BRIEF CONTENT OF LECTURES ON DISCIPLINE “MODERN METHODS OF BIOTECHNOLOGY

Specialty McS Biotechnology

Lector: prof. Kenzhebayeva S.S.

- L. 1. Modern methods of BT
- The subject of modern BT methods
  The use of methods in different BT areas.
  Methods of study of membrane structures
  Development of methods for cell disruption

- BT includes
- new methods of studies of living organisms
- new application of results obtained by using these methods,
  new philosophical and methodological approaches to study the living organisms (often said that biotechnology has created a new "ideology ").
- Modern BT is interdisciplinary science . Methods used in BT directed to study:
- the physical,
- chemical,
- biophysical,
- biochemical properties and principles of the functioning of living systems.
- Among them, there are methods that are used for a long time (centrifugation , spectrophotometry , various types of chromatography , immunological methods , etc.).
- However, these methods still remain effective in the study of biological objectives.
  Innovative BT high-tech methods:
  using of them allows to obtain the results dramatically accelerated progress in biology and BT.
- The most significant of these modern methods
  genetic engineering is changing the target of genes of an organism by manipulating its DNA. Methods of transferring genetic material from one object to another.
  At the present time In experiment any combination of genetic elements obtained from different biological systems can be received.
- 3. Engineering enzymology
- Is a change of the enzymes properties with a view to their use in the food, pharmaceutical and chemical industries.
- For example, with the use of enzyme glycoseizomerase converting glucose into fructose assumed to get commercially the end product (diabetics).
  3. One of the most important approaches - immobilization of enzymes or them containing cells - fastening on cellulose, collagen gel-forming materials, or other media. In the immobilized state enzyme function is improved or modified and increased their stability.
- 4. Methods of sequencing the nucleotide sequences of the NA.
- This allows to study the primary structure of eukaryotic genomes with a size of tens of billions of base pairs.
  5. Methods of analysis of the functional role of genes and other DNA sites.
- This allows to establish differential activity of the genome during ontogeny
- 6. Cultivation of plant or animal cells on nutrient media. It is necessary for mass production of products (for example , medicinal ginseng glycosides cultivator with cells of the plant). From cultured plant cells, whole plants can be obtained which completely identical in hereditary traits (clones) .

- 7. industrial production of
  biological products on a large scale (e.g. , bacterial preparations for animal feed), or vice versa , in miniscule saturating the world market amounts (expensive drugs produced in gram amounts or even milligram),
- the development of industrial biotechnology apparatus (fermentors , bioreactors) and production processes

- 8. Isolation, purification , chemical modification and stabilization of biotechnology products using modern methods (ion exchange , affinity and gel chromatography , electrophoresis, isoelectric focusing, isotachophoresis , immunochemical methods, etc.)
- 9. Biotechnology ecosystem (eco-engineering)
  10 . The enormous potential of computer-based methods of studying NC , proteins and regulatory systems of cells and organisms
-
- So, The rapid progress in various fields of science , related to biology , will lead to an improvement in the appearance of old and new techniques that will significantly accelerate the development of BT .
- In this regard, we must know the basic methods of BT , which determine the development of the scientific and practical activities of man.
- BT methods used in the following areas:
-

  1. Agriculture .
  BT pesticides, replacing pesticides, for example, the use of natural enemies of insects - pests or weeds;
- Development of resistant pathogens , or virus-free plants;
- The creation of new breeds of genetically engineered
- Microbial biomass and other feed additives for animals;
  New methods of prevention (genetically engineered vaccines) and treatment of diseases of farm animals.

• 2 . **Medicine**

• provides new antibiotics, vaccines, therapeutic serum -based monoclonal antibodies, hormones, blood factors synthesized in microbial cultures using genetic engineering techniques and others.

- 3. **Food processing industry**
  • sweeteners , aromatics and flavorings Supplements (добавки), is not obtained by chemical synthesis, using a microorganism or cultured plant cells ;
  • food enzymes ;
  • Regulators - substances prolonging the shelf life of products.
-
  4. Energetics - production of renewable fuels.
- • 5 . **Mining (BIOGEOTECHNOLOGY)** leaching of metals from ores by microorganisms , microbial separation of water-oil emulsions;
  • 6 . **The recovery of residual oil** from the wells by injecting therein viscous solutions microbial biopolymers.
- **5. Горнодобывающая промышленность** (биогеотехнология): выщелачивание металлов из руд с помощью микроорганизмов; микробное разделение водно-нефтяных эмульсий;
- 6. **Извлечение остаточной нефти** из скважин путем закачивания в них вязких растворов микробных биополимеров. **5.**

L. 2. **Methods of identification of subcellular fractions.**

1. MICROSCOPY.

2. Variations to bright field (transmission) microscopy
3.
The microscope is still an extremely important tool in BT research.
The light microscope has a limited capability in regards to the size of a particle that can be examined.
The electron microscope provides additional resolution that allows for the examination of **subcellular structures and even molecules.**
Major components of a light microscope
Sample Preparation
Specimens can be examined by simply placing them on a glass microscope slide under a glass cover slip. However, it is usually necessary to prepare and stain the samples before examination by microscopy.
Fixation is a process by which cells are preserved and stabilized.
Common fixatives include: acids, organic solvents, formaldehyde and glutaraldehyde
The image generated by microscopy depends
upon different components in the sample interacting with and impeding the light wave differentially. Biological samples are fairly homogeneous (i.e., carbon-based polymers) and do not greatly impede light. Therefore, it is often necessary **to stain cells with dyes to provide more contrast.**
Different dyes have different affinities for different subcellular components.
For example, many dyes specifically interact with nucleic acids (i.e., DNA and RNA) and will differentially stain the cytoplasm and nucleus. These stained subcellular components will differentially absorb the light waves and result in less light reaching the eyes or camera, and thus appears darker.
Furthermore, since the dyes only absorb certain wavelengths of light, the various structures within the specimen will exhibit
different colors
Variations to bright field (transmission) microscopy
Dark Field -
• Phase Contrast
• Differential Interference
Contrast (or Normarski)
• Confocal Scanning
• Fluorescence
• Image Enhancement
The images produced by dark-field microscopy are low resolution and details cannot be seen.
Dark-field microscopy is especially useful for **visualization of small particles such as bacteria.**
Phase contrast microscopy and differential-interference-contrast
allow objects that differ slightly in refractive index or thickness to be distinguished within unstained or living cells.
Differences in the thickness or refractive index within the sample result **in a differential retardation of light which shifts the phase or deviates the direction of the light.**
During phase contrast microscopy the phase differences are converted to intensity differences by special objectives and condensers.
In fluorescence microscopy a fluorochrome is excited with ultraviolet light and the resulting visible fluorescence is viewed. This produces a bright image in a dark background.
•
ELECTRON MICROSCOPY
Particles, such as electrons, travelling near the speed of light behave as a wave (i.e., radiation) and their effective wavelength is inversely proportional to electron's velocity. Therefore increased resolution can be achieved by examining a sample with high velocity electrons.
Comparison of Microscope Optics

The general principal of electron microscopy is analogous to light microscopy (Figure) except that **electrons are used to analyze the specimen instead of visible light.**

The illumination source is a white-hot tungsten filament (вольфрамовая нить) which emits high velocity electrons.

The electron beamis focused by a condenser lens onto the specimen.

The electrons interact with the photographic plate or fluorescent screen as if they were photons (i.e., light) and generate an image. The differential loss of electrons due to the substractive action of the sample will generate an image in much the same way as the absorption of light creates an image in light microscopy.

Sample preparation

Fixation

2. Dehydration

3. Embedding

4. Sectioning

5. Staining

Sample preparation

It is not possible to view living material with an electron microscope. Biological samples are usually fixed with glutaraldehyde, which cross-links proteins, and treated with osmium tetroxide, which stabilizes lipid bilayers and proteins.

Osmium tetroxide is reduced by many organic compounds, especially lipids, which results in cross-linking. Since electrons have very little penetrating power, samples must be embedded in special plastic resins and cut into thin sections of 0.05-0.1 μm. Removing all water from the specimen is necessary for the proper polymerization of the plastic resin. Following fixation the samples are dehydrated by exposing them to series of increasing alcohol concentrations until reaching 100%. The

L2-3 NMR spectroscopyNMR vs X-ray crystallography for protein structure determination in an x-ray diffraction pattern, each datum (reflection) contains information about each atom in the asymmetric unit

-each atom contributes information that contributes to the intensity of each reflection

• in an NMR spectrum, each peak contains information about only a single interatomic distance or angle

-the process of determining a solution structure by NMR is one of measuring many small distances and angles “one at a time”

Is the potential technique for structural proteomics.

NMR spectroscopy does not require crystals.

So structure determination of samples can be identified in a short period of time.

Why use NMR?

• can’t get a crystal / want to work in solution
• want to look at binding to other proteins/molecules
• want to understand stability
• want to measure fast dynamics processes

L. 4 MEMBRANES AND DETERGENTS

Membranes are composed of a lipid bilayer and associated proteins. Lipids are amphipathic molecules in that they contain a polar head group and hydrophobic tails. In aqueous solutions lipids will aggregate such that the hydrophobic tails interact with other hydrophobic tails and the polar head groups are exposed to water. The two possible configurations are: 1) a spherical micelles with the hydrophobic tails pointed inward, or 2) a bilayer with the hydrophobic tails sandwiched between the polar head groups. The shape of the lipid and its amphipathic nature cause them to spontaneously form bilayers in aqueous solutions and accounts for the stability of membranes.

Proteins interact with this layer bilayer in several different fashions. Transmembrane proteins pass through the bilayer. A hydrophobic domain, typically an α-helix composed of amino acids with hydrophobic side-chains, interacts with the hydrophobic tails of the lipids and anchors the protein to the bilayer. Some transmembrane proteins will have multiple membrane spanning domains. Other membrane proteins are anchored to the lipid bilayer via fatty acids or lipids that are covalently attached to the protein. Other membrane are attached to the membrane by non-covalent association with other membrane proteins.
The condenser lens, is an **electromagnet instead of a glass.**
These electrons are differentially impeded by the various structures within the sample.
In other words, some of the electrons are scattered or absorbed by the atoms of the specimen.
The electrons which pass through the sample are focused with a series of magnetic objective lens onto either a photographic plate or a fluorescent screen.
Proteins interact with this layer bilayer in several different fashions.
Transmembrane proteins pass through the bilayer.
A hydrophobic domain, typically an α-helix composed of
amino acids with hydrophobic side-chains, interacts with the hydrophobic tails of the lipids and anchors the protein to the bilayer.
Some transmembrane proteins will have multiple membrane
spanning domains.
Other membrane proteins are anchored to the lipid bilayer via fatty acids or lipids that are covalently attached to the protein.
Other membrane are attached to the membrane by non-covalent association with other membrane proteins.
Many cellular processes occur on membranes or in membrane-bound subcellular compartments.
Approximately half of a cell's total protein is associated with membranes or
found in membrane-bound compartments. The study of these membrane associated processes may require the isolation of membranes or membrane proteins. The choice of tissue and membrane fraction (i.e., organelles) will depend in part on the phenomenon being studied.
Some cell types are better sources for certain types of membranes. The cells need to be disrupted by procedures which preserved the activity of interest (see next slide).
Disruption of cells or subcellular compartments will usually result in the
membranes forming vesicles.
The different types of membranes will have different sizes and densities based upon their lipid and protein composition and can be prepared by differential centrifugation, density gradient centrifugation or a combination of the two.
choice of cells or tissue
• choice of membrane fraction
• homogenization conditions
• preparation of membranes
♦ differential centrifugation
♦ density gradient centrifugation
• solubilization of membranes
• isolation of proteins
The exact isolation technique will depend upon the source and type of membrane being isolated as well as the desired purity.

L 5. «Present types of biophysical methods on study of membrane structures. Basic principles of chromotography »

L. 5 **BIOPHYSICAL METHODS**
Interaction of detergents with membrane structures

Methods for detergent removal
Electrophysiology
Hydrodynamics
Microscopy & Imaging
Atomic Force Microscopy (AFM)

Detergents interact with both membrane lipids and proteins.
At low detergent to lipid ratios,
detergent monomers incorporate into the lipid bilayer without disrupting the membrane.
As the detergent:lipid ratio increases,
lipids, because of their similar properties as detergents, will form **mixed micelles with the detergent.**
Therefore, the detergent concentration needs to be above the **CMC.** Detergent molecules also interact with the hydrophobic portions of the proteins and in effect replace the lipids.
This will lead to the proteins being coated with detergents and
prevent protein-protein interactions that would normally result **in protein precipitation.**
Protein solubilization occurs at or near **the CMC** for most detergents.
There are no general rules for choosing among the various types of detergents.
It is not possible to predict which detergent will be most useful for any particular application.
Pilot experiments to determine the optimal detergent as well as the optimal conditions for maximal protein solubilization are carried out.
Possible Detergent Effects
• protein structure
• protein activity
• interference with assays
• separation techniques
DETERGENT REMOVAL
A detergent is solubilized the protein of interest.
It is necessary to remove the detergent.
In general, detergents are difficult to remove.
Therefore, when choosing a detergent is need also to consider how ease a detergent can be removed.
methods for detergent removal
Dialysis
• Chromatography
• Replacement
Biophysical methods
Electrophysiology
Hydrodynamics
Microscopy & Imaging
Atomic Force Microscopy (AFM)
Biophysical methods
Allow to study:
The characterization of molecular structure,
the measurement of molecular properties, and
the observation of molecular behavior
A wide range of biophysical techniques are developed to study molecules in **crystals, in solution, in cells, and in organisms**.
These biophysical techniques provide information about:
the electronic structure,
size,
shape,

dynamics,
polarity, and
modes of interaction of biological molecules.
Some of the most exciting **Biophysical Techniques** provide
images of cells,
subcellular structures, and
even individual molecule
It is now possible, for example,
to directly observe the biological behavior and physical properties of single protein or DNA molecules within a living cell and
determine how the behavior of the single molecule influences the biological function of the organism.

Electrophysiology
Detection of Secretion by Electrochemical Methods
The basic requirements for the types of electrochemical experiments that are used at single cells are two electrodes placed in a common solution, a voltage source, and a **picoammeter.** One of the electrodes, the reference electrode, maintains a constant potential.
Early in the development of electrochemical techniques, the normal hydrogen electrode (NHE) was accepted as a standard.

Electrophysiology
In practice, the NHE is difficult to use and, more commonly, the Ag/AgCl electrode is used. By convention, the NHE has a standard potential, or E^o, of 0 V. and other types of reference electrode's potentials are commonly compared to the
NHE. For example, the Ag/AgCl electrode has an E^o of 0.197 V versus the NHE, or $E^o = 0.197$ V
vs. NHE, in electrochemical terminology. The Ag/AgCl electrode consists of metallic Ag coated with AgCl in contact with an aqueous solution containing a fixed concentration of chloride ion. This electrode configuration generates a stable potential and is insignificantly perturbed when small currents flow through it.
The other electrode in biological applications is typically a **carbon fiber microelectrode** and it is referred to as the *working electrode*. By controlling the surface potential of the working electrode and simultaneously measuring the amount of current passing through it, information concerning the charge transfer processes that occur at the interface of the carbon and solution is obtained.

Hydrodynamics
The behavior of large biomolecules such as :
proteins,
carbohydrates, and
nucleic acids
in solution.
Hydrodynamics is complex and directly related to molecular size, shape, and flexibility of large biomolecules;
The analysis of hydrodynamic behavior of large biomolecules provides important information about the structure, dynamics, and interactions of biomacromolecules.
Hydrodynamics
a branch of physics that deals with the motion of fluids and the forces acting on solid bodies immersed in fluids and in motion relative to them
the study of forces that act on or are produced by liquids. Also called **hydromechanics**. — **hydrodynamic, hydrodynamical,**

Microscopy & Imaging

The most accessible developments in biophysics involved improvements in our ability to generate **images of cellular and molecular structures with dimensions from microns to nanometers.**

It is now possible to "see" individual molecules or cellular structures using atomic force, electron, or confocal fluorescence microscopy.

Confocal Fluorescence Microscopy

Confocal microscopy offers several advantages over conventional optical microscopy, including

1. controllable depth of field,
2. the elimination of image degrading out-of-focus information, and
3. the ability to collect serial optical sections from thick samples.

The key to the confocal approach is the use of spatial filtering to eliminate out-of-focus light or flare in samples that are thicker than the plane of focus.

There has been a tremendous explosion in the popularity of confocal microscopy in recent years, due in part to the relative ease with which extremely high-quality images can be obtained from samples prepared for conventional optical microscopy, and in its great number of applications in many areas of current research interest, including biotechnology.

Confocal Fluorescence Microscopy

Root of Arabidopsis thaliana with green fluorescent protein decorating cell membrane and red fluorescent protein marking nuclei.

Atomic Force Microscopy (AFM)

Allows Measuring Force of Intermolecular Interaction and bonds between a bacterium and another surface.

AFM is very useful in biological sciences because it can be used **on living cells that are immersed in water.**

AFM is particularly useful when the **cantilever (support**) is modified with chemical groups (e.g. amine or carboxylic groups), small beads (e.g. glass or latex), or even a bacterium.

AFM operates on a very different principle than other forms of microscopy, such as optical microscopy or electron microscopy.

The key component of an AFM is a **cantilever**(*консольная*) that bends in response to forces that it experiences as it touches another surface.

Forces as small as a few picoNewtons can be detected and probed with AFM.

Atomic Force Microscopy (AFM)

Atomic-force microscopy (**AFM**) or **scanning-force microscopy** (**SFM**) is a type of scanning probe microscopy (SPM), with demonstrated resolution on the order of fractions of a nanometer, more than 1000 times better than the optical diffraction limit. The information is gathered by "feeling" or "touching" the surface with a mechanical probe. Piezoelectric elements that facilitate tiny but accurate and precise movements on (electronic) command enable very precise scanning.

Abilities of Atomic Force Microscopy AFM

has three major abilities:

1. force measurement,
2. imaging, and
3. manipulation.

In force measurement, AFMs can be used to measure the forces between the probe and the sample as a function of their mutual separation. This can be applied to perform ***force spectroscopy.***

For imaging, the reaction of the probe to the forces that the sample imposes on it can be used to form an image of the three-dimensional shape (topography) of a sample surface at a high resolution. This is achieved by raster scanning the position of the sample with respect to the tip and recording the height of the probe that corresponds to a constant probe-sample interaction (see section topographic imaging in AFM for more details). The surface topography is commonly displayed as a ***pseudocolor*** plot.

Abilities of Atomic Force Microscopy AFM

In manipulation, the forces between tip and sample can also be used to change the properties of the sample in a controlled way. Examples of this include atomic manipulation, scanning probe lithography and local stimulation of cells.

Simultaneous with the acquisition of topographical images, other properties of the sample can be measured locally and displayed as an image, often with similarly high resolution. Examples of such properties are mechanical properties like stiffness or adhesion strength and electrical properties such as conductivity or surface potential. In fact, the majority of SPM techniques are extensions of AFM that use this modality.

Atomic Force Microscopy

L. 6 ANALYSIS AND CHARACTERIZATION OF PROTEINS

- L. 6 Analysis and Characterization of Proteins
- PROTEIN STABILITY
- Factors Affecting Protein Stability
- PROTEIN ASSAYS
- PROTEIN
- Proteins contribute to he structure of a cell and are responsible for cellular functions such as catalysis and molecular recognition.
- Proteins are polymers of L-α-amino acids.
- Twenty different amino acids, distinguished by their side-chain groups, are found in proteins.
- The side-chain groups vary in terms of their chemical properties such as
- polarity, charge and size.
- These various side-chain groups will influence the chemical properties of proteins as well as determine the overall structure of the protein (see Appendix).
- For example, the polar amino acids tend to be on the outside of the protein where they interact with water and the nonpolar amino acids are on the inside forming a hydrophobic core.
- L-α-AMINO ACIDS

Nonpolar	**Polar**
• Alanine	Arginine
• Glycine	Asparagine
• Isoleucine	Aspartic a.
• Leucine	Cysteine
• Methionine	Glutamic a.
• Phenylalanine	Glutamine
• Proline	Histidine
• Tryptophan	Lysine
• Valine	Serine
•	Threonine
•	Tyrosine

- PROTEIN STABILITY

Proteins are often weak molecules that need to be protected during purification and characterization. Protein denaturation is the loss of protein structure due to unfolding. Maintaining biological activity is important and protein denaturation should be avoided in those situations.

- Factors Affecting Protein Stability
- General rules to proteins

, things that destabilize H-bonding and other forces that contribute to secondary and tertiary protein structure will promote

protein denaturation.
Different proteins exhibit different degrees of sensitivity to denaturing agents and some proteins can be re-folded to their correct conformations following denaturation.
Many proteins are especially labile and need to be stored at -20o or -80o. However, repeated freezing and thawing of protein solutions is often deleterious.
Adding 50% glycerol to storage buffers will lower the freezing point and allow storage at -20o.
Solutions for working with proteins will often contain heavy-metal chelators and/or antioxidants as protectants..
In addition, proteases may be released during cell disruption and it may therefore be necessary to include protease inhibitors.

- *PROTEIN ASSAYS*

Numerous spectrophotometric methods have been developed to estimate the amount of protein in a sample.
Proteins are chromophores with absorption maximum in the UV range.
Some proteins, such as cytochromes and hemoglobin, will have distinct spectral characteristics due to prosthetic groups.
In addition, several indirect ways to measure protein concentrations spectrophotometrically have been developed.

- UV Absorption

A simple method to measure protein concentration is to determine the absorption at 280 nm. Tyrosine and tryptophan residues which have Amax at 275 and 280, respectively, are responsible for this absorption. The distribution of tyrosine and tryptophan is fairly constant among proteins so it is not absolutely necessary to determine an extinction coefficient for each individual protein.
Typically, a 1 mg/ml protein solution will result in an A280 of approximately one (1). In the case of purified proteins the exact extinction coefficient will depend on the exact amount of tyrosine and tryptophan in that particular protein.
For example, a 1 mg/ml IgG solution has an A280 of approximately 1.5.
The simplicity and ability to completely recover the sample are the major advantages of this method. The lower limit of sensitivity for UV absorption is 5-10 μg/ml.
A potential problem with using A280 values to calculate protein concentration is the absorption due to contaminating substances, and in particular, nucleic acids which have an Amax at 260 nm.
It is possible to use correction factors that permit the determination of protein concentrations.
A particularly convenient formula is:
(A235 - A280)/2.51 = mg/ml protein.
Indirect spectrophometric assays (eg., Lowry, Bradford) for the determination of protein concentrations overcome some of the problems associated with interfering substances in protein samples.
However, the measured protein cannot be recovered in such assays and they take longer to perform.

- Folin-Ciocalteu or Lowry

Historically, one of the most widely used protein assays was the Lowry assay. This assay is a modification of a previous assay known as the Biuret.
In the Lowry assay proteins react with alkaline Cu^{2+} reducing it to Cu^{+}.
The reduced Cu^{+} and the side-chain (R) groups of tryptophan, tyrosine and cysteine react with the Folin-Ciocalteu reagent (complex of inorganic salts) to form a blue color that is proportional to the amount of protein.
The A600-750 is determined and protein concentration is calculated from a standard curve.

- Bradford or Coomassie Blue G-250

The Bradford assay has replaced the Lowry as the standard protein assay. The major advantage is that it is carried out in **a single step** and that there are very few interfering substances.
The principle of the assay is based on a shift of the Amax of the Coomassie-blue (G-250) dye from 465 nm to 595 nm in the presence of protein due to a stabilization of the anionic form of the dye.
The dye reacts primarily with arginine residues and to lesser extent with his,
lysine, tyrosine, tryptophan and phenylalanine.
Protein concentrations are determined by developing a standard curve with known amounts of proteins and extrapolating the absorbance
values of the samples.

- Assay of Specific Proteins

In addition to measuring the total amount of protein, it is
often necessary to estimate the **amount of a specific protein** in a
mixture of proteins.
Measuring a specific protein will depend upon the availability of an assay that is specific for the protein of interest.
Protein assays should be practical in addition to being specific and accurate.
Specificity
• Sensitivity
• Accuracy
(Quantification)
• Rapid
• Easy to Perform
Typically protein assays are based upon the biological activity of the protein of interest.
For example, enzyme assays will detect the conversion of a substrate to a product.
Enzymes assays can be based upon colorimetric, fluorescent or radioactive substrates (or products).
Many proteins bind to ligands or other substances and this binding activity is measured.
Bioassays measure a change in some biological property (eg., stimulation of cell division).
In cases where the protein of interest has no measurable activity or the activity is unknown it may be possible to generate antibodies against the protein and develop an immunoassay.
If antibodies against such a protein are not available, the assay may simply be the amount of a protein on a Commassie blue-stained gel following electrophoresis.

L. 7 GENERAL STRATEGY OF PROTEIN PURIFICATION.

Proteins are purified by fractionation procedures, a series of independent steps in which the properties of protein of interest are utilized to separate it from other contaminating proteins.
General strategy of protein purification
Characteristic Procedure Solubility 1. Salting in 2. Salting out Ionic charge: 1. Ion exchange chromatography 2. Electrophoresis 3. Isoelectric focusing Polarity: 1. Adsorption chromatography 2. Paper chromatography 3. Hydrophobic interaction chromatography Molecular size: 1. Dialysis and ultrafiltration 2. Gel electrophoresis 3. Gel filtration chromatography 4. Ultracentrifugation Binding specificity: 1. Affinity chromatography .

- Assay of specific Proteins

In addition to measuring the total amount of protein, it is
often necessary to estimate the **amount of a specific protein** in a
mixture of proteins.
Measuring a specific protein will depend upon the availability of an assay that is specific for the protein of interest.
Protein assays should be practical in addition to being specific and accurate.

- Specificity
- • Sensitivity
- • Accuracy

- (Quantification)
- • Rapid
- • Easy to Perform

L. 8 MAIN PRINCIPLES OF ELECTROPHORESIS

- 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 ($SO_4^{\cdot}$). 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., H3PO4).
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.

L 9. «METHODS OF NUCLEIC ACIDS ISOLATION».

- NUCLEIC ACID ISOLATION
- Three major types of techniques
- Isolation of High Molecular Weight Genomic DNA
- Plasmid Minipreps and Adsorption Methods.
- ISOLATION OF RNA
- Three major types of techniques

, or combinations of them, are employed in the isolation of nucleic acids:
differential solubility,
absorption methods, or
density gradient centrifugation.
The choice of method will depend on the type of DNA being isolated (Box) and the application. A major goal of nucleic acid isolation is the removal of proteins.
The separation of nucleic acids from proteins is generally easily accomplished due to their different chemical properties. In particular, the highly charged phosphate backbone makes the nucleic acids rather hydrophilic as compared to proteins which are more hydrophobic. Separating the different types of nucleic acids can be more problematic in that they all have similar chemistries.
On the other hand, though, this similar chemistry results in a few basic procedures which are common to many nucleic acid isolation protocols.
Most nucleic acid isolation protocols involve a cell lysis step, enzymatic treatments, differential solubility (eg., phenol extraction or absorption to a solid support), and precipitation.

- Genomic (chromosomal)
- • Organellar (satellite)
- • Phage/Viral (ds or ss)
- • Plasmid (extrachromosomal)
- • Complementary (mRNA)

Cell Lysis. Nucleic acids must be solubilized from cells or other biological material. This solubilization is usually carried
out under denaturing conditions such as:
SDS, alkali, boiling or
chaotropic agents.
These denaturing conditions efficiently
solubilize the nucleic acids and generally do not adversely affect them.
In addition, the denaturing conditions promote the removal of proteins during the subsequent steps and inhibit the activity of nucleases which will degrade the nucleic acids.

- Cell Lysis
- • ± Enzyme Treatment
- • Phenol Extraction or
- Adsorption Methods
- • EtOH Precipitation
- Enzymatic Treatment.

Another approach in the isolation of nucleic acids is to degrade unwanted components.
For example, inclusion of proteases (eg., proteinase K) in the lysate will
promote the removal of proteins. Proteinase K is still active at 55o in the presence of 0.5% SDS.
The elevated temperature and SDS improve solubility and inhibit any DNAse activity that may be present in the lysate.
Nucleases can also be used to remove unwanted nucleic acids. For

example, many DNA extraction protocols include a RNase treatment step, and visa versa. It is important that the RNase be free of DNase activity.
DNase-free RNase is easily prepared by boiling commercial RNase for 10 minutes. The stability of RNase makes the preparation of RNase-free DNase more difficult.
RNase-free DNase should be purchased from a reliable vendor or tested before it use.

- 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. Suspension

Lysis: grind in Liquid N2 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.

- 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.
- Phenol Extraction

Phenol is an organic solvent that is used to separate proteins from nucleic acids. Proteins are hydrophobic and partition in the organic phase.
Nucleic acids are highly charged and partition in the aqueous phase.
The advantages are of phenol extraction are that it is easy to carry out and can be adapted to many applications. It is also easily applied over a wide range of volumes (40 mkl to several ml). In particular, phenol extraction is widely used for the isolation of high molecular weight genomic DNA.
Phenol extraction is accomplished by mixing the sample with an equal volume of phenol which has been previously saturated with a Tris buffer at pH 8 containing EDTA and NaCl.
The phenol should be molecular biology grade phenol should and store at -20o until preparing the saturated solution.
The saturated solution is stored at 4o. Phenol is easily oxidized, as evidenced by yellowing, and the oxidation products can break DNA.
Oxidized phenol should be discarded.

Depending on the application, the two phases are completely mixed by vortexing, or gently mixed (eg., high molecular weight DNA).
The phases are separated by centrifugation and the upper aqueous phase, which contains the nucleic acids, is retained.
Proteins will often be visible as flocculent material at the top of the phenol phase.
The two phases need to be carefully separated in that the nucleic acids and proteins tend to be at the interface. Leaving too much of the aqueous layer behind will lead to undue loss of material and aspirating too close to the interface can include protein.
The aqueous phase can be re-extracted with phenol to remove more protein.
A common variation of phenol extraction is a mixture of phenol:chloroform: isoamyl alcohol (25:24:1). The more organic chloroform removes lipids, denatures more protein and mixes less with the aqueous phase leading to more efficient extraction.

- Ethanol Precipitation

Nucleic acids can be precipitated from dilute solutions with ethanol. This precipitation can be a concentration step or a means to change buffers, especially after phenol extraction.
Typically either sodium acetate or potassium acetate, pH 5.0-5.5, is added to a final concentration of approximately 0.3 M.
The sodium and acidic pH will neutralize the highly charged phosphate backbone and to promote hydrophobic interactions.
Two-to-two and a half volumes of ethanol are added and the sample is incubated as -20o.
If the nucleic acids are small in size and/or in low concentrations an extended incubation (several hours to overnight) is needed.
The precipitated DNA is collected by centrifugation. The pellet is rinsed with 70% ethanol to remove any excess salt, dried and dissolved in the appropriate buffer.
A variation is to substitute ammonium acetate if the 'hard' salts are a problem. Another modification is to use an equal volume of isopropanol (instead of 2-2½ volumes of ethanol) which minimizes the increase in sample volume.

- Isolation of High Molecular Weight Genomic DNA

High molecular weight chromosomal DNA is usually isolated by multiple rounds of phenol extraction and enzyme treatments as discussed above.
Shear forces, which can break long DNA molecules, need to be avoided during all steps and samples should never be vortexed.
Therefore, the phenol extraction is carried with gentle rocking for several hours.

- Isolation of High Molecular Weight Genomic DNA

These precautions against shear forces are not necessary in the isolation of *low molecular weight DNA*.
Another common modification at the ethanol precipitation step is ‘spool out’ the high molecular weight genomic DNA on the end of a sealed Pastuer pipet.
The precipitated DNA is wrapped *(Завернутый)* around the end of the pipet is then allowed to partially dry and then dissolved in the appropriate buffer.
This minimizes the contamination with RNA and low molecular weight DNA fragments.

- Plasmid Minipreps (*минипродукты*)and Adsorption Methods

Historically, phenol extractions were used for the isolation of most forms of nucleic acids. It is now more common to use techniques based upon adsorption chromatography for the isolation of smaller DNA molecules, such as plasmids.
Various kits are available for the rapid isolation of small quantities of plasmid DNA.
The procedure consists of solubilizing the bacteria in an alkali solution followed by neutralization with sodium acetate.
The neutralization results in the precipitation of some of the protein and the genomic DNA which is removed by centrifugation.

The soluble material is then mixed with a resin (*Смолы*) in the presence of chaotropic agents (usually guanidine hydrochloride). The resins are usually either based on silica or diatomaceous earth. Under these conditions DNA binds to the
matrix, but proteins and RNA do not. The DNA is eluted in a low salt buffer. These methods are rapid and yield a highly purified plasmid DNA which can generally be used directly in most applications without further processing.
Another common application for an adsorption method is the isolation of DNA fragments following gel electrophoresis.
In this case the agarose gel piece containing the DNA is dissolved in NaI, *a chaotropic salt*, and the DNA adsorbed to silica (*Кремнезем*).
The DNA is then eluted with a low salt buffer and sometimes gentle heating.

L. 10. METHODS OF NA ANALYSIS

ISOLATION OF RNA
DENSITY GRADIENT CENTRIFUGATION
ANALYSIS AND QUANTIFICATION.....

ISOLATION OF NUCLEIC ACIDS
Three major types of techniques, or combinations of them, are employed in the isolation of nucleic acids:
differential solubility,
absorption methods, or
density gradient centrifugation.
The choice of method is depended on the type of DNA being isolated (Box) and the application.
A major goal of nucleic acid isolation is **the removal of proteins**.
The separation of nucleic acids from proteins is generally easily accomplished due to their different chemical properties.
In particular, the highly charged phosphate backbone of the nucleic acids makes them rather hydrophilic as compared to proteins which are more hydrophobic
Separating the different types of nucleic acids can be more problematic in that they all have similar chemistries.
On the other hand, though, this similar chemistry results in a few basic procedures which are common to many nucleic acid isolation protocols.
Most nucleic acid isolation protocols involve:
a cell lysis step,
enzymatic treatments,
differential solubility (eg., phenol extraction or absorption to a solid support), and precipitation.
Cell Lysis
• ± Enzyme Treatment
• Phenol Extraction or
Adsorption Methods
• EtOH Precipitation
Cell Lysis. Nucleic acids must be solubilized from cells or other biological material.
This solubilization is usually carried
out under **denaturing conditions such as: SDS, alkali, boiling or**
chaotropic agents.
These denaturing conditions efficiently solubilize the nucleic acids and do not adversely affect them.
In addition, the denaturing conditions promote the removal of proteins during the subsequent steps and inhibit the activity of nucleases which will degrade the nucleic acids.
Enzymatic Treatment.

Another approach in the isolation of nucleic acids is to degrade unwanted components.
For example, inclusion of proteases (eg., proteinase K) in the lysate will promote the removal of proteins. Proteinase K is still active at 55o in the presence of 0.5% SDS.
The elevated temperature and SDS improve solubility and inhibit any DNAse activity that may be present in the lysate.
Nucleases can also be used to remove unwanted nucleic acids.
For example, many DNA extraction protocols include a RNase treatment step, and visa versa.
It is important that the RNase be free of DNase activity.
DNase-free RNase is easily prepared by boiling commercial RNase for 10 minutes. The stability of RNase makes the preparation of RNase-free DNase more difficult.
RNase-free DNase should be purchased from a reliable vendor or tested before it use.
Phenol Extraction
Phenol is an organic solvent that is used to separate
proteins from nucleic acids.
Proteins are hydrophobic and partition in the organic phase.
Nucleic acids are highly charged and partition in the aqueous
phase.
The advantages are of **phenol extraction** are that it is easy to carry out and can be adapted to many applications.
It is also easily applied over a wide range of volumes (40 μl to several ml).
In particular, phenol extraction is widely used for the isolation of high molecular weight genomic DNA
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 (Box).
In particular, precautions against RNase activity must be taken.
RNase is an extremely stable and active enzyme.
Gloves should be worn at all times and sterile plasticware 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.

Precipitation of RNA with LiCl.
LiCl has been used to selectively precipitate RNA.
Large RNAs (rRNA, mRNA) are insoluble at high ionic strength.
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 -20°. 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
usually low salt and high temperature which breaks the base pairing.
Oligo-dT- column
ANALYSIS AND QUANTIFICATION of NUCLEIC ACIDS
The quality and quantity of isolated nucleic acids 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
con

L 11. «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).

DNA can also be digested with **mixtures of restriction enzymes**.
If the two enzymes have the similar optimal reaction conditions, they can be mixed together and the reaction carried out *simultaneously*.
If not, the digestions are carried out sequentially by diluting the sample in the appropriate buffer or by adding the required components.
If necessary, the DNA can be reisolated after digestion with the first enzyme.
The typically procedure is to phenol extract the digestion mixture, precipitate the aqueous phase with ethanol, and dissolve the precipitated
DNA in the appropriate buffer for the second digestion.

- Conditions Promoting Star Activity of endonuclease
- • high glycerol (>5%) concentration
- • high enzyme/DNA ratio (100 units/μg)

- • low ionic strength (<25 mM)
- • high pH (>8)
- • organic solvents
- • substitution of Mg2+
- **Frequency of NA cutting byendonuclease**

- Theoretically restriction sites will **be random within any genome.**
- The number of times a particular restriction enzyme will cut a genomic DNA can be approximated from the size of the genome and its base composition.
- For example the probability of finding any 6-base recognition sequence in a genome with 50% GC content is:
- (¼)(¼)(¼)(¼)(¼)(¼) = 1/4096.
- This means that one would expect on average one restriction site every 4000 bp.

L 12. «METHODS OF STUDY GENES EXPRESSION»

SINGLE TRANSCRIPTS:

MULTIPLE TRANSCRIPTS SIMULTANEOUSLY:

Gene expression is a highly regulated mechanism that controls the function and adaptability of all living cells including prokaryotes and eukaryotes.
Several techniques exist for studying and quantifying gene expression and its regulation. Some of these techniques are old and well established while others are relatively new, multiplex techniques.
The field of gene expression analysis has undergone major advances in biomedical research. Traditional methods focused on measuring the expression of one gene at a time and not in any particular biological context.
However, today, mRNA expression techniques have led to improvements in gene identification.
Cells in all living organisms are continually activating or deactivating genes through gene expression, which contain the information required for producing proteins through proteins synthesis. When a particular protein is required by the cell, the gene coding for that protein is activated.
The first stage in producing a protein involves the production of an RNA copy of the gene's DNA sequence. This RNA copy is the messenger RNA.
The amount of mRNA produced correlates with the amount of protein eventually synthesized and measuring the amount of a particular mRNA produced by a given cell or tissue is often easier than measuring the amount of the final protein. Levels in gene activation may vary between cancerogenic and healthy cells.
Techniques that are low- to mid-plex (*Сплетение*). Reporter gene
A gene contains two functional segments. One is a coding DNA sequence, which contains the instructions for making a protein.
The other is a DNA sequence called a promoter, which is linked to this coding region and regulates the gene’s transcription, either by activating or suppressing its expression.
A reporter gene assay is used to determine the regulatory potential of a DNA sequence that is unknown.

This involves a promoter sequence being linked to a detectable reporter gene such as luciferase, β-galactosidase or β-glucuronidase. Examples of methods used to determine the expressed reporter gene protein are fluorescence, absorbance and luminescence.
Northern blotting. Western blotting
This is a technique used to detect specific RNA molecules present within an RNA mixture. Northern blotting is employed in the analysis of an RNA sample from a cell type or tissue so as to determine the RNA expression of certain genes.
Western blotting is a technique for detecting specific protein molecules within a protein mixture. This mixture might include all the proteins that are associated with a certain cell type or tissue. The technique can help to determine a protein's size, and how much of it is expressed.
NORTHERN BLOT
How can amounts of RNA be quantified?
The slide shows a virtual Northern with two lanes, one with RNA from control cells, the other with RNA from the experimental sample (eg drug treated cells). Let's say that there is 10x the amount of signal in the experimental sample compared to the control sample for the target gene.
This could mean expression of the gene has increased 10-fold, or it could mean that there is 10x as much RNA in the experimental sample.
To check for this one usually does a so-called 'loading control' in which the blot is probed for expression of a gene which does not change (e.g. actin, GAPDH, cyclophilin, RPLP0 mRNAs; ribosomaL RNA).
In this case, let's say that the loading control shows that there is twice as much RNA in the expt lane.
Thus the real change in the target gene is 10/2 =5 fold. We can express this in a more general fashion:
Northern blot
NORTHERN BLOT
1. mRNA isolation and purification
2. electrophorese on a gel
3. The gel is probed by hybridizing with a labeled clone for the gene under study.

Fluorescent in situ hybridization (FISH)

This is a cytogenetic technique that can be used to identify and locate specific gene sequences. FISH can be used to visualize copy number aberrations such as the deletion, translocation or amplification of chromosomes. The technique is used in prenatal diagnosis and also provides a useful tool in the diagnosis and predicted prognosis of various sarcomas. The technique is also used in dermatology to help evaluate atypical moles.

Methods for the Study of Gene Expression

Single Transcripts:
- Northern blot
- RNAse protection assay
- Reverse Transcription (RT)-PCR
- Real-time PCR (qPCR)
- In-situ-hybridization
-

Multiple Transcripts Simultaneously:
Dot-Blot analysis
- Differential Display, SAGE
- Subtractive hybridization (1996)
- DNA Microarrays (1999)
- NanoString
- Second generation Sequencing (NGS)

Reverse transcription polymerase chain reaction (RT-PCR)

This is the most **sensitive** technique available for detecting and quantifying mRNA. Using RT-PCR, extremely small sample sizes can be used in the quantification of mRNA and the technique can in fact do this using just a single cell.

MOLECULAR METHODS TOOL BOX

Subtractive hybridization Methods

Techniques that are higher plex

Serial Analysis of Gene Expression (SAGE)

SAGE is a technique used to create a library of short sequence tags which can each be used to detect a transcript. The expression level of the transcript can be determined by assessing how many times each tag is detected. This technology enables comprehensive expression analysis across the genome.

DNA microarray

Also known of as biochip or DNA chip, a DNA microarray is a solid surface to which a collection of microscopic DNA spots are attached. The microarrays are used to determine expression levels across a large number of genes or to perform genotyping across different regions of a genome

- **L. 13-14. THE USE OF DNA MARKERS IN MOLECULAR BREEDING**

Genetic markers are used for labeling and tracking the genetic variations in DNA samples. Genetic markers are biological compounds which can be determined by **allelic variations** and can be used **as experimental probes or labels** to track an individual, tissue, cell, nucleus, chromosomes or
genes.

The use of DNA markers in plant and animal breeding has opened new territory in agriculture which is called **molecular breeding**.

These markers are widely used because of their high prevalence
and expression in different stages of the organisms.

These markers come from **different classes of DNA mutations:**

1. **such as substitution mutations (point mutations),**
2. **re-assortments (insertions and deletions),**
3. **replication errors and DNA tandem repeats**

These markers are selectively neutral because they are usually **located in non-coding regions of DNA.**

Unlike morphological and biochemical markers, DNA markers are practically unlimited in number and **are not affected by environmental factors and/or the developmental stage of the plant.**

Apart from the use of DNA markers in construction of linkage maps, they have numerous applications in plant breeding such as:

1. assessing the level of genetic diversity within cultivars and
2. fifingerprinting the germplasms.

Many agriculturally important traits such as productivity and
quality, tolerance to environmental stresses, and some of forms of disease resistance are quantitative, are controlled by **polygenes** which complicate the breeding
process since phenotypic performances only partially reflects the genetic values of individuals. These complex traits are referred to **as quantitative traits** (also called as **polygenic or multifactorial traits**) and the regions within genomes that contain genes associated with a particular quantitative trait are known *as quantitative trait loci (QTLs).*
(also called polygenic, continuous, multifactorial, or complex traits) in nature.

- The genetic variation of a quantitative trait is controlled by the collective

effects of numerous genes, known **as quantitative trait loci (QTLs).**

Identification of **QTLs of agronomic importance** and its utilization in a crop improvement requires mapping of these QTLs in the genome of crop species using molecular markers.
Identification of QTLs with genetically linked DNA-markers is useful for incorporating genes into improved cultivars via marker-assisted selection
(MAS), map-based cloning of the tagged genes, and for a better understanding of the genetics of complex traits. Linkage analysis and association
mapping are the two most commonly used methods for QTL mapping. The main items include the two QTL mapping methods with special focus on
mapping population types, linkage map construction, and marker–trait association analysis using different statistical methods and software programs.

- A **quantitative trait locus** (**QTL**)
- is a region of **DNA** that is associated with a particular **phenotypic trait.**
- These QTLs are often found on different **chromosomes**.
- Knowing the number of QTLs that explains variation in the phenotypic trait tells us about the **genetic architecture** of a trait. It may tell us that plant height is controlled by ***many genes*** of small effect, or by a few genes of large effect.
- 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.
- In a recent development, classical QTL analyses are 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
- Principle of QTL Analysis

Identifying a QTL or a gene within a plant genome is like finding the proverbial
needle in a haystack. However, QTL analysis can be used to divide the haystack in manageable piles *(стог сена в управляемые сваи)* and systematically search them.
In simple terms, QTL analysis is based on the principle of detecting an association between phenotype and the genotype of markers.
Markers are used to partition the mapping population into different genotypic groups based on the presence or absence of a particular marker locus and to determine whether significant differences exist between groups with respect to the trait being measured .
A significant difference between phenotypic means of the groups (either 2 or 3), depending on the marker system and type of population, indicates that the marker locus being used to partition the mapping population is linked to a QTL controlling the trait.

- Traditional QTL Mapping (Linkage Mapping)

The general steps involved in a traditional QTL mapping experiment are as follows:

1. select two parental strains that have differences between them in the alleles that affect variation in a trait. The parents need not be different in the mean phenotypic value of the trait as different allelic combinations can yield the same phenotypic mean;

(2) develop an appropriate mapping population by crossing the selected
parents;
(3) phenotype the mapping population for the trait(s) of interest (morphological characters, agronomic traits, disease and pest scores, drought resistance, etc.) under greenhouse, screen-house, and/or field conditions;
(4) generate the molecular data on the population with adequate number of uniformly spaced polymorphic markers;
(5) construct a genetic map; and
(6) identify molecular markers linked to the trait(s) of interest using statistical programs.

- Mapping Populations Used in QTL Mapping Experiments
- Various types of mapping population may be produced from the heterozygous F1 hybrids:

1. **Double haploid lines (DHLs):** plants are regenerated from pollen (which is haploid) of the F1 plants and treated to restore diploid condition in which every locus is homozygous. Since the pollen population has been generated by meiosis, the DHLs represent a direct sample of the segregating gametes.
2. **Backcross (BC) population**: the F1 plants are backcrossed to one of the parents.
3. **F2 population: F1 plants are selfed.**
   - F2:3/F2:4 lines: F3/4 plants tracing back to the same F2 plant, also called F2 families.
   - 5. **Recombinant inbred lines (RILs):** inbred generation derived by selfing individual F2 plants and further single seed descent.
   - A population of RILs represents an 'immortal' or permanent mapping population.

Each of the above mapping populations
(**Double haploid lines (DHLs),**
F2 population: F1 plants are selfed,
Recombinant inbred lines (RILs), possesses advantages and disadvantages.
Hence, the choice of the type of mapping population depends on many factors such as the plant species, type of marker system used, and the trait to be mapped later on. F2 populations, derived from F1 hybrids, and BC populations, derived by crossing the F1 hybrid to one of the parents, are the simplest types of mapping populations developed for self-pollinating species.
Their main advantages are that they are easy to construct and require only a short time to produce.
However, such populations are not fixable due to their inherent heterozygous genetic constitution. This restricts their wide utility in QTL analysis.
The length of time needed for producing RIL population is the major disadvantage, because usually six to eight generations are required.
Development of a DH population takes much less time than RIL; however, the production of DH populations is only possible in species that are amenable to tissue culture (e.g., cereal species such as rice, barley, and wheat).
The major advantages of RIL and DH populations are that they produce homozygous or 'true-breeding' lines that can be multiplied and reproduced without genetic change occurring.
This allows for the conduct of replicated trials across different locations and years. Furthermore, seed from individual RIL or DH lines may be transferred between different laboratories for further linkage analysis and the addition of markers to existing maps. Information provided by co-dominant markers is best exploited by an F2 population, while information obtained by dominant marker systems can be maximized by using RILs or DHLs.
Double haploids, F2 or F3 families, or RILs are advantageous if the trait to be mapped cannot be accurately measured on a single-plant basis but must be assessed in replicated field experiments.
The frequency of genetic information obtained from genetic markers have significant impact on the evolutionary biology, **particularly in understanding the genetic basis of complex traits.**
At the same time, **new concepts, such as quantitative trait loci (QTL)**
mapping followed by the development of statistical tools, have emerged in **quantitative genetics** to identify the genes involved in the genetic variability of complex traits.
The complexity of these traits is influenced by the segregation of alleles at many loci, environmental factors and their interactions.

- A **quantitative trait locus** (**QTL**)
- is a section of **DNA** (the **locus**) that correlates with variation in a **phenotype** (the quantitative trait).
- The QTL typically is linked to, or contains, the **genes** that control that phenotype.
- **QTLs** are mapped by identifying which **molecular markers** (such as **SNPs** **or** **AFLPs**) correlate with an observed trait.

- This is often an early step in identifying and **sequencing** the actual genes that cause the trait variation.
- **Quantitative traits** are **phenotypes** (characteristics) that vary in degree and can be attributed to **polygenic** effects, i.e., the product of two or more **genes**, and their environment.
- Quantitative trait locus mapping.
- a) **Quantitative trait locus (QTL)** mapping requires parental strains (red and blue plots) that differ genetically for the trait, such as lines created by divergent artificial selection
- . b) The parental lines are crossed to create F1 individuals (not shown), which are then crossed among themselves to create an F2, or crossed to one of the parent lines to create backcross progeny.
- Both of these crosses produce individuals or strains that contain different fractions of the genome of each parental line. The phenotype for each of these recombinant individuals or lines is assessed, as is the genotype of markers that vary between the parental strains.
- c) Statistical techniques such as composite interval mapping evaluate the probability that a marker or an interval between two markers is associated with a QTL affecting the trait, while simultaneously controlling for the effects of other markers on the trait.
- The results of such an analysis are presented as a plot of the test statistic against the chromosomal map position, in recombination units (cM). Positions of the markers are shown as triangles. The horizontal line marks the significance threshold. Likelihood ratios above this line are formally significant, with the best estimate of QTL positions given by the chromosomal position corresponding to the highest significant likelihood ratio. Thus, the figure shows five possible QTL, with the best-supported QTL around 10 and 60 cM.
- **Polygenic inheritance** refers to inheritance of a **phenotypic** characteristic (trait) that is attributable to two or more genes and can be measured quantitatively.
- **Multifactorial inheritance** refers to polygenic inheritance that also includes interactions with the environment.
- Unlike **monogenic traits**, polygenic traits do not follow patterns of Mendelian inheritance (discrete categories).
- Instead, their phenotypes typically vary along a continuous gradient depicted by a bell curve.
- An example of a polygenic trait is **human skin color** variation. Several genes factor into determining a person's natural skin color, so modifying only one of those genes can change skin color slightly or in some cases, such as for SLC24A5, moderately.
- Many disorders with **genetic components** are polygenic, including autism, cancer, diabetes and numerous others.
- Most phenotypic characteristics are the result of the interaction of multiple genes.
- **Multifactorial traits in general**
- Traits controlled by the both environment and genetic factors. Usually, multifactorial traits outside of illness result in what we see as **continuous characteristics** in organisms, especially human organisms such as: height, skin color, and body mass.
- All of these phenotypes are complicated by a great deal of give-and-take between genes and environmental effects. The continuous distribution of traits such as height and skin color described above, reflects the action of genes that do not manifest typical patterns of dominance and recessiveness. Instead the contributions of each involved locus are thought to be additive. Writers have distinguished this kind of inheritance as *polygenic*, or *quantitative inheritance*.
- Thus, due to the nature of polygenic traits, inheritance will not follow the same pattern as a simple **monohybrid** or **dihybrid cross**. Polygenic inheritance can be explained as Mendelian inheritance at many loci, resulting in a trait which is normally-distributed. If **n** is the number of involved loci, then the coefficients of the binomial expansion of $(a + b)^{2n}$

will give the frequency of distribution of all n allele combinations. For a sufficiently high n, this binomial distribution will begin to resemble a normal distribution. From this viewpoint, a disease state will become apparent at one of the tails of the distribution, past some threshold value. Disease states of increasing severity will be expected the further one goes past the threshold and away from the mean

The most important applications of QTL analysis **are marker-assisted selection (MAS) and the QTL cloning**.

Using **MAS**, hidden wild (скрытый дикий) alleles can be revealed and new genetic material can be produced and inserted to other plants without using exhausting traditional breeding methods.

Other applications of QTL analysis include understanding the genetics of complex traits and plant genomics.

It should be noted that the QTLs are large chromosomal regions, which in the most cases, cause the problem of linkage between the desirable QTLs
and undesirable traits.

Therefore, one of the main objectives of QTL analysis is limiting the QTL region to smaller chromosomal regions.

The success of this operation depends on many things such as:

1. **The type of experimental design,**
2. **the type of segregating population,**
3. **the rate of markers polymorphism and**
4. **the statistical methods used in QTL analysis.**

Precise selection of the complex traits: polygenic traits are often difficult to select for using conventional breeding approaches.

DNA markers linked to QTL allow them to be treated as single Mendelian factors.

Beside analyzing and selecting the interesting characters, molecular markers allow the researchers also to analyze the wild species with
potential interest for the breeding program.

Numerous articles that consist of DNA markers are available. But, if all markers are useful?

- Applications of Molecular Markers in Plants

1. Construction of linkage maps and QTL mapping

One of the main applications of DNA markers in agricultural research is the construction of linkage maps for different types of crops.

Linkage maps are used to **identify chromosomal regions** that contain
single gene traits (controlled by a single gene) and quantitative traits using QTL analysis.

In other words, in order to use the genetic information that has
been provided by molecular markers, it is important
to know the relative location of molecular markers on
chromosomes.

Linkage map can be considered as a "road map“ of chromosomes from two different parents.

Genetic linkage maps show the relative distances between markers along the chromosomes which are similar to the symptoms and signs along the highway.

The most important application of linkage maps is the identifification of chromosomal regions containing QTLs and genes associated with desired traits. Such maps can be named **QTL maps.**

- **Heritable disease and multifactorial inheritance**
- A mutation resulting in a disease state is often recessive, so both alleles must be mutant in order for the disease to be expressed phenotypically. A disease or syndrome may also be the result of the expression of mutant alleles at more than one locus. When more than one gene is involved, with or without the presence of environmental triggers, we say that the disease is the result of multifactorial inheritance.
- The more genes involved in the cross, the more the distribution of the genotypes will resemble a normal, or Gaussian distribution. This shows that multifactorial inheritance is

polygenic, and genetic frequencies can be predicted by way of a polyhybrid Mendelian cross. Phenotypic frequencies are a different matter, especially if they are complicated by environmental factors.

- The paradigm of polygenic inheritance as being used to define multifactorial disease has encountered much disagreement. Turnpenny (2004) discusses how simple polygenic inheritance cannot explain some diseases such as the onset of Type I diabetes mellitus, and that in cases such as these, not all genes are thought to make an equal contribution.
- The assumption of polygenic inheritance is that all involved loci make an equal contribution to the symptoms of the disease. This should result in a normal curve distribution of genotypes. When it does not, the idea of polygenetic inheritance cannot be supported for that illness.
-
- 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.
- The ANOVA approach for QTL mapping has three important weaknesses. First, 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. Second, we must discard individuals whose genotypes are missing at the marker. Third, 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.
- 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.
- **Interval mapping**
- The principle for QTL mapping is:
- **1) The Likelihood 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 can be established by permutation testing.**
- 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. As a consequence, the power of detection may be compromised, and the estimates of locations and effects of QTLs may be biased.

- Even nonexisting so-called "ghost" (призрак) QTLs may appear.
- 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.
- Why Association Mapping?

The precision of QTL mapping largely depends on the genetic variation (or genetic background) covered by a mapping population, the size of a mapping population, and number of marker loci used.

Once QTLs affecting a trait of interest are accurately tagged using above-outlined approach, marker tags are the most effective tools in a crop improvement that allows the mobilization of the genes of interest from donor lines to the breeding material through MAS. Although traditional QTL mapping will continue being an important tool in gene tagging of crops, it is a 'now classical approach' and suffers from a number of limitations.

First, allelic variation in each cross is usually restricted because typically only two parents are used to create a QTL mapping population.

Second, since early generation crosses are used, the number of recombination events per chromosome is usually small.

Third, a typical QTL identified from a cross consisting of a few hundred offspring can span anywhere between a few to tens of centimorgans encompassing several megabases.

Such large genome regions contain, typically, hundreds if not thousands of genes, making the process of identifying the causal gene in a QTL region a tedious and quite time-consuming task through map-based cloning.

In addition, for many organisms the generation of mapping populations is either not possible or at least very time-consuming.

For instance, the long generation time of most forest trees have completely prevented any progress in elucidating the genetic basis of complex traits using QTL mapping experiments.

These limitations, however, can be reduced with the use of 'association mapping'.

Turning the gene-tagging efforts from biparental crosses to natural populations (or germplasm collections) and from traditional QTL mapping to linkage disequilibrium (LD)-based association mapping can lead to the most effective utilization of ex situ conserved natural genetic diversity of worldwide crop germplasm resources. This approach has many major advantages over conventional QTL mapping.

First, a much larger and more representative gene pool can be surveyed.

Second, it bypasses the expense and time of mapping studies and enables the mapping of many traits in one set of genotypes.

Third, a much finer mapping resolution can be achieved, resulting in small confidence intervals of the detected loci compared to classical mapping, where the identified loci need to be fine-mapped.

Finally, it has the potential not only to identify and map QTLs but also to identify the causal polymorphism within a gene that is responsible for the difference in two alternative phenotypes.

- General Steps in Association Mapping

The performance of association mapping includes the following general steps:

(1) selection of a group of individuals from a natural population or germplasm collection with wide coverage of genetic diversity;

(2) measuring the phenotypic
characteristics (yield, quality, tolerance, resistance, etc.) in the population, preferably, in different environments and multiple replication/trial design;
(3) Genotyping the mapping population individuals with molecular markers;
(4) quantification of the extent of LD of a chosen population genome using molecular marker data;
(5) assessment of the population structure (the level of genetic differentiation among groups within a sampled population individuals) and kinship (coefficient of relatedness between pairs of each individuals within a sample); and
(6) based on information gained through quantification of LD and population structure, correlation
of phenotypic and genotypic data with the application of an appropriate statistical approach that reveals 'marker tags' positioned within close proximity of targeted trait of interest.

- Types of Association Mapping

Association mapping generally falls into two broad categories: (1) candidategene-based association mapping, which relates polymorphisms in selected candidate genes that have purported roles in controlling phenotypic variation for specific traits and (2) genome-wide association mapping (GWAM), which surveys genetic variation in the whole genome to find signals of association for various complex traits (Risch and Merikangas 1996; Zhu et al. 2008).
candidate-gene-based association study is more hypothesis-driven than a genome-wide study because mapping is restricted to genes thought to be good candidates for controlling the trait of interest (Neale and Savolainen 2004; Hall et al. 2010). Although the selection of candidate genes is not always straightforward, genes are usually selected based on information obtained from, for instance, genetic, biochemical, mutation, physiology, or expression studies in both model and non-model plant species. The construction of molecular linkage maps based on the genes [e.g., expressed sequence tags (ESTs), EST-SSRs)] is another way of identifying the candidate genes underlying QTL, instead of time-consuming fine mapping (Sehgal and Yadav 2009). Standard neutrality tests applied to DNA sequence variation data can also be used to select candidate genes or amino acid sites that are putatively under selection for association mapping. This is one of the effective alternative strategies in association mapping that allow reducing the total amount of marker genotyping in less number of individuals. This increases the power and precision of the trait–marker correlations. However, it is important to remember that a candidate gene approach is limited by the choice of candidate genes that are identified and hence always runs the risk of missing out on identifying causal mutations located in non-identified candidate genes.

- Candidate-Gene-based Association Studies in Plants

The candidate gene strategy has shown promise for bridging the gap between quantitative genetic and molecular genetic approaches to study complex traits (Cattivelli et al. 2008; Ingvarsson and Street 2011). Along this line, important studies on association mapping with the candidate gene approach are summarized as follows. Vernalization requirement in wheat is controlled by four major genes, viz. VRN1, VRN2, VRN3, and VRN4, with VRN1 gene copies VRN-A1, VRN-B1, and VRN-D1 located on the long arms of chromosomes 5A, 5B, and 5D, respectively. An association mapping study conducted by Rousset
et al. (2011) on 235 hexaploid wheat collections revealed the effects of the flowering time candidate genes in modulating flowering time in wheat. In that study, genetic variation in VRN-A1, VRN-B1, and VRN-D1 genes explained a large part of phenotypic variation in growth habit. Huang and Brûlé-Babel (2012) studied

genetic diversity, haplotype structure, and association of genes involved in starch biosynthesis in wheat. Genes encoding granule-bound starch synthase (GBSSI, also
known as waxy or Wx genes) and soluble starch synthase (SSIIa) were selected for nucleotide diversity and SNP density study. None of the SNPs within the three
SSIIa genes and Wx-D1 gene was associated with yield-related traits.
However, both SNPs and haplotypes within the Wx-A1 gene were associated with seed number per spike, seed weight per spike and thousand kernel weight. Another study on grain size of wheat also demonstrated the association of haplotype of a grain size gene (TaGW2) with larger grain size, earlier heading date and maturity in hexaploid wheat (Su et al. 2011).
Similarly, transcription factors such as the gibberellin-regulated Myb factor (GAMYB), the barley leucine zippers 1 and 2 (BLZ1, BLZ2), and the barley prolamin box binding factor (BPBF) were evaluated for their association with agronomic traits in barley (Haseneyer et al. 2010).
SNPs within BLZ1 were associated with days to flowering, and its haplotype was also associated with both days to flowering and plant height. The haplotype of BLZ2 was associated with thousand kernel weight, while the haplotype of the BPBF gene was associated with both crude protein and starch in barley endosperm . However, the candidate genes explained only a small portion of the total
genetic variation. Similarly for sorghum and rice, candidate genes involved in
starch biosynthesis were associated with the expected traits and the results were in agreement with QTL studies (Bao et al. 2006; Figueiredo et al. 2010).

L. 15. METHODS OF STUDYING THE SEQUENCES OF THE NUCLEIC ACIDS FRAGMENTS.

Objectives
Compare and contrast the chemical (Maxam-Gilbert) and chain termination (Sanger) sequencing methods.
List the components and molecular reactions that occur in chain termination sequencing.

Sequencing Methods
Maxam-Gilbert chemical sequencing
Sanger chain termination sequencing
Maxam-Gilbert Sequencing
Maxam-Gilbert Sequencing
Chain Termination (Sanger) Sequencing
Chain Termination (Sanger) Sequencing
A modified DNA replication reaction.
Growing chains are terminated by dideoxynucleotides.

Steps:
- Denaturation
- Primer attachment and extension of bases
- Termination
- Gel electrophoresis

Chain Termination (Sanger) Sequencing
The 3′-OH group necessary for formation of the phosphodiester bond is missing in ddNTPs.
Chain Termination (Sanger) Sequencing
A sequencing reaction mix includes labeled primer and template.
Dideoxynucleotides are added separately to each of the four tubes.
Chain Termination (Sanger) Sequencing
A sequencing reaction mix includes labeled primer and template.

Dideoxynucleotides are added separately to each of the four tubes.
Chain Termination (Sanger) Sequencing
Chain Termination (Sanger) Sequencing
Chain Termination (Sanger) Sequencing
With addition of enzyme (DNA polymerase), the primer is extended until a ddNTP is encountered.
The chain will end with the incorporation of the ddNTP.
With the proper dNTP:ddNTP ratio, the chain will terminate throughout the length of the template.
All terminated chains will end in the ddNTP added to that reaction.
Chain Termination (Sanger) Sequencing
The collection of fragments is a sequencing ladder.
The resulting terminated chains are resolved by electrophoresis.
Fragments from each of the four tubes are placed in four separate gel lanes.
Automated Version of the Dideoxy Method
Chain Termination (Sanger) Sequencing
Summary
Genetic information is stored in the order or sequence of nucleotides in DNA.
Chain termination sequencing is the standard method for the determination of nucleotide sequence.
Dideoxy-chain termination sequencing has been facilitated by the development of cycle sequencing and the use of fluorescent dye detection.
The dideoxy method is good only for 500-750bp reactions
Expensive
Takes a while