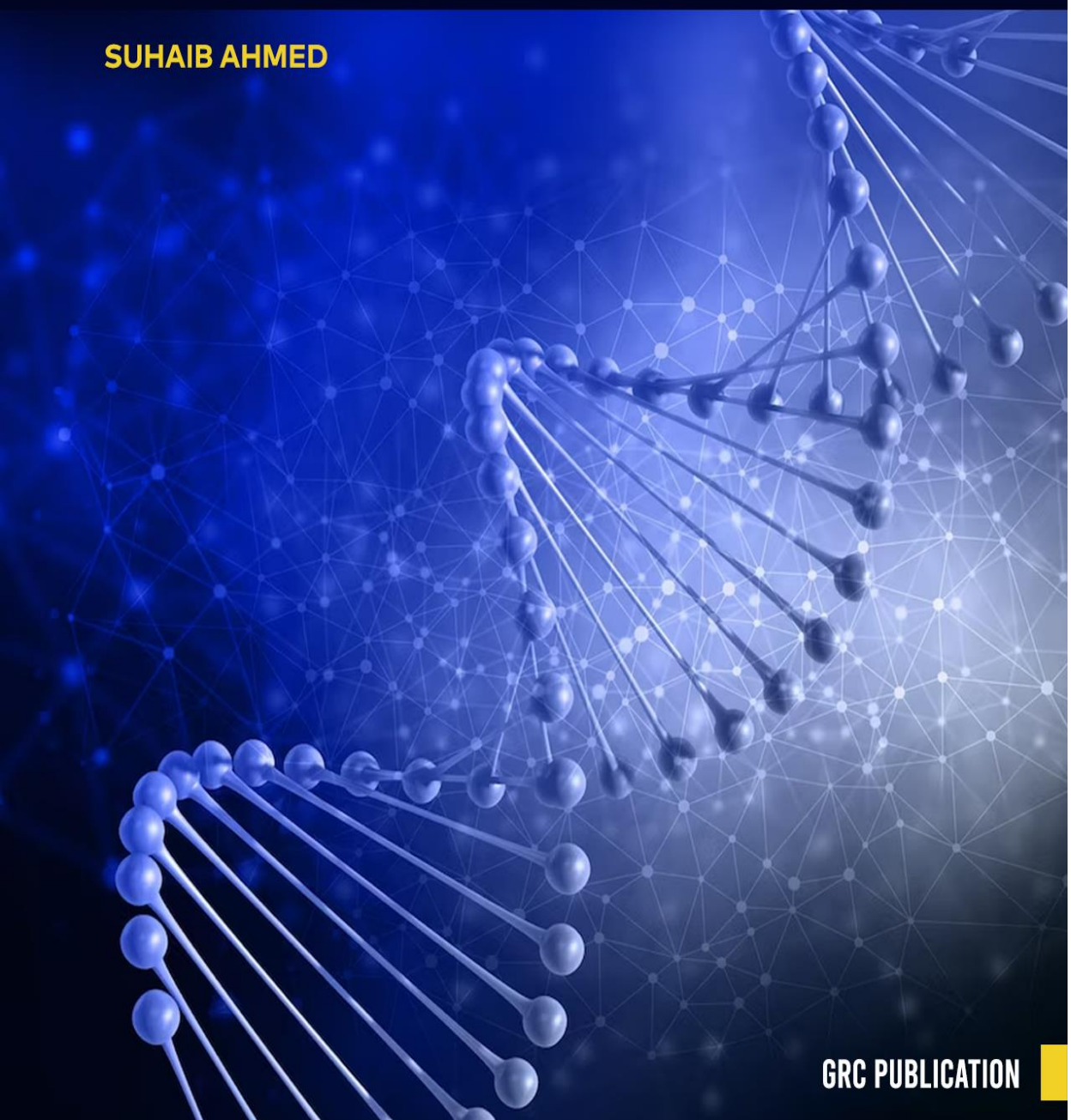
2ND EDITION

MANUAL OF PCR IN DIAGNOSTIC PATHOLOGY

SUHAIB AHMED



GRC PUBLICATION

Manual of PCR
in Diagnostic Pathology

2nd Edition

Suhaib Ahmed
MBBS; FCPS (Pak); PhD (London)

Manual of PCR in Diagnostic Pathology

ISBN: 978-969-23272-1-3

First Published 2013
Second Edition 2023

Produced by:
Genetics Resource Centre (GRC)
1st Floor Rafay Mall
Peshawar Road
Rawalpindi - Pakistan
Tel: +92-51-5167312
E-mail: info@grcpk.com
Web: www.grcpk.com

Table of Contents

Preface to the 1st Edition

Polymerase Chain Reaction (PCR) has made the analysis of genes simple and cheap. A large number of diseases with a genetic basis can be diagnosed by PCR. Prenatal diagnosis of inherited disorders by PCR in early pregnancy has revolutionized the management of common inherited disorders like thalassaemia. PCR has also become an essential tool for the identification of several infectious diseases. The hazardous procedure of viral culture is almost completely replaced by simple accurate and rapid PCR based assays. Real time PCR is being used to precisely quantify the target DNA or RNA of infectious agents or a clone of malignant cells.

There is no shortage of books on PCR. But most such books are written by scientists who have relatively limited knowledge of the clinical applications of PCR. Moreover, the books on PCR mostly focus on a limited range of applications. The residents in Pathology and laboratory technologists require a book that covers a broader range of subjects with reference to the PCR applications in diagnostic Pathology. This book would be equally useful for the pathologists, lab technologists, scientists, researchers and university students.

The book has two sections. The first section deals with the basic techniques used in PCR and related procedures like nucleic acid extraction and post PCR analysis. The second section of the book includes common applications of PCR in inherited, neoplastic and infectious diseases. In addition, chapters on forensic DNA testing and study of donor chimerism are also included.

I would be grateful if the readers could point out the mistakes, errors or omissions in the book. This feedback will help in improving the standard of the future editions of the book.

Maj Gen Suhaib Ahmed (R), HI (M)
Rawalpindi
Jan 2013

Preface to the 2nd Edition

With the expanding scope of molecular pathology PCR is now considered as an essential diagnostic tool for a good diagnostic lab. There is no dearth of information on the basic technique and applications of PCR in various disciplines of pathology. In this book I have tried to explain things in a way that I wanted to know about PCR when I was learning it.

The book mostly focuses on the practical aspects of PCR. It describes several relevant examples that may be helpful in developing low-cost in-house PCR applications. In the second edition of the book almost all chapters have been revised. A lot of new information regarding gene expression studies, next generation sequencing and sequence variant nomenclature etc. has been added or updated.

I hope this book will continue to be a source of information for pathologists, lab technologists, scientists and students of biological sciences.

Suhaib Ahmed
Rawalpindi
Apr 2023

Foreword

Any professional in Pakistan associated with molecular pathology, hemoglobinopathies, or hematology in general, would need no introduction to the author of the book Manual of PCR in Diagnostic Pathology, – Dr. Suhaib Ahmed. Not only is it an absolute delight for me to review this particular book of his, which has been of so much practical assistance to me, but I will also take this opportunity to express my views about the author of the book - one of the most original thinkers I have come across in general, and specifically, change-drivers in the field of hemoglobinopathies and molecular diagnostics in Pakistan. The statement may be cliché, but here it holds true regardless - it is a tremendous honor for me to be writing a review for one of the books authored by Dr. Suhaib Ahmed. In my role as an audience in many of his highly informative workshops and lectures, I have been immensely inspired by his originality of ideas, his knowledge, and even more importantly, translation and implementation of those ideas into services and products in Pakistan. I can say with assurance that he has been a significant driving force behind my ever-accelerating awe with the field of molecular pathology, ultimately culminating in my own pursuance of a doctoral degree in human and molecular genetics. While Dr. Suhaib's numerous publications in renowned international journals and his other scholarly work are enough to testify to his genius, he has translated that genius into scaling up services and applications in Pakistan, which is what truly sets him apart.

The field of molecular pathology is young but, arguably, the most rapidly expanding area of medicine. Research in this discipline started to pick up pace starting the 1970s, when the technology of cloning was discovered, leading to medicinal breakthroughs like recombinant insulin and other hormones. It was not until the 1990s, however, that molecular genetic techniques were purposed, in the earnest, for clinical diagnostics. The turn of the century, with completion of the human genome project and advances in computing and bioinformatics, gave molecular diagnostics and research a phenomenal momentum, a receding of which is nowhere in sight. It can be reasonably argued that many of the current clinical diagnostic techniques will become obsolete in the next few

years, to be superseded by molecular diagnostic techniques. It goes without saying thus that the field needs urgent development in Pakistan, if we are to hope to catch up pace with modern medicine. While some research opportunities in molecular genetics have existed in Pakistan, the general unfamiliarity with the diagnostic aspects of the field can be gauged from the fact that the College of Physicians and Surgeons, the premier medical post-graduate institute in Pakistan has not, to date, initiated a fellowship program in molecular pathology. A lack of structured training opportunities in Pakistan in the specialty, has been an important limiting factor for the growth of this field in the country. In this context, the book Manual of PCR in Diagnostic Pathology, fills an important knowledge gap. Though interesting in its own right and useful for readers of all clinical backgrounds, this book is of particular utility for those intending to initiate new molecular diagnostic services in Pakistan. One important challenge that organizations face when setting up these services are a dearth of pertinent, context-specific, practical, and step-by-step guidelines. The questions that such organizations are looking to be answered are how should we start? Which molecular assays are most frequently needed and of the highest clinical utility in Pakistan? What materials and equipment should we get? What are the most cost-effective, yet sensitive and specific methods for a particular clinical assay? What will be the detailed standard operating procedures of these assays? While there is no dearth of international literature in the field, an initial setter up of molecular diagnostic services in Pakistan may find those overwhelming – with their translation into local implementation challenging - some significant considerations in this regard being cost, and local availability of materials and consumables for unhindered scalability and throughput, as is often needed in clinical services. Overall, the book answers these questions in the form of an exceedingly comprehensive, utilitarian, yet simplified, step by step guide to setting up molecular diagnostic services for any resource-constrained organization.

Dr. Javeria Aijaz
MBBS FCPS (Hematology) PhD (Human Genetics, USA)
Section Head Molecular Biology Clinical Laboratories
Indus Hospital and Health Network,
Karachi

1

Genome and its Resources

Deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins are the three most essential molecules of life. The blue print of protein structure, called "genetic code", resides in the DNA that transcribes to form messenger RNA (Figure 1.1). A subtle change in the "genetic code" can result in profound phenotypic effects by altering the quantity or the quality of the biologically active protein molecules.

Human Genome

Genome refers to the genetic material of an organism. The human genome comprises more than three billion bases and most of it lies in the nucleus. A small portion (16.6 kb) of the genome is also present in the mitochondria (Figure 1.2).

GRCh37 & GRCh38

These are the Genome Reference Consortium Human genome build 37 & 38. As of May 7, 2014 GRCh38 is the standard reference assembly sequence used by NCBI. Unlike other sequences, GRCh38 is not from one individual's genome sequence, but is built from reference sequences of different individuals.

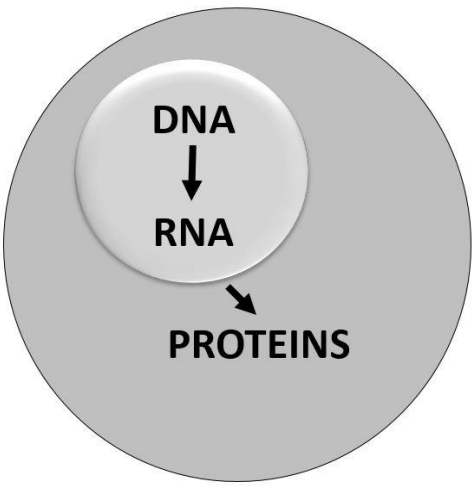


Figure 1.1. Three most essential molecules of life. DNA transcribes to RNA that translates to form protein. The proteins carry out most of the biological functions in a cell.

The nuclear DNA is present in 23 pairs of chromosomes including 22 autosomes and one pair of sex chromosomes. Each chromosome is composed of a single coiled up molecule of DNA supported by a protein backbone (Figure 1.3). The segments of genome that codes for proteins are called genes. Approximately 2% of the human genome contributes to a phenotype/function and is also called exome. The rest of the genome apparently does not perform any function. Most of the extra-genic DNA is unique whereas the rest has tandem and non-tandem repetitive sequences.

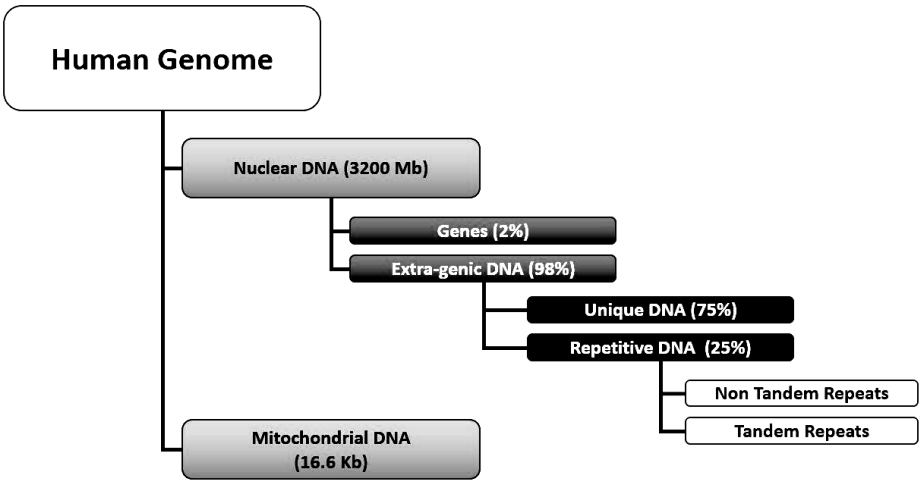


Figure 1.2. Organization of the human genome. Most of the DNA is contained in the nucleus and only a small portion is present in the mitochondria. Genes comprise only 2% of the nuclear DNA.

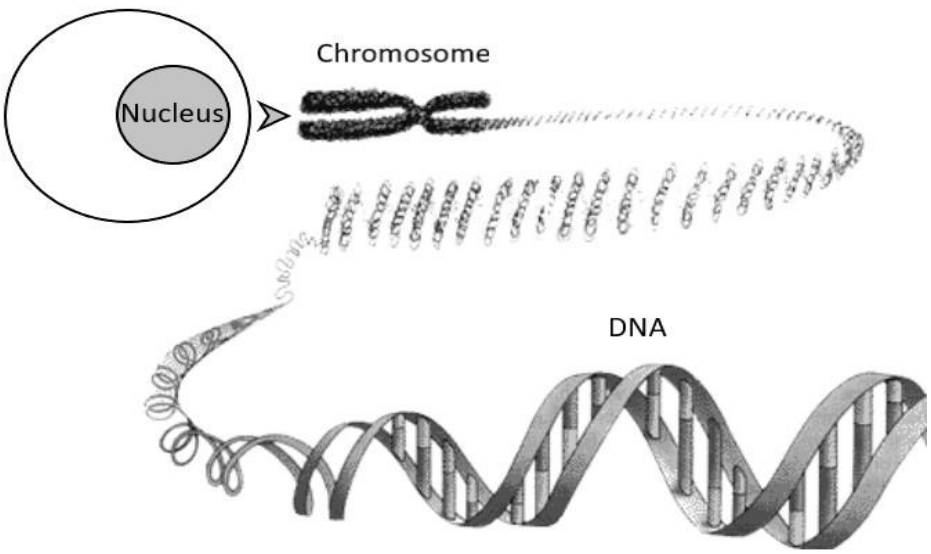


Figure 1.3. Most of the DNA in a cell is contained in the nucleus where it lies in the form of 23 pairs of chromosomes. Each chromosome is composed of a protein backbone and single coiled up molecule of DNA.

Deoxyribonucleic acid (DNA)

DNA is a double stranded molecule. Each strand is made of individual units called nucleotides. The nucleotides consist of deoxyribose (5-carbon sugar), nitrogen containing base attached to the sugar, and phosphate group (Figure 1.4). There are four different nucleotides each differing by the nitrogenous base. The bases are either purines (adenine and guanine) or pyrimidines (cytosine and thymine). The four bases are abbreviated as A, G, C, and T. The deoxyribose sugar has 5 carbon atoms numbered 1', 2', 3', 4', and 5'. A hydroxyl group on the 5’ and 3’ carbons is attached to a phosphate group to form the backbone of DNA. The 5’ end of the strand is the starting point of DNA molecule while the 3’ end is the terminal point. During the synthesis of a new strand nucleotides are added at the 3’ end and the strand elongates from 5’ to the 3’ end.

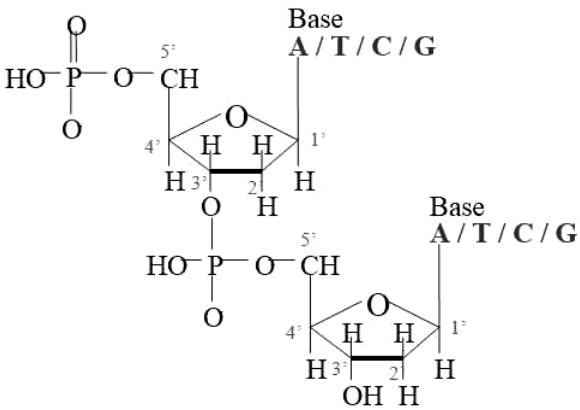


Figure 1.4. Chemical structure of the DNA molecule.

The two strands of DNA run in opposite directions and wind around each other to form a right-handed spiral. By convention the strand that runs from left to right (→) is called the forward strand and the strand running from right to left

(←) is called the reverse strand (Figure 1.5). The nitrogenous bases of the nucleotides face towards the interior of the helix. The two strands are kept together through hydrogen bonds between purines and pyrimidines. Adenine (A) forms two hydrogen bonds with thymine (T) and cytosine (C) forms three hydrogen bonds with guanine (G) on the opposite strand. The G-C bond is stronger than the A-T bond. This makes the G-C rich areas of DNA more stable than the A-T rich areas. The synthesis of DNA starts with separation of the two strands followed by addition of nucleotides on each strand from 5' end to the 3' end.

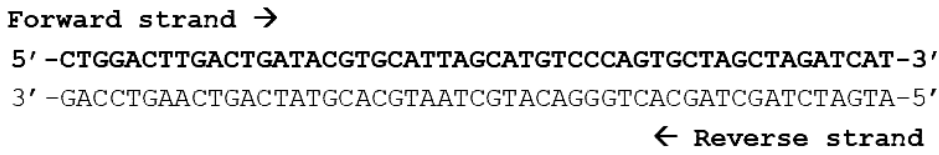


Figure 1.5. Double stranded DNA. The forward and the reverse strands are complementary to each other and run in opposite directions.

DNA sequence

The nucleotide composition of a DNA molecule is called its "sequence". To avoid complexity the DNA sequence of only one of the strands is written (Figure 1.6). The sequence is written from left to right starting from the 5' end to the 3' end. Only the abbreviations of the nucleotides i.e. G, A, T, and C are used. Occasionally other abbreviations are also used e.g. Y for pyrimidines, R for purines, and N for any nucleotide.

Ribonucleic Acid (RNA)

RNA mostly exists as a single stranded chain of nucleotides. The nucleotides of RNA are the same as of DNA but uracil (U) replaces thymine (T) and ribose sugar replaces deoxyribose. Uracil is a pyrimidine that is structurally similar to thymine, and it can also base-pair with adenine. In a cell RNA is seen as messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA). In addition, the cells also contain a variety of other types of RNA like micro-RNA

and antisense-RNA that can cause post-transcriptional modification of protein synthesis. The basic structure of all types of RNA molecules is similar.

Figure 1.6. DNA sequence of the HBB (Haemoglobin β-gene), GenBank accession: NG_000007.3. The left hand column shows the numbring of nucleotide bases. The three coding regions (CDS) are shown as highlighted text.

National Centre for Biotechnology Information (NCBI)

The most common source of genomic information is the National Centre for Biotechnology Information (NCBI) at the National Institute of Health (NIH), Bethesda USA (http://www.ncbi.nlm.nih.gov/). At the NCBI website the information is contained in databases including PubMed, Genes, and Proteins.

GenBank is a genetic sequence database of all publicly available DNA and RNA sequences. The information is submitted to GenBank by the individual

researchers and project groups all over the world. The files in the GenBank are sorted into 'divisions' such as bacteria (BCT), viruses (VRL), primates (PRI) and rodents (ROD) etc. There are twelve taxonomic divisions (BCT, ENV, INV, MAM, PHG, PLN, PRI, ROD, SYN, UNA, VRL, VRT) and five high-throughput divisions (EST, GSS, HTC, HTG, STS). As of Aug 2017, there are 2,635,527,587,818 bp of records in the GenBank and it is annually increasing at 35%.

GenBank is part of the International Nucleotide Sequence Database Collaboration which includes DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

GenBank accession number

When a new sequence is received at the GenBank it is assigned a unique accession number. The accession number format consists of a prefix, including the underscore, followed by 6 numbers followed by the sequence version number. Unique prefixes are reserved for the chromosomal regions, DNA, RNA, mRNA and Protein sequences (Table 1.1). The older as well as the newer versions of the sequences are kept in the database. In the accession number NG_000007.3, NG_ is the prefix, 000007 is the sequence number and the digit 3 after a decimal indicates the version.

Table 1.1. GenBank accession number prefixes.

Prefix	Molecule type	Use context
NC_, AC_, NZ_	DNA	Chromosomes
NG_	DNA	Genomic DNA
NM_, XM_	mRNA	Protein coding transcripts
NR_, XR_	RNA	Non protein coding transcripts
NP_, AP_, XP_, YP_	Protein	Various

Accessing a sequence record

The nucleotide sequence of a record can be accessed by entering the accession number at the NCBI website. For example, the human beta globin complex can be accessed by searching for NG_000007.3 at the NCBI site.

The sequence records can be viewed in the GenBank, FASTA or Graphic format. The GenBank view provides a complete summary of the record with multiple display options. It also provides analysis tools like running BLAST, picking primers, highlighting the sequence features (exons), mRNA, and protein sequence etc. The FASTA format provides option of cut/paste of a region of the sequence or download of the complete sequence. The graphic format provides a complete graphic summary of the record.

It is mandatory that all scientific publications related to genomic data should provide the GenBank accession numbers. The GenBank database is also linked to scientific literature via PubMed and PubMed Central therefore cross reference and searching is greatly facilitated.

Basic Local Alignment Search Tool (BLAST)

BLAST is one of the most frequently used tools of the GenBank. It is a family of search tools that allows comparison and finding similarities between various sequences. BLAST allows comparison of a sequence data with any of the known sequences available in the GenBank. For example, the primer BLAST feature allows searching of the entire human, bacterial, viral or parasitic genomes to know whether a newly designed primer would cross anneal with a homologous sequence elsewhere or not.

Proteins and Genetic Code

Protein are large biomolecules made up of a string of varying numbers of amino acids. The amino acids in a protein are held together by peptide bonds. Each amino acid of a protein chain is coded by a triplet of DNA bases. Most amino acids are coded by more than one triplet. Like DNA and RNA, the

constituent amino acids of a protein are also assigned unique three letter and single letter codes (Table 1.1). The amino acid composition of a protein is called its sequence and is written like the sequence of DNA or RNA (Figure 1.6).

Like DNA and RNA, the GenBank protein database is a collection of amino acid sequences from several sources including the annotated coding regions in the GenBank, each sequence has its unique identifier. For example, NP_000198.1 is the accession no of human pre-pro-insulin (Figure 1.6).

Table: 1.1. Amino acid codes and the codons.

Amino Acids	Codes	Codes	Codons					
Alanine	Ala	A	GCA	GCC	GCU			
Cysteine	Cys	C	UGC	UGU				
Aspartic Acid	Asp	D	GAC	GAU				
Glutamic Acid	Glu	E	GAA	GAU				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Argenine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				
Stop Codon	Stop	-	UAA	UAG	UGA			

```
ORIGIN
        1 malwmrllpl lallalwgpd paaafvnqhl cgshlvealy lvcgergffy tpktrreaed
       61 lqvgqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn
```

Figure 1.6. Amino acid sequence of human pre-pro-insulin (NCBI accession no: NP_000198.1). For convenience the sequence is written in blocks of ten amino acids each.

Bibliography

1. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW (2018) Nucleic Acid Res 46: D41–D47. doi:10.1093/nar/gkx1094
2. Makalowski M (2001) The human genome structure and organization. Acta Biochim Pol 48: 587-598.

2

Extraction of Nucleic Acids

The first step in analysis of DNA or RNA is to get a good quality sample; a process commonly known as “extraction”.

DNA Extraction

DNA is a very long molecule and it can easily break by vigorous shaking during the process of extraction. Therefore, gentle and careful handling in the processing is essential. DNA can also be destroyed by the DNase that is commonly present in the environment or in the bacteria that may contaminate the sample.

There are three basic steps of DNA extraction:

1. Removing the membrane lipids by detergents to expose the DNA in the nucleus of cell.
2. Removal of proteins by protease digestion and precipitation by phenol or other agents.
3. Precipitation of DNA with ethanol or isopropanol.

Sources of DNA

DNA can be extracted from any source that contains nucleated cells. Most commonly it is extracted from the blood collected in EDTA. EDTA has an additional advantage of chelating magnesium ions which reduce the activity of DNA degrading enzymes of cellular or bacterial origin. Heparin is not a good anticoagulant as it interferes with subsequent PCR. The blood may be kept at $4^{o}C$ for a few days or at $-20^{o}C$ for several months without causing significant loss in the yield of DNA. Blood contaminated with bacteria becomes unsuitable for DNA extraction as the bacterial enzymes can easily degrade DNA. DNA can also be extracted from bone marrow aspirates or bone marrow smears on slides. Archival bone marrow slides stored at room temperature for several years can be used to extract good quality DNA.

Buccal smear on cotton swab or mouth wash is another easily available source of DNA. This is especially useful for field work. Solid fresh tissues, like surgical biopsy specimens, chorionic villi and tissues collected at autopsy are also used for DNA extraction but the tissue should not be fixed in formalin. It should preferably be kept in normal saline. DNA can also be extracted from hair root, blood stains, archival bones etc. Fixation of the tissues with formalin can make DNA extraction very difficult. Special processing protocols may be required to extract DNA from formalin fixed and paraffin embedded tissues.

Choice of method

The standard method of DNA extraction uses phenol chloroform for protein precipitation. Keeping in view the toxicity of phenol, methods have been

developed to precipitate proteins without using phenol. A large number of commercial kits are available that are time and cost effective. Some of the methods can also be automated for large scale DNA extraction. A quick method of DNA extraction is by ion exchange resin Chelex 100.

In the subsequent section phenol chloroform method is described in detail. It is robust and cost effective and consistently gives good quality high molecular weight DNA.

DNA extraction from whole blood

1. Collect 3 ml blood in EDTA.
2. Centrifuge at 3000 rpm for 5-10 minutes.
3. Remove the supernatant plasma leaving behind the buffy coat and the RBCs.
4. Add red cell lysis buffer (Table 2.1) 2-3 times the volume of the red cells and mix by inverting a few times.
5. Centrifuge at 3000 rpm for 5-10 minutes and discard the supernatant.
6. Repeat the above step once more if the cell pellet contains too many RBCs.
7. Add 0.7 ml cell lysis buffer (Table 2.1) and transfer the contents to a 1.5 ml Eppendorf tube.
8. Add 20 µl Proteinase-K (Table 2.1) and mix by gentle vortex.
9. Incubate at 37^{o}C overnight or at 56^{o}C for two hours.
10. The cell lysis buffer may be replaced guanidine isothiocyanate (Table 2.1) that can effectively breakdown the proteins and also does not require Proteinase-K. Use 0.7 ml cell lysis buffer with guanidine to lyse the WBC pallet. Shake well on vortex and place at 37^{o}C overnight.
11. Add 250 µl buffered phenol (Table 2.1) and 250 µl chloroform.
12. Vortex for a few seconds.

13. Centrifuge at 10,000 rpm for 2 minutes in a micro-centrifuge.
14. Carefully remove the subnatent phenol layer with a pasture pipette leaving behind the clear watery supernatant.
15. Repeat the above step if cloudiness remains in the supernatant.
16. Add 500 µl chloroform and vortex for a few seconds.
17. Centrifuge at 10,000 rpm for 2 minutes.
18. Remove as much of the subnatent chloroform as is possible leaving behind clear supernatant DNA solution.
19. Add 150µl 7.4M ammonium acetate solution.
20. At this stage the DNA solution left from the previous step should be approximately 500µl. Fill the Eppendorf tube to its top with pure ethanol (about 1 ml). This will make a final concentration of 70% ethanol in which the DNA forms a whitish precipitate.
21. Gently invert the Eppendorf tube 3-4 times and watch for a whitish hairball like precipitate of DNA.
22. Centrifuge at 10,000 rpm for 2 minutes.
23. Remove the ethanol by gently inverting the tube and leaving behind the DNA pellet at the bottom of the tube.
24. Add 500 µl of fresh 70% ethanol and gently vortex to give the DNA pellet a good wash in ethanol.
25. Centrifuge at 10,000 rpm for 1 minute.
26. Remove as much of the top ethanol layer as is possible. Leaving ethanol behind can interfere with subsequent DNA hydration.
27. Dry the DNA pallet for 5-10 minutes by keeping the tube in inverted position on a clean tissue paper.
28. Dissolve the DNA in DNase free water. Ordinary distilled water can also be used if DNase free water is not available. The amount of water to be added depends on the yield as seen in the DNA pellet. On an average the DNA extracted from 3ml of blood with a normal white cell count can be dissolved in 200-300µl of water to give a

final concentration of ~200ng/µl. It is safe to add less water as the concentrated solution can be diluted further whereas a diluted DNA is very difficult to concentrate.

29. Leave the DNA solution at 37°C for 15-30 minutes.
30. DNA may be stored at 4°C for a few weeks, at -20°C for several months and at -80°C for several years.
31. The DNA solution is fairly stable at room temperature for many days. It can be transported from one place to another without being kept in ice.

DNA extraction from CVS and fresh tissues

1. Take approximately 25-50 mg of fresh tissue (chorionic villi, skin, or other solid tissues) in 0.5 ml of cell lysis buffer (Table 2.1).
2. Add 20-40 µl of Proteinase-K (Table 2.1) depending on the amount of tissue.
3. Keep at 37°C overnight. Allow longer incubation or add more Proteinase-K if the tissue is not completely digested/dissolved.
4. Proceed as step 11 onwards of the DNA extraction protocol.

DNA extraction from archival bone marrow slides

1. Take a slide of bone marrow smear that has good number of cells.
2. Layer about 0.7 ml cell lysis buffer (Table 2.1) on the smear.
3. Gently scratch the smear from the slide with a wooden stick and transfer the contents to an Eppendorf tube.
4. Add 20 µl Proteinase-K (Table 2.1) and keep at 37°C overnight.
5. Cell lysis buffer with guanidine (Table 2.1) may be used instead of the standard cell lysis buffer and Proteinase-K.
6. Proceed as step 11 onwards of the DNA extraction protocol.

Chelex method of DNA extraction

1. Make 5-7% solution of chelex, aliquot 300 µl in 1.5 ml Eppendorf tubes and refrigerate.
2. Take 300 µl blood in an Eppendorf tube and add fill up to top with distilled water or RBC lysing solution (Table 2.1) to lyse the red cells.
3. Centrifuge at 5000 rpm for 2 minutes to pallet the white cells.
4. Repeat the red cell lysis step if the white cell pallet contains too many red cells.
5. Add 300 µl 5-7% chelex solution to the white cell pallet and vortex for 15-20 seconds.
6. Place the tube in a heating block at 95^{o}C for 20 minutes.
7. Vortex for 15-20 seconds.
8. Centrifuge at 10,000 rpm for 2 minutes.
9. Transfer the supernatant to a fresh Eppendorf tube and use as source of DNA.
10. The DNA extracted by Chelex method may contain some residual haemoglobin especially when the white cell pallet contains red cells. Such DNA may give excessive background fluorescence in real time PCR applications.

Commercial kits for DNA extraction

A large number of commercial kits are available for DNA extraction. These methods are quick and provide consistently good quality DNA. Methods using silica columns or magnetic beads can be used on automated equipment for high throughput DNA extraction.

DNA Extraction from archival bones

Extraction of DNA from dried bones is always challenging. DNA is present in the osteocytes located in the bony cortex from where the cells are freed by filing and converting the dense bone to fine powder. The archival bones recovered from graves etc. are also heavily contaminated by dust and other PCR inhibitors.

Therefore, the bones must first be cleaned to remove any possible contaminants.

The following protocol gives reasonably good results.

1. Wash the bone surface with 0.5-1 M EDTA and rinse in distilled water. Dry the bone in air and choose a thick cortical portion from a long bone like femur. Spongy soft bones usually harbor contaminants and should be avoided for DNA extraction.
2. Convert the cortical bone to fine powder by using a saw or a file. The process of filing is done gently and carefully to avoid generation of heat that may degrade DNA. Usually, 5-10 grams of bone powder is enough for processing.
3. Decalcify the bone particles by adding sufficient amount of 0.5M EDTA to the bone powder and leaving at room temperature for 48 hours. Frequent agitation or vortexing helps this process.
4. Centrifuge for 3 minutes at 13000 rpm to remove the supernatant EDTA solution.
5. Wash the bone particles twice in distilled water.
6. Add 0.5 ml lysis buffer (Proteinase K 20mg/ml, 10µl 1M Tris-HCl, 2µl 0.5M EDTA, 100µl 10% SDS and 200µl distilled water). Incubate at 56^{o}C overnight. If bone particles are not completely dissolved the step may be repeated with addition of fresh Proteinase K until the bone particles are completely dissolved. This may take 2-3 days.
7. The further steps of DNA extraction of DNA from the dissolved bone are the same as described in the section of DNA extraction from blood by phenol chloroform method.

<u>DNA extraction from paraffin embedded tissues</u>

Extraction of DNA from paraffin embedded tissue blocks that are fixed in formalin is difficult. DNA can be extracted by taking 2-3 microtome sections in xylene to dissolve the wax. The tissue is air dried and can be processed as for

the fresh tissue or the chelex protocol described above. Many commercial kits are also available for extraction of DNA from paraffin embedded tissues that give consistently good quality results.

Measurement of DNA concentration

The concentration of DNA in a PCR is critical. Too much or too little DNA can affect the amplification. Most PCR applications work well at DNA concentration between 100-200ng/µl. This concentration can be achieved by following the guidelines given in the extraction protocol. PCR applications using genetic analyzers and Next Generation Sequencing (NGS) work at very low DNA concentration (~1ng/µl). These applications require exact quantification of DNA should.

Optical density (OD) method

DNA and RNA absorb UV light at 260nm. The OD of DNA solution measured at 260nm can be used to calculate the concentration of DNA or RNA.

Make 1:100 dilution of DNA in distilled water (20µl + 2ml)

Take OD at 260nm

DNA concentration (ng/µl) = 50 x dilution factor x OD

Example:

OD at 260nm: 0.068

Concentration: 50 x 100 x 0.068 =

340 ng/µl or 0.340 µg/ml

The optical density method can also be used to determine the protein content of the DNA. The proteins leftover from the extraction procedure can interfere in PCR. It is sometimes required to know the purity of the extracted DNA. Proteins absorb UV light at 280nm. In a good DNA sample the ratio of OD at 260nm and 280nm should be above 1.8. Ratio below 1.8 indicates protein contamination in the DNA solution.

Fluorometry

Commercial kits based on fluorescent dyes like SYBR Green etc. can be used for DNA quantification. The green fluorescence emitted by the DNA standards of known concentration are used to plot a standard curve against which any unknown DNA concentration in a sample can be measured. Qubit® is a popular fluorometry based instrument that is commonly used for DNA quantification.

Real time PCR

Accurate DNA quantification can be done by real time PCR. The unknown sample is amplified with SYBR green or the TaqMan® probe method along with tenfold dilutions of a DNA standard of known concentration. A standard curve is plotted with Ct values and concentration of the DNA standard. Concentration of the unknown sample is calculated from the standard curve (Chapter 5).

Gel electrophoresis

The DNA may also be quantified after gel electrophoresis but the method only gives a rough estimate of the quantity. The DNA in an unknown sample is amplified. The amplified DNA is run on a gel along with a commercially available 100bp ladder. The intensity of the amplified DNA bands of the unknown sample is compared with that of the bands of the allelic ladder. An approximate estimate of the DNA quantity in the unknown sample is made by comparison with the bands of the known concentration of DNA in the allelic ladder.

Quality of DNA

The DNA being a long and fragile molecule can easily get broken into pieces due to vigorous shaking during extraction. The DNA can also be fragmented due to digestion by enzymes derived from bacterial contamination in a biological sample. The fragmented or degraded DNA is difficult to amplify especially when large amplified products are required. In many applications like forensic casework or sequencing etc. it is sometimes essential to assess the quality of the DNA before amplification.

The best method to assess the quality of DNA is through gel electrophoresis. The DNA is amplified at an easily amplifiable locus and the amplified product is run on a gel. A sample with fragmented DNA would give a smearing effect (Figure 2.1).

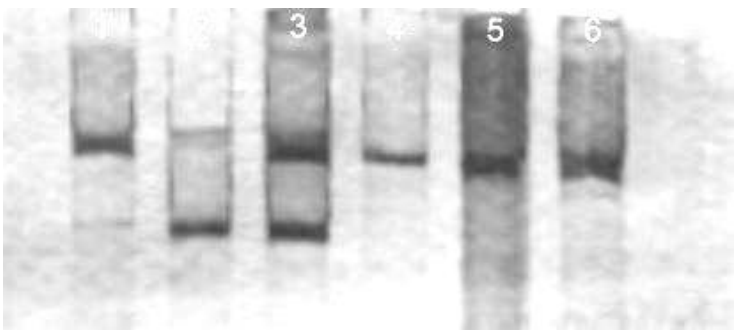


Figure 2.1. The assessment of quality of DNA by PCR amplification and gel electrophoresis. The sample in lane 5 shows fragmented DNA that appears as dark brown smearing in the background of the amplified DNA band. The samples in lane 1, 3 and 6 also show increase in the background that is due to an excess DNA in the sample resulting in over amplified PCR products. This can be improved by diluting the DNA. Lane 4 shows the best result with a clean amplification product (sharp band) and minimal background.

RNA Extraction

RNA is easily degraded by the RNA digesting enzymes (RNase) present in the environment. Since RNA isolation procedure takes place in a strong denaturant that renders RNase inactive, the integrity of RNA is mostly at risk either before or after the extraction. Therefore, handling of the sample prior to extraction and storage of RNA after extraction are critical.

A single step RNA extraction reagent is commercially available as TRI reagent. It is a phenol-based reagent that contains a combination of denaturants and RNase inhibitors. The RNA is separated from DNA by centrifugation after extraction with an acidic solution containing guanidinium isothiocyanate,

sodium acetate, phenol and chloroform. The total RNA remains in the upper aqueous phase, while most of the DNA and the proteins remain either in the inter-phase or in the lower organic phase. The RNA is recovered from the aqueous phase by precipitation with isopropanol. The extracted RNA is re-suspended and stored in RNase free water.

RNA extraction by TRIzol® Reagent

1. Mix 0.75 ml of TRIzol® Reagent (Invitrogen, USA) with 0.25 ml of sample and lyse cells (or cellular debris) suspended in the sample by passing the suspension several times through a pipette. Use at least 0.75 ml of the TRIzol® reagent per 5-10 x 10^6 cells.
2. If the sample volume is <0.25 ml, adjust the volume to 0.25 ml with water. The volume ratio of TRIzol® reagent to sample should be 3:1.
3. Keep the lysate/homogenate for 5 minutes at room temperature to permit complete dissociation of nucleoprotein complexes. At this stage the samples can be stored at -70^{o}C for at least one month.
4. Add 0.2 ml chloroform (free of isoamyl alcohol or any other additive) per 0.75 ml of TRIzol® reagent.
5. Cover the samples tightly and shake vigorously for 15 seconds.
6. Keep the mixture at room temperature for 2-15 minutes depending on the number of cells in the sample.
7. Centrifuge the mixture at 12,000 g for 15 minutes at 4-10^{o}C. Centrifugation at higher temperature may result in DNA contamination of the aqueous RNA phase making it unsuitable for PCR.
8. The mixture separates into a lower red phenol-chloroform phase, interphase and a colourless upper aqueous phase. RNA remains exclusively in the aqueous phase whereas DNA and proteins are in the interphase and organic phase respectively. The volume of the aqueous phase is about 70% of the volume of TRIzol® reagent used for homogenization.

9. Transfer the aqueous phase to a fresh tube. Interphase and organic phase may be used for subsequent isolation of DNA and proteins.
10. Precipitate RNA from the aqueous phase by mixing with isopropanol. Use 0.5 ml of isopropanol per 0.75 ml of TRIzol® reagent used for the initial homogenization.
11. Keep at room temperature for 5-10 minutes and centrifuge at 12,000 g for 8 minutes at 4-25°C. RNA precipitate (often invisible before centrifugation) forms a gel-like or white pellet on the side and bottom of the tube.
12. Discard the supernatant and wash the RNA pellet with 1 ml 75% ethanol. The RNA precipitate can be stored in 75% ethanol at -20°C for up to one year.
13. Vortex and centrifuge at 12,000 g for 5 minutes at 4-25°C.
14. Discard the ethanol and dry the RNA pellet for 3-5 minutes at room temperature. It is important not to completely dry the RNA pellet as this will greatly decrease its solubility.
15. Dissolve RNA in DEPC treated water or 0.5% SDS by passing the solution a few times through a pipette tip and incubating for 10-15 minutes at 55-60°C.
16. The total RNA is essentially free of DNA and proteins and should have a 260/280 ratio of 1.6-1.9.
17. Hands and dust are a major source of the RNase contamination. Use gloves and keep the tubes closed throughout the procedure.

<u>RNA extraction by silica columns and magnetic beads</u>

RNA can be extracted by using commercially available kits that use silica columns or magnetic beads. For samples with very low concentration of RNA, for example viral RNA, carrier RNA is also used. These methods can be used for high throughput automated RNA extraction.

Table 2.1. Reagents used in DNA extraction.

Red cell lysis buffer

- Sucrose: 109.5 g
- Tris (pH 7.6): 1.58 g
- $MgCl_2$: 476 mg
- Triton-X: 10 ml
- Sodium azide: 200 mg
- Distilled water: up to 1L

Cell lysis buffer

- Tris (pH 8.0) 7.85 g
- Disodium EDTA: 6.68 g
- SDS: 20 g
- Distilled water: up to 1L

Cell lysis buffer with guanidine

- Guanidine isothiocyanate: 50 .0 g
- SDS 2.0 g
- IM Sodium citrate (pH 7.0): 2.5 ml
- 2-Mercaptoethanol: 0.7 ml

Buffered phenol

- Phenol 250 g
- Distilled water 40 ml
- Place at 65° C for 1-2 hrs
- Cool and add 300 mg 8-hydroxyquinoline
- Equilibrate with equal volume of 1M Tris buffer (pH 8.0)
- Remove the supernatant after allowing phenol to settle down
- Repeat twice equilibration with 1M Tris
- Add 0.4 ml 2-mercaptoethanol
- Add 100 ml of 0.1M Tris buffer (pH 8.0)
- Store at 4° C in a dark bottle

Proteinase K

- Proteinase K: 20 mg
- Distilled water: 1 ml

- Make aliquot of 0.5 ml and store at -20°C
- Use 20 µl/extraction

7.4 M Ammonium acetate

- Ammonium acetate: 57.0 g
- Distilled water: up to 100ml

Bibliography

1. Chomczynski P (1993) A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. BioTechniques 15: 532-537.
2. Kleines M, Schellenberg K, Ritter K (2003) Efficient extraction of viral DNA and viral RNA by the chemagic viral DNA/RNA kit allows sensitive detection of Cytomegalovirus, Hepatitis B virus, and Hepatitis G virus by PCR. J Clin Microbiol 41: 5273–5276.
3. Green MR, Sambrook J (2012) Molecular Cloning: A Laboratory Manual. (Fourth Edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

3

Polymerase Chain Reaction

Polymerase chain reaction (PCR) is used for amplification of DNA. The double stranded DNA is denatured by heating at ~94^{o}C. The two strands separate from each other and each strand acts as a template for the synthesis of a new strand. The point of synthesis on each strand of the DNA is marked by a primer (short sequence of DNA) complementary to the DNA sequence. The PCR requires at least two primers; one for the forward and the second for the reverse strand of the DNA. The two primers are allowed to anneal with the target regions at around 60^{o}C. In the presence of heat stable Taq DNA polymerase, deoxy-nucleotides (dNTPs), $MgCl_2$ and a reaction buffer, the synthesis of new strands takes place at 72^{o}C. During the synthesis step the dNTPs are added to the target template. At the end of the first cycle two molecules of double stranded

DNA are formed from a single starting molecule (Figure 3.1). If a similar three step cycle of heating and cooling is repeated the newly synthesized double stranded DNA molecules act as templates for the further DNA synthesis and from two molecules four molecules of DNA are formed.

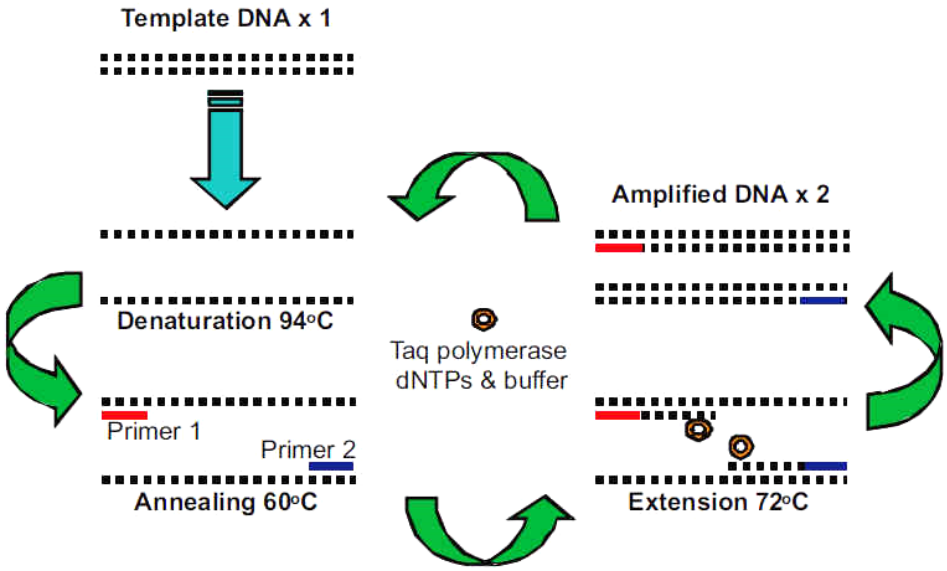


Figure 3.1. The principle of Polymerase Chain Reaction (PCR). At the end of one cycle of sequential heating and cooling one molecule of template DNA is amplified to produce two molecules of DNA. In each successive cycles the newly formed (amplified DNA) molecules act as templates for further amplification.

Sequential repetition of the cycles comprising denaturation, annealing, and extension leads to doubling of the DNA molecules with each cycle. In thirty cycles one molecule of DNA can be amplified to produce approximately one billion molecules of DNA. The number of DNA molecules produced from one molecule of target DNA can be calculated by the formula 2^n, where n is the number of cycles.

Size of the new double stranded DNA molecules is defined by the two primers used for the amplification. In the initial few cycles, the newly synthesized DNA

strands are slightly longer than the actual boundaries defined by the two primers. However, in the subsequent cycles when more and more of the newly synthesized strands are used as templates the amplified DNA fragments of exactly the same size, defined by the two primers, are produced.

The initial outburst of amplification in a PCR is gradually reduced to a plateau. The decreasing efficiency of the PCR is due to the consumption of dNTPs and primers and progressive denaturation of the Taq polymerase at very high temperatures. Many other factors including quality and quantity of the template DNA, presence of inhibitors of PCR, quality of the consumables and accuracy of cycling parameters etc. also play role in the overall performance of PCR. The efficiency of PCR is further discussed in Chapter 5.

PCR Primers

The primer is a short sequence of DNA that is one of the most critical ingredients in a PCR. The primers are also the sole determinants of the specificity of the PCR. There are several variables that can be detrimental for the overall performance of a primer.

Primer design

The PCR amplification starts when a “primer” anneals to its target. The primer, in addition to being the sole determinant of the specificity of the PCR, also has an important bearing on the sensitivity of PCR. Therefore, it is essential to understand the basic principles of designing a PCR primer.

The primers may be designed manually or by computer software packages available for this purpose. The primer designing software also provides simulation data about the performance of the primers.

Primer length

The length of a primer is measured in bases. The usual length ranges from 20-25 bases. The number of bases in a primer is often called “mer” e.g., 20 or 30 mer. A short primer is more likely to anneal non-specifically to the sequences that are even partially identical to it. Increasing the length of the primer increases its specificity but reduces the sensitivity.

Primer ends (5’ and 3’)

The primers, like any other DNA sequence, have a 5’ (5 prime) and a 3’ (3 prime) end. For example, the primer 5'-CAATGTATCATGCCTCTTTGCA begins at “C” (5’ end) and ends at “A” (3’ end).

The 3’ end of the primer is the most critical region as the extension (addition of dNTPs) occurs at this end. The last few bases, especially the last base, at the 3’ end are the most critical in annealing of a primer. A mismatch at the last base could seriously impair the ability of a primer to anneal. It is also important that the last five bases at the 3’ end should not have more than three G or C. The 5’ end of the primer is less important. This end is often used for tagging the primers with fluorescent dyes etc.

Primer direction (forward and reverse primers)

The two strands of DNA run in opposite directions. Similarly, the primer and the template DNA also run in opposite directions. In amplification of double stranded DNA one primer each is used for the two strands i.e., one for the forward strand and the other for the reverse strand. It is customary to name the two primers by their direction. The primer that anneals with the reverse strand of the DNA is called the forward or the sense primer while the primer that anneals with the forward strand is called the reverse or the ante-sense primer.

Example:

DNA is a double stranded molecule but for writing its sequence only one of the strands, usually the forward strand, is written (Chapter 1, Figure 1.4). Figure 3.2 shows marking of the two PCR primers on a single strand of DNA. The primer marked in bold normal is actually meant to anneal with the other strand. The location of the primer for the other strand is marked in bold italic. The actual sequence of this primer can be understood if the opposite strand is also written.

Sequence of primer for the opposite strand should be reversed as follows:

Actual sequence: 5’-GAGTTTTCATCCATTCTGTCCTG

Reversed sequence: 5’-CAGGACAGAATGGATGAAAACTC

TTTAGACATAATTTATTA**GCATGCATGAGCAAATTAAGA**AA
AACAACAACAAATGAATGCATATATATGTATATGTATGTGT
GTA**[C/T]**ATATACACACATATATATATATATTTTTTCTTT
TCTTACCAGAAGGTTTTAATCCAAATAAGGAGAAGATATGC
TTAGAACCGAGGTA***GAGTTTTCATCCATTCTGTCCTG***TAAG

Figure 3.2. The two PCR primers i.e. forward (bold normal) and the reverse (bold italic) are usually marked, although incorrectly, on the same strand of DNA.

Primer sequence

The sequence of nucleotides (G, A, C & T) in a primer that is complementary to the target is called its sequence. In an ideal primer the numbers of G, A, C and T should be equal. But in practice it is not always possible. However, the total GC content in a primer should remain between 40-60%. Too many G and C can lead to formation of secondary structures and impair the performance (Chapter 1, Figure 1.5). The repeats like ATATAT in a primer can lead to nonspecific annealing and should be avoided.

Primer Tm and annealing temperature

The melting temperature (Tm) is the temperature at which half of the primer is annealed to its target. Tm is dependent on the nucleotide sequence of the primer. Greater the numbers of G and C in a primer higher is its Tm. There are many formulae that can be used for calculation of Tm. For example:

Tm = 2 x AT + 4 x GC

(AT is the sum of A and T nucleotides, and GC is the sum of G and C nucleotides).

The primer designing software and primer vendors also provide data on the Tm of a primer. It is essential that two or more primers being used together in a PCR should have nearly equal Tm. The optimum annealing temperature of a primer is generally 3-5^{o}C below the theoretically calculated Tm.

Primer dimers and secondary structures

Partial complementarity in the sequence of primers can make them anneal with each other to form primer dimers. This may include self-dimers (between themselves) or cross-dimers (with other primers) (Figure 3.3). The primer

dimerization may be difficult to control in a multiplex PCR (PCR with more than two primers). Primers may also form secondary structures like loops and hair pins (Figure 3.4). Strong primer dimers and secondary structures reduce the bioavailability of a primer leading to impaired amplification. The problem can become serious when primers are designed manually. The primer designing software is useful in tackling this issue.

Figure 3.3. Primer dimers. (a) cross-dimer between forward and the reverse primers (b) self-dimer between the two copies of the forward primer.

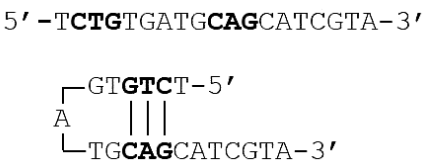


Figure 3.4 Example of hairpin secondary structure of a short sequence of DNA (primer).

Primer cross-homology

A good primer should be specific for its own target. Occasionally a primer may have partial or rarely complete sequence homology with another site in the genome. This would cause nonspecific amplification. The best way to test for the primer cross-homology is through the Primer BLAST tool of the NCBI website (Chapter 1).

Primer designing by computer software

The primers can be designed on computer software or by manual method. For some applications like allele specific PCR the primers are better designed manually. Many commercial software packages are available with numerous options. A simple free of cost online primer designing software “Primer3plus” is available:

http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

The sequence for which the primers are to be designed can be uploaded as a file or it can be pasted in the software window. The software provides options for marking a particular region of the sequence. The software also provides options for a large number of other useful parameters.

Primer labeling

Primers are often required to be labeled with dyes like fluorochromes etc. All such labels are attached at the 5’ end which is also the non performing end of the primer. A typical fluorescent dye “FAM” labeled primer would be written as:

5’ FAM-CAGTAGCATCTGACTTTGAGCCTCAGGGTCT

Primer purification

During the process of primer synthesis many short length sequences are also synthesized. This is especially common with primers longer than 30 bases in length. The short length sequences in a primer can cause nonspecific amplification. The latter also consumes the other reaction ingredients like dNTPs and Taq and lowers the specific amplification. The presence of short length primers can also falsely elevate the final concentration of the working primer solution. In applications where it is critically important to get only the specific amplification the primers may have to be purified before use. The primer manufacturers usually provide free of cost gel-based purification of the primers. In critical applications like diagnostic PCRs extra purification by HPLC may be required especially when the primers are longer than 30 bases.

Primer concentration and dilution

An optimum concentration of the primer is critical in a PCR. Too much primer causes it to anneal non-specifically whereas too little primer reduces its availability for annealing and the amplification. Almost all PCRs are done with at least one pair of primers. It is important that the two primers of a pair are used in equi-molar ratios. The primer concentration is usually determined by the optical density (OD) of the primer at 260 nm. In general, 1µg of a primer with 25 bases has an OD of 25. The primer manufacturers provide the concentration of the primers in µg and OD when reconstituted in one ml of water. The primer is usually supplied in lyophilized form that is reconstituted in water and is diluted as per requirement. The lyophilized primer is stable at room temperature for several weeks. But once reconstituted it is stable at 4^{o}C for a couple of weeks and for over a couple of years at -20^{o}C. The lyophilized primer is usually diluted in 1ml of DNase free water and it is best to store it in aliquots at -20^{o}C.

It is convenient to use primer concentration in micro-moles (µM) or pico-moles per micro-liter (pM/µl). The optimum concentration of a primer in most PCRs ranges between 5-10 pM/µl. The following example can explain the dilution of a lyophilized stock primer to make a final concentration of 5pM/µl:

1. OD of the lyophilized stock primer: 13.5 (provided with the stock primer vial).
2. Reconstitute the lyophilized stock primer by adding 1.0 ml DNase free water.
3. Gently mix by inverting several times.
4. Aliquot and store the stock primers at -20^{o}C.
5. To make 100µL of 5pM/µl primer use the following volumes:
    - Volume of the stock primer:
    - 1/OD x volume required (1/13.5 x 100 = 7.4µl)
    - Volume of distilled water: (100-7.4 = 92.6µl)
6. Using 1µl of the diluted primer will give a final concentration of 5pM per reaction.

7. To make 10pM/µl concentration, double the volume of the stock primer (14.8µl stock primer + 85.2µl water).
8. Two or more primers to be used in a PCR reaction may also be diluted together in one vial. For example, four primers with ODs 10.5, 13.4, 12.5 and 15.1 may be diluted together in one vial, at concentration of 5pM each in 100µL final volume, as follows:
    a. Calculate the volume of the individual primers as above.
    b. Add the volume of the four stock primers (1/10.5 x 100) + (1/13.4 x 100) + (1/12.5 x 100) + (1/15.1 x 100).
    c. Add distilled water equal to 100 – volume of the four stock primers.

Ordering the primer for synthesis

The designed primers are made on a DNA synthesizer. The primers are best synthesized by good commercial company. The request for primer synthesis should include the sequence of the primer, any extra purification, and the synthesis scale. The primers are usually synthesized at a scale of 25, 50, 100 or 200 nM. The labeling of the primer, if required, is also indicated. A synthesis scale of 50 nM is usually optimum for most applications.

Taq polymerase

Synthesis of DNA requires a polymerase. Since the temperature in a PCR is raised to 94^{o}C an ordinary polymerase would be denatured at this temperature. The problem is overcome by using a DNA polymerase derived from *Thermus aquaticus,* that grows in hot water springs. Optimum temperature of the Taq polymerase is 72^{o}C.

The *Taq polymerase* also has 5’ to 3’ exonuclease activity that can remove any dye attached to the 3’ end of a primer. This function is used in removing a dye/quencher attached to the 3’ end of a TaqMan® probe used in real time PCR (Chapter 5).

Tth DNA polymerase

Tth is another heat stable DNA polymerase that is derived from *Thermus thermophilus*. The enzyme has dual function of reverse transcriptase and DNA polymerase. This is useful in applications where cDNA synthesis from RNA and its amplification are required in the same reaction tube.

High fidelity *Taq polymerase*

The *Taq polymerase* lacks the 3' to 5' exonuclease activity and therefore is unable to proof read any errors in DNA synthesis. The enzyme *Pfu DNA polymerase* by virtue of its 3' to 5' exonuclease activity can also do the proof reading. It is especially useful in doing PCR of very long templates (high fidelity amplification) or in sequencing of DNA where errors in PCR can introduce unwanted nucleotide bases in the results.

Deoxynucleotide phosphates (dNTPs)

Deoxynucleotide phosphates (dNTPs) are the building blocks of DNA synthesis. These are available from many commercial sources as a set of four nucleotides i.e., dATP, dTTP, dGTP and dCTP. These are usually available in 10 or 100mM concentrations. The four nucleotides are also available as a premixed solution. If a set of four separate nucleotides is purchased then these should be mixed in equal proportions and stored as aliquots at -20^{o}C or -80^{o}C.

Magnesium chloride

Mg^{++} ions increase the solubility of dNTPs and facilitate their incorporation in the template. These ions also stimulate the Taq polymerase. Therefore, the concentration of $MgCl_2$ in a PCR is critical. The amplification can be enhanced by increasing the concentration of Mg^{++}. The optimum results are achieved at $MgCl_2$ concentration between 1.5-2.0 mM.

PCR buffer

PCR buffer is mostly supplied with the Taq polymerase. It can also be prepared in the lab. The following buffer formulation gives consistently good results in a wide range of PCR applications:

Stock Solutions:

2M KCl:	14.9 g/dl
1M Tris:	12.1 g/dl
1M $MgCl_2$:	9.52 g/dl
1M Spermidine:	14.52 g/dl

10 X PCR Buffer

2M KCl:	1.25 ml
1M Tris (pH 8.3):	0.5 ml
1M $MgCl_2$:	75 µl
Gelatin (300 bloom):	5 mg
Distilled water:	3.2 ml

PCR mix (1 ml)

10 X PCR buffer:	50 µl
100 mM dNTPs (all):	2.5 µl
1M Spermidine:	1.0 µl
Distilled water:	946.5 µl

Inhibitors and enhancers of PCR

Haemoglobin, heparin, porphyrin, SDS, phenol, and Proteinase-K are potent inhibitors of PCR. SDS, and phenol must be completely removed during DNA extraction. Proteinase K can be inactivated by heating. PCR may also be inhibited by chocolate that can be present in the DNA extracted from a mouthwash. Enhancers of PCR include formamide, DMSO, tetra methyl-ammonium chloride (TMAC), polyethylene glycol (PEG), glycerol, Tween 20 and 7-deaza-dGTP.

Setting up a PCR

An example of amplification of β-globin gene containing C-T polymorphism is described. The sequence of the gene was downloaded from the NCBI website (GenBank accession no. NG_000007.3) and the primers flanking the C-T polymorphism were designed using the Primer3plus software (Figure 3.2).

Locus: β-globin gene

GenBank accession: NG_000007.3

Forward primer: 5'-GCATGCATGAGCAAATTAAGA

(Length: 21, Tm: 59°C, GC 38%)

Reverse primer: 5'-CAGGACAGAATGGATGAAAACTC

(Length: 23, Tm: 60°C, GC 43%)

Amplified product size: 179bp

PCR master mix

For 10 reactions (7 test samples + 1 positive control + 1 negative control + 1 reagent blank) make 260 µl master mix (25 µl/reaction + 10 µL extra for pipetting errors). Mix the following in a 0.5 ml plastic tube:

- PCR buffer with dNTPs: 240 µL (25 µL per reaction)
- Forward primer 10 µL (5 pmol/µL) (1 µL per reaction)
- Reverse primer 10 µL (5 pmol/µL) (1 µL per reaction)
- Taq polymerase (5 units/µL) 1.0 µL (0.1µL/reaction)
- Template DNA (~200 ng/µL): 1 µL/reaction

1. Take 10 x 0.2 ml PCR reaction tubes in a rack and label these 1-10.
2. Dispense 25µL master mix to each of the ten labeled tubes.
3. Add 1 µL of the test DNA samples to the tubes 1-7.
4. Add 1 µL of control DNA positive for the C-T polymorphism in tube 8.
5. Add 1 µL control DNA Negative for the C-T polymorphism in tube 9.
6. Add 1µL water instead of DNA in tube 10 (reagent blank).
7. Firmly close the lids of the tubes and put these in the thermal cycler.
8. Programme the thermal cycler and run PCR as follows:
    a. Initial denaturation 94°C for 1 minute.
    b. Thirty cycles each of denaturation at 94°C for 1 minute, annealing at 60°C for 1 minute and extension at 72°C for 1 minute.
    c. Final extension at 72°C for 3 minutes.

9. Electrophoresis:
    - 10 x 10 cm 6% polyacrylamide, 30 minutes at 150 volts.
10. Staining:
    - 0.1% silver nitrate

DNA Thermal Cycler

DNA thermal cycler is a machine that can be programmed to rapidly heat and cool the PCR reaction vials. A wide variety of machines are available from different sources. The choice of machine depends on the requirement of a lab. An average thermal cycler can heat or cool from 1-5^{o}C per second. More expensive machines with very fast heating or cooling (ramping) are also available. Very fast ramping can shorten the time required in thermal cycling. The fast machines also require special thin walled PCR tubes that can allow heat transfer at an equally fast rate. Most PCR machines have a 96 well format. Smaller machines with a capacity of 8, 16, 32, or 48 well are also available (Figure 3.5).

Figure 3.5. A typical DNA thermal cycler has a heating block, heated lid, keypad and an LCD display.

Unusual Forms of PCR

Multiplex PCR

PCR is normally done with a pair of primers targeting the sequence of interest. In multiplex PCR multiple targets are amplified by using more than one pair of primers or one forward and multiple reverse primers. However, when more than two primers are used then their cross homology may cause excessive primer dimerization and reduced bioavailability. This may result in reduced amplification or even false negative results. Therefore, multiplexing should be done with great care. The main advantage of multiplex PCR is reduction in cost and time.

Hot start PCR

At the start of a PCR when temperature in reaction vial is not high the primers may anneal nonspecifically to the partially denatured template resulting in nonspecific amplification. This can be prevented by "Hot Start PCR". The Taq polymerase is tagged with an anti-Taq antibody making it unavailable for the amplification. The PCR begins with an initial step of denaturation at 95^{o}C for a few minutes. The anti-Taq antibody is denatured at this temperature and the Taq becomes available for PCR. The Taq may also be inhibited by some covalently bound inhibitors that only dissociate after a high temperature activation step. The "hot start" step effectively prevents bioavailability of the Taq at lower temperatures and prevents non-specific amplification.

Nested PCR

When the target DNA is very low in concentration the amplification can be done in two steps. In the first step a pair of primers is used to amplify the target. In the second step the amplified product of the first step is re-amplified with another set of primers located within the outer/flanking primers used in the first step. Nested PCR is extremely sensitive and needs special handling protocol. The first step amplified product can easily contaminate the other PCR reagents giving false positive results. Such a situation can be very difficult to handle and it is best avoided.

Asymmetric PCR

Some PCR applications, like sequencing, require yield of only the single stranded DNA. This is achieved by using only one of the pair of primers. With each successive PCR cycle only one of the strands is reproduced.

Touchdown PCR

In a touchdown PCR the annealing temperature of the primer is kept 3-4^{o}C higher in the initial few cycles. This improves the specificity of the reaction. In the later cycles the annealing temperature is lowered to the actual allowing more efficient amplification.

Digital Droplet PCR

A traditional PCR is carried out in a single reaction per sample. Whereas in a digital droplet PCR the sample is separated into a large number of partitions (droplets) and the reaction is carried out in each partition individually. The separation into droplets allows greater sensitivity in nucleic acid quantification.

The method is useful for studying:

1. Absolute Quantification
2. Detection of rare target sequences
3. Gene expression and micro-RNA analysis
4. NGS
5. Single cell analysis (low copy number analysis)

Rapid PCR

A conventional PCR takes about 1-2 hours to complete. The overall duration of PCR can be reduced by increasing the concentration of the reactants (buffer, Taq and primers) and by decreasing the ramp and the step times. An important limitation in reducing the ramp time is the plastic ware that does not allow rapid conduction of heat. In some PCR instruments substituting the plastic tubes by glass capillaries has reduced the PCR times to around 30 minutes. In an extreme example PCR has been shown to be completed in just 15-60 seconds.

In-situ PCR

In-situ PCR is a technique by which the cellular DNA and RNA targets in a frozen or paraffin embedded tissue section can be amplified. The amplified DNA can be visualized by hybridization with suitably labelled probes. The in-situ PCR, because of the technical difficulties, has been largely replaced by PCR amplification of DNA or RNA extracted from the micro-dissected tissue areas in a slide section. The amplified DNA can be used for further analysis like gene profiling and sequencing etc.

Bibliography

1. Bagasra O. Protocols for the in-situ PCR-amplification and detection of mRNA and DNA sequences. Nature protocols. 2007 Nov;2 (11):2782.
2. Beck J, Bierau S, Balzer S, Andag R, Kanzow P, Schmitz J, Gaedcke J, Moerer O, Slotta JE, Walson P, Kollmar O. Digital droplet PCR for rapid quantification of donor DNA in the circulation of transplant recipients as a potential universal biomarker of graft injury. Clinical chemistry. 2013 Jan 1: clinchem-2013.
3. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 'Touchdown'PCR to circumvent spurious priming during gene amplification. Nucleic acids research. 1991 Jul 25;19(14):4008.
4. Farrar JS, Wittwer CT. Extreme PCR: efficient and specific DNA amplification in 15–60 seconds. Clinical chemistry. 2014 Jan 1: Clinchem-2014.
5. Kellogg DE, Rybalkin I, Chen S, Mukhamedova N, Vlasik T, Siebert PD, Chenchik A. TaqStart Antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. BioTechniques. 1994 Jun; 16(6):1134-7.
6. Newton CR and Graham A (1994) PCR. Bios Scientific Publishers, Oxford, UK.
7. Nuovo GJ. PCR in situ hybridization: protocols and applications. Raven Press; 1994.
8. Saiki RK, Gelfand DH, Stofel S, Scharf R, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487
9. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35: W71-W74.

4

Electrophoresis of DNA

Electrophoresis is used for the separation of DNA, RNA, or other charged molecules like proteins by an electric current applied to a gel matrix. In PCR applications it is usually done for separation and visualization of amplified DNA. The electrophoresis gel is a porous matrix that allows movement of DNA through its pores. DNA fragments with large differences in size are typically separated on agarose gels whereas small differences in DNA fragment sizes are resolved on polyacrylamide gels.

The nucleic acids have a negative charge due to the sugar phosphate backbone of the molecule. This allows their movement from the negative to the positive

electrode. The double-stranded DNA fragments naturally behave as long rods, so their migration through the gel corresponds to their size. On the other hand single-stranded DNA or RNA tends to form secondary structures with complex shapes and these migrate through the gel in a complicated manner. The formation of secondary structures can be overcome by adding denaturing agents like formamide or urea in the gel. As a result, the single stranded nucleic acids behave like straight rods and their movement becomes more predictable and size dependent.

Agarose gel electrophoresis

Agarose is composed of long un-branched chains of carbohydrate. Since there are no cross links between chains these gels have large pores and are suitable for separation of larger molecules of DNA. NuSieve™ agarose is a special product with smaller pore size and better resolving power. The agarose gels are typically run in horizontal tanks after submerging in a buffer solution hence the name submarine gels (Figure 4.1). The agarose gels are unable to withstand high voltage because the heat generated during electrophoresis may cause melting of the gel.

Preparing 2% agarose gel (14 x 10 x 0.5 cm)

1. In a 500 ml beaker take:
    - Agarose 1.6 g
    - 1 x AGB buffer 80 ml
2. Cover with a cling film.
3. Heat in a microwave oven for 2 minutes at high power.
4. Seal the ends of a gel tray with cello-tape.
5. Pour the molten agar on the gel tray.
6. Place the appropriate sample comb in the gel tray.
7. Leave to cool for 30-60 minutes. The gel will solidify and appear opaque grayish in colour.

8. Remove the scotch tape before running the gel.

To prepare 4% NuSieve agarose gel (14 x 10 x 0.5 cm) take the following in a 500 ml beaker:

a. Agarose: 1.6 g
b. NuSieve agarose: 1.6 g
c. 1 X AGB buffer 80 ml

Running the agarose gel

1. Fill the electrophoresis tank with 1 x TBE buffer.
2. Place the gel in the electrophoresis tank while it is still in the casting tray.
3. Carefully remove the comb(s) from the gel. If the comb is removed hurriedly or the gel has not solidified completely the well floor may be damaged.
2. Prepare the amplified DNA for loading by mixing 20µl amplified product and 5µl loading dye. The amplified DNA must be handled with care as it is a potent source of contamination for other PCR reactions. It is advisable to use a separate pipette for handling amplified DNA that should be marked “Amplified DNA only”.
3. Carefully load 20-25µl of the DNA and loading dye mixture in to the wells. The DNA in the sample quickly sinks to the floor of the well because of the glycerol in the loading dye.
4. Start electrophoresis at ~150 volts for 30-60 minutes depending on the size of the DNA fragments to be resolved. Use of higher voltage can cause melting of the gel.
5. Progress of electrophoresis can be monitored by movement of the blue coloured loading dye. Bromophenol blue moves in the agarose gel at approximately the speed of a 200bp DNA fragment.

Staining of agarose gels

The agarose gels are stained with ethidium bromide which is an intercalating dye that binds to double stranded DNA. It gives reddish orange fluorescence when exposed to ultraviolet light. The fluorescence of ethidium bromide increases 20-fold after binding to double stranded DNA. Ethidium bromide is a carcinogen and mutagen therefore it should be handled after wearing gloves.

1. Add 3-5 drops of 1% ethidium bromide solution to 500 ml of water in a staining tray.
2. Carefully remove the gel from the tray and submerge it in the stain.
3. Keep in the stain for 30 minutes.
4. Transfer the gel to 500 ml clean water and keep for 5-10 minutes. This will wash the excess background satin from the gel.
5. Place the gel on UV trans-illuminator. The amplified DNA fragments are seen as reddish orange bands.
6. A permanent record of the gel is kept by taking a picture with a camera or gel documentation system (Figure 4.2).

AGB Buffer (10 X)

Tris:	48.4 g
$NaAC.3H_2O$:	27.2 g
Disodium EDTA:	0.744 g
Glacial acetic acid:	12 ml
Deionized water:	up to 1L

TBE Buffer (10 X)

Tris:	108 g
Boric acid:	55 g
Na_2EDTA:	9.3 g
Deionized water:	up to 1L

Loading dye for agarose gels

- Bromophenol blue 5 mg
- Glycerol 10 ml
- Deionized water up to 100 ml

Ethidium bromide (1% stock solution)

- Ethidium bromide 1 g
- Distilled water 100 ml

Acrylamide gel electrophoresis

Acrylamide is a white crystalline powder that forms polymers after dissolving in water. The acrylamide polymers can be cross-linked by addition of bis-acrylamide. As a result of cross-linking the pore size of the gel also decreases. This makes acrylamide gels most suitable for separation and resolution of DNA molecules that differ in size by only one base pair. Acrylamide gels can withstand high temperatures and hence are suitable for applications requiring high voltage. Acrylamide gels are typically run in a vertical position using discontinuous buffer compartments (Figure 4.3).

Acrylamide in solution becomes viscous but does not form a gel. The cross-linked acrylamide polymers can be solidified to form a gel by adding ammonium persulphate (APS) and Tetra-methyl-ethylene-diamine (TEMED) in equimolar ratio. The concentration of APS and TMED determine the rate of gel formation and its turbidity and elasticity.

Acrylamide should be stored in a cool dark and dry place to reduce auto-polymerization and hydrolysis. Acrylamide is neurotoxic that is absorbed through skin. Its effect is cumulative and toxicity may develop over prolonged exposure. Gloves must be worn when using acrylamide.

Preparation of polyacrylamide mini gel

1. Make 6% acrylamide solution by mixing the following:

- Acrylamide 28.5 g
- N, N' methyl bis-acrylamide 1.5 g
- TBE buffer (10 X) 50 ml
- Distilled water up to 500 ml

Figure 4.1. Submarine gel electrophoresis tank for running agarose gels.

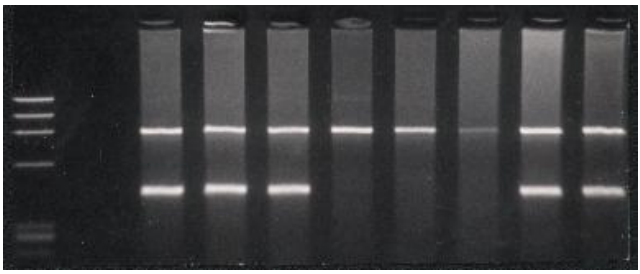


Figure 4.2. The results of ARMS PCR for thalassaemia run by agarose gel electrophoresis.

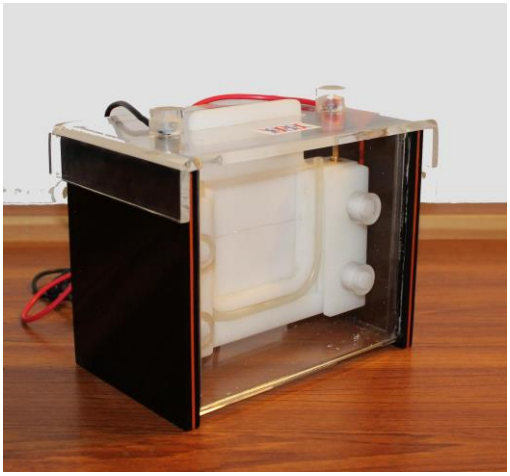


Figure 4.3. Vertical polyacrylamide gel electrophoresis tank.

2. Wash 10 x 10 cm glass plates with soap and water.
3. Dry the plates completely with cotton gauze.
4. Put 1-2 ml ethanol on the surface of the plates facing the gel and wipe the surfaces thoroughly with dry piece of gauze to remove any grease etc.
5. Assemble the glass plates as per instructions of the manufacturer.
6. Place the required comb between the plates.
7. Take 10 ml 6% acrylamide in a small beaker.
8. Add 100µl 10% ammonium persulphate (APS) not older than one week.
9. Add 20µl TEMED.
10. Gently mix and fill the polymer in a 10 ml syringe.
11. Attach a 21-gauge butterfly needle set and gently fill the polymer in the tubing and remove any bubbles in the tube.
12. Gently pour the acrylamide between the plates before the gel polymerization starts (within 2-3 minutes).
13. Remove bubbles if any by gently tapping the plates.
14. Allow up to 30 minutes for complete polymerization to occur.

15. For larger polyacrylamide gels increase the polymer and the catalysts accordingly. For example, 16 x 20 x 0.1 cm gel would take 32 ml polymer. Make about 35 ml polymer in a beaker and add 350µl APS and 70µl TMED.
16. The larger gel plates must be cleaned thoroughly to leave no trace of grease or dirt on the plates. When preparing very large gels one of the plates (not both) may be coated with silicon to allow easier separation at later stage.

Running acrylamide gel

1. Remove the gel plates from the casting assembly and fix it on the gel tank as recommended by the manufacturer.
2. Fill the upper and the lower buffer compartments with 1 X AGB buffer.
3. Gently pull the comb out and wash the wells with buffer to remove any residual un-polymerized acrylamide. A syringe with 21-gauge butterfly needle may be used for washing wells.
4. Prepare the samples for loading by mixing 2-4µl amplified product (depending on the quality of amplification) and 2-3µl of loading dye (0.05% xylene cyanol and bromophenol blue in formamide). The samples may be prepared in 96 well ELISA plates or 0.2ml plastic tubes.
5. Load 3-4µl of the prepared sample with a long nose plastic tip. The amplified DNA must be handled with care as it is a potent source of contamination for other PCR reactions. Use separate pipette marked as "Amplified DNA only".
6. Run the gel at 150-200 volts for 15-45 minutes depending on the fragment size of the amplified products.

Staining of acrylamide gels

<u>Silver staining</u>

1. Remove the gel from casting assembly.
2. Carefully separate the plates taking care to avoid tearing of the gel.

3. Mark the first well or the side of the gel by cutting a small piece from the left lower corner of the gel.
4. Put the gel in 0.1% silver nitrate solution for 15-20 minutes. The gel may be kept in stain while it is still on the glass plate.
5. Discard the stain and wash the gel in plenty of tap water.
6. The stain may be reused if kept in dark brown bottles. However, it loses potency after exposure to bright light.
7. Prepare fresh developing solution by adding 75µl formaldehyde to approximately 100 ml 1.5% NaOH. Stock NaOH is stable at room temperature but once formaldehyde is added it must be used within one hour.
8. Submerge the gel completely in developing solution.
9. In approximately 5-10 minutes the bands of amplified DNA can be seen on the gel. The background of the gel also becomes light yellowish brown. The background colour could become very dark if the gel is kept for too long in the developer solution. The exact developing time can be learnt by trial and error.
10. Discard the developing solution and wash the gel in plenty of water when the DNA bands are clearly seen.
11. Cut a piece of filter paper slightly larger than the gel itself and lay it flat on the gel surface. Gently pick the filter paper along with the gel that sticks to its surface.
12. Place the gel and the filter paper on a gel dryer making sure that the gel faces towards the front.
13. Dry the gel under vacuum for 20-30 minutes at 80^{o}C.
14. The dried gel can be pasted in a record book for long term storage after trimming its margins. It may be photographed if a gel dryer is not available (Figure 4.4).

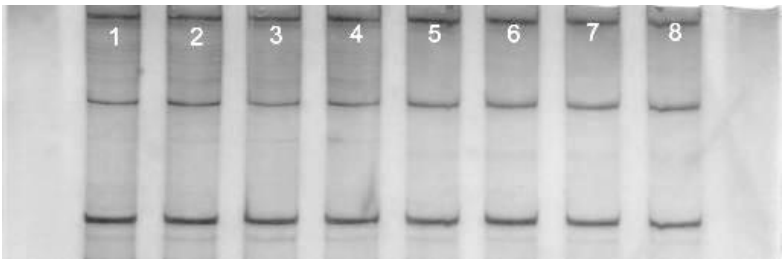


Figure 4.4. Polyacrylamide Gel Electrophoresis (PAGE) after staining in silver nitrate.

Ethidium bromide staining of acrylamide gels

The acrylamide gels can also be stained in ethidium bromide as described in section of staining agarose gels.

Capillary electrophoresis

Electrophoresis of DNA can be done in very narrow bore long and flexible capillaries. The capillaries are filled with an electrolytic solution and its ends are dipped in buffer chambers. The sample injection and running is done by electric current. The electrophoresis is usually done for short periods at 5000 to 30,000 volts (Figure 4.5).

Capillary electrophoresis is especially useful for automated fragment analysis and genomic sequencing of fluorescent labeled amplified products. Many automated genetic analyzers are commercially available that allow simultaneous running of several samples. These analyzers are extremely efficient in fragment analysis and genomic sequencing applications.

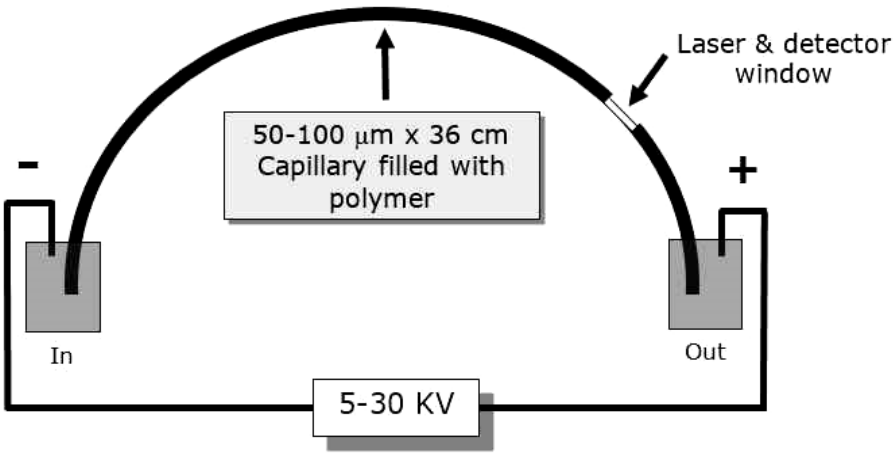


Figure 4.6. Diagrammatic representation of capillary electrophoresis.

Bibliography

1. Cai SP, Kan YW (1990) Identification of the multiple β-thalassaemia mutations by Denaturing Gradient Gel Electrophoresis. J Clin Invest 85: 550-553.
2. Constantinos Z (2009, January 18) Capillary Electrophoresis. SciTopics. Retrieved July 2, 2012, from http://www.scitopics.com/Capillary_Electrophoresis.html
3. Fodde R and Losekoot M (1994) Mutation detection by denaturing gradient gel electrophoresis (DGGE). Hum Mutation 3: 83-94.
4. Grompe M (1993) The rapid detection of unknown mutations in nucleic acids. Nature Genetics 5: 111-117.

5

Quantitative PCR (Q PCR)

PCR can be used to determine the concentration of nucleic acid in a sample. This is also called Quantitative-PCR or Q-PCR. In the past Q-PCR was done with endpoint electrophoresis of the amplified DNA. Quantification was done by examining the intensity of the amplified DNA bands on stained electrophoresis gels. Labelling of the amplified product with radioactive material followed by auto-radiography has also been used for Q-PCR. The endpoint PCR, being inaccurate, has largely been replaced by real time PCR.

Real time PCR

In a conventional PCR the amplified DNA is examined at the end of PCR whereas in a real time PCR the amplification is monitored (visualized) after each cycle and the amount of amplification is plotted against the cycle numbers. The cycle number at which the amplification becomes detectable is dependent on the quantity of the target DNA. It appears earlier when the DNA concentration is high and it appears late when the target DNA concentration is low. This makes real time PCR an excellent tool for quantitative estimation of DNA. The concentration of an unknown DNA can be calculated by comparing its result with that of a known concentration of DNA.

The amplified product of PCR is a potent source of contamination for the other PCR reactions. In a real time PCR since there is no need for the post PCR processing the chances of cross contamination are also minimized.

Monitoring of amplification

The real time monitoring of PCR is done by incorporation of fluorescent dyes in the amplified product. The fluorescence emitted from the amplified product is recorded at each cycle and the results are plotted against the cycle numbers. There are two types of methods that can be used for fluorescent labelling of the amplified product. In the non-specific method of labelling amplification is done in the presence of intercalating dyes like SYBR green that emits green fluorescence when it binds to double stranded DNA. In the more specific method fluorescent labelled molecular probes are used.

SYBR green method

SYBR green is a fluorescent dye that binds to the minor groove of double stranded DNA. It emits a strong green fluorescent signal after binding with double-stranded DNA. The SYBR green is added to the PCR reaction mixture. As the amplified double stranded DNA is produced it binds to the SYBR green and starts emitting fluorescence. The fluorescence is directly proportional to the amount of amplification. The main drawback of SYBR green is that it also binds

to DNA that is the non-specifically amplified or even the primer dimers. The SYBR green method is also not suitable for analysis of multiplex PCR products. The analysis of mixtures of DNA of various sizes or sequences by the SYBR green method can be done by a technique called DNA melting curve analysis.

Melting curve analysis

The real time PCR machines can determine the melting point (Tm) of the amplified product through a process called melting curve analysis. The Tm of a double stranded DNA depends on its base composition and the size. In the SYBR green method the specific and the nonspecific amplified products or primer dimers can be differentiated from each other by measuring the Tm through melting curve analysis. At the end of the PCR the temperature of the heating block or the heating chamber is raised in increments of 0.1-1^{o}C. At each step the Tm of the reaction products is measured by the amount of fluorescence emitted. At the end of the analysis the software plots the fluorescence readings against the temperatures (Figure 5.1).

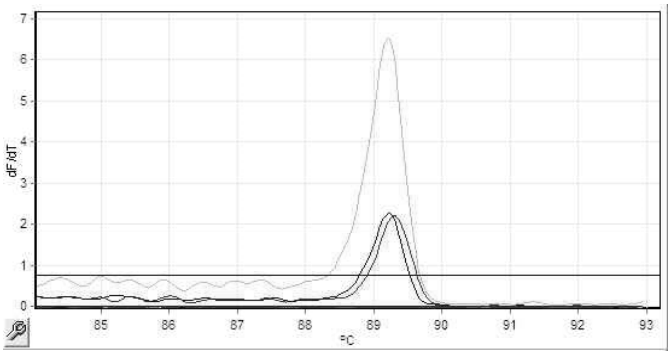


Figure 5.1. Melting curve analysis after SYBR Green real time PCR. Temperature and fluorescence are shown on the X and the Y axis respectively.

Molecular (hydrolysis) probe methods

The PCR amplification can also be monitored by a more specific method employing fluorescent labeled short sequences of DNA (probes) complementary to the target DNA. Several molecular probe based methods are available that

can be used effectively for real time monitoring of PCR. The molecular probes, because of the chemical nature of the reaction, are also called hydrolysis probes.

The fluorescence emitted in the molecular probe methods is directly proportional to the annealing of the probe. Therefore, the methods are highly specific unless the target has a cross homology with another sequence in the template DNA.

TaqMan® probe method

The TaqMan® probe is a short sequence of DNA complementary to the internal region of the PCR target DNA. The 5' end of the probe is labeled with a fluorescent dye (reporter) while its 3' end is labeled with a fluorescence quencher (suppressor). The quencher may be a high energy fluorescent dye like rhodamine (TAMRA) or a non-fluorescent chemical quencher e.g. DABCYL and black hole quenchers (BHQ). In the intact probe the reporter dye and the quencher are in close proximity to each other. This suppresses the fluorescence of the reporter dye.

In the TaqMan® method of real time PCR the target DNA is amplified by a pair of primers. The reaction mixture also contains the TaqMan® probe that emits no fluorescence when it is not annealed to its target. If the target DNA is absent there is no amplification and the probe does not emit any fluorescence. When the target DNA is present it would be amplified. During amplification the probe, being complementary to the internal region of the target, also anneals to the target. The Taq polymerase used in PCR has 5' to 3' exonuclease activity which removes the probe from the target in a stepwise manner. First the reporter dye is removed then the probe itself and finally the quencher is also removed (Figure 5.2). As a result the reporter dye is relieved from the quencher and starts emitting fluorescence on excitation by the light of appropriate wavelength. The net increase in fluorescence from the reaction is directly proportional to the amount of probe annealed to the target and hence the amplification.

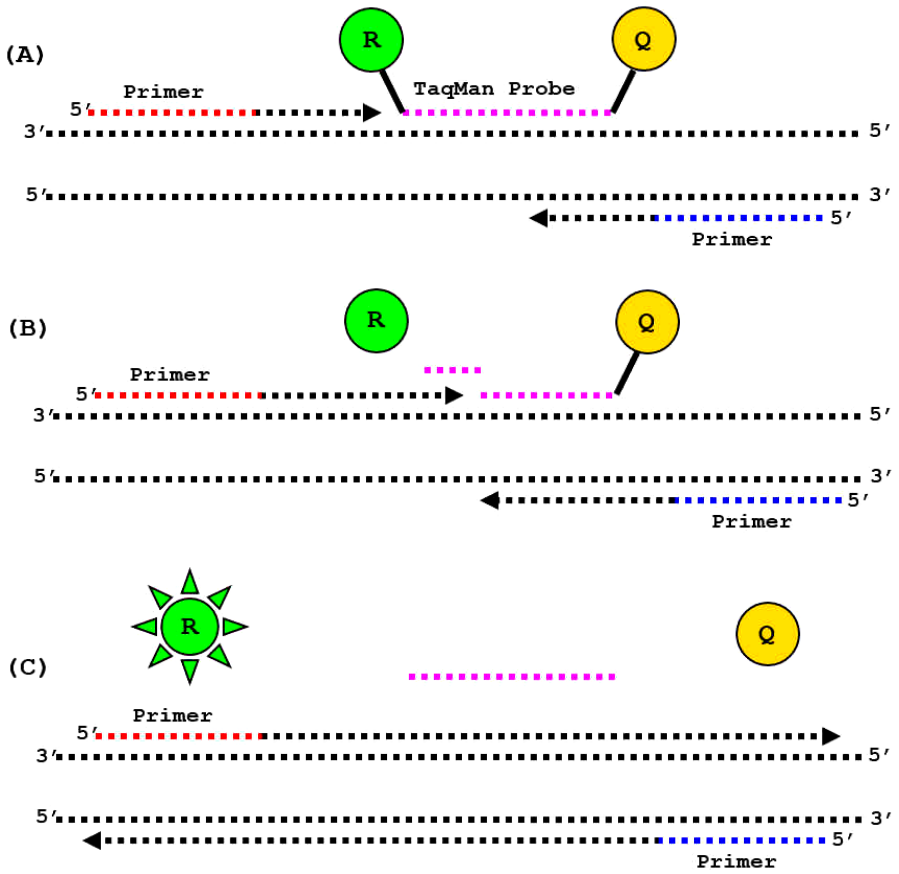


Figure 5.2. Real time PCR by TaqMan® probe method. (A) The target is amplified by primers F & R. TaqMan® probe is annealed to the target and there is no fluorescence from the reporter (R) because of the close proximity to the quencher (Q). (B) As the primer extension takes place and reaches the 5' end of the probe, the reporter (R) is released due to the 5' to 3' exonuclease activity of Taq polymerase. (C) Continued extension of the new strand removes the probe and finally the quencher (Q). As the distance between the reporter and the quencher increases the reporter dye gives a bright fluorescence on excitation that is measured by the instrument. The quantity of fluorescence is directly proportional to the amount of target and the amplification.

TaqMan® probe designing

The TaqMan® probe is designed while keeping the following parameters in mind:

1. Tm of the probe should be 10°C higher than that of the primers.

2. Runs of identical nucleotides especially “Gs” should be avoided.
3. The G+C content should be 30-80%.
4. There should be more Cs than Gs.
5. There should be no G at the 5' end.

Primers for TaqMan® assay

In addition to the general requirements of primer designing (Chapter 3) the primer pair used with the TaqMan® probe should have the following additional properties:

1. Should have no runs of four or more Gs (any nucleotide).
2. Should have no more than two G+C at the 3’ end.
3. Should have no G at the 5' end (A or C is preferred).
4. The amplicon size should range between 50-150bp (max 400).
5. Preferably should span the exon-exon junctions in cDNA.

The TaqMan® probes are best designed by computer software specially made for this purpose. Many such commercial software are available. The real time PCR machine vendors may also provide the software with the machine. The sequence of interest is uploaded in the FASTA format or it can be cut and pasted in the software window. The software offers several design options and provides a list of possible primers and the probe combinations. The best combination is chosen for a particular application.

Choice of reporter and quencher

The reporter and the quencher pair for a probe should be compatible with each other chemically as well as for the wavelength of fluorescent light. A wide variety of reporter fluorescent dyes are available to choose from. Table 5.1 gives a list of the reporter dyes with the wavelengths of incident and emerging light. An appropriate dye can be chosen depending on the capability of the machine on which the probe would be used.

The quencher may be fluorescent e.g., TAMRA or chemical e.g., DABSYL and black hole quenchers (BHQ). TAMRA as a quencher has a drawback of giving high background fluorescence. DABCYL also has a limitation of having poor spectral overlap between itself and the reporter. Black hole quenchers (BHQ1, BHQ2 and BHQ3 etc.) have been developed to overcome these problems. Table 5.1 also provides a list of compatible quenchers for the relevant reporter dyes.

When choosing probes for a multiplex real time PCR the reporter dyes should be selected so that their excitation wavelengths are clearly separate with minimal overlap.

Passive reference dye

Some real time PCR machines require addition of a passive reference dye in the reaction mixture. The passive reference dye e.g., ROX is included in the reaction master mix. It does not participate in the 5' nuclease reaction but it serves as an internal reference for background fluorescence emission.

Fluorescence Resonance Energy Transfer (FRET) Probe method

In the FRET probe method, the target is amplified by the usual pair of primers. The FRET probes are a pair of probes that anneal to the inner region of the target in a head to tail configuration. The upstream probe has a fluorescent dye (donor) at the 3’ end whereas the downstream probe has another fluorescent dye (acceptor) attached at the 5’ end. If the target DNA is present it is amplified by the two primers and the two FRET probes also anneal to the target. The head to tail configuration of the FRET probes ensures that the fluorescent dye at the 3’ end of the upstream probe comes in close proximity to the fluorescent dye at the 5’ end of the downstream probe. On excitation by the light of appropriate wavelength the energy from the upstream probe (λ_1) is transferred to the downstream probe (λ_2). The fluorescent dye on the downstream probe after absorbing energy emits this in the form of light of yet another wavelength (λ_3). The latter is measured by the instrument and is directly proportional to the amount of amplification.

Table 5.1. Wavelengths of the incident and the emergent light of some commonly used reporters and the quenchers for TaqMan® probes.

Reporter dye	Incident light	Emergent light	Compatible quenchers
FAM/SYBR Green	492	516	TAMRA, DABCYL, BHQ1
TET	517	538	TAMRA, DABCYL, BHQ1
HEX/JOE/VIC	535	555	TAMRA, DABCYL, BHQ1
Cy3	545	568	BHQ2
TAMRA	556	580	BHQ2
ROX/Texas Red	585	610	BHQ2
Cy5	635	665	BHQ3

Molecular Beacon method

The target DNA is amplified by the usual pair of primers. The molecular beacon probe is designed to be complementary to the inner region of the target DNA. At each end of the probe 5-7 nucleotides, complementary to each other, are inserted to ensure a hairpin structure (stem) of the probe (Figure 5.3). At each end of the probe a fluorescent reporter dye and a quencher are also attached. In the non-annealed probe, the reporter and the quencher are very close to each other and the fluorescence from the reporter dye is quenched. As a result of the annealing of the probe to its target the hairpin structure opens up and separates the reporter from the quencher. The net result is that on excitation by the light of appropriate wavelength the reporter dye starts emitting fluorescence. The latter is directly proportional to the amount of amplification.

Real time PCR machine

The hardware

The real time PCR machine consists of a thermal cycler, light source, interchangeable light filters and a sensitive light recording device. The light source is either halogen lamp or LED and the recording devices may be CCD

camera, photodiode, or photomultiplier tube. The machines vary in the capability of detecting the number of colours. A basic machine can detect two colours while more advanced machines can detect up to five or more colours. The choice of machine with more colour detection is useful for simultaneous detection of greater number of probes with different fluorochromes in a multiplex PCR.

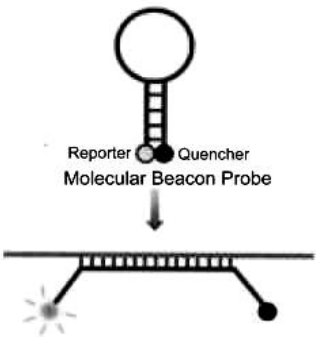


Figure 5.3. Molecular Beacon probe. In the un-annealed state it does not emit any fluorescence. In the annealed state the reporter is relieved from the quencher and starts emmitting fluorescent signal.

The software

The instrument software provides an interface for entering thermal cycling parameters and choice of dye (fluorochrome) and quencher selection. As the run starts the real time graphic record of amplification after each cycle is displayed on the screen. At the end of the run the amplification plot can be seen on linear or logarithmic scale.

Threshold cycle (Ct)

The first significant increase in the amount of PCR product as seen in the real time graphic recording of a PCR is called “Cycle-threshold” or “Ct”. It correlates indirectly with the starting amount of template DNA. Higher the amount of

template DNA smaller is the Ct. The real time PCR software provides options for calculating Ct automatically or manually (Figure 5.4).

Rn

Rn is an indicator of the amount of fluorescence generated in a PCR. The amount of fluorescence (Rn) of a test reaction and a non target control (NTC) are called Rn^+ and Rn^- respectively. The difference between Rn^+ and R^- is called ΔRn. The Ct values are derived from plotting the cycle numbers against the ΔRn (Figure 5.4).

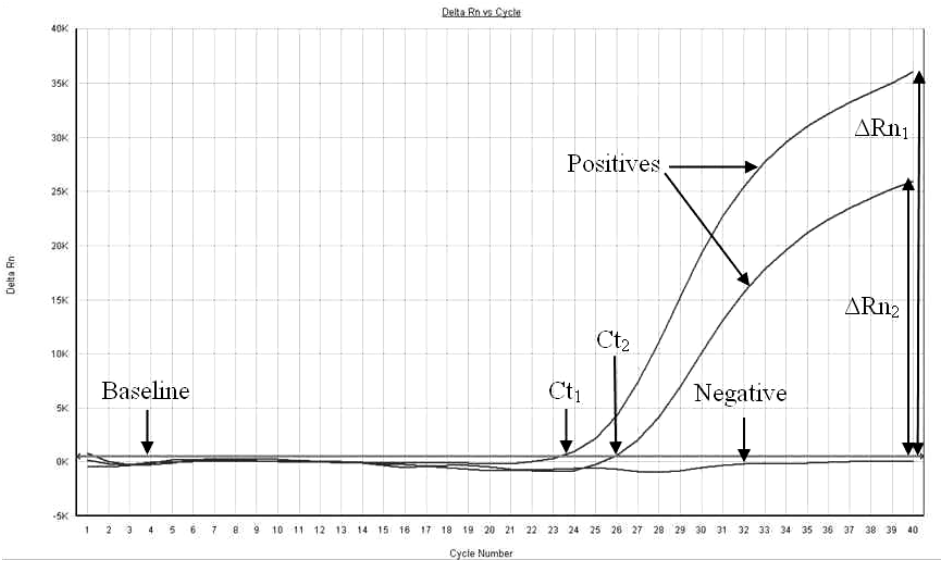


Figure 5.4. Real time PCR plot on a linear scale. The cycle numbers are shown on X-axis while the amount of fluorescence is shown on Y-axis. The Ct_1 (23.6) and Ct_2 (26.0) indicate that there is greater quantity of target DNA in the sample 1 than in the sample 2. This is also reflected in the ΔRn values of the two samples.

Linear versus logarithmic recording

In PCR one molecule of DNA is doubled with every cycle. In the initial 15-20 cycles the DNA molecules are doubling in number but their total number is far less as compared to the number of DNA molecules produced in the later cycles. In a real time PCR, our main interest is in knowing the first detectable change in

the amount of fluorescence (increase in DNA). If the result of real time PCR is plotted on a linear scale it becomes difficult to appreciate the smaller change in fluorescent signal in the earlier cycles. However, if the same result is plotted on a log scale the change in the earlier cycles becomes more pronounced and easy to recognize. The software of the real time PCR machines provides option for displaying the results in linear as well as log scale (Figure 5.5).

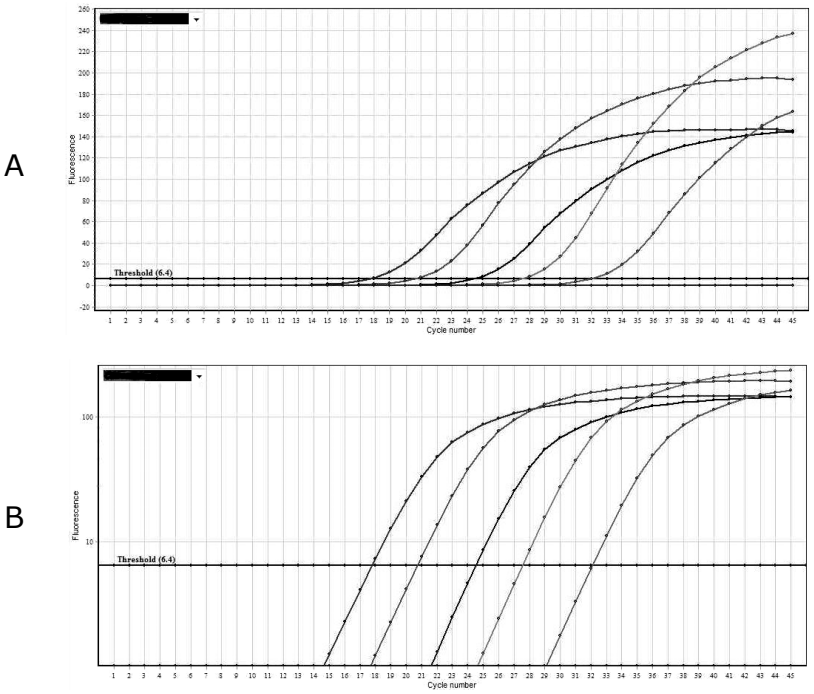


Figure 5.5. Real time PCR recording of the same samples on linear (A) and logarithmic (B) scale.

Absolute quantification of DNA by real time PCR

Absolute quantification of DNA is done by plotting a “standard curve” as follows:

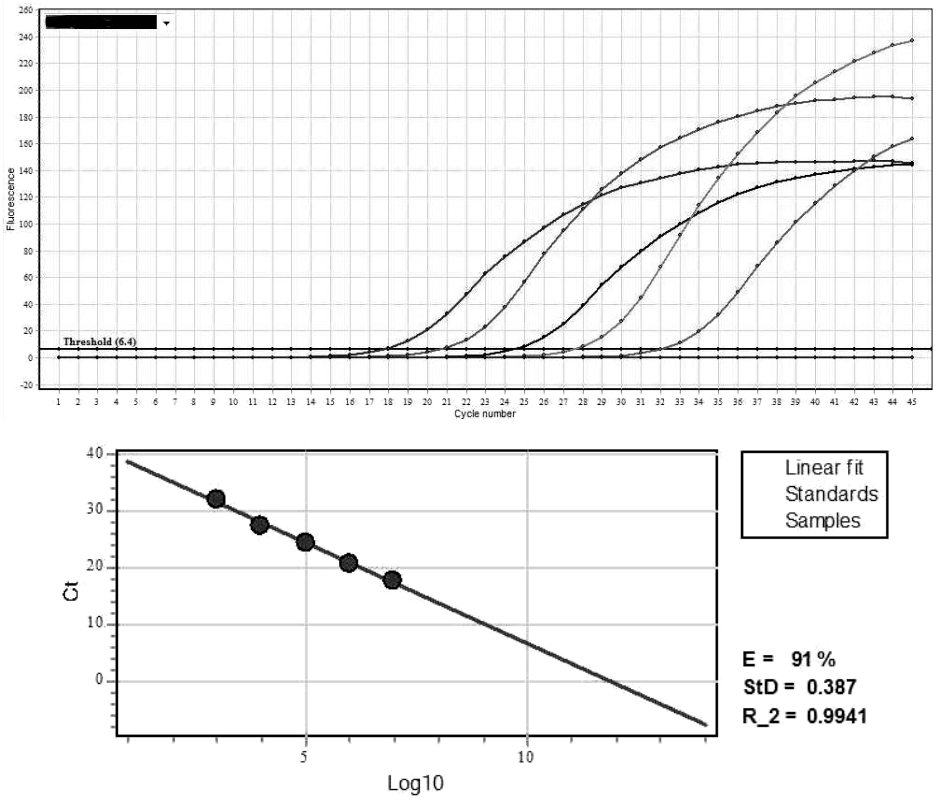
1. Take four or five DNA standards of known concentration or make tenfold dilutions of one high value DNA standard (1/1, 1/10, 1/100, 1/1000).
2. Enter the concentration of each DNA standard or its dilution in the real time PCR and complete the run.
3. Use the real time PCR software to plot the standard curve. The software uses the concentration and the Ct value of each DNA standard or its dilution to plot the standard curve (Figure 5.6). The standard curve may be saved for use with future runs.
4. The samples with unknown concentration of DNA may be run with the DNA standards. The real time PCR software automatically uses the standard curve to calculate the results of the unknown samples.
5. The unknown samples may be run separately and the previously saved standard curve may be imported for use with the new batch of unknown samples. If the results of unknown samples are calculated from a previously saved standard curve then it must be ensured that the run conditions of the standards and the unknown samples are identical.

PCR efficiency

Efficiency of PCR means the rate at which the amplified product is generated. The PCR efficiency is highest in the initial “exponential phase” when the number of DNA molecules is doubling with each cycle. The efficiency of PCR decreases in the latter “linear” and “plateau” phases because of the consumption and degradation of the reaction components.

In a PCR the number of DNA molecules double with each cycle. The doubling of DNA cannot go on indefinitely because the efficiency of PCR decreases due to progressive consumption of dNTPs and primers and denaturation of Taq polymerase at high temperatures. The initial outburst of amplification is gradually reduced to a plateau in the last few cycles of the PCR. If the process of

amplification is monitored in real time the following four phases can be recognized (Figure 5.7).



Standard	Ct	Concentration
QS-1	17.7	1.0 X 10^7 copies/ml
QS-2	20.7	1.0 X 10^6 copies/ml
QS-3	24.4	1.0 X 10^5 copies/ml
QS-4	27.5	1.0 X 10^4 copies/ml
QS-5	31.1	1.0 X 10^3 copies/ml

Figure 5.6. Real time PCR plot of five DNA standards and the standard curve.

Lag phase

In the initial 10-15 cycles of PCR the DNA molecules are doubling very efficiently but their number is too insignificant to be visualized.

Exponential phase

It is a continuation of the lag phase in which the DNA molecules are doubling very efficiently with each cycle and they also reach the threshold of detection. Exponential phase of PCR is short lived and is the most efficient phase of reaction for monitoring.

Linear phase

The exponential phase of PCR slows down as the reaction components are consumed. In the linear phase of PCR the doubling of DNA molecules progressively decreases and it takes more than one cycle for the DNA molecules to double.

Plateau phase

When the reaction components are consumed to a large extent the doubling of DNA molecules almost comes to a halt. If the reaction is allowed to proceed indefinitely the number of amplified DNA may actually decrease due to their degradation. This is the most inefficient phase of the reaction.

The efficiency of PCR is usually gauged by the number of cycles in which a given amount of DNA is increased tenfold. In 100% efficient PCR tenfold increase in the target DNA is achieved on an average in 3.32 cycles. The efficiency of PCR is usually measured from the slope of a tenfold diluted standard curve ($E = 10^{-1/slope}$). A steep slope indicates low efficiency whereas a gradual slope indicates a more efficient PCR.

The efficiency of PCR is mostly dependent on composition of PCR mix. Several additives that enhance PCR (Chapter 3) can improve its efficiency. PCR efficiency between 90-110% is considered acceptable. A greater efficiency of PCR is more likely to detect smaller amount of the target and vice versa.

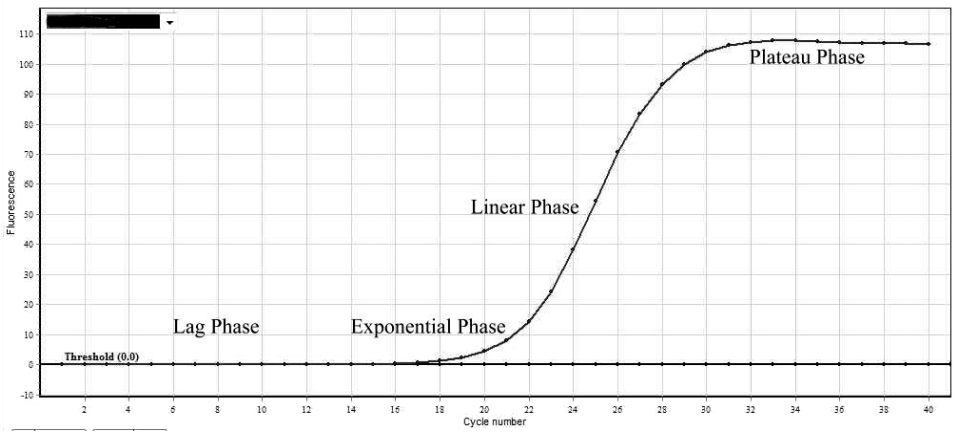


Figure 5.7. Real time monitoring of PCR showing four distinct phases.

Coefficient of correlation (R^2)

In a real time PCR there is a linear correlation between Ct value and the concentration of DNA. Under perfect conditions the Ct values of tenfold dilution of DNA standard when plotted against the concentration should show a straight line (100% correlation or R^2=1.0). However, inconsistencies in carrying out PCR, notably the pipetting errors, may cause varying degrees of deviation from the straight line. This may result in R^2 values below 1.0.

Relative Quantification of DNA by real time PCR

The amount of target DNA in a sample can also be calculated from the difference in the Ct values of the target DNA and another sample. The other sample can be internal reference (internal control DNA) or DNA standard with known concentration. In relative quantification the quality of RNA or DNA of the two samples should be similar. The target and the reference RNA or DNA can be run in the same or two different reaction tubes. The difference in the Ct value between the unknown (target) and the standard (reference) is called ΔCt. A major drawback of the relative quantification by using a single standard is that it

does not take into account the PCR efficiency that is calculated from the slope of the standard curve. Since PCR efficiency is higher in the earlier than in the later cycles a single standard is unlikely to give accurate results in the very high or the very low range.

Example:

ΔCt = Ct target – Ct reference (standard)

The amount of target is given by the mathematical formula $2^{-\Delta Ct}$

For example, at ΔCt of +2.0 the amount of target would be:

$2^{2.0} = 4.0$

Or the unknown sample has four-fold (times) DNA or RNA as compared to the reference (standard).

At ΔCt of -2.0 the amount of target would be:

$2^{-2.0} = 0.25$

Or the target sample has 0.25 (1/4) fold (times) DNA or RNA as compared to the reference (standard).

A scientific calculator or Excel spreadsheet can be used to calculate complex values of ΔCt like 2.7 or -3.4 etc.

Q-PCR by measurement of endpoint fluorescence

Real time PCR is the gold standard for Q-PCR. But it requires expensive equipment. A cheap alternate to Q-PCR is by measurement of the end point fluorescence (EPF) after amplification by fluorescent dye labeled probe and PCR primers. This method does not require expensive equipment and is also free from the hazard of carryover of amplified products because the PCR vial remains closed at the time of reading. The PCR is done in an ordinary thermal cycler with the standard pair of primers and fluorescent labeled probe. At the end of the PCR EPF is measured on a mini-fluorometer.

EPF by molecular probe method is as sensitive as the real time PCR. The molecular probe methods are highly specific and give almost no non-specific fluorescence. The fluorochrome molecules are released at each cycle of the

PCR. The fluorescence measured at the end of the PCR is the accumulated fluorescence produced during the entire PCR process. Some inconsistency may be seen in the EPF method because of variable PCR efficiency in the later amplification cycles. This makes EPF unsuitable for applications requiring precise measurement. Inconsistency in the EPF measurement can be minimized by using good quality consumables and by keeping the number of PCR cycles to bare minimum. The problems in real time PCR like inconsistency in RNA to cDNA conversion, lower amplification when the target template is low and background noise subtraction may also be seen with the EPF method.

EPF detection in the PCR reaction tube, like real time PCR, does not require the tube to be opened. This is a clear advantage as compared to the endpoint analysis by gel electrophoresis. It avoids the risk of cross-contamination by the amplified product of PCR which is a potent source of contamination for other PCR reactions.

DNA quantification by gel electrophoresis

DNA quantification by densitometry or visual inspection of the amplified product on gel electrophoresis is a very crude method of Q-PCR and is not recommended.

Digital droplet PCR (ddPCR)

In a digital droplet PCR (ddPCR) the traditional real time PCR reaction mix with a fluorescent probe/quencher the reaction volume is portioned into tens of thousands of nanoliter sized droplets. Each droplet behaves as an individual PCR reaction. The reaction can be used to amplify DNA or cDNA made from RNA. The distribution of the target DNA molecules in the droplets is assumed to follow a Poisson distribution. Depending on the concentration of the target DNA in the sample each droplet would be either devoid of the target or would have one or more target DNA molecules. If the target DNA concentration is high then most of the droplets would contain the target whereas only a few droplets would have the target if its concentration is low. The samples are portioned by making

oil emulsion in the PCR reaction tubes, micro well plates, capillaries, or miniaturized arrays with nucleic acid binding surfaces.

The droplets are amplified like any other endpoint PCR. At the end of the PCR the result is read by imaging the droplets. Each droplet emitting a fluorescent signal is read as positive. The proportion of droplets reading positive against the negatives give a precise measure of the exact quantity of the target nucleic acid in the sample.

The ddPCR method does not require comparison with any standards as in the real time PCR. It is also not subject to the variations related to PCR efficiency.

ddPCR is a very useful tool for qPCR. Because of its extreme sensitivity it can also be used to measure trace amounts of target DNA like that of cancer etc.

Bibliography

1. Dorak MT (ed) (2006) Real Time PCR. Taylor & Francis Group, New York.
2. Naeem MA, Ahmed S, Khan SA (2018) Detection of asymptomatic carriers of malaria in Kohat district of Pakistan. Malar J 17: 44 doi.org/10.1186/s12936-018-2191-y.
3. Ponchel F, Toomes C, Bransfield K, Leong FT, Douglas SH, Field SL, Bell SM, Combaret V, Puisieux A, Mighell AJ, Robinson PA, Inglehearn CF, Isaacs JD, Markham AF (2003) Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. BMC Biotechnology 3: 18-31.

6

Identification of Sequence Variants

An alteration in the DNA or RNA sequence is called “variant”. The Human Genome Variation Society (HGVS) has recommended to replace the previously used terms mutation and polymorphism with “sequence variant”. The sequence variants involving one or just a few nucleotides are called point variants. These may occur in the form of insertions, deletions, or substitutions. Most sequence variants whether point or large can be detected by PCR.

Point variants

The known point variants are mostly detected by modifications of PCR or by restriction enzyme digestion of the PCR amplified DNA. The unknown sequence variants are detected by genomic sequencing of the PCR amplified DNA.

Detection of known point variants

Known point variants can be detected by modifications of PCR for example use of allele specific primers and probes, single nucleotide primer extension and oligonucleotide ligation assay etc. All of these methods work on the principle of match or mismatch between the variant allele and the primer/probe. The two most commonly employed methods include Amplification Refractory Mutation System (ARMS) and Reverse Dot Blot (RDB). Some known variants may also be identified by restriction enzyme digestion of PCR amplified DNA if the variant creates or abolishes a restriction site.

Amplification Refractory Mutation System (ARMS)

The DNA is amplified by primers that are specific for the variant allele. ARMS PCR is also known as allele specific PCR. In PCR a mismatch at the 3' end of a primer and its target can drastically reduce the annealing of the primer and hence the amplification. This is due to the absence of 3' to 5' exonuclease proofreading activity of Taq polymerase. High fidelity DNA polymerases, having 3' to 5' exonuclease activity, cannot be used in ARMS PCR. Heterozygotes and homozygotes of an ARMS positive sequence variant can be differentiated by an additional step of amplification with a primer specific for the normal allele.

ARMS primer design

ARMS PCR is done with a pair of primers including a common and an ARMS primer. General principles of primer designing (chapter 3) also apply to the ARMS primers. The common primer is like any other PCR primer. But the ARMS primer has the following additional features:

1. The primer is usually 25-30 bases in length.
2. The nucleotide at the 3' end of the primer should be complementary to the target nucleotide i.e. G for C and T for A and vice versa. Mismatch at this position can drastically reduce the amplification. A:G, G:A, and C:C mismatches have the worst effect whereas the other mismatches have varying degrees of effect. For example in a variant with A>T substitution the ARMS primer for the variant allele (T) the last nucleotide should be complementary to the nucleotide T i.e. it should have A. The primer for the normal allele at the same position should be complementary to the nucleotide A i.e. it should have T (Figure 6.1).
3. An additional mismatch at one of the last five nucleotides of the ARMS primer further increases its specificity (Figure 6.1).
4. It is essential to include an internal PCR control in ARMS reactions. A pair of primers is included to amplify an unrelated region of the genome. Amplification of the internal control region and no amplification by the ARMS primer indicate a true negative. In a false negative result neither the ARMS primers nor the internal control primers show any amplification. There could be several reasons for the false negative result e.g. too little or too much DNA, poor quality of target DNA, failure to add primer, Taq, or other reagents and presence of PCR inhibitors (Chapter 3).
5. Sensitivity and specificity of an ARMS PCR can be ensured by using stringent reaction conditions. Good primer design, higher annealing temperature and limited number of cycles are important in avoiding false results. The number of cycles should be just enough to give a clear positive result. Increasing the number of cycles un-necessarily can cause false positives. The usual length of the ARMS primer should be 25-30 bases and these should have high Tm and annealing temperature.

PCR conditions for ARMS

Locus: β-globin gene

GenBank accession: NG_000007.3

Allele: HBB:c.92+5G>C

- Forward primer: 5’-ACCTCACCCTGTGGAGCCAC
- Reverse primer (ARMS)
  5’-CTCCTTAAACCTGTCTTGTAACCTTGTTAG
- Amplified product: 285 bp
- Control primer (forward):
  5'-CAATGTATCATGCCTCTTTGCACC
- Control primer (reverse):
  5'-GAGTCAAGGCTGAGAGATGCAGGA
- Amplified product: 861 bp

- Reaction volume: 25µl
- PCR mix: 22µl
- Primer concentration: 1µl (5 pmol each/µl) (Chapter 3)
- Taq polymerase: 0.5 units (0.1µl)
- Template DNA: 2µl (~300ng)

Thermal cycling:

- Initial denaturation: 1 minute at 94°C
- No. of cycles: 25
- Denaturation: 1 minute at 94°C
- Annealing: 1 minute at 65°C
- Extension: 1 minute 30 seconds at 72°C
- Final extension: 3 minute at 72°C

Electrophoresis: 10 X 10 cm 6% polyacrylamide, 40 minutes at 150 volts.

Staining: 0.1% silver nitrate

Result: Figure 6.2.

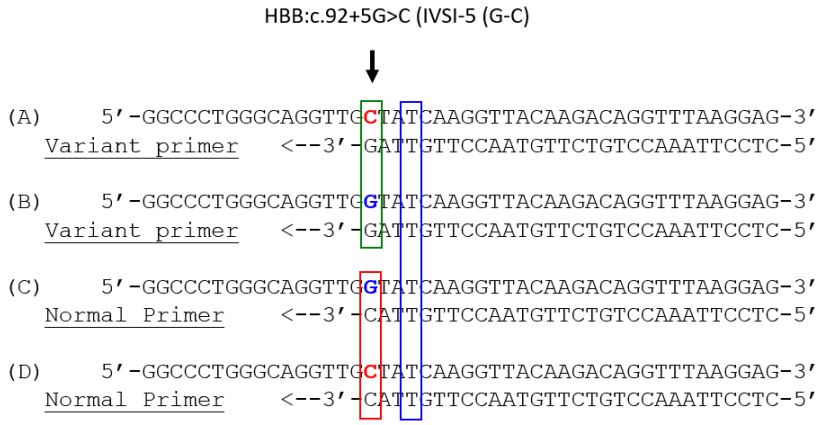


Figure 6.1. ARMS primers for the variant and the normal alleles of HBB:c.92+5G>C (IVSI-5 (G-C). (A) ARMS primer for the variant 'C', (B) mismatch between the variant ARMS primer and the normal sequence 'G', (C) matched ARMS primer for the normal sequence 'G', (D) mismatch between the normal ARMS primer and the variant 'C'. An additional mismatch (T:T) near the 3' end (blue box) increases the specificity of the primer.

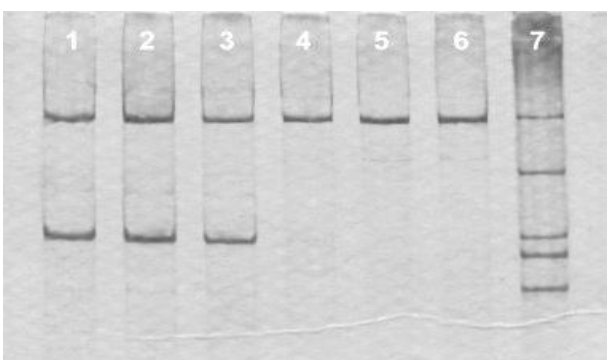


Figure 6.2. Silver stained polyacrylamide gel electrophoresis after ARMS PCR. All lanes show 861bp internal control fragment. Lanes 1-3 show 285bp fragment of HBB:c.92+5G>C. Lanes 4-5 are negative for the variant. Lane 7 shows allelic ladder for various thalassaemia variants.

Real time PCR

High Resolution Melting Curve (HRMC) analysis

The DNA sequence containing the variant(s) is amplified in multiple overlapping fragments by the real time PCR method in the presence of DNA intercalating dyes like SYBR green. At the end of the PCR high resolution melting curve analysis (HRMCA) step is added that measures fluorescence reading at temperature increments of 0.1°C. The Tm of an amplified DNA fragment containing a point variant differs from that of its normal sequence. The difference in the Tm of the two fragments generates different melting curves (Figure 6.3).

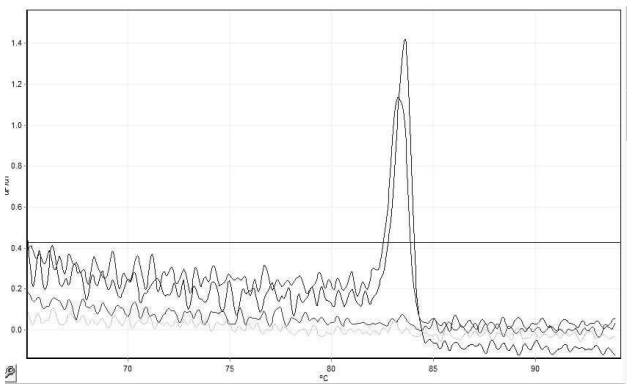


Figure 6.3. HRMCA of β-globin gene. The two peaks with slightly differing Tm are of the wild type (Normal) DNA and HBB:c.27_28insG (Fr 8-9 (+G) variant.

In another type of HRMCA the region of DNA containing the variant(s) is amplified by a pair of primers. Instead of the SYBR green the reaction tube contains two fluorescently labeled probes that hybridize to the adjacent sequences in the target DNA. One of the probes (sensor probe) covers the region expected to contain the variant(s) and is labelled with a fluorochrome at its 3' end. The second probe (anchor probe) lies a few bases upstream to the sensor probe. Close proximity of the annealed probes facilitates fluorescence resonance energy transfer between them. Probes are designed to have different melting

temperatures (Tm), so that the sensor probe with lower Tm lies over the variant site(s). Monitoring of the emitted fluorescent signals as the temperature increases detects loss of fluorescence as the probe with the lower *T*m melts off the template. A single base mismatch under this probe produces a Tm shift of 5–10°C, allowing easy distinction between the wild-type and the variant alleles. The ability to detect base mismatches under the probe with the lower Tm (variant detection probe) and the use of two, different-colored probes allows more than one mutation to be screened in a single PCR reaction. Because of its low costs, reproducibility, and ease of handling, the assay is potentially suitable for a routine clinical laboratory.

HRMC analysis software is used to generate DNA melting curves. Presence of each point mutation generates its own specific melting curve that can be identified by simultaneous running of known positive and negative DNA controls. Extensive optimization of the procedure is required to define HRM curves of known mutations in the target population. The patterns generated by heterozygotes, homozygotes and all possible compound heterozygote combinations in the target population are required.

Hydrolysis probe method

Segment of a gene containing the variant is amplified by a pair of primers and a hydrolysis probe (Taqman® probe). A separate pair of probe is required for identification of each mutation (one for the variant and the second for the normal allele). The hydrolysis probes are typically labelled at the 5'-end by a fluorescent dye and a quencher at the 3'-end. When a probe anneals to its target the fluorochrome is detached from the probe and starts emitting fluorescence. The procedure can be multiplexed for several variants in one reaction if multi coloured fluorescent dyes are used on each probe. Real time PCR by hydrolysis probes is efficient but expensive because it requires two separate probes for each variant i.e. one for the variant and the other for the normal allele.

Reverse dot blot (RDB)

In reverse dot blot (RDB) multiple allele specific oligonucleotide (ASO) probes for the known variants and their normal alleles are immobilized on a nylon membrane. PCR is done with biotin labelled primers and the amplified product is layered on the ASO probes bound to the nylon membrane. The results are read after incubation and washing at 45°C. ASO probes immobilized on the nylon membrane would bind to their target allele if it is present otherwise the amplified product would be washed away.

Procedure

1. Target DNA is amplified with 5' biotin labelled primers.
2. ASO probes are prepared with 5'-NH_2 modification.
3. ASO probes for the mutant and the normal alleles are spotted on the surface of a nylon membrane.
4. Amplified DNA is added to the nylon membrane after dilution in hybridization buffer and denaturing at 95°C for 10 minutes.
5. Nylon membranes are incubated for 60 minutes at 45°C and then washed for 20 min at 45°C. The washing temperatures for different probes may vary depending on its length and amount. It should be determined by trial and error.
6. The washed membrane is immersed in streptavidin labelled alkaline phosphatase solution for 30 min at room temperature.
7. The membranes are washed for 15 min at room temperature.
8. Finally, the membranes are incubated in a solution containing substrate for alkaline phosphatase (NBT/BCIP) and the color is developed for 30-45 minutes.
9. Positive results are shown by dark spots whereas negative results are shown by light grey dots (Figure 6.4).
10. The membrane strips with immobilized ASO probes for different variants can be prepared in large batches and stored at room temperature for up to six months.

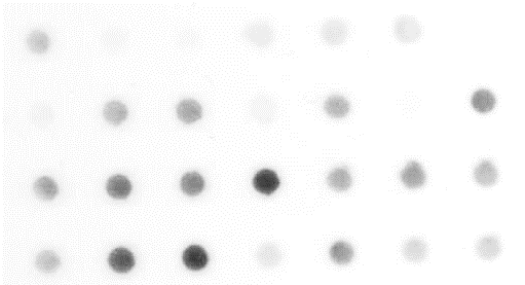


Figure 6.4. Results of reverse dot blot. Dark spots represent positive results whereas light grey dots are negative results.

Restriction enzyme method

Restriction endonucleases are enzymes of bacterial origin that can cut DNA at specific sequences called "restriction sites". As a result of digestion by a restriction enzyme the DNA sequence is cut into fragments of varying lengths. This forms the basis of restriction fragment length polymorphism (RFLP). It is a very useful technique in identification of single nucleotide variants (SNV). The SNVs may be recognized by the presence or the absence of a restriction enzyme site (restriction site).

By convention the presence of a restriction site is written as "+" and its absence as "-". Any DNA sequence containing the SNV is amplified by a pair of flanking primers and the amplified DNA is incubated with the restriction enzyme. If the SNV is present the amplified DNA is cut in to two pieces at a point where the SNV is present (+). If the SNV is not present the DNA fragment remains as single piece (-). The case would be reversed if the SNV abolishes the restriction site. A heterozygote of SNV would have the cut and the uncut fragments (+/-). The homozygote would have only the cut (+/+) or the uncut (-/-) fragments.

Example 1

DNA sequence in the 5' un-translated region of β-globin gene contains a C>T SNV (GenBank NG_000007.3) (Figure 6.5). The SNV can be identified by digestion with restriction enzyme Rsa-I that cuts the GTAC but not the GTAT sequence (Figure

6.6). A 213bp fragment of DNA containing the C>T SNV is amplified by a pair of primers (Figure 6.4). The amplified DNA is incubated overnight at 37°C with the restriction enzyme Rsa-I. The C>T variant when present results in cutting of the 213bp fragment into two pieces of 66bp and 147bp each.

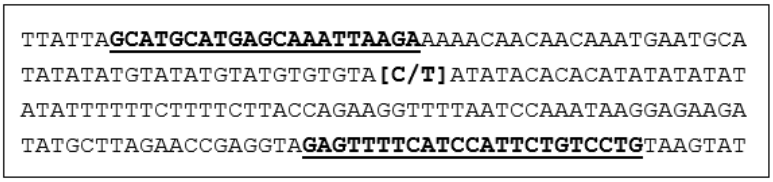


Figure 6.5. PCR primers for the C>T SNV in the 5' untranslated region of β-globin gene.

Figure 6.6. The C>T Variant can be recognized by a restriction enzyme "Rsa-I" that will cut the "C" allele into two pieces. The "T" allele is not cut by Rsa-I and therefore remains as a single piece.

PCR Protocol

- Locus: β-globin gene
- GenBank accession: NG_000007.3
- Forward primer: 5'-GCATGCATGAGCAAATTAAGA
- Reverse primer: 5'-TCTTCCTGCGTCTCCAGAAT
- Amplified product: 213 bp
- Restriction enzyme: Rsa-I
- Restriction fragments: 66bp and 147bp
- Reaction volume: 25µl
- Primer concentration: 1µl (5 pmol each/µl)
- Taq polymerase: 0.5 units (0.1µl)
- Template DNA: 1µl (~300ng)
- Thermal cycling:
    - Initial denaturation: 1 minute at 94^{o}C
    - No. of cycles: 30
        - Denaturation: 1 minute at 94^{o}C
        - Annealing: 1 minute at 57^{o}C
        - Extension: 1 minute 30 seconds at 72^{o}C
    - Final extension: 3 minute at 72^{o}C
- Restriction enzyme (Rsa-I): 1µl
- Amplified product: 15 µl
- Incubation: Overnight at 37^{o}C
- Electrophoresis: 10 X 10 cm 6% polyacrylamide
  Run for 20 minutes at 150 volts
- Staining: 0.1% silver nitrate

Example 2

Sickle haemoglobin is caused by A-T substitution in Codon 6 of β-globin gene. The variant can be identified by RFLP using the restriction enzyme Dde-I. The enzyme recognizes the sequence CTNAG where N stands for any nucleotide (Figure 6.7).

The 597bp fragment of β-globin gene containing the A>T substitution (Figure 6.8) is amplified by the following pair of primers:

- Forward Primer: 5'-GGCCAATCTACTCCCAGGAG
- Reverse Primer: 5'-ACATCAAGCGTCCCATAGAC

The amplified fragment contains three restriction sites identified by the enzyme Dde-I. The first site, present at 145bp downstream from the forward primer, is altered when the Hb-S variant is present. The other two sites serve as internal controls for the restriction enzyme Dde-I. The net result of Dde-I digestion of the 597bp fragment is generation of different patterns of restriction fragments when the Hb-S mutation is present or absent.

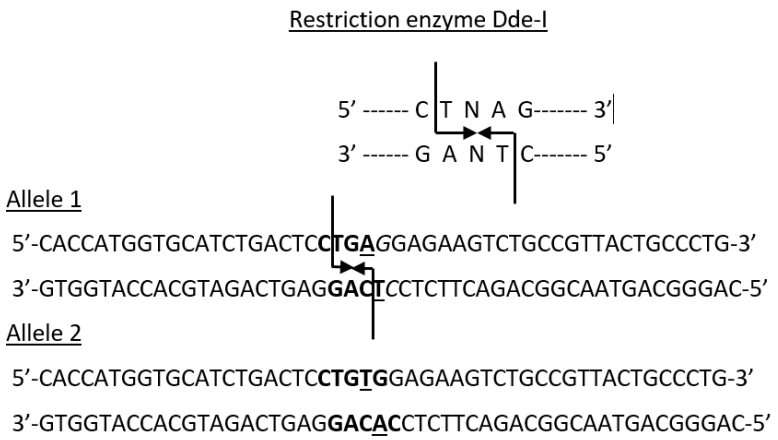


Figure 6.7. Identification of the C>T SNV in human globin gene with restriction enzyme Dde-I.

TAGACCTCACCCTGTGGAGCCACACCCTAGGGTT**GGCCAATCTACTCCCAGGAG**CA
GGGAGGGCAGGAGCCAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT
ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTG
CATCTGACTC***CTG[A/T]G***GAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGA
ACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGAC
AGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGT
TTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTG
CTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCAC
TCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCG
GTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACA
CTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGATC***CTGAG***AACTTCAGGGTGA**G**
TCTATGGGACGCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTC

Figure 6.8. The primer pair (bold underline) for amplification of A>T substitution in codon 6 of β-globin gene. The 597bp amplified product contains three restriction sites (bold italic) recognized by Dde-I at positions +145, +523 and +560 from the forward primer. The first of the three restriction sites is abolished by the presence of A>T substitution (Hb-S).

PCR Protocol:

- Reaction volume: 25µl
- Primer concentration: 1µl (5 pmol each/µl)
- Taq polymerase: 0.5 units (0.1µl)
- Template DNA: 1µl (~300ng)
- Thermal cycling:
    - Initial denaturation: 1 minute at 94°C
    - No. of cycles: 30
        - Denaturation: 1 minute at 94°C
        - Annealing: 1 minute at 57°C
        - Extension: 1 minute 30 seconds at 72°C
    - Final extension: 3 minute at 72°C
- Restriction enzyme (Dde-I): 1µl
- Amplified product: 15 µl
- Incubation: Overnight at 37°C

- Electrophoresis: 10 X 10 cm 6% polyacrylamide
  Run for 20 minutes at 150 volts
- Staining: 0.1% silver nitrate

Examples 3

H63D variant in the HFE gene on chromosome 6 can be detected by RFLP. The restriction enzyme Mbo-I cuts the DNA sequence NGATCN:

Restriction enzyme Mbo-I

5'...**G A T C**...3'

3'...**C T A G**...5'

A 208bp fragment of DNA sequence containing the H63D variant (GenBank NT_007592.15) is amplified by the following primers:

H63D-F 5'-ACATGGTTAAGGCCTGTTGC

H63D-R 5'-GCCACATCTGGCTTGAAATT

The normal (wild type) DNA is cut by the restriction enzyme (Mbo-1) resulting in two fragments of 138 and 70bp. The presence of H63D variant abolishes the restriction site. In a heterozygote (-/+) the uncut 208bp fragment and the cut 138bp and 70bp fragments are present. Homozygotes (-/-) of the variant show only 208bp fragment whereas the homozygotes of the normal allele (+/+) show two fragments of 138bp and 70bp (Figure 6.9).

PCR Protocol

- Reaction volume: 25µl
- Primer concentration: 1µl (5 pmol each/µl)
- Taq polymerase: 0.5 units (0.1µl)
- Template DNA: 1µl (~300ng)
- Thermal cycling:
  - Initial denaturation: 1 minute at 94^{o}C
  - No. of cycles: 30

- Denaturation: 1 minute at 94°C
- Annealing: 1 minute at 57°C
- Extension: 1 minute 30 seconds at 72°C

- Final extension: 3 minute at 72°C

- Restriction enzyme (Mbo-1): 1µl
- Amplified product: 15 µl
- Incubation: Overnight at 37°C
- Electrophoresis: 10 X 10 cm 6% polyacrylamide
  Run for 20 minutes at 150 volts
- Staining: 0.1% silver nitrate

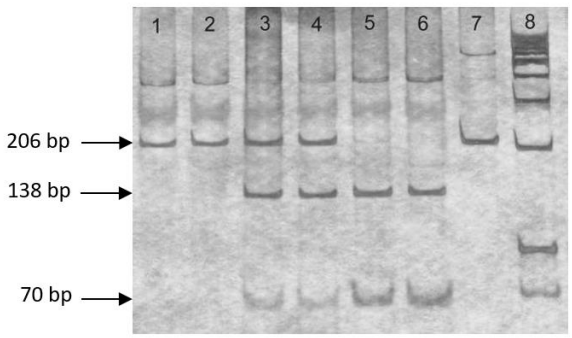


Figure 6.9. Silver stained polyacrylamide gel electrophoresis of the Mbo-I restriction enzyme digested fragments of the HFE gene showing the H63D variant. The -/- (lanes 1 & 2 showing 206 bp fragment) indicates homozygous and -/+ (lanes 3 & 4 showing 206, 138 & 70 bp fragments) indicates heterozygous for the variant. The +/+ (lane 5 & 6 showing 138 & 70 bp fragments) indicates normal genotype. Lane 7 shows 206 bp uncut fragment.

Detection of gene deletions and inversions

Large deletions in a gene can be detected by gap PCR. The target DNA is amplified by a pair of primers flanking the deletion. In the absence of deletion PCR primers are far apart and no amplification is observed. But the presence of deletion brings

the same pair of primers close enough to enable amplification. Most gene deletions causing α-thalassaemia can be detected by gap PCR (Figure 6.10).

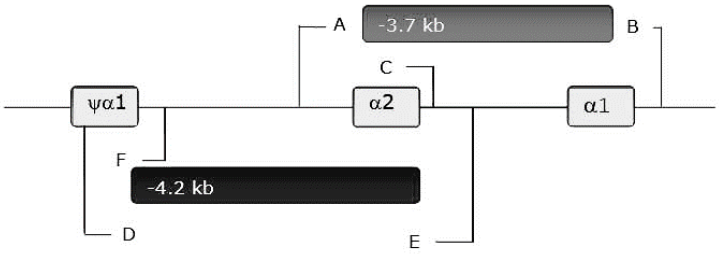


Figure 6.10. Gap PCR for α-thalassaemia-2 deletions. The functional α2 and α1 and ψα1 genes are shown as open boxes and the solid boxes indicate the extent of deletions. Primers A and C amplify a control fragment whereas primers A and B amplify only when -3.7kb deletion is present. Similarly, primers D and F amplify a control fragment while primers D and E amplify a fragment only when -4.2kb deletion is present. The amplified products can be visualized by gel electrophoresis.

If the gene deletion is less than 1kb in length the flanking PCR primers generate an amplification product that is larger than the amplified product when the deletion is present. In a heterozygote of such deletion two amplified products are generated i.e. short (deletion fragment) and long (wild-type fragment). In homozygous deletion only the short fragment is generated. The 619bp deletion in the distal part of β-globin gene is a typical example.

PCR can also be used to detect inversions in the genome. For example Inv/Del $G_{\gamma}(A_{\gamma\delta\beta})^{o}$ in the β-gene complex can be detected as follows:

Primers H1 & H2 and H4 & H5 amplify 1,195bp and 665bp fragments respectively of the normal β-gene cluster. The presence of Inv/Del $G_{\gamma}(A_{\gamma\delta\beta})^{o}$ brings Primers H1 & H3 and H2 & H5 target sequences in line and sufficiently close to allow amplification of 327 bp and 371 bp fragments of the β-gene complex respectively (Figure 6.11). The heterozygotes show both the normal and the abnormal amplified fragments whereas the homozygotes show only the abnormal amplified product while the normal amplified products are absent.

Figure 6.11. PCR screening strategy and location of primers for detection of Inv/Del $G_{\gamma}(A_{\gamma\delta\beta})^{o}$.

The following five primers are used for the amplification in two different combinations comprising H1, H2, & H3 and H2, H4 & H5.

H-1 5'-ATG CCA TAA AGC ACC TGG ATG

H-2 5'-GAG CTG AAG AAA ATC ATG TGT GA

H-3 5'-TAA CCA TAT GCA TGT ATT GCC

H-4 5'-CAA TGT ATC ATG CCT CTT TGC AC

H-5 5'-GCA GCC TCA CCT TCT TTC ATG G

Multiplex Ligation-dependent Probe Amplification (MLPA)

MLPA is a very sensitive method for detection of very small to large deletions, insertions, and duplications. Multiple pairs of probes containing sequences complementary to the target are used. A large number of probes spanning the entire target gene are used. In each pair of probes one probe is synthetic whereas the other is cloned in M13 vector. Each pair of probes after hybridization are ligated with DNA ligase. All probe pairs have identical end sequences that allow subsequent amplification by a single pair of primers (Figure 6.12). The amplification results in several amplification products varying in size from 130 to 480 bases. The amplified products are separated by electrophoresis. If one of the two amplification primers are labelled with a fluorochrome separation may be done by capillary electrophoresis on a genetic analyzer. Several commercial kits for MLPA applications in many genetic and neoplastic disorders are available.

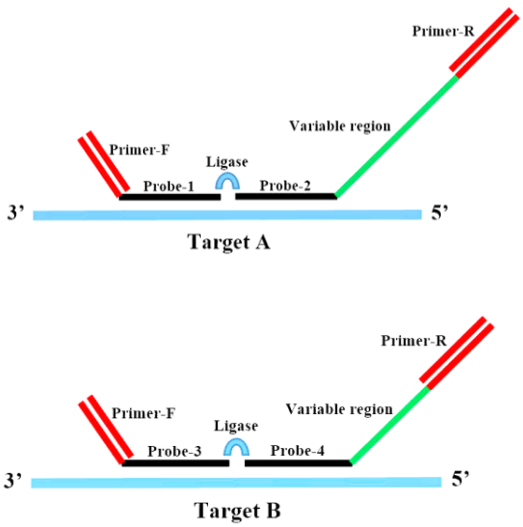


Figure 6.12. MLPA Probes for detection of gene deletions. Multiple sets of probes amplify several targets (A, B etc.) along the length of the gene. Each set of probes has two components (black). At the ends of each probe same leader sequences are attached (red). After hybridization two probes of a set are joined by DNA ligase followed by amplification with a single pair of primers (F & R). The amplified products of each set of probes are different because of the variable lengths of a specific regions in each probe (green).

Detection of unknown Variants

The unknown DNA variants may be screened by non-specific methods like Denaturing Gradient Gel Electrophoresis (DGGE) or Single Strand Conformation Polymorphism (SSCP). These methods only indicate whether a variant is present in the DNA or not. The unknown variant if indicated may be confirmed by genomic sequencing.

Denaturing Gradient Gel Electrophoresis (DGGE)

Movement of single stranded DNA in acrylamide gels, in addition to the size of the fragments, is also dependent on its base composition. Single stranded DNA tends to form secondary structures. The amount and shape of secondary structures is dependent on the base composition of DNA.

When a DNA fragment 200-700bp is run on acrylamide gel with increasing concentration (gradient) of a denaturant like urea it would initially move according to the molecular weight. As it moves into the higher concentration of the denaturant it reaches a point where the DNA starts to melt (the two strands begin to separate). The resulting single strands of DNA also start to develop secondary structures. The partial melting and formation of secondary structures severely retard the movement of DNA molecule in the gel. A single base pair change in the DNA can cause significant mobility shift. In this way different alleles can be identified by the differences in mobility on a gel with a gradually increasing denaturant concentration.

Constant temperature of 60^{o}C, formamide (0-40%) and gradually increasing concentration of urea from 0-7 M are usually used as denaturants in DGGE.

DGGE is a useful technique for identification of point mutations. But these gels are difficult to make and run. Consequently it is mostly replaced by genomic sequencing. The technical details of making DGGE gels can be seen in the previous edition of this book.

Single Strand Conformation Polymorphism (SSCP)

The secondary structures in single stranded DNA interfere with the mobility of the DNA strand in a polyacrylamide gel. Two single-stranded DNA fragments with only a single nucleotide difference would migrate at different rates. This forms the basis of SSCP.

Single stranded DNA can be produced by first amplifying DNA and then denaturing it to form single strands. Alternatively single stranded DNA can be produced by asymmetric PCR in which one primer of a pair is used in excess. In the initial cycles of PCR double stranded DNA is produced but as the primer with lower concentration is consumed the primer in excess causes extension of only a single strand. The usual fragment size for SSCP is between 150-300 bases. SSCP is done on non-denaturing polyacrylamide gels that are stained in silver nitrate.

SSCP is a useful technique for screening of known and unknown SNVs. SSCP is unable to tell the base composition of DNA unless control DNA fragments of known composition are also run parallel to the unknown samples.

Genomic sequencing

The process of determining the sequence of an unknown DNA is called sequencing. In the last three decades automated Sanger's di-deoxy chain termination method has dominated the scene. It has played a pivotal role in the human genome project. More recently the next generation sequencing (NGS) technologies have emerged to produce enormous amount of sequencing data at low cost. Up to one billion short reads of sequence can be produced in one instrument run. Details of the sequencing technologies are discussed in Chapter 9.

Bibliography

1. Geever RF, Wilson LB, Nallaseth FS, Milner PF, Bittner M, Wilson JT (1981) Direct identification of sickle cell anemia by blot hybridization. Proc Natl Acad Sci USA 8: 5081-5085.
2. Kho SL, Chua KH, George E, Tan JAMA. High throughput molecular confirmation of β-thalassemia mutations using novel TaqMan probes. Sensors 2013; 13: 2506-2514. doi:10.3390/s130202506.
3. Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sininsky JJ (1989) Effect of primer template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res* 18: 999-1005.
4. Chassanidis, C., Boutou, E., Voskaridou, E., & Balassopoulou, A. (2016). Development of a high-resolution melting approach for scanning beta globin gene point mutations in the Greek and other Mediterranean populations. *PloS one*, *11*(6), e0157393.
5. Moreno I, Bolufer P, Perez ML, Barragan E, Sanz MA. Rapid detection of the major Mediterranean β-thalassaemia mutations by real-time polymerase chain reaction using fluorophore-labelled hybridization probes. Br J Haematol 2002; 119: 554–557.
6. Newton CP, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The

amplification refractory mutation system (ARMS). *Nucleic Acids Res* 17: 2503-2516.

7. Schouton JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002; 30(12):e57.
8. Tsui WMS, Lam PWY, Lee KC, Ma KF, Chan YK, Wong MWY, Yip SP, Wong CSC, Chow ASF, Lo STH (2000) The C282Y mutation of the *HFE* gene is not found in Chinese haemochromatotic patients: multicentre retrospective study. HKMJ 6: 153-8.

7

Gene Expression Studies

Proteins are the most important effector molecules in a cell that perform most of the biological functions. The blueprint of protein structure resides in DNA and is called the “genetic code”. The genetic code transcribes to messenger RNA (mRNA) which is exported to cytoplasm where protein synthesis takes place. The study of expression of the genetic code through quantitation of messenger RNA (mRNA) is called gene expression study. It can provide very useful information about the physiological and pathological processes. Gene expression study is done by quantitation of mRNA in a cell.

Northern blotting

Northern blotting, first described in 1977, uses electrophoretic separation of cellular RNA and using radiolabelled hybridization probes to the mRNA of interest. The process is only of historical interest and is mostly replaced by PCR based methods.

Ribonuclease protection assay

Cellular RNA containing a mixture of different RNA molecules is mixed with antisense RNA or DNA probes complementary to the sequence or sequences of interest. The complementary strands hybridized to form double-stranded RNA or DNA-RNA hybrids. The mixture is exposed to ribonuclease that cleaves the single-stranded RNA but has no activity against the double-stranded RNA. On completion of the reaction susceptible RNA regions are degraded to very short oligomers or to individual nucleotides; the surviving RNA fragments are those that were complementary to the added antisense strand and thus contained the sequence of interest. The process can be used to quantify any mRNA of interest.

Polymerase Chain Reaction (PCR)

RNA quantitation by PCR is the simplest and the most sensitive method of studying gene expression. Another advantage of PCR is its ability to discriminate closely related mRNAs. Conventional PCR and gel electrophoresis, once popular, has almost completely been replaced by real time PCR.

PCR based RNA quantification involves the following steps:

1. Extraction of RNA.
2. Inactivation of contaminating DNA in the RNA sample.
3. Synthesis of cDNA from the target and the reference gene mRNA.
4. PCR amplification of the target and the reference cDNA.
5. Mathematical calculation.

Extraction of RNA

Details of RNA extraction are given in Chapter 2.

Inactivation of contaminating DNA

The extracted RNA samples are usually contaminated with DNA left from the extraction procedure. This DNA can be amplified in subsequent PCR and may give false positive results. The unwanted DNA may be removed by incubating the extracted RNA with DNase-I. The incubation is done at 37^{o}C for 10-30

minutes followed by inactivation of DNase-I at 75^{o}C for 10 minutes. A cDNA minus control may also be included to look for non-specific amplification of the contaminating DNA in the sample.

Synthesis of cDNA

Since RNA cannot be amplified like DNA, it is first converted to copy DNA or cDNA. The cDNA is then used as template for amplification like any other DNA sequence. The cDNA synthesis requires:

1. RNA template.
2. Primer complementary to the template RNA sequence.
3. Reverse transcriptase.
4. Deoxy-nucleotides (dNTPs).
5. Magnesium ions.
6. Reaction buffer.

The primer

cDNA may be synthesized by any of the three primers:

1. Gene specific primer.
2. Random hexamers.
3. Oligo dT primer.

Gene specific primers are complementary to the target itself. These can be used in all types of applications but are especially suitable for cDNA synthesis and PCR amplification in the same reaction tube. These are also useful for synthesis of long stretches of cDNA. The forward or the reverse primer used for amplification is often used for cDNA synthesis.

Random hexamers are a mixture of random combinations of six nucleotide long pieces of DNA. These are useful when only small stretches of cDNA or multiple target cDNAs are to be synthesized (Figure 7.1).

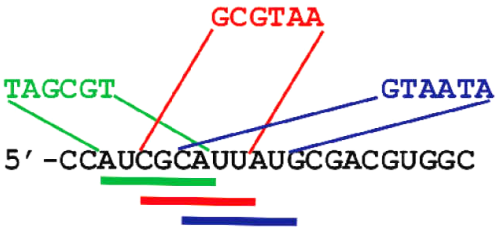


Figure 7.1. Random hexamers are a mixture of random combinations of six nucleotides. These anneal randomly to their target RNA and produce relatively short stretches of cDNA.

Oligo dT primers are designed to anneal with the poly-A tail at the 3' end of mRNA molecules. For example 5′-d (TTT TTT TTT TTT TTT TTT)-3. Since mRNA constitutes only 1-2% of the total RNA, the yield of cDNA from oligo dT is low. These are useful for synthesizing long stretches of cDNA or when multiple targets are to be synthesized.

Reverse transcriptase

Reverse transcriptase is an enzyme that is used for synthesizing DNA copy of RNA. The most commonly used reverse transcriptase is derived from Moloney murine leukaemia virus (MMLV). It is a RNA dependent DNA polymerase that can synthesize DNA complementary to the RNA sequence. In the presence of the primer DNA synthesis starts from 5' to 3' direction. A thermostable enzyme derived from *Thermus thermophilus (Tth)* has dual activity of reverse transcriptase and DNA polymerase. It can be used for reverse transcription and subsequent PCR in the same reaction tube.

Example:

Osteopontin (OP) gene expression in murine osteoblast cell line MC3T3-E1:

The cell line was exposed to varying concentration of hydroxyapatite (HA) during culture. The murine glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was used as a reference. Murine osteoblasts were cultured in a medium containing discs with varying amount of HA. After seven days of cell

culture RNA was extracted from the osteoblasts sticking to the discs. The level of gene expression was determined by PCR based method as follows:

Target gene: Murine Osteopontin (OP) gene (GenBank: X51834.1).

Reference gene: Murine GAPDH gene (GenBank NM_008084.2).

PCR Primers:

- Murine OP-F: 5'-TCTGATGAGACCGTCACTGC
- Murine OP-R 5'-AGGTCCTCATCTGTGGCATC
- Murine GAPDH-F 5'-AAGGTCATCCCAGAGCTGAA
- Murine GAPDH-R 5'-CTGCTTCACCACCTTCTTGA

cDNA synthesis

Primer (OP gene specific): 5'-AGGTCCTCATCTGTGGCATC

Primer (GAPDH gene specific): 5'-CTGCTTCACCACCTTCTTGA

Reaction conditions:

- 5 X reaction buffer: 4 µl
- 10 mmol dNTPs mix: 2 µl
- Primer (10 pmol/µl): 1 µl
- MMLV- Reverse Transcriptase (200 U/µl): 1 µl
- RNase inhibitor (20 U/µl) 1 µl
- Deionized water: 13 µl
- RNA: 2 µl
- Incubation: 42^{o}C for 30 minutes
- RT inactivation: 70^{o}C for 5 minutes

The reaction volumes may be reduced to suit individual requirements.

PCR amplification

1. Murine OP-F: 5'-TCTGATGAGACCGTCACTGC
   Murine OP-R 5'-AGGTCCTCATCTGTGGCATC
   Amplified product size: 170 bp
2. Murine GAPDH-F 5'-AAGGTCATCCCAGAGCTGAA
   Murine GAPDH-R 5'-CTGCTTCACCACCTTCTTGA

Amplified product size: 138 bp

3. PCR conditions:
    - SYBR green PCR mix (Thermo Fisher Scientific): 10 µl
    - Forward primer: 1 µl (10 pmol/µl)
    - Reverse primer: 1 µl (10 pmol/µl)
    - cDNA 3 µl
    - DEPC treated water: 5 µl
    - Thermal cycling (real time):
        - Initial denaturation 95^{o}C for 5 minutes
        - Forty cycles of:
            - Denaturation: 95^{o}C for 20 seconds
            - Annealing/reading: 60^{o}C for 30 seconds
            - Extension: 72^{o}C for 30 seconds

PCR may also be done in ordinary PCR buffer without SYBR green using the same protocol. The amplified product may be run on 6% polyacrylamide gel and silver staining or 1% agarose gel and ethidium bromide staining (Chapter 3).

<u>High Resolution Melting Curve (HRM) analysis:</u>

When using SYBR green method of real time PCR the distinction between specific and nonspecific amplified products becomes difficult. To control for the non-specific amplified products, it is advisable to add HRM step at the end of the real time PCR (Chapter 5).

<u>Mathematical calculations:</u>

Gene expression study is mostly done to compare the level of expression between different samples. For example, two or more samples with and without treatment (exposure) etc. The above example of OP gene expression in a murine osteoblast cell line exposed to varying concentration of hydroxyapatite (HA) is used to describe the mathematical calculations.

Real time PCR (Pfaffl method)

In the Pfaffl method change in gene expression is expressed as ratio. For example, ratio between samples with and without treatment (exposure) or before and after treatment etc. The target genes in the exposed (test) and the unexposed (control) are amplified and their Ct values are noted. A reference gene, unrelated to the target gene, is also amplified to control for the variation in the RNA concentration between the test and the control samples. The difference in the Ct values of the control and the test is called ΔCt. The relative expression ratio (R) of the target gene is calculated based on the ΔCt of the target and the reference gene as follows:

$$\Delta Ct = Ct\ control - Ct\ test$$

$$Ratio\ (R) = 2.0^{\Delta Ct\ (target)} / 2.0^{\Delta Ct\ (reference)}$$

Table 7.1. Real time PCR results of murine OP and GAPDH genes of RNA samples extracted from the osteoblasts exposed to HA concentrations of 100%, 70%, 50% and 0%.

HA	Gene	Ct test	Ct control	ΔCt[#]	$2.0^{\Delta Ct}$	Ratio (R)[@]
100%	Target (OP)	19.2	26.3	7.1	137	137/0.81= 169
	Reference (GAPDH)	20.4	20.1	-0.3	0.81	
70%	Target (OP)	21.3	26.9	5.6	48	48/0.71= 67
	Reference (GAPDH)	20.9	20.4	-0.5	0.71	
50%	Target (OP)	23.1	26.2	3.1	8.6	8.6/1.07= 8.03
	Reference (GAPDH)	20.3	20.4	0.1	1.07	
0%	Target (OP)	22.4	22.1	-0.3	0.81	0.81/1.07= 0.75
	Reference (GAPDH)	20.2	20.3	0.1	1.07	

(Ct control – Ct test)

@ $2.0^{\Delta Ct\ (target)}/2.0^{\Delta Ct\ (reference)}$

Effect of PCR efficiency

The above calculations are based on an assumption that PCR efficiency in the test and the control samples remains 100%. In the Pfaffl method an increase or decrease in the PCR efficiency may be catered for as follows:

120% PCR efficiency = $2.2^{\Delta Ct}$

110% PCR efficiency = $2.1^{\Delta Ct}$

100% PCR efficiency = $2.0^{\Delta Ct}$

90% PCR efficiency = $1.9^{\Delta Ct}$

80% PCR efficiency = $1.8^{\Delta Ct}$

Choice of reference genes

A reference gene in the gene expression studies is required for normalization of the expression of the gene under study. It serves as an internal reaction control.

An ideal reference gene should have the following qualities:

1. It should express itself without any stimulation and at a constant rate throughout the experiment.
2. It should have minimal variability of expression between tissues and physiological states of the organism under study.
3. It should be amplifiable with the same efficiency as of the test gene.

The most commonly used reference genes include the housekeeping genes (HKGs) that maintain the basic metabolic function of the cell and so are typically expressed in all cell types of an organism.

Commonly used reference genes:

1. GAPDH (glyceraldehyde-3-phosphate dehydrogenase).
2. 18S rRNA (18S ribosomal RNA).
3. 28S rRNA (28S ribosomal RNA).
4. TUBA (α-tubulin).
5. ACTB (β-actin).
6. β2M (β2-microglobulin).
7. ALB (albumin).

8. RPL32 (ribosomal protein L32).
9. TBP (TATA sequence binding protein).
10. CYCC (cyclophilin C).
11. EF1A (elongation factor 1α).
12. HPRT (hypoxanthine phosphoribosyl transferase).
13. RPII (RNA polymerase II).

Gel electrophoresis method

Gel electrophoresis for QPCR is an obsolete method because of its insensitivity therefore it is not recommended for use in gene expression studies.

Bibliography

1. Green MR, Sambrook J (2012) Molecular Cloning: A Laboratory Manual. (Fourth Edition). Cold Spring Harbor Laboratory Press.
2. Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. Journal of applied genetics, 54(4), 391-406.
3. Lowe K (2001) First strand cDNA synthesis. Oxford Practical Series. Oxford University Press.
4. Pfaffl M W. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001; 29: 2002-2009.

8

Short Tandem Repeat (STR) Analysis

Genomic DNA contains repetitive sequences at numerous places. Apparently, these sequences perform no function in the genome. The repetitive DNA sequences may be long or short. The short (2-6) base pair repetitive DNA sequences are called “Short Tandem Repeats” (STR). The STRs are randomly distributed throughout the genome and are located on all chromosomes including autosomes and sex chromosomes.

The number of repeats in a STR at a particular locus vary between individuals and the alleles are named according to the number of repeat units these may contain. The STRs are inherited in a simple Mendelian fashion. Every individual inherits one allele each from its parents. The two alleles may be identical (homozygous) or different (heterozygous or compound heterozygous).

STR nomenclature

The STR loci are named according to the number and the site on a chromosome. For example, the STR locus D5S818 is located on chromosome 5 at site 818. The

STR alleles at a locus are named on the basis of number of repeats. For example, alleles (11,15) mean 11 repeats on the chromosome from one parent and 15 repeats on its homologous pair from the other parent.

Example

An example of a tetra-nucleotide repeat (four base pair repeat) STR on chromosome 5 is shown in Figure 8.1 The sequence "GATA" is repeated several times.

PCR amplification of STRs

STRs can be amplified by a pair of primers flanking the tandem repeats. The resulting amplified products vary in length depending on the number of repeat units present (Figure 8.2). In a compound heterozygote with different number of repeats on the two chromosomes, two different sized amplified products are formed. In a homozygote with the same number of repeats on the two chromosomes two amplified fragments of the same size are produced. Whereas in a compound heterozygote with two different alleles on the two chromosomes two different sized amplified fragments would be produced. If at a tetra-nucleotide repeat locus one allele has 5 repeats of 4 bp and the other allele has 8 repeats the two amplified fragments would differ by 12 bp (4X3).

ATATTTTAATAGCAAGTATGTGACAAGGGTGATTTTCCTCTTTGGTAT
CCTTATGTAATATTTTGAA**GATA**GATA**GATA**GATA**GATA**GATA**GATA**G
ATA**GATA**GATA**GATA**GGTAGATAGAGGTATAAATAAGGATACAGATAT
AGNTACAAATGTTGTAAACTGTGGCTATGATTGGAATCACTTGGCTAA

Figure 8.1. GATA repeats at the D5S818 locus (GenBank: G08446.1) The sequence GATA is repeated 11 times.

Polyacrylamide gel electrophoresis

The amplified fragments of STRs at any given locus are only a few base pairs different. Such small differences are best resolved by Polyacrylamide gel electrophoresis. After PCR amplification the STRs are analyzed by relatively long

polyacrylamide gels. The gels are usually run for several hours. Various alleles are recognized by simultaneously running accurate DNA size standards or allelic ladders (Figure 8.2).

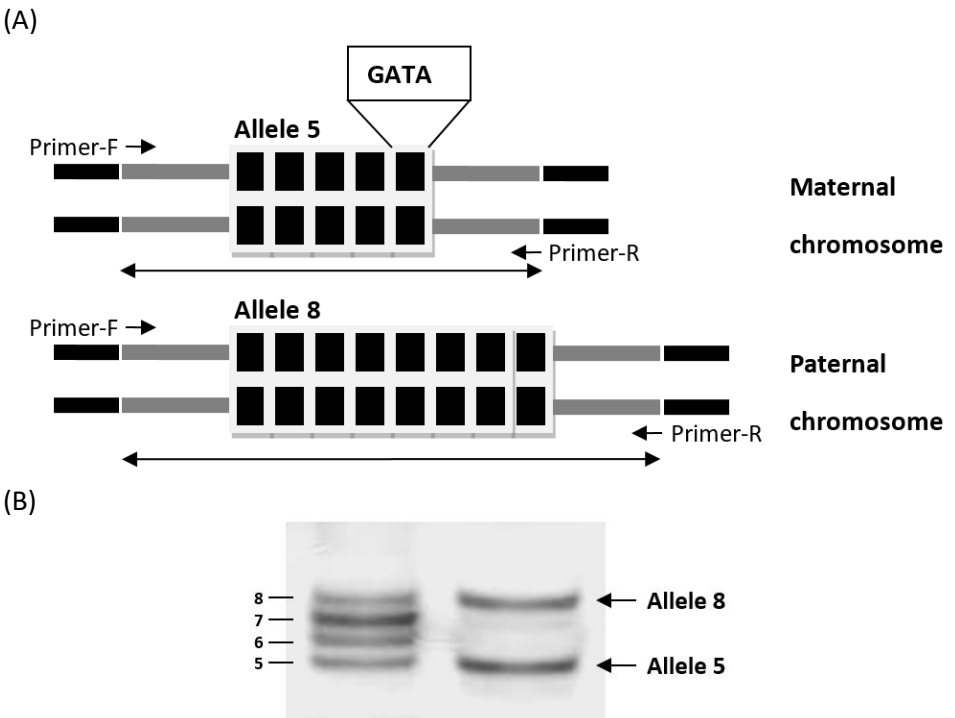


Figure 8.2. (A) PCR amplification of STR locus D5S818 having "GATA" repeats. Tandem repeats on the maternal and the paternal chromosomes are amplified by a pair of primers (F & R) flanking the repeat sequences. (B) Polyacrylamide gel electrophoresis of the amplified product and allelic ladder shows the presence of allele 5 and 8.

Genetic analyzer

The manual sizing of STR alleles is laborious and time consuming. Genetic analyzer with high voltage capillary electrophoresis provides an extremely efficient but expensive alternate to the manual method. It uses a long thin bored capillary for electrophoresis of the amplified DNA (Chapter 4). The STRs are amplified by fluorescent labeled primers and the amplified products are run with a fluorescent labeled DNA size standard in the same capillary. The genetic

analyzer detects the amplified products by laser excitation of the fluorescent dyes. It is a very sensitive and accurate method of STR allele sizing. Very small amounts of target DNA (<0.5ng) can be detected. The other advantages of genetic analyzer include high speed and analysis of multiple amplified products labeled with different fluorescent dyes in the same tube. Commercial kits with amplification of a large number of STR loci are also available.

STR analysis by manual method

Locus:	D5S818 (Figure 8.3)
GenBank accession:	G08446.1
Forward primer:	5'-AAGGGTGATTTTCCTCTTTGG
Reverse primer:	5'- AGCGCTTTTTAGCCAAGTGA
Amplified product:	136-176 bp (1-11 repeats)

ATATCTAATTAAAGTGGTGTCCCAGATAATCTGTACTAATAAAAGTATATTT
TAATAGCAAGTATGTGAC**AAGGGTGATTTTCCTCTTTGG**TATCCTTATGTAA
TATTTTGAA**GATA**GATA**GATA**GATA**GATA**GATA**GATA**GATA**GATA**GATA**GAT**
AGATAGATAGAGGTATAAATAAGGATACAGATATAGNTACAAATGTTGTAAA
CTGTGGCTATGATTGGAA**TCACTTGGCTAAAAAGCGCT**NAAGCNTTCCTCTG

Figure 8.3. PCR primers for amplification of the GATA repeat at D5S818 locus (GenBank accession: G08446.1).

Reaction conditions

- Reaction volume: 25 µl
- Primer concentration: 1 µl (5 pmol each/µl)
- Taq polymerase: 0.5 units (0.1µl)
- Template DNA: 1µl (200ng)
- Thermal cycling:
    - Initial denaturation: 1 minute at 94°C
    - No. of cycles: 30

      Denaturation: 48 seconds at 93°C

Annealing:	48 seconds at 60°C
Extension:	1 minute at 72°C
Final extension:	3 minutes at 72°C

- Electrophoresis: 16 X 20 X 0.1cm 6% polyacrylamide gel.
  Load 3µl amplified product in 3µl loading dye
  Include 3µl allelic ladder in 3µl loading dye
  Run for 5 hours at 200 volts.
- Staining: 0.1% Silver nitrate.
- Result: Figure 8.4.

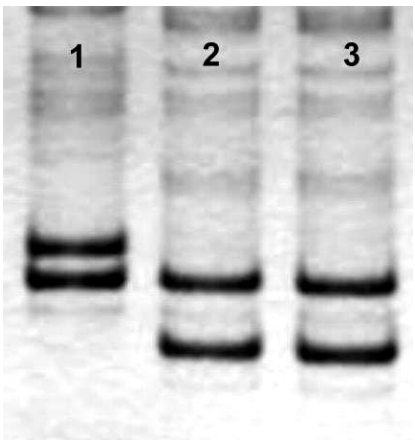


Figure 8.4. Polyacrylamide gel electrophoresis of PCR amplification products at D5S818 locus.

Preparation of allelic ladder for STR analysis by gel electrophoresis

1. Select DNA samples of known genotypes at different loci previously analyzed on genetic analyzer. The selected samples should represent the alleles usually found in the target population.

2. Amplify each DNA sample in multiples of five at the locus for which the ladder is required. Use the amplification protocol described above.
3. Pool the amplified products of all the samples.
4. The amplified DNA is a potent source of cross contamination therefore it should be handled only with pipettes marked as “amplified DNA only”.
5. Mix the pooled amplified products and make aliquots of 100µl.
6. Store at -20^{o}C.
7. Use 3µl of the pooled amplified products (allelic ladder) for each electrophoresis run (Figure 8.5).

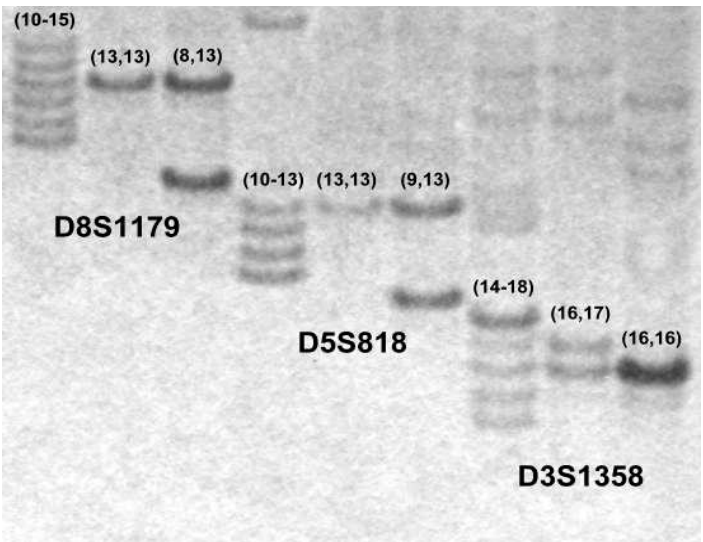


Figure 8.5. Silver-stained polyacrylamide gel electrophoresis of PCR amplified products at D8S1179, D5S818 and D3S1358 loci and the allelic ladders. The allele sizes in each allelic ladder and two unknown samples are shown in brackets.

PCR primers for amplification of STR loci

The lists of primers for amplification of autosomal and Y-chromosomal STRs are given in Table 8.1 & 8.2.

Table 8.1. PCR primers for CODIS core STR Loci.

Locus	Primer sequence
D3S1358-F	5'-ACTGCAGTCCAATCTGGGT
D3S1358-R	5'-ATGAAATCAACAGAGGCTTG
D5S818-F	5'-GGGTGATTTTCCTCTTTGGT
D5S818-R	5'-TGATTCCAATCATAGCCACA
D7S820-F	5'-TGTCATAGTTTAGAACGAACTAACG
D7S820-R	5'-CTGAGGTATCAAAAACTCAGAGG
D8S1179-F	5'-TTTTTGTATTTCATGTGTACATTCG
D8S1179-R	5'-CGTAGCTATAATTAGTTCATTTTCA
D13S317-F	5'-ACAGAAGTCTGGGATGTGGA
D13S317-R	5'-GCCCAAAAAGACAGACAGAA
D16S539-F	5'-GATCCCAAGCTCTTCCTCTT
D16S539-R	5'-ACGTTTGTGTGTGCATCTGT
D18S51-F	5'-CAAACCCGACTACCAGCAAC
D18S51-R	5'-GAGCCATGTTCATGCCACTG
FGA-F	5'-GCCCCATAGGTTTTGAACTCA
FGA-R	5'-TGATTTGTCTGTAATTGCCAGC
TPOX-F	5'-CACTAGCACCCAGAACCGTC
TPOX-R	5'-CCTTGTCAGCGTTTATTTGCC
CSF1PO-F	5'-AACCTGAGTCTGCCAAGGACTAGC
CSF1PO-R	5'-TTCCACACACCACTGGCCATCTTC
TH01-F	5'-GTGGGCTGAAAAGCTCCCGATTAT
TH01-R	5'-ATTCAAAGGGTATCTGGGCTCTGG
vWA-F	5'-CCCTAGTGGATAAGAATAATC
vWA-R	5'-GGACAGATGATAAATACATAGGATGGATGG
Amgl-F	5'-ACCTCATCCTGGGCACCCTGG
Amgl-R	5'-AGGCTTGAGGCCAACCATCAG

Table: 8.2. PCR primers for Y-STR loci.

Locus	Primer sequence
DYS19-F	5’-CTACTGAGTTTCTGTTATAGT
DYS19-R	5’-ATGGCCATGTAGTGAGGACA
DYS385a/b-F	5’-AGCATGGGTGACAGAGCTA
DYS385a/b-R	5’-TGGGATGCTAGGTAAAGCTG
DYS389-I/II-F	5’-CCAACTCTCATCTGTATTATCTAT
DYS389-I/II-R	5’-TCTTATCTCCACCCAGA
DYS390-F	5’-TATATTTTACACATTTTTGGGCC
DYS390-R	5’-TGACAGTAAAATGAACACATTGC
DYS391-F	5’-CTATTCATTCAATCATACACCCA
DYS391-R	5’-GATTCTTTGTGGTGGGTCTG
DYS393-F	5’-GTGGTCTTCTACTTGTGTCAATAC
DYS393-R	5’-AACTCAAGTCCAAAAAATGAGG
DYS438-F	5’-TGGGGAATAGTTGAACGGTAA
DYS438-R	5’-GTGGCAGACGCCTATAATCC

STR analysis by Genetic analyzer

A method for multiplex analysis of three STR loci (D13S631, D18S51 and D21S11) is described. Primers for the three loci are described in Figure 8.5. One primer of each pair is labeled with a fluorescent dye. Since the amplified products of the three loci do not overlap the three primers are labeled with the same fluorescent dye (FAM).

Loci:	D13S631, D18S51, D21S11	
Primer mix:		
D13S317-F	6 FAM 5’-ACAGAAGTCTGGGATGTGGA	10 pmol/µl
D13S317-R	5'-GCCCAAAAAGACAGACAGAA	10 pmol/µl
D18S51-F	6 FAM 5'-CAAACCCGACTACCAGCAAC	10 pmol/µl
D18S51-R	5'-GAGCCATGTTCATGCCACTG	10 pmol/µl
D21S11-F	6 FAM 5’-TGAGTCAATTCCCCAAG TGAA	10 pmol/µl

D21S11-R	5’-AGTCAATGTTCTCCAGAGACAGAC	10 pmol/µl
Amgl-F	6 FAM 5’-CTGATGGTTGGCCTCAAGCCT	10 pmol/µl
Amgl-R	5’- ATGAGGAAACCAGGGTTCCA	10 pmol/µl

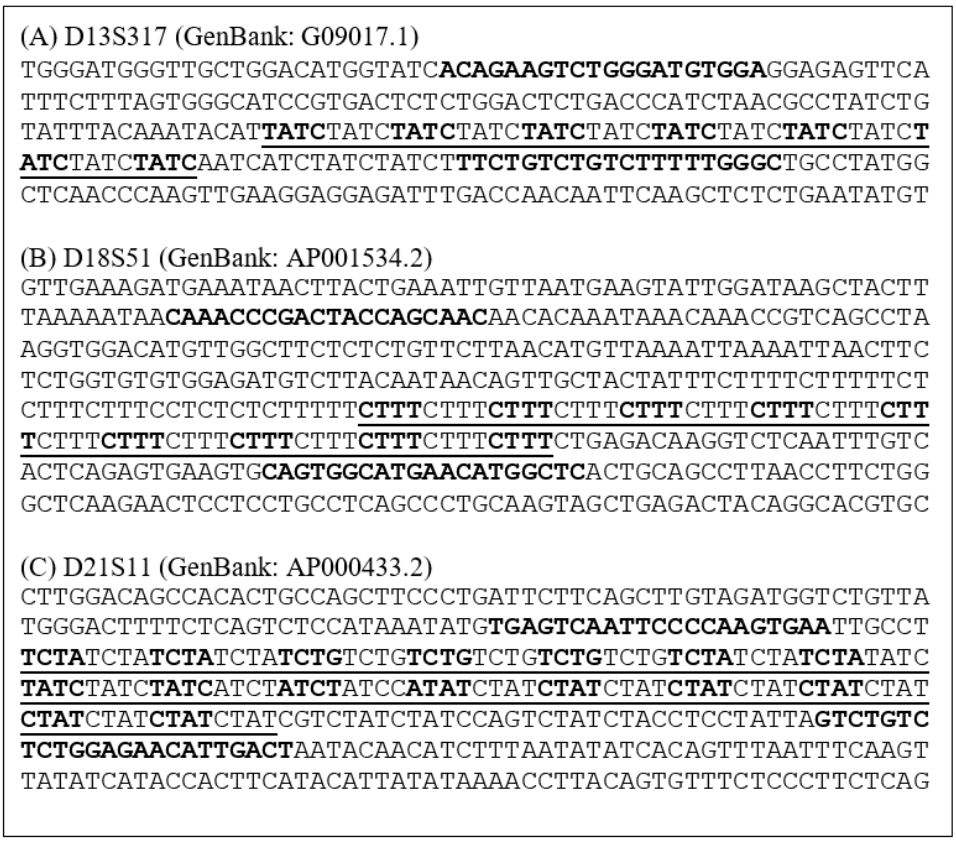


Figure 8.5. PCR primers for D13S631, D18S51, D21S11 STR loci.

For each sample label four tubes A, B, C and D (one each for the three STR loci and one for Amgl).

Reactions conditions:

PCR mix:	23 µl
Respective primer mix:	1 µl
Taq polymerase:	0.1 µl (0.5 units)
DNA:	1 µl (20ng)

Thermal cycling:

Initial denaturation:	1 minute at 94°C
No. of cycles:	25
Denaturation:	48 seconds at 93°C
Annealing:	48 seconds at 60°C
Extension:	1 minute at 72°C
Final extension:	3 minute at 72°C

Pool 5 µl amplified product from each of the 4 reactions and use 1.3 µl of the pool.

HiDi formamide:	10 µl/sample
GeneScan Liz:	0.5 µl/sample
Denature in thermal cycler:	94°C for 4 minutes

Run on Genetic Analyzer as per the manufacturer instructions.

Some technical problems with STR analysis

PCR stutter products

In amplification of STRs the elongating DNA strand can slip one or more steps forward resulting in formation of end products that are one or more repeat units smaller in length. This is also called slippage synthesis. On gel electrophoresis the shorter amplified products appears as shadow bands or peaks often called stutter bands or peaks. The stutter products longer than the actual product can also result by backward slippage but these are uncommon.

Stutter bands/peaks are most commonly seen in amplification of di-nucleotide repeats but these are also common in PCR of tetra-nucleotide repeats.

Twin peaks of "+A" and "–A" products

Taq polymerase has the property of adding +A at the 3' end of the elongating strand that results in formation of products one base pair longer. The latter may appear as twin peaks on a genetic analyzer.

Quantity of DNA

Too much or too little DNA in PCR can give false negative results. The quantity of DNA is especially important for analysis on a genetic analyzer. The analyzer works best with DNA quantity ranging from 0.5-1 ng per reaction.

The quantity of DNA is less critical in gel electrophoresis and silver nitrate staining. It gives good results at 50-200ng DNA per reaction. The minimum amount of DNA that can be seen on a gel by this method is ~20ng (Figure 8.6).

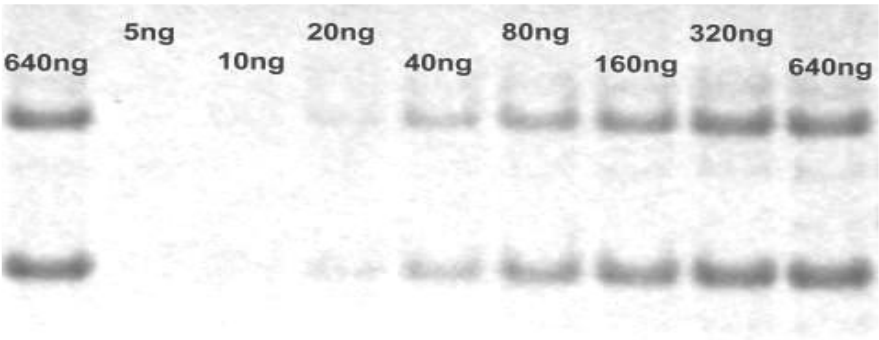


Figure 8.6. Amplification of STR locus D8S1179 by using decreasing concentration of DNA. The 6% polyacrylamide gel was stained in silver nitrate. Minimum detectable amount of DNA is ~20ng.

Bibliography

1. Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 1998 62: 1408-1415.
2. Short Tandem Repeat DNA Internet DataBase, http://www.cstl.nist.gov/strbase/
3. Smith RN (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. *Biotechniques* 18: 122-128.

9

Genomic Sequencing

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

Sequencing by automated Sanger's method

In a diagnostic lab sequencing is mostly done to see the presence of a variant in an already known sequence of DNA. This is also called "re-sequencing".

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 the 3' end with the next dNTP. In a sequencing reaction the process is randomly terminated and results 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 fluorochromes 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 the 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

In 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 (Figure 9.1). At the end of cycling the DNA strands of varying lengths are analyzed on a genetic analyzer by capillary electrophoresis.

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 (10 pmol/µl)
- Taq polymerase: 1 unit (0.2µl)
- Template DNA: 2 µl (~300ng)
- Thermal cycling:
    - Initial denaturation: 1 minute at 94°C
    - Cycles: 25
    - Denaturation: 1 minute at 94°C
    - Annealing: 1 minute at 65°C

- Extension: 1 minute 30 seconds at 72°C
- Final extension: 3 minutes at 72°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

3'-GAGTCGTACGCTACCATGTCATATCAACCATAGTTCCGATCCTAGTACTAAC-5'
5'-**CATGCGATGGTACAGTATAG**-3'→

3'-GAGTCGTACGCTACCATGTCATATCAACCATAGTTCCGATCCTAGTACTAAC-5'
5'-**CATGCGATGGTACAGTATAG**-**ddTTP**
5'-**CATGCGATGGTACAGTATAG**T-**ddTTP**
5'-**CATGCGATGGTACAGTATAG**TT-**ddGTP**
5'-**CATGCGATGGTACAGTATAG**TTG-**ddGTP**
5'-**CATGCGATGGTACAGTATAG**TTGG-**ddTTP**
5'-**CATGCGATGGTACAGTATAG**TTGGT-**ddATP**
5'-**CATGCGATGGTACAGTATAG**TTGGTA-**ddTTP**
5'-**CATGCGATGGTACAGTATAG**TTGGTAT-**ddCTP**
5'-**CATGCGATGGTACAGTATAG**TTGGTATC-**ddATP**
5'-**CATGCGATGGTACAGTATAG**TTGGTATCA-**ddATP**
5'-**CATGCGATGGTACAGTATAG**TTGGTATCAA-**ddGTP**
5'-**CATGCGATGGTACAGTATAG**TTGGTATCAAG-**ddGTP**

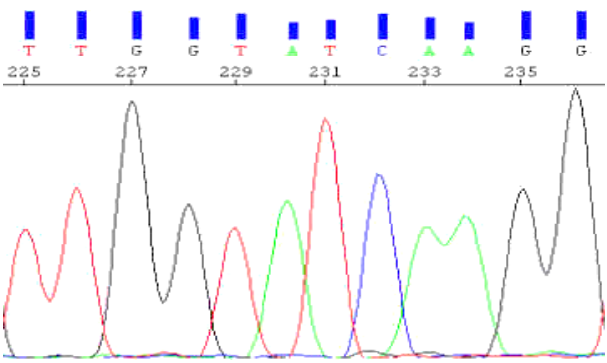


Figure 9.1. DNA template and the sequencing primer. The newly formed DNA strand is terminated when a ddNTP instead of the dNTP is incorporated. The four ddNTPs are tagged with different fluorescent dyes. The DNA strands of varying lengths are separated by capillary electrophoresis. Each fragment is represented by a coloured peak on electropherogram.

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.

Gel purification

1. It is tedious but efficient in removing non-specific amplified products.
2. 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.

Enzymatic digestion

1. 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.
2. 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.

Column purification

1. 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.
2. 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.

3. 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. DNA in the template may be quantified using different methods (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. These may cause dye blobs to appear in the electropherograms and should 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.

Next generation sequencing (NGS)

NGS involves sequencing of DNA in a massively parallel fashion. The sequencing chemistries are similar to the cycle sequencing in which a DNA polymerase incorporates fluorescently labelled dNTPs on a DNA template strand in successive cycles of synthesis. The addition of each dNTP is recognized by fluorophore excitation. Instead of a single template in the cycle sequencing, NGS involves sequencing of millions of overlapping templates in parallel.

The NGS technology involves library preparation, cluster generation, sequencing, and data analysis (Figure 9.2).

Library Preparation

The library or template preparation involves controlled ultrasonic shredding of genomic DNA or cDNA into smaller pieces of 90-500 bp length depending on the type of application. The DNA fragments are ligated with synthetic short pieces of DNA called adapter molecules. The adapter ligated fragments are then PCR amplified and purified either by gel electrophoresis or other methods. A more

efficient process, called tagmentation, combines fragmentation and adapter ligation in a single step.

Library preparation is the most laborious part of NGS. It is also the most critical step as only a neat and clean sample can ensure quality results. Commercial kits for various steps in library preparation are available from the instrument manufacturers and other sources. The process of library preparation can also be automated to avoid errors.

Multiplexing

A larger number of libraries from different test samples can be multiplexed (pooled) by tagging with unique adapter molecules. The results from all such samples can be sorted out in the final data analysis. The multiplexing of samples makes NGS more time and cost effective.

Cluster Generation and Bridge Amplification

In cluster generation the library of templates is captured through oligonucleotides complementary to the adapter molecules bound to either a solid surface, or individual agarose beads that are later converted to form an oil emulsion. The net result is creation of millions of spatially distinct partitions. The templates anchoured to a surface are amplified through bridge amplification that results in spatially distinct clusters of billions of molecules. The clusters are now ready to be sequenced.

Sequencing

The clusters are sequenced by fluorescent labeled dNTPs that are incorporated specifically. The addition of each dNTP generates a specific fluorescence that is imaged to produce the sequencing data.

Data Analysis

The data analysis involves aligning of the multiple short sequence reads with a reference sequence (Figure 9.3). Any variation in the sequence is identified and reported (Figure 9.4). The NGS technology also allows sequencing of both ends

of the fragments. This allows greater accuracy especially for the single nucleotide variants (SNVs).

NGS Coverage

The average number of reads that align to, or "cover," known reference bases is called NGS coverage. For example 50 Mb of sequence data for a 5 Mb genome would be 10X coverage. It determines whether a variant can be identified with a certain degree of confidence at particular base position. A higher coverage would increase the accuracy but would also increase the amount of data to be analyzed. For publishing human genome variants 10× to 30× depth of coverage is recommended.

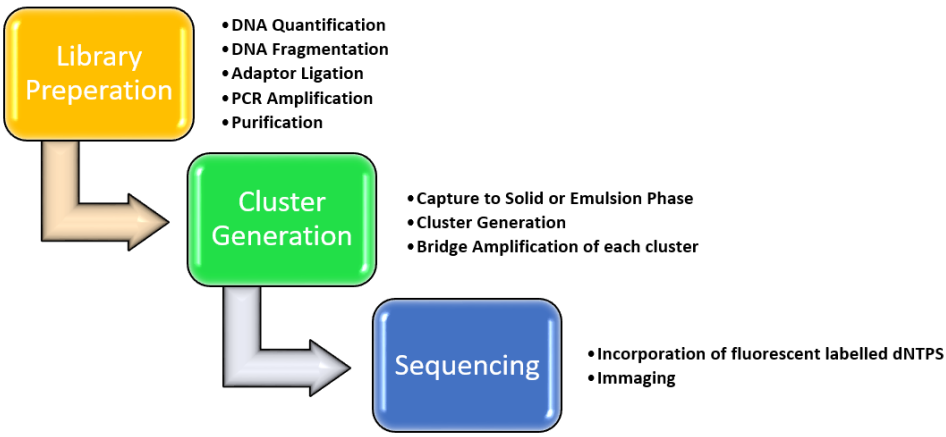


Figure 9.2. The process of NGS comprises library preparation, cluster generation and sequencing. The sequenced data is further analyzed by bioinformatics software.

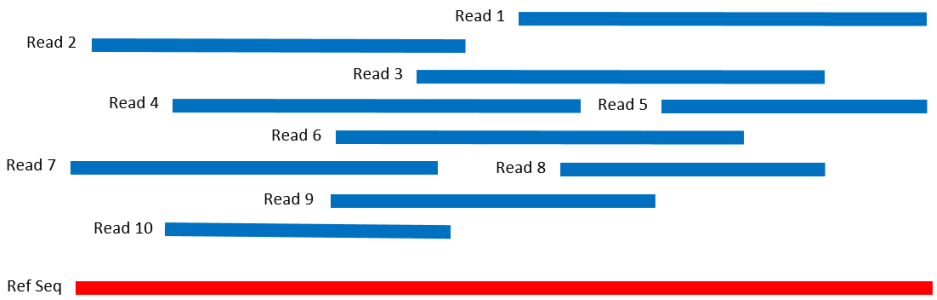


Figure 9.3. The individual overlapping NGS reads (blue) are aligned with a reference sequence (Ref Seq) as shown in red.

```
Read 1:                  AGCTTGAGCGACTTAGCAATT
Read 2:                   GCTTGCGCGACTTA
Read 3:                      GAGCGACTTAGCAAT
Read 4:                TAAGCTTGAGCGACTTAGC
Read 5:                   GCTTGAGCGACTTAGCAA
Reference Sequence: CGATAGCTTGCGCGACTTAGCAATTT
```

Figure 9.4. The difference between the reference sequence and the NGS reads (shown in blue and red) are picked by the bioinformatics software.

NGS Applications

NGS is being used in a wide variety of applications including genomics, transcriptiomics, and epigenomics.

Genomics (DNA Sequencing)

Whole Genome Sequencing (WGS)

NGS can be used to sequence the entire 3.2 billion bases of human genome. Similarly, the whole genomes of all known species can also be sequenced. This has opened new vistas in research and diagnostics.

Whole Exome Sequencing (WES)

The protein coding regions of DNA are called exomes. There are about 180,000 exons that constitute about 1% of the human genome. NGS can be used to selectively sequence only the exomes and the process is called whole exome sequencing (WES). This provides a cost-effective method of screening for the disease-causing genes and their variants. A disadvantage of WES is that it is unable to provide information about variants in the introns (intervening sequences) that may be of clinical significance.

The WES involves an initial step of capturing the fraction of DNA containing the exomes. Genomic DNA is first fragmented to pieces and the fragments of interest are captured to microarrays layered with specific probes. The microarray-based capturing has been replaced by more efficient in-solution capture techniques in which the DNA is captured with probes labelled to beads in solution. Several commercial kits are available for exome capturing. The captured DNA is amplified and sequenced.

Targeted Sequencing (amplicon sequencing)

The targeted NGS has the best potential for use in diagnostic pathology. Specific areas of the genome or groups of genes can be isolated by using biotinylated probes and capture through magnetic beads. Target enrichment can also be achieved by “amplicon sequencing” that involves amplification and purification of the regions of interest by highly multiplexed PCR oligo sets. This can allow sequencing of 16-1536 targets at a time. The approach is highly suitable for discovery, validation or screening of genetic variants. Several targeted NGS kits are available with preformed content. The kits can also be ordered as custom designed panels. Besides being cost-effective targeted NGS has the advantages of easy data analysis, high resolution, and accuracy. The targeted NGS can easily achieve 500x – 1000x coverage.

Amplicon sequencing targeting the bacterial 16S rRNA gene present in multiple species is commonly used for specie identification in a diverse metagenomic sample.

De-novo NGS

In de-novo NGS the sequencing data is analyzed without a reference sequence. It is mostly required in research.

RNA Sequencing (Transcriptomics)

RNA sequencing begins with isolation of total cellular RNA. This is followed by steps to remove the unwanted fraction of RNA. In the mixture of cellular RNA selective removal of ribosomal RNA, mRNA, small RNA, and micro-RNA etc. one can focus on the RNA of interest. The RNA of interest is then converted to cDNA which is subjected to NGS.

Total RNA and mRNA sequencing provide a complete snapshot of gene expression of the entire genome. RNA sequencing by NGS can also be focused on expression of few selected genes or sequencing of micro-RNA.

Epigenetics

Epigenetics is the study of heritable changes in gene activity without involving changes in the DNA. For example, the expression of genes may be affected by their methylation status, regulation through micro-RNA or protein DNA interaction etc. The epigenetic phenomena can be studied through modifications of the NGS technologies.

NGS Service Providers

The NGS technology is evolving at a very rapid pace and the instruments may become obsolete in a relatively short time. Therefore, in countries with resource constraints it may not be advisable to invest heavily in such instrumentation.

There are many good quality NGS service providers whose services may be used. The DNA samples can be submitted to a service provider who can analyze the samples and return the data files that may be analyzed by suitable software.

Bibliography

1. An Introduction to Next Generation Sequencing Technology. Illumina Publication No. 770-2012-008 25 May 2016. (www.illumina.com/technology/next-generation-sequencing.html)
2. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-termination inhibitors. Proc Natl Acad Sci, USA 74: 5463-5467.
3. Murray V (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res 17: 8889.
4. Metzker ML (2010) Sequencing technologies-the next generation. Nat Rev Genet 11: 31-46.

10

Quality Control in Diagnostic PCR

In a PCR lab quality control is mainly concerned with the control of errors in performance of tests and verification of test results. It includes internal and external quality control. The internal quality control includes monitoring of laboratory performance by using control material and/or repeat measurements.

Errors in sample collection, transport and extraction of nucleic acids

DNA is fairly stable when it is in the cell. The cellular DNA may get degraded by DNAse derived from bacterial contamination of the sample. The sample for DNA extraction should be processed before bacteria grow in it. The sample should be collected in a sterile container. But if the sample is already contaminated then it's collection in s sterile container would not help. DNA is a very large and fragile molecule. It can easily get broken by vigorous shaking during extraction.

Fragmented or degraded DNA is not good for PCR. It is a common reason for poor quality of results.

RNA is far more sensitive to degradation than DNA. It can easily be degraded by RNAse present in the environment especially that derived from the cells from where the RNA is being extracted. Since RNA extraction takes place in a strong denaturant that renders RNAse inactive, the integrity of RNA is mostly at risk prior to and after its extraction. Therefore, in RNA extraction the handling of sample prior to isolation and the storage of isolated RNA are very critical. The extracted RNA should be analysed as early as is possible. RNA can be stored at -20^{o}C for a few days but longer storage should be done at -80^{o}C.

Inhibitors of PCR

Heparin, porphyrin, SDS, phenol, and proteinase-K are potent inhibitors of PCR. SDS and phenol must be completely removed from the sample in nucleic acid extraction. The blood samples should be collected in EDTA. Heparin should not be used as anticoagulant as it inhibits PCR. Proteinase K can be inactivated by heating. PCR may also be inhibited by chocolate that might contaminate the DNA extracted from a mouthwash. The effect of inhibitors can be reduced or abolished by diluting the sample as it also dilutes the inhibitor.

Errors in PCR

There are numerous factors that may affect the efficiency of PCR. The quality of reagents, primers, buffer pH, quality and quantity of DNA and PCR tubes etc. are all important.

1. The quality of Taq polymerase is important in diagnostic work. The cheaper brands of Taq may be alright for research work but in diagnostic work it may give false positive results.
2. The primer, especially its design, is the most important determinant in specificity of PCR (Chapter 3). A poorly designed primer may give false positive or a false negative result.
3. The concentration of DNA in PCR is also important. Apart from the quality, too much or too little DNA can give false negative results.

4. While setting up PCR it is always better to premix reagents. If ten samples are to be amplified prepare reaction mix in one master tube and then transfer it to ten individual tubes labeled 1-10. This helps in avoiding inconsistency in pipetting etc. The DNA should always be added at the end.
5. The quality of PCR tubes is often ignored. The tube should be of very thin wall that allows quick transfer of heat. The tube should fit well in the sample block. A loose-fitting tube will have air around it that can interfere with heat transfer. Poor quality PCR tubes may collapse at high temperature. It's not uncommon for the lid of the tube to open leading to evaporation of the contents.

False positive PCR results

PCR is an extremely sensitive technique. In thirty cycles one molecule of DNA can be amplified a billion-fold. The amplified DNA is an excellent template for re-amplification. Carryover of the amplified DNA from one sample to another is a potent source for further amplification and false positive PCR results. It is a serious problem that is very difficult to handle.

Identifying false positives

The false positives are identified by including good quality negative, positive and non-target controls (NTC). In diagnostic PCR it is mandatory to include at least one negative and a positive DNA control. It is also essential to include NTC that contains everything except DNA (reagent blank). The NTC is included to exclude contamination of the reagents by extraneous DNA. If the NTC shows amplification the result of complete batch of samples becomes null and void.

Preventing false positives

1. Sample preparation, amplification and end point analysis should be physically separated. The samples may be prepared in a separate room or in a bio-safety cabinet.

2. The place should be irradiated with UV light from time to time or when contamination is suspected. UV light can destroy DNA rendering it unsuitable for further amplification. The lab benches, pipettes and other plastic ware should be cleaned periodically with fresh 10% bleach solution.
3. Separate sets of pipettes should be used for the three areas. The extracted DNA or RNA is less likely to be carried over from one sample to another. The carryover is mostly caused by amplified DNA. In a PCR lab with resource constraints if separate pipettes cannot be used for each area, then at least the amplified DNA should be handled with separate pipettes clearly marked “amplified DNA only”. Barrier filter tips are also useful in preventing carryover due to aerosol spray of the amplified products.
4. Disposable gloves should be worn to prevent carryover of amplified DNA from one area to another especially the sample preparation area.
5. Re-amplification of any contaminating amplified DNA from a previous PCR can be avoided if dUTP instead of dTTP is used in PCR. The dUTP containing amplified products can be rendered un-amplifiable by enzymatic degradation of uracil with Uracil DNA Glycosylase (UNG).

Handling false positives

Contamination with carryover amplified DNA is the commonest and the most serious hazard in a PCR Lab. Prevention is better than cure. All possible precautions should be taken in preventing carryover contamination. Once developed it is difficult and time consuming to get rid of it. To investigate errors due to contamination by extraneous DNA proceed as follows:

1. The most often contaminated reagents include working solutions of DNA/RNA extraction, primers, PCR buffer and Taq polymerase. Less often the stock solutions may also be contaminated.
2. In a stepwise manner carry out the PCR with “in-use” reagents but replacing one item at a time from the fresh stock solutions.
3. Discard the contaminated reagent if, by luck, one is able to find it!

4. In a busy diagnostic lab there may not be enough time to investigate the source of contamination. In that case the best choice would be to discard all remaining quantities of the working reagents and prepare fresh from the stock solutions. Use fresh vial of Taq polymerase. In the mean while de-contaminate the work area etc. with 10% bleach solution and UV light for 30-60 minutes. The pipettes are the usual source of carry over and should be thoroughly cleaned with bleach and irradiated with UV light.

False negatives

The false negatives are also serious problems that often do not get enough attention. These are best detected by inclusion of good quality positive, negative and internal controls. Inclusion of an internal PCR control is always good to reduce the risk of false negatives. This is usually done by including a separate set of primers in the reaction to amplify an unrelated part of the DNA. A PCR result that shows positive internal control but no amplification of the target DNA is called a true negative. Failure of the internal control and the target to amplify indicates a false negative result.

The false negatives may be caused by any of the following:

1. Degraded, too little or too much DNA.
2. Poor quality of reagents.
3. Degraded primers are the most common cause of poor or no amplification. The reconstituted primers tend to degrade over time especially when these are frequently thawed and frozen. It is better to make fresh dilution of primers every 2-3 weeks.
4. Failure to add any of the reaction constituents like primer, Taq polymerase or DNA.
5. Problem with the PCR machine.

<u>Avoiding false negatives</u>

1. Observe precautions in extraction to avoid degradation of DNA or RNA.
2. Use optimum concentration of DNA/RNA.

3. Use good quality of reagents that are not expired. In resource constrained labs reagents may be used beyond expiry if their potency is proven by running positive controls.
4. The reagents like PCR buffers and primers etc. should be stored in aliquots. This helps in reducing loss of quality due to repeated freeze thawing. Moreover, if an aliquot of reagents is contaminated it can be conveniently discarded.
5. Wearing disposable gloves helps in preventing DNAse present on the skin from degrading DNA.
6. The PCR mix contains several ingredients whose quality and quantity can be detrimental. For example, insufficient amount of Mg^{++} can result in poor amplification. Increasing the concentration of Mg^{++} can increase amplification but too much of Mg^{++} can result in false positives. Similarly poor quality of dNTPs, degraded primers, and Taq polymerase can also cause false negatives.

PCR Machine

In a PCR lab thermal cycler gets the least attention. Thermal cycler like any other lab equipment requires performance evaluation from time to time. The commonest error is the difference in the temperature of the block/reaction tube and that displayed on the screen. If the block temperature is lower than the displayed temperature primers can anneal non-specifically. The result would be non-specific amplification. Higher block temperature than the displayed would result in higher denaturation and annealing temperatures. The former could rapidly denature Taq polymerase while the latter would result in difficulties in primer annealing. The net result is reduction in the amplification.

The thermal cyclers should be periodically checked for the block and the displayed temperatures. Good quality digital temperature meters should be kept in a PCR lab. Use the following procedure to check the temperature in the reaction tube:

1. Place five PCR tubes in the heating block of the thermal cycler (four tubes in each corner well and the fifth tube in the centre well).

2. Add 50µl water to each of the tubes and keep their lids open.
3. Turn on the thermal cycler and set to hold temperature for five minutes at 65°C.
4. Once the temperature has reached 65°C allow one minute for temperature of water in the tubes to reach 65°C.
5. Measure temperature in each of the five tubes by dipping the sensor in the water in each tube.
6. The temperature in all of the tubes should be equal and at 65°C.
7. The machine needs calibration if there is any difference of temperature between the tubes and that displayed on the screen. Calibration is done by entering the new calibration factor in the software if this right is given to the user. Otherwise, the help of a service engineer would be required. If none of the options are available one could use temperature settings in the thermal cycling programme making "+" or "-" adjustment for the difference in temperatures of the tubes and the display.
8. If there is a difference of >1.0°C in the temperature of individual tubes the machine needs servicing by a qualified engineer.

Bibliography

1. Kwok S, Higuchi R (1989) Avoiding false positives with PCR. Nature 339: 237-238.
2. Rys PN, Persing DH (1993) Preventing false positives: Quantitative evaluation of three protocols for inactivation of polymerase chain reaction amplification products. J Clin Microbiol 31: 2356-2360.

11

Setting-up a PCR Lab

PCR is a relatively new technique for most diagnostic pathology laboratories. Therefore, one is required to create a space for PCR facility in an already established lab. No matter how small the PCR setup is it should preferably be housed in a separate room or a cabin. The size of the lab depends on the range of applications and the quantum of workload.

Whether the PCR lab is small, medium or large its basic format remains the same. It is essential to divide or at least restrict the flow of work to three separate areas including specimen preparation, PCR setup and post PCR processing. A suggested layout for a medium sized PCR lab is shown in Figure 11.1.

Specimen preparation area

The specimen preparation area is dedicated for receiving and processing all specimens. Special precautions are also required if any of the infectious samples are being received. No handling of amplified product should be done in this area. The equipment used in this area especially the pipettes etc. should not be used in the other areas of the lab. The area should be regularly cleaned and decontaminated with bleach and UV light.

PCR setup area

This area is meant for setting up the PCR. It should be maintained free of contamination from sample preparation and post PCR processing areas. The pipettes and the other plastic ware etc. should not be used in the other areas. If possible, each worker should have its own bench space and separate set of pipettes etc. Handling of the amplified products should be strictly forbidden in this area. This area may also be used for reagent preparation. The PCR setup area should be regularly cleaned and decontaminated with bleach and UV light. A safety cabinet with built in UV light source may be used if resources are available. The thermal cyclers should preferably be kept in this area.

Post PCR processing area

This area is used for post PCR processing of the amplified products. If the lab is using only the real time PCR then there is no need to keep this area. It is mostly required for gel electrophoresis of the amplified products. This is potentially the most dangerous area that may spread contamination to the rest of the work areas. It should be the most isolated place in the PCR lab and may conveniently be kept outside the PCR room close to a water basin. Ideally the person handling the PCR amplified products should not be allowed to work in the sample preparation or the PCR setup area. The pipettes used in this area must be marked “amplified DNA only” and must not be used anywhere else in the PCR lab. In a small lab the entire procedure of PCR can be done in a single room. The areas may be demarcated or physically separated by creating partitions. Whether PCR is done in separate rooms or separate areas in a single room the flow of work must remain uni-directional.

Equipment required for a PCR lab

The basic equipment required for setting up a PCR lab is listed in Table 11.1.

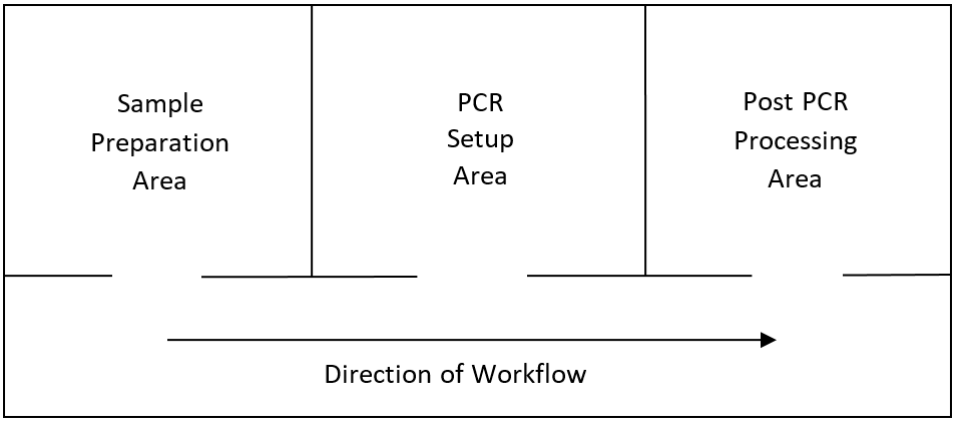


Figure 11.1. Suggested layout of a PCR laboratory.

Table 11.1. List of basic equipment for a medium sized PCR lab.

Equipment	Quantity
DNA Thermal cycler	2
Micro-centrifuge (up to 15000 rpm)	2
Vortex mixer	2
Analytical balance (0.00g)	1
Adjustable pipettes (full range)	4 sets
Poly-acrylamide gel electrophoresis apparatus	2
Submarine agarose gel electrophoresis tanks	2
DC Power Supply (100-300 volts)	2
Water bath	1
Gel dryer with vacuum pump	1
Gel documentation system	1
Refrigerator (-20^{o}C)	1
Real time PCR machine (optional)	1

Thermal cycler

It is the most important piece of equipment in a PCR lab. A wide variety of instruments are available with widely ranging prices. In general machines that can accommodate greater numbers of tubes and have higher ramping rates (ability to rapidly cool or heat) are more expensive. Such machines are usually suitable for labs with very large quantum of work.

One should buy a machine that best suits its requirements. If the daily work load of a lab is between 10-20 samples there is no point in buying an expensive machine with capacity to hold 96 samples or to have a machine with very high ramping rates. In a diagnostic lab a spare PCR machine should be available as backup. A thermal cycler with the option of thermal gradient may be useful in research work but it is rarely required in a diagnostic lab.

Real time PCR machine

A real time PCR machine with capacity to detect greater number of fluorescent dyes (colours) is more expensive. While choosing a real time PCR machine one should keep the requirements in mind. A diagnostic lab mostly involved in real time PCR of infectious diseases should buy a machine with two colour detection capability. It is good enough for doing TaqMan® probe based real time PCR. The four/five colour machines are usually required for genotyping and detection of multiple targets etc.

Measures to protect against power breakdowns

1. Frequent power breakdowns can adversely affect the overall functioning of the lab. All sensitive equipment especially the thermal cyclers must have a backup electric supply. Un-interrupted Power Supply (UPS) that generates square wave electric current can be harmful for the equipment with moving parts. All such equipment should have a sine wave UPS. The UPS should be of appropriate power. Lab equipment like incubators, water baths, and water distillation plants that use maximum electric power should not be placed on UPS.

2. Most DNA thermal cyclers and Real time PCR machines tend to loose data in case of interruption due to power failure. This could result in considerable wastage of expensive consumables. Therefore, all such equipment must be run on a suitable UPS with a battery backup of at least two hours.

3. Loose power connections can be a source of instrument malfunction. Adequate attention must be paid to electric wiring. Loose connections and inappropriate electric wiring can also be a fire hazard in a lab.

4. In a large or medium sized lab a central generator supply with UPSs for individual instruments can be a good combination to tackle power breakdowns. Refrigerators used for storage of reagents/kits should also be put on generator supply. The generator used should have the capacity well in excess of the total power load. It may be advisable to keep the high-power consuming equipment like air conditioners and water baths etc. off the generator supply.

Low-cost PCR Lab

Most of the equipment used in a PCR lab are imported from abroad and are expensive. Keeping in view the cost constraints the essential equipment for PCR, including DNA thermal Cycler, PCR Reader, Heating Block, PAGE System, DC Power Supply, and Micro-centrifuge, was developed at the Genetic Technology Instrumentation (GTI) division of the Genetics Resource Centre (Figure 11.2) (www.grcpk.com).

The GTI-16 DNA thermal cycler has a 16 well format with real time graphic LCD display. It is programmable with a capacity to store 100 PCR programmes. The instrument has been extensively validated for all types of PCR applications.

The GTI-PCR Reader-G is a unique fluorometer that is designed for endpoint analysis of FAM labelled PCR products. The fluorescence is measured through a data acquisition software installed on a computer. The instrument has been validated for qualitative and semi-quantitative endpoint analysis of PCR products amplified by the TaqMan® probe-based chemistry.

A DNA lab equipped with the GTI equipment was established at Fatimid Thalassaemia Centre Karachi in 2014 (Figure 11.3). The lab is providing regular service for prenatal diagnosis.

Figure 11.2. Low-cost PCR equipment developed at the Genetic Technology Instrumentation (GTI) division of the Genetics Resource Centre (GRC), Rawalpindi, Pakistan.

Figure 11.3. A lab for prenatal diagnosis of thalassaemia, equipped with the GTI equipment, was established at the Fatimid Thalassaemia Centre Karachi in 2014. This lab is fully operational till date.

12

PCR in Inherited Disorders

There are over 5000 known genetic disorders and vast majority of these can be diagnosed by PCR and its applications. Because of the simplicity of PCR even small labs can develop the expertise to diagnose several genetic disorders. Molecular genetic testing has been successfully used in the prenatal and postnatal diagnosis of genetic disorders. In addition, the PCR applications have been instrumental in the development of non-invasive prenatal testing, pre-implantation genetic diagnosis, pre-symptomatic testing and predictive testing. In this chapter some examples of common genetic disorders are discussed.

β-thalassaemia

β-thalassaemia is a typical single gene disorder whose molecular genetics is almost completely known. There are over 200 different variants that cause β-

thalassaemia. However, each ethnic population has its own set of common and uncommon variants. The first step in carrying out PCR based diagnosis of β-thalassaemia is to know the pattern of variants in the target population. This knowledge is essential for genetic diagnosis.

Screening for β-thalassaemia variants

Most of the β-thalassaemias are caused by point variants that are best detected by ARMS (Chapter 6). ARMS can be done in separate reactions for each variant or as multiplex PCR.

Standard ARMS (separate reactions)

1. In the first round of PCR common variants are tested.
2. The samples that do not show any of the common variants are tested in the next rounds for the uncommon and the rare variants.
3. The list and the sequences of primers for the β-thalassaemia variants in the major world populations is given in Table 12.1. The primers may be given a serial number because the numbers are easier to remember and document than the full nomenclature of the allele.
4. In each reaction the ARMS primer for a variant is run with another primer called “common primer”.
5. Most of the ARMS primers are designed as complementary to the forward strand of DNA. All of these are used with the same “common primer” which is complementary to the reverse strand (primer 3).
6. ARMS primers for some variants are complementary to the reverse strand and these are used with another “common primer” complementary to the forward strand (primer 4).
7. In each reaction two additional primers (1 & 2) are used to amplify 861bp fragment of the distal end of β-globin gene. This works as an internal control in the ARMS PCR. The location of the primers 1 & 2 is such that it also allows identification of 619bp deletion in the β-globin gene. If the del 619bp variant is present size of the control fragment, instead of the usual 861bp, is reduced to 242bp.

Multiplex ARMS

1. A multiplex ARMS PCR for 12 β-thalassaemia variants found in the Pakistani population is described. Similar multiplexes for the other ethnic populations can be developed.
2. Three combinations of primers for multiplex ARMS of β-thalassaemia are shown in Table 12.2. The multiplexes AD-1 and AD-2 contain the ARMS primers used with "common primer 3" while the AD-3 multiplex contains the ARMS primers that are used with "common primer 4".
3. The amplified products of all the variants tested in each multiplex are sufficiently different for resolution by mini polyacrylamide gel electrophoresis. However, the size of the fragments generated by IVSI-1 and IVSI-5 differ by only 5bp and are difficult to resolve on a mini gel. Similarly there is no difference between the fragment sizes of Cd30 (G-C), Cd30 (G-A) and IVSI-1 variants. The problem of differentiating IVSI-1 and IVSI-5 is overcome by adding IVSI-1 primer to AD-1 and AD-2 multiplexes. IVSI-5 results in amplification with AD-1, but IVSI-1 causes amplification with AD-1 and AD-2. Amplification with AD-2 but not AD-1 indicates Cd30. The difference between Cd30 (G-C) and Cd30 (G-A) is only of academic interest because the same normal primer is used to differentiate between the homozygotes and heterozygotes of the two variants. The AD-3 primer combination includes Cd15 and Cap+1 and both of these primers are used with "common primer 4". An allelic ladder for the respective variants is prepared by pooling the PCR products of separately amplified reactions of various variants. The allelic ladder is kept frozen in aliquots. 5μl of the pooled product is used in all polyacrylamide gel electrophoresis runs.
4. The amplified products of the multiplex ARMS and the respective allelic ladders are run on 6% polyacrylamide mini gels.
5. Interpretation of the multiplex ARMS results is shown in Figure 12.1.

Table 12.1. List and sequence of primers for the Asian and Far-Eastern β-thalassaemia variants.

Control and common primers:			
1. Control-F	5'-CAATGTATCATGCCTCTTTGCACC		
2. Control-R	5'-GAGTCAAGGCTGAGAGATGCAGGA		
3. Common-1	5′-ACCTCACCCTGTGGAGCCA		
4. Common-2	5′-CCCCTTCCTATGACATGAACTTAA		
Asian variants			
Allele	**Primer sequence**	**Used with**	**Size**
Fr 8-9 (+G) M	5′-CCTTGCCCCACAGGGCAGTAACGGCACACC	3	215
Fr 8-9 N	5′-CCTTGCCCCACAGGGCAGTAACGGCACACT	3	215
IVSI-5 (G-C) M	5′-CTCCTTAAACCTGTCTTGTAACCTTGTTAG	3	285
IVSI-5 N	5′-CTCCTTAAACCTGTCTTGTAACCTTGTTAC	3	285
Fr 41-42 (-TTCT) M	5′-GAGTGGACAGATCCCCAAAGGACTCAACCT	3	439
Fr 41-42 N	5′-GAGTGGACAGATCCCCAAAGGACTCAAAGA	3	443
Del 619bp-F	5'-CAATGTATCATGCCTCTTTGCACC	2	242
Del 619bp-R	5'-GAGTCAAGGCTGAGAGATGCAGGA	1	242
Cd 15 (G-A) M	5′-TGAGGAGAAGTCTGCCGTTACTGCCCAGTA	4	500
Cd 15 N	5′-TGAGGAGAAGTCTGCCGTTACTGCCCAGTG	4	500
Cd 5 (-CT) M	5′-ACAGGGCAGTAACGGCAGACTTCTCCGCGA	3	205
Cd 5 N	5′-ACAGGGCAGTAACGGCAGACTTCTCCGCAG	3	205
IVSI-1 (G-T) M	5′-TTAAACCTGTCTTGTAACCTTGATACGAAA	3	281
IVSI-1 N	5′-GATGAAGTTGGTGGTGAGGCCCTGGGTAGG	4	450
Cd 30 (G-C) M	5′-TAAACCTGTCTTGTAACCTTGATACCTACG	3	280
Cd 30 (G-A) M	5′-TAAACCTGTCTTGTAACCTTGATACCTACT	3	280
Cd30 N	5′-TAAACCTGTCTTGTAACCTTGATACCTACC	3	280
Fr 16 (-C) M	5′-TCACCACCAACTTCATCCACGTTCACGTTC	3	238
Fr 16 N	5′-TCACCACCAACTTCATCCACGTTCACGTTG	3	239
IVSII-1 (G-A) M	5′-AAGAAAACATCAAGGGTCCCATAGACTGAT	3	634
IVSII-1 N	5′-AAGAAAACATCAAGGGTCCCATAGACTGAC	3	634
Cap+1 (A-C) M	5′-ATAAGTCAGGGCAGAGCCATCTATTGGTTC	4	567
Cd 48 (+ATCT) M	5′-ATAACAGCATCAGGAGTGGACAGATAGATC	3	467
IVSI-25 M	5′-CTCTGGGTCCAAGGGTAGACCACCAGCATA	3	354
-88 (C-T) M	5′-TCACCTAGACCTCACCCTGTGGAGCCTCAT	4	655
Far Eastern variants			
Fr 41-42 (-TTCT) M	5′-GAGTGGACAGATCCCCAAAGGACTCAACCT	3	439
Fr 41-42 N	5′-GAGTGGACAGATCCCCAAAGGACTCAAAGA	3	439
C17 M (A-T)	5′-CTCACCACCAACTTCATCCACGTTCACATA	3	211
IVSII-654 M (C-T)	5′-GAATAACAGTGATAATTTCTGGGTTAACGT	2	830
Hb-E Cd 26 (G-A)(M)	5′-TAACCTTGATACCAACCTGCCCAGGGCGTT	3	267

Table 12.1. (Continued) List and sequence of primers for the Mediterranean and the African β-thalassaemia variants.

Mediterranean variants			
Allele	**Primer sequence**	**Used with**	**Size**
IVSI-110 M (G-A)	5′-ACCAGCAGCCTAAGGGTGGGAAAATAGAGT	3	390
IVSI-110 N	5′-ACCAGCAGCCTAAGGGTGGGAAAATACACC	3	390
IVSI-1 (G-A) M	5′-TTAAACCTGTCTTGTAACCTTGATACGAAT	3	281
IVSI-1 N	5′-TTAAACCTGTCTTGTAACCTTGATACGAAC	3	281
C 39 M (C-T)	5′-CAGATCCCCAAAGGACTCAAAGAACCTGTA	3	436
C 39 N	5′-TTAGGCTGCTGGTGGTCTACCCTTGGTCCC	4	436
IVSI-6 M (T-C)	5′-TCTCCTTAAACCTGTCTTGTAACCTTCATG	3	286
IVSI-6 N	5′-TCTCCTTAAACCTGTCTTGTAACCTTCATA	3	286
IVSII-1M (G-C)	5′-AAGAAAACATCAAGGGTCCCATAGACGCAG	3	634
IVSII-1 N	5′-AAGAAAACATCAAGGGTCCCATAGACTGAC	3	634
IVSII-745 M (C-G)	5′-TCATATTGCTAATAGCAGCTACAATCGAGG	2	738
IVSII-745 N	5′-TCATATTGCTAATAGCAGCTACAATCGAGC	2	738
African variants			
-28 M (A-G)	5′-AGGGAGGGCAGGAGCCAGGGCTGGGCTTAG	4	624
-88 M (C-T)	5′-TCACTTAGACCTCACCCTGTGGAGCCTCAT	4	655
C 24 M (T-A)	5′-TGGGGCAAGGTGAACGTGGATGAAGTTAGA	4	472
Hb-S-M	5′-CCCACAGGGCAG TAACGGCAGACTTCTGCA	3	208
Hb-S-N	5′-CCCACAGGGCAGTAACGGCAGACTTCTGCT	3	208
Hb-C-M	5′-CCACAGGGCAGTAACGGCAGACTTCTCGTT	3	207
Hb-C-N	5′-CCACAGGGCAGTAACGGCAGACTTCTCGTC	3	207

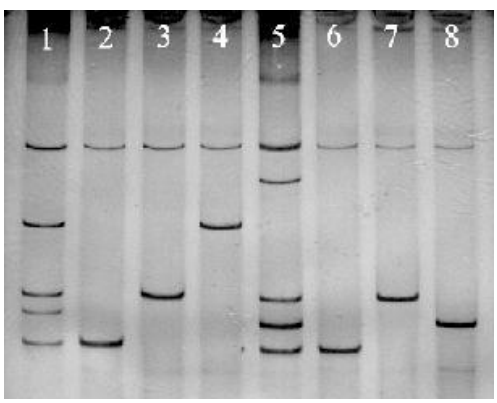


Figure 12.1. Silver stained mini polyacrylamide gel electrophoresis of multiplex ARMS PCR products. Lane 1 and 5 show allelic ladders for the AD-1 and AD-2 multiplexes. All lanes show the 861bp internal control fragment. The sample in lane 2, 3 and 4 are positive for Fr 8-9 (+G), IVSI-5 (G-C) and Fr 41-42 (-TTCT) variants respectively. Lanes 6, 7 and 8 show Cd5 (-CT), IVSI-1 (G-T) and Fr 16 (-C) variants respectively.

Table 12.2. Multiplex ARMS primer combinations for β-thalassaemia variants found in the Pakistani population.

Primer ID:	Variant	Product size
AD-1	Fr 8-9 (+G)	215 bp
	IVSI-5 (G-C)	285 bp
	Fr 41-42 (-TTCT)	439 bp
	IVSI-1 (G-T)	280 bp
	Del 619bp	242 bp
AD-2	Cd 5 (-CT)	205 bp
	Fr 16 (-C)	238 bp
	IVSI-1 (G-T)	280 bp
	Cd 30 (G-C)	280 bp
	Cd 30 (G-A)	280 bp
	IVSII-1 (G-A)	634 bp
AD-3	Cd 15 (G-A)	500 bp
	Cap+1 (A-C)	567 bp

Homozygous or heterozygous variant

1. Once the variant is identified the next step is to find whether the variant is homozygous or heterozygous. This is done by setting up a separate PCR reaction in which the normal allele of the respective variant is tested. The list and the sequences of the normal ARMS primers are shown in Table 12.1.
2. In a homozygote either two copies of the same variant (true homozygote) or two different variants (compound heterozygote) are seen.
3. The true homozygote shows only one variant. Its homozygosity is ascertained by testing for the normal allele checked in a separate PCR reaction. The compound heterozygote shows two different variants.
4. A heterozygote has the variant as well as the normal allele. There is no need to test for the normal allele if the individual is known to have thalassaemia trait.

Sequencing of β-globin gene

The β-globin gene sequencing is done to identify an unknown variant or to confirm the results of ARMS PCR. The gene is approximately 1.5 kb in length. On the 36 cm capillary of genetic analyzer approximately 600 bases can be sequenced in one go. The entire β-globin gene can be sequenced in three overlapping segments.

Sequencing strategy

Most of the β-thalasaemia variants are located in the first two exons and the intervening sequence (segment-I). The segment-II mostly covers the IVS-II which is an unusual site for the variants. The segment-III covers the third exon and the terminal portion of the gene. It also contains only a few uncommon variants.

- Segment-I
  - Target region: Promoter region & Exon-I
  - Forward primer: 5'-AGACATAATTTATTAGCATGCATG
  - Reverse primer: 5'-CTCCTTAAACCTGTCTTGTAACCT
  - Fragment size: 809 bp
  - Sequencing primer: 5'-AGACATAATTTATTAGCATGCATG
- Segment-II
  - Target region: Exon-I, IVS-I & Exon-II
  - Forward primer: 5'-ACCTCACCCTGTGGAGCCAC
  - Reverse primer: 5'-CCCCTTCCTATGACATGAACTTAA
  - Fragment size: 676 bp
  - Sequencing primer: 5'-ACCTCACCCTGTGGAGCCAC
- Segment-III
  - Target region: Exon-III & Poly-A site
  - Forward primer: 5'-CAATGTATCATGCCTCTTTGCACC
  - Reverse primer: 5'-GAGTCAAGGCTGAGAGATGCAGGA
  - Amplified fragment: 861 bp
  - Sequencing primer: 5'-CAATGTATCATGCCTCTTTGCACC

The details of the sequencing protocol are given in Chapter 9.

Results

The results of normal sequence, IVSI-5 (G-C) and Fr 8-9 (+G) variant are shown in Figure 12.2 and 12.3.

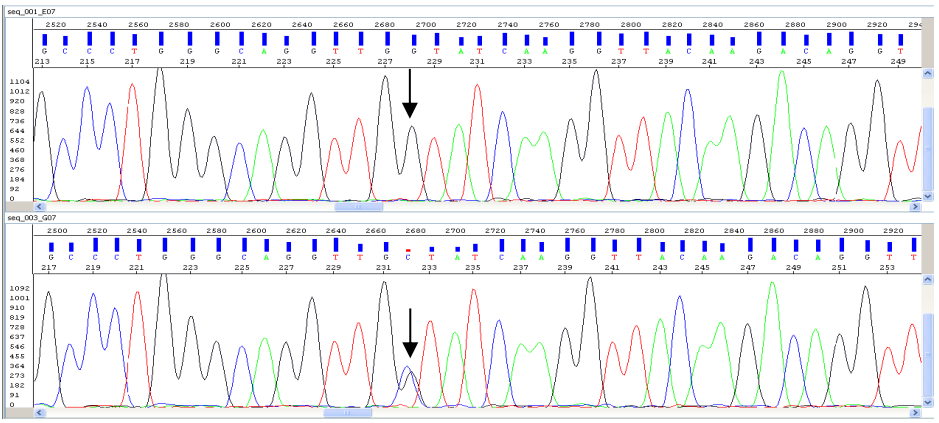


Figure 12.2. Electropherogram of β-globin gene showing DNA sequence around the first exon-intron junction. The upper half of the picture shows the normal sequence whereas the lower part shows IVSI-5 (G-C) variant (arrow) seen as two overlapping peaks (black and blue).

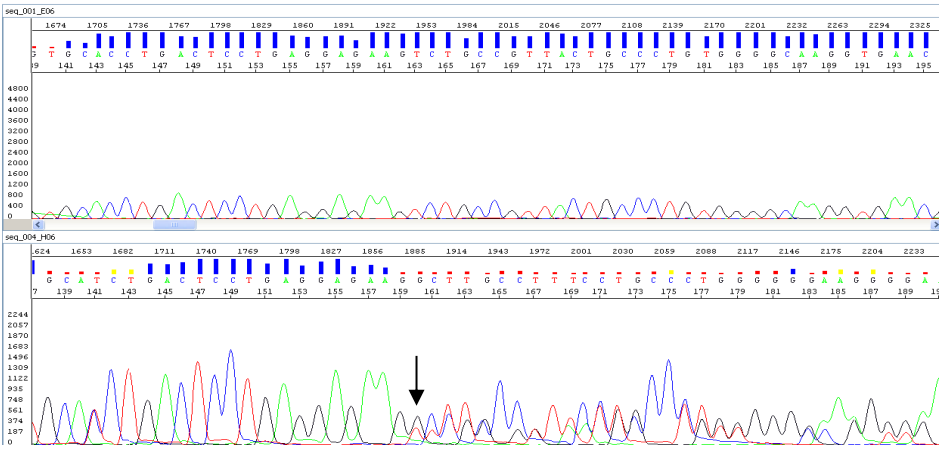


Figure 12.3. Upper half of the picture shows normal sequence of β-globin gene whereas Fr 8-9 (+G) variant is seen to cause complete disruption of the sequence beyond codon eight (arrow) in the lower part of the picture.

Prenatal Diagnosis (PND)

PND by direct variant analysis

1. Identify the parent's variants by testing the DNA of each parent or an affected child if the DNA sample of a parent is not available.
2. Carefully dissect the chorionic villus sample (CVS) to remove any maternal decidua.
3. Extract DNA from the CVS (fetus) (Chapter 2).
4. Setup ARMS PCR as follows (Chapter 6):
    1. If both parents have the same variant label ten tubes and add:
        i. Tubes 1-5: primers 1+2 + common primer + variant primer.
        ii. Tubes 6-10: primers 1+2 + common primer + normal primer for the variant.
        iii. In tubes 1-5 add DNA of father, mother, CVS (in duplicate), and negative control for the variant.
        iv. In tubes 6-10 add DNA of CVS (in duplicate), positive control for normal allele (normal DNA), negative control for normal allele (homozygous for the variant) and no DNA (reagent blank).
        v. The results of gel electrophoresis and their interpretation is shown in Figure 12.4.
    2. If the two parents have different variants label nine tubes and add:
        i. Tubes 1-4: primers 1+2 + common primer + primer of father's variant
        ii. Tubes 5-9: primers 1+2 + common primer + primer of mother's variant
        iii. In tubes 1-4 add DNA of father, CVS (in duplicate) and negative control for father's variant.

iv. In tubes 5-9 add DNA of mother, CVS (in duplicate), negative control for mother's variant and no DNA (reagent blank).

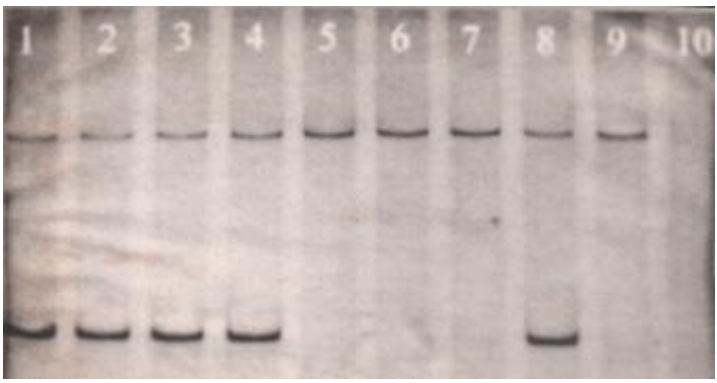


Figure 12.4. Polyacrylamide gel electrophoresis of prenatal diagnosis for β-thalassaemia. All lanes except lane 10 show 861bp internal control bands. Absence of any result in the reagent blank (Lane 10) excludes false positive results. Lanes 1 & 2 shows the parent's variants (IVSI-5). Lanes 3 & 4 show the same variant (IVSI-5) tested in duplicate in the fetal DNA. Lane 5 is a negative control for IVSI-5 variant. Lanes 6 & 7 shows normal allele of IVSI-5. Its absence in the fetal DNA confirms that the fetus has homozygous IVSI-5. Lanes 7 and 8 are negative and positive controls for the normal allele of IVSI-5 respectively. Lane 8 is a negative control for IVSI-5.

Exclusion of maternal contamination

1. Presence of maternal tissue in the CVS can be a potential source of error in prenatal diagnosis. Since the mother is a carrier who has the mutant as well as the normal alleles, contamination of CVS by maternal tissue would make all normal or homozygous fetal results to appear as thalassaemia trait.
2. When the parents have the same variant and the fetal diagnosis is "normal" or "homozygous (thalassaemia major)" maternal contamination is ruled out.

3. When the parents have two different variants and the mother's variant is not present in the CVS, maternal contamination is ruled out.
4. When the parents have the same variant and the fetal diagnosis is "heterozygous (thalassaemia trait)" maternal contamination is not ruled out.
5. Ideally each CVS should be tested for the presence of maternal contamination by Short Tandem Repeat (STR) analysis. But this adds to the overall cost of prenatal diagnosis. Experience has shown that meticulous cleaning of the CVS is sufficient to safeguard against errors due to maternal contamination.
6. The testing for maternal contamination by STR analysis is done by PCR amplification at various loci (Chapter 8 & 15).
7. The STR marker is called "informative" when its alleles can distinguish between the maternal and the fetal DNA. For example, at D21S11 locus if the mother has alleles 28,29 and the fetus (CVS) has 28,31. The marker is informative because allele 29 and 31 are exclusive for the mother and the fetus respectively. It also indicates that there is no maternal contamination. It would have indicated maternal contamination had the exclusive maternal allele (29) been present in the fetus and it is expressed as 28,31(29).
8. STR locus D21S11 is commonly used because of its high degree of polymorphism. Its limitation is that it cannot differentiate between maternal contamination and trisomy 21 in the fetus.
9. If the SRT is not informative at any locus the analysis is done at other loci.

PND of thalassaemia by linkage analysis

If the parent's variant is not identified prenatal diagnosis can be done by linkage analysis. There are several single nucleotide variants (SNV) that are closely linked to the β-globin gene. The SNVs are inherited en-block with the β-globin gene and these can be used to track the inheritance of any variant in the gene.

The SNVs are usually recognized by restriction enzymes. In the first step DNA of the parents and the affected child are tested to establish the linkage of the SNV with the chromosome carrying the variant. Once the informative markers are

identified the fetal DNA is tested to determine its genotype. By convention the SNV is written "+" when it is present and "-" when it is absent.

Pre-requisites of linkage based prenatal diagnosis

1. The couple must have a previously affected child.
2. At least one fully informative or two partially informative markers are present. The marker is called fully informative when both of the parents are heterozygous (-/+). The marker is called non-informative if one or both of the parents are homozygous (+/+, or -/-) for the marker. Fetal diagnosis can be done by using more than one partially informative markers.
3. A cross over during meiosis may result in dissociation of the marker and the gene of interest resulting in an error in diagnosis.

SNV markers linked to the β-globin gene

At least eight SNVs are closely linked to the β-globin gene (Table 12.4). These markers can be used for the linkage based prenatal diagnosis of β-thalassaemia.

Amplification Protocol for RFLP analysis

PCR amplification is done in 15-20 µl reaction mixture containing 10 pM of each primer, 0.5 units of Taq polymerase, 30 µM of each dNTP, 10 mM Tris HCl (pH 8.3), 50 mM KCL, 1.5 mM $MgCl_2$, 100 mg/ml gelatin, and ~200 ng of genomic DNA. Thermal cycling consists of an initial denaturation at 95°C for 2 minutes followed by 30 cycles of the following programme:

- Denaturation: 94°C for 30 seconds
- Annealing:
    - 5'ψβ, 3'ψβ, Ava II, Hinf I: 56°C for 1 minute
    - Gγ, Aγ, RSA, Xmn-I: 60°C for 1 minute
- Extension: 72°C for 1 minute 30 second

Enzyme digestion

1. Take 1 µl of the restriction enzyme in 4 µl of reaction buffer in a 0.2 ml PCR tube.

2. Add 5 µl of the amplified product to the enzyme buffer mix and incubated at 37°C overnight.

Electrophoresis

The enzyme digested amplified products are separated on 2% agarose gel run for 45 minutes at 150 volts or 6% mini-polyacrylamide gel for 30 minutes at 150 volts. The agarose gels are stained in ethidium bromide while the polyacrylamide gels are stained in 0.1% silver nitrate.

The results of genotyping at an SNP recognized by restriction enzyme Hinc-II in a couple with an affected child are shown in Figure 12.5.

Table 12.4. SNV markers linked to β-globin gene, restriction enzymes, primers for amplification and size of the restriction fragments.

Locus	Enzyme	Primer	Fragment Size	(-)	(+)	Common
Gγ	Hind III	F-AGTGCTGCAAGAAGAACAACTACC R-CTCTGCATCATGGGCAGTGAGCTC	328	328	237 91	-
Aγ	Hind III	F-ATGCTGCTAATGCTTCATTAC R- TCATGTGTGATCTCTCAGCAG	635	635	327 308	-
5'ψβ	Hinc II	F-TCCTATCCATTACTGTTCCTTGAA R- ATTGTCTTATTCTAGAGACGATTT	794	690	690 104	-
3'ψβ	Hinc II	F-GTACTCATACTTTAAGTCCTAACT R- TAAGCAAGATTATTTCTGGTCTCT	914	914	435 479	-
β-	Hinf I	F-TGGATTCTGCCTAATAAAA R- GGGCCTATGATAGGGTAAT	742	341	128 213	244 154
β-	Ava II	F-ACTCCCAGGAGCAGGGAGGGCAGG R- TTCGTCTGTTTCCCATTCTAAACT	794	794	438 356	-
β-	RSA	F-AGACATAATTTATTAGCATGCATG R- ACATCAAGGGTCCCATAGAC	1152	411	330 81	646 95
Gγ	Xmn-I	F-GAACTTAAGAGATAATGGCCTAA R-ATGACCCATGGCGTCTGGACTAG	641	641	418 223	-

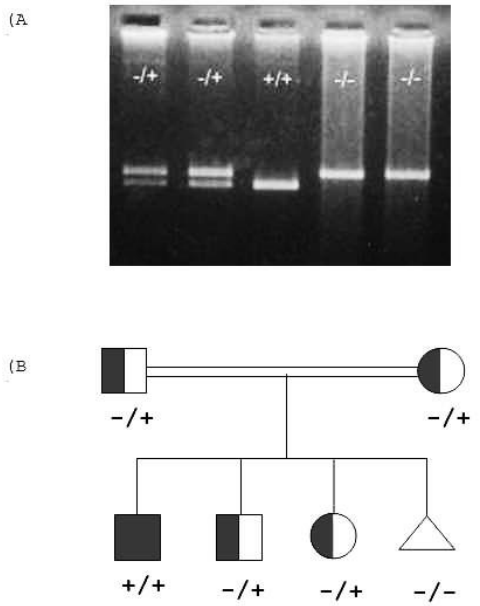


Figure 12.5. (A) Ethidium bromide stained agarose gel electrophoresis of Hinc-II digested fragments of β-globin gene. The father and the mother have -/+ genotype. The affected child of the couple has +/+ genotype indicating that the chromosome carrying the "+" sites in the father and the mother carry the β-thalassaemia variant. The fetal DNA has -/- genotype indicating that the fetus is unaffected by β-thalassaemia.

Limitations of linkage based PND

1. The couple must have a living affected child.
2. At least one informative marker should be present.
3. There is approximately 1% chance that during meiotic cross-over the marker and the abnormal gene could dissociate causing error in the diagnosis.
4. The presence of variant out-side the β-globin gene can cause error in diagnosis.

Sources of error in PND

1. Maternal contamination in fetal sample
2. PCR failure
3. Clerical mistakes

4. Meiotic crossover in linkage analysis
5. Non paternity

Duchene Muscular Dystrophy (DMD)

DMD is mostly caused by large deletions and inversions in the dystrophin gene. These variants can be detected by Multiplex Ligation-dependent Probe Amplification (MLPA) based methods (Chapter 6). The inversions/deletions causing DMD can also be detected by amplification of di-nucleotide short tandem repeats (STR) closely linked to the Introns 44, 45, 49 and 50 of the dystrophin gene on the X-chromosome. Failure to amplify any of the STR alleles indicates the presence of inversion/deletion variant. The STR based method has an added advantage of acting as a linkage marker for the inheritance of the mutant DMD allele if an affected child is also available (Figure 12.6).

Example

Primers for amplification of di-nucleotide STRs in the four introns of dystrophin gene.

DMD-44 F	5'-TCCAACATTGGAAATCACATTTCAA
DMD-44 R	5'-TCATCACAAATAGATGTTTCACAG
DMD-45 F	5'-GAGGCTATAATTCTTTAACTTTGGC
DMD-45 R	5'-CTCTTTCCCTCTTTATTCATGTTAC
DMD-49 F	5'-CGTTTACCAGCTCAAAATCTCAAC
DMD-49 R	5'-CATATGATACGATTCGTGTTTTGC
DMD-50 F	5'-AAGGTTCCTCCAGTAACAGATTTGG
DMD-50 R	5'-TATGCTACATAGTATGTCCTCAGAC

Each pair of primer for the respective intron is used for amplification in a separate tube as follows:

- Reaction volume: 25 µl
- Primer concentration: 1.0 µl (5 pmol)
- Taq polymerase: 0.1 µl (0.5 units)
- Template DNA: 1.0 µl (200ng)

- Thermal cycling:
    - Initial denaturation: 1 minute at 94°C
    - 30 cycles of:

      Denaturation: 30 seconds at 94°C

      Annealing: 30 seconds at 60°C

      Extension: 2 minutes at 72°C
    - Electrophoresis: 6% polyacrylamide gel and run at 100 Volts for 4-5 hours
- Staining: 0.1% Silver nitrate.
- Result: Figure 12.6

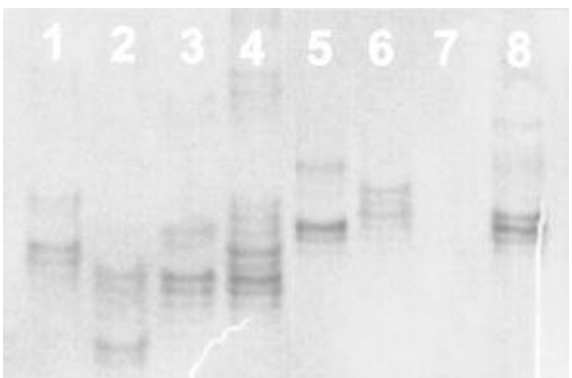


Figure 12.6. PAGE after PCR amplification of di-nucleotide repeats in Intron 44 of dystrophin gene (Lane 1-4) in a family with an affected child of DMD. The father (lane 1) has one allele due to one X chromosome (a faint band of PCR stutter product can also be seen). The mother (lane 2) shows two alleles due to two X chromosomes. The affected child (lane 3) being a male has one allele that is inherited from the mother. This indicates that out of the mother's two alleles the larger allele corresponding to that of the affected child can be used as a marker for the abnormal X chromosome. The fetal sample (lane 4) shows two alleles one inherited from the father and the other from the mother. The result is consistent with female fetus carrier for DMD. Lanes 5-8 show dinucleotide repeat in intron 45 in another family with DMD. Lane 5 shows father's sample, lane 6 mother's sample, lanes 7 child affected by DMD, and lane 8 shows fetal DNA. Sample of the affected child in lane 7 does not show any amplification indicating a deletion in the dystrophin gene involving intron 45. The fetal DNA (8) does not show the deletion in intron 45.

PCR based detection of Trisomies

Trisomies of chromosome 13, 18 and 21 constitute over 90% of the clinically significant abnormalities. The three trisomies can be screened by PCR amplification of polymorphic STR loci on the respective chromosomes. Each chromosome carries one allele of a STR and a normal individual inherits two alleles (one each on a chromosome). In case of a trisomy three, instead of the usual two, alleles are seen. A technical problem often arises if two or all three of the alleles on the three chromosomes have the same number of repeat units that appear as a single band on electrophoresis. The problem is solved either by testing additional loci or by densitometric reading after electrophoresis or by automated STR analysis on genetic analyzer (Chapter 8).

The following allele patterns may be seen at an STR locus (example D21S11):

1. Normal
    a. Disomic di-allelic (28,29)
    b. Disomic mono-allelic (28,28)
2. Abnormal
    a. Trisomic tri-allelic (28,29,30)
    b. Trisomic di-allelic (28,29,29)
    c. Trisomic monoallelic (28,28,28)

The STR allele sizing is done by PCR amplification of the locus followed by gel electrophoresis. In manual allele sizing gel electrophoresis is done with an allelic ladder. The results may be interpreted by visual inspection of the gel or by densitometry. The trisomy screening can be done without allele sizing if the amplified DNA samples of the father, mother and the subject are run simultaneously on a polyacrylamide gel (Figure 12.7) The DNA amplified with a fluorescent labeled primer when run on genetic analyzer also provides information on the allele sizes (Figure 12.8).

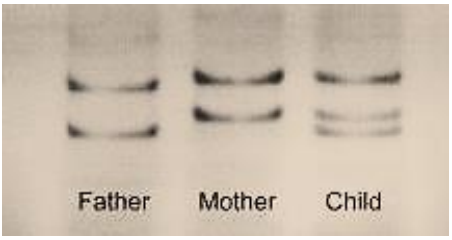


Figure 12.7. Polyacrylamide gel electrophoresis of PCR amplification at D21S11 locus in a couple suspected to have a child with trisomy 21. The father and the mother have disomic diallelic pattern whereas the child shows trisomic triallelic pattern confirming trisomy 21. In addition it also shows that the child inherited the extra chromosome 21 from the mother.

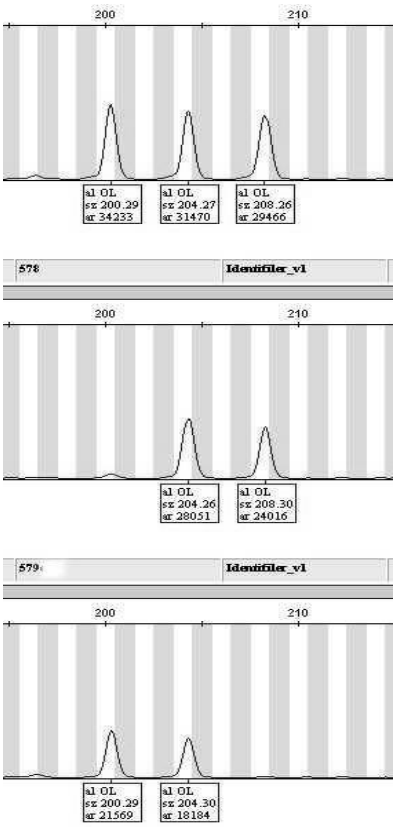


Figure 12.8. PCR amplification at D21S11 locus ran on a genetic analyzer. The father and the mother have a disomic diallelic pattern whereas the affected child (top) shows trisomic triallelic pattern confirming trisomy 21.

Primers for screening of trisomies:

Trisomy 13:

D13S317-F	5'-ACAGAAGTCTGGGATGTGGA
D13S317-R	5'-GCCCAAAAAGACAGACAGAA
D13S634-F	5'-TCCAGATAGGCAGATGATTCAAT
D13S634-R	5'-CCTTCTTCTTCCCATTGATA

Trisomy 18

D18S51-F	5'-CAAACCCGACTACCAGCAAC
D18S51-R	5'-GAGCCATGTTCATGCCACTG
D18S535-F	5’-TCATGTGACAAAAGCCACAC
D18S535-R	5’-AGACAGAAATATAGATGAGAATGCA

Trisomy 21

D21S11-F	5’-GTGAGTCAATTCCCCAAG
D21S11-R	5’-GTTGTATTAGTCAATGTTCTCC
D21S1411-F	5’-ATGATGAATGCATAGATGGATG
D21S1411-R	5’-AATGTGTGTCCTTCCAGGC
D21S1412-F	5’-CGGAGGTTGCAGTGAGTT
D21S1412-R	5’-GGGAAGGCTATGGAGGAGA
D21S1414-F	5’-AAATTAGTGTCTGGCACCCAGTA
D21S1414-R	5’-CAATTCCCCAAGTGAATTGCCTTC

Limitations of the PCR based screening of trisomies

Detection of trisomies by PCR is only a screening method and it is not a replacement of cytogenetic analysis. Its major limitation is that in PCR the smaller alleles may be preferentially amplified sometimes giving the false impression of trisomy. Maternal contamination in the CVS DNA may also give a false positive of trisomy. But this will give a trisomic pattern with all STR markers.

Testing for rare genetic disorders

There are more than 5000 known genetic disorders and most of these can be diagnosed molecular genetic analysis. PCR, microarrays and more recently NGS has opened up new vistas in the field of genetic (inherited) disorders.

Extensive NGS gene panels for newborn screening and screening in the latter life are available. The best strategy for a developing country with no NGS facility would be to get the initial genetic testing in the affected child from a commercial vendor. Once the disease causing genetic defect (variant) is identified a PCR based test or Sanger sequencing protocol for detection of the variant can be developed. The PCR based test or Sanger sequencing can be used for prenatal diagnosis and carrier screening in the extended family.

Non Invasive Prenatal Testing (NIPT)

NIPT is a relatively recent entry in the field of prenatal genetic testing. The procedure is based on the fact that minute quantities of cell-free DNA (cfDNA) of fetal origin is present in a mother in early pregnancy. NIPT was initially used for the determination of Rh-D blood-group status, fetal sex and autosomal dominant disorders of paternal inheritance. The use of NIPT to screen for the fetal aneuploidies became feasible with the development of microarrays and next generation sequencing (NGS) technologies. Currently most of the NIPT work is being done by NGS.

NIPT for screening of common autosomal aneuploidies

In a recent pooled meta-analysis of a large number of studies NIPT was found to have a sensitivity of 99% for trisomy 21, and a specificity of 99.92%. For trisomy 18, the reported figures were 96.8% (sensitivity) and 99.85% (specificity). For trisomy 13, they were 92.1 and 99.80% respectively.

A major reason why NIPT is less than 100% accurate for the common autosomal aneuploidies is that the fetal cfDNA actually comes from the placenta. A positive result indicating an aneuploidy may be generated by factors other than an aneuploid fetal karyotype, including placental mosaicism, a vanishing twin or a maternal tumor.

In a pregnant woman cfDNA is mostly of maternal origin and only a small proportion (~10%) derives from the fetus (placenta). Most labs working on NIPT require the 'fetal fraction' of cfDNA to be above 4%. Although cfDNA of fetal origin can be found in maternal blood as early as 4 weeks of pregnancy but its adequate amount is not reached before nine or ten weeks of gestation. An earlier testing may therefore lead to an inaccurate or failed result. A low fetal cfDNA may also be seen due to dilution factor resulting from higher maternal body weight. The NIPT failure rates vary between laboratories, ranging from 0 to 5%. NIPT may be repeated after a failed result or an invasive prenatal testing may be considered. The NIPT data on multiple gestations is also limited.

NIPT can also be done for sex chromosome abnormalities like Turner syndrome (45, X) and Klinefelter (XXY) etc. But the available data indicate a lower accuracy for the sex chromosome aneuploidies than for trisomies 21 and 18.

The currently available data indicate that NIPT is only a screening test for the detection of chromosomal aneuploidies. All positive NIPT results should be confirmed by invasive testing like chorionic villus sampling or amniocentesis.

NIPT for chromosomal micro-deletion syndromes

The available NGS technologies are sensitive enough to provide sequence data on selected micro-deletion syndromes like DiGeorge, Prader Willi/Angelman, Cri-du-chat, Wolf–Hirschhorn etc. But there is not enough evidence based data on the accuracy of NIPT for such disorders.

NIPT and single gene disorders

The ever increasing resolving power of NGS is likely to provide accurate diagnostic information on the single gene disorders in the foreseeable future. But keeping in view the high cost NIPT, especially in the developing countries, is unlikely to replace the invasive prenatal diagnosis.

Pre-implantation Genetic Diagnosis (PGD)

PGD is done to know the genetic status of an embryo prior to its implantation. This involves in-vitro fertilization, embryo biopsy, genetic analysis and selective implantation. It is a complex and expensive procedure that involves an

obstetrician, embryologist, and geneticist. It is mostly done to screen for chromosomal aneuploidies and gender selection. It may also be done for single gene disorders in couples who do not accept termination of pregnancy.

An in-vitro fertilized embryo can be biopsied at the eight cell stage to obtain a blastomere without harming the embryo. Alternatively, several cells can be obtained from the older embryo by tropho-ectoderm biopsy. Genetic analysis on the single or a few cells may be done by fluorescent in-situ hybridization (FISH), PCR or NGS. There are not enough data on the long-term complications of embryo biopsy especially at the eight cell stage. Another limitation is the fairly high degree of mosaicism.

Bibliography

1. Adinolfi M, Pertl B, Sherlock J (1997) Rapid detection of aneuploidies by microsatellites and the quantitative fluorescent polymerase chain reaction. Prenat Diagn 17: 1299-1311.
2. Ahmed S, Petrou M, Saleem M (1996) Molecular genetics of β-thalassaemia in Pakistan: a basis for prenatal diagnosis. Br J Haematol 94: 476-481.
3. Ahmed S, Saleem M, Petrou M, Sultana N, Raashid Y, Waqar A, Anwar M, Karamat KA (2000) Prenatal diagnosis of β-thalassaemia in Pakistan: experience in a Muslim country. Prenat Diagn 20: 378-383.
4. Clemens PR, Fenwick RG, Chamberlain JS, Gibbs RA, Andrade M, Chakraborty R (1991) Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. Am J Hum Genet 49: 951-960.
5. Dondorp W, De Wert G, Bombard Y, Bianchi DW, Bergmann C, Borry P, Chitty LS, Fellmann F, Forzano F, Hall A, Henneman L (2015) Non-invasive prenatal testing for aneuploidy and beyond: challenges of responsible innovation in prenatal screening. Eu J Hum Genet 23:1438-1450.
6. Huisman TJH, Carver MFH (1998) The δ- and β-thalassaemia repository. Haemoglobin 22: 169-195.
7. Katsanis SH, Katsanis N (2014) Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 14: 415-426. doi:10.1038/nrg3493Old JM, Varawalla NY, Weatherall DJ (1990) Rapid detection and prenatal diagnosis of β-thalassaemia: studies in Indian and Cypriot populations in the UK. Lancet 336: 834-837.
8. Old JM (1996) Haemoglobinopathies. Prenat Diagn 16: 1181-1186.

13

PCR in Neoplastic Disorders

Traditional anatomic pathology aided by molecular techniques has evolved in to a new discipline called “molecular pathology”. This has led to a better understanding of the disease through gene expression studies, identification of the causative agents, more accurate diagnosis, prediction of prognosis and providing basis for individualized approach to treatment.

At the sub-cellular level cancer is almost certainly a genetic disorder. The molecular genetics of cancer is typically marked by variant(s) in the oncogenes, tumour suppressor genes or micro RNA genes. The genetic changes in a cancer carry the potential for diagnosis by PCR that can be used in several different situations:

1. Diagnosis in obscure cases
2. Pre-symptomatic detection
3. Detection of secondaries

4. Prediction of prognosis
5. Monitoring of disease progression and response to treatment
6. Demonstrate association of malignancies and viruses
7. Research

Some of the commonly used molecular genetic techniques in the investigation of a neoplastic disorder include quantitative PCR (Chapter 5), gene expression studies (Chapter 7), DNA sequencing and NGS (Chapter 9). In addition fluorescent in-situ hybridization (FISH) and DNA microarrays are also used.

This chapter describes a few typical examples of PCR in the diagnosis of common malignancies.

Ig gene rearrangement

Immunoglobulin and T-cell receptor genes in the germ-line configuration are made of several individual parts called variable (V), diversity (D) and joining (J) segments. On antigenic challenge the germ line segments rearrange to form a much smaller gene. In a clonal proliferation of B or T lymphocytes the entire population of the cells carry a specific Ig or T-cell receptor gene rearrangement.

The Ig or the T-cell receptor genes without rearrangements are too large for amplification by PCR. In case of a clonal gene rearrangement the genes become much smaller and suitably placed PCR primers can be used to amplify the gene. The assembly of IgH genes from the variable (VH), diversity (DH), and joining (JH) regions creates a DNA target that is complex and unique to each patient.

A simple broad spectrum PCR assay is described that can pick up to 90% of the IgH gene rearrangements in patients of lymphoproliferative disorders. A forward primer complementary to consensus JH sequence is used with a reverse primer in the VH region.

Forward primer (JH) (5pmol):	5’-AACTGCAGAGGAGACGGTGACC
Reverse primer (VH) (5pmol):	5’-CTGTCGACACGGCCGTGTATTACT
Reaction volume:	25µl
PCR buffer:	22µl
Taq polymerase:	0.5 units (0.1µl)

Primer mix: 1µl (5pmol of each primer/µl)

DNA: 2µl (~200ng)

Thermal cycling

Initial denaturation 94°C for 5 minutes.

Thirty cycles each of:

- Denaturation at 94°C for 1 minute
- Annealing at 60°C for 1 minute
- Extension at 72°C for 1 minute 30 seconds
- Final extension at 72°C for 3 minutes

Electrophoresis may be done on 2% agarose at 150 volts for 60 minutes or 6% mini-polyacrylamide gel at 150 volts for 30 minutes (Figure 13.1).

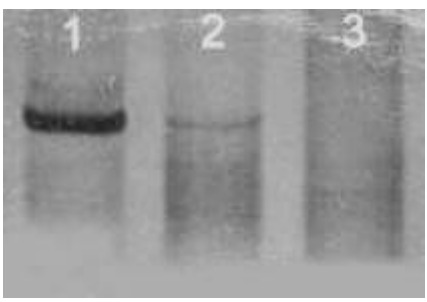


Figure 13.1. PCR for Immunoglobulin heavy chain gene rearrangement. Lane 1 shows a sharp band of clonal Ig gene rearrangement whereas lane 3 shows a diffuse pattern of polyclonal gene rearrangements. Lane 2 shows a faint clonal band of minimal residual disease along with a diffuse polyclonal background in the post treatment sample of the patient at lane 1.

Real time PCR for IgH gene rearrangement

Forward primer (JH) (5pmol): 5'-AACTGCAGAGGAGACGGTGACC

Reverse primer (VH) (5pmol): 5'-CTGTCGACACGGCCGTGTATTACT

SYBR green ready reaction mix: 23µl

DNA: 2µl (~200ng)

SYBR green ready reaction mix containing dNTPs and Taq polymerase is available from many commercial sources.

Thermal cycling

Initial denaturation 94°C for 5 minutes.

Thirty cycles each of:

Denaturation at 94°C for 1 minute

Annealing at 60°C for 1 minute (read fluorescence)

Extension at 72°C for 1 minute 30 seconds

Melting curve analysis may be added at the end of the real time PCR protocol to differentiate between specific and non-specific amplification products.

Bcl-II gene rearrangement

Follicular lymphoma is characterized by translocation of bcl-II gene from chromosome 18 to IgH gene on chromosome 14. The fusion gene is a lymphoma specific marker that can be used in diagnosis and monitoring of disease activity.

mbr	5'-GAGTTGTACGTGGCCTG
mcr	5'-CGCTTGACTCCTTTACGTGC
s-icr	5'-TCGTTCTCAGTAAGTGAGAGTGC
LJH	5'-TGAGGAGACGGTGACC

Real Time PCR

Primers:

Mbr-F	5'-TTAGAGAGTTGCTTTACGTGGCC
IgH-R	5'-ACTCACCTGAGGAGACGGTGAC
Mbr-Probe	6 FAM 5'-TTTCAACACAGACCCACCCAGAGCC-TAMRA

Primer mix: Make a mix of the four primers and two probes at concentration of 5pmol each (Chapter 3).

PCR protocol:

Reaction volume:	25µl
PCR mix:	22µl
Primer mix:	1µl

Taq polymerase: 0.5 units (0.1µl)

DNA: 2µl

Thermal cycling

Initial denaturation 95°C for 10 minutes

Forty cycles each of:

Denaturation at 95°C for 15 seconds

Annealing/extension at 60°C for 1 minute

Read fluorescence at 60°C step

Bcr-abl gene rearrangements

A reciprocal translocation between the long arms of chromosomes 9 and 22 (Philadelphia chromosome) is found in >90% patients of chronic myeloid leukemia and 15–25% patients with acute lymphoblastic leukemia. This translocation transposes c-abl oncogene from chromosome 9q34 to bcr gene on chromosome 22q11. The fused bcr/abl gene provides specific marker for diagnosis and disease monitoring. The translocation can be detected by first making cDNA of the bcr-abl mRNA followed by PCR amplification.

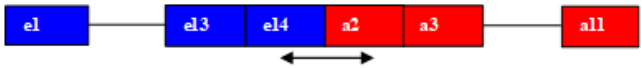
Common bcr-abl gene rearrangements

Over 95% of the bcr-abl transcripts in CML are between introns 13/14 on bcr gene and intron 2 on abl gene (e13a2 & e14a2). The transcripts result in protein called p210. In ALL 70% of the transcripts are e1a2 that result in protein called p190. About 25% of transcripts in ALL are e13a2 or e14a2 (Figure 13.1). In a minority of the patients atypical transcripts may also be observed.

PCR screening strategy

Sequence of the bcr-abl gene transcripts and the PCR primers with TaqMan® probe are shown in Figure 13.2. Three separate forward primers E-1, E-13 and E-14 are used with a common reverse primer A-2. A common bcr-abl TaqMan® probe complementary to the abl gene is used. A fragment of the abl gene just close to the breakpoint region is amplified as an internal control by a separate pair of primers.

A. Homo sapiens partial mRNA for bcr-abl e14a2 chimeric protein (GenBank: AJ131466.1).



ACTGTCCACAGCATTCCGCTGACCATCAATAAGGAAGATGATGAGTCTCCGGGGCTCTATGGGTTTCTG
AATGTCATCGTCCACTCAGCCACTGGATTTAAGCAGAGTTCAAAAGCCCTTCAGCGGCCAGTAGCATCT
GACTTTGAGCCTCAGGGTCTGAGTGAAGCCGCTCGTTGGAACTCCAAGGAAAACCTTCTCGCTGGACCC
AGTGAAAATGACCCCAACCTTTTCGTTGCACTGTATGATTTTGTGGCCAGTGGAGATAACACTCTAAGC
ATAACTAAAGGTGAAAAGCTCCGGGTCTTAGGCTATAATCACAATGGGGAATGGTGTGAAGCCCAAACC
AAAAATGGCCAAGGCTGGGTCCCAAGCAACTACATCACGCCAGTCAACAGTCTGGAGAAACACTCCTGG
TACCATGGGCCTGTGTCCCGCAATGCCGCTGAGTATCTGCTGAGCAGCGGGATCAATGGCAGCTTCTTG
GTGCGTGAGAGTGAGAGCAGTCCTGGCCAGAGGTCCATCTCGCTGAGATACGAAGGGAGGGTGTACCAT

B. Homo sapiens partial mRNA for bcr-abl e13a2 chimeric protein (GenBank: AJ131467.1).

AAGAAGTGTTTCAGAAGCTTCTCCCTGACATCCGTGGAGCTGCAGATGCTGACCAACTCGTGTGTGAAA
CTCCAGACTGTCCACAGCATTCCGCTGACCATCAATAAGGAAGAAGCCCTTCAGCGGCCAGTAGCATCT
GACTTTGAGCCTCAGGGTCTGAGTGAAGCCGCTCGTTGGAACTCCAAGGAAAACCTTCTCGCTGGACCC
AGTGAAAATGACCCCAACCTTTTCGTTGCACTGTATGATTTTGTGGCCAGTGGAGATAACACTCTAAGC
ATAACTAAAGGTGAAAAGCTCCGGGTCTTAGGCTATAATCACAATGGGGAATGGTGTGAAGCCCAAACC
AAAAATGGCCAAGGCTGGGTCCCAAGCAACTACATCACGCCAGTCAACAGTCTGGAGAAACACTCCTGG
TACCATGGGCCTGTGTCCCGCAATGCCGCTGAGTATCTGCTGAGCAGCGGGATCAATGGCAGCTTCTTG
GTGCGTGAGAGTGAGAGCAGTCCTGGCCAGAGGTCCATCTCGCTGAGATACGAAGGGAGGGTGTACCAT

C. Homo sapiens bcr-abl e1a2 chimeric protein mRNA, partial cds (GenBank: AF113911.1).

GACTTCTCCTCTGGCCAGTCCAGCCGCGTGTCCCCAAGCCCCACCACCTACCGCATGTTCCGGGACAAA
AGCCGCTCTCCCTCGCAGAACTCGCAACAGTCCTTCGACAGCAGCAGTCCCCCCACGCCGCAGTGCCAT
AAGCGGCACCGGCACTGCCCGGTTGTCGTGTCCGAGGCCACCATCGTGGGCGTCCGCAAGACCGGGCAG
ATCTGGCCCAACGATGGCGAGGGCGCCTTCCATGGAGACGCAGAAGCCCTTCAGCGGCCAGTAGCATCT
GACTTTGAGCCTCAGGGTCTGAGTGAAGCCGCTCGTTGGAACTCCAAGGAAAACCTTCTCGCTGGACCC
AGTGAAAATGACCCCAACCTTTTCGTTGCACTGTATGATTTTGTGGCCAGTGGAGATAACACTCTAAGC
ATAACTAAAGGTGAAAAGCTCCGGGTCTTAGGCTATAATCACAATGGGGAATGGTGTGAAGCCCAAACC
AAAAATGGCCAAGGCTGGGTCCCAAGCAACTACATCACGCCAGTCAACAGTCTGGAGAAACACTCCTGG
TACCATGGGCCTGTGTCCCGCAATGCCGCTGAGTATCTGCTGAGCAGCGGGATCAATGGCAGCTTCTTG
GTGCGTGAGAGTGAGAGCAGTCCTGGCCAGAGGTCCATCTCGCTGAGATACGAAGGGAGGGTGTACCAT

Figure 13.2. Homo sapiens partial mRNAs of bcr-abl chimeric proteins (A-C). The PCR primers and TaqMan probes for the rearranged bcr-abl gene and a control fragment of the abl gene are shown. The bcr gene sequences are shown in blue whereas the abl gene sequences are shown in red.

RNA extraction

For RNA extraction commercial TRIzol® Reagent® (Invitrogen, USA) as described in Chapter 2 gives good quality results. Several commercial kits based on silica columns or magnetic beads are also available for manual as well as automated RNA extraction.

Preparation of cDNA

Gene specific primer "ABL-R" may be used for preparing cDNA. Commercially available Oligo dT primer or random hexamers may also be used for this purpose. In this experiment RNA extracted from a patient known to have bcr-abl gene rearrangement was serially diluted ranging from 100% to 50%, 25%, 12.5%, 6.25%, 3.12%, and 1.56%. A known negative RNA was also included as a negative control. cDNA was prepared from each dilution and the negative control as per the following protocol:

Primer (ABL-R) 5'-GGCCACAAAATCATACAGTGCA

Reaction conditions (20 µl per reaction):

- 5 X reaction buffer: 4 µl
- 10 mmol dNTPs mix: 2 µl
- RT Primer (10 pmol/µl): 1 µl
- MMLV- Reverse Transcriptase (200 U/µl): 1 µl
- RNAse inhibitor (20 U/µl): 1 µl
- Deionized water: 9 µl
- RNA: 2 µl
- Incubation: 42^{o}C for 60 minutes
- RT inactivation: 70^{o}C for 5 minutes

Real Time PCR

Primers and Probes:

BCR-ABL

E-1 5'-GCAGATCTGGCCCAACGAT

E-13 5'-GCATTCCGCTGACCATCAATAA

E-14	5’-CAGCCACTGGATTTAAGCAGAGT
A-2 (R)	5’-TCCAACGAGCGGATTCACT
BCR-ABL-P	6 FAM 5’-AAGCCCTTCAGCGGCCAGTAGCATCT-BHQ1
Internal control	
ABL-F	5’-GCTGGGTCCCAAGCAACTAC
ABL-R	5’-ACACAGGCCCATGGTACCA
ABL-P	JOE 5’-TCACGCCAGTCAACAGTCTGGAGAAACA-BHQ1

Primer mix: Make a mix of the six primers and the two probes at concentration of 5pmol each (Chapter 3).

PCR protocol (per reaction):

Reaction volume:	25µl
PCR mix:	21µl
Primer mix:	1µl
Taq polymerase:	0.5 units (0.1µl)
cDNA:	3µl

Thermal cycling

Initial denaturation 95^{o}C for 10 minutes

Forty cycles each of:

Denaturation at 95^{o}C for 15 seconds

Annealing/extension at 60^{o}C for 1 minute

Read fluorescence at 60^{o}C step

Result

The real time PCR plot of the serially diluted sample and the negative control is shown in Figure 13.3.

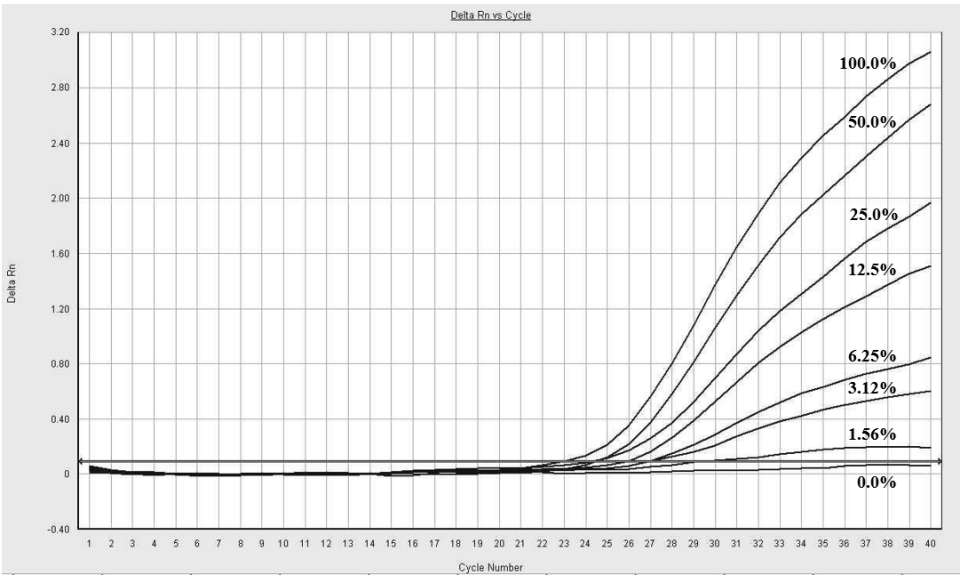


Figure 13.3. Real time plot of the serially diluted RNA sample (100-1.56%) positive for the bcr-abl gene rearrangement and a known negative sample.

A strategy for comprehensive molecular diagnosis of leukaemia

Molecular diagnosis of leukaemia may be helpful in diagnosis, prognostic stratification and detection of minimal residual disease. The WHO classification of acute leukaemia is based on demonstration of a large number of gene rearrangements that also correspond to gross chromosomal abnormalities. A comprehensive strategy for detection of various molecular genetic abnormalities in Acute Lymphoblastic Leukaemia (ALL) and Acute Myeloid Leukaemia (AML) is described.

RNA extraction

Commercial TRIzol® Reagent® (Invitrogen, USA) as described in Chapter 2. Several commercial kits based on silica columns or magnetic beads are also available for manual as well as automated RNA extraction.

Preparation of cDNA

Primer: Random hexamers at 0.2 µg/µL

Reaction conditions (20 µl per reaction):

- 5 X reaction buffer: 4 µl
- 10 mmol dNTPs mix: 2 µl
- RT Primer (10 pmol/µl): 1 µl
- MMLV- Reverse Transcriptase (200 U/µl): 1 µl
- RNAse inhibitor (20 U/µl): 1 µl
- Deionized water: 9 µl
- RNA: 2 µl
- Incubation: 42°C for 60 minutes
- RT inactivation: 70°C for 5 minutes

Real Time PCR

Primers and Probes

ALL

E2A-PBX1	t(1;19) (GenBank: M31222 & M86546)	
Forward:	ENF-E2A	5'-CCAGCCTCATGCACAACCA
Reverse:	ENR-PBX1	5'-GGGCTCCTCGGATACTCAAAA
Probe:	ENP E2A	FAM-5'-CCCTCCCTGACCTGTCTCGGCC-BHQ1
MLL-AF4	t(4;11) (GenBank: L04284 & L13773)	
Forward1:	ENF-MLL	5'-CCCAAGTATCCCTGTAAAACAAAAA
Forward2:	ENF-MLL	5'-GATGGAGTCCACAGGATCAGAGT
Reverse:	ENR-AF4	5'-GAAAGGAAACTTGGATGGCTCA
Probe:	ENP-AF4	FAM-5'-CATGGCCGCCTCCTTTGACAGC-BHQ1
TEL-AML1	t(12;21) (GenBank: U11732 & D43969)	
Forward:	ENF-TEL	5'-CTCTGTCTCCCCGCCTGAA
Reverse:	ENR-AML1	5'-CGGCTCGTGCTGGCAT
Probe:	ENP-TEL	FAM-5'-TCCCAATGGGCATGGCGTGC-BHQ1
m-bcr	t(9;22) p190 (GenBank: X02596 & X16416)	
Forward:	ENF-BCR	5'-CTGGCCCAACGATGGCGA
Reverse:	ENR-ABL	5'-CACTCAGACCCTGAGGCTCAA

Probe:	ENP-ABL	FAM-5'-CCCTTCAGCGGCCAGTAGCATCTGA-BHQ1

M-bcr t(9;22) p210 (GenBank: X02596 & X16416)

Forward:	ENF-BCR	5'-TCCGCTGACCATCAAYAAGGA
Reverse:	ENR-ABL	5'-CACTCAGACCCTGAGGCTCAA
Probe:	ENP-ABL	FAM-5'-CCCTTCAGCGGCCAGTAGCATCTGA-BHQ1

AML

SIL-TAL1 del(1) (GenBank: M74558 & S53245)

Forward:	ENF-SIL	5'-CGCTCCTACCCTGCAAACA
Reverse:	ENR-TAL	5'-CCGAGGAAGAGGATGCACA
Probe:	ENP-SIL	FAM-5'-ACCTCAGCTCCGCGGAAGTTGC-BHQ1

PML-RARA t(15;17) (GenBank: M73778 & X06538)

Forward1:	ENF-PML	5'-TCTTCCTGCCCAACAGCAA
Forward2:	ENF-PML	5'-ACCTGGATGGACCGCCTAG
Forward3:	ENF-PML	5'-CCGATGGCTTCGACGAGTT
Reverse:	ENR-RARA	5'- GCTTGTAGATGCGGGGTAGAG
Probe:	ENP-RARA	FAM-5'-AGTGCCCAGCCCTCCCTCGC-BHQ1

CBFB-MYH1 inv(16) (GenBank: L20298 & D10667)

Forward:	ENF-CBFB	5'-CATTAGCACAACAGGCCTTTGA
Reverse1:	ENR-MYH11	5'-AGGGCCCGCTTGGACTT
Reverse2:	ENR-MYH11	5'-CCTCGTTAAGCATCCCTGTGA
Reverse3:	ENR-MYH11	5'-CTCTTTCTCCAGCGTCTGCTTAT
Probe:	ENP-CBFB	FAM-5'-TCGCGTGTCCTTCTCCGAGCCT-BHQ1

AML1-ETO t(8;21) (GenBank: D43969 & D14289)

Forward:	ENF-AML1	5'-CACCTACCACAGAGCCATCAAA
Reverse:	ENR-ETO	5'-ATCCACAGGTGAGTCTGGCATT
Probe:	ENP-AML1	FAM-5'-AACCTCGAAATCGTACTGAGAAGCACTCCA-BHQ1

Primer and Probe mixes:

Make the respective primer and probe mixes of each molecular defect in a separate tube at concentration of 5pmol each (Chapter 3).

PCR protocol (25 µl per reaction):

PCR mix:	21µl
Primer mix:	1µl
Taq polymerase:	0.5 units (0.1µl)
cDNA:	3µl

Thermal cycling

Initial denaturation 95°C for 10 minutes

Forty cycles each of:

Denaturation at 95°C for 15 seconds

Annealing/extension at 60°C for 1 minute

Read fluorescence at 60°C step

P53 gene variants

The gene for tumor suppressor protein p53 is located on chromosome 17. It is mutated and/or deleted in >50% of human cancers. A large number of p53 gene variants and deletions have been described in various cancers.

The s in p53 gene may be screened by DGGE, SSCP or melting curve analysis. Final confirmation of variants is done by genomic sequencing (Chapter 9).

A strategy for screening of variants in exon 5-8 of p53 gene by melting curve analysis is described:

Primers	Sequence and annealing conditions
Exon5a-F	5’-CAACTCTGTCTCCTTCCTCTTCCTAC
Exon5a-R	5’-AGCCATGGCACGGACGCG
	65–60°C touchdown 0.5°C/cycle for 10 cycles
Exon5b-F	5’-CTCCTGCCCGGCACCCGC
Exon5b-R	5’-CTAAGAGCAATCAGTGAGGAATCAGA
	65–60°C touchdown 0.5°C/cycle for 10 cycles

Exon6-F	5’-CAACCACCCTTAACCCCTCCT
Exon6-R	5’-AGACGACAGGGCTGGTTGC
	68–58°C touchdown 1.0°C/cycle for 10 cycles
Exon7-F	5’-AGGCGCACTGGCCTCATC
Exon7-R	5’-GAGGCTGGGGCACAGCA
	68–58°C touchdown 1.0°C/cycle for 10 cycles
Exon8-F	5’-GACCTGATTTCCTTACTGCCTCTTG
Exon8-R	5’-AATCTGAGGCATAACTGCACCCTT
	63.5–58.5°C touchdown 0.5°C/cycle for 10 cycles

For screening of variants in each exon use the respective pair of primers in separate tubes.

Forward primer concentration: 5pmol

Reverse primer concentration: 5pmol

SYBR green ready reaction mix: 23µl

DNA: 2µl (~200ng)

Thermal cycling programme

Initial denaturation 94^{o}C for 5 minutes.

Thirty cycles each of:

Denaturation at 94^{o}C for 1 minute

Annealing at 60^{o}C for 1 minute (read fluorescence)

Extension at 72^{o}C for 1 minute 30 seconds

Select melting curve analysis form software of real time thermal cycler

BRCA1 and BRCA2 variants

BRCA1 & BRCA2 (BReast CAncer) tumour suppressor genes are located on chromosomes 17 and 13 respectively. A large number of variants have been described in the two genes that are strongly associated with familial breast and ovarian cancer.

The variants may be screened by DGGE, SSCP and melting curve analysis. Final confirmation of the variants is done by genomic sequencing (Chapter 9).

A strategy for screening of variants in exon 2 & 20 of BRCA1 and exon 11 of BRCA2 genes by melting curve analysis is described:

Primers

BRCA1-F	5'-AAAAGATATAGATGTATGTTTTGCTAATGTGT
Exon 2-R	5'-TCCCAAATTAATACACTCTTGTGCTGA
BRCA1-F	5'-GAGTGGTGGGGTGAGATTTTTGTC
Exon 20-R	5'-CCTGATGGGTTGTGTTTGGTTTCT
BRCA2-F	5'-CGAAAATTATGGCAGGTTGTTACG
Exon 11-R	5'-GCTTTCCACTTGCTGTACTAAATCCA

For screening of variants in each exon use the respective pair of primers in separate tubes.

Forward primer concentration: 5pmol

Reverse primer concentration: 5pmol

SYBR green ready reaction mix: 23µl

DNA: 2µl (~200ng)

Thermal cycling programme

Initial denaturation 94^{o}C for 5 minutes.

Thirty cycles each of:

Denaturation at 94^{o}C for 1 minute

Annealing at 60^{o}C for 1 minute (read fluorescence)

Extension at 72^{o}C for 1 minute 30 seconds

Select melting curve analysis form software of real time thermal cycler

JAK-2 variant (V617F)

A simple PCR for identification of JAK-2 V617F variant by ARMS is described:

Primers:

JAK2-Variant-F (5pmol)	5’-AGCATTTGGTTTTAAATTATGGAGTATATT
JAK2-Control-F (5pmol)	5’-ATCTATAGTCATGCTGAAAGTAGGAGAAAAG
JAK2-Common-R (10pmol)	5’-CTGAATAGTCCTACAGTGTTTTCAGTTTCA
Amplified products:	Variant: 203bp

Control: 364bp

Primer mix: Make a mix of the three primers at a concentration of 5pmol each (Chapter 3).

Reaction volume: 25µl

PCR buffer: 22µl

Taq polymerase: 0.5 units (0.1µl)

Primer mix: 1µl (5pmol of each primer/µl)

DNA: 2µl (~200ng)

Thermal cycling

- Initial denaturation 94°C for 5 minutes
- Twenty five cycles of:

  Denaturation at 94°C for 40 seconds

  Annealing at 58°C for 40 seconds

  Extension at 72°C for 1 minute

Electrophoresis: Mini 6% polyacrylamide gels at 150V for 40 minutes

Staining: Silver nitrate

Result: Figure 13.4.

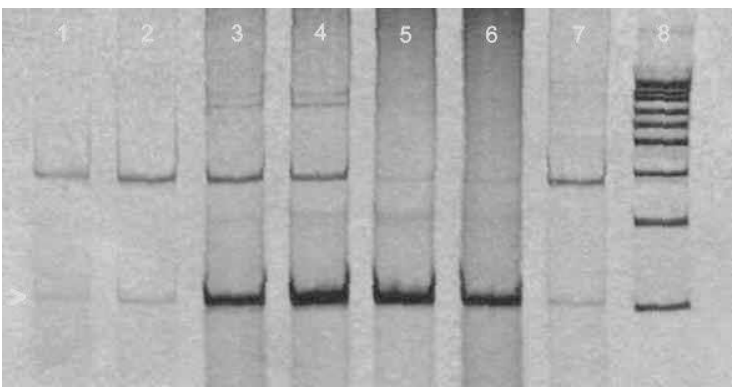


Figure 13.4. Silver stained PAGE of PCR for JAK-2 variant. Arrow pointing at the lanes 3-6 show positive while the lanes 1, 2 & 7 show negative result.

Gene profiling in cancer

DNA Microarrays and more recently NGS has enabled analysis of a very large number of sequence variants in a wide variety of cancers. Such extensive gene profiling is helpful in prediction of prognosis and a more individualized approach towards treatment protocols. However, in developing countries gene profiling may be of limited use because the individualized treatment protocols are usually not an option.

Molecular testing on circulating tumour cells (Liquid Biopsy)

The circulating tumour cells (CTC) in blood can be separated by physical or immunological methods. The nucleic acids extracted from the CTC can be analyzed by RT-PCR. The technique can be useful in estimation of prognostic information, real-time monitoring of therapies, and research related to therapeutic targets, resistance mechanisms, and development of metastasis in cancer patients.

Bibliography

1. Alix-Panabières C, Pantel K. Circulating tumor cells: liquid biopsy of cancer (2013) Clin Chem 59: 110-8.
2. Baxter EJ, Scott LM, Campbell PJ, et al. (2005) Acquired mutation of the tyrosine kinase *JAK2* in human myeloproliferative disorders. Lancet 365: 1054-1061.
3. Gabert J, Beillard E, van der Velden VHJ, Bi W, Grimwade D, Pallisgaard N, Barbany G, Cazzaniga G, Cayuela JM, Cave H, Pane F, Aerts JLE, De Micheli D, Thirion X, Pradel V, Gonza´lez M, Viehmann S, Malec M, Saglio G, van Dongen JJM (2003) Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia – A Europe Against Cancer Program. Leukemia 17: 2318–2357.
4. Gleißner B, Rieder H, Thiel E, Fonatsch C, Janssen LAJ, Heinze B, Janssen JWG, Schoch C, Goekbuget N, Maurer J, Hoelzer D, Bartram CR (2001) Prospective BCR-ABL analysis by polymerase chain reaction (RT-PCR) in adult acute B-lineage lymphoblastic leukemia: reliability of RT-nested-PCR and comparison to cytogenetic data. Leukaemia 15: 1834-1840.
5. Hostein I, Menard A, Soubeyran I, Eghbali H, Debled M, Gastaldello B, Soubeyran P (2001) A 1-kb Bcl-2-PCR fragment detection in a patient with follicular lymphoma and development of a new diagnostic PCR method. Diagn Mol Pathol 10: 89-94.

6. Krypuy M, Ahmed AA, Etemadmoghadam D, Hyland SJ, Australian Ovarian Cancer Study Group, DeFazio A, Fox SB, Brenton JD, Bowtell DD, Dobrovic A (2007) High resolution melting for mutation scanning of *TP53* exons 5–8. BMC Cancer 7: 168-181.
7. Khalil SH, Siegrist K, Akhtar M (1997) Application of Polymerase Chain Reaction to detect rearrangement of Immunoglobulin heavy chain genes in lymphoproliferative disease. Ann Saudi Med 17: 395-398.
8. Potter MN, Steward CG, Oakhill A (1993) The significance of detection of minimal residual disease in childhood acute lymphoblastic leukaemia. Br J Haematol 83: 412-418.
9. Takano EA, Mitchell G, Fox SB, Dobrovic A (2008) Rapid detection of carriers with *BRCA1* and *BRCA2* mutations using high resolution melting analysis. BMC Cancer 8: 59-66.

14

PCR in Infectious Diseases

The PCR based methods for detection of pathogens are rapid and very sensitive. In a clinical sample targets as small as few molecules of DNA or RNA can be detected. PCR is particularly useful in the diagnosis of tuberculosis where culture takes long time or leprosy where culture may not be possible. It is also used for rapid diagnosis of life threatening common bacterial infections and for quick identification of the agents of bioterrorism like *Bacillus anthracis* and *Yersinia pestis*. Bacterial antibiotic resistance genes like MRSA (*Staph aureus*) and VRE (*enterococcus*), and MDR tuberculosis (*katG* and *rpoB* genes) can also be identified by PCR.

PCR is now the method of choice for identification of viral genomes like hepatitis B & C, SARS, CMV, EBV and HIV etc. An interesting application in viral

diseases is the in-situ PCR. The virus particles, for example hepatitis-B virus in the liver cells, CMV in lung, and EBV in association with lymphoma, and HPV in cervical cancer can be demonstrated in a smear or tissue specimen. PCR is also being used for many fungal and parasitic infections.

Points to remember in using PCR for infectious agents:

1. The PCR of infectious agents may give false positive or false negative results as discussed in Chapter 10.
2. With the passage of time the genome of infectious agents tends to develop subtle changes called polymorphisms. The development of DNA or RNA polymorphism may make a previously designed primer or probe ineffective. Similarly, all sub-species/types of an organism or virus may not be detected by the primer or probe due to DNA polymorphisms within the species. However, most infectious agents have some conserved regions of the genome. While designing a primer/probe it is important to choose only the conserved regions of the genome.
3. The presence of cross-homology (similarity) between DNA or RNA of the infectious agent and the host genome can give false positive result. More commonly cross-homology may be present between genome of the infectious agent and the organisms that may be normally present in the clinical sample. The problem of cross-homology is best addressed by BLAST searching of the primer/probe before putting them in use (Chapter 1 & 3).
4. PCR would give a positive result whether the organism in a clinical sample is dead or alive. For example, a patient on anti-tuberculosis treatment may still be having dead *mycobacteria* in the sample that can give positive result on PCR. Similarly, the nucleic acid fragments of a virus like SARS-Cov2 may remain in the host during convalescence giving a positive PCR result.
5. PCR in itself is a very sensitive technique. But the threshold of detection of a bacterial or viral genome in a clinical sample is largely dependent on the method of extraction of DNA or RNA from the sample. The inconsistencies in nucleic acid extraction from the infectious agents may

be overcome by using good quality commercial kits made for this purpose.

6. The transport and storage of clinical samples for PCR is also critical. The samples with organisms/viruses where PCR is done on DNA targets are fairly stable under ordinary transport and storage conditions. However, clinical samples containing RNA viruses are highly prone to degradation by RNases in the environment. Such samples must be processed as early as is possible. When delay is unavoidable the samples like serum/plasma must be separated and stored at -20^{o}C or below. Repeated freeze thawing of the samples is also damaging for the RNA.
7. Viral transport medium (VTM) may be used for the transport of nasopharyngeal swabs suspected to have SARS-Cov2.
8. If the bacteria in a clinical sample are low in number, they may be enriched by culture for a short time before DNA extraction.

There are numerous applications of PCR in the diagnosis of infectious disorders. But here only selected protocols involving common pathogens are described as examples.

PCR for *Mycobacterium tuberculosis*

DNA extraction

DNA extraction from clinical samples like sputum etc. containing M. tuberculosis is difficult. In order to overcome the inconsistencies in extraction of microbial DNA it is advisable to use good quality commercial kits.

Real time PCR

A real time PCR for *Mycobacterium tuberculosis* is described. The PCR primers and the TaqMan® probe (Figure 14.1) for the insertion sequence element IS986/IS6110 of *Mycobacterium tuberculosis* CDC1551 (GenBank accession: AE000516.2) were designed by the Primer Express® software (Applied Biosystems, USA).

CGGAGCTGCGCGATGGCGAACTCAAGGAGCACATCAGCCGCGTCCACGCCGCCAAC
TACGGTGTTTACGGTGCCCGCAAAGTGTGGCTAACCCT**GAACCGTGAGGCATCGA**G
GTGG***CCAGATGCACCGTCGAACGGC***T**GATGACCAAACTCGGCCTGT**CCGGGACCAC
CCGCGGCAAAGCCCGCAGGACCACGATCGCTGATCCGGCCACAGCCCGTCCCGCCG

Figure 14.1. The primers and the TaqMan® probe for the *Mycobacterium tuberculosis* insertion sequence (IS6110): (GenBank accession: AE000516.2).

Primers:

Forward primer:	5'-GAACCGTGAGGGCATCGA
Reverse primer:	5'-ACAGGCCGAGTTTGGTCATC
TaqMan® Probe:	6 FAM 5'-CCAGATGCACCGTCGAACGGC-BHQ1
Amplified product:	64bp
Primer mix:	Make a mix of the two primers and the probe at concentration of 5pmol each (Chapter 3).

PCR protocol (per sample):

Reaction volume:	25µl
PCR mix:	20µl
Primer mix:	1µl
Taq polymerase:	0.5 units (0.1µl)
DNA:	4µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR for *Mycobacterium tuberculosis*

The forward and the reverse primers without the probe can also be used for conventional endpoint PCR of *Mycobacterium tuberculosis*. The amplification is

done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by 64bp amplified fragment.

PCR for *Hepatitis B virus*

DNA extraction

For consistent results the viral DNA is best extracted by commercially available kits using silica columns or magnet beads.

Real time PCR

A TaqMan® probe based real time PCR for *HBV* targeting the conserved 5'Untranslated region of the virus (Figure 14.2) is described.

TCAATCTTCTCGAGGACTGGGGACCCTGCACCGAACATGGAGAGCACAACATCAGG
ATTCCTA**GGACCCCTGCTCGTGTTACA**GGCGGGGTTTTTCT***TGTTGACAAGAATCC***
TCACAATACCACAGAG**TCTAGACTCGTGGTGGACTTCTCTC**AATTTTCTAGGGGGA
GCACCCACGTGTCCTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAAC

Figure 14.2. Real time PCR primers and TaqMan® probe for the conserved 5'-Untranslated region of *Hepatitis-B virus* DNA (GenBank accession: NC_003977.1).

Primers and probe:

Forward primer: 5'- GGACCCCTGCTCGTGTTACA

Reverse primer: 5'- GAGAGAAGTCCACCMCGAGTCTAGA

Probe: 6 FAM 5'- TGTTGACAARAATCCTCACCATACCRCAGA-BHQ1

Primer mix: Make a mix of the two primers and the probe at concentration of 5pmol each (Chapter 3).

PCR protocol (per sample):

Reaction volume: 25µl

PCR mix: 20µl

Primer mix: 1µl

Taq polymerase: 0.5 units (0.1µl)

DNA: 4µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR for *HBV*:

The forward and the reverse primers without the probe can also be used for conventional endpoint PCR of *HBV*. The amplification is done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by 90bp amplified fragment.

PCR for *Epstein-Barr virus (EBV)*

DNA extraction

For consistent results the viral DNA is best extracted by commercially available kits using silica columns or magnet beads.

Real time PCR

A TaqMan® probe based real time PCR for *EBV* targeting a conserved region of the virus is described (Figure 14.3).

GCCCCGTCACGGTGACGTAGTCTGTCTTGAGGAGATGTAGACTTGTAGACACTGCA**AAAC**
CTCAGGACCTACGCTGCCC***TAGAGGTTTTGCTAGGGAGGAGACGTGTG***TGGCTGTAGCCA
CCCGTCCCGGGTACAAGTCCCGG**GTGGTGAGGACGGTGTCTGT**GGTTGTCTTCCCAGACT

Figure 14.3. DNA sequence of *Epstein-Barr virus (EBV)* genome (GenBank accession: V01555.2) showing the real time PCR primers and the TaqMan® probe.

Primers:

Forward primer:	5'-AAACCTCAGGACCTACGCTGC
Reverse primer:	5'-ACAGACACCGTCCTCACCAC
Probe:	FAM 5'-TAGAGGTTTTGCTAGGGAGGAGACGTGTG-BHQ1
Primer mix:	Make a mix of the two primers and the probe at concentration of 5pmol each (Chapter 3).

PCR protocol (per sample):

Reaction volume:	25µl
PCR mix:	20µl
Primer mix:	1µl
Taq polymerase:	0.5 units (0.1µl)
DNA:	4µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR

The forward and the reverse primers without the probe can be used for the conventional endpoint PCR of EBV. The amplification is done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by 107bp amplified fragment.

PCR for *Cytomegalo Virus (CMV)*

DNA extraction

For consistent results the viral DNA is best extracted by commercially available kits using silica columns or magnet beads.

Real time PCR

A TaqMan® probe based real time PCR for *CMV* targeting the human cytomegalovirus glycoprotein B gene of the virus is described (Figure 14.4).

CTTCTTCCCAAACGGTCAGCCATGGTGTTAACGAGACCATCTACAACACTACCCTCAAGT
ACGGAGATGTGGTGGGGGTCAATACCACC**AAGTACCCCTATCGCGTGTG**TTCTA**TGGCCC**
AGGGTACGGATCTTATTCGCTTTGAACGTAATATCGTCTGCACCTCGATGAAGCCCATCA
ATGAA**GACCTGGACGAGGGCATCAT**GGTGGTCTACAAACGCAACATCGTCGCGCACACCT

Figure 14.4. DNA sequence of C*ytomegalo virus* glycoprotein B gene (GenBank accession: M60929.1) showing the real time PCR primers and the TaqMan® probe.

Primers:

Forward primer: 5'-AAGTACCCCTATCGCGTGTG

Reverse primer: 5'-ATGATGCCCTCGTCCAGGTC

Probe: 6 FAM 5'-TGGCCCAGGGTACGGATCTTATTCG-BHQ1

Primer mix: Make a mix of the two primers and the probe at concentration of 5pmol each (Chapter 3).

PCR protocol (per sample):

Reaction volume: 25µl

PCR mix: 20µl

Primer mix: 1µl

Taq polymerase: 0.5 units (0.1µl)

DNA: 4µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR

The forward and the reverse primers without the probe can be used for the conventional endpoint PCR of *CMV*. The amplification is done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by 116 bp amplified fragment.

PCR for *HCV*

PCR of *HCV*, being RNA, is done in two steps. In the first stem *HCV* RNA is converted to cDNA and in the second step cDNA is amplified by PCR. A method targeting the highly conserved 5'-untranslated region of the virus is described. Figure 14.5 shows a set of primers and TaqMan® probes for the 5' untranslated region of HCV genome.

CACTCCACCATGAATCACTCCCCTGTGAGGAACTACTGTCTTCACGCAGAA**AGCGTCTAGC**
CATGGCGTTAGTATGAGTGTCGTGCAGCCTCCAGGAC**CCCCCCTCCCGGGAGAGCCATAGT**
GGTCTGCGGAACCGGT**GAGTACACCGGAATTGCCAGGA**CGACCGGGTCCTTTCTTGGATCA

Figure 14.5. Sequence of 5' untranslated region of *HCV* genome (GenBank: EU678744.1) showing a set of primers and TaqMan® probe for real time PCR.

RNA Extraction

For consistent results the viral RNA is best extracted by commercially available kits using silica columns or magnet beads. Carrier RNA is also required for the stability of viral RNA especially when the copy number is low.

Preparation of cDNA:

Reverse transcription (RT):

Primer (HCV-R) 5'-TCCTCGCAATTCCGGTGTACTC

Reactions conditions:

- 5 X reaction buffer: 4 µl
- 10 mmol dNTPs mix: 2 µl

- RT Primer (10 pmol/µl): 1 µl
- MMLV- Reverse Transcriptase (200 U/µl): 1 µl
- RNase inhibitor (20 U/µl): 1 µl
- Deionized water: 11 µl
- RNA: 2 µl
- Incubation: 42°C for 60 minutes
- RT inactivation: 70°C for 5 minutes

Real Time PCR

PCR Primers:

HCV-F 5'-AGCGTCTAGCCATGGCGTTAGTAT

HCV-R 5'- TCCTCGCAATTCCGGTGTACTC

HCV Probe 6 FAM-CCCCCCTCCCGGGAGAGCCATAGT-BHQ1

Amplified Product: 109bp

Reaction volume: 25µl

PCR mix: 21µl

Primer mix: 1µl (containing each primer at 5pmol/µl concentration)

Taq polymerase: 0.5 units (0.1µl)

cDNA: 3µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

One Step PCR for Viral RNA

Several commercial kits are available for one step synthesis of cDNA and PCR amplification in the same reaction vial. The primers for HCV described above can be used with the one step kits.

Endpoint PCR for *HCV*

The forward and the reverse primers without the probe can also be used for conventional endpoint PCR of *HCV*. cDNA is prepared as described above using the *HCV*-R primer. The amplification is also done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by a 109 bp amplified fragment.

PCR for SARS-Cov2 (Covid-19):

A real time PCR for SARS-Cov2 causing the pandemic that started in 2019 is described. The primers and the probe targeting the open reading frame *1ab* gene of SARS-CoV2 are used. For greater accuracy additional primers/probe targeting an additional conserved region of the *SARS-Cov2* may also be used. The PCR is done in two steps. In the first step the viral RNA is converted to cDNA and in the second step cDNA is amplified by PCR.

RNA Extraction

For best and consistent results RNA extraction by commercial kits using silica columns or magnetic beads is recommended.

Preparation of cDNA:

Reverse transcription (RT):

Primer Cov-19-R 5'-TTCCATCTCTAATTGAGGTTGAACC

Reactions conditions:

- 5 X reaction buffer: 4 µl
- 10 mmol dNTPs mix: 2 µl
- RT Primer (10 pmol/µl): 1 µl
- MMLV- Reverse Transcriptase (200 U/µl): 1 µl
- RNase inhibitor (20 U/µl): 1 µl
- Deionized water: 9 µl
- RNA: 2 µl

- Incubation: 42°C for 60 minutes
- RT inactivation: 70°C for 5 minutes

TGGTGCCACTTCTGCTGCTCTTCAACCTGAAGAAGAGCAAGAAGAAGATTGGTTAG
ATGATGATAGTCAACAAACTGT**TGGTCAACAAGACGGCAGTGAGGA**CAATCAGACA
ACTACTATTCAAACAATTGTTGA**GGTTCAACCTCAATTAGAGATGGAA**CTTACACC
AGTTGTTCAGACTATTGAAGTGAATAGTTTTAGTGGTTATTTAAAACTTACTGACA

Figure 14.7. The primers and the TaqMan® probe for real time PCR targeting the open reading frame *1ab* gene of *SARS-CoV2*.

Real Time PCR

PCR Primers:

Cov-19-F 5′-AGAAGATTGGTTAGATGATGATAGT

Cov-19-R 5′-TTCCATCTCTAATTGAGGTTGAACC

Cov-19-Probe 5′-FAM-TCCTCACTGCCGTCTTGTTGACCA-BHQ1

Amplified Product: 118bp

Reaction volume: 25µl

PCR mix: 21µl

Primer mix: 1µl (each primer at 5pmol/µl concentration)

Taq polymerase: 0.5 units (0.1µl)

cDNA: 3µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR for *SARS-Cov2*

The forward and the reverse primers without the probe can be used for the conventional endpoint PCR of *SARS-Cov2*. cDNA is prepared as described above using the Cov-19-R primer. The amplification is also done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose gel at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 20 minutes (Chapter 5). Positive result is shown by a 118 bp amplified fragment.

PCR for *Dengue virus*:

A pan dengue PCR is described that can identify all four genotypes of the virus. The PCR of *Dengue virus*, being RNA virus, is done in two steps. In the first step the viral RNA is converted to cDNA and in the second step cDNA is amplified by PCR. A method using the primers and TaqMan® probe (Figure 14.6) for the highly conserved 3' untranslated region shared by the four genotypes of the *Dengue virus* is described.

RNA Extraction

For best and consistent results RNA extraction by commercial kits using silica columns or magnetic beads is recommended.

Preparation of cDNA:

Reverse transcription (RT):

Primer (DENV-R) 5'-ACCATTCCATTTTCTGGCGTT

Reactions conditions:

- 5 X reaction buffer: 4 µl
- 10 mmol dNTPs mix: 2 µl
- RT Primer (10 pmol/µl): 1 µl
- MMLV- Reverse Transcriptase (200 U/µl): 1 µl
- RNase inhibitor (20 U/µl): 1 µl
- Deionized water: 9 µl
- RNA: 2 µl
- Incubation: 42°C for 60 minutes

- RT inactivation: 70°C for 5 minutes

DENV-1
TAGAGGTTAGAGGAGACCCCCCGCACAACAACAAACAGCATATTGACGCTGGGAGAGAC
CAGAGATCCTGCTGTCTCTACAGCATCATTCCAGGCACAGAACGCCAAAAAATGGAATG
GTGCTGTTGAATCAACAGGTTCT

DENV-2
CTAGAGGTTAGAGGAGACCCCCCCAAAACAAAAACAGCATATTGACGCTGGGAAAGAC
CAGAGATCCTGCTGTCTCCTCAGCATCATTCCAGGCACAGAACGCCAGAAAATGGAATG
GTGCTGTTGAATCAACAGGTTCT

DENV-3
ACTAGAGGTTAGAGGAGACCCCCCGCAAATAAAAACAGCATATTGACGCTGGGAGAGAC
CAGAGATCCTGCTGTCTCCTCAGCATCATTCCAGGCACAGAACGCCAGAAAATGGAATG
GTGCTGTTGAATCAACAGGTTCT

DENV-4
ACTAGAGGTTAGAGGAGACCCCCCCAACACAAAAACAGCATATTGACGCTGGGAAAGAC
CAGAGATCCTGCTGTCTCTGCAACATCAATCCAGGCACAGAGCGCCGCAAGATGGATTG
GTGTTGTTGATCCACAACAGGTTCT

Figure 14.6. The primers and the TaqMan® probe for real time PCR targeting the 3' un-translated region of four genotypes of *Dengue virus*. Minor differences in the sequence of the four dengue virus genotypes are shown as gray.

Real Time PCR

PCR Primers:

DENV-F 5'-AGAGACCAGAGATCCTGCTGTCTC

DENV-R 5'-TGATTCAACAGCACCATTCCAT

DENV-Probe 5'-FAM-AGCATCATTCCAGGCACAGAACGCC-BHQ1

Amplified Product: 68bp

Reaction volume: 25µl

PCR mix: 21µl

Primer mix: 1µl (containing each primer at 5pmol/µl concentration)

Taq polymerase: 0.5 units (0.1µl)

cDNA: 3µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR for *Dengue virus*

The forward and the reverse primers without the probe can be used for the conventional endpoint PCR of Dengue virus. cDNA is prepared as described above using the DENV-R primer. The amplification is also done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by a 68 bp amplified fragment.

PCR for *Malaria*

A TaqMan® probe based real time PCR for malaria targeting a conserved region of *Plasmodium falciparum, P. vivax, P. malariae*, and *P. ovale* is described (Figure 14.7).

DNA extraction:

Commercial kits or Chelex based method on whole blood collected in EDTA (Chapter 2) can be used.

Real Time PCR

Forward primer:	5'-ACATGGCTATGACGGGTAACG
Reverse primer:	5'-TGCCTTCCTTAGATGTGGTAGCTA
Probe:	6 FAM 5'-TCAGGCTCCCTCTCCGGAATCGA-BHQ1
Primer mix:	Make a mix of the two primers and the probe at concentration of 5pmol each (Chapter 3).

P. falciparum (GenBank M19172.1)
TAGTGTGTATCAATCGAGTTTCTGACCTATCAGCTTTTGATGTTAGGGTATTGGCCTA**ACAT**
GGCTATGACGGGTAACGGGGAATTAGAGT**TCGATTCCGGAGAGGGAGCCTGA**GAAA**TAGCTA**
CCACATCTAAGGAAGGCAGCAGGCGCGTAAATTACCCAATTCTAAAGAAGAGAGGTAGTGAC
AAGAAATAACAATGCAAGGCCAATTTTTGGTTTTGTAA

P. vivax (GenBank X13926.1)
TTAGTGTGTATCAATCGAGTTCTGACCTATCAGCTTTTGATGTTAGGGTATTGGCCTA**ACAT**
GGCTATGACGGGTAACGGGGAATTAGAGT**TCGATTCCGGAGAGGGAGCCTGA**GAAA**TAGCTA**
CCACATCTAAGGAAGGCAGCAGGCGCGTAAATTACCCAATTCTAAAGAAGAGAGGTAGTGAC
AAGAAATAACAATACAAGGCCAATCTGGCTTTGTAATT
P. malariae (GenBank M54897.1)
TAGTGTGTATCAATCGAGTTTCTGACCTATCAGCTTTTGATGTTAGGGTATTGGCCTA**ACAT**
GGCTATGACGGGTAACGGGGAATTAGAGT**TCGATTCCGGAGAGGGAGCCTGA**GAAA**TAGCTA**
CCACATCTAAGGAAGGCAGCAGGCGCGTAAATTACCCAATTCTAAAGAAGAGAGGTAGTGAC
AAGAAATAACAATGCAAGGCCAAATTTTGGTTTTGCAA

P. ovale (GenBank: L48987.1)
TAGTGTGTATCAATCGAGTTTCTGACCTATCAGCTTTTGATGTTAGGGTATTGGCCTA**ACAT**
GGCTATGACGGGTAACGGGGAATTAGAGT**TCGATTCCGGAGAGGGAGCCTGA**GAAA**TAGCTA**
CCACATCTAAGGAAGGCAGCAGGCGCGTAAATTACCCAATTCTAAAGAAGAGAGGTAGTGAC
AAGAAATAACAATACAAGGCCATTTCATGGTTTTGTAA

Figure. 14.7. Sequence and the primers for real time PCR targeting the Small Subunit (SSU) RNA gene of the four subspecies of malarial parasite.

PCR protocol (per sample):

Reaction volume:	25µl
PCR mix:	22µl
Primer mix:	1µl
Taq polymerase:	0.5 units (0.1µl)
DNA:	2µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
    i. Denaturation at 95°C for 15 seconds
    ii. Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR

The forward and the reverse primers without the probe can be used for the conventional endpoint PCR of Malaria. The amplification is done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by 84 bp amplified fragment.

PCR for *Leishmania donovani*

A TaqMan® probe based real time PCR for *Leishmania donovani* glucose phosphate isomerase gene (Figure 14.8) is described.

GTACGGCAGCACCGCCTCTGTCTCCGCGCCGAAAAAGTTGTTGTA**CCAGATGCCGACCAAAG**
CCAGCATC**ATCGGCAGGTTCTGCTCCGTCG**GTGCAGACGCAAAGTG**GTTATCCATCACGTGC**
GCGCCAGTCAGGAACTCCACAAAGTTGTCGTAGCCGATCGAAAGCATCACGGAGAGACCGAT

Figure 14.8. DNA sequence of *Leishmania donovani* isolate WR 378 glucose phosphate isomerase gene (GenBank: AY974201.1).

DNA extraction

DNA extraction can be done by commercial kits or Chelex method from 0.2 ml bone marrow aspirate or archival bone marrow smears (Chapter 2).

Real Time PCR

Forward primer:	5'-CCAGATGCCGACCAAAGC
Reverse primer:	5'-CGCGCACGTGATGGATAAC
Probe:	6 FAM 5'-ATCGGCAGGTTCTGCTCCGTCG-BHQ1
Primer mix:	Make a mix of the two primers and the probe at concentration of 5pmol each (Chapter 3).

PCR protocol (per sample):

Reaction volume:	25µl
PCR mix:	22µl
Primer mix:	1µl

Taq polymerase: 0.5 units (0.1µl)

DNA: 2µl

Thermal cycling

- Initial denaturation 95°C for 10 minutes
- Forty cycles each of:
  - Denaturation at 95°C for 15 seconds
  - Annealing/extension at 60°C for 1 minute
- Take fluorescence reading at 60°C step

Endpoint PCR for *Leishmania donovani*

The forward and the reverse primers without the probe can be used for the conventional endpoint PCR of *Leishmania donovani*. The amplification is done as described above except that the probe is omitted from the primer mix. At the end of the PCR the amplified product may be run on 2% agarose at 100 volts for 30 minutes or 6% mini-polyacrylamide gel at 150 volts for 25 minutes (Chapter 5). Positive result is shown by 82 bp amplified fragment.

Point of care PCR for infectious diseases

Development in microfluidic devices using channels to transport small amounts of reagents has led to the development of on-chip PCR allowing multiplex analysis and high throughput screening for infectious diseases in 15-30 minutes. It miniaturizes and integrates the various steps, including extraction and purification of nucleic acids, PCR amplification, and on-chip detection of reaction products.

Direct processing of a clinical sample and the high speed of processing has enabled this technique to be used at the point of care (POC). The POC PCR is done in a special cartridge that allows direct loading of the clinical sample. The cartridge contains a micro-fluidic device that allows extraction of nucleic acid, and real time PCR amplification. The cartridges from different patients are run on a specially designed real time thermal cycler.

Specially designed cartridges for the pathogens found in different clinical samples like sputum, blood, CSF, stool and urine etc. are available. A high cost of the cartridges is a limitation in the use of POC PCR in developing countries.

Bibliography

1. Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, Yao JDC, Wengenack NL, Rosenblatt JE, Cockerill III FR, Smith TF (2006) Real-Time PCR in Clinical Microbiology: Applications for routine laboratory testing. Clin Microbiol Rev 19: 165–256.
2. Gurukumar KR, Priyadarshini D, Patil JA, Bhagat A, Singh A, Shah PS, Cecilia D (2009) Development of real time PCR for detection and quantitation of Dengue viruses. Virology Journal 6: 10-17.
3. Jebbink J, Bai X, Rogers BB, Dawson DB, Scheuermann RH, Domiati-Saad R (2003) Development of Real-Time PCR Assays for the Quantitative Detection of Epstein-Barr Virus and Cytomegalovirus, Comparison of TaqMan Probes, and Molecular Beacons. J Mol Diagn 5: 15–20.
4. Kozel TR, Burnham-Marusich AR (2017) Point-of-care testing for infectious diseases: past, present, and future. J Clin Microbiol 55: 2313-20.
5. Lee MA, Tan CH, Aw LT, Tang CS, Singh M, Lee SH, Chia HP, Yap EPH (2002) Real-Time Fluorescence-Based PCR for Detection of Malaria Parasites. J Clin Microbiol 40: 4343–4345.
6. Meng S and Li J (2010) A novel duplex real-time reverse transcriptase-polymerase chain reaction assay for the detection of hepatitis C viral RNA with armored RNA as internal control. Virology Journal 7: 117-125.
7. Pas SD, Fries E, DeMan RA, Osterhaus ADME, Niesters HGM (2000) Development of a quantitative real-time detection assay for hepatitis B virus DNA and comparison with two commercial assays. J Clin Microbiol 38: 2897-2901.
8. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA*. Published February 7, 2020. doi:10.1001/jama.2020.1585
9. Wortmann G, Hochberg L, Houng HH, Sweeney C, Zapor M, Aronson N, Weina P, Ockenhouse CF (2005) Rapid identification of Leishmania complexes by real time PCR assay. Am J Trop Med Hyg 73: 999–1004.

15

Forensic DNA Testing

DNA from no two individuals, except identical twins, is alike. The person to person differences in DNA can be analyzed by PCR amplification or genomic sequencing. The main advantages of DNA typing are the universality of application, an almost unlimited power to discriminate, extreme sensitivity, and reasonable resistance to degradation by environmental factors. The typing of DNA can be used in:

- Linking suspect to a crime
- Excluding falsely accused person
- Recognizing serial crimes
- Resolving parentage disputes
- Identification of the remains of victims

The differences in DNA are either in the form of variable number of tandem repeats (VNTR) or single nucleotide polymorphisms (SNP). Approximately 20% of the human genome comprises tandem repeat sequences. The tandem repeats

are called micro-satellites or short tandem repeats (STR) when the repeats are 2-6 bp in length. When the repeat sequences are 7-80 bp in length these are called mini-satellite or variable number of tandem repeats (VNTR). The number of STR repeat units tends to vary between individuals. This variation (polymorphism) makes them extremely useful in applications like human identification and linkage analysis for diagnosis of genetic disorders.

STR analysis

There are over a million sites (loci) in the human genome that have STRs of which over 20,000 have been characterized. Considering the highly polymorphic nature and the ease with which the STRs can be analyzed these provide an extremely powerful tool for human identification. The discrimination power of STRs increases with increase in the number of loci tested. Therefore it is common to use a battery of STRs. The number of bases in a repeat unit of STRs range from 2-6. Due to technical reasons the four base pair (tetra-nucleotide) repeats are best suited for forensic case work.

STR loci

Most STRs are located in the non-coding DNA between genes (inter-genic DNA). Some STRs are also present in the intervening sequences (introns) of known genes. The STRs found in the inter-genic DNA are named according to the chromosome number and the site. For example in “D5S818” “D” stands for DNA, “5” is the chromosome number, “S” stands for single copy sequence and “818” is the locus number. The STRs in the introns of the known genes are identified by their location e.g. TH01 is present in the intron-1 of human tyrosine hydroxylase gene and TPOX is located in the thyroid peroxidase gene.

In a typical tetra-nucleotide STR the repeat units comprise four base pairs. For example STR with “GATA” repeat would look like:

“GTATCCTTATGTAATATTTTGAA**GATA**GATA**GATA**GATA**GATA**GATA**GATA**GATA**G**ATA**GATA**GATA**GGTAGATAGAGGTATAAATAAGGATACAGATATAG”

Autosomal STRs

In most forensic DNA applications STRs located on autosomes are used. A list of 15 core STR loci and their chromosomal locations are shown in Table 15.1 (Short Tandem Repeat DNA Internet DataBase).

Table. 15.1. Core STR loci used in forensic DNA testing.

Name	Chromosome	Location	Type	GenBank
TPOX	2p23-2pter	Thyroid Peroxidase Gene	Simple	M68651
D3S1358	3p21	-	Complex	11449919
FGA	4q28	Alpha Fibrinogen Gene	Complex	M64982
D5S818	5q21-q31	-	Simple	G08446, AC008512
CSF1PO	5q33.3-34	CSF-1 receptor Gene	Simple	X14720
D7S820	7q	-	Simple	G08616, AC004848
D8S1179	8q24.1-24.2	-	Simple	GO8710, AF216671
Th01	11p15-15.5	Tyrosine Hydroxylase Gene	Complex	D00269
vWA	12p12-pter	von Willebrand antigen Gene	Complex	M25858
D13S317	13q22-q31	-	Simple	G09017, AL353628.2
D16S539	16q22-24	-	Simple	G07925, AC024591.3
D18S51	18q21.3	-	Simple	X91254, AP001534
D21S11	21q21.1	-	Complex	M84567, AP000433
D2S1338	2q35	-	Complex	AC010136, G08202
D19S433	19q12	-	Complex	G08036, AC008507.6

Y-Chromosome STRs

Several STRs have also been identified on Y-chromosome. These are useful in tracing male DNA in investigations like sexual assault. These may also be used in investigation of paternal lineage inheritance. A list of 15 Y-STR markers available as a commercial kit (Life Technologies, USA) is shown in Table 15.2 (Short Tandem Repeat DNA Internet DataBase).

STR allele nomenclature

The STR alleles are named according to the number of repeat units it contains e.g. 7, 8, 9, 10, 11 etc. Some STRs are more complex than simple repeats. The complexity may be present both in the sequence as well as the number of bases in the repeat unit. The variation in number of bases in a repeat results in different sizes that can be picked on gel electrophoresis. Whereas the variation in

sequence of the repeats can only be found by genomic sequencing. For example, at the TH01 locus allele 9 has nine simple repeats i.e. $[AATG]_9$. Another allele at the same locus has an additional triplet i.e. $[AATG]_6ATG[AATG]_3$. The resulting allele is 3bp longer and is written as 9.3. The D21S11 locus contains numerous complex alleles like 32.1, 32.2 etc. The analysis of complex loci is difficult but being more polymorphic these are more informative.

The STRs are inherited in a simple Mendelian fashion. An individual inherits an allele each from its father and the mother. A person may be homozygous (the same allele on the maternal and the paternal chromosomes) or compound heterozygous (different alleles on the two chromosomes).

Table 15.2. Fifteen Y-STRs available as a commercial kit (Life Technologies, USA).

Locus	Repeat Numbers	Repeat Motif	GenBank Accession	Reference Allele
DYS19	10-19	TAGA	AC017019	15
DYS385 a/b	7-28	GAAA	AC022486	11
DYS389 I	DYS389I: 9-17	(TCTG) (TCTA)	AC004617	12, 29
DYS389 II	DYS389II:24-34	(TCTG) (TCTA)	AC004617	-
DYS390	17-28	(TCTA) (TCTG)	AC011289	24
DYS391	6-14	TCTA	AC011302	11
DYS392	6-17	TAT	AC011745	13
DYS393	9-17	AGAT	AC006152	12
DYS437	13-17	TCTA	AC002992	16
DYS438	6-14	TTTTC	AC002531	10
DYS439	9-14	AGAT	AC002992	13
DYS448	20-26	AGAGAT	AC025227	22
DYS456	13-18	AGAT	AC010106	15
DYS458	13-20	GAAA	AC010902	16
DYS635	17-27	TSTA compound	AC004772	23

STR profile

The STR profile of an individual is referred to the combination of various genotypes at several loci. A typical STR profile comprising genotypes at several loci is shown in Table 15.3.

Table 15.3. The STR profile at seven different loci in individuals A, B, and C.

Individuals	D3S1358	D5S818	D7S820	D8S117	D21S11	TH01	TPOX
A.	15,18	12,12	9,10	12,16	28,29	7,9	9,11
B.	14,14	10,12	8,10	11,15	30,32.2	7,9.3	8,11
C.	15,17	10,10	9,11	13,15	29,32.2	8,10	8,11

Allele frequencies

In a given population the STR allele frequencies are calculated by simple counting. Each individual has two alleles at each locus. Genotyping of 100 unrelated individuals from a population would mean examination of 200 chromosomes (alleles). Each allele of a compound heterozygote is counted as one and homozygotes are counted as two. For example if 20/100 people have allele 8, including 2 homozygotes (8,8), the frequency of allele 8 would be 18+2+2 = 22/200 i.e. 11% or 0.11.

Minimum allele threshold

In a population survey the uncommon or the rare alleles are expected to have an under-representation. In order to overcome this problem it has been recommended to inflate the frequency of rare alleles (<5 counts) to 5. The 5/2N formula is used for this purpose where N is the number of individuals examined. The N is doubled because each individual has two chromosomes (alleles). For example if allele 12 is observed in 2/100 people its frequency by the conventional method would be 2/2x100 = 1% (0.01). However, by the 5/2N formula its frequency would be 5/2x100 (2.5%) or 0.025.

Genotype frequencies

The combination of alleles in an individual at a locus is called its genotype e.g. “7,10” or “7,7” etc. The number of possible genotypes increases with an increase in the number of alleles. The possible genotypes can be calculated by the formula [n(n+1)/2] where n is the number of different alleles found in the population. The observed genotype frequencies are calculated by simple counting. The expected genotype frequencies can be calculated from the observed allele frequencies by Hardy Weinberg equation ($p^2 + q^2 + 2pq = 1$). The homozygotes of two alleles

with frequencies of p and q would be equal to p^2 and q^2 respectively while the heterozygotes (compound heterozygotes) would be equal to $2pq$.

Example

Table 15.4 describes an example of calculation of expected genotype frequencies of alleles 10 and 11 with observed frequencies of p and q respectively.

Table 15.4. The expected genotypes of two alleles at D5S818 locus.

D5S818	Frequency	p^2	q^2	$2pq$
Allele 10	p = 0.108	0.0117	-	0.0687
Allele 11	q = 0.318	-	0.101	0.0687

Profile frequencies

The STR profile of an individual is the combination of genotypes at several loci. Larger the number of loci tested rarer would be the combination in the population. The frequency of a profile is calculated by the multiplication rule. The individual genotype frequencies are multiplied to get the combined frequency (see example). The combined frequency of 15 core STR loci in the US Caucasian population is 1 in 2.46 quadrillion (10^{15}). Table 15.5 gives an example of how a profile frequency can be calculated from various allele and genotype frequencies.

Forensic calculations and consanguineous marriage

Hardy-Weinberg equation for the calculation of expected genotype frequencies is for populations where mating is random. In a population where consanguineous marriage and marriage between tribe members is very common, the population has several substructures (smaller groups). Therefore Hardy-Weinberg equation would not be applicable as such.

The main genetic effect of consanguineous marriage is an increase in the proportion of homozygotes and a corresponding reduction in the heterozygotes. The increase in homozygotes as compared to Hardy-Weinberg proportions is by an amount Fpq, while heterozygotes are reduced by $2Fpq$ where F is the

inbreeding coefficient, and *p* and *q* are the frequencies of the alleles under consideration.

Homozygotes = $p^2 + Fpq$ or $q^2 + Fpq$

Heterozygotes = $2pq\ (1 - F)$

Table 15.5. Example of calculation of profile frequency from the genotype frequencies in a given population.

Loci	Alleles	Allele frequency	p^2	q^2	$2pq$
D3S1358	15	p = 0.299	0.0894	-	0.0466
	18	q = 0.078	-	0.0061	0.0466
D5S818	10	p = 0.108	0.0117	-	0.0687
	11	q = 0.318	-	0.1011	0.0687
D8S1179	12	p = 0.103	0.0107	-	0.0132
	16	q = 0.064	-	0.0041	0.0132
D21S11	30.2	p = 0.059	0.0035	-	-
	30.2	-	-	-	-
Profile		Profile frequency			
D3 & D5		0.0466 X 0.0687 = 0.0032 (1 in 312)			
D3, D5 & D8		0.0466 X 0.0687 X 0.0132 = 0.000042 (1 in 23,809)			
D3, D5, D8 & D21*		0.0466 X 0.0687 X 0.0132 X 0.0035 = 0.00000015 (1 in 6666,666)			

*additional loci can also be added to this calculation

The coefficient of inbreeding is the probability that an individual receives at a given locus two genes that are identical by descent (copies of a single gene carried by a common ancestor). The value of *F* for a first cousin marriage is 0.0625 i.e. 6.25% of the genes are identical by descent. In 1½ cousin and 2nd cousin marriage *F* is 0.0313 and 0.0156 respectively. The average coefficient of inbreeding in the Pakistani population is 0.0280 that may be as high as 0.0350 in selected populations/tribes.

Example

The correction for consanguineous marriage applied to the genotype and profile frequency in the example shown in Table 15.5 is presented in Table 15.6.

Table 15.6. Correction for consanguineous marriage applied to the genotype and profile frequency of the example presented in Table 15.5.

Loci	Alleles	Allele frequency	$p^2 + Fpq$	$q^2 + Fpq$	$2pq\ (1 - F)$
D3S1358	15	$p = 0.299$	0.0900	-	0.0453
	18	$q = 0.078$	-	0.0068	0.0453
D5S818	10	$p = 0.108$	0.0127	-	0.0668
	11	$q = 0.318$	-	0.1021	0.0668
D8S1179	12	$p = 0.103$	0.0109	-	0.0128
	16	$q = 0.064$	-	0.0059	0.0128
D21S11	30.2	$p = 0.059$	0.0036*	-	-
	30.2	-	-	-	-
Profile		Profile frequency			
D3 & D5		0.0453 X 0.0668 = 0.00303 (1 in 330)			
D3, D5 & D8		0.0453 X 0.0668 X 0.0128 = 0.000039 (1 in 25,641)			
D3, D5, D8 & D21		0.0453 X 0.0668 X 0.0128 X 0.0036 = 0.0000014 (1 in 7142,857)			

* $p^2 + Fp(1-p)$

Spontaneous mutations in STRs

With the passage of time genomic DNA is likely to develop spontaneous mutations. In fact the highly polymorphic STRs are thought to have developed due to spontaneous mutations. The STR mutations may become relevant if these are encountered in solving a case of inheritance.

The rate of spontaneous mutations at the STR loci ranges from 1 in 500 to 1 in 1000. The rates of mutations at the core STR loci used in forensic DNA testing are given in Table 15.7. A mutation may be suspected if a disagreement is found between alleles of the parents and the offspring at one odd locus out of the several tested. The mutation can be confirmed by genomic sequencing. It may be pointed out that matching between DNA samples from a crime scene and a suspect is not affected by spontaneous mutations.

Core STR loci used in human identification

The core STR loci are sets of DNA markers that are globally accepted for human identification (Table 15.8). The uniformity is adopted to share and compare genetic information between different labs and the legal systems. The loci have

been carefully selected to avoid their linkage to any physical character or genetic disease. In addition to the STRs a DNA marker for sex determination, usually the amelogenin gene, is also included in the profile.

Table 15.7. Mutation rates at the STR loci commonly used in forensic DNA testing.

Locus	Mutation Rate	Locus	Mutation Rate	Locus	Mutation Rate
CSF1PO	0.16%	D3S1358	0.12%	D16S539	0.11%
FGA	0.28%	D5S818	0.11%	D18S51	0.22%
TH01	0.01%	D7S820	0.10%	D21S11	0.19%
TPOX	0.01%	D8S1179	0.14%	D2S1338	0.12%
VWA	0.17%	D13S317	0.14%	D19S433	0.11%

Table 15.8. Core STR loci required by various countries and the legal systems.

Countries	Core STR Loci
US	CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Amelogenin
UK/European	FGA, TH01, vWA, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, Amelogenin Recommended Loci: D1S1656, D2S441, D10S1248, D12S391, D22S1045, TPOX
Interpol	FGA, TH01, vWA, D3S1358, D8S1179, D18S51, D21S11, Amelogenin

COmbined DNA Index System (CODIS)

United States Federal Bureau of Investigation has created a database, called CODIS, that stores the DNA profiles of convicted offenders and the biological material found at crime scenes. The database contains DNA profiles comprising 15 STR loci including CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, THO1, TPOX, vWA and Amelogenin to determine sex. Several million entries exist in the database making it the largest in the world. The database has helped in solving a very large number of criminal investigations.

Single Nucleotide Variants (SNV)

SNVs, previously called Single Nucleotide Polymorphisms (SNPs), are DNA sequence variations that are randomly distributed throughout the genome and are found after every 300 to 500 base pairs. Almost all SNVs have two alleles. By convention the base change is called polymorphism when the frequency of its alleles in a population is more than 1%. SNVs are present in the non-coding and the coding sequences of DNA.

The SNVs may be identified by restriction enzyme analysis or allele specific primers and gel electrophoresis (Chapter 8). Real time PCR can also be used to analyze multiple SNVs in a single tube by using multiplex allele specific probes. The automated SNV genotyping by micro arrays also allow analysis of very large number of SNVs on a single gene chip. The latter can be especially useful for human identification in mass disasters.

Mitochondrial DNA

Each mitochondrion contains 2-10 copies of a circular piece of DNA, 16,569 base pairs in length. It is inherited from the mother. Mitochondrial DNA contains two hyper variable regions (HV1 and HV2) that contain many SNPs. The regions can be amplified and sequenced. Since mitochondrial DNA is inherited from the mother it can be useful in tracing the maternal inheritance. Mitochondrial DNA can be extracted from structures like hair shafts, old bones and teeth etc.

PCR for sex determination

Determination of sex is an essential part of human identification by DNA testing. Most commonly it is done by amplifying a sequence from the amelogenin gene whose length varies between a male and a female.

Degraded DNA and “Mini STRs”

The DNA is a very large molecule that can be easily broken to smaller fragments by shearing force or bacterial enzymes (Chapter 2). Such degraded DNA samples are poorly amplified. The problem is most marked in highly degraded DNA samples. In precious and trace forensic samples degradation can completely jeopardize its analysis.

The problem of degraded DNA can be solved to some extent by using "Mini STRs". In a mini STR analysis the forward and the reverse amplification primers flanking the repeat units are brought to the nearest possible distance from the repeat units. The resultant amplified fragments are of smallest possible size. Many degraded DNA samples can be analyzed by the mini STR methods to provide sufficient information.

Limitations of DNA test for human identification

Some of the limitations of forensic DNA test are:

1. It cannot tell the age of the person.
2. In some situations it may provide information about predisposition to disease, color of eyes, height or hair color.

Applications of forensic DNA testing

Matching suspect with evidence:

The profiles of the two or more DNA samples are aligned to see if there is any difference or not. There could be three possible outcomes:

1. Match: When the two or more samples have the same genotypes. The statistical significance of the match is calculated as described in a subsequent section.
2. Exclusion: When the two samples originate from two different sources. This does not require prior knowledge of the allele frequencies in the population.
3. Inconclusive: Neither of the above two outcomes.

Example

The STR profiles of four DNA samples are shown in Table 15.9. Sample 1 is picked from crime scene whereas samples 2-4 are of the suspects. The samples 1 and 2 are completely matching whereas the samples 3 and 4 do not match and are therefore excluded.

Table 15.9. Comparison of the STR profile of a DNA sample collected from a crime scene (serial 1) and three suspects (serial 2-4). The profile of sample at serial 2 completely matches with that of the crime scene DNA whereas the samples at serial 2 & 3 are excluded from the match.

ID	D2S1338	D3S1358	D5S818	D7S820	D8S1179	D13S317	D16S539	D18S51
1	22,25	17,17	11,11	8,11	10,16	9,11	12,13	13,13
2	22,25	17,17	11,11	8,11	10,16	9,11	12,13	13,13
3	24,24	18,18	12,14	9,11	8,13	11,11	11,13	14,14
4	20,23	15,15	9,11	8,12	13,16	11,12	10,10	14,17
ID	**D19S433**	**D21S11**	**FGA**	**TH01**	**TPOX**	**vWA**	**CSF1PO**	**Amel**
1	13,15	29,30	21,25	6,9	8,9	16,16	11,11	XY
2	13,15	29,30	21,25	6,9	8,9	16,16	11,11	XY
3	13,14	28,31.2	21,23	7,9	8,9	21,21	11,12	XY
4	12,14	28,32.2	24,24	6,8	8,11	16,16	8,11	XY

Probability of match

The example shown in Table 15.9 shows a complete match between the samples at serials 1 & 2. However, there is a remote possibility that the match could be by chance. The probability of match by chances is inversely proportional to the number of loci tested. Its calculation in a given population requires comprehensive knowledge of the allele frequencies at all the loci tested. An example of the profile frequency (probability of match by chance) is discussed in Table 15.5.

Sexual assault

In a sexual assault the DNA testing is done to demonstrate male DNA from the perpetrator in a sample collected from the victim. In a vaginal swab the victim's own DNA is found to be mixed with the DNA of the perpetrator. Less often there may be more than one perpetrator and in that case complex mixtures might be found. If the perpetrator has used condom his DNA would not be found in the vaginal secretions/swab. In that case the evidence material may have to be collected from other places like seminal stains on the victim's clothing or some other objects.

Example

The STR profiles of DNA extracted from high vaginal swab of a victim of sexual assault (sample 1) and the suspects (sample 2 & 3) are shown in Table 15.10.

Table 15.10. DNA profiles of victim (sample 1) and two suspects (samples 2 &3) in a case of sexual assault. The vaginal swab shows mixture of two DNAs. The minor component has a male genotype and matches exactly with that of the suspect 2 whereas the suspect 3 is excluded.

ID	D3S1358	D5S818	D7S820	D8S1179	D13S317	D18S51
1	16,17(16,18)	11,13(10,11)	11,11(10,12)	11,14(13,15)	11,11(11,11)	12,18(13,15)
2	16,18	10,11	10,12	13,15	11,11	13,15
3	15,16	10,13	11,13	13,15	10,13	12,15
ID	**D21S11**	**TPOX**	**FGA**	**Th01**	**Amgl XY**	**-**
1	29,30(29,32.2)	8,8(8,8)	23,24(22,23)	9.3,9.3(6,9)	XX(XY)	-
2	29,32.2	8,8	22,23	6,9	XY	-
3	28,31.2	8,8	21,22	6,9.3	XY	-

Resolving mixtures of DNA

Mixture of DNA from more than one source is typically encountered in investigation of sexual assault.

1. Complex mixtures are best resolved by analysis on genetic analyzer.
2. In a mixture of DNA the major component is usually of the victim.
3. The minor component may be from one or more individuals.
4. Identify all possible alleles.
5. A difficult issue is to differentiate between stutter products and the minor component of a mixture.
    a. The stutter products are usually of smaller size than the main PCR product and therefore would be seen moving ahead of the main product.

   b. The height of the stutter peak is usually less than 10% of the main peak.
   c. The mixture allele peak when present in the stutter peak area is considered significant when its height is more than 15% of the main peak.
   d. An allele peak seen in the larger size area is unlikely to be a stutter peak.
   e. A peak height less than 10% present in an area where stutter peaks are not expected is considered significant.
6. The alleles once identified are sorted out.
7. The victim's alleles are identified by aligning/comparing them with his/her own DNA extracted from blood.
8. The DNA samples of the suspect(s) are run and the results are aligned/compared to find any match.
9. In a sexual assault male DNA can also be demonstrated by Y-STR profiling.
10. Resolution of complex mixtures may require probabilistic genotyping computer software like STRmix and TrueAllele. These software use thousands of mathematical computations to generate statistical likelihoods in a mixture.

Paternity testing

At any autosomal locus an individual inherits one allele each from the biological parents (mother and father). The child inherits mitochondrial DNA only from the mother while Y-chromosome is transmitted from the father to the son. The exceptions to the rules include development of spontaneous mutations or chromosomal aneuploidies (trisomy or monosomy). Most parentage disputes are of paternity. However, occasionally maternity may also be questioned e.g. exchanged babies in a labour room.

In solving a paternity dispute a battery of STR markers are used. The step wise procedure includes:

1. Profiling of the alleged father and the child or the products of conception is done. The mother's profile is usually not required.
2. Of the two paternal alleles at each locus the allele transmitted to the child, called the "obligate allele", is identified.
3. If none of the alleles of the alleged father are present in the child at any of the loci tested the paternity is excluded. For example, if the child has genotype 14,18 and the father has 13,15 the paternity is excluded. However, keeping in view the possibility of spontaneous mutations it is recommended to test at more than one loci before excluding paternity.
4. Frequencies of the obligate paternal alleles are noted from the table of allele frequencies in the reference population.
5. The Paternity Index (PI) is calculated by dividing the prior probability with the frequency of the allele in question. The prior probability is the chance of transmitting the obligate allele by the alleged father to the child. If the alleged father is homozygous for the allele the prior probability is 1.0 and it is 0.5 if he is heterozygous.
6. Combined Paternity Index (CPI) is calculated by multiplying the PI values calculated at each locus.
7. The Probability of Paternity (POP) is calculated by the formula:

   CPI /CPI + (1-prior probability) x 100

Example

An example of the calculation of paternity index and the combined paternity index is shown in Table 15.11.

Identification in mass disasters

In mass disasters the DNA testing is primarily done to identify bodies that are beyond recognition. It may also be required to put the pieces of bodies together.

The process is done as follows:

1. The DNA profiles of the dead bodies or their remains are entered in a computer database.

2. The DNA profiles of the bodies is matched with those of the parents or the offsprings. If these are not available then DNA may be matched with that recovered from the personal effects of the victims.
3. The matched profiles are put together after sorting (Table 15.12).

Table 15.11. Calculation of Paternity Index and Probability of Paternity.

Locus	Genotype			Obligate Allele	Frequency	Paternity Index (PI) = Prior probability*/frequency
	Mother	Child	AI/Father			
D3S1358	14,15	14,18	15,18	18	0.078	0.5/0.078 = 6.41
D5S818	9,12	9,9	9,9	9	0.049	1.0/0.049 = 20.41
D7S820	8,11	8,9	9,11	9	0.108	0.5/0.108 = 4.63
D8S1179	10,12	10,15	11,15	15	0.211	0.5/0.211 = 2.37
D21S11	29,32.2	29,31.2	28,31.2	31.2	0.123	0.5/0.123 = 4.06
TPOX	9,11	9,11	8,11	11	0.333	0.5/0.333 = 1.50
TH01	8,9.3	9,9.3	6,9	9	0.250	0.5/0.250 = 2.00
FGA	21,21	20,21	20,24	20	0.093	0.5/0.093 = 5.38
Combined Paternity Index (CPI)			6.41 X 20.41 X 4.63 X 2.37 X 4.06 X 1.50 X 2.00 X 5.38 = 94,072 or 1 in 94,072			
Probability of Paternity			CPI/CPI + (1-prior probability) X 100 94072/(94072 + 0.5) X 100 = 99.999 %			

*If the father is heterozygous for the allele Prior Probability = 0.5
*If the father is homozygous for the allele Prior Probability = 1.0

Table. 15.12. Comparison of the DNA profiles of three dead bodies from a mass disaster and a father. The DNA profile of the "father" shares at least one allele at all fifteen loci with the DNA profile of the body number 7 and is therefore a proof of identity.

Sample	D2S1338	D3S1358	D5S818	D7S820	D8S1179	D13S317	D16S539	D18S51
3	20,25	14,15	8,10	8,12	14,16	10,11	10,10	13,16
7	20,23	15,15	9,11	8,12	13,16	11,12	10,10	14,17
16	22,23	14,14	10,11	8,11	13,16	10,12	10,12	15,17
Father	21,23	15,16	9,11	7,12	13,14	12,14	10,11	13,17

Sample	D19S433	D21S11	FGA	TH01	TPOX	vWA	CSF1PO	Amel
3	11,14	29,32.2	23,23	6,9.3	8,11	15,16	8,12	XY
7	12,14	28,32.2	24,24	6,8	8,11	16,16	8,11	XY
16	13,14	30,32.2	22,24	6,9	8,10	14,14	8,11	XY
Father	13,14	32.2,32.2	21,24	7,8	8,10	15,16	8,11	XY

Determining the ethnic origin of a person

Can DNA testing provide information on the ethnic origin of a person? The question may be asked when no clue is available to identify the perpetrator. There are some STR loci that have more significant differences in allele frequencies between major world populations like Caucasians, Blacks, Hispanics, Asians and East Asians etc. The differences are more marked in the frequency of the less common alleles than the common ones. An important requirement would be to have a prior knowledge of the allele frequencies in a representative sample of the target ethnic groups with ethnically pure individuals and not the ones with self-declared ethnicity. In this context analysis of SNP is considered more informative than the STR markers.

In tribal populations with high frequency of consanguinity and marriage within the same tribe founder effect and genetic drift might cause an unexpectedly higher or lower frequency of alleles than in the rest of the population.

OmniPop 200.1 software is available on the internet free of cost. It is basically a database of the STR allele frequencies published in over 200 studies in the major world populations. By entering the CODIS STR profile of an individual it provides statistical information on finding a similar profile in various world populations. Similar software can be developed for one's own requirements if comprehensive knowledge of the allele frequencies in the smaller target groups and populations is available.

Collection storage and dispatch of samples for forensic DNA testing

1. Collection and storage of samples
    a. DNA can be extracted from any biological material containing nucleated cells. Since DNA is susceptible to degradation by nucleases from the cells and bacterial contamination the sample should remain as clean and free of bacterial contamination as possible.
    b. DNA is usually extracted from whole or dried blood (stains), buccal smear, wet or dried semen and other body secretions, hair roots, soft tissues, fresh and dried bones etc.

c. The blood stains or body secretions should first be air dried and then kept in a paper envelop rather than plastic bag.

d. Fresh blood: 2-3 ml of venous blood should be collected in EDTA. The sample should be stored at 4°C if there is delay in dispatch.

e. Buccal smear obtained on a clean swab on stick contains mucosal cells to yield sufficient quantity of good quality DNA. The swab can also be used to obtain DNA from a dead body. The swab should be rubbed several times over clean part of buccal mucosa. It may be air dried before storage or dispatch.

f. Semen and other body fluids containing nucleated cells are another good source of DNA. The stained clothes or objects should be air dried and treated as blood stains.

g. Hair that come out with roots can be used to extract DNA. Sufficient DNA can be extracted from 2-3 hair with roots. Hair shafts can be used to extract mitochondrial DNA.

h. Soft tissues are a rich source of DNA. Skin is an easily accessible tissue that can be used to collect DNA from a dead body. A full thickness piece of skin measuring 2x2 cm from a clean part of the body or its remains can be taken. If skin is not available any other available soft tissue should be collected. The soft tissue samples provide an excellent medium for bacterial growth. In a putrefied or heavily contaminated soft tissue sample the yield as well as the quality of DNA can be very poor. The soft tissue samples can be stored as such in a suitable container at -20°C for several days. The sample may be transported in normal saline. Do not put the samples for DNA testing in formalin at any stage.

i. Bones can also be used as a source of DNA. But extraction of DNA from a bone is difficult therefore the yield and the quality of DNA is also variable. DNA is best collected from compact bones like humerus or femur. The spongy bones with thin cortex are usually contaminated with mud etc. that may inhibit PCR.

2. Dispatch of samples

    a. The samples should be properly labeled and sealed.

    b. The request form should contain all available details of the individual(s) to be tested along with a brief summary of the incident and what is required to be solved by the DNA test.

    c. Delay in transport can affect the quality of the sample. Avoid exposing the sample to extreme heat or direct sunlight.

3. Chain of custody

    a. Forensic DNA testing is done for medico-legal purpose therefore it is essential to maintain the chain of custody. A record of the individuals receiving and handing over the samples must be maintained at all steps as they may be called by the court as a witness.

Bibliography

1. Almeida M, Betancor E, Fregel R, Suárez NM, Pestano J (2011) Efficient DNA extraction from hair shafts. Forensic Science International: Genetics Supplement Series 3: e319-320.
2. Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., and Rolf, B. (1998). Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. Am J Hum Genet 62: 1408-1415.
3. Elston RC (1986) Probability and paternity testing. Am J Hum Genet 39: 112-122.
4. Kaiser L and Sever G (1983) Paternity testing: I. Calculation of paternity indexes. Am J Med Genet 15: 323-329.
5. Short Tandem Repeat DNA Internet DataBase, http://www.cstl.nist.gov/strbase/Smith RN (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. *Biotechniques* 18: 122-128.

16

Study of Donor Chimerism

The outcome of a stem cell transplant can be complete donor chimerism (100% donor cells) or mixed donor chimerism of varying proportions of donor and recipient cells. Another type of chimerism called split chimerism may also exist in which one or more whole lineage is of recipient and the other of donor in origin.

The haematopoietic cells of donor origin can be detected in the host. The test may be done on peripheral blood, bone marrow or lineage specific cells e.g., T cells, B cells and granulocytes. Peripheral blood is equally sensitive in detection of chimerism than bone marrow. The study of donor chimerism may be done to know:

- Whether the donor engraftment has occurred or not?
- Whether there is mixed chimerism? If present then how much?

- If there is mixed chimerism then which lineages are mixed and which are fully donor?
- Whether there is chimerism in the lymphoid and the myeloid compartments?

Applications

- Myeloablative Stem Cell Transplant
  - Donor engraftment
  - Graft rejection
  - Prediction of GVHD
- Non-Myeloablative Stem Cell Transplant (Mini transplant)
- Relapse prediction

Techniques

Donor chimerism is usually tested by cytogenetics/FISH, real time PCR or STR analysis. STR analysis is basically an extension of the analysis of DNA mixtures discussed in Chapter 8 and 15.

Samples required

1. Recipient's pre-transplant blood sample in EDTA. If this is not available recipient's buccal mucosal cells, skin biopsy or hair roots may be collected on a stick swab to represent the pre-transplant status. Care should be taken to avoid contamination of the swab with recipient's blood.
2. Donor blood in EDTA.
3. Recipient's post transplant blood or bone marrow sample.

Procedure

1. Extract DNA from the three samples.
2. The STR profiling may be done by manual method or by genetic analyzer (Chapter 8).

3. Out of the many STR markers "informative" marker(s) are chosen. The STR marker is called "informative" when its alleles can distinguish between the recipient and the donor DNA. For example, if at the D21S11 locus the recipient has alleles 28,29 and the donor has 28,31 the marker is informative because allele 29 and 31 are exclusive for the recipient and the donor respectively.
4. In the presence of complete donor chimerism the recipient's post transplant DNA shows the donor's genotype. In mixed chimerism mixture of recipient and donor genotypes is seen (Figure 16.1).
5. In the manual method quantitative estimation of amplified products is done by densitometry of polyacrylamide gels (Figure 16.2).
6. If genetic analyzer is used for STR genotyping the analysis is done by measuring the peak heights and area under the curve.

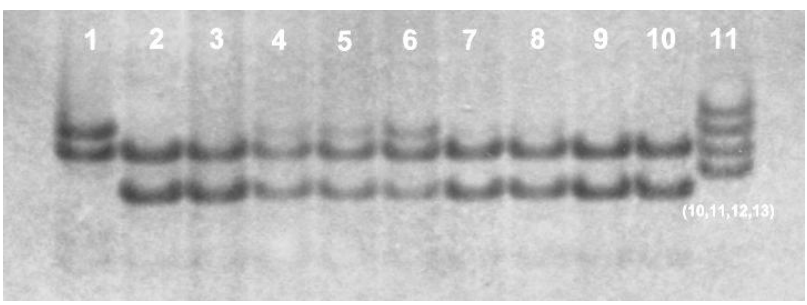


Figure 16.1. PAGE of PCR amplification at D5S818 locus. Lane 11 shows allelic ladder with alleles 10-13. The lane 1 shows recipient's pre-transplant sample (alleles 11,12) and lane 2 shows donor sample (alleles 9,11). Lane 3-10 show serially collected recipient's post-transplant samples. Lanes 4-6 show gradually appearing recipient's exclusive allele (12) and a gradually decreasing strength of the donor' exclusive allele (9) indicating graft failure. The patient received an infusion of donor lymphocytes that resulted in disappearance of the recipient's allele (lanes 7-10).

Calculation of donor and recipient component

Calculation of the donor and recipient components is usually done by analyzing the samples on a genetic analyzer and then measuring the peak areas. A cost-effective alternate is by doing PAGE followed by densitometry. Figures 16.2 and 16.3 show an example of the calculation of donor chimerism by PAGE and its densitometry at the D5S818 locus.

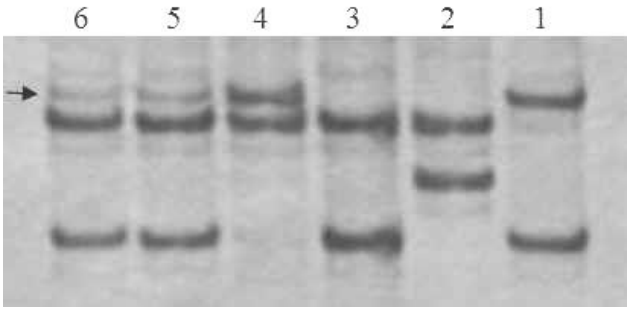


Figure 16.2. PAGE of STR amplification at D5S818 locus. Lanes 1-4 show the samples of father, mother, donor and recipient (pre-transplant) respectively. Lanes 5 & 6 show the recipient's post-transplant samples in duplicate. The sample shows mixed donor chimerism represented by reappearance of the recipient's exclusive allele (arrow).

Real time PCR for assessing donor chimerism

As discussed in Chapter 5 end point analysis of PCR products by gel electrophoresis, including analysis on a genetic analyzer, is not very good for quantitative assessment of DNA. Quantitative assessment of donor chimerism by real time PCR can give a more accurate measurement of the donor or recipient components. SNP genotyping by real time PCR using TaqMan® probe assay has been used for this purpose.

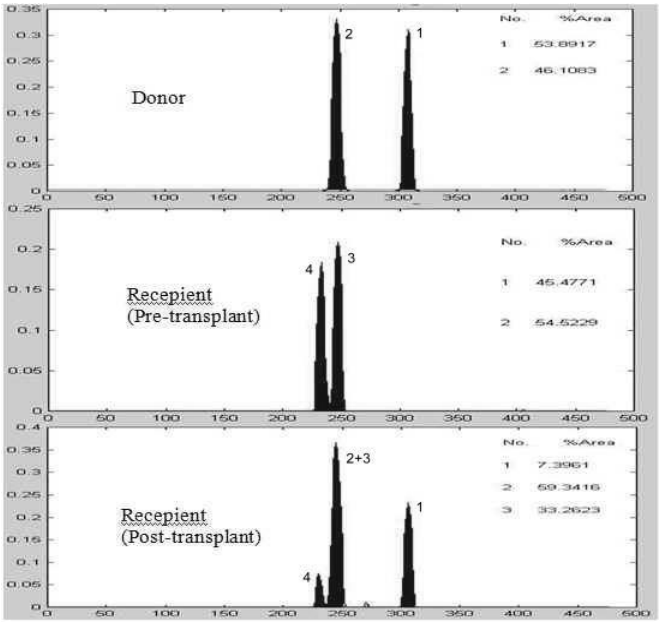


Figure 16.3. Densitometric recording of STR amplified products run on polyacrylamide gel shown in Fig. 16.2. The recipient's post transplant sample shows mixed donor chimerism. The recipient's pre transplant allele peak 4 represents 7.3%. The total recipient's component is calculated by doubling the allele peak 4 component (3+4) i.e. 14.6%.

Bibliography

1. Alizadeh M, Bernard M, Danic B, Dauriac C, Birebent B, Lapart C, Lamy T, Le Prise´ P, Beauplet A, Bories D, Semana G, Quelvennec E (2002) Quantitative assessment of hematopoietic chimerism after bone marrow transplantation by real-time quantitative polymerase chain reaction. Blood 99: 4618-4625.
2. Antin JH, Childs R, Filipovich AH, Giralt S, Mackinnon S, Spitzer T, Weisdorf D (2001) Establishment of complete and mixed donor chimerism after allogeneic lymphohematopoietic transplantation: Recommendations from a workshop at the 2001 tandem meetings. Biology of Blood and Marrow Transplantation 7: 473-485.
3. Scharf SJ, Smith AG, Hansen JA, McFarland C, Erlich HA (1995) Quantitative determination of bone marrow transplant engraftment using fluorescent polymerase chain reaction primers for human identity markers. Blood 85: 1954-1963.
4. Thiede C, Lion T (2001) Quantitative analysis of chimerism after allogeneic stem cell transplantation using multiplex PCR amplification of short tandem repeat markers and fluorescence detection. Leukemia 15: 303–306.

17

Molecular Pathology in the Developing Countries

Pathology has traditionally been a descriptive science. Pathologists take pride in giving elaborate description of the gross and microscopic picture of a pathological lesion. With the availability of techniques to analyze proteins and more recently the nucleic acids we now know that at the submicroscopic level most pathological processes are caused by changes in the proteins or the genetic code. For example the clonal proliferation of seemingly heterogeneous population of granulocytic cells in a patient of chronic myeloid leukaemia can also be recognized by genetic analysis of the *bcr-abl* gene. The genetic changes mostly remain beyond the visible range of a microscope. Understanding of the molecular basis of disease provides a vast potential for diagnosis by using molecular analytical techniques. Consequently molecular pathology is now recognized as a distinct entity.

The proteins as markers of disease are in use for the past several decades. The use of molecular genetic techniques in diagnostic pathology is a relatively new development. The past couple of decades are dominated by an ever increasing use of PCR and its applications in the diagnosis of a wide range of diseases. PCR has made DNA analysis so simple that even small diagnostic labs can routinely offer sophisticated molecular genetic testing.

The major areas where molecular genetic techniques have made an impact include inherited, neoplastic and infectious disorders. In addition, it has also provided a strong basis for DNA profiling in forensic pathology and tissue typing for organ transplantation.

The Developing Country Scenario

There is a general feeling that the molecular genetic techniques are expensive, complex, and difficult to practice. Therefore in a developing country like Pakistan the full potential of their use in diagnostic pathology has remained limited. Unfortunately the development of newer and more complex techniques like droplet PCR, microarrays and NGS have further increased the gap between technology and its utility. In this challenging scenario the applications of molecular genetic techniques in a resource constrained setting need to focus on:

1. Identification of high priority areas.
2. Development of low cost PCR applications.
3. Development of high tech equipment at low cost.
4. Outsourcing of complex molecular genetic testing requiring expensive equipment.

High priority areas

Genetic disorders

There are more than 5000 known genetic disorders. Majority of these disorders can be diagnosed by molecular methods. Most genetic disorders are either not treatable or their treatment is very expensive. There is a consensus of opinion that the births of children with a genetic disorder can be prevented by prenatal diagnosis or premarital carrier screening and genetic counselling.

Genetic haemoglobin disorders are typical example of a success story. These disorders are found at high frequency in a large number of countries in Africa, Europe and Asia. Many developed countries in the Mediterranean region have brought down the birth incidence of thalassaemia to almost zero. Prenatal diagnosis with very simple molecular genetic methods has played a key role in the prevention of these disorders. Thalassaemia and sickle haemoglobin continue to pose a severe health burden in most developing countries of the region including Pakistan. As the incidence of common infectious diseases comes under control a high mortality due to genetic disorders is likely to become increasingly obvious. The developing countries need to follow the example of the Mediterranean countries. They should focus on the development of cost effective methods for prenatal diagnosis of common genetic disorders and their service delivery. Similarly there are many other relatively less common or uncommon genetic disorders for which cost effective molecular methods can be developed.

Infectious diseases

Tuberculosis, hepatitis and malaria together constitute nearly 50% of the health burden due to infectious diseases in many developing countries. Cost effective molecular methods can be developed for rapid diagnosis and follow-up of several common and uncommon infections rampant in these populations.

Neoplastic disorders

At the subcellular level cancer is almost certainly a genetic disorder. The altered DNA carries a potential for diagnosis by PCR amplification. PCR being exceptionally sensitive can be used for the initial diagnosis of cancer, predict its prognosis or monitor the disease progress and response to treatment if a cancer specific DNA marker is identified. PCR can also be used to demonstrate the association of some malignancies and viruses e.g. human papilloma virus and cervical cancer etc.

The availability of NGS applications like whole exome sequencing and custom designed gene panels have provided new opportunities for looking at the genetic makeup of a large number of cancers. The extensive gene profiling is done to tailor the treatment regimen for an individual patient. This leaves the countries

with resource constraints in a dilemma as to how they should cope with the ever expanding knowledge about the genetic markers in cancer and the limited choices available to treat such patients.

Short of the high tech genetic profiling a large number of very useful and simple PCR applications to look for the gene rearrangements and cancer specific mutations in the common malignancies can be developed locally.

Development of low cost PCR applications

Besides technical difficulties and complexities cost is a major limiting factor in the widespread use of molecular genetic testing. Most diagnostic kits for genetic testing are imported from abroad. There is a need to focus on the development of cost-effective diagnostic kits locally.

Development of high tech equipment at low cost

Most of the equipment used in molecular genetic testing is also imported from abroad. Thermal cycler is a key instrument required for this purpose. A few locally developed versions of thermal cycler and other related equipment are available in Pakistan. These instruments are cheap and come with a complete backup support that is often lacking for the imported equipment. Another useful option could be to buy used equipment that are also available at a very reasonable price in the local market.

Outsourcing for complex genetic testing

The high tech genetic testing like NGS is evolving at a very rapid pace and the instruments may become obsolete in a relatively short period of time. Secondly the maintenance of these instruments is also demanding. Therefore in the countries with resource constraints it may not be advisable to invest heavily in such equipment. There are many good quality NGS service providers whose services may be used. The DNA samples can be submitted to a service provider who can analyze these and return the data that may be analyzed by a suitable software.

18

Sequence Variant Nomenclature

The new system for sequence variant nomenclature was proposed in 2000 by the Human Genome Variation Society (HGVS). It has been widely adopted and has become an internationally accepted standard method of describing sequence variations. The recommendations are currently commissioned through a Sequence Variant Description Working Group (SVD-WG). Any modifications and extensions go through the SVD-WG after a community consultation step. The HGVS recommendations are designed to be meaningful, stable, memorable, and unequivocal. The sequence variants are reported with a consistent and unambiguous method. The uniformity of nomenclature is also essential for sharing of information on the analysis of genes.

The traditional terms “mutation” and “polymorphism” have been replaced by a common terminology “variant” because it is less confusing and carries a positive value in genetic counselling and other discussions.

All sequence variants are reported with reference to a stable and standard sequences of DNA (RefSeq). For reporting all clinically significant variants it is mandatory to use Locus Reference Genomic (LRG) reference sequences.

Reference Sequence (RefSeq)

The RefSeq project at the National Center for Biotechnology Information (NCBI) is a publicly available database of annotated genomic, transcript, and protein sequence records (http://www.ncbi.nlm.nih.gov/refseq/). The project controls the submitted data against a combination of computation, manual curation, and collaboration to produce a standard set of stable, non-redundant reference sequences. The database currently represents sequences from more than 55,000 organisms ranging from a single record to complete genomes.

Locus Reference Genomic (LRG)

LRG is a collection of manually curated records of reference sequences that have some clinical significance. These reference sequences are designed specifically for reporting sequence variants with clinical implications. The LRG project was created to avoid the problems of multiple sequences for a given locus as well as confusion over versions.

The LRG database aims to record every locus with clinical significance. LRG sequences are selected in collaboration with the diagnostic and research communities. At present 1315 LRGs have been created, of which 1163 are publicly available.

Each LRG record is allotted a unique identifier. For example the LRG_199 and LRG_1232 are allotted to DMD (Duchene Muscular Dystrophy gene) and HBB (Haemoglobin β-globin gene) respectively.

HGNC Gene Symbols

In order to bring uniformity in the gene symbols Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) has formulated guidelines for creating gene symbols.

The symbol should be a unique short representation or abbreviation of the descriptive gene name. It should only contain Latin letters and Arabic numerals. It should not contain any reference to the species, for example, "H/h" for human etc. All characters of the symbol are written on the same line without any superscripts or subscripts. Any previously used Greek or Roman letters etc. are changed to their Latin or Arabic counterparts.

Examples

HBB: Hemoglobin β-gene

HBA: Hemoglobin α-gene

INS: Insulin gene

GLB: Galactosidase β-gene

DMD: Duchene Muscular Dystrophy gene

Prefixes for reference sequences

The types and the prefixes of the reference sequences are as follows:

g. Genomic DNA:	Starts with the first nucleotide of the reference sequence.
c. Coding DNA:	Starts with the first nucleotide of the initiation codon of the reference sequence.
n. Non-coding DNA:	Starts with the first nucleotide of the non-coding DNA reference sequence.
m. Mitochondrial DNA:	Starts with the first nucleotide of the mitochondrial DNA reference sequence.
r. RNA	First nucleotide of the translation start codon of the RNA reference sequence or first nucleotide of the noncoding RNA reference sequence.

p. Protein First amino acid of the protein reference sequence.

Choice of reference sequence

The most commonly used reference sequences include genomic DNA (g), and the coding DNA (c). When genomic DNA is sequenced it should preferably be reported against a genomic reference sequence (g). In clinical diagnostics, reporting based on a coding DNA reference sequence (c) is popular because:

1. It gives a clear indication whether the variant is in the exons or the introns.
2. It tells whether it is 5' of the initiation codon (ATG) or 3' of the stop codon.
3. Dividing the nucleotide number by 3 can tell about the number of the amino acid residue affected.

Variant nomenclature

The HGVS variant nomenclature consists of the following components:

1. The nomenclature may start with the Reference Sequence accession number, LRG_ID, or the HGNC Gene Symbol e.g. NG_000007.3, LRG_1232, or HBB.
2. Give the reference sequence prefix e.g. for genomic DNA HBB:g. and for coding DNA HBB:c. etc.
3. The variant position is given as follows:
    a. For genomic DNA reference sequence give position of the variant with reference to the first nucleotide of the reference genomic DNA e.g. NG_000007.3:g.70595A>T
    b. For the coding DNA reference sequence give position of the variant with reference to the first base of the initiation codon (ATG) e.g. HBB:c.1A>T or LRG_1232:c.1A>T
    c. If the variant position is 5' to the initiation codon (ATG) a minus (-) sign is used e.g. HBB:c.-50
    d. If the variant position is 3' to the initiation codon (ATG) the position is given without any sign e.g. HBB:c.92

e. If the variant is located in the intron count from the last base of the exon and indicate the position with a + sign e.g. position 5 of the intron would be written as HBB:c.92+5

f. If the variant is located at the 3' end of an intron count backwards from the first nucleotide of the next exon and give position with a minus (-) sign e.g. HBB:c.93-1

g. The stop codon is represented by * sign. If the variant position is 3' to stop codon indicate the position after * sign e.g. HBB:c.*108

4. Define the variant as follows (Table 18.1):

   a. Substitution (>):HBB:c.47G>A; HBB:c.92+5G>C

   b. Deletion (del): HBB:c.396_397delGA; HBB:c.68_74delAAGTTGG
   Large deletions are usually reported with the prefix g. e.g. NG_000007.3:g.71609_72227del619

   c. Insertion (ins): HBB:c.27_28insG

   d. Deletion-insertion (delins): HBB:c.74delinsCAC
   HBB:c.301_302delCCinsTCTGAGAACTT

Definitions

The sequence variants are classified as substitution, deletion, insertion, inversion, duplication, deletion-insertion, and conversion. (Table 18.1).

Molecular Path Reports

In addition to the general requirements of a lab report a diagnostic molecular pathology report should include:

- Date of report (and time, if appropriate)
- Referring physician or lab
- Indications for testing (reason for referral) pertinent clinical history, ethnicity/race, pedigree diagram and/or family history, previous molecular/genetic studies, and other relevant clinic-pathologic findings.
- Type of sample
- Demographic information Accession number, and specimen number from referring laboratory.

- Test performed (PCR, RT-PCR of RNA, Q-PCR, linkage analysis, STR analysis, in situ hybridization, gene dosage array, RNA expression array, and sequencing etc.
- Test procedure: Pertinent details of procedure, for example analyte-specific reagent or kit version and manufacturer, instrument type. Disclaimer on non-FDA approved tests in which a commercial analyte-specific reagent was used.
- Reference Sequence (RefSeq) preferably LRG number)
- Name of analyte tested such as gene, locus, or genetic defect; use HUGO-approved gene nomenclature. Use standardized gene nomenclature and standard units of measure
- Test result
- Opinion
- Comments: Significance of the result in general or in relation to this patient. Residual risk of disease (or carrier status). Answer specific questions posed by the requesting clinician (eg, rule out CML etc.).
- Cite peer-reviewed medical literature or reliable Web sites on the assay and its clinical utility.
- Recommend additional measures (further testing, genetic counseling etc.).
- Condition of specimen that may limit adequacy of testing (sample received thawed, partially degraded DNA etc.).
- Reason specimen rejected or not processed to completion
- Disposition of residual sample (eg, sample repository) Chain of custody documentation, if needed
- Signature and printed name of reporting physician, for any test having a physician interpretation. Signature of lab director or designee when interpretation is performed (Reports may be signed electronically).

Table 18.1. Definitions of various types of sequence variants.

Nomenclature	Example	Definition
Substitution (>)	g.1318G>T HBB:c.47G>A; HBB:c.92+5G>C	One nucleotide replaced by another nucleotide
Deletion (del)	g.3661_3706del HBB:c.396_397delGA; HBB:c.68_74delAAGTTGG	One or more nucleotides are deleted
Insertion (ins)	g.7339_7340insTAGG HBB:c.27_28insG	One or more nucleotides are inserted but is/are not a copy of a sequence immediately 5'
Inversion (inv)	g.495_499inv	More than one nucleotide replaces the original sequence and is the reverse-complement of the original sequence (e.g., CTCGA to TCGAG)
Duplication (dup)	g.3661_3706dup	A copy of one or more nucleotides are inserted directly 3' of the original copy of that sequence
Deletion-insertion (delins/indel)	g.112_117delinsTG	One or more nucleotides are replaced by one or more other nucleotides and which is not a substitution, inversion, or conversion
Conversion (con)	g.333_590con1844_2101	A range of nucleotides replacing the original sequence are a copy of a sequence from another site in the genome

Bibliography

1. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PE, Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organisation (HUGO). HGVS recommendations for the description of sequence variants: 2016 update. Human mutation. 2016 Jun;37(6):564-9.
2. Gulley ML, Braziel RM, Halling KC, Hsi ED, Kant JA, Nikiforova MN, Nowak JA, Ogino S, Oliveira A, Polesky HF, Silverman L. Clinical laboratory reports in molecular pathology. Archives of pathology & laboratory medicine. 2007 Jun;131(6):852-63.
3. MacArthur JA, Morales J, Tully RE, Astashyn A, Gil L, Bruford EA, Larsson P, Flicek P, Dalgleish R, Maglott DR, Cunningham F. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. Nucleic acids research. 2013 Nov 26;42(D1):D873-8.

4. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research. 2015 Nov 8;44(D1):D733-45.
5. Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S. Guidelines for human gene nomenclature. Genomics. 2002 Apr 30;79(4):464-70.

Glossary of Terminologies

Allele: one of the two or more versions of a genetic sequence at a particular location in the genome.

Allele frequency: proportion of an allele among the individuals in a population.

Allelic ladder: mixture of amplified DNA products of known sizes used for size comparison in gel electrophoresis of DNA.

Amplicon sequencing: sequencing of a specific region of the genome. Also called targeted NGS.

Amplification: increase in the number of copies of a specific DNA fragment.

Amplification Refractory Mutation System (ARMS): application of PCR for detection of point mutations and SNP in which the DNA is amplified by allele specific primers.

Annealing: process of alignment/attachment of a short piece of DNA (primer or probe) to its complementary region on the target DNA.

Autosome: chromosome not involved in sex determination. The diploid human genome consists of 46 chromosomes including 22 pairs of autosomes, and one pair of sex chromosomes.

Base pair (bp): two nitrogenous bases paired together in double-stranded DNA by weak bonds (adenine with thymine and guanine with cytosine).

Base sequence: order of nucleotides in a DNA molecule.

BLAST: Basic Local Alignment Search Tool available at the NCBI website.

cDNA: DNA copy of RNA.

Chromosome: coiled up molecule of DNA with a supporting protein structure that carries the genetic information which is physically transmitted from one generation to the next.

Chromosome region: defined as a genomic region that has been associated with a particular syndrome or phenotype, particularly when several genes within it may be involved in the phenotype.

CODIS (COmbined DNA Index System): database created by FBI that stores the DNA profiles of convicted offenders and the biological material found at crime scenes.

Coefficient of inbreeding: probability that an individual receives at a given locus two genes that are identical by descent (copies of a single gene carried by a common ancestor).

Combined Paternity Index (CPI): calculated by multiplying the PI values calculated at each locus.

Complementary sequences: DNA sequences that form a double-stranded structure by matching base pairs e.g., C-A-T-G is complementary to G-T-A-C.

Compound heterozygote: two different alleles at the same locus on the maternal and the paternal chromosomes (homologous chromosomes).

Conversion (con): a range of nucleotides replacing the original sequence are a copy of a sequence from another site in the genome.

Cross-homology: two DNA (or protein) sequences are said to be homologous if they are sufficiently similar to suggest that they are derived from a common ancestral gene.

Ct (Cycle threshold): first significant increase in the amount of amplified product as seen in the real time graphic recording of a PCR.

Cycle sequencing: technique in which 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”.

Degraded DNA: DNA broken into smaller fragments by shearing force or bacterial enzymes.

Deletion (del): one or more nucleotides are deleted.

Deletion insertion (del/ins): one or more nucleotides are replaced by one or more other nucleotides and which is not a substitution, inversion, or conversion.

Denaturation: physical process of splitting the complementary strands of DNA to form single strands e.g., by heat or chemicals.

De-novo sequencing: NGS done without using any reference sequence.

Deoxyribonucleic acid (DNA): genetic material of organisms, usually double-stranded; that contains deoxyribose, and four nucleobases.

DNA sequence: order of the base pairs in a DNA molecule.

DNAse: enzyme, usually of bacterial origin, that can degrade DNA into smaller pieces.

Donor chimerism: existence of two genetically different tissues in the same individual for example after receiving organ transplant.

Duplication (dup): a copy of one or more nucleotides are inserted directly 3' of the original copy of that sequence.

ΔRn: difference between Rn^+ and R^-.

Electrophoresis: technique of separating charged molecules by electric current.

Entrez: an integrated, text-based search engine and retrieval system for the NCBI website.

Exon: portion of a gene that encodes amino acids.

FASTA format: text-based format for writing the nucleotide or peptide sequences using single letter codes.

Forward and reverse strands: two strands of a double-stranded DNA molecule each runs in opposite direction.

FRET probe: a pair of probes that anneal to the inner region of the target DNA in head to tail configuration. Two florescent dyes are attached at the 3' and 5' ends of the upstream and the downstream probes respectively. On annealing the close proximity of the two fluorescent dyes ensures transfer of energy from the former to the later that in turn emits light of yet another wavelength.

Gene: an ordered sequence of nucleotides located at particular position on particular chromosome and encodes specific functional product like protein or RNA.

Gene profiling: DNA microarrays and more recently NGS has enabled analysis of a very large number of sequence variants in a wide variety of cancers.

Gel: semisolid matrix used in electrophoresis to separate molecules e.g. agarose and acrylamide etc.

Gene: basic unit of heredity composed of nucleotides on a chromosome.

GenBank: an annotated collection of all publicly available DNA sequences in the form of database at the National Institute of Health (NIH), Bethesda USA.

GenBank accession no: a unique number assigned to a new sequence received at the NCBI.

Gene frequency: relative frequency of a particular allele in a population.

Genome: total genetic material found in a cell; its size is generally given as number of base pairs.

Genotype: an organism's complete collection of genes. The term also refers to the two alleles inherited for a particular gene.

Gene variant: any heritable change in DNA sequence that may or may not be associated with harmful effect on the phenotype.

GRCh37 & 38: Genome Reference Consortium Human genome build 37 & 38. GRCh38 is the standard reference assembly sequence used by NCBI.

Heterozygote: different alleles at the same locus on the maternal and the paternal chromosomes (homologous chromosomes).

HGNC gene symbols: In order to bring uniformity in the gene symbols Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) has formulated guidelines for creating gene symbols.

Homozygote: similar alleles at the same locus on the maternal and the paternal chromosomes (homologous chromosomes).

Insertion (ins): one or more nucleotides inserted but is/are not a copy of a sequence immediately 5'.

Intron: intervening sequence of DNA between two coding regions (exons).

Inversion (inv): more than one nucleotide replaces the original sequence and is the reverse-complement of the original sequence (e.g., CTCGA to TCGAG).

Linkage: proximity of a marker and a gene on a chromosome; when the marker and the linked gene are close together the probability of their separation during meiotic crossover is low and the two are inherited together as a unit.

Liquid biopsy: nucleic acids extracted from the circulating tumour cells analyzed by molecular technique can be useful in estimation of prognostic information,

real-time monitoring of therapies, and research related to therapeutic targets, resistance mechanisms, and development of metastasis in cancer patients.

Locus: specific physical location on a chromosome.

A locus is a point in the genome which can be mapped by some means for example a marker. It refers to a map position and is not synonym for gene. A single gene may have several loci within it (each defined by different markers).

Locus Reference Genomic (LRG): a collection of manually curated records of reference sequences that have some clinical significance.

Mini STRs: STR analysis in a degraded DNA sample can be improved by bringing the forward and the reverse amplification primers flanking the repeat units to the nearest possible distance from the repeat units.

Mitochondrial DNA : each mitochondrion contains 2-10 copies of a circular piece of DNA, 16,569 base pairs in length. It is inherited from the mother.

MLPA: (Multiplex ligation-dependent probe amplification) is a very sensitive method for detection of very small to large deletions, insertions, and duplications.

Molecular beacon probe: short sequence of DNA complementary to the internal region of the PCR target DNA. The un-annealed probe assumes the shape of a hairpin due to the presence of short DNA sequences complementary to each other. The later also ensures quenching of the attached fluorescent dye and the quencher. On annealing the hairpin structure opens up relieving the fluorescent dye from the effect of quencher.

Multiplex PCR: PCR done with more than two primers.

Mutation: now called variant, is a heritable change in DNA sequence that is associated with a harmful effect on the phenotype.

Next generation sequencing (NGS): new technique of extremely high throughput genomic sequencing involving template preparation, sequencing, imaging, and data analysis.

NGS bridge amplification: templates anchoured to a surface are amplified through bridge amplification that results in spatially distinct clusters of billions of molecules.

NGS cluster generation: the library of templates is captured through oligonucleotides complementary to the adapter molecules bound to either a solid surface, or individual agarose beads that are later converted to form an oil emulsion.

NGS coverage: average number of reads that align to, or "cover," known reference bases is called NGS coverage. For example, 50 Mb of sequence data for a 5 Mb genome would be 10X coverage.

NGS library preparation: NGS library or template preparation involves controlled ultrasonic shredding of genomic DNA or cDNA into smaller pieces of 90-500 bp length depending on the type of application. The DNA fragments are ligated with synthetic short pieces of DNA called adapter molecules. The adapter ligated fragments are then PCR amplified and purified either by gel electrophoresis or other methods.

Non-Invasive Prenatal Testing (NIPT): minute quantities of cell-free DNA (cfDNA) of fetal origin tested in a mother in early pregnancy.

Nucleotide: basic building block of DNA comprising phosphate, ribose or deoxyribose, and a purine or pyrimidine base.

Obligate allele: of the two parental alleles at each locus the allele transmitted to a child.

Oligo dT: primers that are designed to anneal with the poly-A tail attached at the 3’ end of mRNA molecules.

Oligonucleotide: short, typically synthetic, polymer of nucleotides.

Oncogene: gene which is associated with cancer.

Paternity Index (PI): calculated by dividing the prior probability with the frequency of the allele in question.

Phenotype: observable traits of an individual such as height, eye color, and blood group etc.

Point of care PCR: development of on-chip PCR allowing multiplex analysis and high throughput screening for infectious diseases in 15-30 minutes.

Point variant: alteration in DNA sequence caused by a single-nucleotide base change, insertion, or deletion.

Polymerase chain reaction (PCR): in-vitro technique in which a DNA molecule can be amplified in sequential steps of denaturation, primer annealing and extension to produce millions of copies of desired length.

Polymorphism: old terminology (now called variant) for inheritable change in DNA sequence that is present in more than 1% individuals of a population. Unlike mutation polymorphism is not associated with any harmful effect on the phenotype.

Pre-implantation genetic diagnosis (PGD): done to know the genetic status of an embryo prior to its implantation involves in-vitro fertilization, embryo biopsy, genetic analysis and selective implantation.

Primer: short sequence of single-stranded DNA of a specific base composition to which new deoxyribonucleotides can be added by DNA polymerase.

Prior probability: probability of transmitting the obligate allele to the offspring. It is 1.0 (100%) when the parent is homozygous and 0.5 (50%) when the parent is heterozygous for an allele.

Probability of paternity (POP): is calculated by the formula: CPI/CPI + (1-prior probability) x 100.

Probe: short sequence of single-stranded DNA or RNA of a specific base composition labeled with radioactive, chemi-lumiscent or fluorescent dye.

Proteinase-K: broad spectrum protein cleaving enzyme derived from extracts of fungus *Engyodontium album* and commonly used in nucleic acid extractions.

Quencher: compound that has the ability to suppress the light emitting from a fluorescent dye. The quencher may be a high energy fluorescent dye like rhodamine (TAMRA) or a non-fluorescent chemical quencher e.g., DABCYL and black hole quenchers (BHQ).

Random hexamers: mixture of all possible combinations of six nucleotide pieces of DNA.

Real time PCR: in-vitro technique of DNA amplification in which the process is monitored by evaluating amplification at the end of every cycle. The amplification is monitored by fluorescent dyes.

Reference gene: in the gene expression studies a reference gene is required for normalization of the expression of the gene under study.

Reference sequence (RefSeq): project at the National Center for Biotechnology Information (NCBI) is a publicly available database of annotated genomic, transcript, and protein sequence records (http://www.ncbi.nlm.nih.gov/refseq/).

Restriction enzyme: enzyme of bacterial origin that can cut DNA at pre-defined sequences.

Restriction fragment length polymorphism (RFLP): variation between individuals in DNA fragment sizes cut by specific restriction enzymes.

Reverse transcriptase: enzyme used for synthesizing DNA copy of RNA.

Rn: indicator of the amount of fluorescence generated in a PCR. Rn of a test reaction and a non- target control (NTC) are called Rn^+ and Rn^- respectively.

RNase: enzyme that can degrade RNA into smaller pieces. It is present in all organisms.

Sequencing: in-vitro technique to determine the order of nucleotides in DNA or RNA molecule or amino acids in a protein.

Short tandem repeats (STR): tandemly repeated short stretches (units) of DNA (2-6bp) that are randomly distributed in the genome. The number of repeat units vary between individuals of a population.

Single nucleotide variant (SNV): previously called single nucleotide polymorphism is a naturally occurring variation of DNA present after every 300-500 nucleotides in the genome.

Stutter bands: in PCR amplification of STRs the elongating DNA strand can slip one or more steps forward resulting in formation of end products that are one or more repeat units smaller in length.

Substitution (>): one nucleotide replaced by another nucleotide.

SYBR green: dye that binds to minor groove of double stranded DNA. It gives little fluorescence when in solution but emits a strong fluorescent signal after binding with double-stranded DNA. It can be used to monitor real time PCR.

Tm: temperature at which 50% of the primer/probe is annealed to its target.

TaqMan® probe: short sequence of DNA complementary to the internal region of the PCR target DNA. The 5' end of the probe is labeled with a fluorescent dye

(reporter) while its 3' end is labeled with a fluorescence quencher (suppressor). It is used for real time monitoring of PCR.

Trisomy: presence of three homologous chromosomes instead of the usual two.

Twin peaks "+A" and "–A" products: Taq polymerase has the property of adding +A at the 3' end of the elongating strand that results in formation of products one base pair longer. These products appear as twin peaks on a genetic analyzer.

Variant: a heritable change in DNA sequence.

VNTR: Variable Number of Tandem Repeats are tandemly repeated sequences of DNA that range in length between 7-80bp. These are also called mini-satellites.

Whole exome sequencing (WES): sequencing of the protein coding regions of the genome.

Whole genome sequencing (WGS): sequencing of the entire genome of an organism.

Index

ABOUT THE AUTHOR

Suhaib Ahmed, graduated from King Edward Medical College in 1980. He got fellowship in Haematology from the College of Physicians & Surgeons, Pakistan in 1986 and PhD from University College London in 1998. While working at the prestigious Armed Forces Institute of Pathology, Rawalpindi he started in 1994 the first clinical service for prenatal diagnosis of thalassaemia and other genetic disorders in Pakistan. Since then, he has done over 15,000 prenatal diagnoses. He is renowned for developing several cost-effective PCRs in neoplastic and infectious diseases. He has also developed a complete range of instruments used in DNA analysis. He has vast experience of STR analysis for diagnostic and forensic use. He has over 70 publications in the National and some of the most well reputed International Journals. He is the founding President of Association for Molecular Pathology in Pakistan. He is the past President of Pakistan Society of Haematology and is the current President of Pakistan Thalassaemia Welfare Society and Autism Society of Pakistan.

978-969-23272-1-3

GRC PUBLICATION