,000,000 or 10^{15}) microbial cells or about a trillion (1,000,000,000,000 or 10^{12}) human cells (roughly all the cells in an average human body) to generate a gram of DNA. While scientists do not necessarily need a gram of DNA for most of their experiments, they do need a lot more than a femtogram or a picogram. As such, scientists need a way to increase – *to amplify* – the amount of DNA from a cell without changing the sequence of that cell's DNA.

In the mid-1980s, several groups of scientists were working to develop a way of amplifying an organism's DNA. Most of these scientists were researching an enzyme that evolved to naturally copy the DNA of a cell before it divides and creates two daughter cells. This enzyme is called DNA polymerase and it is found in all forms of life. From a biological perspective, this makes intuitive sense because we know that all cells need to have a genome and it must be perfectly copied before it is passed to daughter cells during the process of cell division. It is this DNA polymerase enzyme that forms the heart of the PCR.

DNA is double-stranded: think of a ladder, but one that twists in the form of a helix. The side supports of the ladder are made up of so-called sugar phosphate backbone, the components of which are constant throughout the DNA molecule. However, the rungs of the ladder do not consist of a single piece of wood or metal but rather of two pieces. In the case of DNA, each rung is made up of two compounds, called bases, that vary up and down the ladder and provide the information that is encoded in the genes.

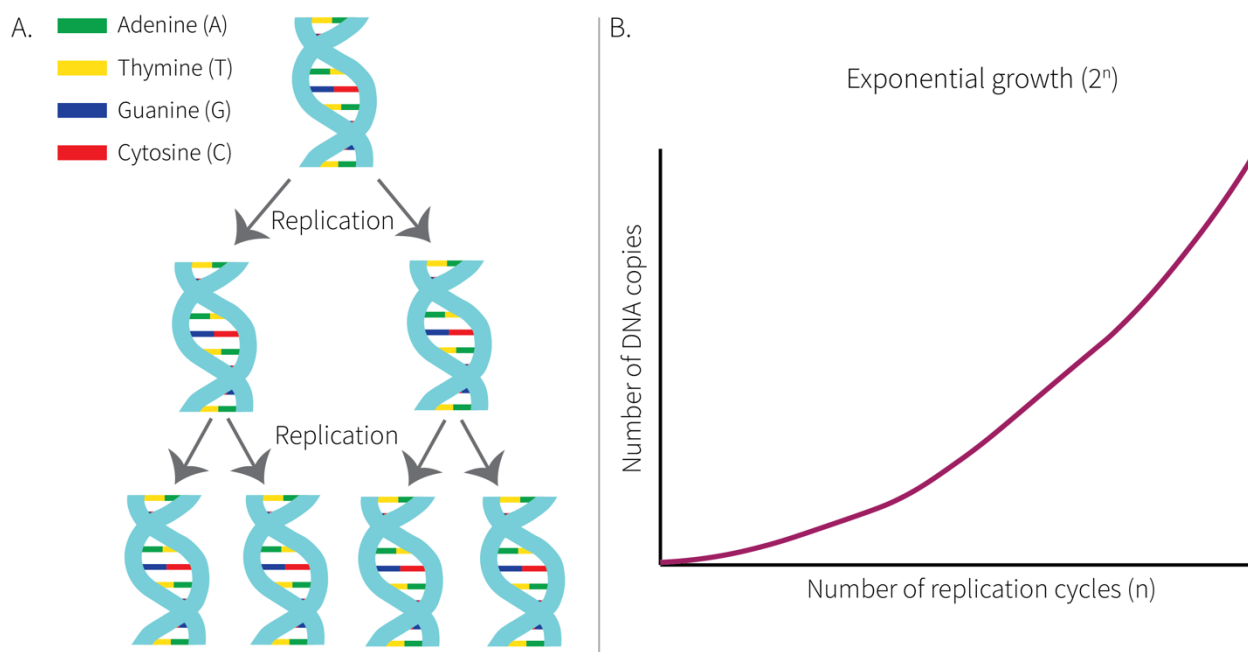
There are 4 bases of different sizes in DNA: adenine (A), cytosine (C), guanine (G), and thymine (T). These bases have 2 key properties: Adenine (A) on one strand chemically forms hydrogen bonds with thymine (T) in the other strand, whereas cytosine (C) in one strand chemically forms hydrogen bonds with guanine (G) in the other strand. *A sticks to T, and G sticks to C*. These pairings are specific, and other combinations are not possible. Secondly, although individual bases have different sizes, the base pairs A+T and G+C occupy identical spaces, which

A child-centric microbiology education framework

means that the rungs of the ladder – A+T or G+C – always have the same length – the side supports are always the same distance from one another, as in all good ladders. And these two features of the bases neatly provide a mechanism for duplication of the DNA molecule by DNA polymerase, because separation of the two strands allows them to be exactly copied, with an A on the strand to be copied always specifying a T on the new strand being synthesized. Each strand serves as a *template* for the synthesis of a new daughter strand. This finding by Drs. James Watson and Francis Crick was rewarded with the Nobel Prize.

This pattern of base pairing – the complementarity of the two strands of the DNA double helix – represents a zipper-like structure. Prior to DNA polymerase attaching to and copying DNA in a cell, the strands of DNA must be separated or “unzipped” by specialized enzymes called helicases. Once the double strand of DNA is separated into its two strands, each strand is copied by DNA polymerase to form a total of four single strands of DNA, or two double strands of DNA. This is what happens in the cell before it divides, so that each daughter cell receives a daughter genome.

When we want to amplify DNA to obtain large quantities, we do the same thing but in the “test tube” (i.e., outside of the cell), and we do not carry out one round of replication each time but rather multiple rounds: through a second round of copying, two double strands of DNA become four double strands, and, after a third round of copying, four double strands of DNA become eight double strands. This is called an exponential function and represents a doubling in the number of the DNA strands during each round of replication.



Structure and replication of DNA. (A) The DNA double helix undergoing replication. After one round of replication a single copy of DNA becomes two. After a second round of replication, these two copies become four, and so on. Each round of replication results in a doubling of the number of copies, known as exponential growth and in this case 2^n , where the 2 corresponds to the two strands of DNA and n is the number of rounds of replication, depicted in (B).

A child-centric microbiology education framework

Scientists have also learned that an efficient way to unzip the strands of DNA is by simply increasing the temperature to near boiling, or $\sim 94^{\circ}\text{C}$ (201°F). This is the preferred method of separating DNA strands in a test tube in the laboratory since it does not require a special helicase enzyme. However, these temperatures are hard on proteins and can cause them to denature and become non-functional. For example, when you cook a chicken egg ($\sim 71^{\circ}\text{C}$, $\sim 160^{\circ}\text{F}$), the proteins and enzymes that form the liquid components denature to form a more firm, solid texture that can be eaten. This also means that the egg itself can no longer develop into a chick, since those proteins and enzymes are denatured and no longer functional. A similar thing happens to a typical DNA polymerase enzyme when it was subjected to heating to 94°C – it becomes denatured and non-functional. This was a dilemma for scientists: in order to copy DNA using DNA polymerase, they first needed to heat the mixture to separate the strands. But, this essential heat step denatured and deactivated the DNA polymerase itself. Thus, for each doubling of DNA – each replication cycle – scientists had to add more DNA polymerase, a costly and time demanding task. What were scientists to do?

The Solution

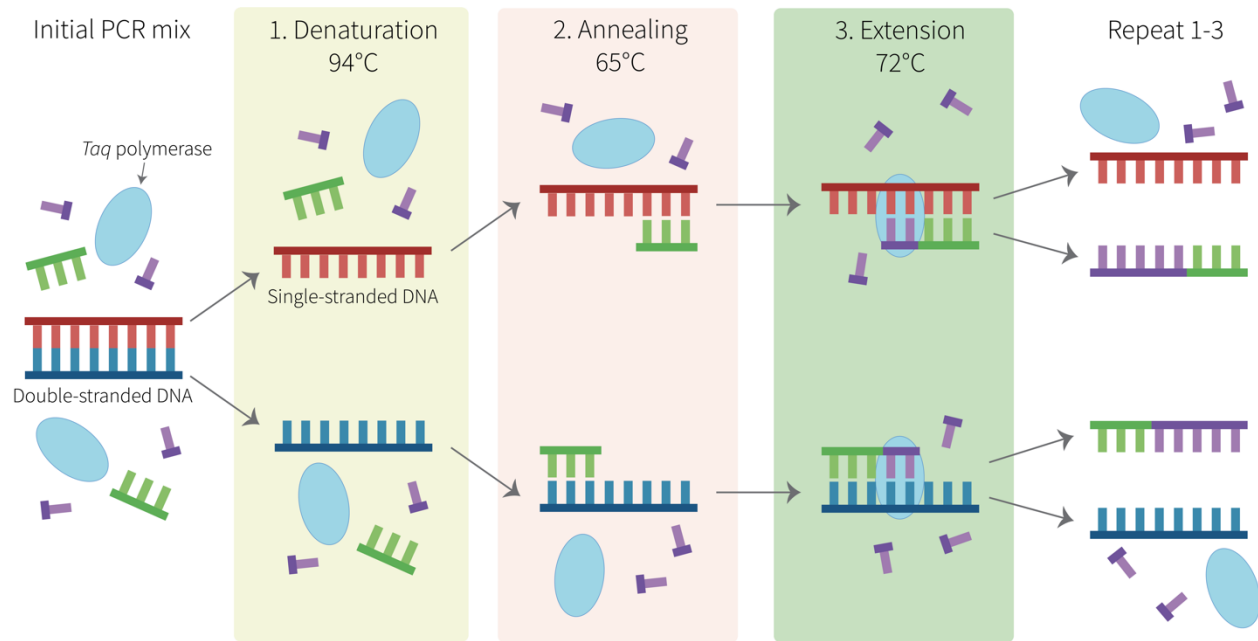
About the same time that scientists were developing methods to copy DNA, Dr. Brock and his colleagues had isolated the first thermophilic microorganisms, including *T. aquaticus*, from hot springs in Yellowstone. These organisms not only tolerated high temperatures ($>90^{\circ}\text{C}$), but they grew best at those extreme temperatures. At the time, no one suspected that any organism could thrive at these temperatures. For these cells to grow at high temperature, they need to be able to effectively copy their DNA at these temperatures. This meant that their DNA polymerase should remain functional at high temperature.

Scientists proceeded to isolate the DNA polymerase from *T. aquaticus* (called *Taq* polymerase) and test its utility in the repeated cycles of DNA strand unzipping and copying, all of which took place at temperatures much higher than the body temperature of humans or other animals. Remarkably, *Taq* polymerase remained functional each time the DNA was heated to 94°C to separate the strands prior to replication, meaning that the reaction could be automated and numerous cycles (rounds) could be run in sequence.

DNA polymerases cannot initiate the synthesis of a complementary DNA on a naked template DNA strand: they can only extend an existing complementary piece of DNA. The cell solves this by providing a short piece of DNA (or RNA), called a primer, on the template. Scientists figured out that by adding to the DNA mix a short, synthetic primer that could be extended by DNA polymerase, a particular region of DNA could be specifically targeted for precise amplification of a single gene of interest.

Soon, instruments that could rapidly heat and cool reactions were invented, allowing for the exponential replication of 1 molecule of double stranded DNA into nearly 500 million identical copies in just 30 cycles of PCR that together took only a few hours. This modern-day PCR technology ushered in a new generation of applications.

A child-centric microbiology education framework



PCR. Through a series of heating and cooling steps and with the addition of *Taq* DNA polymerase and primers, DNA can be replicated exponentially through the polymerase chain reaction (PCR). The first step of the PCR is to denature the double-stranded DNA into single-stranded DNA. This allows for short DNA sequences, called primers, to attach, or anneal, at 65°C to the single-stranded DNA in the second step. Finally, the *Taq* DNA polymerase extends the complementary single-stranded DNA starting at the primers (step 3). This is repeated upwards of 30 cycles to generate millions of copies of a single DNA fragment.

But how does PCR impact our daily lives? Some key examples

The PCR is an essential tool in clinical, pharmaceutical, forensic, and research laboratories because it can be used to create millions of identical copies of a gene or DNA fragment. In some cases, clinicians and researchers want to simply determine whether the gene is present in a sample or not. However, to do this, they need a sensitive way to detect a gene even if it is present in extremely low numbers. This approach might be used to identify an infectious microbe in the body of a sick patient. This use of PCR might enable the monitoring of the spread of an emerging contagious disease or to detect biothreat agents or antimicrobial resistance genes. In other cases, clinicians and researchers use the PCR to create enough identical copies of a gene to enable the determination of the sequence of that gene to identify an organism or individual. This is common in the research laboratory where scientists are aiming to describe the number of sequence variants of a gene (termed diversity) in a sample. Still in other cases, researchers use PCR to generate enough identical copies of a gene so that it can be effectively used in molecular biology applications to modify what an organism can do. Let's take a closer look at a few key examples of these applications.

1. Medicine. One of the most substantive ways that the PCR has impacted human society is in its use in the pharmaceutical industry. For example, it is estimated that nearly 6% of the world's population is affected by diabetes, a condition where the body does not produce adequate amounts of the hormone insulin to regulate the amount of glucose in a person's blood stream.

A child-centric microbiology education framework

This disease is especially prevalent among low income and impoverished populations, meaning that any treatment for this disease would not only need to be widely available but also low cost.

The primary treatment for diabetic patients is to provide them with insulin made from a source from outside of their body. In the past, insulin was obtained from the pancreas (the organ that makes insulin) from cows or pigs. However, today, nearly 500 million people are afflicted with diabetes while there are only ~ 1 billion cows and ~ 700 million pigs on the planet that could serve as a source of insulin. There are simply not enough cows or pigs on Earth to provide enough insulin to provide daily doses to diabetic individuals.

To overcome this limitation, scientists used the PCR and other molecular biology tools to turn microbial cells into insulin production factories. The genes that are required to produce insulin were copied by the PCR and introduced into the bacterium *Escherichia coli* and the fungus *Saccharomyces cerevisiae*. These genetically modified organisms (GMOs) are easy and inexpensive to grow in large quantities. This means that these GMOs can be used to produce large amounts of insulin for relatively low cost. We owe a debt of gratitude to the PCR, in particular *Taq* polymerase, for making it possible to generate synthetic insulin to treat the global diabetes epidemic.

2. Pathogen ecology and disease epidemiology. Between December 2019 and April 2022, nearly 500 million people were infected by a novel coronavirus (SARS-CoV-2), resulting in over 6 million deaths, and counting. We all remember this as the COVID-19 pandemic. However, how do clinicians and scientists know how many cases of COVID-19 have occurred globally? In other words, where did this number of estimated infections come from?

The SARS-CoV-2 virus is an RNA-virus, meaning that its genome is made of single stranded RNA rather than DNA. (Note: most scientists do not refer to viruses as living organisms because they cannot replicate on their own, thus the earlier statement that all living organisms have DNA genomes remains true). The most sensitive way to detect SARS-CoV-2 is to first convert the RNA of the virus into DNA through a process called reverse transcription. Then, clinicians and scientists apply the PCR to copy the SARS-CoV-2 genetic material to levels high enough to be identified in the samples. If you have had a PCR-based COVID test, you know these samples typically come from your nasopharynx, or your nose. This test is so sensitive that it can detect the presence of as few as 100 SARS-CoV-2 RNA molecules per milliliter of specimen. In this case, PCR is being used to identify and help stop the spread of the deadly virus that causes COVID-19. Of course, PCR is key to identifying new variants of concern of pathogens, mutants that have important new properties, such as heightened virulence or transmissibility, and was crucial for monitoring and modelling the successive waves of COVID-19 variants.

PCR was used as a diagnostic tool long before COVID-19; it is also commonly used to identify the presence of other human pathogens such as *Chlamydia trachomatis* (a bacterium that causes the sexually transmitted disease chlamydia), *Streptococcus* (the bacterium that causes strep throat), HIV (the virus that causes acquired immunodeficiency syndrome or AIDs), *Mycobacterium tuberculosis* (the bacterium that causes tuberculosis), *Neisseria gonorrhoeae* (the bacterium that causes the sexually transmitted disease gonorrhea), and influenza (the virus that causes the flu), among countless others. Once detected, appropriate treatments can be quickly administered to control these infections. Our ability to rapidly detect and identify infectious agents using the PCR

has had a large impact on how clinicians treat diseases and public health agencies monitor and control their spread through our communities.

3. *Forensic science.* PCR also plays an important role in criminal justice, particularly for forensic scientists. Forensic scientists are tasked with collecting and analyzing evidence from crime scenes. Oftentimes, the evidence that is collected is not in and of itself capable of being used to identify a criminal or a victim. For example, a hair left at a crime scene on its own cannot be used to identify someone since there are only so many varieties and colors of hair. The same is true for blood collected at a crime scene. Further, those hair and blood samples could be from an entirely different type of animal! Fortunately for forensic scientists, the root of a hair (where it attaches to your skin) and blood contain cells, and these cells contain genomic DNA unique to the individual who left them. However, criminals are often keen to not leave evidence behind and thus samples of their hair and/or blood are likely to be minimal.

To overcome this limitation, forensic scientists use the PCR to create millions of identical copies of the DNA from those pieces of evidence. With millions of identical DNA copies, forensic scientists can obtain the DNA sequence from the piece of evidence and compare it to DNA sequences obtained from suspects to find a match. Remarkably, the use of the PCR by forensic scientists has assisted in thousands of criminal investigations and prosecutions of perpetrators. It has also been used to prove that a suspect is innocent of a crime. In both instances, it has been used to retrospectively identify perpetrators/show innocence in old cases that have been re-opened (i.e., cold cases).

4. *Basic research.* The PCR is routinely used by the research community, and perhaps nowhere has it made more of an impact than on environmental microbiology. Environmental microbiologists have long known that the organisms that can be grown in the laboratory represent just a fraction of the organisms that are known to exist in nature. But it was not until the PCR was applied to DNA obtained from environmental samples that they were able to prove this unequivocally.

In the first application, scientists extracted DNA from Obsidian Pool, a hot spring in Yellowstone National Park. The small amount of DNA that was obtained was used in a PCR to generate enough copies of a gene to allow for its sequencing. What the scientists discovered was an entirely new world of microorganisms (identified by their unique DNA gene sequence) that was previously unknown to science. The sheer enormity of the newfound biodiversity is equivalent to walking outside and discovering plants (phylum Plantae), which represent a single phylum of organisms, for the first time but doing this close to a dozen times – nearly 20 new bacterial and archaeal phyla were discovered in that single hot spring sample.

In the three decades since its first use in environmental microbiology, DNA from thousands of samples from across the world have been subjected to PCR and sequencing and, amazingly, scientists are continuing to discover new microorganisms. These studies have revolutionized our understanding of the evolution of life on Earth and the role of microorganisms in the global biogeochemical cycles that sustain plant, animal, and human health. And let's not forget, each one of these newly identified organisms could host the next major discovery, just like *T. aquaticus* and the PCR.

Consequence for Humanity

The PCR has clearly had a huge effect on humanity. While it is difficult to quantify this effect overall given the diversity of ways that the PCR benefits humanity, it is worth noting that millions of PCR-based tests have been performed to help mitigate the spread of COVID-19 and to identify the right treatments for other diseases. Similarly, use of the PCR to develop alternative forms of insulin has made treatment of diabetes possible for millions of people. Aside from medical applications, many criminals have been convicted and innocent people exonerated using PCR-based forensic evidence. And, the PCR has made it possible for us to learn about the vast diversity of life with which we share this planet. It is difficult to imagine our world without the immeasurable benefits that the PCR has provided.