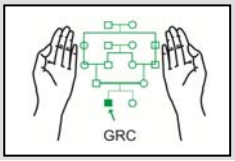
Genomic Sequencing

Maj Gen (R) Suhaib Ahmed, HI (M)



The process of determining the sequence of an unknown DNA is called sequencing. There are many approaches for DNA sequencing. In the last couple of decades automated Sanger's di-deoxy chain termination method has dominated the scene. It has played a pivotal role in the completion of human genome project. More recently the new generation sequencing (NGS) technologies are emerging to produce enormous amount of sequencing data at very low cost. Up to one billion short reads of sequence can be produced in one instrument run.

Next generation sequencing (NGS)

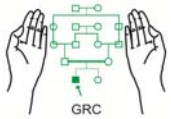
The NGS technology involves template preparation, sequencing, imaging, and data analysis. The template preparation is done by randomly breaking genomic DNA into smaller pieces from which fragment or mate-pair templates are created by emulsion PCR (emPCR) or solid phase amplification. The templates are immobilized on a solid surface as spatially distinct sites on which thousands to billions of separate sequencing reactions can take place. The sequencing is done by dye (fluorescent) labeled nucleotides. These nucleotides are incorporated specifically and are cleaved during fluorescence imaging. DNA synthesis is terminated by reversible terminators. The imaging of fluorescent labeled nucleotides produces data that are aligned with a known reference sequence.

Sequencing by automated Sanger's method

In a diagnostic lab sequencing is mostly done to see the presence of any mutation or polymorphism in an already known sequence of DNA. The later is also called re-sequencing. Sequencing of an unknown DNA is a tedious process and its description is beyond the scope of this book.

During synthesis of DNA deoxy-nucleotides (dNTPs) are incorporated in the DNA strand. In Sanger's method DNA strand synthesis is terminated by incorporation of di-deoxy-nucleotides (ddNTPs). The ddNTPs are synthetic analogs of dNTPs that can bind at 5' end with the preceding dNTP but are unable to form bond at 3' end with the next dNTP. In DNA sequencing the reactions are terminated randomly resulting in generation of varying lengths of DNA strands each terminating at the respective ddNTP. Previously the sequencing was done by radio-labeled ddNTPs in four separate reaction vials and the end products were electrophoresed in four separate lanes. In the automated genetic analyzers ddNTPs are labeled with four different flurochromes and the reaction product is run in a single capillary tube.

In a typical sequencing reaction the target DNA is first amplified by a pair of primers. The amplified product is purified to remove excess dNTPs, and unused primers. The sequencing is initiated by a single "sequencing primer" which is complementary to the target region of interest. The sequencing primer may be one of the primers used in the initial PCR amplification of the fragment or any other primer complementary to region of



interest in the amplified product. The reaction is done in a PCR tube that contains the amplified target DNA, the sequencing primer, dNTPs, fluorescent labeled ddNTPs, sequencing polymerase, and a reaction buffer. The end products of the reaction are purified and are analyzed on a genetic analyzer.

Direct sequencing

DNA can be sequenced directly by using single stranded or double stranded templates. Single stranded DNA is generated by asymmetric PCR in which the two primers are used in unequal proportions (one primer is in excess of the other). Double stranded DNA amplified by a pair of primers can be sequenced directly after separating its strands by heat denaturation and snap cooling.

Cycle sequencing

The DNA to be sequenced is first amplified by a pair of primers. The amplified target is then subjected to linear or asymmetric amplification by a single primer i.e. "sequencing primer". In a thermal cycler the target DNA and the reaction mixture are subjected to repeat cycling of denaturation, primer annealing, and extension. In each cycle DNA strands of varying lengths are produced that terminate at each ddNTP (Fig. 10.1). At the end of cycling the DNA strands of varying lengths are analyzed by capillary electrophoresis in a genetic analyzer.

(A)

3′-GAGTCGTACGCTACCATGTCATATCAACCATAGTTCCGATCCTAGTACTAAC-5′
5′-CATGCGATGGTACAGTATAG-3′→
3′-GAGTCGTACGCTACCATGTCATATCAACCATAGTTCCGATCCTAGTACTAAC-5′
5′-CATGCGATGGTACAGTATAG-ddTTP
5′-CATGCGATGGTACAGTATAGT-ddTTP
5′-CATGCGATGGTACAGTATAGTT-ddGTP
5′-CATGCGATGGTACAGTATAGTTG-ddGTP
5′-CATGCGATGGTACAGTATAGTTGG-ddTTP
5′-CATGCGATGGTACAGTATAGTTGGT-ddATP
5′-CATGCGATGGTACAGTATAGTTGGTA-ddTTP
5′-CATGCGATGGTACAGTATAGTTGGTAT-ddCTP
5′-CATGCGATGGTACAGTATAGTTGGTATC-ddATP
5′-CATGCGATGGTACAGTATAGTTGGTATCA-ddATP
5′-CATGCGATGGTACAGTATAGTTGGTATCAA-ddGTP
5′-CATGCGATGGTACAGTATAGTTGGTATCAAG-ddGTP

(B)

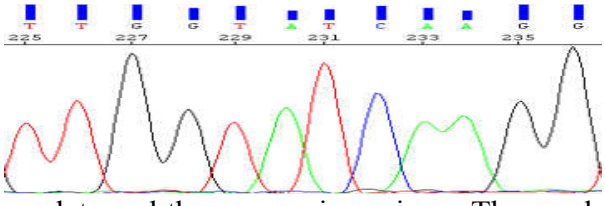


Fig. 10.1. (A) DNA template and the sequencing primer. The newly formed DNA strand is terminated when a ddNTP instead of the dNTP is incorporated. The four ddNTPs are tagged with different fluorescent dyes. (B) The DNA strands of varying lengths are

separated by capillary electrophoresis. Each fragment is represented by a coloured peak on electropherogram shown in the lower half of the picture.

Example

Locus: β-globin gene

GenBank accession: NG_000007.3

PCR Amplification:

- Forward primer 5'-ACCTCACCCTGTGGAGCCAC
- Reverse primer 5'-CCCCTTCCTATGACATGAACTTAA
- Amplified product: 676bp
- Reaction volume 50 µl
- Primer concentration: 2 µl each of the forward and the reverse primer (10 pmol/µl)
- Taq polymerase: 1 unit (0.2µl)
- Template DNA: 2 µl (~500ng)
- Thermal cycling:
  - Initial denaturation: 1 minute at 94^{o}C
  - Cycles: 25
  - Denaturation: 1 minute at 94^{o}C
  - Annealing: 1 minute at 65^{o}C
  - Extension: 1 minute 30 seconds at 72^{o}C
  - Final extension: 3 minute at 72^{o}C
- Electrophoresis: To see the quality of amplification run 3µl of amplified product on 6% mini-polyacrylamide gel for 40 minutes at 150 volts.
- Staining: 0.1% silver nitrate

Purification of amplified product

The amplified product contains unused dNTPs and primers that must be removed before sequencing. This may be done by gel purification, enzyme digestion or column purification.

1. Gel purification
   a. It is tedious but efficient in removing non-specific amplified products.
   b. The amplified DNA is run on 2% agarose gel at 150 volts for one hour. The gel is stained in ethidium bromide and is visualized by UV light. The bands of amplified DNA are localized and are carefully cut with knife. The piece of gel is transferred to an Eppendorf tube. The DNA may be extracted from the gel by a commercial spin column.

2. Enzymatic digestion

    a. The amplified DNA is incubated in a single step with exonuclease and alkaline phosphatase. However, the method is unable to remove non-specific amplified products.

    b. In a PCR tube 2-5µl of amplified product is incubated with 1µl each of exonuclease and alkaline phosphatase at 37^{o}C for 15 minutes. The enzymes are inactivated at 80^{o}C for another 15 minutes. The product is ready for use.

3. Column purification

    a. Micro-columns with silica gel filters are efficient in removing unwanted contaminants in amplified DNA. These may also be used to concentrate DNA. However, these are unable to remove non-specific amplified products.

    b. The micro-columns usually take 100µl of the amplified product. The initial amplification is done in duplicate of 50µl each and the end products are pooled.

    c. Purification is done as per instruction of the commercial column manufacturer.

Cycle sequencing

1. Since the cycle sequencing reaction requires extensive optimization the procedure is best done by a commercial sequencing kit.
2. The choice of sequencing primer depends on the application. Usual choice is the "forward primer" used in the initial amplification. However, any primer complementary to the amplified target sequence can be used. The sequencing primer is used at a concentration recommended by the sequencing kit manufacturer.
3. Volume of the purified amplified product (template) for use in cycle sequencing reaction depends on the quality and quantity of initial amplification. It ranges from 2µl for very good amplification to 8µl for poor amplification. For better results DNA in the template may be quantified (Chapter 2).
4. Thermal cycling parameters are also set as recommended by the kit manufacturer.

Purification of sequencing reaction products

The cycle sequencing reaction product contains many unwanted substances like unused dNTPs and fluorescent labeled ddNTPs. The later may cause dye blobs to appear in the electropherograms. These unwanted substances must be completely removed before the sample is run on a genetic analyzer.

Commercial kits are available for clean up of the sequencing reaction products. A good low cost method based on ethanol precipitation is described below.

Ethanol precipitation method

1. Add the following to the tube containing sequencing reaction product:
    a. Absolute ethanol, 50µl
    b. EDTA 125mM, 2µl
    c. Sodium acetate 3M, 2µl
2. Vortex gently to mix and centrifuge briefly.
3. Incubate at room temperature for 15 minutes.
4. Centrifuge at 14000 rpm for 15-25 minutes.
5. Carefully remove the supernatant using a fine tipped pipette.
6. Add 60µl 70% ethanol.
7. Vortex briefly and centrifuge at 14000 rpm for 5 minutes.
8. Carefully remove the supernatant using a fine tipped pipette.
9. Add 10µl Hi-Di formamide and use for electrophoresis.

Run conditions for genetic analyzer

The run conditions are chosen as per manufacturer's instructions.

Bibliography

1. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-termination inhibitors. Proc Natl Acad Sci, USA 74: 5463-5467.
2. Murray V (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res 17: 8889.
3. Metzker ML (2010) Sequencing technologies-the next generation. Nature Rev Genet 11: 31-46.