**\subsection{Sanger dideoxy sequencing: the beginning}**

The definitive identification of DNA as the genetic that material is responsible for transformation of R strain of bacteria into S strain bacteria by Oswald Avery in 1944\cite{} and the discovery of the structure of the DNA by James Watson and Francis Crick in 1953\cite{} has sparked interest to develop DNA sequencing technologies to determine the nucleotide composition of organisms and to better understand genetic variations. In 1977, Frederick Sanger developed chain termination method\cite{}, and Allan Maxam and Walter Gilbert developed chemical cleavage method\cite{} for DNA sequencing. The chain termination method, also referred to as Sanger dideoxy sequencing, became the preferred method for DNA sequencing thanks to scalable implementation of the method using fluorescent nucleotides by Applied Biosystems (ABI) \cite{} and was the primary sequencing method for the Human Genome Project.

The original chain termination method starts with a single-stranded template DNA and a primer designed to bind to the start of the single-stranded template DNA. DNA polymerase (DNAP) binds to the primer and initiates DNA elongation using the free nucleotides in the mixture. Sanger mixed used both deoxyribonucleotides and dideoxyribonucleotides where the concentration of deoxyribonucleotides were higher than the concentration of dideoxyribonucleotides such that DNA elongation will be preferred. Until the incorporation of the dideoxyribonucleotide, DNAP will use the template DNA and perform DNA elgonation. Upon the incorporation of dideoxyribonucleotide, DNA elongation is terminated as the dideoxyribonucleotide does not have the 3’-OH to form phosphodiester bond with next nucleotide. Repeat chain-termination experiments results in DNA fragments of varying sizes and these fragments are ordered by their size through gel electrophoresis. This sequencing experiment is repeated with the four dideoxynucleotides (ddATP, ddGTP, ddCTP and ddTTP) such that the DNA sequence can be determined from reading the gel image from top to bottom \cite{}. Sanger and colleagues used this method to determine the sequence of the 5,375 bp long $\Phi$X174 bacteriophage. ABI modified the chain-termination method such that fluorophore-labelled dideoxynucleotide is used and this allowed sequencing from reading the fluorescence emitted from the chain-terminating nucleotide. The chain-termination method produces reads that are typically 500bp to 1000 bp long.

Why can it not read longer than 1000bp?

*Following rounds of template DNA extension from the bound primer, the resulting DNA fragments are heat*[*denatured*](https://en.wikipedia.org/wiki/DNA_denaturation)*and separated by size using*[*gel electrophoresis*](https://en.wikipedia.org/wiki/Gel_electrophoresis)*.*

*The DNA sample is divided into four separate sequencing reactions, containing all four of the standard*[*deoxynucleotides*](https://en.wikipedia.org/wiki/Deoxynucleotides-triphosphate)*(dATP, dGTP, dCTP and dTTP) and the DNA polymerase.  To each reaction is added only one of the four [dideoxynucleotides](https://en.wikipedia.org/wiki/Dideoxynucleotides" \o "Dideoxynucleotides) (ddATP, ddGTP, ddCTP, or ddTTP), while the other added nucleotides are ordinary ones. The deoxynucleotide concentration should be approximately 100-fold higher than that of the corresponding dideoxynucleotide (e.g. 0.5mM dTTP : 0.005mM ddTTP) to allow enough fragments to be produced while still transcribing the complete sequence (but the concentration of ddNTP also depends on the desired length of sequence).*[*[4]*](https://en.wikipedia.org/wiki/Sanger_sequencing#cite_note-Sanger1977-4)

**During sample preparation, different-sized fragments of DNA are generated each starting from the same location.**

**Each fragment ends with a particular base that is labeled with one of four fluorescent dyes corresponding to that particular base. Then all of the fragments are distributed in the order of their length via capillary electrophoresis. Information regarding the last base is used to determine the original sequence.**

**\subsection{Genome assembly}**

The ability to determine the nucleotide composition of organisms at scale with ABI capillary sequencing platform initiated a race to sequence the organisms around us to understand the structure and organisation of genomes and to the evolutionary relationship between organisms through comparative genomics.

**\subsection{Human Genome Project}**