Brief content of lecture

Discipline *"Methods of molecular biotechnology"*

Lecture 1. Physical and Structural techniques Nucleic acids

DNA extraction methods

DNA Extraction

- **Purpose of DNA extraction**
- **Review the main steps in**

the DNA extraction protocol

and the chemistry

Purpose of DNA Extraction

To obtain DNA in a relatively purified form which can be used for further investigations, i.e. PCR, sequencing, etc

Basic Protocol

- **Most DNA extraction protocols consist of two parts**
  1. **A technique to lyse the cells gently and solubilize the DNA**
  2. **Enzymatic or chemical methods to remove contaminating proteins, RNA, or macromolecules**
- **In plants, the nucleus is protected within a nuclear membrane which is surrounded by a cell membrane and a cell wall. Four steps are used to remove and purify the DNA from the rest of the cell.**
  1. **Lysis**
  2. **Precipitation**
  3. **Wash**
  4. **Resuspension**

A comparison of DNA extraction methods used in research labs as opposed to "classroom labs"

<u>Research</u>

***Lysis*: grind in Liquid N_2 and use detergent**

***Precipitation Part I*: phenol/chloroform extraction to get rid of proteins**

***Precipitation Part II*: addition of salts to interrupt hydrogen bonding between water and phosphates on the DNA**

***Precipitation Part III*: addition of ethanol to pull DNA out of solution**

***Wash and resuspend:* DNA is washed in ethanol, dried, and resuspended in H20 or TE buffer.**

"Classroom"

***Lysis*: grind in mortar/pestel and use detergent**

***Precipitation Part I*: NONE (chemical are too dangerous!)**

***Precipitation Part II*: addition of salts to interrupt hydrogen bonding between water and phosphates on the DNA**

***Precipitation Part III*: addition of ethanol to pull DNA out of solution**

***Wash and resuspend:* DNA is washed in ethanol, dried, and resuspended in H20 or TE buffer.**

LYSIS:
In DNA extraction from plants,
this step commonly refers to the breaking
of the cell wall and cellular membranes (most importantly, the plasma and nuclear membranes)

- **The cell wall (made of cellulose) is disrupted by mechanical force (for example, grinding the leaves)**
- **Then the addition of a detergent in the which breaks down the cell membranes**
    - **Detergents are able to disrupt membranes due to the amphipathic (having both hydrophilic and hydrophobic regions) nature of both cellular membranes and detergent molecules. The detergent molecules are able to pull apart the membranes**
- **The end result of LYSIS is that the contents of the plant cells are distributed in solution.**

Why add detergent?

Blending separated the pea cells, but each cell is surrounded by a sack (the cell membrane). DNA is found inside a second sack (the nucleus) within each cell.

To see the DNA, we have to break open these two sacks.

Why add detergent?

We do this with detergent.

Think about why you use soap to wash dishes or your hands. To remove grease and dirt, right?

Why add detergent?

Soap molecules and grease molecules are made of two parts:

Heads, which like water

Tails, which hate water.

Why add detergent?

Both soap and grease molecules organize themselves in bubbles (spheres) with their heads outside to face the water and their tails inside to hide from the water.

Why add detergent?

When soap comes close to grease, their similar structures cause them to combine, forming a greasy soapy ball.

Why add detergent?

A cell's membranes have two layers of lipid (fat) molecules with proteins going through them.

Why add detergent?

When detergent comes close to the cell, it captures the lipids and proteins.

What is an enzyme?

After the detergent step, the last question was: what do you have now in your pea soup?

The cell and nuclear membranes have been broken apart,
as well as all of the organelle membranes.

What is an enzyme?

The DNA in the nucleus of the cell is molded, folded, and protected by proteins. The meat tenderizer cuts the proteins away from the DNA.

Extracting DNA

DNA will rise into the alcohol layer from the pea layer

Use a wooden stick draw DNA into the alcohol

What is the stringy stuff?

Alcohol is less dense than water, so it floats on top.

Since two separate layers are formed, all of the grease and the protein that we broke up in the first two steps and the DNA have to decide which layer to go to.

What is the stringy stuff?

In this case, the protein and grease parts find the bottom, watery layer the most comfortable place, while the DNA prefers the top, alcohol layer.

DNA is a long, stringy molecule that likes to clump together.

What is the stringy stuff?

DNA is a long, stringy molecule. The salt you added in step one helps it stick together. So what you see are clumps of tangled DNA molecules!

DNA normally stays dissolved in water, but when salty DNA comes in contact with alcohol it becomes undissolved. This is called precipitation. The physical force of the DNA clumping together as it precipitates pulls more strands along with it as it rises into the alcohol.

You can use a wooden stick or a straw to collect the DNA. If you want to save your DNA, you can transfer it to a small container filled with alcohol.

PRECIPITATION (In a lab):
This a series of steps where DNA is separated from the rest of the cellular components

In a research lab, the first part of precipitation uses phenol/chloroform to remove the proteins from the DNA

Phenol denatures proteins and dissolves denatured proteins.

Chloroform is also a protein denaturant

THIS STEP CANNOT BE PERFORMED IN "CLASSROOM LABS"!!

The second part of research lab DNA precipitation is the addition of salts

The salts interrupt the hydrogen bonds between the water and DNA molecules.

The DNA is then precipitated from the protein in a subsequent step with isopropanol or ethanol

In the presence of cations, ethanol induces a structural change in DNA molecules that causes them to aggregate and precipitate out of solution.

The DNA is pelleted by spinning with a centrifuge and the supernatant removed

PRECIPITATION (In a "classroom lab"):
This a series of steps where DNA is separated from the rest of the cellular components

In a classroom lab, DNA precipitation involves the addition of salts

The salts interrupt the hydrogen bonds between the water and DNA molecules.

The DNA is then precipitated from the protein in a subsequent step with isopropanol or ethanol

In the presence of cations, ethanol induces a structural change in DNA molecules that causes them to aggregate and precipitate out of solution.

The DNA is pelleted by spinning with a centrifuge and the supernatant removed

Note: because this protocol does not use phenol/chloroform, the DNA extracted in a classroom lab is not as "clean" as the DNA extracted in a research lab!

Washing:

The precipitated DNA is laden with acetate salts. It is "washed" with a 70% ethanol solution to remove salts and other water soluble impurities but not resuspend the DNA.

Resuspension:

The clean DNA is now resuspended in a buffer to ensure stability and long term storage.

The most commonly used buffer for resuspension is called 1xTE

Checking the Quality of your DNA

- **The product of DNA extraction will be used in subsequent experiments**
- **Poor quality DNA will not perform well in PCR**
- **To assess the quality of DNA extraction use the following simple protocol:**
    - **Mix 10 µL of DNA with 10 µL of loading buffer**
    - **Load this mixture into a 1% agarose gel**

▶ **Analyze results (the following slides provide guidance)**

Expected Results in a
"Classroom Lab"

Using the protocol in the Cereal Genomics module, the genomic DNA extracted will look different than the optimized DNA extraction on the previous slide (*this is mainly due to the missing phenol/chloroform step*)

This is expected. Even though this genomic DNA preparation is not perfect, it is suitable for use as a PCR template

Lane A: Barley

Lane B: Corn

Lane C: Oat

Lane D: Rice

Lane E: Wheat

LECTURE 2 Extraction of RNA

1. **Main principes of RNA Extraction**
2. **A Rapid, Quantitative Assay for Direct Detection of MicroRNAs and Other Small RNAs Using Splinted Ligation**

- ISOLATION OF RNA

Most RNA isolation protocols also involve phenol extractions and are similar to DNA isolations.

However, there are some differences and special considerations.

In particular, precautions *(Меры предосторожности)* against RNase activity must be taken.

RNase is an extremely stable and active enzyme.

Gloves should be worn at all times and sterile plastic ware should be used whenever possible to avoid introducing exogenous RNase to the sample.

Glassware needs to treated with DEPC-water and autoclaved to inactivate any RNase. Buffers should be prepared from DEPC-water or RNase inhibitors included.

- RNase inhibitors!
- • extraction in guanidine salts
- • phenol extractions at pH 5-6
- • treatment with RNase-free DNase
- • precipitation with LiCl
- • oligo-dT column

The cell lysis and solubilization of RNA will typically be carried out in guanidine salts (especially guanidine thiocyanate). **Guanidine is a strong chaotropic agent** and will inhibit RNases.

The strong denaturing effect of this salt will also promote better phenol extraction.

The phenol extraction is the same as the DNA isolation except that the phenol is usually saturated with a buffer of pH 5-6.

The lower pH will result in some DNA partitioning in the organic phase.

If DNA contamination is a problem, it is possible to purchase RNase-free DNase. Long term storage is best as a precipitate in 70% ethanol at –20o.

- Precipitation of RNA with LiCl.

LiCl has been used to selectively precipitate RNA.

Large RNAs (rRNA, mRNA) are insoluble at high ionic strength, whereas small RNAs (tRNA

and 5S rRNA) and DNA generally remain soluble. Following either phenol or guanidine

extraction, an equal volume of 8 M LiCl is added. The sample is mixed vigorously and

incubated at -20o.

The precipitate is collected by centrifugation and reprecipitated if necessary.

- Affinity chromatography

Most eukaryotic mRNA contains a stretch of A residues at its 3' end which added post-transcriptionally.

It is possible to isolate mRNA by affinity **chromatography on oligodT columns** (Figure).

The RNA solution is passed

over an oligo-dT column under conditions which promote base pairing.

Only RNA with a polyA tail binds. The polyA RNA is eluted under

conditions (usually low salt and high temperature) which breaks the base pairing.

- DENSITY GRADIENT CENTRIFUGATION

Density gradient centrifugation can also be used in the analysis and isolation of NA.

Double-stranded DNA, single-stranded DNA, RNA and proteins all have different densities in **CsCl**

and therefore can be separated by **isopycnic** (i.e., equilibrium) centrifugation.

- CsCl is the standard medium for the density gradient centrifugation of nucleic acids and are especially useful for the purification of large amounts of highly purified DNA.

The gradients are carried out in the presence of ethidium bromide which fluoresces when bound to DNA. The DNA bands are detected by illumination with ultraviolet light and easily recovered with a syringe and needle by puncturing the wall of the disposable tube and aspirating the DNA.

The CsCl can be removed by dialysis or by precipitating the DNA.

Density in CsCl

- DNA ~ 1.7 g/cm3
- Protein ~ 1.3 g/cm3
- RNA > DNA
- ssDNA > dsDNA

The %G:C content affects the density of DNA (Figure).

This can result in multiple bands on CsCl gradients if DNA composed different **G:C** compositions are centrifuged.

For example, minor bands, called **satellite DNA**, are often observed when total DNA from an organism is analyzed by CsCl gradients.

These **satellite bands** are usually due to highly repetitive DNA or organellar DNA.

Mitochondrial DNA has high **A:T content**.

These satellite DNA bands can be purified from the genomic DNA by *density gradient centrifugation*.

- density (g/cm3)
- Nucleic acids can also be separated according to **size** by *rate zonal centrifugation on sucrose gradients*.
-
- However, this method is not widely used since **gel electrophoresis i**s generally a more convenient method for the size fractionation of nucleic acids.
- A Rapid, Quantitative Assay for Direct Detection
  of MicroRNAs and Other Small RNAs Using Splinted Ligation

(miRNAs) are **endogenous small noncoding RNAs** that regulate gene expression by binding to target messenger RNAs and inducing translational repression, cleavage, or destabilization of the target.

Each miRNA can potentially regulate expression of a distinct set of genes and therefore miRNAs is ideally suited to rapidly adjusting protein concentrations in cells.

miRNAs is required during cell differentiation. In fact, certain miRNAs are differentially expressed, both spatially and temporally, in many types of immune cells.

Moreover, consistent with the discovery that miRNAs modulate gene expression, altered miRNA expression has been associated with various types of diseases,

including cancer.

Northern blotting is the standard method for the detection of small RNAs, because it allows direct comparison of the quantity of small RNA between different samples.

However, major drawbacks (*недостатки*) of Northern blotting are:

1. the time-consuming procedures and,
2. poor sensitivity, especially when monitoring expression of short nucleotide sequences.

3. Northern blotting requires relatively large amounts of starting material and involves multiple handling steps (Table 1).

- Direct Detection of MicroRNAs by ***splinted ligation***

This protocol describes a method that uses ***splinted ligation*** (*шинирование*) for in-solution, direct labeling of small RNAs from total RNA.

In mol. Biology ***ligation*** *is* the covalent linking of two ends of DNA or RNA molecules.

The liquid phase hybridization method makes it possible to achieve sensitive, specific, and quantitative detection while eliminating a number of time-consuming and labor-intensive steps required for the standard **Northern blot assay**.

The assay uses a small **RNA-specific bridge oligonucleotide (**РНК-специфический мостиковый олигонуклеотид to form **base pairs with the small RNA** and a 5′ end radiolabeled ligation oligonucleotide.

The captured (*захваченный*) small RNA is internally labeled by ligation.

Detection of the labeled small RNAs is performed by **denaturing gel electrophoresis** and autoradiography or phosphorimaging.

This protocol has been successfully used to study expression of various classes of biological small RNAs from nanogram to microgram amounts of total RNA **without an amplification step** and is significantly more simple and more sensitive than Northern blotting or ribonuclease protection assays.

Once the oligonucleotides have been synthesized and total RNA has been extracted, the procedure can be completed in 6 h.

- **Table. Comparison of the amplification-independent assays for small RNA detection**

- The splinted-ligation technique

(Fig. 1) is a nucleic acid hybridization assay that uses a bridge oligonucleotide with perfect Watson–Crick complementarity to a target small RNA and a

5/ end radiolabeled ligation oligonucleotide.

Simultaneous base pairing between both the small RNA and ligation oligonucleotide to the bridge oligonucleotide yields a double-stranded structure with a nick on one strand, which can be ligated with T4 DNA

Ligase, thus labeling the target small RNA.

In addition, because the labeled phosphate provided by the ligation oligonucleotide is

- ***The protocol is divided into five steps:***

1, labeling of the ligation oligonucleotide;

2, capturing of the ligation oligonucleotide and small RNA on a bridge oligonucleotide,

3. linking of the ligation oligonucleotide to the small RNA using T4 DNA ligase;

step 4, partial removal of labeled phosphate from the unligated oligonucleotide;

step 5, fractionation on a denaturing gel and detection by a phosphorimager.

The gel image shows detection of *miR-21* by splinted ligation.

Assay reactions were performed with the indicated amounts of HeLa cell total RNA.

Lanes designated “neg” is a no RNA negative control and “pos” is a synthetic *miR-21* positive control.

These controls were complete reactions in which the RNA samples were replaced by water and 2.5

femtomoles synthetic *miR-21* RNA, respectively. *Lane M* is 5¢ end-labeled oligodeoxynucleotides markers. The *top arrow*

indicates the position of *miR-21* ligated to the ligation oligonucleotide. The *bottom arrow* indicates residual radiolabeled

14 nt ligation oligonucleotide that is present due to incomplete removal of the 5¢ end-32P.

- **The gel image of mir-RNA detection by splinted-ligation technique**

The gel image shows detection of *miR-21* by splinted ligation.

Assay reactions were performed with the indicated amounts of HeLa cell total RNA.

Lanes designated “*neg*” is a no RNA negative control and “*pos*” is a synthetic *miR-21* positive control.

These controls were complete reactions in which the RNA samples were replaced by water and 2.5 *femtomoles (1^{-15}mol)* synthetic *miR-21* RNA, respectively. *Lane M* is 5/ end-labeled oligodeoxynucleotides markers.

The *top arrow* indicates the position of *miR-21* ligated to the ligation oligonucleotide.

The *bottom arrow* indicates residual radiolabeled 14 nt ligation oligonucleotide that is present due to incomplete removal of the 5/ end-32P.

- ***ANALYSIS AND QUANTIFICATION of NA***

The quality and quantity of isolated NA can be determined spectrophotometrically

Nucleic acids have an Amax of 260 nm and proteins have Amax of 280 nm.

The A260/A280 ratio is therefore indicative of the degree of purity of the nucleic acid. A260/A280 ratios

of 1.6-1.8 or 1.8-2.0 are usually acceptable for DNA and RNA, respectively.

The standard extinction coefficient used for ssDNA, dsDNA and RNA are 0.03 ml/µg, 0.02 ml/µg and 0.025

ml/µg, respectively.

Formulas which take into account protein and other contaminants are also available.

Indirect spectrophometric assays for DNA quantification are also available, but rarely used. In some instances fluorometry using fluorescent dyes that bind DNA and/or RNA is used to determine nucleic acid concentrations.

- DNA A260 1.0 ≈ 50 µg/ml A260/A280 1.6 - 1.8
- RNA A260 1.0 ≈ 40 µg/ml A260/A280 ~2.0
- Spectrophotometric Conversions
- 1 A260 unit of double-stranded DNA = 50 µg/ml

- 1 A260 unit of single-stranded DNA = 33 μg/ml
- 1 A260 unit of single-stranded RNA = 40 μg/ml
- **DNA Molar Conversions**
- 1 μg of 1,000 bp DNA = 1.52 pmole (3.03 pmoles of ends)
- 1 pmole of 1,000 bp DNA = 0.66 μg
- **Protein Molar Conversions**
- 100 pmoles of 100,000 dalton protein = 10 μg
- 100 pmoles of 50,000 dalton protein = 5 μg
- 100 pmoles of 10,000 dalton protein = 1 μg
- **LECTURE 3. Hybridization Conditions and Melting Temperature**
- ***Considerations for better oligonucleotide design***
- Hybridization is a common step of many molecular biology protocols.
- Examples of techniques that include nucleic acid hybridization are:

1. northern and Southern analysis,
2. PCR/qPCR,
3. cloning,
4. *in situ* hybridization,
5. array analysis, gene knockdown, and
6. next generation sequencing (NGS).

- **Important Factors That Affect Stringency and Hybridization**
- *Temperature of hybridization and salt concentration*
    - Increasing the hybridization temperature or decreasing the amount of salt in the buffer increases probe specificity and decreases hybridization of the probe to sequences that are not 100% the same.
- *Concentration of the denaturant in the buffer*
    - For example: Deionized Formamide and SDS can be used to reduce non-specific binding of the probe
- *Length and nature of the probe sequence*
- The criteria for hybridization are based on **nucleic acid strand melting**.
- Therefore, an understanding of melting temperature (T_m) provides information on when and how the DNA or RNA strands are going to hybridize and defines the rules for hybridization.
- It is very important to understand this process so that you can better design and optimize the oligos for your experiments.

- **Oligonucleotide concentration**
- T_m varies with oligonucleotide concentration.
- While folding of a single oligonucleotide is concentration independent, the T_m is strongly influenced by oligo concentration when 2 or more nucleic acid strands interact.
- Oligo concentration alone can cause T_m to vary by ±10°C. The strand that is in excess determines T_m.
- For example, in PCR/qPCR, the target concentration is usually designed to be much lower than that of the probe.
- In such situations, T_m is determined by the probe because it is in excess.
- **Salt environment**
- The concentrations of monovalent (sodium, potassium), divalent (magnesium), and polyvalent cations affect the stability of hybridized oligonucleotides.
- Divalent cations have the biggest impact on T_m—changes in the millimolar range are significant. Increasing the concentration of monovalent cations, such as Na^+, up to 1–2 M stabilizes oligos.
- However, these higher concentrations can significantly impact Tm. "A change from 20–30 mM Na^+ to 1 M Na^+ can cause oligonucleotide T_m to vary by as much as 20°C," "We have worked through the calculations for sodium and magnesium ion concentrations and have come up with complicated models to predict T_m.
- **Salt environment**
- This formula is used in the IDT SciTools® programs to ensure your calculations are correct." (See the formula in the above sidebar, *Accurate T_m Calculation*, and read a detailed explanation in the Technical Report, *Calculation of T_m for Oligonucleotide Duplexes*, listed in the Related Reading box below). The free online SciTools programs, such as the OligoAnalyzer® tool, are available on the IDT website at www.idtdna.com/scitools.
- Only free Mg^{2+} in solution reacts with DNA; therefore, it is also important to consider the presence of any compounds that bind magnesium ions.
- **Salt environment**
- For example, PCR reactions require deoxynucleoside triphosphates (dNTPs) in a mixture with short oligomer primers, probes, and longer nucleic acid targets. Magnesium ions bind to all of these components, thus decreasing the concentration of free Mg^{2+}.
- As DNA synthesis proceeds during PCR, dNTPs are incorporated into the products and pyrophosphate is released.
- Pyrophosphate also binds Mg^{2+}. Ideally, all of this should be considered when estimating the concentration of free Mg^{2+} in the hybridization solution.
- ***Mismatches*** Несоответствия ***and single nucleotide polymorphisms (SNPs)***
- Mismatches between hybridizing oligos have a profound (*глубокий*) effect on T_m. The effect depends both on sequence context and solution composition (while salt is the main factor considered, other additives, like urea, DMSO, and even SYBR® Green, can shift T_m).

- A single mismatch can cause T_m to vary between 1 and 18°C in PCR applications. The identity of the mismatch, its position in the sequence, and its context all impact the degree of the mismatches effect. A-A and A-C are among the least stable mismatch pairs, causing the largest T_m variation, as compared to G-T, one of the more stable mismatch pairs.
- The context of the mismatch (e.g., whether a G-T mismatch is adjacent to an A-T or G-C base pair) also affects T_m.
- **Mismatches and single nucleotide polymorphisms (SNPs)**
- SNPs that underlie primer and probe hybridization sites, cause mismatches.
- Thus, "SNPs can affect PCR/qPCR experimental outcomes.
- There are a lot of SNPs out there, and their numbers are growing exponentially!
- Thus, we must design our PCR/qPCR assays intelligently, with SNPs in mind. It is important to check NCBI's dbSNP database (www.ncbi.nlm.nih.gov/snp) to determine whether SNPs have been identified within your amplicon.
- Then you can attempt to design your primer and probe locations around those SNPs, as the mismatches they create could likely affect your PCR results,". "However, by positioning them towards the 5' end of the probe or primer sequences, thus, away from the 3' end, these effects can be minimized or even prevented."
- Stringency is a term that many molecular technologists are all very familiar with. It is a term that describes the combination of conditions under which a target is exposed to the probe. Typically, conditions that exhibit high stringency are more demanding of probe to target complementarity and length. Low stringency conditions are much more forgiving.
- If conditions of stringency are too HIGH → Probe doesn't bind to the target
- If conditions of stringency are too LOW → Probe binds to unrelated targets
- **Melting Temperature (T_m) Long Probes**
- The ideal hybridization conditions are estimated from the calculation of the T_m.
- The T_m of the probe sequence is a way to express the amount of energy required to separate the hybridized strands of a given sequence.
- At the T_m: Half of the sequence is double stranded and half of the sequence is single stranded.
- $Tm = 81.5°C + 16.6\log M + 0.41(\%G+C) – 0.61(\%formamide) – (600/n)$
- Where M = Sodium concentration in mol/L
- n = number of base pairs in smallest duplex
- If we keep in mind that RNA is single stranded (ss) and DNA is double stranded (ds), then the following must be true:
- RNA : DNA Hybrids More stable
- DNA : RNA Hybrids ↓
- DNA : DNA Hybrids Less stable

- T_m of RNA probes is higher, therefore RNA : DNA hybrids increase the Tm by 20 – 25°C
- **Calculating the T_m for Short Probes (14 – 20 base pairs)**
- Tm = 4°C x number of G/C pairs + 2°C x number of A/T pairs
- The hybridization temperature (annealing temp) of oligonucleotide probes is approximately 5°C below the melting temperature.
- **Sequence Complexity (C_ot)**
- Sequence complexity refers to the length of unique, non-repetitive nucleotide sequences.
- C_ot = Initial DNA Concentration (C_o) x time required to reanneal it (t)
- $C_o t_{1/2}$ = Time required for half of the double-stranded sequence to anneal under a given set of conditions.
- Short probes can hybridize in 1 – 2 hours, where long probes require more time.

- **LECTURE 4.** MODIFICATION OF NUCLEIC ACIDS
- All DNA molecules are similar in terms of their **biochemical and physical properties.**

Therefore, unlike proteins, DNA techniques are not highly dependent upon the particular gene being studied.

Genomic DNA is an extremely large molecule. For example, the

human genome contains approximately 2 x 109 base pairs (bp).

The size of a gene for a 50 kDa protein might be as small as 2000 bp (or 2 kb). Therefore, it can be quite difficult to identify and characterize specific genes from an organism.

- **The study of specific genes involves**
- manipulating nucleic acids.
- In particular, it is possible to break DNA into smaller fragments, identify fragments of interest and to amplify these fragments so that they can be analyzed.
- Enzymes are used to carry out these manipulations of DNA and RNA. DNA modifying enzymes include:

1. **polymerases,**
2. **ligases and**
3. **nucleases**.

Polymerases synthesize nucleic acids in a template mediated fashion.

Ligases join fragments of DNA.

Nucleases cleave the **phosphodiester bond** between nucleotides. A wide range of specificities are exhibited by nucleases.

Exonucleases remove nucleotides one at a time from either the 5' or 3' end.

In addition, some exonucleases exhibit substrate specificities in terms of preferences for double-stranded (ds) or single-stranded (ss) DNA or RNA.

Endonucleases cleave the phosphodiester bond in the middle of a oligonucleotide and produce fragments.

Such exo- and endonucleases have some applications primarily directed at removing unwanted types of nucleic acids or removing single stranded overhangs from dsDNA.

- RESTRICTION ENDONUCLEASES
- **Three distinct classes of restriction endonucleases, designated types I, II and III:**
- **Type I** cleavage occurs 400-7000 bp from recognition site
- **Type II** cleavage occurs adjacent or within recognition site
- **Type III** cleavage occurs 25-27 bp from recognition site.
- Class II restriction enzymes are the most useful since they cleave DNA at the recognition site.

The recognition sequences of type II restriction enzymes are generally *4-8 consecutive (последовательный)nucleotides.*

Some restriction enzymes do allow for some *degeneracy* (*вырождение*) at a particular nucleotide.

This degeneracy can be restricted to **a purines (R)** or **pyrimidines (Y),** or be completely

unrestricted (N).

Most recognition sites are ***palindromes*** in that both strands exhibit the same sequence.

The sequence complementarity and opposite orientations of the two strands leads to a dyad symmetry.

Hydrolysis of the phosphodiester bond results in the phosphate on the 5'-carbon and the hydroxyl on the 3'-carbon.

Both DNA strands are cut between the same two residues.

This will result in blunt ends if the cleavage site is in the exact center of the recognition site.

Alternatively, 5' overhangs (or extensions) or 3' overhangs of varying length will be

produced if the cleavage site is not in the center of the recognition site.

- **Isoschizomers**
- are restriction enzymes isolated from different sources that recognize ***the same sequence***, but cleave at different positions resulting in different overhangs.
- In addition, different recognition sites can produce identical overhangs (called compatible ends).

LECTURE 5. Main principles of electrophoresis for analysis of nucleic acids

- GEL ELECTROPHORESIS
- SDS-PAGE
- Practical considerations.
- Electrophoresis,
- like centrifugation, is a **hydrodynamic technique.**

- A charged particle (i.e., molecule) in an electric field experiences a force that is proportional to the potential difference (E), or voltage, of the electric field and inversely proportional to the distance **(d)** between the electrodes.
- (The potential difference divided by the distance **(E/d**) is referred to as the field strength.)
-
- The force will also be proportional to the net charge of the molecule **(q).**
- Therefore, the force experienced by the molecule can be expressed by the following equation:
- The force of charged particle (i.e., molecule) experienced them in an electric field will by opposed by a frictional force (*сила трения*) **(= fv),**
- where **f** is a frictional coefficient and **v** is the velocity of the particle
- The frictional coefficient depends on the size (eg., **r = radius**) and shape of the molecule and the **viscosity (η)** of the medium.
- For example, in the case of a sphere the frictional force is:
- A particle will move at a velocity **(v)** so that these two forces are equal, therefore:
- or solving for v
- This equation indicates that the mobility (i.e., velocity) of a molecule in an electric field
- is proportional to the electric field **(E/d),** or more simply the applied voltage, and the net charge of the molecule.
- **The mobility of particle in** an electric field **is inversely proportional to a frictional coefficient (i.e., size and shape of the molecule and the viscosity of the medium), as indicated by the following equation:**
- **mobility = (applied voltage)(net charge/(friction coefficient**
- **Therefore, it is possible to derive information about the charge, size and shape of a molecule by its mobility in an electric field.**
- GEL ELECTROPHORESIS
- Electrophoresis of macromolecules can be carried out in solution.
- However, the ability to separate molecules is compromised by their diffusion.
- **Greater resolution is achieved if electrophoresis is carried out on semi-solid supports such as polyacrylamide or agarose gels**.
- Gels are formed by cross-linking polymers in aqueous medium.
- This will form a 3-dimensional meshwork which the molecules must pass through.
- Polyacrylamide is a common gel for protein electrophoresis whereas agarose is more commonly used for nucleic acids.
- ***Agarose gels have a larger pore size than acrylamide gels and are better suited for larger macromolecules. However, either type of gel can be applied to either nucleic acids or proteins depending on the application.***
- Gels are formed from long polymers in a cross-linked lattice.
- The space
- between the polymers are the pores. Higher concentrations of the polymer will result in smaller
- average pore sizes. Polyacrylamide gels are formed by covalently cross-linking acrylamide
- monomers with bis-acrylamide with a free radical like persulfate ($SO4\cdot$). The cross-linking of
- the acrylamide polymers results in 'pores' of a defined size. The total acrylamide concentration
- and the ratio of bis-acrylamide to acrylamide will determine the average pore size. The
- polyacrylamide solution is poured into a mold and polymerized. This mold can be a cylindrical
- tube, but is usually a 'slab' poured between two glass plates
- EQUIPMENT. Equipment to conduct gel electrophoresis is relatively simple. They consist of a mold to form the gels, an apparatus to hold the gel and contain buffers, and a power supply capable of delivering the required voltage or current.
- Discontinuous or "disc" electrophoresis.

- The Laemmli discontinuous buffers are extensively used in gel electrophoresis. Discontinuous gels consist of two distinct gel regions referred
- to as **stacking gel (**Штабелирующий) and separating gel and a Tris-glycine tank buffer.
- **The stacking gel has a lower acrylamide concentration, a lower pH and a lower ionic strength than the separating**
- Composition of Laemmli Gels
- The lower ionic strength of the stacking gel results in a greater local electric field strength than in the separating gel.
- The field strength difference combined with the lower acrylamide concentration results in
- proteins having a higher mobility in the stacking gel than in the separating gel.
- In addition, the glycine in the tank buffer has a higher mobility in the separating gel than in the stacking gel
- because of the pH differences.
- Therefore, proteins will migrate faster than the glycine in the stacking gel.
- When proteins reach the separating gel **their mobility is decreased** because of the
- increased acrylamide concentration and decreased field strength, whereas the increase in pH
- results in glycine having a higher mobility.
- All of these factors result in the proteins becoming compressed at the interface between the two gels and thus increasing resolution.
- Resolution in non-discontinuous electrophoresis depends partially on **the volume of the sample.**
- However, stacking also occurs at the interface of the sample and gel, especially if a high voltage is applied.
- SDS-PAGE
- Polyacrylamide gel electrophoresis in the presence of SDS (sodium dodecyl sulfate) is the most common form of protein gel electrophoresis.
- **SDS completely disrupts protein-protein interactions and denatures almost all proteins resulting in a complete unfolding of proteins.**
- In addition, ß-mercaptoethanol (or other reducing agents) is often used **to break disulfide bonds.**
- The SDS binds to the unfolded proteins giving all proteins a similar shape (i.e., random coil or extend conformation) and an uniform charge-to-mass ratio.
- In other words, coating proteins with a negatively charged detergent minimizes the effects of a protein's net charge.
- Therefore, during electrophoresis in the presence of SDS the mobility of a protein now depends primarily upon its size (i.e., mobility is inversely proportional to protein mass).
- SDS-PAGE
- Mobility in SDS gel electrophoresis is expressed as a **relative mobility (Rf).**
- The distance the protein migrated is compared to the length of the gel, or:
- The length of the gel is often defined by the migration of a substance which is not impeded by the matrix such a small molecular weight tracking dye (eg., bromophenol blue).
- This mobility can then be used to calculate the size of proteins.
- Protein standards of known size are used to generate a standard curve by plotting the log of the molecular weight against the Rf values.
- Practical considerations.
- 1. Pour separating gel.
- 2. Pour stacking gel.
- 3. Load samples.
- 4. Apply electric field.
- 5. Stain or process gel.
- Proteins to be analyzed by SDS-PAGE are solubilized in **a sample buffer.**

- Typically contains 2% SDS and 5% β-mercaptoethanol and then boiled.
- The reducing agent is omitted in situations where disulfide bonds need to be preserved.
- When an enzyme activity will be measure following electrophoresis, a lower SDS concentration is used and the sample is not boiled.
- The amount of protein that can be loaded onto a gel is limited. Overloading the gels results in the pores becoming plugged (*закупоривать*) and has an adverse effect on the electrophoresis.
- ISOELECTRIC FOCUSING
- Isoelectric focusing (IEF) separates proteins based on their **isoelectric points.**
- The isoelectric point is defined as the pH at which a protein has no net charge (i.e., the number of negative and positive charges are equal) and is a measure of the protein's net charge.
- Separating proteins according to their net charge is accomplished by generating a **pH gradient** in an electric field.
- The effect of protein size on mobility is minimized by carrying out the electrophoresis gels with **large pore sizes** such as low acrylamide concentrations (eg., 3.5%) or agarose.
- This large pore size minimizes the molecular sieving.
- A pH gradient is generated with **carrier ampholytes**.
- These ampholytes are a mixture of aliphatic amines and either carboxylic or sulfonic
- acid. They have a high buffering capacity, low molecular weight (300-600 Da) and a range of pKa values. Initially the pH of an ampholyte
- solution will be the average of the pKa values of the mixture.
- Application of an electric current causes the ampholytes to migrate toward the electrodes according to their charges.
- Ampholytes that have pKa values above the pH will be positively charged and those with pKa values below the pH will be negatively charged. As the ampholytes migrate this will result in changes in the local pH due to the buffering action of the ampholytes.
- .
- This change in the local pH will affect the charge on the ampholytes depending upon the pKa.
- The ampholytes will continue to migrate until they reach a position in which the local pH equals their pKa (i.e., no net charge). The end result is a pH gradient in which the most basic ampholytes are found at the cathode, a dilute alkali solution (eg., NaOH), and the most acidic ampholytes are at the anode, a dilute acid solution (eg., H_3PO_4).
- Carrier ampholytes with defined pH ranges can be purchased or prepared by isoelectric focusing.
- Proteins are also ampholytes and will migrate within the pH gradient until they reach a pH equal to their isoelectric point.
- The carrier ampholytes are needed since the protein concentration is generally not high enough to establish a stable pH gradient and the isoelectric points of the proteins may not be uniformly distributed along a pH gradient
- IEF is an equilibrium phenomenon since the components of the system migrate until
- they have no net charge. As the system approaches equilibrium the resistance approaches
- infinity since there are no ions to conduct the current. However, the pH gradient will start to
- break down before true equilibrium is reached and the ampholytes will migrate into the anode
- and cathode buffers.
- This gradient breakdown is accompanied by a lowering of the resistance.
- Therefore, the progress of IEF can be followed by performing the electrophoresis under
- constant voltage and monitoring the current. Initially the current will rapidly drop in concordance with the rapid migration of the ampholytes.
- As the ampholytes **lose their net charge**, the resistance increases and the current decreases (E = IR).
- The rate at which the current decreases levels off as the system approaches equilibrium. The current will start to rise again when the pH gradient starts to break down.

- IEF needs to be discontinued before this point.
- The pH gradient can be determined with **marker proteins** with known isoelectric points or by measuring the pH along the gel.
- This is accomplished by slicing the gel into pieces,
- eluting the ampholytes into distilled water and measuring the pH.

LECTURE 6. *Nucleic Acid Detection.*

LECTURE 6. NUCLEIC ACID DETECTION

- Autoradiography
- *in situ* hybridization
- DNA Separation Techniques
- PCR and RT-PCR
- 3.5 Nucleic Acid Detection
- Hybridization of a labeled nucleic acid to complementary sequences can identify specific nucleic acids.
- probe – A radioactive nucleic acid, DNA or RNA, used to identify a complementary fragment.
- 3.5 Nucleic Acid Detection
- autoradiography – A method of capturing an image of radioactive materials on film.
- 3.5 Nucleic Acid Detection
- *in situ* hybridization – Hybridization of a probe to intact tissue to locate its complementary strand by autoradiography.
- DNA Separation Techniques
- Gel electrophoresis separates DNA fragments by size, using an electric current to cause the DNA to migrate toward a positive charge.
- DNA Separation Techniques
- DNA can also be isolated using density gradient centrifugation.
- ***DNA Sequencing***
- Classical chain termination sequencing uses dideoxynucleotides (ddNTPs) to terminate DNA synthesis at particular nucleotides.
- Primer - A single stranded nucleic acid molecule with a 3′ –OH used to initiate DNA polymerase replication of a paired template strand.
- ***DNA Sequencing***
- Fluorescently tagged ddNTPs and capillary gel electrophoresis allow automated, high-throughput DNA sequencing.

- The next generations of sequencing techniques aim to increase automation and decrease time and cost of sequencing.
- ***PCR and RT-PCR***
- Polymerase chain reaction (PCR) permits the exponential amplification of a desired sequence, using primers that anneal to the sequence of interest.
- ***PCR and RT-PCR***
- RT-PCR uses reverse transcriptase to convert RNA to DNA for use in a PCR reaction.
- ***PCR and RT-PCR***
- Real-time, or quantitative, PCR detects the products of PCR amplification during their synthesis, and is more sensitive and quantitative than conventional PCR.
- PCR depends on the use of thermostable DNA polymerases that can withstand multiple cycles of template denaturation.
- ***PCR and RT-PCR***
- fluorescence resonant energy transfer (FRET) – A process whereby the emission from an excited fluorophore is captured and reemitted at a longer wavelength by a nearby second fluorophore whose excitation spectrum matches the emission frequency of the first fluorophore.
- Blotting Methods
- Southern blotting involves the transfer of DNA from a gel to a membrane, followed by detection of specific sequences by hybridization with a labeled probe.
- ***Blotting Methods***
- Northern blotting is similar to Southern blotting, but involves the transfer of RNA from a gel to a membrane.
- Western blotting entails separation of proteins on a sodium dodecyl sulfate (SDS) gel, transfer to a nitrocellulose membrane, and detection of proteins of interest using antibodies.
- ***Blotting Methods***
- epitope tag – A short peptide sequence that encodes a recognition site (“epitope”) for an antibody, typically fused to a protein of interest for detection or purification by the antibody.
- ***DNA Microarrays***
- DNA microarrays comprise known DNA sequences spotted or synthesized on a small chip.
- DNA Microarrays
-
- Genome-wide transcription analysis is performed using labeled cDNA from experimental samples hybridized to a microarray containing sequences from all ORFs of the organism being used.
- SNP arrays permit genome-wide genotyping of single-nucleotide polymorphisms.
- Array comparative genome hybridization (array-CGH) allows the detection of copy number changes in any DNA sequence compared between two samples.

- **LECTURE 7.** ***DNA cloning techniques***
- 3.1 Introduction
- ***Cloning vector*** **–** DNA (often derived from a plasmid or a bacteriophage genome) that can be used to propagate an incorporated DNA sequence in a host cell.
    - Vectors contain selectable markers and replication origins to allow identification and maintenance of the vector in the host.
- 3.2 Nucleases
- **Nucleases** hydrolyze an ester bond within a phosphodiester bond.
- **Phosphatases** hydrolyze the ester bond in a phosphomonoester bond.
- 3.2 Nucleases
- **endonuclease –** Nuclease that cleaves phosphoester bonds within a nucleic acid chain.
    - It may be specific for RNA or for single-stranded or double-stranded DNA.
- **exonuclease –** Nuclease that cleaves phosphoester bonds one at a time from the end of a polynucleotide chain.
    - It may be specific for either the 5′ or 3′ end of DNA or RNA.
- 3.2 Nucleases
- Restriction endonucleases can be used to cleave DNA into defined fragments.
- 3.2 Nucleases
- A map can be generated by using the overlaps between the fragments generated by different restriction enzymes.
- 3.3 Cloning
- **Cloning** a fragment of DNA requires a specially engineered **vector**.
- **recombinant DNA** – A DNA molecule that has been created by joining together two or more molecules from different sources.
- **ligating (or ligation)** – The process of joining together two DNA fragments.
- 3.3 Cloning
- **subclone –** The process of breaking a cloned fragment into smaller fragments for further cloning.
- **multiple cloning site (MCS) –** A sequence of DNA containing a series of tandem restriction endonuclease sites used in cloning vectors for creating recombinant molecules.
- 3.3 Cloning
- 3.3 Cloning
- **transformation** – The acquisition of new genetic material by incorporation of added exogenous, nonviral DNA.

- Blue/white selection allows the identification of bacteria that contain the vector plasmid and vector plasmids that contain an **insert**.
- 3.4 Cloning Vectors Can Be Specialized for Different Purposes
- 3.4 Cloning Vectors Can Be Specialized for Different Purposes
- 3.4 Cloning Vectors Can Be Specialized for Different Purposes
- Cloning vectors may be bacterial plasmids, phages, **cosmids**, or **yeast artificial chromosomes**.
- **Shuttle vectors** can be propagated in more than one type of host cell.
- **Expression vectors** contain promoters that allow transcription of any cloned gene.
- 3.4 Cloning Vectors Can Be Specialized for Different Purposes
- **Reporter genes** can be used to measure promoter activity or tissue-specific expression.
- 3.4 Cloning Vectors Can Be Specialized for Different Purposes
- Numerous methods exist to introduce DNA into different target cells.
- 3.7 DNA Sequencing
- Classical chain termination sequencing uses **dideoxynucleotides** (ddNTPs) to terminate DNA synthesis at particular nucleotides.
- **Primer -** A single stranded nucleic acid molecule with a 3′ –OH used to initiate DNA polymerase replication of a paired template strand.
- 3.7 DNA Sequencing
- Fluorescently tagged ddNTPs and capillary gel electrophoresis allow automated, high-throughput DNA sequencing.
- The next generations of sequencing techniques aim to increase automation and decrease time and cost of sequencing.
- 3.8 PCR and RT-PCR
- **Polymerase chain reaction (PCR)** permits the exponential amplification of a desired sequence, using primers that anneal to the sequence of interest.
- 3.8 PCR and RT-PCR
- **RT-PCR** uses reverse transcriptase to convert RNA to DNA for use in a PCR reaction.
- 3.8 PCR and RT-PCR
- **Real-time**, or **quantitative**, **PCR** detects the products of PCR amplification during their synthesis, and is more sensitive and quantitative than conventional PCR.
- PCR depends on the use of thermostable DNA polymerases that can withstand multiple cycles of template denaturation.
- 3.8 PCR and RT-PCR

- **fluorescence resonant energy transfer** (**FRET**) – A process whereby the emission from an excited fluorophore is captured and reemitted at a longer wavelength by a nearby second fluorophore whose excitation spectrum matches the emission frequency of the first fluorophore.
- 3.9 Blotting Methods
- **Southern blotting** involves the transfer of DNA from a gel to a membrane, followed by detection of specific sequences by hybridization with a labeled probe.
- 3.9 Blotting Methods
- Northern blotting is similar to Southern blotting, but involves the transfer of RNA from a gel to a membrane.
- Western blotting entails separation of proteins on a sodium dodecyl sulfate (SDS) gel, transfer to a nitrocellulose membrane, and detection of proteins of interest using antibodies.
- 3.9 Blotting Methods
- **epitope tag –** A short peptide sequence that encodes a recognition site ("epitope") for an antibody, typically fused to a protein of interest for detection or purification by the antibody.
- 3.10 DNA Microarrays
- DNA microarrays comprise known DNA sequences spotted or synthesized on a small chip.
- 3.10 DNA Microarrays
- Genome-wide transcription analysis is performed using labeled cDNA from experimental samples hybridized to a microarray containing sequences from all ORFs of the organism being used.
- SNP arrays permit genome-wide genotyping of **single-nucleotide polymorphisms**.
- Array comparative genome hybridization (array-CGH) allows the detection of copy number changes in any DNA sequence compared between two samples.
- **LECTURE 8.** ***DNA Sequencing technques***
- Classical chain termination sequencing uses **dideoxynucleotides** (ddNTPs) to terminate DNA synthesis at particular nucleotides.
- **Primer -** A single stranded nucleic acid molecule with a 3′ –OH used to initiate DNA polymerase replication of a paired template strand.
- 3.7 DNA Sequencing
- Fluorescently tagged ddNTPs and capillary gel electrophoresis allow automated, high-throughput DNA sequencing.
- The next generations of sequencing techniques aim to increase automation and decrease time and cost of sequencing.
- 3.8 PCR and RT-PCR
- **Polymerase chain reaction (PCR)** permits the exponential amplification of a desired sequence, using primers that anneal to the sequence of interest.
- 3.8 PCR and RT-PCR

- **RT-PCR** uses reverse transcriptase to convert RNA to DNA for use in a PCR reaction.
- 3.8 PCR and RT-PCR
- **Real-time**, or **quantitative**, **PCR** detects the products of PCR amplification during their synthesis, and is more sensitive and quantitative than conventional PCR.
- PCR depends on the use of thermostable DNA polymerases that can withstand multiple cycles of template denaturation.
- 3.8 PCR and RT-PCR
- **fluorescence resonant energy transfer** (**FRET**) – A process whereby the emission from an excited fluorophore is captured and reemitted at a longer wavelength by a nearby second fluorophore whose excitation spectrum matches the emission frequency of the first fluorophore.
- 3.9 Blotting Methods
- **Southern blotting** involves the transfer of DNA from a gel to a membrane, followed by detection of specific sequences by hybridization with a labeled probe.
- 3.9 Blotting Methods
- Northern blotting is similar to Southern blotting, but involves the transfer of RNA from a gel to a membrane.
- Western blotting entails separation of proteins on a sodium dodecyl sulfate (SDS) gel, transfer to a nitrocellulose membrane, and detection of proteins of interest using antibodies.
- 3.9 Blotting Methods
- **epitope tag –** A short peptide sequence that encodes a recognition site ("epitope") for an antibody, typically fused to a protein of interest for detection or purification by the antibody.
- 3.10 DNA Microarrays
- DNA microarrays comprise known DNA sequences spotted or synthesized on a small chip.
- 3.10 DNA Microarrays
- Genome-wide transcription analysis is performed using labeled cDNA from experimental samples hybridized to a microarray containing sequences from all ORFs of the organism being used.
- SNP arrays permit genome-wide genotyping of **single-nucleotide polymorphisms**.
- Array comparative genome hybridization (array-CGH) allows the detection of copy number changes in any DNA sequence compared between two samples.

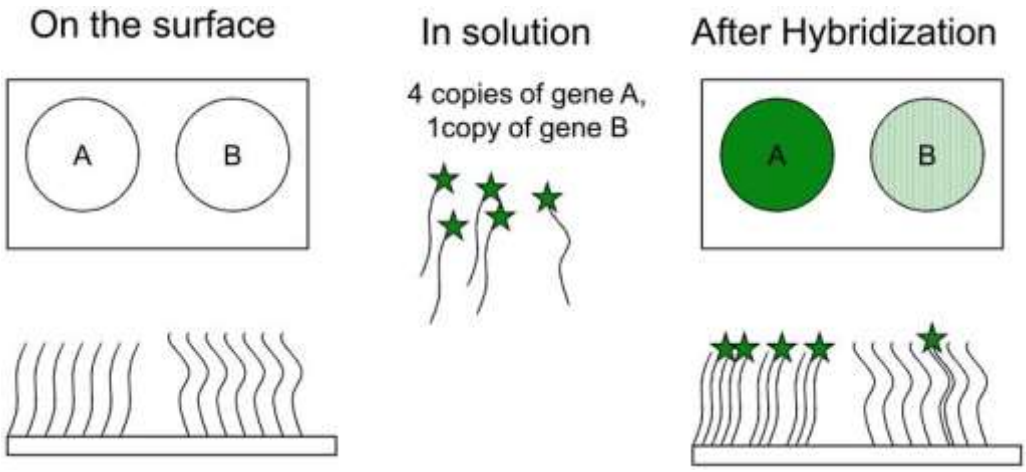
- **Lecture 8. DNA Sequencing techniques**
- 3.7 DNA Sequencing
- Classical chain termination sequencing uses **dideoxynucleotides** (ddNTPs) to terminate DNA synthesis at particular nucleotides.
- **Primer -** A single stranded nucleic acid molecule with a 3′ –OH used to initiate DNA polymerase replication of a paired template strand.

- 3.7 DNA Sequencing
- Fluorescently tagged ddNTPs and capillary gel electrophoresis allow automated, high-throughput DNA sequencing.
- The next generations of sequencing techniques aim to increase automation and decrease time and cost of sequencing.
- 3.8 PCR and RT-PCR
- **Polymerase chain reaction (PCR)** permits the exponential amplification of a desired sequence, using primers that anneal to the sequence of interest.
- 3.8 PCR and RT-PCR
- **RT-PCR** uses reverse transcriptase to convert RNA to DNA for use in a PCR reaction.
- 3.8 PCR and RT-PCR
- **Real-time**, or **quantitative**, **PCR** detects the products of PCR amplification during their synthesis, and is more sensitive and quantitative than conventional PCR.
- PCR depends on the use of thermostable DNA polymerases that can withstand multiple cycles of template denaturation.
- 3.8 PCR and RT-PCR
- **fluorescence resonant energy transfer** (**FRET**) – A process whereby the emission from an excited fluorophore is captured and reemitted at a longer wavelength by a nearby second fluorophore whose excitation spectrum matches the emission frequency of the first fluorophore.
- 3.9 Blotting Methods
- **Southern blotting** involves the transfer of DNA from a gel to a membrane, followed by detection of specific sequences by hybridization with a labeled probe.
- 3.9 Blotting Methods
- Northern blotting is similar to Southern blotting, but involves the transfer of RNA from a gel to a membrane.
- Western blotting entails separation of proteins on a sodium dodecyl sulfate (SDS) gel, transfer to a nitrocellulose membrane, and detection of proteins of interest using antibodies.
- 3.9 Blotting Methods
- **epitope tag –** A short peptide sequence that encodes a recognition site ("epitope") for an antibody, typically fused to a protein of interest for detection or purification by the antibody.
- 3.10 DNA Microarrays
- DNA microarrays comprise known DNA sequences spotted or synthesized on a small chip.
- 3.10 DNA Microarrays
- Genome-wide transcription analysis is performed using labeled cDNA from experimental samples hybridized to a microarray containing sequences from all ORFs of the organism being used.
- SNP arrays permit genome-wide genotyping of **single-nucleotide polymorphisms**.

- Array comparative genome hybridization (array-CGH) allows the detection of copy number changes in any DNA sequence compared between two samples.

Lecture 9-10. ***DNA Microarrays as new molecular biotechnology tool.***

- A **DNA microarray** (also commonly known as DNA chip or biochip) is a collection of microscopic DNA spots attached to a solid surface.
- Scientists use ***DNA microarrays*** to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. Each DNA spot contains picomoles (10^{-12} moles) of a specific DNA sequence, known as *probes* (or *reporters* or *oligos*). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA (also called antisense RNA) sample (called *target*) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. The original nucleic acid arrays were macro arrays approximately 9 cm × 12 cm and the first computerized image based analysis was published in 1981.
- ***Figure 1***
- Simplified view of a DNA array. The upper rectangles show two spots of DNA on a solid surface (sequences "A" and "B") prior to and after hybridization. The lower rectangles show highly idealized side views of the same surfaces.
-



The early history of DNA arrays

One could argue that the original DNA array was created with the colony hybridization method of Grunstein and Hogness, 1975). In this procedure, DNA of interest was randomly cloned into *E. coli* plasmids that were plated onto agar petri plates covered with nitrocellulose filters. Replica plating was used to produce additional agar plates. The colonies on the filters were lysed and their DNA's were denatured and fixed to the filter to produce a random and unordered collection of DNA spots that represented the cloned fragments. Hybridization of a radiolabeled probe of an DNA or RNA of interest was used to rapidly screen 1000's of colonies to identify clones containing DNA that was complimentary to the probe (Unit 6.3).

The birth of the modern DNA array

In the late 90's and 2000's, DNA array technology progressed rapidly as both new methods of production and fluorescent detection were adapted to the task. In addition, increases in our knowledge of the DNA sequences of multiple genomes provided the raw information necessary to assure that arrays could be made which fully represented the genes in a genome, all the sequence in a genome or a large fraction of the sequence variation in a genome. It should also be noted that during this time, there was a gradual transition from spotting relatively long DNA's on arrays to producing arrays using 25-60bp oligos. The transition to oligo arrays was made possible by the increasing amounts of publicly available DNA sequence information. The use of oligos (as opposed to longer sequences) also provided an increase in specificity for the intended binding target as oligos could be designed to target regions of genes or the genome that were most dissimilar from other genes or regions. Three basic types of arrays came into play during this time frame, spotted arrays on glass, *in-situ* synthesized arrays and self assembled arrays ***(Figure 2).***

Spotted arrays

In 1996 Derisi et. al. published a method which allowed very high-density DNA arrays to be made on glass substrates(DeRisi et al., 1996). Poly-lysine coated glass microscope slides provided good binding of DNA and a robotic spotter was designed to spot multiple glass slide arrays from DNA stored in microtiter dishes. By using slotted pins (similar to fountain pens in design) a single dip of a pin in DNA solution could spot multiple slides. Spotting onto glass, allowed one to fluorescently label the sample. Fluorescent detection provided several advantages relative to the radioactive or chemilluminescent labels common to filter based arrays. First, fluorescent detection is quite sensitive and has a fairly large dynamic range. Second, fluorescent labeling is generally less expensive and less complicated than radioactive or chemilluminescent labeling. Third, fluorescent labeling allowed one to label two (or potentially more) samples in different colors and co-hybridize the samples to the same array. As it was very difficult to reproducibly produce spotted arrays, comparisons of individually hybridized samples to ostensibly identical arrays would result in false differences due to array-to-array variation. However, a two-color approach in which the ratio of signals on the same array are measured is much more reproducible.

In-situ, Synthesized arrays

In 1991 Fodor et.al. published a method for light directed, spatially addressable chemical synthesis which combined photolabile protecting groups with photolithography to perform chemical synthesis on a solid substrate(Fodor et al., 1991). In this initial work, the authors demonstrated the production of arrays of 10-amino acid peptides and, separately, arrays of di-nucleotides. In 1994, Fodor et.al. at the recently formed company of Affymetrix demonstrated the ability to use this technology to generate DNA arrays consisting of 256 different octa-nucleotides (Pease et al., 1994). By 1995-1996, Affymetrix arrays were being used to detect mutations in the reverse transcriptase and protease genes of the highly polymorphic HIV-1 genome(Lipshutz et al., 1995) and to measure variation in the human mitochondrial genome(Chee et al., 1996). Eventually, Affymetrix used this technology to develop a wide catalogue of DNA arrays for use in expression analysis(Lockhart et al., 1996; Wodicka et al., 1997), genotyping (Chee et al., 1996; Hacia et al., 1996) and sequencing (G Wallraff, 1997)(see www.Affymetrix.com for the current catalog of arrays).

A major advantage of the Affymetrix technology is that because the DNA sequences are directly synthesized on the surface, only a small collection of reagents (the 4 modified nucleotides, plus a small handful of reagents necessary for the de-blocking and coupling steps) are needed to construct an arbitrarily complex array. This contrasts with the spotted array technologies in which one needed to construct or obtain all the sequences that one wished to deposit on the array in advance of array construction. However, the initial Affymetrix technology was limited in flexibility as each model of array required the construction of a unique set of photolithographic masks in order to direct the light to the array at each step of the synthesis process. In 2002, authors from Nimblegen Systems Inc., published a method in which the photo-deprotection step of Fodor et. al (Fodor et al., 1991; Lipshutz et al., 1999) is accomplished using micro-mirrors (similar to those in video computer projectors) to direct light at the pixels on the array(Nuwaysir et al., 2002). This allows for

custom arrays to be manufactured in small volumes at much lower cost than by photolithographic methods using masks to direct light (which are cheaper for large volume production). One constraint with this method is that the total number of addressable pixels (e.g. unit oligos that can be synthesized) is limited to the number of addressable positions in the micro-mirror device (of order 1M).

In 1996, Blanchard et.al. proposed a method use inkjet printing technology and standard oligo synthesis chemistry to produce oligo arrays(Blanchard et al., 1996). In brief, inkjet printer heads were adapted to deliver to the four different nucleotide phosphoramidites to a glass slide that was pre-patterned to contain regions containing hydrophilic regions (with exposed hydroxyl groups) surrounded by hydrophobic regions. The hydroxylated regions provided a surface to which the phosphoramidites could couple, while the surrounding hydrophobic regions contained the droplet(s) emitted by the inkjets to defined regions. This technology was eventually commercialized by Rosetta Inpharmatics (Hughes et al., 2001) and licensed to Agilent Technologies who produces these arrays at present. The inkjet array approach shares the advantage of the Affymetrix/Nimblegen approach in that one only need to have available a small number of reagents to produce an array. In addition, similar to the Nimblegen approach, the production of a new type of array only requires that a different set of sequence information is delivered to the printer. Hence, the inkjet array technology has been particularly useful for the design of custom arrays that are produced in low volume.

Self assembled arrays

An alternative approach to the construction of arrays was created by the group of David Walt at Tufts University(Ferguson et al., 2000; Michael et al., 1998; Steemers et al., 2000; Walt, 2000) and ultimately licensed to Illumina. Their method involved synthesizing DNA on small polystryrene beads and depositing those beads on the end of a fiber optic array in which the ends of the fibers were etched to provide a well that is slightly larger than one bead. Different types of DNA would be synthesized on different beads and applying a mixture of beads to the fiber optic cable would result in a randomly assembled array. In early versions of these arrays, the beads were optically encoded with different fluorophore combinations in order to allow one to determine which oligo was in which position on the array (referred to as “decoding the array”)(Ferguson et al., 2000; Michael et al., 1998; Steemers et al., 2000; Walt, 2000). Optical decoding by fluorescent labeling limited the total number of unique beads that could be distinguished. Hence, the later and present day methods for decoding the beads involve hybridizing and detecting a number of short, fluorescently labeled oligos in a sequential series of steps(Gunderson et al., 2004). This not only allows for an extremely large number of different types of beads to be used on a single array but also functionally tests the array prior to its use in a biological assay. Later versions of the Illumina arrays used a pitted glass surface to contain the beads instead of a fiber option arrays.

The above is not intended to be a comprehensive history or survey of all DNA microarray technologies. However, it does cover the major advances in the field and the predominate methods of manufacture of arrays.

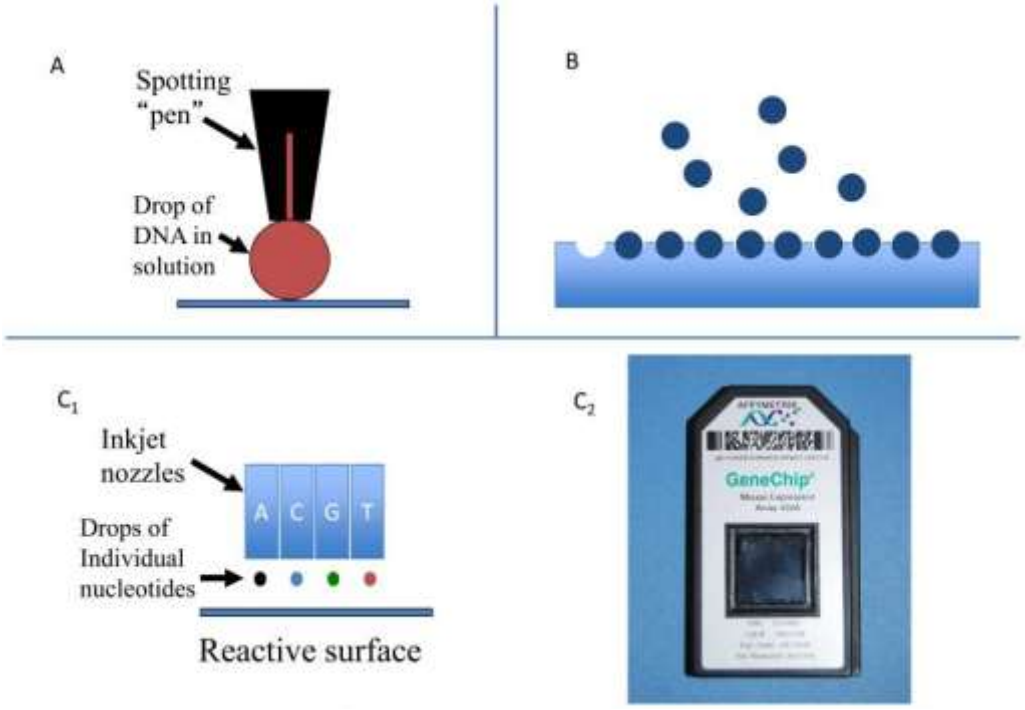


Figure 2

Three basic types of microarrays: (A) Spotted arrays on glass, (B) self assembled *arrays* and (C) *in-situ* synthesized arrays.

A. With spotted arrays, a "pen" (or multiple pens) are dipped into solutions containing the DNA of interest and physically deposited on a 1"x 3" glass microscope slide. Typically the glass slide surface is coated with something to help retain the DNA such as polylysine {DeRisi, 1997 #28191}, a silane {Call, 2001 #28277} or a chemically reactive surface {Rogers, 1999 #28278} (to which chemically reactive oligos or PCR products would be added).

B. Self assembled arrays can be created by applying a collection of beads containing a diverse set of oligos to a surface with pits the size of the beads. After the array is constructed a series of hybridizations determine which oligo is in what position on each unique array (Ferguson et al., 2000; Michael et al., 1998; Steemers et al., 2000; Walt, 2000) (Gunderson et al., 2004).

C1 and C2. In-situ synthesized arrays can be produced by inkjet oligo synthesis methods (C1) or by photolithographic methods such as used by Affymetrix (C2).

Spotted arrays

In 1996 Derisi et. al. published a method which allowed very high-density DNA arrays to be made on glass substrates(DeRisi et al., 1996). Poly-lysine coated glass microscope slides provided good binding of DNA and a robotic spotter was designed to spot multiple glass slide arrays from DNA stored in microtiter dishes. By using slotted pins (similar to fountain pens in design) a single dip of a pin in DNA solution could spot multiple slides. Spotting onto glass, allowed one to fluorescently label the sample. Fluorescent detection provided several advantages relative to the radioactive or chemilluminescent labels common to filter based arrays. First, fluorescent detection is quite sensitive and has a fairly large dynamic range. Second, fluorescent labeling is generally less expensive and less complicated than radioactive or chemilluminescent labeling. Third, fluorescent labeling allowed one to label two (or potentially more) samples in different colors and co-

hybridize the samples to the same array. As it was very difficult to reproducibly produce spotted arrays, comparisons of individually hybridized samples to ostensibly identical arrays would result in false differences due to array-to-array variation. However, a two-color approach in which the ratio of signals on the same array are measured is much more reproducible.

In-situ, Synthesized arrays

In 1991 Fodor et.al. published a method for light directed, spatially addressable chemical synthesis which combined photolabile protecting groups with photolithography to perform chemical synthesis on a solid substrate(Fodor et al., 1991). In this initial work, the authors demonstrated the production of arrays of 10-amino acid peptides and, separately, arrays of di-nucleotides. In 1994, Fodor et.al. at the recently formed company of Affymetrix demonstrated the ability to use this technology to generate DNA arrays consisting of 256 different octa-nucleotides (Pease et al., 1994). By 1995-1996, Affymetrix arrays were being used to detect mutations in the reverse transcriptase and protease genes of the highly polymorphic HIV-1 genome(Lipshutz et al., 1995) and to measure variation in the human mitochondrial genome(Chee et al., 1996). Eventually, Affymetrix used this technology to develop a wide catalogue of DNA arrays for use in expression analysis(Lockhart et al., 1996; Wodicka et al., 1997), genotyping (Chee et al., 1996; Hacia et al., 1996) and sequencing (G Wallraff, 1997)(see www.Affymetrix.com for the current catalog of arrays).

A major advantage of the Affymetrix technology is that because the DNA sequences are directly synthesized on the surface, only a small collection of reagents (the 4 modified nucleotides, plus a small handful of reagents necessary for the de-blocking and coupling steps) are needed to construct an arbitrarily complex array. This contrasts with the spotted array technologies in which one needed to construct or obtain all the sequences that one wished to deposit on the array in advance of array construction. However, the initial Affymetrix technology was limited in flexibility as each model of array required the construction of a unique set of photolithographic masks in order to direct the light to the array at each step of the synthesis process. In 2002, authors from Nimblegen Systems Inc., published a method in which the photo-deprotection step of Fodor et. al (Fodor et al., 1991; Lipshutz et al., 1999) is accomplished using micro-mirrors (similar to those in video computer projectors) to direct light at the pixels on the array(Nuwaysir et al., 2002). This allows for custom arrays to be manufactured in small volumes at much lower cost than by photolithographic methods using masks to direct light (which are cheaper for large volume production). One constraint with this method is that the total number of addressable pixels (e.g. unit oligos that can be synthesized) is limited to the number of addressable positions in the micro-mirror device (of order 1M).

In 1996, Blanchard et.al. proposed a method use inkjet printing technology and standard oligo synthesis chemistry to produce oligo arrays(Blanchard et al., 1996). In brief, inkjet printer heads were adapted to deliver to the four different nucleotide phosphoramidites to a glass slide that was pre-patterned to contain regions containing hydrophilic regions (with exposed hydroxyl groups) surrounded by hydrophobic regions. The hydroxylated regions provided a surface to which the phosphoramidites could couple, while the surrounding hydrophobic regions contained the droplet(s) emitted by the inkjets to defined regions. This technology was eventually commercialized by Rosetta Inpharmatics (Hughes et al., 2001) and licensed to Agilent Technologies who produces these arrays at present. The inkjet array approach shares the advantage of the Affymetrix/Nimblegen approach in that one only need to have available a small number of reagents to produce an array. In addition, similar to the Nimblegen approach, the production of a new type of array only requires that a different set of sequence information is delivered to the printer. Hence, the inkjet array technology has been particularly useful for the design of custom arrays that are produced in low volume.

Self assembled arrays

An alternative approach to the construction of arrays was created by the group of David Walt at Tufts University(Ferguson et al., 2000; Michael et al., 1998; Steemers et al., 2000; Walt, 2000) and ultimately licensed to Illumina. Their method involved synthesizing DNA on small polystryrene beads and depositing those beads

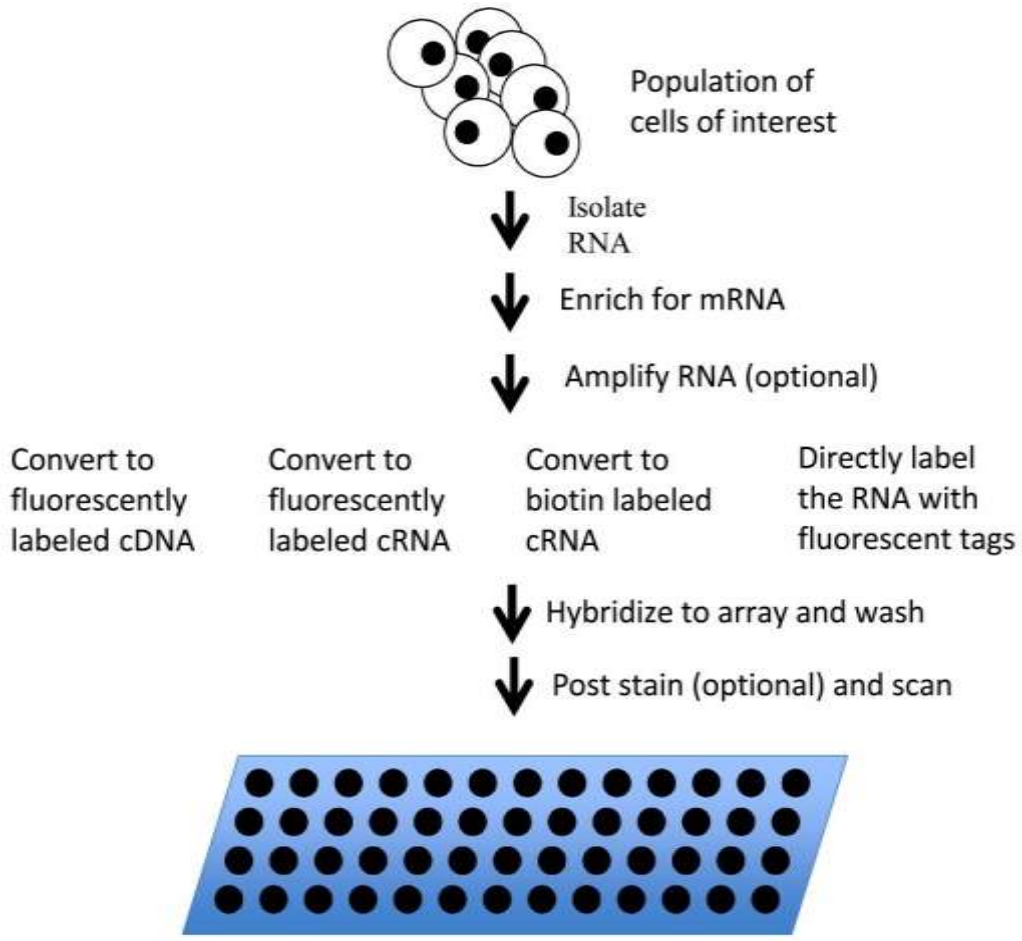
on the end of a fiber optic array in which the ends of the fibers were etched to provide a well that is slightly larger than one bead. Different types of DNA would be synthesized on different beads and applying a mixture of beads to the fiber optic cable would result in a randomly assembled array. In early versions of these arrays, the beads were optically encoded with different fluorophore combinations in order to allow one to determine which oligo was in which position on the array (referred to as "decoding the array")(Ferguson et al., 2000; Michael et al., 1998; Steemers et al., 2000; Walt, 2000). Optical decoding by fluorescent labeling limited the total number of unique beads that could be distinguished. Hence, the later and present day methods for decoding the beads involve hybridizing and detecting a number of short, fluorescently labeled oligos in a sequential series of steps(Gunderson et al., 2004). This not only allows for an extremely large number of different types of beads to be used on a single array but also functionally tests the array prior to its use in a biological assay. Later versions of the Illumina arrays used a pitted glass surface to contain the beads instead of a fiber option arrays.

The above is not intended to be a comprehensive history or survey of all DNA microarray technologies. However, it does cover the major advances in the field and the predominate methods of manufacture of arrays.

Lecture 11 . Applications of microarrays in gene expression analysis

Gene expression analysis

The predominate application of DNA microarrays has been to measure gene expression levels (Figure 3). In this application, RNA is extracted from the cells of interest and either, labeled directly, converted to a labeled cDNA or converted to a T7 RNA promoter tailed cDNA which is further converted to cRNA through the Eberwine amplification process (Van Gelder et al., 1990). A wide variety of methods have been developed for labeling of the cDNA or cRNA including: incorporation of fluorescently labeled nucleotides during the synthesis, incorporation of biotin labeled nucleotide which is subsequently stained fluorescently labeled streptavidin, incorporation of a modified reactive nucleotide to which a fluorescent tag is added later, and a variety of signal amplification methods (an early review of different labeling methods is provided in (Richter et al., 2002)). The two most frequently used methods are the incorporation of fluorescently labeled nucleotides in the cRNA or cDNA synthesis step or the incorporation of a biotin labeled nucleotide in the cRNA synthesis step (as is done by Affymetrix).



The labeled cRNA or cDNA are then hybridized to the microarray, the array is washed and the signal is detected by measuring fluorescence at each spot. In the case of biotin labeled samples, the array is stained post-hybridization with fluorescently labeled streptavidin. Laser induced fluorescence is typically measured with a scanning confocal microscope. The intensity of the signal(s) on each spot is taken as a measure of the expression level of the corresponding gene. Gene expression analysis is described in more detail in Units 22.2-22.4.

Transcription factor binding analysis

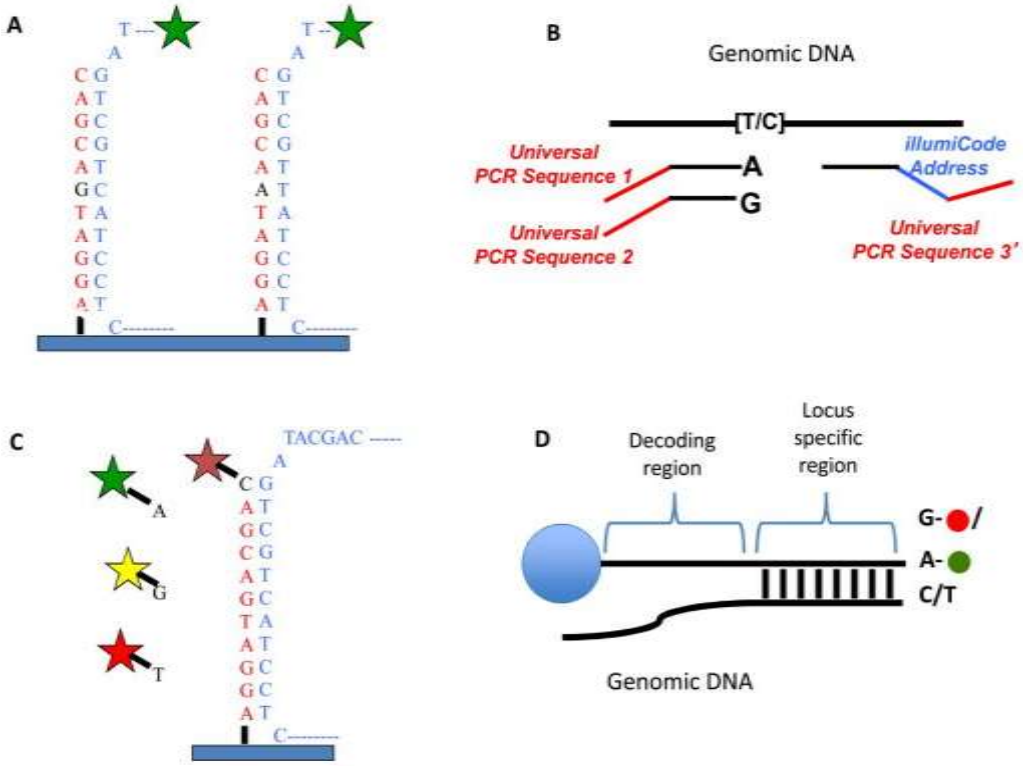
Microarrays have also been used in combination with chromatin immunoprecipitation (Solomon et al., 1988) to determine the binding sites of transcription factors (Horak and Snyder, 2002; Iyer et al., 2001). In brief, transcription factors (TFs) are cross linked to DNA with formaldehyde and the DNA is fragmented. The TF(s) of interest (with the DNA to which they were boud still attached) are affinity purified using either an antibody to the TF or by tagging the transcription factor with peptide that's amenable to affinity chromatography (for example a FLAG-, HIS-, myc or HA-tag). After purification, the DNA is released from the TF, amplified, labeled and hybridized to the array. This technique is commonly referred to as "ChIP-chip" for Chromatin Immuno-Precipitation on a "chip" or microarray.

As TF's often bind quite a distance away from the genes that they regulate, the design of the array and size distribution of the fragment length are interrelated. E.g. the array must contain probes that will interrogate the region of DNA bound to the transcription factor. For bacteria or yeast, the intergenic regions are fairly small and the same arrays used for gene expression work can be applied to ChIP-chip. For mammalian genomes, the intergenic regions are large and the TF often bind many kbp away from the gene of interest. Hence, for

mammalian genomes, oligo arrays with oligo's spaced evenly across the entire genome are typically used for ChIP-chip experiments. Buck et. al. provide a good review(Buck and Lieb, 2004) of the considerations for the design and analysis of ChIP-chip experiments and the technique is discussed in detail in Units 21.9 and 21.13.

Genotyping

Microarrays have been widely used as single-nucleotide-polymorphism (SNP) genotyping platforms. Several alternative approaches have been used to detect SNP's but the most commonly used are allele discrimination by hybridization as used by Affymetrix (Wang et al., 1998), allele specific extension and ligation to a "bar-code" oligo which is hybridized to a universal array (the Illumina "Golden Gate Assay"(Fan et al., 2003)) or approaches in which the arrayed DNA is extended across the SNP in a single nucleotide extension reaction (the Arrayed Primer Extension assay of Kurg et.al. (Kurg et al., 2000) or the Infinium Assay of Illumina (Gunderson et al., 2006)). Figure 4 explains the detection approaches in more detail. Allelic discrimination by hybridization suffers background due to non-specific hybridization in complex genomes. In order to reduce this background, Affymetrix developed a PCR based approach to reduce genomic complexity.



igure 4

SNP detections strategies for arrays. A) Allele discrimination by hybridization – Oligos that are complimentary to each allele are placed on the array and labeled genomic DNA is hybridized to the array. The variant position is placed in the center of the oligo (typically 25bp on Affymetrix arrays) as this position has the greatest affect on hybridization. Typically, multiple array positions are used for each allele to improve signal to noise. B) Illumina's "Golden Gate Assay"- two allele specific oligos are each tailed with a different universal primer (1 and 2) and hybridized in solution to genomic DNA. A third oligo that is complementary to the same locus is tailed with a "barcode" sequence and a third universal primer (3). Polymerase is used to extend the allele specific primers across the genomic sequence and the extended products are ligated to the third oligo. PCR is performed using primers complimentary to universal sequences 1, 2 and 3. The PCR primers complimentary to the universal sequences 1 and 2 are labeled with a unique fluorophore. The barcode se-

quence on the third oligo allows the PCR product to be uniquely detected on an array containing oligos complimentary to the barcode sequence. The use of multiple barcodes (one for each locus of interest) allows the assay to be multiplexed to sample many loci. C) Arrayed primer extension (APEX) – In this assay, the array contains DNA oriented with the 5′ end attached to the array and the 3′ end stopping one nucleotide short of the SNP. Genomic DNA is fragmented and hybridized to the array and the oligo on the array is extended in single nucleotide dye terminator sequencing reaction. D) Illumina's Infinium assay – This assay is similar to the APEX assay except that the oligo to be extended is on a bead and the single nucleotide that is added is labeled with a nucleotide specific hapten as opposed to a fluorophore. The haptens are then detected by staining with fluorescently labeled proteins that bind each hapten.

Collection and analysis

The collection of data is done by using a microarray scanner. This scanner consists of a laser, a computer, and a camera. The laser excites fluorescence of the cDNA, generating signals.

When the laser scans the array, the camera records the images produced. Then the computer stores the data and provides the results immediately. The data thus produced are then analyzed. The difference in the intensity of the colors for each spot determines the character of the gene in that particular spot.

1. **Creation of labeled cDNA -** to create cDNA (complementary DNA strand), reverse transcription of the mRNA is done. Both the samples are then incorporated with different fluorescent dyes for producing fluorescent cDNA strands. This helps in distinguishing the sample category of the cDNAs.
2. **Hybridization -** the labeled cDNAs from both the samples are placed in the DNA microarray so that each cDNA gets hybridized to its complementary strand; they are also thoroughly washed to remove unbounded sequences

Applications

To study transcriptomes and proteomes
To diagnose pathogenic as well as genetic diseases in man
To identify microbes in the environment with the help of species-specific probes
To genotype genomes through single nucleotide polymorphism (SNP) analysis
To detect gene expression of mRNAs of a particular cell at different times
To measure changes in the level of gene expression
To observe DNA mutations
To study genomic gains and losses

Types of DNA microarrays

DNA microarrays are of four types:

a. cDNA microarrays: uses complementary DNA strands formed by transcription of mRNA
b. Oligo DNA microarrays: uses chemically synthesized oligo DNA as probes
c. BAC microarrays: uses template amplified by polymerase chain reaction as the probe
d. SNP microarrays: used to detect polymorphisms within a population

Data standards and data exchange

With the exception of DNA sequencing, microarrays were perhaps the earliest technology that allowed biologists to vast amounts of complex digital data. As the technology came into use, it rapidly became apparent that in order for others to be able to reproduce a given microarray experiment a detailed description of the array, the sample, the protocols and the data analysis methods needed to be available. Moreover, it also became apparent that access to the raw and processed data would allow others to perform analyses and meta analyses (on combinations of data) that the original data producers had not conceived. To address these issues of reproducible science and data exchange, members of the Microarray Gene Expression Data Society (now the Function Genomics Data Society – www.FGED.org) created the MIAME (Minimum Information About a Microarray Experiment) standards for the description of microarray experiments(Ball and Brazma, 2006; Brazma et al., 2001) and for the exchange of microarray data(Rayner et al., 2006; Spellman et al., 2002). These efforts influenced the creation public databases for microarray data (Barrett et al., 2007; Brazma et al., 2006; Brazma et al., 2003) and subsequent standards efforts in other areas (Deutsch et al., 2008; Field et al., 2008; Taylor et al., 2007).

Transcription factor binding analysis

<

LECTURE 13. ***Different types of PCR in molecular biotechnology.***

Polymerase chain reaction (PCR) is an efficient and cost-effective molecular tool to copy or amplify small segments of DNA or RNA. PCR combines the principles of complementary nucleic acid hybridization with those of nucleic acid replication that are applied repeatedly through numerous cycles. It results in the exponential production of the specific target DNA/RNA sequences **by a factor of 10^7 within a relatively short period.**

This **in vitro amplification technique** can amplify a single copy of nucleic acid target by using two synthetic oligonucleotide "primers" that bind to the target genomic sequence, which are extended by a Taq polymerase (a thermostable DNA polymerase) . An automated process of repeated cycles (usually 25 to 40) of denaturation of the template DNA (at 94°C), annealing of primers to their complementary sequences (50°C), and primer extension (70°C) is employed for the amplification of target sequence.

PCR was originally developed in 1983 by the American biochemist and Nobel Laureate Kary Mullis.

Primer: *A short segment of nucleotides, which is complementary to a section of the DNA or RNA, which is to be amplified in the PCR. Two short DNA sequences designed to bind to the start (forward primer) and end (reverse primer) of the target sequence is used in PCR.*

Taq polymerase: *A thermally stable DNA polymerase originally isolated from the thermophilic bacterium Thermus aquaticus, which resist inactivation during denaturation temperatures and allows primer extension at high temperature.*

Components of Polymerase Chain Reactions (PCR)

- DNA template (the sample DNA that contains the target sequence to amplify)
- Deoxyribonucleoside triphosphates (dNTPs)
- PCR buffer
- Primers (forward and reverse)
- Taq polymerase

Steps of polymerase chain reaction-PCR

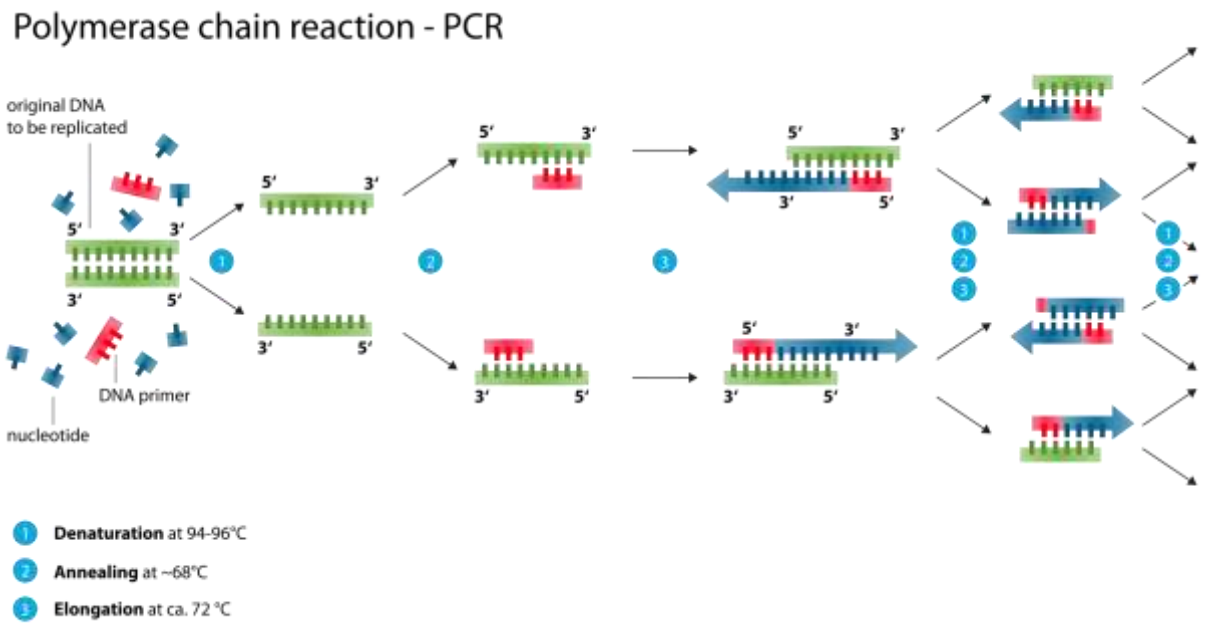
To perform PCR, extracted sample (which contains target DNA template) is added to a tube containing primers, free nucleotides (dNTPs), and Taq polymerase. The PCR mixture is placed in a PCR machine. PCR machine increases and decreases the temperature of the PCR mixture in automatic, programmed steps which generates copies of the target sequence exponentially.

Polymerase Chain Reaction (PCR) has **three major steps.**

Steps of polymerase chain reaction-PCR

To perform PCR, extracted sample (which contains target DNA template) is added to a tube containing primers, free nucleotides (dNTPs), and Taq polymerase. The PCR mixture is placed in a PCR machine. PCR machine increases and decreases the temperature of the PCR mixture in automatic, programmed steps which generates copies of the target sequence exponentially.

Polymerase Chain Reaction (PCR) has **three major steps.**



Types of polymerase chain reaction-PCR

Steps of Polymerase Chain Reactions (PCR)

1. **Denaturation (strand separation)** : The separation of the two hydrogen-bonded complementary chains of DNA into a pair of single stranded polynucleotide molecules by a process of heating (94°C to 96°C)
2. **Annealing (primer binding):** The temperature is lowered (45-60 °C) so the primers can attach themselves to the single stranded DNA strands.
3. **Extension (synthesis of new DNA):** It starts at the annealed primer and works its way along the DNA strand (72°C).

Once the first round is completed, the process is repeated by cycling back to the first reaction temperature and next round of denaturation, annealing and extension is started*(an automatic process in thermocycler)*. This 3 steps temperature cycle is repeated approximately 30 times which results exponential amplification of target gene sequence.

Several modification of PCR methods have been developed to enhance the utility of this method in diagnostic settings based on their applications. Some of the common types of PCR are;

1. Real-Time PCR
2. Nested PCR
3. Multiplex PCR
4. Quantitative PCR
5. Arbitrary Primed PCR

Detection of PCR products

Labeled probe that is specific for the target gene sequence is used to detect PCR amplified gene product (also known as amplicon). Based on the nature of the reporter molecule used, probe generates radioactive, colorimetric, fluorometric, or chemiluminescent signals. Probe based detection of amplicons serves two purposes

1. It allows visualization of the PCR product
2. It provides specificity by ensuring that the amplicon is the target sequence of interest and not the result of non-specific amplification.

Apart from DNA based hybridization method, sometimes simple gel electrophoresis method is sufficient to confirm the presence of specific amplicons.

Applications of PCR

- Identification and characterization of infectious agents
    - Direct detection of microorganisms in patient specimens
    - Identification of microorganisms grown in culture
    - Detection of antimicrobial resistance
    - Investigation of strain relatedness of pathogen of interest
- Genetic fingerprinting (forensic application/paternity testing)
- Detection of mutation (investigation of genetic diseases)
- Cloning genes
- PCR sequencing

LECTURE 14 . *Molecular markers types and applications*

1. 1. Utilization of Molecular Markers for PGRFA Characterization and Pre-Breeding for Climate Changes Aug. 31st- Sept. 4th, 2014
2. Based on Hybridization Markers Morphological Biochemical Molecular Based on PCR RFLP Mini-satallite Microsatallite Isozyme Protein Banding Pattern RAPD SSRISSR AFLPSTS SCARSCoT ESTAP-PCR
3. Molecular Cytogenetic Techniques Types of Molecular Markers Molecular Markers Techniques FISH Extended DNA fiber–FISH RAPD SSR AFLP SCoT RFLP EST
4. Morphological markers
5. • Assessment Genetic diversity in a crop species is fundamental to its improvements. • Genetic variability is considered the reservoir that plant breeders fall upon in their continuous strive to develop improved varieties and hybrids. • Knowledge of germplasm diversity and of relationship among elite breeding materials has a significant impact improvement of crop plants. • This information is useful in planning crosses for hybrid and line development, in assigning lines to heterotic groups. Genetic Diversity
6. •Limited genomic coverage
7. Molecular Markers They may be due to: • Base pair changes. • Rearrangements (translocation or inversion). • Insertions or deletions. • Variation in the number of tandem repeats. Reflect heritable differences in homologous DNA sequences among individuals.
8. ⎪ Ubiquitous. ⎪ Stably inherited. ⎪ Multiple alleles for each marker. ⎪ Devoid of pleiotropic effects. ⎪ Detectable in all tissues, at all ages. ⎪ Long shelf life of the DNA samples. Advantages of Molecular Markers
9. Hybridization based (non-PCR) Technique RFLPs Restriction Fragment Length Polymorphism analysis Botstein et al. (1980)
10. Genetic markers resulting from the variation or change in the length of defined DNA fragments produced by digestion of the DNA sample with restriction endonucleases RFLPs :
11. RFLPs (restriction fragment length polymorphisms) Electrophoretic comparison of the size of defined restriction fragments derived from genomic DNA 1. Isolate high quality DNA 2. Digest with a combination of restriction enzymes 3. Fractionate digested samples by electrophoresis 4. Transfer fragments to membrane 5. Hybridize with radioactively labeled DNA probe(s); detect by autoradiography. Can also use non-radioactive labeling systems
12. RFLP analysis Polymorphism revealed by different probe/enzyme combinations among 13 different accessions.
13. Considerations for use of RFLPs - Relatively slow process -Use of radioisotopes has limited RFLP use to certified laboratories (but non-radioactive labeling systems are now in wide use) - Co-dominant markers; often species-specific - Need high quality DNA - Need to develop polymorphic probes - expensive
14. PCR based techniques (RAPD, ISSR, SSR, AFLP, EST ,SCoT)
15. Random Amplified Polymorphic DNA (RAPD) Randomly Amplified Polymorphic DNA (RAPDs) are genetic markers resulting from PCR amplification of genomic DNA sequences recognized by ten-mer random primers of arbitrary nucleotide sequence (Williams et al., 1990). RAPDs are dominant markers that require no prior knowledge of the DNA sequence, which makes them very suitable for investigation of species that are not well known (Williams et al. 1993).
16. RAPD profiles for the 14 Date Palm accessions as detected with primers OPB-06 (A), OPB-08 (B), OPB-11 (C), and OPO-07 (D). Lanes 1 to 14 represent: SAK-AK, SAK-AB, BRT-AK, BRT-AB, MLK-AK, MLK-AB, GND-AK, GND-AB, SIW-KH, SIW-DK, SIW-HB, SIW-TZ, FRA-HB and FRA-TZ. M: 1 Kb ladder DNA marker.
17. Inter-Simple Sequence Repeats (ISSR) The generation of ISSR markers involve PCR amplification of DNA using a single primer composed of a microsatellite repeated sequence and in some cases primer also contains 1- 4 base anchor at either 3′ or 5′ or at both ends, which target a subset of ‘simple sequence repeats’ (SSRs) and amplify the region between two closely spaced and oppositely oriented SSRs (Fang et al., 1997; Fang and Roose, 1997; Moreno et al., 1998). ISSR technique permits the detection of polymorphisms in microsatellites and inter-microsatellites loci without previous knowledge of the DNA sequence (Moreno et al., 1998).

18. ISSR profiles of the 14 Date palm accessions using the primers: IS3 (A), IS4 (B), IS7 (C), IS9 (D). M: 1 Kb ladder DNA marker. Lanes 1 to 14 represent: SAK-AK, SAK-AB, BRT-AK, BRT-AB, MLK-AK, MLK-AB, GND-AK, GND-AB, SIW- KH, SIW-DK, SIW-HB, SIW-TZ, FRA-HB and FRA-TZ.
19. Are DNA sequences with repeat lengths of a few base pairs. Variation in the number of repeats can be detected with PCR by developing primers for the conserved DNA sequence flanking the SSR. As molecular markers, SSR combine many desirable marker properties including high levels of polymorphism and information content, unambiguous designation of alleles, even dispersal, selective neutrality, high reproducibility, co-dominance, and rapid and simple genotyping assays. Microsatellites have become the molecular markers of choice for a wide range of applications in genetic mapping and genome, genotype identification and variety protection, seed purity evaluation and germplasm conservation, diversity studies, paternity determination and pedigree analysis, gene and quantitative trait locus analysis, and marker-assisted breeding. Simple Sequence Repeats (SSR)
20. A B C D FE M 1 20 M 1 20 200 bp 100 bp Photograph of EtBr stained polyacrylamide gels of polymorphic SSR products from 20 maize inbred lines as detected by SSR primers (A: M28, B: M27, C: M25, D: M20, E: M18 and F: M22). M (100bp DNA ladder)
21. AFLP (Amplified Fragment Length Polymorphisms) ¬ A combination of PCR and RFLP ¬ Informative fingerprints of amplified fragments
22. Amplified Fragment Length Polymorphism (AFLP) AFLP process 1. Digest genomic DNA with restriction enzymes 2. Ligate commercial adaptors (defined sequences) to both ends of the fragments 3. Carry out PCR on the adaptor-ligated mixture, using primers that target the adaptor, but that vary in the base(s) at the 3' end of the primer. AFLP technology is a DNA fingerprinting technique that combines RFLP and PCR. It is based on the selective amplification of a subset of genomic restriction fragments using PCR.
23. AFLP profiles of the 14 Date Palm accessions as revealed by the primer combination Eact X Mcta. (A) and the primer combination Eagc X Mcaa. (B). Lanes 1 to 14 represent: SAK-AK, SAK-AB, BRT-AK, BRT-AB, MLK-AK, MLK-AB , GND-AK, GND-AB, SIW-KH, SIW-DK, SIW-HB, SIW-TZ, FRA-HB and FRA-TZ. M: DNA molecular weight marker (100 bp Ladder).
24. Advantages of AFLP's ¬ Very sensitive ¬ Good reproducibility but technically demanding ¬ Relatively expensive technology ¬ Discriminating homozygotes from heterozygotes ¬ Requires band quantitation (comparison of pixel density in images from a gel scanner) ¬ Bands are anonymous
25. SCOT (Start codon targeted Marker)
26. Start Codon Targeted (SCoT) Polymorphism analysis SCoT is a novel method for generating plant DNA markers. This method was developed based on the short conserved region flanking the ATG start codon in plant genes. SCoT uses single 18-mer primers in polymerase chain reaction (PCR) and an annealing temperature of 50°C. PCR amplicons are resolved using standard agarose gel electrophoresis. This method was validated in rice using a genetically diverse set of genotypes and a backcross population. Diagram showing principle of SCoT PCR amplification
27. SCoT Analysis SCoT analyses were performed as described by Collard and Mackil (2009). Table : SCoT primers and their sequences
28. SCoT profiles of the two parental genotypes as revealed by different primers M P1 P1 P2 P2 P1 P1 P2 P2 P1 P1 P2 P2 P1 P1 P2 P2 P1 P1 P2 P2 P1 P1 P2 P2 P1 P1 P2 P2 P1 P1 P2 P2 P1 P1 P2 P2
29. ESTs (Expressed Sequence Tags) Single-pass sequencing reads from randomly selected cDNA clones 3'EST sequence 5'EST sequence cDNA Clone dbEST May 7, 2003 21,265,083 ESTs from 611 species
30. Steps for EST's •cDNA libraries (containing many of the expressed genes of an organism) •pick cDNA clones randomly •rapidly determine some of the sequence of nucleotides from the end of each clone. •These ESTs could then be compared to all known sequences using a program called BLAST.
31. An exact match to a sequenced gene means that the gene encoding that EST is already known. If the match was close but not exact one could conclude that the EST is derived from a gene with a function similar to that of the known gene. The EST sequences with their putative identification are then deposited in the GenBank and the clones from which they were derived are kept in a freezer for later use.

32. Overview of the EST sequencing process Clones are picked from petrie dishes into microtitre plates, and archived for later use. All subsequent manipulations (PCR, clean up and sequencing) are carried out in microtitre plates to yield medium-throughput.
33. Features RFLP PCR- RFLP DFP RAPD Microsatellite SNP Detection method Hybridization PCR Hybridization PCR PCR PCR Type of probe/primer used g DNA/ cDNA sequence of structural genes Sequence specific primers Mini satellite synthetic oligos Arbitrarily design primer Sequence specific primers Sequence specific primers Requirement of radioactivity Yes No/Yes Yes No/Yes No/Yes No/Yes Extant of genomic coverage Limited Limited Extensive Extensive Extensive Extensive Degree of polymorphisms Low Low High Medium to High High High Phenotype expression Co dominant Co dominant Co dominant Co dominant/D ominant Dominant Co dominant Possibility of automation No Yes No Yes Yes Yes
34. DATA ANALYSIS
35. 765432 000110 011000 111101 100111 111111 011000 111111 100111 111111 000111 111111 Scoring of bands
36. 653742 1002 10094.14 10087.593.37 10080.094.187.53 10053.371.462.566.75 10092.357.161.553.357.16 Genetic Similarity matrix calculated according to Jaccard's coefficient based on marker data.
37. Dendrogram constructed with UPGMA cluster analysis of marker data showing the genetic relationships among the different samples. 0.50 0.70 0.80 0.90 1.00 2 3 4 55 6 7
38. ▪ Fingerprinting . ▪ Diversity studies . ▪ Marker-assisted selection . ▪ Genetic maps . ▪ Gene tagging . ▪ Novel allele detection . ▪ Map-based gene colning . ▪ F1 identification . ▪ Comparative maps . ▪ Bulk segregant analysis . ▪ Seed testing . DNA marker applications

- **LECTURE 15 .** QTL analysis
- A quantitative trait locus (QTL)
- is a section of DNA (the **locus**) which correlates with variation in a phenotype (the quantitative trait).
- QTLs are mapped by identifying which molecular markers (such as ***SNPs*** or ***AFLPs***) correlate with an observed trait.
- This is an early step in identifying and sequencing the actual genes that cause the trait variation.
- A **quantitative trait locus** (**QTL**)
- **is a region of** *DNA* **which is associated with a particular phenotypic trait, which varies in degree and which can be attributed to polygenic effects.**
- **Polygenic effects is the product of two or more genes, and their environment.**
- **These QTLs are often found on different chromosomes.**
- **The number of QTLs which explain variation in the phenotypic trait indicates the genetic architecture of a trait.**
- **It may indicate that plant height is controlled by many genes of small effect, or by a few genes of large effect.**
- A **quantitative trait locus** (**QTL**)
- Typically, QTLs underlie continuous traits (those traits which vary continuously, e.g. height) as opposed to discrete traits (traits that have two or several character values, e.g. red hair in humans, a recessive trait, or smooth vs. wrinkled peas used by Mendel in his experiments).

- Moreover, a single phenotypic trait is usually determined by many genes.
- Consequently, many QTLs are associated with a single trait.
- **Another use of QTLs is to identify candidate genes underlying a trait.**
- **Once a region of DNA is identified as contributing to a phenotype, it can be sequenced.**
- **The DNA sequence of any genes in this region can then be compared to a database of DNA for genes whose function is already known.**
- **Quantitative traits**
- **Polygenic inheritance** is to inheritance of a ***phenotypic*** characteristic (trait) that is attributable to two or more genes and can be measured quantitatively.
- **Multifactorial inheritance** is to polygenic inheritance that also includes interactions with the environment.
- Unlike monogenic traits, polygenic traits do not follow patterns of Mendelian inheritance (discrete categories).
-
- **QTL mapping**
- For organisms whose genomes are known, one might now try to exclude genes in the identified region whose function is known with some certainty not to be connected with the trait in question.
- If the genome is not available, it may be an option to sequence the identified region and determine the putative functions of genes by their similarity to genes with known function, usually in other genomes.
- This can be done using ***BLAST***, an online tool.
- BLAST allows to enter a primary sequence and search for similar sequences within the BLAST database of genes from various organisms.
- It is often not the actual gene underlying the phenotypic trait, but rather a region of DNA that is closely linked with the gene.
- **BLAST for Basic Local Alignment Search Tool is an algorithm for comparing primary biological sequence information,**
- **such as the amino-acid sequences of proteins or**
- **the nucleotides of DNA sequences**
- **QTL mapping**
- **In a recent development, classical QTL analyses were combined with gene expression profiling i.e. by DNA microarrays.**
- **Such expression QTLs (eQTLs) describe cis- and trans-controlling elements for the expression of often disease-associated genes.**
- **Observed epistatic effects have been found beneficial to identify the gene responsible by a cross-validation of genes within the interacting loci with metabolic pathway- and scientific literature databases.**

- **Epistasis** is the phenomenon where the effect of one gene (locus) is dependent on the presence of one or more 'modifier genes', i.e. the **genetic background**.
- Originally the term meant that the phenotypic effect of one gene is masked by a different gene (locus).
- **Analysis of variance**
- The simplest method for QTL mapping is analysis of variance (ANOVA, sometimes called "marker regression") at the marker loci.
- In this method, in a backcross, one may calculate a t-statistic to compare the averages of the two marker genotype groups.
- For other types of crosses (such as the intercross), where there are more than two possible genotypes, one uses a more general form of ANOVA, which provides a so-called F-statistic.
- **Analysis of variance**
- The ANOVA approach for QTL mapping has three important weaknesses.
- 1) we do not receive separate estimates of QTL location and QTL effect. QTL location is indicated only by looking at which markers give the greatest differences between genotype group averages, and the apparent QTL effect at a marker will be smaller than the true QTL effect as a result of recombination between the marker and the QTL. 2), we must discard individuals whose genotypes are missing at the marker.
- 3) , when the markers are widely spaced, the QTL may be quite far from all markers, and so the power for QTL detection will decrease.
- **Interval mapping**
- Lander and Botstein developed interval mapping, which overcomes the three disadvantages of analysis of variance at marker loci. Interval mapping is currently the most popular approach for QTL mapping in experimental crosses.
- The method makes use of **a genetic map** of the typed markers, and, like analysis of variance, assumes the presence of a single QTL.
- In interval mapping, each locus is considered one at a time and the logarithm of the odds ratio (LOD score) is calculated for the model that the given locus is a true QTL.
- The odds ratio is related to the Pearson correlation coefficient between the phenotype and the marker genotype for each individual in the experimental cross.
- **Interval mapping**
- The term ‘interval mapping’ is used for estimating the position of a QTL within two markers (often indicated as ‘marker-bracket’). Interval mapping is originally based on the maximum likelihood but there are also very good approximations possible with simple regression.
- The principle for QTL mapping is:
- 1) The Likelihood (*Вероятность*) can be calculated for a given set of parameters (particularly QTL effect and QTL position) given the observed data on phenotypes and marker genotypes.
- 2) The estimates for the parameters are those where the likelihood are highest.

- 3) A significance threshold (*порог*) can be established by permutation (*перестановка*) testing.
- **Interval mapping**
- **Conventional methods for the detection of quantitative trait loci (QTLs) are based on a comparison of single QTL models with a model assuming no QTL. For instance in the "interval mapping" method the likelihood for a single putative QTL is assessed at each location on the genome. However, QTLs located elsewhere on the genome can have an interfering effect.**
- **Interval mapping**
- As a consequence, the power of detection may be compromised, and the estimates of locations and effects of QTLs may be biased (LANDER and BOTSTEIN 1989; KNAPP 1991). Even nonexisting so-called "ghost" QTLs may appear (HALEY and KNOTT 1992; MARTINEZ and CURNOW 1992). Therefore, it is obvious that multiple QTLs could be mapped more efficiently and more accurately by using multiple QTL models. One popular approach to handle QTL mapping where multiple QTL contribute to a trait is to iteratively scan the genome and add known QTL to the regression model as QTLs are identified. This method, termed composite interval mapping determine both the location and effects size of QTL more accurately than single-QTL approaches, especially in small mapping populations where the effect of correlation between genotypes in the mapping population may be problematic
- Introduction
- QTL analyses are specialized techniques that construct genetic linkage maps to locate loci (QTLs) that affect a quantitative trait and estimate the genetic effect of QTLs on the trait.
- These techniques range from the relatively simple Single Marker Analysis, to the more complex Composite Interval Mapping, where multiple linked markers are taken into consideration.
- JMP Genomics has three different methodologies for quantitative trait loci (QTL) analysis:
- Single Marker Analysis
- Interval Mapping (IM)
- Composite Interval Mapping (CIM).