

Computing Is in Your DNA!

Paul S. Wang, Sofpower.com

November 13, 2024

In modern times, the term **DNA** is one of the most well-known and for good reason. We use DNA to store and pass on genetic information to the next generation. In fact, all known living organisms do it with DNA, including bacteria, viruses, plants, and animals.

The term ‘DNA’ has transcended language barriers and evolved into a universal symbol that carries profound meanings, both scientifically and culturally. But, our in-depth understanding of DNA would not have been achieved without the help of digital computing.

In this article, we’ll look at **DNA computing** and **digital computing** and the wondrous relations between the two.

This article is part of our *Computational Thinking* (CT) blog. You can find other interesting articles in *aroundKent* (aroundkent.net), an online magazine. You can also find many such articles in the author’s book *Becoming A Computational Thinker: Success in the Digital Age*. See the website computize.org/CTer for more information.

DNA Construction

Let’s begin with DNA’s composition and structure. A *DNA* (*Deoxyribonucleic Acid*) *molecule* consists of two strands (continuous polymers) of nucleotides (basic building blocks). Each DNA nucleotide is composed of a five-carbon sugar (deoxyribose), a phosphate group, and one of four nitrogenous **bases**:

- (A) Adenine,
- (T) Thymine,
- (C) Cytosine, and
- (G) Guanine.

From a computation viewpoint, these bases are used to encode information.

Think of one single strand in the DNA as a long string of beads where each bead represents a nucleotide and the beads form a sequence to encode information (Figure 1).

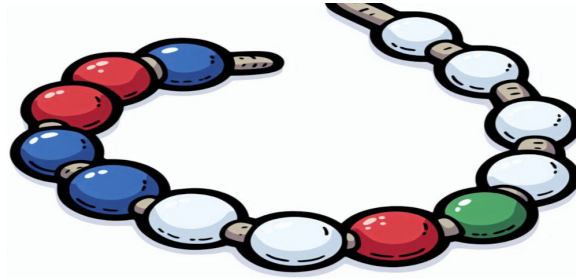


Figure 1: DNA Single Strand Concept

Data Representation in DNA

In a single *DNA strand*, the *specific order of the bases* encodes genetic information. Just as 0 and 1 (two symbols) are used to encode data in digital computers, A, T, C, and G (four bases) are used in DNA strands by organisms to encode aminoacid information as well as operational instructions. DNA contains information for the reproduction, evolution, and function of an organism. Essentially, it contains the genetic instructions for all aspects of life, including growth, traits, and cellular processes. How does DNA do it? Let's look at its structure.

DNA Double Helix

The two strands in the DNA are connected at the base positions to form a “*double helix*” structure (Figure 2)—the connected strands are twisted together in a spiral shape similar to a rope ladder or spiral staircase. The DNA nucleotide **base connections** form the steps of the ladder, and the deoxyribose and phosphate molecules form the sides of the ladder.

The two strands in the DNA are connected at the base positions following the *base paring rule*: **A pairs with T, and C pairs with G**. The pairings



Figure 2: Code in the DNA Double Helix

are known as *complementary base pairs*. The base pairing is very important because it contributes to the stability of DNA. And, it can be used to duplicate a strand using the paired strand as a mirror. From a computing viewpoint, the DNA structure enables information redundancy, stability, and error correction.

The double helix structure of DNA, discovered by James Watson and Francis Crick in 1953, is an iconic image that is immediately recognizable around the world. It visually symbolizes the complexity and beauty of life itself, making DNA not just a scientific term but also a powerful visual symbol.

Actual DNA Code

Let's see how information is actually encoded in DNA and how similar it is to bit patterns in digital computing.

In either single DNA strand of the double helix, a three-base combination of A, T, G, and C forms a *codon*. DNA code is always read as a sequence of codons. A codon may specify a particular amino acid, contribute to a functional protein or regulatory process, or mark the start/end of a code sequence.

For example:

- **ATG** is the *start codon* for protein synthesis (codes for methionine).
- **TAA**, **TAG**, and **TGA** are *stop codons* signaling the end of a functioning sequence.

- **ATGTTCCA**ACTGGTTGCGCAG..., is the beginning of the gene BRCA1 that encodes a protein involved in DNA repair.

For multicellular organisms, DNA is stored in the cell nucleus in the form of *chromosomes*. A chromosome is one separate DNA double helix tightly wrapped around proteins called *histones* to form a compact structure. The total number of chromosomes depends on the organism.

The cell can be seen as a kind of biological CPU that can execute the instructions encoded in DNA strands. And the DNA can be regarded as **firmware** that comes with the creation of the cell. The cell uses RNA (messenger RNA, in particular) to transcribe the DNA's instructions, similar to how a CPU reads and executes firmware.

DNA Compact Storage

Information is stored in DNA securely and in an extremely compact way. Each double-stranded DNA molecule is stored in a separate *chromosome* in the cell nucleus. For humans, there are 23 pairs of chromosomes. Each pair consists of one chromosome from the mother and a corresponding chromosome from the father. Thus, there are 46 separate individual chromosomes in a human cell nucleus.

Each chromosome is coiled, compacted, and organized in 3D space inside the nucleus. But, if we unwind and measure them, the total length of all the individual DNA molecules across all 46 chromosomes is about **2 meters**. Yet, the diameter of a human chromosome in its most compact form is only $1.4\ \mu\text{m}$ (10^{-6} meters).

A chromosome is extremely compact. In fact, it is one of the most efficient and densely packed ways to store information. The ingenuity of nature is amazing indeed. For comparison, one gram of chromosome (DNA) can theoretically store about 215 petabytes of data (215 million gigabytes), far exceeding the capacity of any current electronic storage media like hard drives or flash drives.

Moreover, DNA is much more stable than traditional digital data storage, with the potential to retain information for thousands of years. This is why we can recover DNA information from ancient remains. The stability is consistent with DNA's mission to reliably store and transmit *genetic information*.

DNA contains many **genes** (about 20,000 for human DNA). The genes

contain codes for building proteins and for creating RNA for other functions. We now know a lot about our DNA, its structure, genes, and many other details. In fact we have a complete map of our DNA thanks to the human genome project.

The Human Genome Project

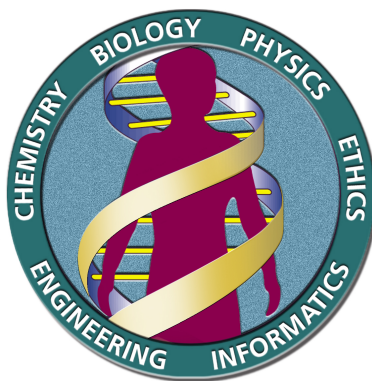


Figure 3: Human Genome Project Logo

The term *genome* refers to the complete set of DNA in an organism. The *Human Genome Project* (HGP) was a monumental, international research initiative aimed at sequencing the entire human genome, which consists of around 3 billion DNA base pairs. The project, a world-wide scientific collaboration, began in 1990 and was completed in 2003.

The HGP was one of the most ambitious scientific projects ever undertaken. Building on earlier developments and discoveries, the HGP produced a map of the entire sequence of human DNA, identifying all the genes and their positions on the 23 pairs of chromosomes.

HGP had many additional achievements—it identified how genes vary among individuals and populations and how these variations relate to diseases and traits. It also created tools for further DNA research.

Computers and DNA

Today we have a well-rounded understanding of DNA and have applied that knowledge to good ends. It is fair to say that our in-depth understanding of DNA would not have been achieved without the help of modern computing.

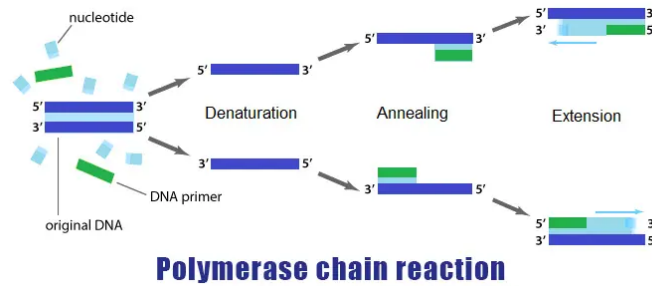


Figure 4: PCR Steps

In the first place, DNA research and application both rely heavily on Polymerase Chain Reaction (PCR, Figure 4), a technique developed in 1983 by Kary Mullis to amplify any desired DNA segments. PCR allows scientists to create millions of copies of a DNA sequence from a tiny sample, enabling more detailed analysis of genetic material. Dr. Mullies was awarded the 1993 Nobel Prize in Chemistry for his role in the PCR invention.

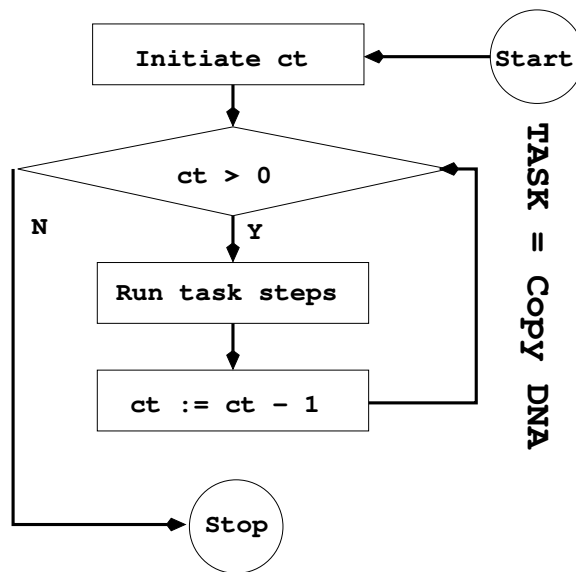


Figure 5: DNA Copying Iteration

In his book, Kary Mullis noted that the concept of PCR was inspired by looping principles in computer programming. The idea

of repeating cycles of heating and cooling to replicate DNA mirrors how computers perform repetitive tasks in loops. This is a **concrete example of applying computational thinking in a different discipline**.

In computer programming, repeating the same set of steps is called *iteration*. Starting with one copy of a DNA sequence, each heating and cooling cycle doubles the number of copies. After 32 iterations, say, you would get $2^{32} = 4294967296$ copies. That is amazing. It is almost impossible to imagine the HGP without PCR.



Figure 6: A PCR Machine

For easy visualization, Figure 5 shows a flowchart of *iteration* where `ct` is a variable used to keep count of the number of iterations still to be performed. When the task is to “*copy the given DNA sequences*”, you can see what Dr. Mullis meant. Figure 6 shows a modern PCR machine, inexpensive and available to order online.

Furthermore, the complexity of the DNA molecule, the vast amount of data it encodes, and the sophisticated analyses required to study it all rely heavily on the power of computing.

DNA+Computing Innovations

Computing allows us not only to better understand DNA but also to apply that knowledge in many areas. Notable examples include: DNA sequence and structure analysis, forensics and crime solving, CRISPR/Cas9 gene editing and therapies, targeted medicine, agriculture improvements, and more.

DNA technology combined with digital computing becomes such a powerful tool that it has changed our world significantly. The technology was

key in the fight against COVID-19 (Figure 7). DNA sequencing provided critical insights into the virus’s genome, while computing technologies enabled real-time data sharing, mutation tracking, vaccine design, diagnostic testing, and epidemiological modeling. These innovations allowed for a rapid and coordinated global response to the pandemic, demonstrating the power of genomics and computing in public health and disease management.

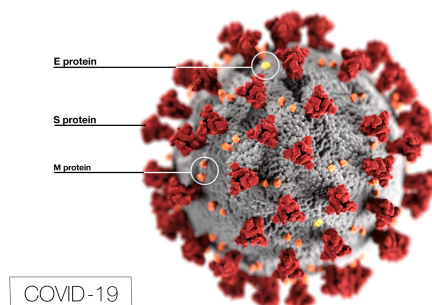


Figure 7: COVID-19 Virus

Without the synergy of DNA analysis and computational power, the response to COVID-19 would have been slower and less targeted. This integration of biotechnology and data science has not only been pivotal in managing COVID-19 but has also set a new standard for addressing future pandemics.

Digital Computing and Biological Computing

Alan Turing, the father of computer science, has described the essence of computing with the *Turing Machine* model (Figure 8)—a device with internal states capable of processing input and output following stored program instructions. Thus, a living cell can be considered as a biological Turing machine, where its DNA acts as the “program” that dictates its functions and behaviors. From this viewpoint, **digital computers and cells offer two very different and effective ways to perform computation.**

DNA stores genetic information and programming in the cells of organisms. Each cell can act as a *biological computer* and run DNA data as a program. Our understanding of DNA is deeply intertwined with advances in modern computing. Without computational tools, it would have been impossible to sequence genomes, analyze genetic variation, predict gene function, or make strides in gene editing technologies like CRISPR. Digital computers

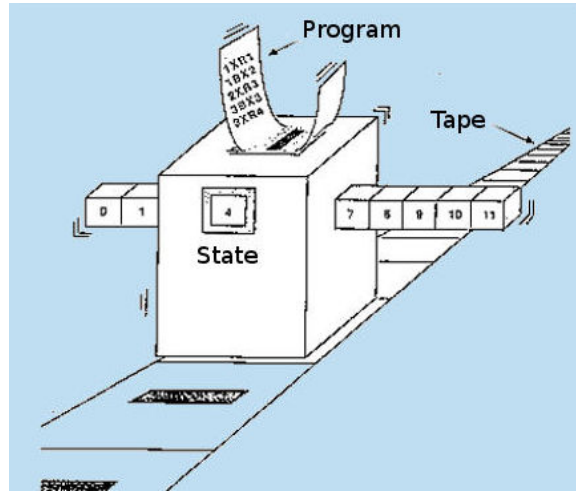


Figure 8: The Turing Machine Model

have not only enabled us to process vast amounts of genetic data but also to interpret, model, and apply this information in ways that are transforming science, medicine, and our understanding of life itself.

DNA and modern computing have a symbiotic relationship that drives breakthroughs in genomics, personalized medicine, biotechnology, and public health. On the other hand, biological computers invented by nature have many great features that man-made digital computers cannot begin to match—data stability, compactness, self-repairing, massive parallelism, and complex interactions. Computer science has a good model to study for insights. Thus, **DNA’s connection to computing is both intimate and inherent.**

Finally

The impact of DNA knowledge and applications is so great that the term “DNA” has transcended language barriers and evolved into a universal symbol that carries profound meanings, both scientifically and culturally. Regardless of the language, “DNA” is often used directly, without translation. It transcends linguistic differences because it represents a fundamental concept of life that is understood globally, much like mathematical symbols (e.g., π or $\sqrt{}$).

It is indeed remarkable that all living organisms share the same funda-

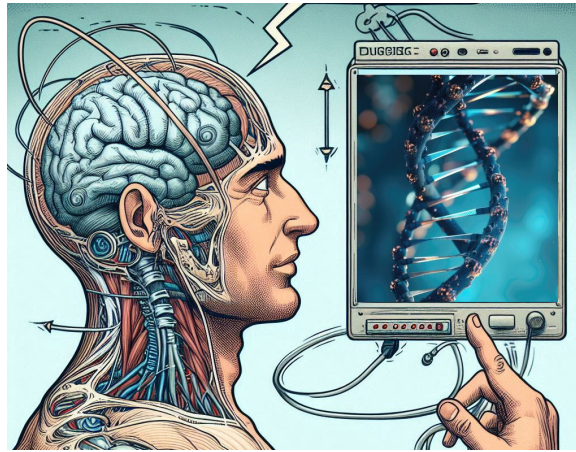


Figure 9: DNA Self Analysis

mental genetic code in DNA, yet only humans, with their advanced cognitive abilities, computing skills and tools, have achieved the scientific understanding to not only fully comprehend its structure and function but also manipulate it through genetic engineering.

Picture this (Figure 9)—a human being, a biological computer, using a powerful digital computer that it has invented to understand and analyze its own instructions. Does this image make your head spin?