11.2: DNA Sequencing

DNA sequencing determines the order of nucleotide bases within a given fragment of DNA. This information can be used to infer the RNA or protein sequence encoded by the gene, from which further inferences may be made about the gene’s function and its relationship to other genes and gene products. DNA sequence information is also useful in studying the regulation of gene expression. If DNA sequencing is applied to the study of many genes, or even a whole genome, it is considered an example of genomics.

11.2.1 Dideoxy sequencing

Recall that DNA polymerases incorporate nucleotides (dNTPs) into a growing strand of DNA, based on the sequence of a template strand. DNA polymerases add a new base only to the 3’-OH group of an existing strand of DNA; this is why primers are required in natural DNA synthesis and in techniques such as PCR. Most of the currently used DNA sequencing techniques rely on the random incorporation of modified nucleotides called terminators. Examples of terminators are the dideoxy nucleotides (ddNTPs), which lack a 3’-OH group and therefore cannot serve as an attachment site for the addition of new bases to a growing strand of DNA (Figure \(\PageIndex{3}\)). After a ddNTP is incorporated into a strand of DNA, no further elongation can occur. Terminators are labeled with one of four fluorescent dyes, each specific for one of the four nucleotide bases.
To sequence a DNA fragment, you need many copies of that fragment (Figure \(\PageIndex{4}\)). Unlike PCR, DNA sequencing does not amplify the target sequence and only one primer is used. This primer is hybridized to the denatured template DNA, and determines where on the template strand the sequencing reaction will begin. A mixture of dNTPs, fluorescently labeled terminators, and DNA polymerase is added to a tube containing the primer-template hybrid. The DNA polymerase will then synthesize a new strand of DNA until a fluorescently labeled nucleotide is incorporated, at which point extension is terminated. Because the reaction contains millions of template molecules, a corresponding number of shorter molecules is synthesized, each ending in a fluorescent label that corresponds to the last base incorporated.

The newly synthesized strands can be denatured from the template, and then separated electrophoretically based on their length (Figure \(\PageIndex{5}\)). Since each band differs in length by one nucleotide, and the identity of that nucleotide is known from its fluorescence, the DNA sequence can be read simply from the order of the colors in successive bands. In practice, the maximum length of sequence that can be read from a single sequencing reaction is about 700 bp.
A particularly sensitive electrophoresis method used in the analysis of DNA sequencing reactions is called **capillary electrophoresis** (Figure \(\PageIndex{6}\)). In this method, a current pulls the sequencing products through a gel-like matrix that is encased in a fine tube of clear plastic. As in conventional electrophoresis, the smallest fragments move through the capillary the fastest. As they pass through a point near the end of the capillary, the fluorescent intensity of each dye is read. This produces a graph called a chromatogram. The sequence is determined by identifying the highest peak (i.e. the dye with the most intense fluorescent signal) at each position.

*Figure \(\PageIndex{6}\)*: Fluorescently labeled products can be separated by capillary electrophoresis, generating a chromatogram from which the sequence can be read. (Wikipedia-Abizar Lakdawalla-PD)

### 11.2.2 Next-generation sequencing

Advances in technology over the past two decades have increased the speed and quality of sequencing, while decreasing the cost. This has become especially true with the most recently developed methods called next-generation sequencing. Not all of these new methods rely on terminators, but one that does is a method used in instruments sold by a company called **Illumina**. Illumina sequencers use a special variant of PCR called bridge PCR to make many thousands of copies of a short (45bp) template fragment. Each of these short template fragments is attached in a cluster in a small spot on a reaction surface. Millions of other clusters, each made by different template fragment, are located at other positions on the reaction surface. DNA synthesis at each template strand then proceeds using dye-labeled terminators that are used are reversible. Synthesis is therefore terminated (temporarily) after the incorporation of each nucleotide. Thus, after the first nucleotide is incorporated in each strand, a camera records the color of fluorescence emitted from each cluster. The terminators are then modified, and a second nucleotide is incorporated in each strand, and again the reaction surface is photographed. This cycle is repeated a total of 45 times. Because millions of 45 bp templates are sequenced in parallel in a single process, Illumina sequencing is very efficient compared to other sequencing techniques. However, the short length of the templates currently limits the application of this technology.

**Contributors and Attributions**

- [Dr. Todd Nickle and Isabelle Barrette-Ng](https://bio.libretexts.org/Bookshelves/Genetics/Book%3A_Online_Open_Genetics_(Nickle_and_Barrette-Ng)/11%3A_Genome_Sequencing_and_Bioinformatics) (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.