Biyani's Think Tank

Concept based notes

Genome Analysis

[B.Sc. Biotech Part-III]

Pragya Dhakar

M.Sc.

Assoc. Prof.

Department of Biotechnology

Biyani Girls College, Jaipur

Published by :
Think Tanks
Biyani Group of Colleges

Sector-3, Vidhyadhar Nagar,
Jaipur-302 023 (Rajasthan)
Ph : 0141-2338371, 2338591-95 • Fax : 0141-2338007
E-mail : acad@biyanicolleges.org
Website :www.gurukpo.com; www.biyanicolleges.org

ISBN: 978-93-81254-28-8

Edition : 2011
Price:-

Leaser Type Setted by :
Biyani College Printing Department

Preface

I am glad to present this book, especially designed to serve the needs of the students. The book has been written keeping in mind the general weakness in understanding the fundamental concepts of the topics. The book is self-explanatory and adopts the "Teach Yourself" style. It is based on question-answer pattern. The language of book is quite easy and understandable based on scientific approach.

Any further improvement in the contents of the book by making corrections, omission and inclusion is keen to be achieved based on suggestions from the readers for which the author shall be obliged.

I acknowledge special thanks to Mr. Rajeev Biyani, *Chairman* & Dr. Sanjay Biyani, *Director* (*Acad.*) Biyani Group of Colleges, who are the backbones and main concept provider and also have been constant source of motivation throughout this Endeavour. They played an active role in coordinating the various stages of this Endeavour and spearheaded the publishing work.

I look forward to receiving valuable suggestions from professors of various educational institutions, other faculty members and students for improvement of the quality of the book. The reader may feel free to send in their comments and suggestions to the under mentioned address.

Note: **A feedback form is enclosed along with think tank. Kindly fill the feedback form and submit it at the time of submitting to books of library, else NOC from Library will not be given.**

Pragya Dhakar

Syllabus

Section-A

- Basic structure of prokaryotic and eukaryotic genome.
- Chromatin model, concept of gene, linkage and crossing over, linkage analysis in *Drosophilla* and *Neurospora*.

Section-B

- Pedigree analysis in human genetic mapping and its tools, genetic mapping of complex character.
- Multigene families in human genome and repetitive DNA
- C-Value paradox and complexity of genome.

Section-C

- Physical map of genome
- Chromosome Walking and Chromosome painting
- FISH, GISH, Zoo Blot, Dot Blot, VNTR, RFLP's RAPD, RACE, SNP's, OTL's, EST, CPG island identification, Exon trapping and sequence analysis

Section-D

- Rice genome project and its applicability.
- Human Genome project and its application to future of mankind.

Content

Section A

Structure of Prokaryotic and Eukaryotic Genome.

Q.1 How are genes and other sequence features organized in a typical prokaryotic genome?

Ans Prokaryotes are organisms whose cells lack extensive internal compartments compared to eukaryotes.

Organization of genome in prokaryotes:-

Genome organization in prokaryotes has been widely studied in *E.coli* according to current models *E.Coli* DNA consist of 40-50 supercoiled loops attached to a protein core. Each loop consist of 100 Kb of super coiled DNA, the protein component of the nucleoid includes DNA gyrase and DNA topoisomeraseI these two are responsible for maintaining super coiled state of DNA. Instead of histones the packaging proteins include HU, which are structurally different to eukaryotic histones but acts in a similar way forming a tetramer around which ~60 up of DNA becomes wound.

Linear or multipartite genome:-

Vast majority of bacteria consist of single, circular DNA molecular while there are some prokaryotes with multipartite genomes.

Multipartite genomes refer to the type of genome which are divided into 2 or more DNA molecules .e.g. *Borelli burgdorferi B31* with 7-8 circular molecules and eleven linear molecules.

Integrons:-

Plasmid contains a set of genes and other DNA sequence that enable plasmids to capture genes from Bacteriophages and other plasmids these are called integrons.

Organization of genes in prokaryotes:-

Bacterial genomes have compact genetic organizations with very little space between genes. Taking *E.coli* as an example which consist of 50 kb segment where genes are

organized very closed to each other with very less or no space in b/w. About 43 genes of *E.Coli* taking up 85.9% of the segment.

A characteriotic feature of gene organization in prokaryotic in the presence of Operon- a group of genes involved in a single biochemical pathway and expressed in conjunction with one another e.g. Thereonin operon where three genes namely thr A ,Thr B(separated by a single nucleotide) and thr C are involved in synthesis of amino acid thronine.

Prokaryotic genes are comparatively shorter than eukaryotes ever after intrans have been removed. The average length of a bacterial gene being about 2/3 rd that of eukaryotic gene.

There are certain other features of prokaryotic genes first is the absence of introns.In E.Coli there are no discontinuous genes, leaving some archeas.

Second feature is the infrequency of repetitive sequence prokaryotic don't posses high copy number genome like eukaryotes but they are consist of certain sequences that might be repeated elsewhere in the genome e.g. in the in certain sequence like IS and ISI86 present in 50Kb segment of *E.Coli*.

These may vary in their position or might be absent in other isolates.

Gene organization in Operons :-

An operon in a group of genes that are located adjacent to one another in the genome, with just one or two nucleotides b/w the end of one gene and start of the next gene. All the genes in an operon are expressed as a single unit.

E.coli Lactose operon was the first operon to be discovered consisting of 3 genes involved in conversion of the disaccharide sugar lactose into its monoaccuharide unit- glucose and galactose . Glucose and galactose form outstare for glycotic pathway, so the function of genes in the lactose operon is to convert lactose in a form that can be utilized by *E.Coli* as an energy source. Not being a common componentof *E.Coli* natural environment the operon is not expressed maximum time and so enzymes for lactose utilization are not synthesized. Operon becomes functional only when lactose is available.

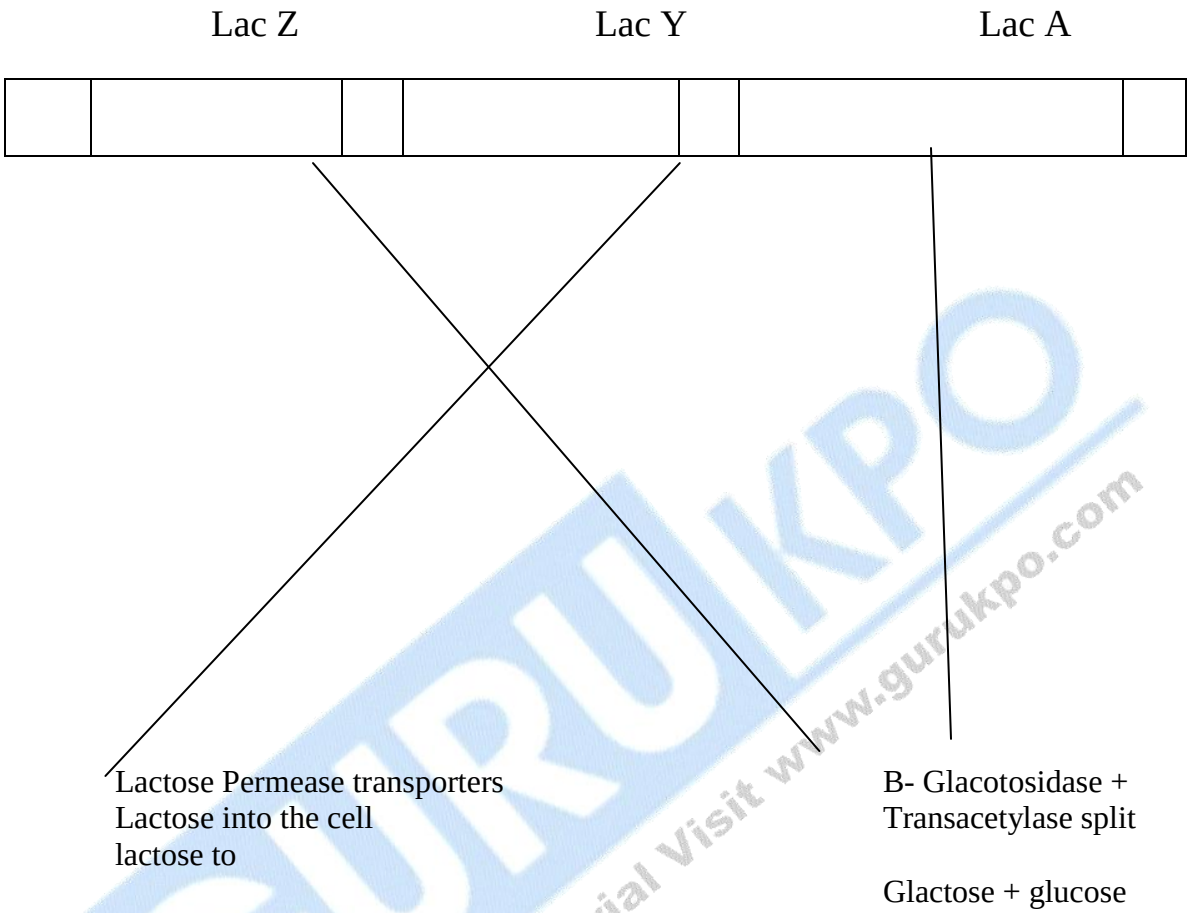


Figure- Lactose Operon

The three genes are called lac Z,lacY and lac A, the first two separated by 52bp and the second two by 64 bp. All three are expressed together, lac Y coding for the lactose permease that transports lactose into the cell and lac Z and lac A coding for enzyme that split lactose into its component sugars- glucose and galactose.

Q.2 Describe Eukaryotic genome giving a suitable example?

Ans Eukaryotic genome is split into a set of linear DNA molecule, each moleculecontained in chromosomes. All Eukaryoties that have been studied have at least two chromosomes and the DNA molecules are always linear. The only variability at this level of Eukaryotic genome structure lies with chromosomes number, which appears to be up related to the biological feature of the organism. For example yeast has its chromosomes 4 times that of fruit fly.

Genetic features of Eukaryotic nuclear genomes:-

Gene study:-

In most eukaryotes genes are distributed randomly with substantial variation in gene density at different position within a chromosomes. The average gene density in *Arabidopsis* is 25 genes per 100kb, but even outside of the centromeres and telomeres the density varies from 1 to 38 genes for 100kb, for the largest of the plant five chromosomes. The same is true for human chromosomes, where the density ranges from 0 to 64 genes per 100kb.

Genes are distributed unevenly within human chromosomes. This has been shown by banding patterns that are produced when chromosomes are stained. This has also been evidenced by Isochore Model. According to this model the genome of vertebrates and plants (and possibly of other eukaryotes are mosaics of segments of DNA, each at least 300kb in length, with each segment having a uniform base composition that differs from that of the adjacent segments. Some experiments close in support of isochore model comes from experiments in which genomic DNA is broken into fragments of approximately 100kb treated with dyes that binds specifically to AT or GC rich regions, and the pieces separated by density gradient centrifugation. When this experiment was carried with human DNA, five fractions are seen each representing a different isochore type with a distinctive base composition: two AT-rich isochores called L1 and L2 and three GC-rich classes called H1, H2and H3. The last of these, H3 is the least abundant in the human genome, making up only 3 % of the total, but contains over 25% of the genes.

Organization of genes within a nuclear genome:-

The over all pattern of gene organization varies greatly between different eukaryotes as they reflect important distinctions between the genetic features and evolutionary histories of these genomes.

- Genes make up only a small part of the human genome. Taking an e.g. of 50kb segment of chromosomes 12 to understand the genes organization.

It consists of mainly 4 genes:

(a) PKP 2:- Codes for plakophilin 2 protein involved in synthesis of desmosomes, structures that act as connections point's b/w adjacent mammalian cells.

(b) SYBI: - specifying a vesicle- associated membrane protein whose role is to ensure that vesicle fuse with their correct target membranes within the cell.

(c) A gene whose function has not yet been identified, called FL JI10143.

(d) CD 27, coding for a member of the tumor necrosis factor receptor super family, a group of proteins that regulate signal transduction pathways involved in apoptosin and cell differentiation.

- All these 4 genes are discontinuous, the no. of introns ranging from two for SYBI to eight for PKP2.
- Consist of 88 genome wide repeat sequence. These are sequences that recur at many places in the genome. There are four main types of genome wide repeats called LINES (Long interspersed elements), LTR (Long terminal repeat) elements and DNA transposons. Most of the genome- wide repeats are located in the intergenic region but several lies within introns.
- 7 microsatellites are present. These are sequences in which a short motif is repeated in tandem. Four of the 7 microsatellites are located within introns.
- Lastly ~ 30% of our 50Kb segment of the human genome is made up of stretches of monogenic, nonrepetetive, single –copy DNA of no k/n function or significance.
- This 50 kb segment of human genome consists of relatively small amount of space taken up by genes. When added together the total length of the exon the part of the 4 genes that contains the biological information is 4745bp, equivalent to 9.5% of the 50kb segment. But compared to other segments is richer where all the exons in the human genome makeup only 48mb, just 1.5% of the total. In contrast, 44% of the genome is taken up by genome-wide repeats. e.g. yeast (S.cerevisiae) genome:-

  The nuclear genome of the yeast is about 12 Mb which is 0.004 times that of human nuclear genomes.

  It consist of 6000 genes.

Taking example of chromosomes III of yeast genome-has following distinctive features:-

(a) Contains 26 genes thought to code for first and two that code for transfer RNA's.

(b) Few of the yeast genes are discontinuous. But this segment does not possess any such gene. In entire yeast genome there are only 239 introns, compared with over 300000 in the human genome.

(c) There are fewer genome-wide repeat. This part of chromosome contains a single LTR elements c/a Y2 and 4 truncated LTR elements called delta sequence. These five genome-wide repeats make up 13.5% of the 50kb segment, but this figure is not typical of the yeast genome as a whole. When all 16 yeast chromosomes are considered.The total amount of sequence taken up by genome-wide repeat is only 3.4% of the total.

Q.3 Differentiate between Prokaryotic and Eukaryotic genome?

Ans Major differences between prokaryotic and Eukaryotic genome are marked under following heads:

	Prokaryotic Genome	**Eukaryotic Genome**
1	Chromosomal DNA	
	circular	Linear
2	**Mechanism of duplication through replication:-** Chromosomal DNA in prokaryotes posses' single origin of replication that is discrete to each chromosomes and which is recognized by no. of factors. For replication of DNA there are three enzymes known as DNA pol I, II, and III.	DNA replication occurs at multiple nodes along the DNA sequence, while the origin of each initiation site is not much specific, the multiple sites at which replication initiates allows for a fundamentally more rapid genome replication. Replication in eukaryotes there are at least four distinct DNA for (a, b, d and g) which are involved in different aspects of DNA replication but in addition require lot more Eukaryotic genome. Replication factors that enhance specificity and activity of the different DNA dependent DNA pol.
3	**Difference at the level of chromosomes** • Prokaryotic chromosomes do not show the features of mitosis and meiosis which are a type of cell division. Instead during cell cycle chromosomes attach to specific sites on the cytoplasmic membrane; the growth of the membrane between the sites brings about the separation of the daughter chromosomes. The same mechanism may	• Eukaryotic chromosomes are linear DNA molecules and undergo specific rearrangement i.e. mitosis and meiosis .Certain other features are specific to eukaryotes includes the centromeres which are present on the chromosomes, but one per chromosome. They are the

	be involved in plasmid segregation, each plasmid having its own membrane attachment site.	attachment sites for microtubules of spindle fibers, to allow for movement of the chromosomes towards the poles during cell division.
	Histones:- Histones are absent in prokaryotes but in E.Coli a single type of histone like protein called HU is found. Chromosomes of prokaryotes contain DNA molecules with unique (Non-repeated) base pair sequences. That is each gene is present only once in the genome. (gene for rRNA are exception)	They are associated with DNA in chromosomes. They are of 5 type namely H1,H2A,H2B,H3 and H4 In eukaryotic genome chromosomes are much more complex. In then certain base sequences are repeated many times.In the haploid chromosomes complement sometimes so many or as million times in mammals. Such DNA sequence are termed repetitive DNA which represents (20-50%) of eukaryotic genome.
4	**DNA Packaging** In Prokaryotes single circular chromosomes is present where DNA packaging is not as complex that of eukaryotes and are not associated with proteins like Histones.	In case of eukaryotes DNA packaging is very compact as DNA molecule is greater in size then chromosomes. DNA is associated with non-specifically inding proteins c/a Histones. These are basic in nature are of 5 types H1, H2A, H2B, H3 and H4. They help in coding and packing. The second type units involved in packing are nucleosomes which are consists of core particle and a small spacer /linker DNA. the core particle consist of histones. These units together form a complex and a highly condensed structure for DNA packaging.

Q.4 Underline the important feature of Chromatin model with well labeled diagrams?

Ans **Chromatin**:- In Eukaryotic cells genetic material is organized into a complex structure composed of DNA and protein and localized in a compartment of the nucleus. This structure is called chromatin.

The term chromatin has been derived from Greek “Khroma” meaning coloured and “soma” meaning body.

Important features of chromatin model:-

The fundamental units of chromatin are termed as nucleosomes which are composed of DNA and Histone protein. This structure provides the first level of compaction of DNA into the nucleus. Nucleosomes are regularly spaced along the genome to form a nucleofilament which gives a higher level of compaction, ultimately resulting in the highly condensed metaphase chromosomes. Within interphase nucleus chromatin is organized into functional territories.

Chromatin has been divided into two types:-

(a) Euchromatin - is decondensed during interphase.
(b) Heterochromatin is localized principally on the periphery of the nucleus and euchromatin in the interior of the nucleoplasm.
(c) Heterochromatin can be defined as a structure that does not alters and remain condensed throughout the cell cycle.

Constitutive Heterochromatin:-

Contain few genes and formed principally of repetitive sequences located in large region coincident with centromeres and telomere.

Facultative Heterochromatin:-

Composed of transcriptionaly active requires that can adopt the structural and functional character of heterochromatin s/a the inactive x chromosomes of mammals.

Nucleosomes:-

These are fundamental packing unit of chromatin and give chromatin the "beads on a string" appearance.
Each nucleosome is disc-shaped about 10nm in diameter. It consists of a core particle and a small spacer or linker DNA.

(a) Core Particle:-

Consists of octamer of histones, having two copies each H2A, H2B, H3 and H4. It is about 11nm in diameter and 6nm in height. A DNA strand of 146bp is tightly wrapped around this core forming two circles.

(b) Spacer DNA:-

It is a small segment of DNA having just four bp one unit of histone H1 is associated with it.

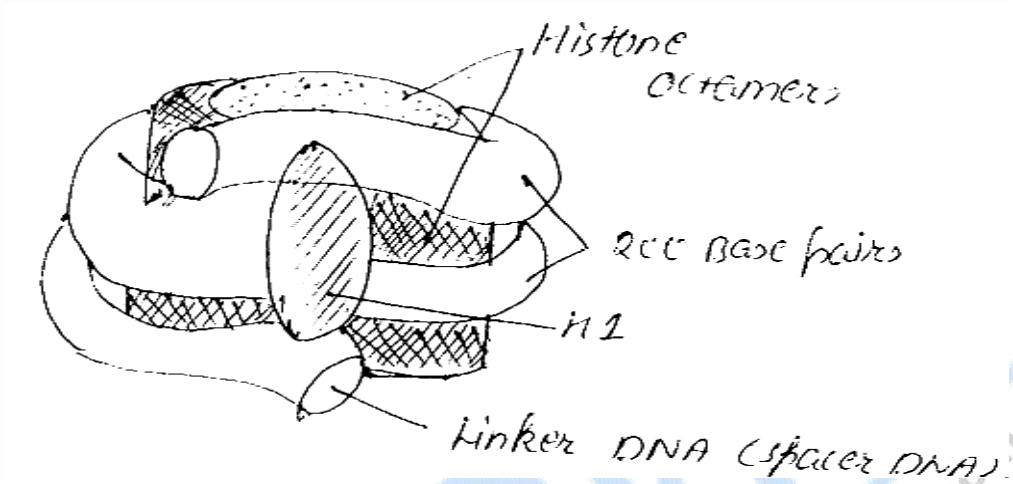


Figure:- 200 A single nucleosomes showing octamer histones,H1 histones and base pairs including DNA.

The length of the linker region varies b/w species and cell type. It is within this region that the variable linker histones are incorporated. Therefore the total length of DNA in the nucleosome can vary with species from 160 to 241 base pairs.

Histone Proteins:-

(a) Core Histones:-

H3, H4, H2A and H2B are small basic protein highly conserved in evolution. These are responsible for coiling DNA into nucleosome. Each of them is formed of about 102-135 amino acids. These form the inner core of nucleosome. Nucleosomal histones are very similar in different species (i.e. these have not changed during evolution). These are present in equivalent amounts two of each type being present at every 200 bp thus forming a histone octamer or core particle.

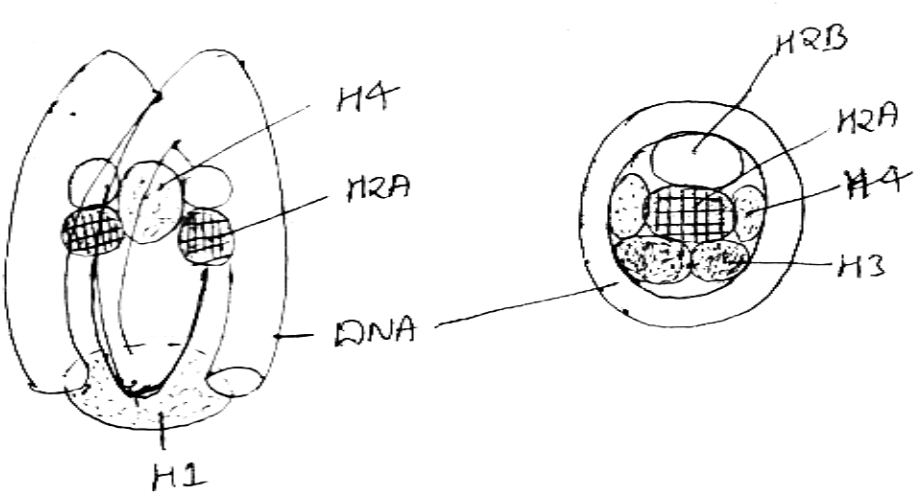


Figure: - Schematic diagram of a nucleosome showing structure of octamer of histones with DNA helix. H1 is not a member of inner circle and technically not a part of core particle.

Structure:-

Structurally are composed of "histone fold domain" consisting of 3-a helices separated by 2 loops regions. Their N terminal tails are more variable and unstructured. The tails are rich in lysine and arginine residues making them extremely basic.

(b) Linker Histones(H-1):-

H-1 are associated with the linker region of DNA between two nucleosome core and unlike the core histones, they are not well conserved between species. They are composed of 3 domains in higher eukaryotes a globular, non-polar central domain essential for interaction of DNA and two non-structural N and C terminal tails that are highly basic and proposed to be the site of post translational modifications. The linker Histones are associated with DNA have a role in spacing nucleosomes and can modulate higher order compaction by providing and interaction region between adjacent nucleosomes.

Nucleosomes packing or formation of chromatin fibre:-

Nucleosomes are packed upon one another to form a fibre with a diameter of about 30 nm chromatin fibre or nucleoprotein fibre. It appears to have a beaded appearance. The beads are nucleosomes. These are connected by linker DNA.

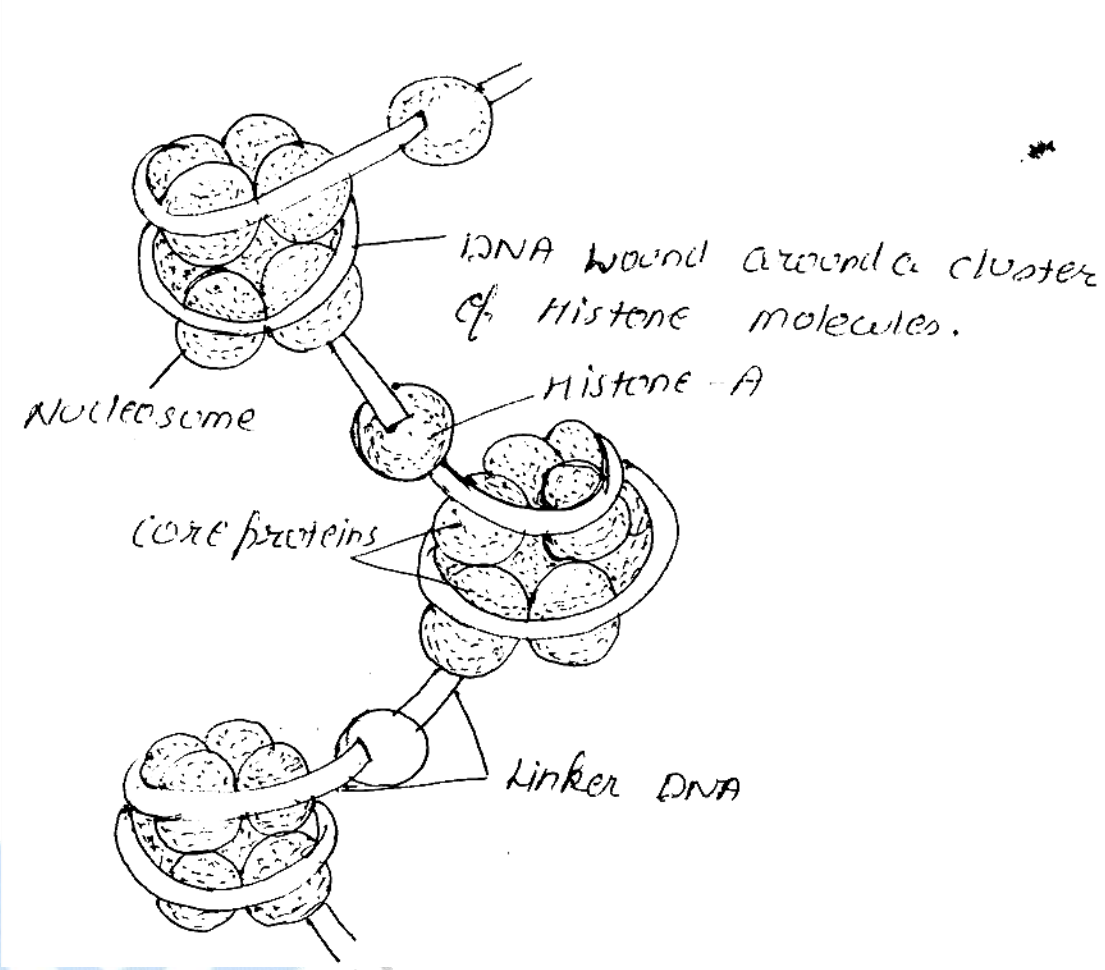


Fig:- Beads on a string, appearance of nucleoprotein fibre.

Q.5 Give a notes on gene concept?

Ans Gene Concept:-

It was introduced by Sutton, studied by Morgan, Bridge and Muller etc. elaborated it. The essential feature of modern concept of genes is:-

- Genes determine the physical as well as physiological characteristics. These are transmitted from parents to the offsprings generations after generations.
- Genes are situated in the chromosomes.

- Since the number of genes in each organism is very large in comparison to the number of chromosomes, several genes are located in each chromosomes. In man about 40000 genes are known to be located on 23 pairs if chromosomes.
- Each gene occupies a specific position in specific chromosomes. This position is k/a locus. The chromosomal aberration (like translocation, inversion) may bring relative change in the specific position of genes and these help in the origin of chromosomes with new sequence of gene and help in the origin of new species.
- Genes in the chromosomes are arranged in a single linear order like the arrangement of beads on a string.
- A single gene may occur in several form or in several functional states. The form other than normal is known as alleles. Many genes have only two alleles one of them is normal and other one is its mutant e.g. red and white colour of flowers is pea plant are controlled by two different alleles of the gene controlling flower colour.
- Only those genes are known which have their alternative alleles.
- The alleles may be related as dominant or recessive but not always.
- Some genes mutate more than once and have more than two alleles. These are known as multiple alleles. Whatever may be the number of alleles in a multiple series only two of them are found in an individual because of the presence of two homologous chromosomes of each type.
- The gene may undergo sudden change in expression due to change in its composition. The changed gene is known as mutant gene and the phenomenon of change in known as mutation.
- Rarely genes from one chromosome may be exchanged or transferred to other chromosomes which may be its homologous counter part (crossing over) or a non-homologous (translocation).
- Genes duplicate themselves very accurately. The phenomenon is k/a replication self duplication of gene heads to chromosomal duplication.
- Genes express themselves by providing enzymes which are protein. It means each gene synthesis a particular protein which acts as enzyme and brings about an appropriate change.
- According to the recent information a gene is a segment of DNA which contains the information for one enzyme or one polypeptide chain coded in the language of nitrogenous bases or the nucleotide. The language of nitrogenous base or the nucleoticle. The sequence of three nucleotide read for one amino acid (codon).

Q.6 Write short notes on:-

(a) Linkage and Crossing over

(b) Linkage Analysis

Ans

(a) Linkage and Crossing over

Linkage:-
The phenomenon of inheritance of genes together and to retain their parental combination even in the offspring is k/a linkage.

The genes located in the same chromosomes and being inherited together is known as linked genes and the characters controlled by these are linked characters.

All those genes which are located in the single chromosome constitute a linkage group. The total number of linkage group in an organism is equal to the number of chromosomes pairs.
For example there are 4 linkage group in *Drosophila melanogaster*, 23 in man and 7 in sweet pea.

The theory of linkage was propounded by T.M. Morgan in 1911.

Chromosome theory of linkage:-
Morgan and Castle formulated this theory having following characteristics:-

(I) Genes that show linkage are situated in the same chromosome.
(II) Genes are arranged in a linear fashion in the chromosomes i.e. linkage of genes is linear.
(III) Distance b/w linked gene is inversely proportional to the strength of linkage. The genes which are closely located show strong linkage, whereas those which are widely reparted, have more chances to get separated, have more chances to get separated by crossing over.
(IV) Linked genes remain in their original combination during course of inheritance.

Type of Linkage:-

Based on absence or presence of non-parental or new combinations of linked genes.

a. Complete Linkage:-

When linked genes are transmitted together to the offspring only in their original or parental combination for two or more or reveral generation exhibit complete linkage In such cases the linked genes don't separate to form the new or non-parental combination. This phenomenon is very rare. Same characteristics in males of Drophilla are found to exhibit complete linkage.

b. **Incomplete Linkage:-**

The homologous chromosomes undergo breakage and reunion during gametogenesis.

During reunion the broken pieces of the chromatin are exchanged, producing some nonparental or new combination. Therefore the linkage is rendered incomplete.

The phenomenon of interchange of chromosome segment between two homologous chromosomes is called crossing over.
Incomplete Linkage is very common and has been osudied in almost all the organism.

e.g. Incomplete Linkage in man – linked genes have been studied in man also. Genes for ABO blood group are linked to the genes for nail papilla syndrome with 10% crossing over. The gene complex for Rh-blood group and gene for elliptocytosis are also linked with a crossing value of 3%

Factor affecting strength of linkage:-

(I) Age of animal:-
With increasing age the chances of crossing over are reduced and therefore strength of linkage increases.

(II) Temperature:-
The increase in temperature increases the frequency of chiasmata formation and therefore it decreases the strength of linkage.

(III) X-Rays:-
Exposure to X-Rays reduces the strength of linkage.

Crossing Over:-
The recombination of linked genes brought about as a result of interchange of corresponding parts b/w the chromatids of a homologous pair of chromosomes, so as to produce new combination of old genes.

E.g. recombination in Drosophilla

A cross b/w a grey bodied, vestigical winged (BBVV) and bloack bodied long winged (bbVV) Drosophilla produced F1 hybrids all of them having grey body and long winged (BbVv). When female flies of F1 generation (BbVv)were crossed with double recessive males having black body and vestigial wings (bbvv) – a test cross four types of offspring were produced as follows-

1.	Grey Vestigial	41.5%	Non cross overs 83%
2.	Black Long	41.5%	
3.	Black Vestigial	8.5%	Cross overs 17%
4.	Grey Long	8.5%	

The recombination of genes or the appearance of nonparental combination is found in 17 % of the offspring.

Essential features of crossing over:-

(I) Only two of the four chromatin cross over, while the other two chromatic preserve their original constitution.

(II) The crossing over occurs b/w the nonsister chromatin of the homologous pair of chromosomes. The two chromatids of the same chromosome never exchange parts.

(III) The number of chiasmata per set of chromosomes depends upon the length of chromosomes. The longer are the chromosomes the greater is the number of chiasmata.

(IV) Chances of crossing over are more if the genes are located farther apart. The more is the distance b/w the genes is the opportunity for chiasma formation.

Type of crossing over:-

1. Single Crossing Over :-

When only one chiasma is formed all along the length of a chromosome pair, it is known as single cross over. The chromatin of homologous chromosomes contact and break only at one point along their entire length.

2. Double Crossing Over :-

In double crossing over chromatids break and rejoin at two points. It means two chiasmata are formed along the entire length of the chromosomes.

They are of three different types:-

(I) Two Strand Double Cross overs:-

The same two chromatide are involved at both the crossing over points. The chiasmata thus formed are k/a reciprocal chiasmata.

(II) Three Strand Double Cross overs:-

In this type three chromatids are involved i.e. the second chiasma contains one of the same chromatid which has crossed at first chiasma but the other chromatid is different. Only one chromatid is non-cross over, while the other three exchange their parts.

(III) Four Strand Double Cross overs:-

In such cross overs all the four chromatids of the tetrad are involved, two of them exchange parts at first chiasma and other two are involved

in the second chiasma. Such chiasmata are k/a complementary chiasmata. These produce 4 single crossovers but no non cross overs.

3. Multiple Cross Over :-

When crossing over occurs at more than two places in the same chromosomes pair and more than two chiasmata are formed, this type of crossing over is k/a multiple crossing over.

Factors affecting crossing over:-

It mainly depends upon the distance between the linked genes, but a no of genetic, environment and physiological factors also affect it.

- Temperature
- X-Ray
- Age
- Chemicals (Mutagens)

- Temperature: - High and low temperature increases the frequency of crossing over.
- X-Ray:- Muller has discovered that exposure to X-Ray and other radiations increases frequency of crossing over.
- Age:- the frequency of crossing over decreases with increasing age in female Drosophilla.
- Chemicals (Mutagens):- it depends on the chemical which can increase or decrease the frequency of crossing over .

Relation b/w linkage and crossing over:-

- Crossing over has led to the construction of linkage maps or genetic maps of chromosomes; crossing over percentage is used to determine the distance between the genes.
- Crossing over between the genes depends upon the distance between linked genes in a chromosome. The chances of crossing over are more between two distantly placed genes as compared to the genes present in close proximity. The inverse in the case with linkage where strength of linkage in inversely proportional to the distance b/w linked genes.

Linkage Analysis:-

Linkage analysis is a method or technique through which correct deduction about gene order, and distance estimates are sufficiently accurate to generate Gene's maps.

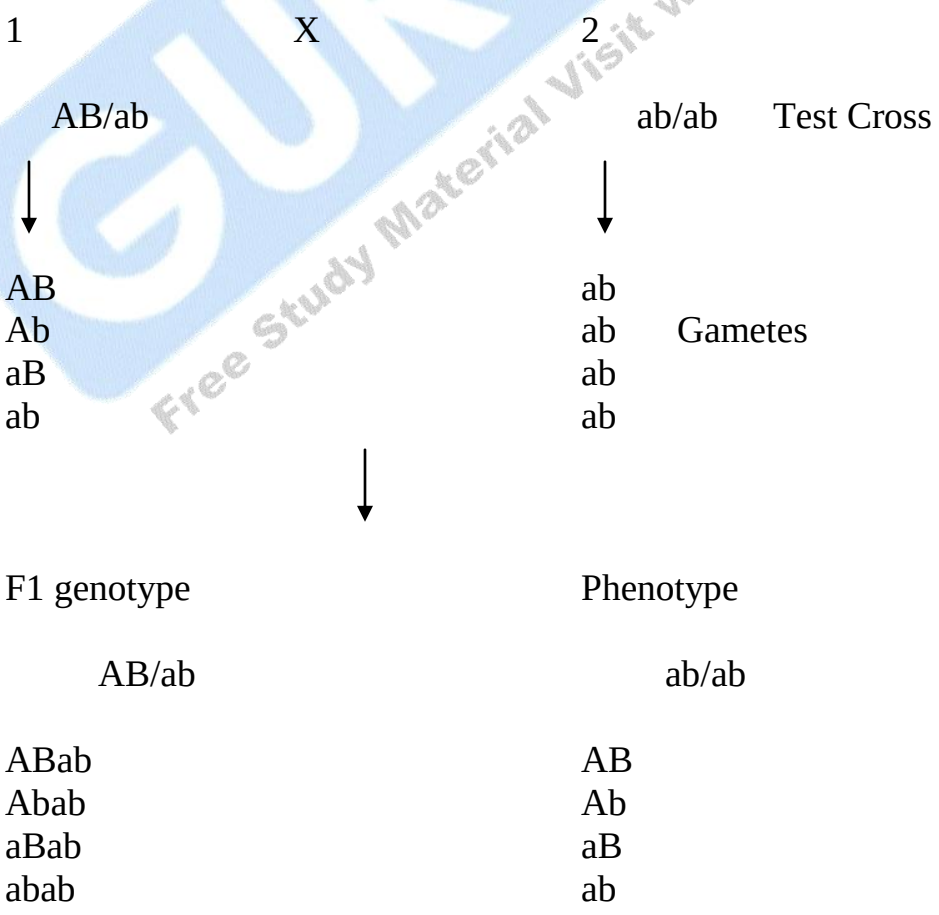
Linkage analysis can be understood by taking a very apt example of fruit flies.

The standard procedure for linkage analysis by using test form. There are 2 genes namely gene A (having alleles A and a) and gene B (alleles B and b) both lie on the chromosomes two of Drosophilla. The critical features of a test cross is the genotype of the two parents.

One parent is a double heterozygous. This means the all four alleles are present in the parent. Its genotype is AB/ab. This nutation indicates that one pair of the homologous chromosomes has alleles A and B, and the other has a and b. Double heterozygotes can be obtain by crossing two pure-breeding strains, e.g. AB/AB * ab/ab.

The second parent is a pure breeding double homozygote. In this parent, both homologous copies of chromosomes 2 are the same: in the e.g. (Fig) both have alleles a and b and the genotype of the parents is ab/ab.

Fig :- A and B are dominant over a and b parents.



Each phenotype is the same as the genotype of the parent 1 gamete.

Fig:- Test Cross b/w alleles displaying dominance and recessiveness.

A and B markers with alleles A and B, a and b. the resulting progeny are scored by examining their phenotype. Because the double homozygous parent (parent 2) has both recessive alleles a and b it effectively makes no contribution to the phenotype of the progeny. The phenotype of each individual in the F1 generation is therefore the same as the genotype of the gamete from parent 1 that gave rise to that individual.

All the gametes produced by the second parent (the double homozygote) will have the genotype ab regardless of whether they are parental or recombinant gametes. Alleles a and b are both recessive, so meiosis in this parent is in effect, invisible when the phenotype of the progeny are examined.

Genes

M	miniature wings
V	vermilion eyes
W	white eyes
Y	yellow body

↓

Recombination frequencies

Between M and V	3.0%
Between M and Y	33.7%
Between V and W	29.4%
Between W and Y	1.3%

Deduced map position

Y	W	V	M
0	1.3	30.7	33.7

Fig: - Working out a genetic map from recombination frequencies.

All four genes are on the X chromosome of the fruit fly.This type of linkage in the figure above phenotype of the diploid progeny can be converted into the genotype of the gametes from the double heterogenous parent. The test cross therefore enable us to make a direct examination of a single meiosis and hence to calculate a recombination frequency and map distance for the two genes being studied.

This type of linkage analysis is the modern counterpart of the method developed by Morgan and his colleagues. The above method is based on the analysis of the progeny of experimental crosses setup b/w parents of k/n gene type and in applicable to all eukaryotes.

Q.5 Describe linkage analysis in *Neurospora*?

Ans Linkage analysis in *Neurospora*:-

Neurospora is a fungi and is one of the must favorites material with genetistics as:

- The life cycle of *Neurospora* enables to analyze the products of a single meiosis and life cycle is short too.
- Its meiotic products are linearly arranged in ascus as 8 ascospores as ordered tetrads in the same order in which chromatids were on the meiotic metaphase plates.
- For understanding linkage analysis in Neurospora analysis of ordered tetrads can be taken as an example.

Tetrad analysis in *Neurospora:-*

In Neurospora the nuclei from hyphae of opposite mating type (+) and (-) fuse to form a diploid Zygote. The zygote is the only diploid stage in the life cycle of Neurospora. The zygote nucleus divides meiotically providing four haploid nuclei, each of which then undergoes mitosis. The eight cells so produced for 8 haploid ascospores enclosed in the ascus. The three divisions proceed along the longitudinal axis, so that the ascopores are arranged in a line in a specific order that indicates the order of arrangement of chromatids on the meiotic metaphase spindle. This is called ordered tetrad. Each of the four products of meiosis can be cut separately to study their phenotype thereby inferring their genotype. This is called tetrad analysis.

a) First division or segregation between centromere and Gene-a

A cross b/w two cultures of Neurospora, one normal (a+) and other mutant (a). in the linear arrangement of ascospores in ascus is 4a+: 4a. it indicates the absence of crossing over between gene-a and centromere. This is described as first division segregation.

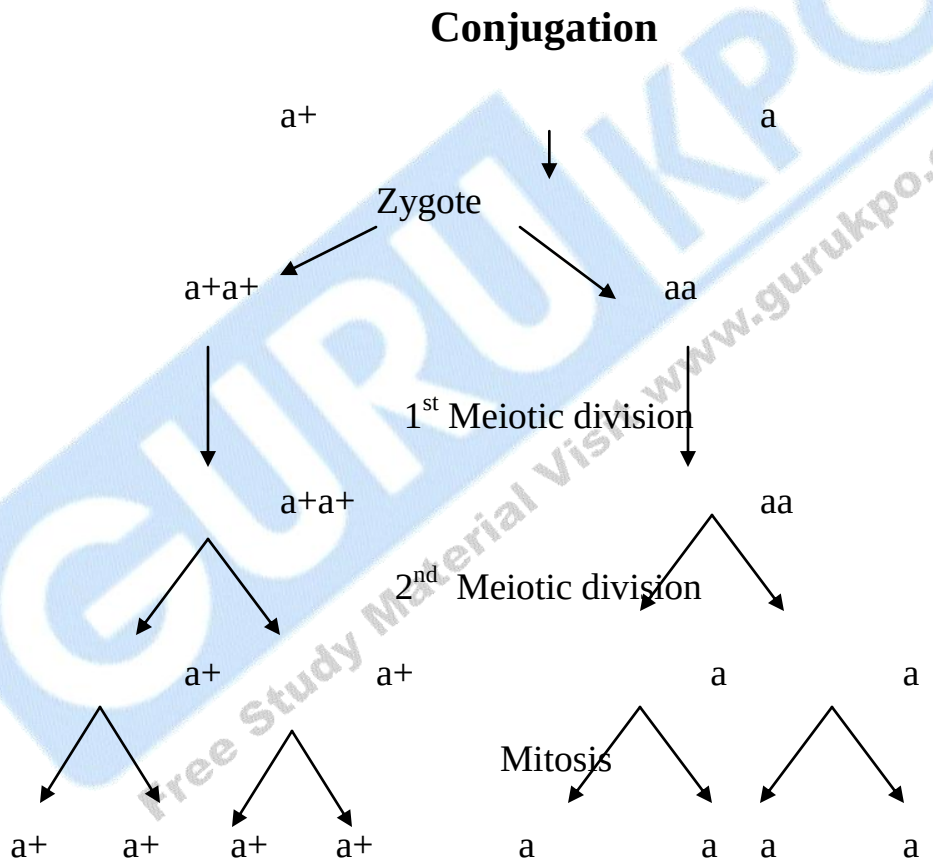


Fig:- Result of first division segregation in a cross between normal (a+) and (a) strain of Neurospora without crossing over between locus a and centromere. b) Second division or segregation between centromere and Gene-a:-

In such type of cross if crossing over taken place leading to paired arrangement of ascospores with a particular gene, it is described as second division crossing over. The arrangement of ascospores in the sequence (2:2:2:2) is as follows:

(I) a+: a+ : a: a : a : a : a+: a+

(II) a : a : a+ : a+: a+: a+: a : a

(III) a+: a+: a: a : a+: a+: a : a

Depends on the orientation of chromosomes at the equatorial plate at second anaphase.

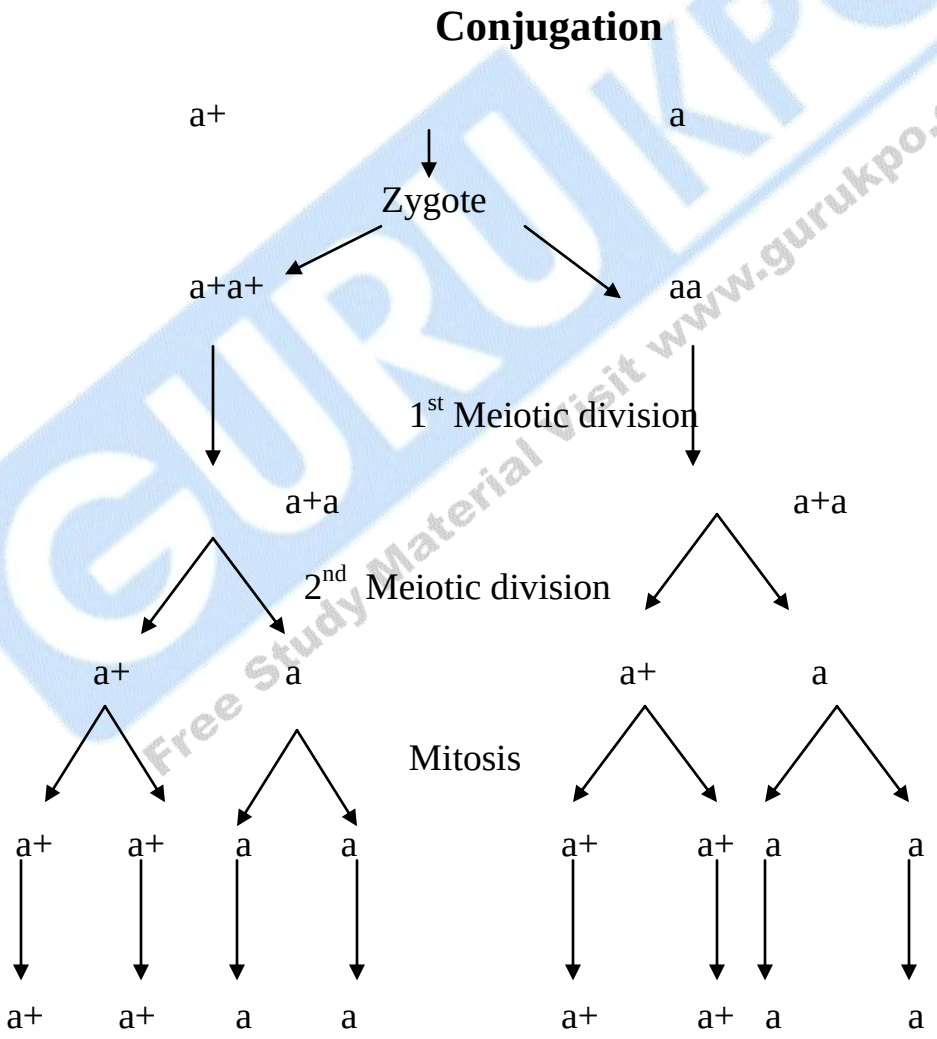


Fig:- Result of second division segregation in a cross between normal (a+) and mutant(a) strains of Neurospora showing crossing over between gene a and centromere resulting in a linear order of ascospore (2:2:2:2).

Detection of linkage between 2 genes:-

If a zygote has two linked genes a and b and heterogenous for these two loci(AB/ab), the recombination frequency for these two genes in a tetrad is calculated as follows:

(I) Parental ditype (PD):-
These are either non-crossover or two strand double crossover, 50% of the meiotic products resemble either of the parental type, i.e. (2AB:2ab). Such tetrads are called parental ditypes.

(II) Non-Parental ditype (PD):-
These are formed as a result of four strand double cross over between the two genes. These present two types of non parental recombination in equal frequency their genic composition is 50% BB (2Ab:2ab)

(III) Tetra type :-
The tetra type asci (i.e. asci with four combinations of genes) are formed either as a result of single crossing over. These asci possess ascospores two parental and two non-parental types (i.e. l AB: lAb: laB: lab).

In all these cases recombination frequency between these two genes in 50%. The PD and NPD are not significantly different. The tetra types tetrads are fewer than PD or NPD. The tetra types can be produced only by the gene that linked. Their ratio of PD or NPD will be governed by the distance between the genes.

Multiple Choice Question

Q1. Genome is
(i) Haploid set of chromosomes
(ii) Diploid set of chromosomes
(iii) Polyploid set of chromosomes
(iv)None of these (i)

Q2. Total no. of genes found in human genome
(i) 20000-25000
(ii) 50000-58000
(iii)100000
(iv)35000-40000 (i)

Q3. Which of the following statements about an organism's genome is false?

(i) The genome contains the genetic information to construct and maintain a living organism.
(ii) Genomes of cellular organism are composed of DNA.
(iii) The genome is able to express its own information without the activity of enzymes and proteins.
(iv) Eukaryotic genomes are composed of both nuclear and mitochondrial DNA.

(iii)

Q4. Somatic cells are those cells that
(i) Contain a haploid set of chromosomes.
(ii) Give rise to the gametes
(iii) Lack mitochondria
(iv) Contain a diploid set of chromosomes and make up majority of human cells. (iv)

Q5.Which of the following protein is used in DNA packaging?
(i) Histone
(ii)Albumin
(iii)Globin
(iv)Nuclease (i)

Q6.Histones are found in
(i) Eukaryotes
(ii) Prokaryotes
(iii)Both
(iv)None of these (i)

Q7. Histones are
(i) Basic proteins
(ii) Acidic proteins
(iii)Neutral proteins
(iv)None of these (i)

Q8.Histones consist of following amino acids
(i) Arginine and Lysine
(ii) Glutamine and Aspartate
(iii)Cystiene and Methionine
(iv)All of these (i)

Q9. Nucleosomes are composed of
(i) Core DNA and Histones
(ii) Linker DNA and Histones
(iii)Linker DNA, Core DNA and Histones
(iv)None of these (i)

Q10. Core of Nucleosomes is consist of Histone proteins
(i) H2A,H2B,H3 and H4
(ii)H1, H2A,H2B and H3
(iii) H2A,H2B and H3
(iv) H1,H2A,H2B,H3 and H4 (i)

Q11. Which of the following histone protein attach with linker DNA
(i)H1
(ii)H2A and H2B
(iii)H3
(iv) H4 (i)

Q12. Which of the following is a common feature of eukaryotic genome?
(i)Presence of Introns
(ii) Presence of Transposons
(iii)Even distribution of genes
(iv)All of these (i)

Q13. Gene density in eukaryotes is
(i) Higher than prokaryotes
(ii) Lower than prokaryotes
(iii)Equal to prokaryotes
(iv) None of these (ii)

Q14. Size of Human Genome is
(i) 125 Mb
(ii) 3300 Mb
(iii) 3200 Mb
(iv)120,000Mb (iii)

Q15. Truncated genes are
(i)Genes that lack a segment or stretch from one end of complete gene
(ii) Genes that lack a segment or stretch from between of a complete gene
(iii) Genes that are short isolated regions of from within a gene
(iv) All of these. (i)

Q16. Following are the features of satellite DNA

(i) Consist of repetitive DNA
(ii) Form a separate band during centrifugation with density different from bulk DNA.
(iii) Consist of long Tandem repeats
(iv)All of these (iv)

Q17. Linkage increases when
(i)Distance between genes increases
(ii) Distance between genes decreases
(iii) Distance between genes is equal
(iv) Is independent of distance (ii)

Q18. Crossing over increases when
(i)Distance between genes increases
(ii) Distance between genes decreases
(iii) Distance between genes is equal
(iv) Is independent of distance (i)

Q19. For the analysis of linkage *Neurospora* is a good model as
(i) It is easy to grow and has a haploid life cycle
(ii) Ordered arrangement of the products of meiosis in *Neurospora* ascospores
(iii) Its entire genome of seven chromosomes has been sequenced
(iv) All of these (iv)

Q20. Genes are said to be linked when
(i) Their loci are located to each other
(ii) They inherit together for two or more generations
(iii)Their chromatids do not separate during chromosomal cross over
(iv)All of these (iv)

Section B

Genetic Mapping & Multigenes Families

Q.1 What is genetic mapping? Describe pedigree analysis in human genetic mapping?

Ans Genetic Mapping:-

Genetic Mapping is a technique based on the use of genetic technique to construct maps showing the position of genes and other sequence features on a genome.

Important features of genetic mapping:-

- Genetic technique include cross breeding experiments or in case of humans , pedigree or family history are examined for construction of GM.
- A genetic map is consist of distinctive features shown according to their relative positions. For construction of genetic map markers are used the first marker were the genes.
- These genes were used as markers for fruit fly.
- A gene to be used as a marker for genetic analysis must exist in at least two forms or alleles each specifying a different form e.g. tall or short stems in the pea plants originally studied by Gregor Mondel.
- For the construction of GM for larger genomes like those of vertebrates and flowering plants, gene markers are by no means ideal as the map based on gene markers is not very detailed.
- DNA markers are then used for GM also need to have at least two alleles to be useful. There are three type of DNA markers used for the purpose RFLP's, SSLP's, SNP's.
- Linkage Analysis:-

Most of the techniques used for GM are based on linkage analysis.

Pedigree analysis in human genetic mapping:-

With human presentation of the genotype of parent and set up crosses designed for mapping purposes is impossible. So for gene mapping instead, data for the calculation of recombination frequencies have to be obtained by examining the genotype of the members of successive generation of existing families. This means that only limited data are available, and their interpretation is often difficult as human marriages rarely result in a convenient cross and often the genotype of one or more

family members are unobtainable because those individuals are dead or unwilling to cooperate.

Taking an e.g. of genetic disease found in a family consist of two parents and 6 children, where genetic diseases act as a gene marker.

Genetic diseases are frequently used as a gene marker in human the case state being one alleles and the healthy state being a second allele.

Example of Pedigree analysis

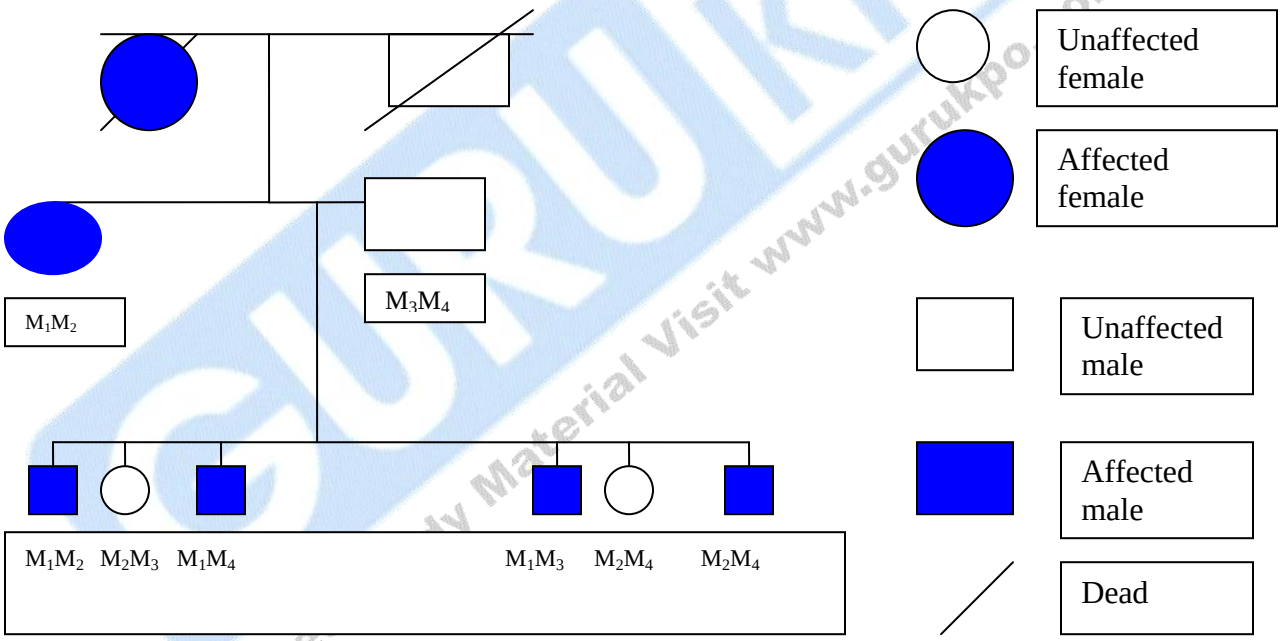


Figure (a): **The pedigree shown inheritance of a genetic disease in a family of two living parents and six children, with information about the maternal grandfather available from family record.**

The disease allele (closed symbols) is dominant over the healthy allele (open symbol). The objective is to determine the degree of linkage between the diseased gene and microsatellites M by typing the allele for (M1, M2, etc.) in living membrane of the family.

Figure (b): Possible Interpretation of the pedigree

Mother's chromosomes

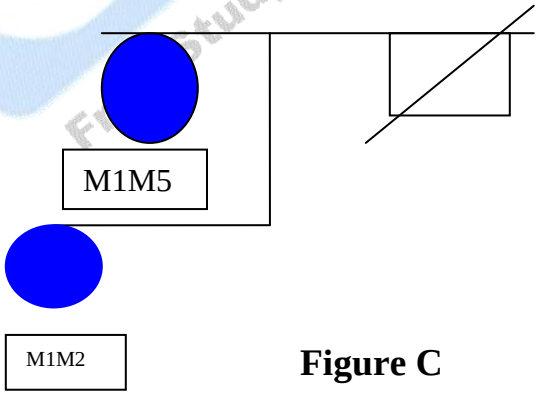
Hypothesis 1	Hypothesis 2
Disease M 1	Healthy M1
Healthy M2	Disease M2

Child 1	Disease M 1	Parental	Recombinant
Child 2	Healthy M2	Parental	Recombinant
Child 3	Disease M 1	Parental	Recombinant
Child 4	Disease M 1	Parental	Recombinant
Child 5	Healthy M2	Parental	Recombinant
Child 6	Disease M 2	Recombinant	Parental

- Recombination – 1/6 = 16.7% 5/6 = 83.3%

The pedigree can be interpreted in two different ways Hypothesis 1 gives a low recombination frequency & indicates that the disease gene is tightly linked to microsatellites.
Hypothesis 2 suggests that the disease gene and microsatellites are much less closely linked.

Fig c: Resurrection of the maternal grandmother disease allele must be linked to M1, Hyp. 1 is correct.



Disease allele must be Linked to M_1, Hyp.I is Correct.

Figure C

In the above example of pedigree it has been shown that the mother is affected by the disease as per her four children. From family accounts it is clear that the maternal grandmother also suffered from this disease but both maternal grandmother and her husband (grandfather) are dead. They have been shown by slashes in the diagram and no further information can be obtained from either of two their genotypes. The

disease gene is located on the same chromosomes as a microsatellite c/d 'M' four alleles of C-M1,M3,M3 and M4 are presence vig. Pedigree analysis our motive is to map position of debased gene relative to microstates to establish a recombination frequency b/w the disease gene and microsatellite M, it is necessary to determine the no of children that are recombinant. Looking at the 6 children that are recombinant. Looking at the 6 children genotype 3 & 4 children have a disease allele and microsatellite allele M1. No. 2 &5 have healthy allele and M2. Now two hypothesis can be made (I) – the copies of the relevant homologous chromosomes in the mother have genotypes Disease –M1 and healthy M2, therefore children 1, 2, 3, 4& 5 have parental genotypes while 6 child is a recombinant. This hypothesis suggests that the diseased gene and the microsatellite are relatively closely linked and that crossover between them occur in frequently.

The alternative hypothesis is that the mother's chromosomes have the genotypes healthy M1 and Disease M2; this would mean that 1-5 are recombinants, and child has the parental genotype. This shows that the gene and microsatellites are relatively far apart on the chromosome.

To determine which hypothesis is correct is not an easy task. The most satisfying solution by pedigree would be to know the hypothesis of the grandmother. Assuming that grandmother is alive, her genotype for microsatellite M turn out to be M1 M5. This shows that the chromosomes inherited by the mother has the genotype disease-M1. Now it can be concluded that Hyp-1, is correct and that only child 6 is a recombinant.

Lod Score:-
Imperfect pedigree are analyzed statistically using a measure c/d lod score. It stands for logarithms of odds that the genes are linked, is used primarilly to determine if the two markers being studied lie on the same chromosome.

Q.2 Describe various tools and techniques used to construct a genetic map?

Ans Genetic Mapping is a technique for the construction of genetic map using genetic techniques. The various techniques used for genetic map construction include:

1. Markers:-

A marker for genetic mapping must be polymorphic i.e., it must exists in different forms so that chromosome carrying the mutant gene can be distinguished from the chromosome with normal gene by the form of marker it also carries. This polymorphin in the marker can be detected at three levels:-

a. Phenotype or morphological markers.
b. Difference in the proteins or biochemical

c. Difference in the nucleotide sequence of DNA or molecular but out of the three types most satisfying and used markers are DNA or molecular markers.

Molecular Markers:-

They are DNA sequence that is readily detected and whose inheritance can easily be monitored. The marker must be polymorphic i.e. must exist in different forms so that chromosome carrying the mutant gene can be distinguished from the chromosome with the normal gene by a marker it also carriers.

Mapped feature that are not genes are called DNA markers. There are three types of DNA sequenced feature that satisfy this requirement:

a. RFLP's- Restriction fragment length polymorphism
b. SSLP's- Simple sequence length polymorphism
c. SNP's- Single Nucleotic Poymorphism

a. RFLP's- Restriction fragment length polymorphism:-

The treatment of a DNA molecule with a restriction enzymes should always produce the same set of fragments. This is not the case with genomic DNA as some restriction sites are polymorphic, existing as two alleles, with one alleles displaying correct sequence for the restriction site and therefore being cut when the DNA is treated with the enzyme and second allele longer recognized. The result of the seq. alteration is that the two adjacent restriction fragments remain linked together after treatment with the enzyme, leading to a length polymorphism. This is an RFLP's and its position on a genome map can be worked out by following the inheritance of its alleles, just as is done when genes are used as markers. There are about 10^5 RFLP's in a mammalian genome.

Methods for scoring an RFLP:

a. RFLP's can scored by southern hybridization. The DNA is digested with the appropriate restriction enzyme and separated in an agarose gel. The smear of restriction fragments is transferred to a nylon membrane and probed with a piece of DNA that spans the polymorphic restriction site. If the site is absent then a single restriction fragment is detected; if the site is present then two fragment are detected.

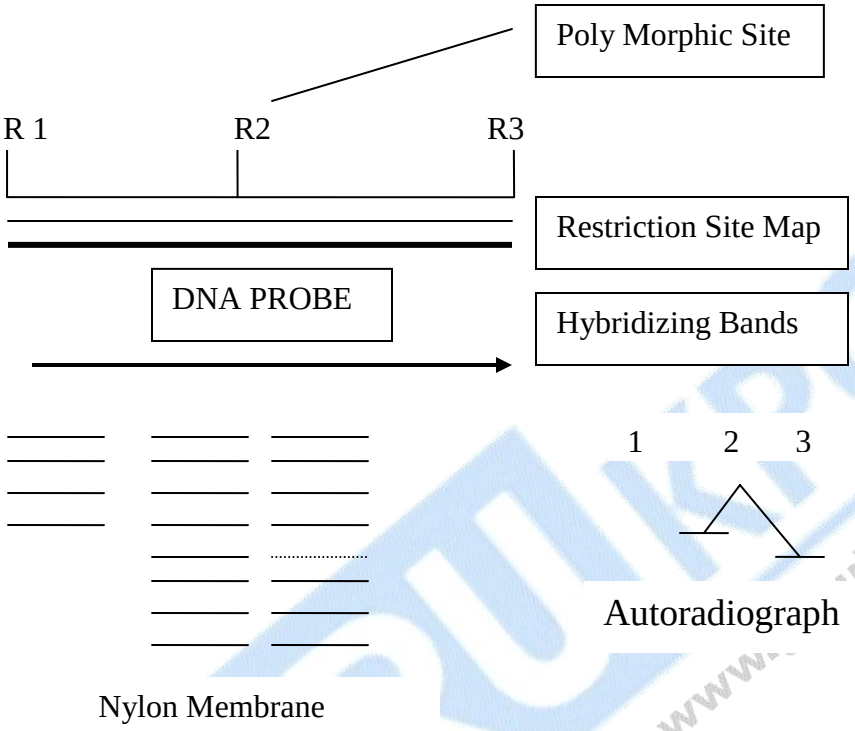


Figure:- Southern Hybridization

b. The RFLP can also be typed by PCR, using primers that anneal either side of the polymorphic restriction site. After the PCR, the products are treated with the appropriate restriction enzymes and then analyzed by agarose gel electrophoresis. If the site is absent then one band is seen on the agarose gel; if the site is present then two bands are seen

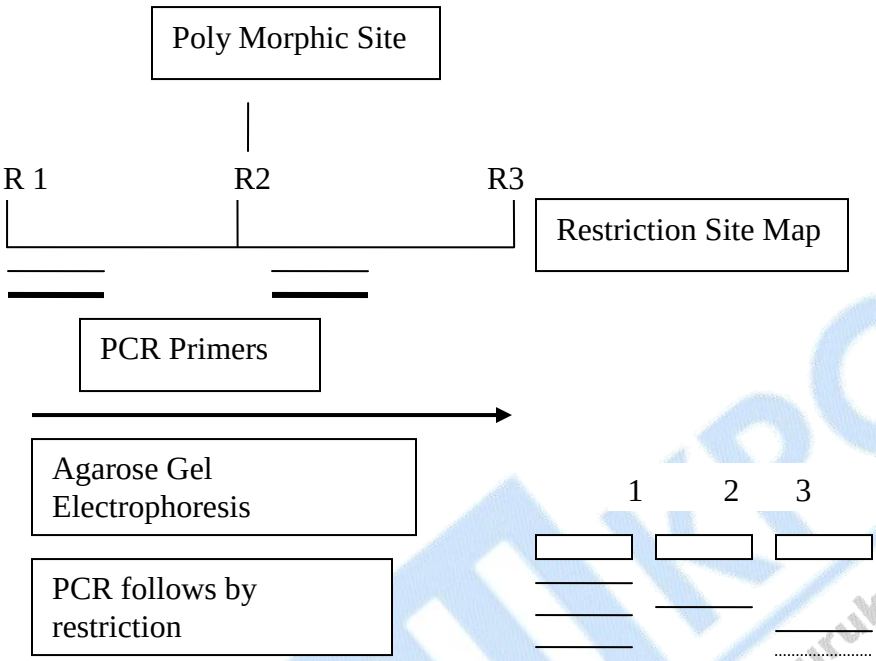


Figure:- PCR

b. SSLP's- Simple sequence length polymorphins:-

- These are array of repeat sequence that display length variation, different alleles containing different number of repeat units.
- Unlike RFLP's SSLP's can be multiallelic as each SSLP can have a number of different length variants.

Types of SSLP's **:-**

Minisatellites: - also called VNTRs (Variable number of tandem repeats), in which the repeat unit is up to 25 bp in length.

Microsatellites:- also called STRs (Simple Tandem Repeats) where repeats are shorter, usually 13 bp or less.

For DNA markers microsatellites are used more compared to minisatellitte. As Minisatellites are not spread evenly around the genome but tends to be found more frequently in the telomeric region of chromosomes while microsatellites are more conveniently spaced throughout the genome. A second reason for more use of microsatellites is that it is fastest way to type a length polymorphism by PCR, but PCR typing is much quicker and more accurate with sequence less than 300bp in length Most Minisatellittes are longer than this because the repeat units are relatively

large and there tend to be many of them in a single array, so PCR products several kilo bases in length are needed to type them. Microsatellites used as DNA markers typically consist of 10-30 copies of a repeat that is no longer than 6 bp in length and so are much more amenable to analysis by PCR.
In human genome there are $5x10^5$ microsatellites with repeat units of 6bp. for analysis of genome STRs are can be visualized by AGE (Agarose Gel Electrophoresis) but being a cumbersome process instead STRs are typed by capillary electrophoresis in a polyacrylamide gel.

Most capillary electrophoresis make use of fluorescent label is attached to one or both of the primers before PCR's carried out. After PCR, the product is located into the capillary system and run past a fluorescent detector. A computer attached to the detector correlates the time of passage of the PCR product with equivalent data for a set of a size markers, and hence identifies the precise length of the product.

c. SNP's- Single Nucleotide Poymorphism:-

SNP's are positions in a genome where some individuals have one nucleotic (e.g., a'G') and others have different nucleotide (e.g., a'C').
In every genome there are vast number of SNP's (around 4 million in human genome), some of which also give rise to RFLP's. but mostly not as the sequence in which they lie is not recognized by any restriction enzyme.
SNP's originates as a result of point mutation occurring in a genome converting one nucleotide into another. They are generally biallelic.
In most cuk, at least one SNP for every 10kb of DNA is found. SNP's therefore enable very detailed genome maps to be constructed.

Role of genome analysis:-

The importance that SNP's have acquired in genome research has stimulated the development of rapid methods for their typing. Several of these methods are based on oligonucleotide hybridization analysis. Oligonucleotide hybridization can discriminate between the two alleles of an SNP. Various screening strategies based on oligonucleotide hybridization are:

(I) DNA chip technology:-

It makes use of a wafer of glass or silicon 2.0 cm^2 or less in area, carrying many different oligonucleotides in a high-density array. The DNA to be tested is labeled with a fluorescent marker and pipetted onto the surface of the chip. Hybridization is detected by examining the chip with a fluorescent microscope.

(II) Solution Hybridization technique :-

Technique is carried out in microtiter tray, where each well contain a different oligonucleotide, This detection system can discriminate between unhybridized single stranded DNA and double strands product that results when oligonucleotide hybridizes to the test DNA.
Other typing method makes use of an oligonucleotide whose mismatch with the SNP occurs at its extreme 5' or 3' end. Under the appropriate condition an oligonucleotide of this type will hybridized to the mismatched template DNA with a short. Non-base paired "tail". This feature is utilized in two different ways.

(a) Oligonucleotide Ligation Assay (OLA):-
It makes use of two oligonucleotide that anneal adjacent to one another, with the 3' end of one of this oligonucleotide positioned exactly at the SNP's. This oligonucleotide will form a complete base paired structure if one version of the SNP's is present in the template DNA, and when this occurs the oligonucleotide can be legated to its partner. The alleles is therefore typed by examining and determining if the ligation product is synthesized usually by running the post reaction mixture in a capillary electrophoresis system.

(b) Amplification refractory mutation system or ARM's test:-

It is based on the same principle as OLA's but in this method the test oligonucleotide is one of a pair of PCR primers. If the test primer anneals to the SNP then it can be extended by taq polymerase and the PCR can take place, but if it does not anneal, because the alternative version of the SNPs is present, then no PCR product is generated.

Linkage Analysis :-

Linkage Analysis is the basis of genetic mapping technique used for the construction of genetic map are all based on genetic linkage.

Linkage :-

The genes located on the same chromosome cannot assort independently, rather these tend to be inherited together. This phenomenon of inheritance of genes together and to retain their parental combination even in the offsprings is called linkage.

- Linkage analysis usually makes correct deduction about gene order, and distance estimates are sufficiently accurate to generate genetic maps that are of value as frame work for genome sequencing projects.

There are three situation in which linkage analysis can be carried out:

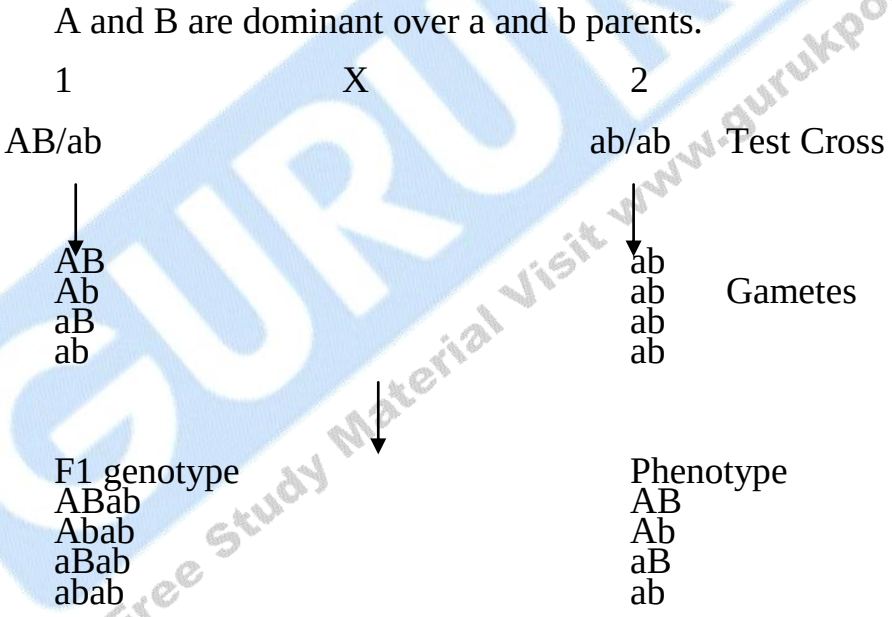
- Linkage analysis with species s/a fruit flies and mice, with which we can carryout planned breeding experiments.

- Linkage analysis with human with whom we cannot carry out planned experiment but instead can make use of family pedigree.

- Linkage analysis with bacteria, which do not undergo meiosis.

Now taking first condition of linkage analysis in fruit fly.:-

The method is based on different the analysis of the progress of experiment crosses set up between parents of genotype and in theory applicable to all eukaryotes.
The std. procedure include the use of test cross (Fig) where a test cross is made to map the two genes, gene A (alleles A & a) and gene BC (alleles B & b) both on the chromosome 2 of the fruit fly.



Each phenotype is the same as the genotype of the parent 1 gametes.

Fig:- A test cross between alleles displaying dominance and recessiveness.

A and B are markers with alleles A, a, B and b. the resulting progeny are scored by examining their examining their phenotypes. Because the double homozygous parent (parent 2) has both recessive alleles- a and b it effectively makes no contribution to the phenotype of the progeny. The phenotype of each individual in the F1 generation is therefore the same as the genotype of the gametes from parent 1 that gave rise to that individual.

The critical feature of a test cross is the genotype of the 2 parents:-

- One parent is double heterozygote. This means that all four alleles are present in this parent. Its genotype is AB/ab. This mutation indicates that one pair of the homologous chromosomes has alleles A and B, and the other has a and b. Double heterozygous can be obtained by crossing two pure breeding strain. E.g. AB/AB X ab/ab.
- The second parent is a pure breeding double homozygote. In this parent both homologous copies of chromosome 2 are the same. In the above fig. both have alleles a and b and the genotype of the parent is ab/ab.
- The double heterozygote has the same genotype as the cell whose meiosis is has to be followed. The main aim therefore is to infer the genotype of the gametes produced by this parent and to calculate the fraction that are recombinants. All the gametes produced by their 2^{nd} parent (the double homozygote) will have genotype ab regardless alleles a and b are both recessive, so meiosis in this parent in effect, invisible when the phenotype of the progeny are examined.
- The phenotype of the diploid progeny can be converted into genotype of the gametes from the double heterozygote parent. The test cross therefore enables us to make a direct examination of a single meiosis and hence to calculate a recombination frequency and map distance for the genes being studied.
- The power of this type of linkage analysis is enhanced if more than two markers are followed in a single cross. This generates recombination frequency more faster and also enables relative order of markers on a chromosome to be determined by simple inspection of data.

Q.3 Write short notes on:-

1. C-Value Paradox
2. Complexity of Genomes
3. Multigene families in human genome

Ans 1. C-Valve Paradox and complexity of genomes:-

The total amount of DNA in the genome is a characteristic of each living species called its C-Value.

There are enormous variations in the range of C-Values, from as little as $<10^6$ bp for a mycoplasma to as much as 10^{11} bp for some plant and amphibians.

C-Valve paradox form correlation between complexity of an organism and the size of its genome there is an increase in the minimum genome size found in each group as the complexity increases. Along with the increase in absolute amount of DNA in the higher eukaryotes some wide variations in the genome. Size within some phyla is also observed (FigA) it is necessary to increase the genome size in order to make insects, birds or

amphibians and mammals. But this point donot holds good relation between genome size and morphological complexity of the organism.

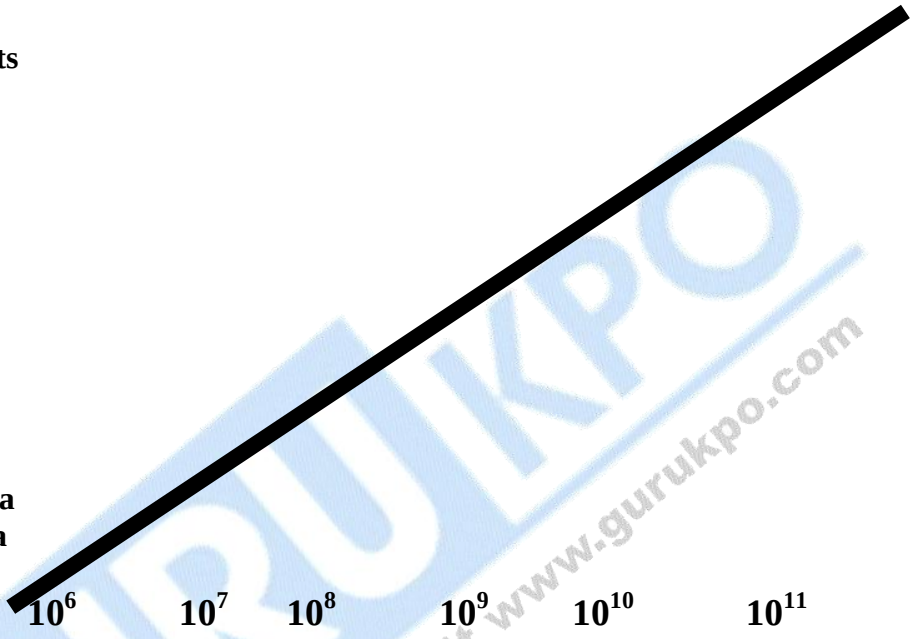


Fig:- DNA content of the haploid genome is related to the morphological Complexity of lower eukaryotes but varies extensively among the higher eukaryotes. The range of DNA values within a phylum is indicated by the shaded area.

Plotting the min amount of DNA required for a member of each phylum suggests that an increase in the genome size is required to make more complex prok and lower cuk.

The smallest genome get identified for a living cell actually belongs to a eukaryotic, the unicellular algae *Pyrenomas salina*, at $6.6x10^5$ bp. these organisms may not be true eukaryotes, but could be an intermediate stage of evolution, representing a primitive presence of a nucleus and chloroplast excluding these algae, mycoplasma are the smallest types of organism k/n , and also may be constructed from very small genomes. Thus organism may have genome sizes only ~3x size of the large bacteriophase.(T_4 is $1.7x10^6$ bp) Bacteria start at ~$2x10^6$ bp unicellular eukaryotes (whose life-styles may resemble the prokaryotic) get by with genomes that are also small although larger than those of the bacteria . Being eukaryotic per does not imply a vast increase in genome size, a yeast may have a genome size of ~$1.3*10^7$ bp only about twice the size of the largest bacterial genomes.

A further two fold increase in genome size in adequate to support the slime mold discoideum, able to live in either unicellular or multicellular modes.

Another increase in complexity is necessary to produce the first fully multicellular organism; the nematode worm C-elegans has a DNA content of $8*10^7$ bp.

Moving further along the evolutionary tree, the relationship between complexity of the organism and content of DNA becomes obscure, although it is necessary to increase the genome. Size in order make insects, birds or amphibians and mammals.

There occurs a steady increase in genome size with complexity of some of the most commonly analyze organism, although there are some real puzzles. The toad * laevals have same genetic complexity as man, as one might suppose naively from their common genome sizes.

In some phyla, the spread of genome sizes in narrow. Birds, reptiles, and mammals all show little variation within the phylum. With a range of genome sizes in each case about two fold. But in other cases, most notably insects, amphibians and plants there is a wide range of valves, often more than tenfold. One wonder whether the common housefly (*Musca domestica*) with a genome size of $8.6x10^8$ really has a corresponding increase in complexity over common fruit fly. (*D. malenogaster*) with its 6 time smaller genome of $1.4x10^8$.

The C-Value paradox takes its name from our inability to account for the content of the genome in terms of the k/n function. It expresses the existence of two features:-

(I) There is an excess of DNA compared with the amount that could be expected to code for protein.
(II) There are large variations in C-Values between certain species which do not show that much variation in complexity.

The C-value paradox refers to the lack of correlation between genome size and genetic complexity.
In amphibians, the smallest genomes are just below $10^{9\ bp.}$ While the largest are almost 10^{11} bp.

Q.4 Multigene families in human genome.

Or

What are multigenes families? Describe giving suitable examples?

Ans Multigene Families:-

Groups of genes of identical or similar sequence commonly found in many genomes and are called multigene families.

Every eukaryote that has been studied (as well as all but the simplest bacteria) has multiple copies of the genes for the rRNA. This is illustrated by the human genome, which contains ~2000 genes for 5srRNA. These all are located in a single cluster on

chromosome. There are also about 280 copies of a repeat unit containing the 28s, 5.8s and 18s rRNA genes, grouped into five clusters of 50-70 repeats one on each of chromosomes 13,14,15,21&22. Ribosomal RNA's are components of the protein synthesizing particles called ribosome, and it is presumed that their genes are present in multiple copies because there is a heavy demand for rRNA synthesis during cell division, when several tens of thousands of new ribosome must be assembled. The rRNA genes are example of simple or classical multigenes families. In which all the members have identical or nearly identical sequence

Origin:- these families are believed to have arisen by gene duplication with the sequence of the individual members kept identical by an evolutionary process that has yet not been fully described.

Other multigene families, more common in higher eukaryotes that in lower eukaryotes are called common because the individual members although similar in sequence are sufficiently different for the gene products to have distinctive properties. One of the best examples of this type of multigene family are the mammalian globin genes.

The globin are the blood protein that combine to make hemoglobin, each molecule of hemoglobin being made up of two α types and two β type globins. In human the α types globins are coded by a small multigene family an chromosome 16 and the β type globins by a second family on chromosomes . These genes were among the first to be sequenced.
The sequence data showed that the genes in each family are similar to one another, but by no means identical . The nucleotide sequences of the two most different genes in the β type cluster coding for the β and α globin display only 79.1% identity.

The members of the globin gene families so different from one another reason being that the genes are expressed at different stages in human development for e.g., in the β type cluster is expressed in the early embryos G_r and A_r (whose protein products differ by just one amino acid) in the fetus and α and β in the adult. The different biochemical properties of the resulting globin proteins are through to reflect slight changes in the physiological role that H6 plays during the course of human development.

In some multigene families, the individual members are clustered as with the globin genes, but in others the genes are dispersed around the genome. On example of such genes for aldolase, an enzyme involved in energy generation which are located on chromosomes 3, 9, 10, 16 and 17. Even though dispersed, the member of multigene families have sequence similarities that point to a common evolutionary origin. When these sequence comparisons are made it is sometimes possible to see relationships not only within a single gene family but also between different families. All of the genes in α and β globin families for example have sequence similarity and are thought to have evolved from a single ancestral globin gene. Thus these two multi gene families

can be referred to as gene super family comprising a single globin gene super family. and from the similarities between the individual genes the duplication events can be list out that has given rise to the series of genes.

Q.5 How are restriction enzymes used to generate both genetic and physical map of a genome?

Ans Introduction:-

Genetic mapping using RFLP's as DNA markers can locate the position of polymorphic restriction sites within a genome, but very few of the restriction sites in a genome are polymorphic, so many sites are not mapped by this technique. By the help of restriction enzymes the marker density on a genome map can be increased to located the position of some of the non polymorphic restriction sites.

Restriction mapping:-
Mapping technique which locates the relative position on a DNA molecule of the recognition sequences for restriction endonucleases.

Methodology:-
The simplest way to construct a restriction map is to compare the fragment sizes produced when a DNA molecule is digested with two different restriction enzymes that recognize different target sequences.

E.g. of restriction enzymes are EcoR1and Bam HI.
The DNA molecules is digested with just one of the enzymes and the sizes of the resulting fragments measured by agarose gel electrophoresis.

In the next step the molecule is digested with the second enzymes and the resulting fragment s are again sized in an agarose gel. Till this stage the numbers of restriction sites for each enzyme are worked out but determination of their relative position is still not done.

For additional information DNA molecules are cut with both enzymes together. This double restriction enables there of the sites to be mapped.

In case of EcoR1 fragment there are two Bam H1 site and there are two alternative possibilities for the map location of the outer one of these. To overcome this problem the original DNA molecule is treated again with Bam HI but this time the digestion is not completed instead reaction in incubated before the completion for a short duration or using a suboptimal incubation temperature. This is called partial restriction. it gives a complex set of products , the complete restriction products now being supplemented with partially restricted fragments that still contain one or more uncut Bam H1 sites.

A partial restriction usually gives the information needed to complete a map, but if there are many restricting sites then this type of analysis becomes unweildy, because there are so many different fragments to consider to simplify the analysis. An alternative strategy is used where markers are attached to the ends of the DNA molecules before digestion (Fig).

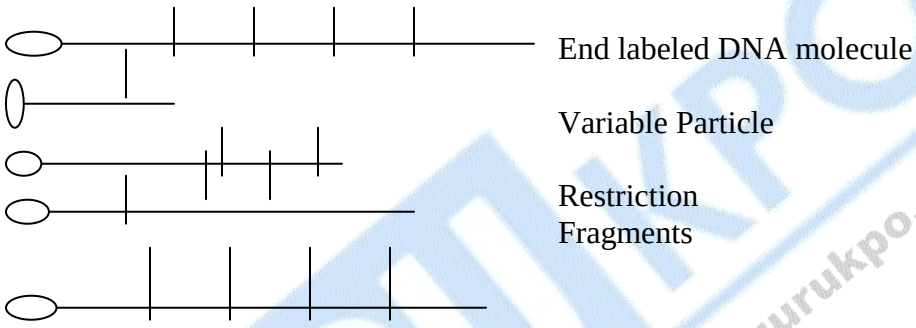


Fig:- Simplifying the analysis of a particle restriction by attaching markers to the ends of the DNA molecule before.

One end of an end –labeled DNA molecule is shown after partial restriction only those products that include an end fragment are detected. This greatly simplifies the analysis, enabling the position of restriction sites to be determined directly from the lengths of labeled products.

Application of restriction mapping:-

Restriction maps are easy to generate if there are relatively few cut sites increases for the enzymes being used.
It is more applicable to small rather than larger molecules, with the upper limit for the technique depending on the frequency of the restriction sites in the molecule being mapped. In practice a DNA molecule is less than 50kb in length is usually possible to construct. But for constructing larger maps there are enzymes called "rare cutters having infrequent cut sites are used. These falls in two categories:-

a. A few restriction enzymes cut at 7 or 8 nucleotide recognition sequences. E.g. Sap I (S'-GCTCTTC-3') and SQFI (S'-GCGATCGC-3'). The enzyme with 7 nucleotide recognition sequence would be expected, on aug to cut a DNA molecule with a GC content of 50% once every 4^7= 16384 bp. The enzyme with eight nucleotide recognition sequences should cut at once 4^8 = 65,336bp.
b. Enzyme whose recognition sequences contain motifs that are rare in the target DNA.

e.g. SmaI (5'-CCCGGC-3'), which cuts human DNA once every 78 kb on average and BSSHII (5'-GCGCGC-3'), which cuts once every 390 kb. Note that not I , a cuts with an eight-nucleotide recognition sequences, also targets 5'-CG-3' sequences (recognition sequences 5'-GCGGCCGC-3') and cuts human DNA very rarely-approximately once every 10Mb.

Restriction fragments are equally useful after bacteria eukaryotic genomic DNA has been cloned, of the cloned fragments are less than 50kb in length, because a detailed restriction map can then be built up as a preliminary to sequencing the cloned region. This is an important application of restriction mapping in projects sequencing larger genome.

Direct examination of DNA molecules for restriction sites:-

For mapping restriction sites in DNA molecules technique other than electrophoresis are also used like optical mapping. In this technique, restriction sites are directly located by looking at the cut DNA molecules with a microscope. The DNA must first be attached to a glass slide in such a way that the individual molecules become stretched out, rather than clumped together in a mass.
There are various ways by which this can be done including gel stretching and molecular combing. To prepare gel stretched DNA fibers, chromosomal DNA is suspended in molten agarose and placed on a microscope slide. As the gel cools and solidifies, the DNA molecules become extended. In molecular combing, the DNA fibres are prepared by dipping a silicone coated cover slip into a solution of DNA, leaving it for 5 min. (during this time the DNA molecules attach to the cover slip by their ends.) then removing the slip at a constant speed of 0.3mm $s^{-1.}$

Fig:-

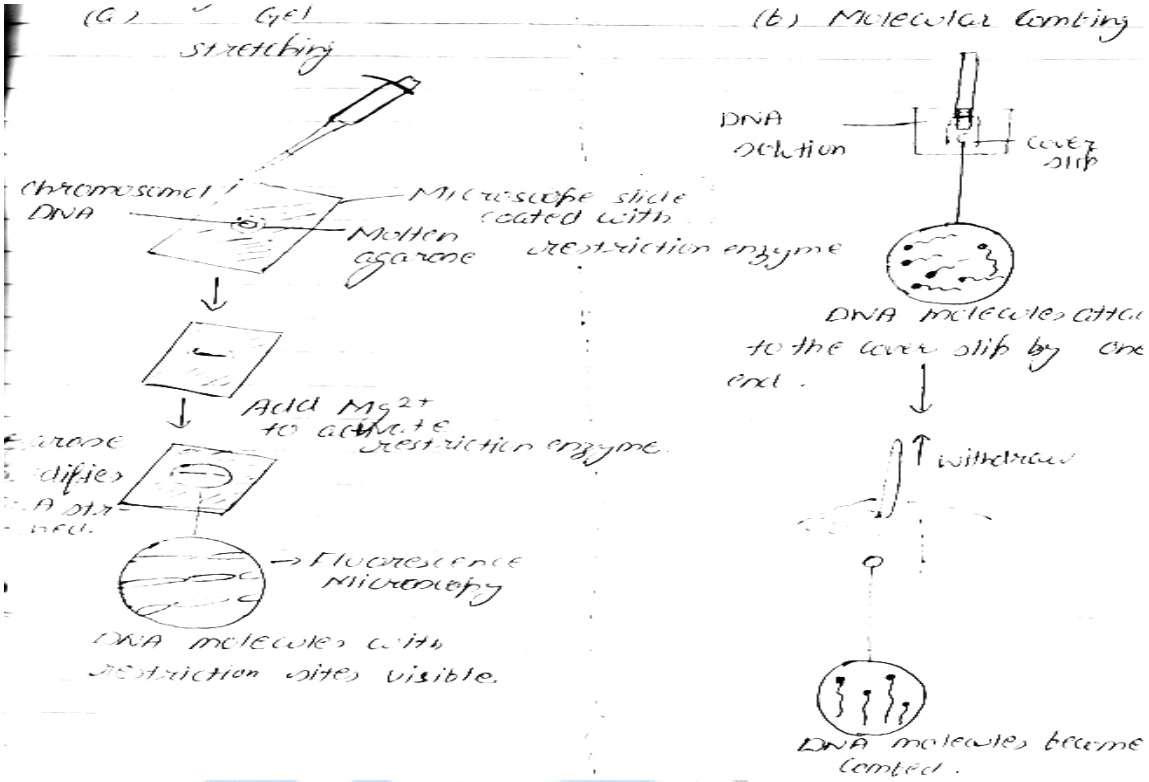


Fig a. Gel Stretching –

To carry out gel stretching molter agarose containing chromosomal DNA molecules is pipetted onto a microscope slide coated with a restriction enzyme. As the gel solidifies, the DNA molecules become stretched. Addition of magnesium chloride activate the restriction enzyme, which cuts the DNA molecules. As the molecules gradually coil up, the gap representing the cut sites become visible.

b. Molecular combing :-

A cover slip is dipped into a solution of DNA. The DNA molecules attach to the cover slip by their ends, and the slip is withdrawn from the solution at a rate of 0.3mm s^{-1} which produce a comb of parallel molecules.

- The force required to pull the molecule through meniscus causes them to line up. Once in the air, the surface of the cover dries, retaining the DNA molecules as an array of parallel fibers. After stretching or combing the immobilized DNA molecules are treated with a restriction enzyme and then visualized by adding a fluorescent dye. s/a DAPI (4,6-diamino-2-phenylindole) dihydrochlroide, which stain the DNA so that the fibers can be seen when the slide is examined with a high powered florescent microscope. The restriction site in the extended molecule gradually become gaps as the degree of fiber extension is reduced by the natural springiness of the DNA, enabling the relative position of the cuts to be recorded.

Q.6 Which is more useful a genetic or a physical map?

Ans Genetic mapping is based on the use of genetic techniques to construct maps showing the positions of genes and other sequence features on a genome. On the contrary physical mapping uses molecular biology techniques to examining DNA molecules directly in order to construct maps showing positions of sequence features, including genes.

Physical mapping compared to genetic mapping is more useful as the map generated by genetic techniques is rarely different for directing the sequencing phase of a genome projects. There are two reasons for this:

a. The resolution of a genetic map depends on the number of crosses that have been scored. This is not a major problem for microorganism because these can be obtained in huge number, enabling many crossovers to be studied, resulting in a highly detailed genetic map in which the markers are just a few kilo bases apart. For e.g. when the E.Coli genome sequencing project began in 1990, the latest genetic map for this organism comprised over 1400 markers, an avg. of 1 per 10 kb.

In case of humans and most other eukaryotes is that it is not possible to obtain large numbers of progeny, so relatively few meiosis can be studied and the resolving power of linkage analysis is restricted. This means that genes that are several tens of kilo bases apart may appear at the same position on the genetic map.

b. Genetic maps have limited accuracy:-

According to Sturtevant's assumption crossovers occur at random along chromosome. This assumption is partly correct because the presence of recombination hotspots means that crossovers are more likely to occur at some points rather than at others.

The accuracy of genetic map was illustrated when the complete sequence for *Saccharomyces cerevisiae* chromosomes III was published, enabling the first direct comparison to be mode between a genetic map and actual position of markers as shown by DNA sequencing. There were considerable discrepancies; even to the extent that one pair of genes had been ordered incorrectly by genetic analysis.

Saccharimyces cervisiae is one of the 2 eukaryotes whose genomes have been subjected to intensive genetic mapping. If the yeast genetic map is inaccurate then how precise are the genetic maps of organisms subjected to less detailed analysis.

These two limitations of genetic mapping mean that for most eukaryotes genetic map must be checked and supplemented by alternative mapping procedures before large scale DNA sequencing begins. To overcome these physical mapping techniques have been developed like restriction mapping, Fish is an sequence tagged mapping.

Q.7 Give a note on repetitive DNA content found in eukaryotes genome?

Ans Repetitive DNA content of eukaryotes nuclear genome:-

The human genome sequence revealed that approximately 62% of the human genome comprises intergenic regions the parts of the genome that lie between genes and which have no k/n function. These sequences used to be called junk DNA.

Tandemly repeated DNA sequences are thought to have arisen by expansion of a progenitor sequence, either by replication slippage as in case of microsatellites, or by DNA recombination processes. Both of these events are likely to result in a series of linked repeat, rather than individual repeat units scattered around the genome interspersed repeats must therefore have arisen by a different mechanism, one that can result in a copy of a repeat unit offspring in the genome at a position distant from the location of the original sequences. The most frequent way in which this occurs is by transposition, and most interspersed repeats have inherent transpositional activity. Transposition is also a feature of some viral genome which is able to insert into genome of the infected cell and then move from place to place within that genome.

Types of repetitive genome:-

In most organisms the bulk of the intergenic DNA is make up of repeated sequences of one type of another repetitive DNA can be divided into two categories:-

Genome wide/ interspersed repeats:-

Whose individual repeats units are distributed around the genome in an apparently random fashion.

Tandemly repeated DNA:-

Whose repeat units are distributed or placed next to each other in an array.

Tandemly repeated DNA is also called satellite DNA because DNA fragments containing tendemly repeated sequences from "satellite" bonds when genomic DNA is fractionated by density gradient centrifugation. For e.g. when broken into fragments 50-100 kb in length, human DNA forms a main band and three satellite bands. The main band contains DNA fragment made up mostly of single copy sequences with GC compositions close to 40.3%, the average value for human genome. The satellite bands contain fragments of repetitive DNA, and hence have GC contents and buoyant densities that are a typical of the genome as a whole. This repetitive DNA is made up of long series of tandem repeats, about 100 of kb's in length. A single genome can contain several different types of satellite DNA, each with a different repeat unit; these units can be from less than 5 to more than 200bp in length. The three satellite band in human DNA includes at least four different repeat types.

Some satellites DNA is scattered around the genome, most is located in the centromeres where it may play a structural role possibly as binding sites for one or more of the special centromeric proteins.(Fig)

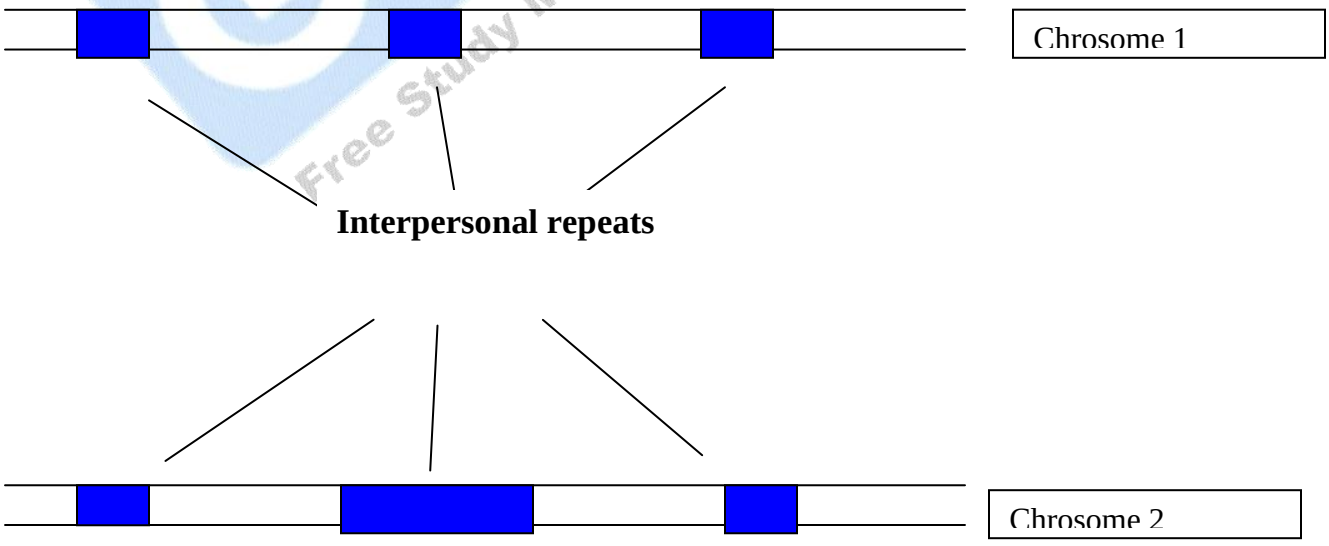


Fig:- The two types of repetitive DNA , interspersed repeat and tandemly repeated

DNA.

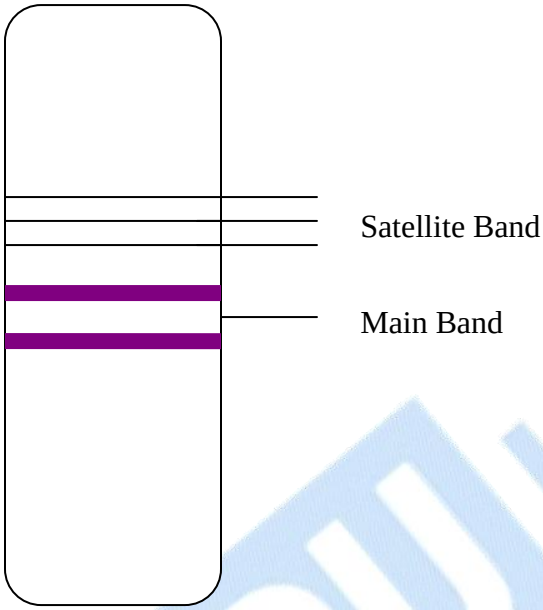


Fig :- Satellite DNA from the human genome.

There are two other types of tandemly repeated DNA:-

a. **Minisatellittes**:-

They form clusters up to 20 kb in length, with repeat units up to 25 bp in length while microsettelites clusters are shorter, usually less then 150 bp and the repeat unit is usually 13bp or less.

MinisatelliteDNA is a second type of repetitive DNA and is associated with structured features of chromosomes. Telomeric DNA which in humans comprises hundreds of copies of the motif 5'-TTAGGG-3' is an e.g. of minisatelites. Some eukaryotic genomescontain varous other clusters of minisatellite DNA many not all near the ends of chromosomes.

b. Microsatellites:-

Is also an example of tandemly repeated DNA. In a microsatellite the repeat units usually short up to 13 bp in length. The commenest type of human microsatellites is dinucleotide repeats, with ~1, 40,000 copies in the genome as a whole, about half of

these being repeats of the motif "A" single nucleotide repeat. (e.g. AAAAA) are the next most common (about 1,20,000 copies in total) function of microsatellites is not clear. It is k/n that they arise through an error in the process responsible for copying of the genome during cell division, and they might simply be unavoidable products of genome replication.

Many Microsatellites are variable and have a wide application. Their variability refers to the no of repeat units in the array is different in different members of the species. This is because "slippage" sometimes occurs when a microsetellites is copied during DNA replication. Leading to insertion or less frequently deletion of one or more of the repeat units. No two humans alive today have exactly the same combination of microsatellite length variation: if enough Microsatellite are examined then a unique genetic profile can be established for every person. The only exceptions are genetically identical twins. It also important in identification of criminals acting as a tool for forensic science.

Multiple Choice Questions

Q1.Which of the following is a method for analyzing human genetics?
(i)LOD score
(ii) Pedigree
(iii) Cross breeding

(iv)All of these (ii)

Q2.Pedigree is
(i)Diagram that represents family relationship
(ii) Represent diseases in family
(iii) Represent superhuman traits

(iv)Represent only dominant traits (i)

Q3.If a trait is dominant then in pedigree chart
(i)Will appear in every generation
(ii) Will appear in alternate generation
(iii) Will appear in distant generations

(iv) Will not inherit further (i)

Q4. If a trait is autosomal recessive then
(i) Both parents are affected; none of the children will be affected.
(ii) Both parents are affected; the children might or might not be affected.
(iii) Both parents are affected; all of the children will be affected.
(iv) None of these (iii)

Q5. If the trait is an X-linked recessive which of the following statements are true

(a) children will not have the trait.
(b) children might or might not have the trait.
(c) all of the children will have the trait.
(d) the females will have the trait but males will only have the trait if their father also has the trait.
(e) males will have the trait, but females will only have the trait if their father also has the trait.

(i) a, b, c, d (ii) a, b, c (iii) b, e (iv) c, d, b (iii)

Q6. Proposita in pedigree is

(i) Male, who was first in which the disease was identified
(ii) Female, who was first in which the disease was identified
(iii) Female, who carries the genes for disease
(iv) Male, who carries the genes for disease (ii)

Q7.Which of the following is an example of autosomal recessive trait?

(i) Albinism
(ii) Alzheimer
(iii) Color- blindness
(iv) Hemophilia (i)

Q8.Which of the following is an example of X-linked recessive trait?

(i) Duchenn muscular dystrophy
(ii)Testicular feminization syndrome
(iii)Hemophilia
(iv)All of these (iv)

Q9.Which of the following is an example of genome wide repeats?
(i)LINEs
(ii) SINEs

(iii) LTR
(iv)All of these. (iv)

Q10. The interspersed repeats are
(i)Repeats whose individual repeat units are distributed in genome in random fashion
(ii)Repeats distributed alternately in genome
(iii)Are distributed evenly throughout the genome
(iv) All of these (i)

Q11.Molecular markers are

(i) Genes that do not inherit
(ii) DNA sequences similar to genes but not genes
(iii) DNA sequences that are not genes but inheritable

(iv) None of these (ii)

Q12. Which of the following is an example of markers used for genetic mapping?
(i)RFLPs
(ii)SSLPs
(iii)SNPs
(iv)All of these (iv)

Q13. Which of the following are the variants of SSLP
(i)VNTRs
(ii)STRs
(iii)SNPs
(iv)(i) and (ii) both (iv)

Q14.Multigene families are
(i) Cluster of several similar genes
(ii) Formed by duplication of a single original gene
(iii) Generally with similar biochemical functions

(iv) All of these (iv)

Q15.Which of the following multigene family expressed only during a particular state of development
(i) Simple multigene family
(ii)Complex multigene family
(iii)Developmentally controlled complex multigene family
(iv) All of these (iii)

Q16. Which of the following is an example of Developmentally controlled complex multigene family?
(i) Histone family of *Xenopus levis*
(ii)Immunoglobin superfamily of humans
(iii)Hemoglobin family of humans
(iv)None of these (iii)

Q17. Which of the following represents repetitive DNA?
i) SINEs
(ii)LINEs
(iii)VNTRs
(iv)All of these (iv)

Q18.Which of the following mechanism generate repetitive DNA?
(i) Transposition
(ii)Deletion
(iii)Duplication
(iv)(i) and (iii) both (iv)

Q19.C- value refers to
(i)Amount of DNA contained within a haploid nucleus
(ii) Amount of DNA contained within whole genome
(iii) Amount of DNA contained within the cell
(iv)None of these (i)

Section C

Physical Mapping

Q.1 Describe the role of FISH for the construction of a physical map?

Ans **FISH**:- Fluorescent in situ hybridization

Definition:

A technique for locating markers on chromosomes by observing the hybridization position of fluorescent labels.

For the construction of a physical map FISH plays an important role where mapping of marker location are done by hybridizing a probe containing the marker to intact chromosomes.

It is a type of optical mapping method which enables the position of a marker on a chromosomes or extended DNA molecule to be directly visualized. In optical mapping the marker is a restriction site and it is visualized as a gap in an extended DNA fiber. In FISH the marker is a DNA sequence that is visualized by hybridization with a fluorescent probe.

Techniques:-

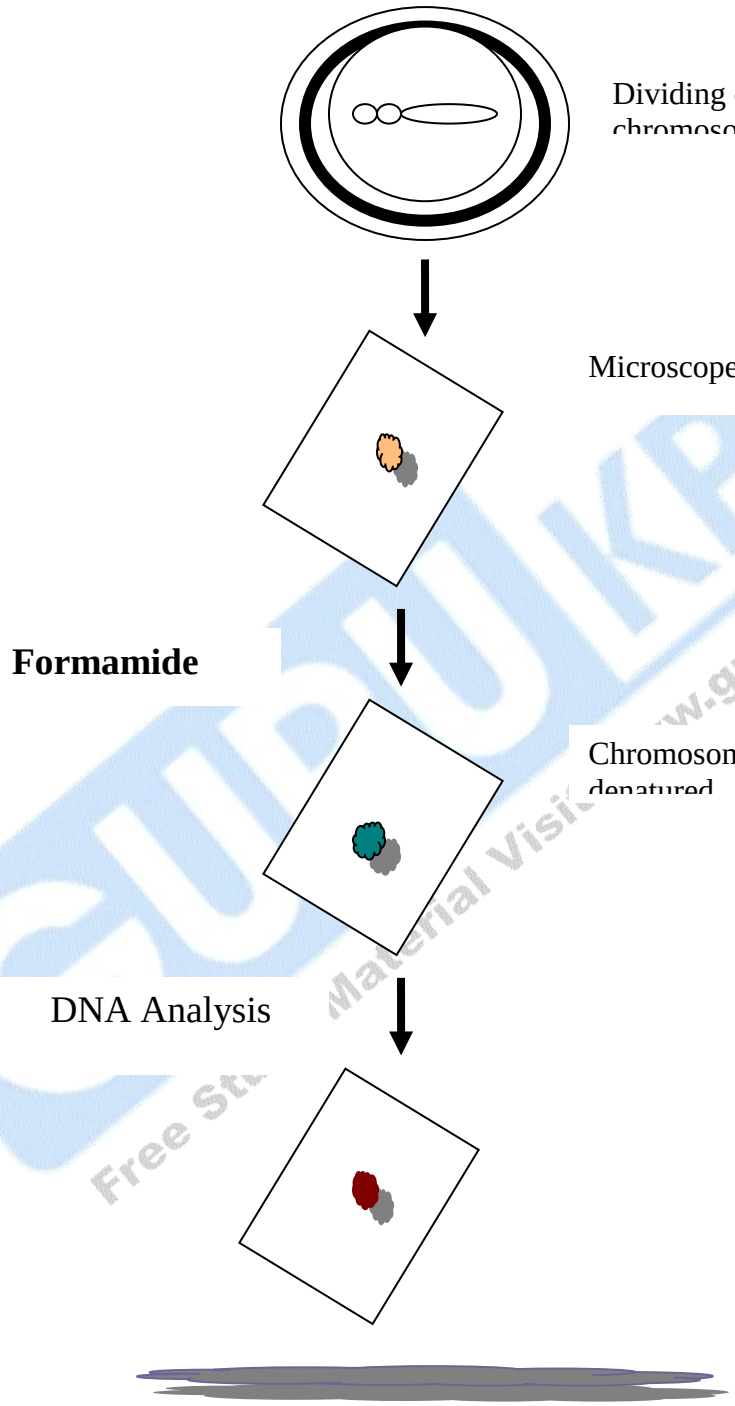
FISH is a version of hybridization analysis where the position on the chromosomes at which hybridization analysis where the position on the chromosomes at which hybridization occurs provides information about the map location of the DNA sequences used as the probe.

For the method to work, the DNA in the chromosomes must be made single stranded (denatured) by breaking the base pairs that hold the double helix together. Only then the chromosomal DNA is able to hybridize with the probe.

The standard method for denaturing chromosomal DNA without destroying the morphology of the chromosome is to dry the preparation on to a glass. Microscope slide and then treat with formamide. For FISH to be successful two requirements must be filled.

1. Sensitivity
2. Resolution

- Sensitivity requires that the radioactive label has a high emission energy(e.g.32 $_{P}$), but if radiolabel has a high emission energy then it scatters its signal and so gives poor resolution.
- High resolution is possible if a radiolabel with low emission energy like ^{3}H is used but these radiolabels have such low sensitivity that lengthy exposures are needed, leading to a high background and difficulties in discerning the genuine signal but now a days fluorescent labels which are non radioactive are used gives high resolution and sensitivity.
- To map the relative positions of the probes sequence fluorolabels with different colored emission have been designed, making it possible to hybridized a no. of different probes to a single chromosome and distinguish their individual hybridization signals.
- To maximize sensitivity the probes must be labeled as heavily as possible which in the past has meant that they must be quite lengthy DNA molecules usually cloned DNA fragments of at least 40kb .
- For the construction of a physical map a cloned DNA fragment can be looked as another type of marker clones as marker can be used because the cloned DNA in the material from which the DNA sequence in determined.
- Mapping the positions of clones therefore provides a direct link between genome map and its DNA sequence



Single Frone Probe

Fig: - FISH: - A sample of dividing cells is dried on to a microscope slide and treated with formamide so that the chromosomes become denatured but do not loose their characteristic metaphase morphologies. The position at which the probe hybridizes to the chromosomal DNA is visualized by detecting the fluorescent signal emitted by the labeled DNA.

Mapping via using FISH with Metaphase chromosomes:-

FISH was originally used with metaphase chromosomes; a fluorescent signal obtained by FISH mapped measuring its position relative to the end of the short arm of the chromosome.

A disadvantage of this method is that the highly condensed nature of metaphase chromosomes means that only low resolution mapping is possible, two markers must be at least 1 MB apart to be resolved as separated hybridization signals. This degree of resolution is insufficient for the construction of useful chromosomes maps, and the main application of metaphase FISH has been in determining the chromosome on which a new marker is located and providing a rough idea of its map position as a preliminary to finer scale mapping by other methods.

As metaphase chromosomes are too condensed for fine scale mapping then we must use chromosome that are more extended. There are two ways of doing this:

a. Mechanically stretched chromosomes :-

Can be obtained by modifying the preparative method used to isolate chromosomes from metaphase nuclei. The inclusion of a centrifugation step generate shear forces which can result in the chromosomes becoming stretched to up to 20 times their normal length FISH signals can me mapped in the same way as with normal metaphase chromosomes. The resolution is significantly improved and markers that are 200-300 kb apart can be distinguished.

b. Non- metaphase chromosomes :-

Can be used because it is only during metaphase that chromosomes are highly condensed at other stages of the cell cycle the chromosomes are naturally unpacked. Attempts are made to use prophase nuclei because in these the chromosomes are still sufficiently condensed for individual chromosomes to be identified. However these preparations provide no advantage over mechanically stretched chromosomes. Interphase chromosomes are more useful because this stage of the cell cycle (between nuclear divisions) is when

the chromosomes are most unpacked. Resolution down to 25kb impossible but chromosome morphology is lost so there are no external reference points against which to map the position of the probe. This technique is therefore used after preliminary map information has been obtained, usually as a means of determining the order of a series of markers a small region of a chromosome.

Interphase chromosomes contain the most unpacked of all cellular DNA molecules. To improve the resolution of FISH to better than 25 kb it is therefore necessary to abandon intact chromosomes and instead use purified DNA. This approach called fiber FISH makes use of DNA prepared by gel stretching or molecular combing and can distinguish marker that are less than 10 kb apart.

Q.2 Explain what do you understand by the following?

a. Chromosomal walking

b. Chromosomal painting

Ans. **<u>Chromosomal walking</u> :-**

It was the first method devised for assembly of clone contigs.

The simplest way to build up an overlapping series of cloned DNA fragments is to begin with one clone from a library, identify a second clone whose insert overlaps with the insert in the first clone. Then identify a third clone whose insert overlaps with the second clone, and so on. This forms the basis of chromosome walking.

Chromosome walking was originally used to move relatively short distance along DNA molecules, using clone libraries prepared with or comid vectors. The most straightforward approach is to use the insert DNA from the starting clone as a hybridization probe to screen all the other clones in the library. Clones whose inserts overlap with the probe gives positive hybridization signal, and their inserts can be used as new probes to continue the walk.

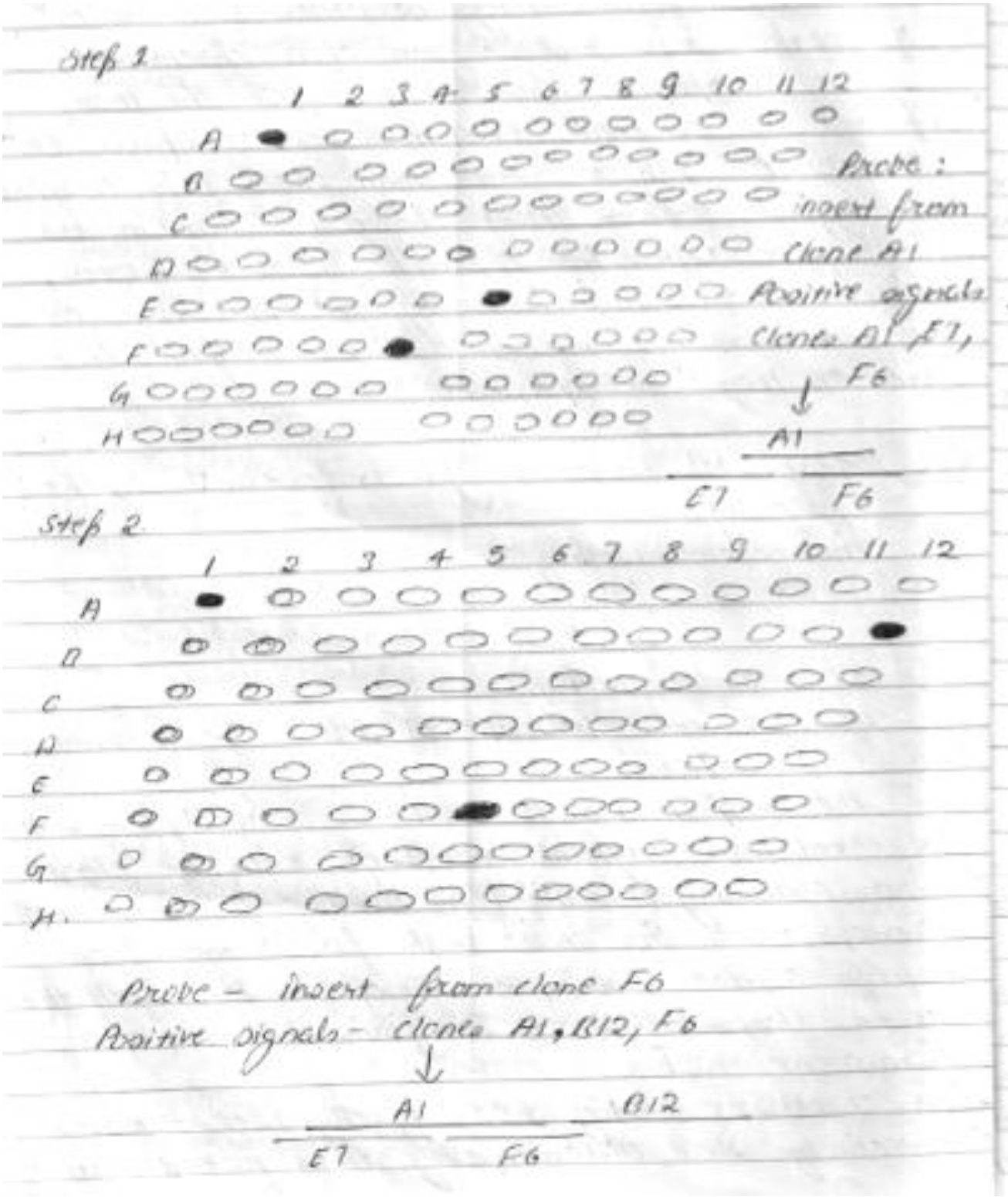


Fig: - chromosome walking: - the library comprises 96 clones, each containing a different insert. To begin the walk the insert from one of the clones is used as a

hybridization probe against all the other clones in the library .In the example shown cloneA1 is the probe it hybridizes to itself and to clones E7 and F6. The inserts from last two clones must therefore overlap with the insert form clones A1. To continue the walk, the probing is repeated but this time with the insert from clone F6 .Thevhybridizing lones are A1 F6 and B12 showing that the insert from B12 overlaps with the insert from F6.

- The main problem that arises is that if the probe contains a repeat sequence then it will hybridize not only to overlapping clones but also to non overlapping clones whose inserts contain copies of the repeat. The extent of this non-specific hybridization can be overcome by blocking repeat sequences by prehybridization with unlabeled genomic DNA. But these don't completely solve the problem, especially If long inserts from a high capacity vector s/a BAC is used for chromosome walk. Therefore intact inserts are rarely used for the purpose with human DNA and similar DNAs, which have a high frequency of repeat sequences. Instead a fragment from the end of an insert is used as the probe to minimize. The chance of a repeat occurring in a short end- fragment compared with the insert as a whole.
- After sequencing the end-fragment walk can be speeded up by using PCR. PCRs provide enough information to identify which of the total clones give products and which do not.

a. <u>Chromosome painting</u> :-

It is a version of FISH in which the hybridization probe is a mixture of DNA molecules, each specified for different regions of a single chromosome.

When applied to interphase nuclei, chromosome painting reveals territories occupied by individual chromosomes.These territories take up the majority of the space within the nucleus, but are separated from, one another by non chromatin regions within which the enzymes and other proteins involved in expression of the genome are located.

Chromosome painting allows the visualization of individual chromosomes in metaphase or interphase cells and the identification of both numerical and structural chromosomal aberrations in human pathology with high sensitivity and specificity.

In addition to human chromosomes specific probe pools painting probes have become available for an increasing range of different species. They can be applied to cross

species comparison as well as to the study of chromosomal rearrangements in animal models of human diseases. The simultaneous hybridization of multiple chromosome painting probes each tagged with a specific fluorochrome or fluorochrome combination has resulted in the different color display of human(and mouse) chromosome i.e. colour karyotyping.

Multicolor Chromosome Painting:-

One of the most attractive features of FISH is the possibility to distinguish in a single experiments, multiple chromosomes or chromosomal targets simultaneously.

The research on multicolor FISH was started in Netherlands. Using 3 fluorochromes, 7-amino-4methylocumarine-3 acetic acid (AMCA), fluorescein isothiocyanate (FITC) and rhodamine, nederlof et al could show in 1989 that simultaneous visualization of differentially labeled FISH probes in indeed possible.

There are two alternative approaches were used to develop colour karyotyping chromosomes.

(I) Fluorochrome specific optical fitters, termed m-FISH and

(II) Interferometer based spectral imaging (introduced as spectral karyotyping or SKY.

The filter based approach was developed by Speicher and colleagues and employs the sequential image acquisition with five different fluorochrome specific optical filters.

Narrow band pass filters were designed to allow maximum distinction. Of fluorescent dyer.

After a computer based image shift correction the image analysis software calculates a chromosome segmentation mask based on the 4(prime), 6-amino-2-phenyl (DAPI) image. In this pre-defined area, the intensities for all subsequently acquired images will be recorded. Based on the labeling schemes, the program then decides wheather a single pixel contains signals from one fluorescent dye or from 2 or more fluorochromes. This information forms the basis of chromosome identification and colour assignment.

Probes for chromosome painting:-

For chromosome painting probes are used the initial of first generation probes were based on chromosome specific phage librariesbut were cumbersome to use due to low insert to vector ratio which frequently resulted in a relatively high background staining.But now a days new improved probes based on plasmid libraries are used where an improved and easier probe generation enhanced the painting quality considerably. However in some instances sub regions of chromosomes notoriously were stained more weakly than others (e.g. the tip of chromosome 1p), which made the interpretation of hybridization results more difficult. Now 2 additional protocols for generations of chromosome painting probes have become available:-

(I) Based on chromosome flow sorting and subsequent PCR amplification using degenerates primers, high quality probes are accessible revealing a considerably improved signal to noise ratio, along with an improved staining homogeneity.

(II) The same applies to probes generated by micro manipulated dissection of normal metaphase chromosomes, again followed by sequence independent DNA amplification.

Both approaches can be extended to produced probes for reverse chromosomes were either flow sorted or micro dissected and subsequently used as painting probes on normal metaphase chromosomal in order to establish the origin of chromosomal material in marker chromosomes. Microdissection probes come with the distinct advantage that in addition to whole chromosome painting probes, region specific probes for chromosomal arms or chromosomal bands can be generated. Flow sorted chromosome painting probes however has the advantage that the target number for subsequent DNA amplification can be extended easily. Thus ensuring a high complexity of the painting probes. Chromosome painting probes are now also available for ever increasing number of species, most notably for the mouse and the rat, allowing the expansion of chromosome painting analysis to animal model of human diseases. Lastly the widespread use of chromosome painting in non-specialized laboratories has also become possible due to improved microscope hardware (microscope and optical filters), the use of sensitive digital imaging devices (CCD camera confocal laser scanning microscopes) and an increasing no of suitable DNA heptenization and fluorescent labeling systems.

Q.2 Write short notes on:-

a. RAPD

b. VNTR

c. SNP
d. RACE

Ans.

a. RAPD:-

It refers to randomly amplified polymorphic DNA.

RAPD analysis is a PCR-based molecular marker technique. Here single short oligonucleotide primers are arbitrarily selected to amplify a set of DNA segments distributed randomly throughout the genome.

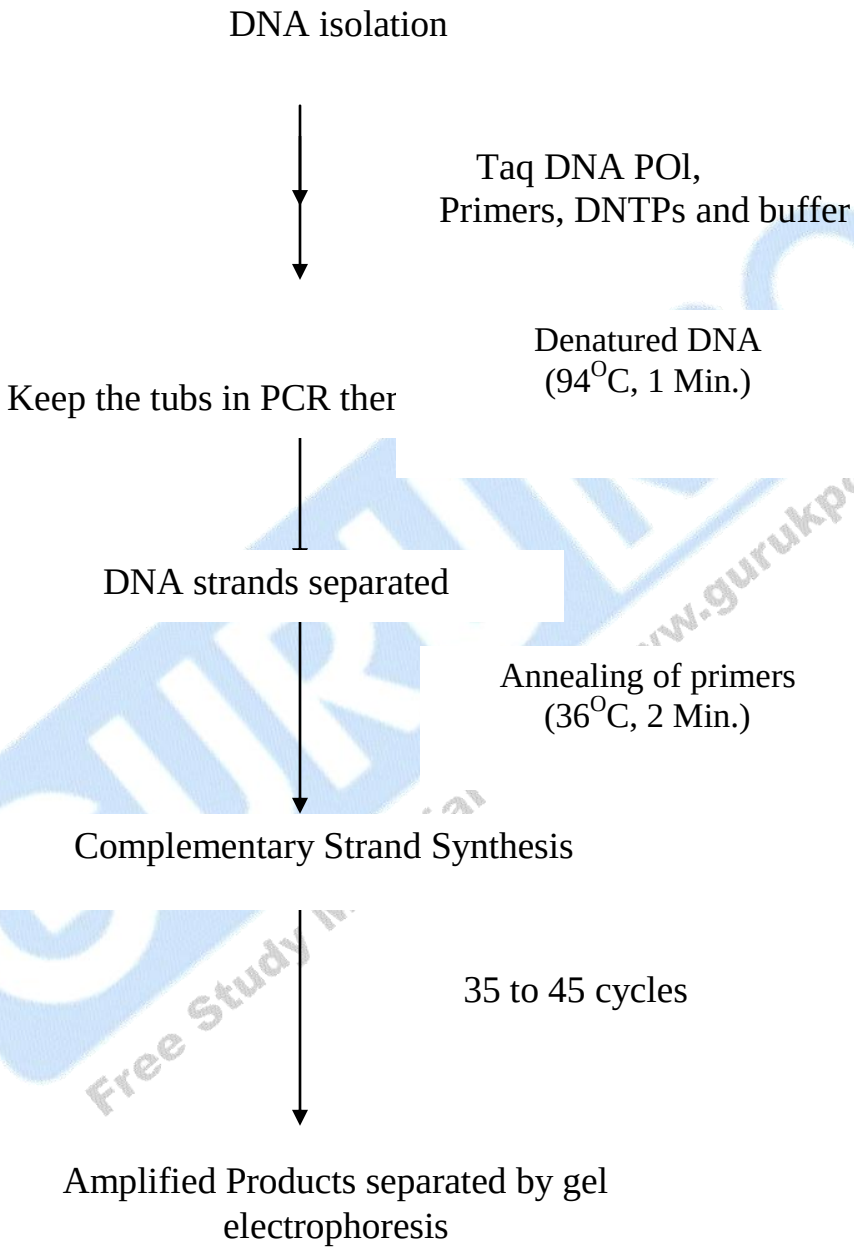
RAPD amplification is preformed in conditions resembling those of PCR, using genomic DNA form the species of interest and a single short oligonucleotide (usually a 10-base primer).

The DNA amplification product is generated from a region that is flanked by a part of 10bp priming site in the appropriate orientation.

Genomic DNA from two different individuals often produces different amplification patterns.

A particular fragment generated for one individual but not for others represents DNA polymorphism and can be used as a genetic marker. Using different combinations of nucleotides, many random oligonucleotide primers have been designed and such primers can be synthesized in an oligonucleotide synthesizing facility based on sequences chosen at random. No separate information is required from the plant to be studied. The choice of single primers or rapid primers to use is done operationally. Since each random primer will anneal to different region of the DNA, theoretically many different loci can be analyzed.

RAPD Analysis:-



Analysis of RAPD markers:-

The reaction products are conveniently analyzed on agarose gels, stained with ethidium bromide and seen under UV light.

Presence of a RAPD based corresponds to a dominant allele against absence of band that corresponds. To a recessive allele.

Thus heterozygote and homozygote dominant individuals cannot be differentiated with RAPD marker. The segregating F_2 population may therefore. Be scored as follows:

Band presents	AA (Homozygote)
	Aa (Heterozygote)
Band absent	aa (Recessive Homezygote)

In F_2, these two classes would be expected to segregate in a 3:1 ratio. In case of RFLP this causes a loss of information as it shows co dominance. Backcross, recombinant inbred, and the doubled haploid population do not suffer this loss of mapping information as complete information available from a backcross can be obtained from a scoring the presence or absence of a polymorphic markers. This fact makes it easy to select the populations best suited for the construction of genetic maps with RAPDs markers.

Advantage of RAPD markers:-

1. Need for a small amount of DNA (15-25ng) makes it possible to work with populations which are inaccessible for RFLP analysis.
2. It involves nonradioactive assays.
3. It needs a simple experimental setup requiring only a thermocycler and an agarose assembly.
4. It does not require species specific probe libraries; thus work can be conducted on a large variety of species where such probe libraries are not available.
5. it provides a quick and efficient screening for DNA sequence based polymorphism at many loci.
6. it does not involve blotting or hybridization steps.

Applications:-

1. For construction of genetic maps used as genetic markers for construction of genetic map in *Arabidopsis*, pine, *Helianthus* etc.

2. Mapping of traits:-RAPD markers may also be used for indirect selection in segregating populations during plant breeding programs. Also employed for tagging genes of economic value.
3. Used for analysis of genetic str. of populations.
4. Fingerprinting of individuals.
5. Identification of somatic hybrids.
6. Useful system for evaluation and characterization of genetic resources.

Limitations:-

1. Are inherited as dominant- recessive characters. These causes a loss of information relative to markers which show co dominance.
2. They are relatively short; a mismatch of even a single nucleotide can often prevent the primer from annealing hence loss of a band.
3. Sensitive to changes in PCR conditions, resulting in changes in some of the amplified fragments.

b. VNTRs:-Variable number tandem repeats**.**

Definition:-

It is s type of simple sequence length polymorphism comprising tandem copies of repeats that are a few tens of nucleotides in length. Also c/d Minisatellites.

They have a repeat unit up to 25kb in length and form clusters up to 20 kb in length, it is associated with structural features of chromosomes telomeric DNA. Which in human comprises hundreds of copies of the motif 5' –TTAGGG-3', is an e.g. of a minisatellite telomeric DNA, some eukaryotic genome contain various other clusters of minisatellite DNA, many although not all near the ends of chromosomes. The functions of these other minisatellite sequences has not been identified.

c. SNPs:-

Definition:- Single nucleotide polymorphism is a type of point mutation that is carried by some individuals of a population. SNPs are a position in a genome where some individuals have one nucleotide (e.g. a G) and others have a different nucleotide (e.g. a C). There are vast no. of SNPs in a genome (over 4 million in human

genome), some of which also give rise to RFLPs, but many of which do not because the sequence in which they lie is not recognized by any restriction enzyme.

Origin of SNPs:-

Any one of the 4 nucleotide could present at any one position in the genome, so it might be imagined. That each SNP should have four alleles. Theoretically it is possible but in practice most SNP's exists just as two variants. This is because each SNP originates when a point mutation occur in a genome, converting one nucleotide into another. if the mutation is in the reproductive cells of an individual, then one or more of that individuals offspring might inherit the mutation and after many generations the SNP may eventually become established in the population . But there are just two alleles the original sequence and mutated version.

For a third allele to arise a new mutation must occur at the same position in the genome in another individual and this individual and his or her offspring must reproduce in such a way that the new allele becomes established. This scenario is not impossible but it is unlikely: consequently the vast majority of SNPs are ballelic. This disadvantage is overcome by the huge no. Of SNPs present in each genome in most eukaryotes, at least one for every l0 Kb of DNA. SNPs therefore enable very detailed genome maps to be constructed.

Applications of SNPs:-

Importance that SNPs have acquired in genome research has stimulated the development of rapid method for their typing. Most of these methods are based on oligonucleotide hybridization analysis.

(a) DNA Chip Technology:-

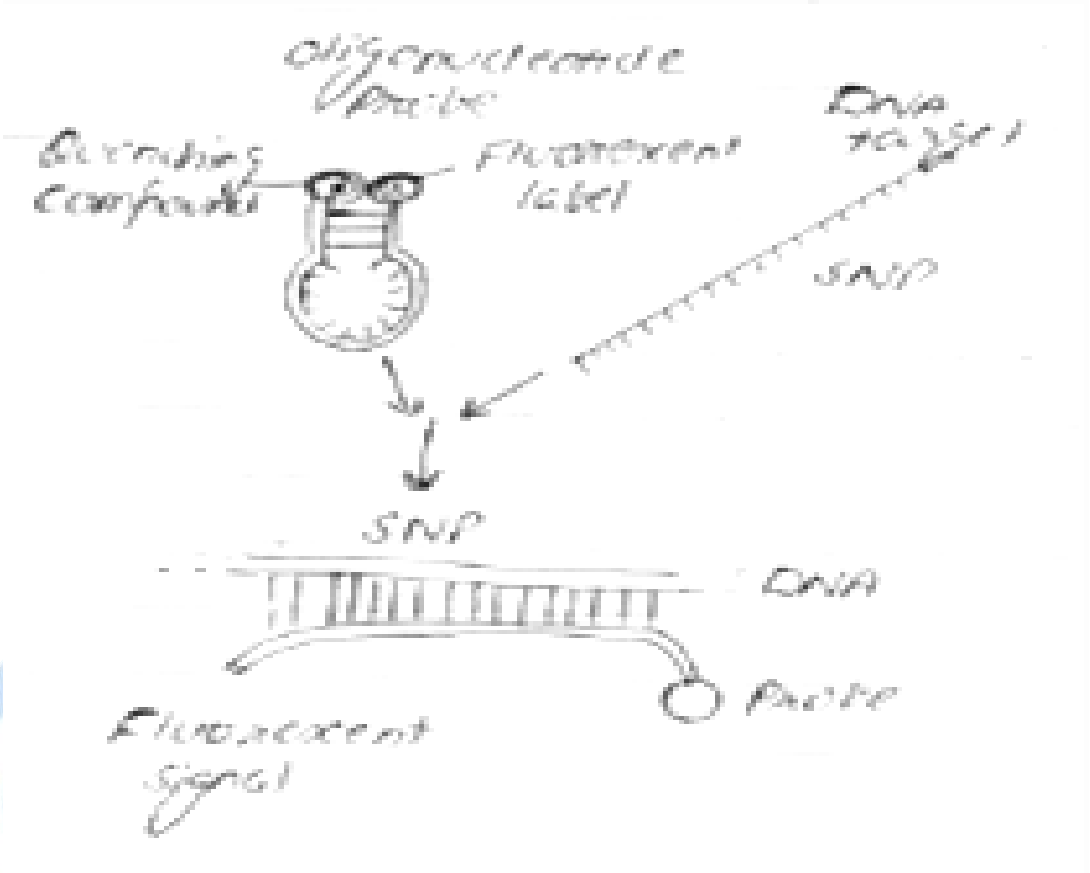
It makes use of a wafer of glass or silicon 2.0 cm^2 or less in area, carrying many different oligonucleots in a high – density array. The DNA to be tested is labeled with a fluorescent marker and pipetted on to the surface of the chip. Hybridization detected by examining the chip with a fluorescent microscope, the position at which the fluorescent signal is emitted indicating which oligonucleotides have hybridized with the test DNA. Therefore, many SNPs can be scored in a single experiment.

(b) **Solution hybridization techniques**:-

The technique is carried out in the wells of a micro titer tray, each well containing a different oligonucleotides using detection system that can discriminate b/w

unhybridized, single stranded DNA and the double stranded product that results when an oligonucleotides hybridize to test DNA.

Fig: - Detecting hybridization by dye – quenching.



One way of typing an SNP by solution hybridization. The oligonucleotide probe has two end labels. One of these is a fluorescent dye and the other is a quenching compound. The two ends of the oligonucleotide base-pair to one another, the fluorescent signals are quenched .When the probes hybridizes to its target DNA, the ends of the molecule becomes separated, enabling the fluorescent dye to emit its signal. The two labels are c/d molecular beacones'.

Other typing method makes use of an oligonucleotide whose mismatch with SNP accurse at its extreme 5'or 3'end. Under appropriate conditions oligonuleotide of this type will hybridize to the mismatched template DNA with a short, non-base paired "tail" this feature is utilized in 2 ways.

(a) **Oligonucleotide ligation assay (OLA)**:-

It make use of two oligoncleotides that anneal adjacent to one another, with the 3'end of one of these oligonucleotides positioned exactly at the SNP, This oligonucleotide will form a completely base paired structure if one version of the SNP is present in the template DNA.

(b) **Amplification refractory mutation system or ARMS test**:-

Is based on the same principle as OLA, but in this method the test oligonucleotide is one of a pair of PCR Primers. If the test primer anneals to the SNP then it can be extended by Taq polymerase and the PCR can take – place, but if it does not anneal, because the alternative version of the SNP is present then no PCR product is generated.

(IV) **RACE**:-

Rapid amplification of cDNA ends:-

For incomplete cDNA more robust methods are needed for locating the precise start and end points of gene transcripts one way is using a special type of PCR that uses RNA rather than DNA as the starting materials the first step in this type of PCR is to convert the RNA into cDNA with reverse transcriptase, after which the cDNA is amplified with taq polymerase in the same way as in a normal PCR. This method is c/d RT-PCR or reverse transcriptase PCR.

RT-PCR can be possible used to amplify a complete CDNA coding region. But in many applications there may be insufficient information for a straight forward strategy. In general, cDNA synthesis and library screening can lead to the isolation of gene sequences that represent all or part of a processed mRNA transcript There are always problems when trying to isolate the 5'end of low abundance transcripts because (DNA libraries often do not contain a full representation of all mRNA sequences. This is especially true of oligo dt primed libraries. RACE protocols generates cDNA fragments primers by using PCR to amplify sequence of a k/n gene b/w a single region in the mRNA and either the 3' or the 5' end of the transcript. Thus it is necessary to know or to deduce a single stretch sequence within mRNA to choose a one gene specific primer complementarty to the sequence of cDNA segment and a second "anchored primer" that anneal to a sequence that has been covalently attaches to the newly created cDNA terminus. This strategy was first described by Frohman etal 1988.

RACE PROTOCOLS :- There are two RACE protocol in the RACE protocols, extension of the (DNA) from the ends of the transcript to the specific primers is accomplished by using primers that hybridize either at the natural 3' poly (A) tail of the mRNA or at a synthetic poly (A) tail added to the 5' end of the first – strand of cDNA.

In the RACE 3' end protocol, mRNA is reverse transcribed using an olio (dT17) primer, which has a 17- nucleotide extension at its 5'end because of the natural 3' poly (A) tail of the mRNA. In this extension, the anchor sequences are designed to contain the restriction sites for subsequent cloning. Amplification is performed using the anchor 17- mer and a primer specific for the sought – after cDNA.

In the RACE 5' end protocol, the mRNA is reverse transcribed from a gene-specific primer. The resultant (DNA is then extended by terminal deoxytidyl transferase (TdT) to create a poly (dA) tail at the 3' end of the CDNA. Amplification is performed with the oligo dT 17) anchor system as well for the 3' protocol and the specific primer. 5' RACE is used to amplify uncloned upstream transcript sequences using a gene specific primer.

Modifications of RACE Protocol:-

Several modifications of original RACE protocol have been described. One of these, called RNA ligase mediated RACE (RLM-RACE), involves the use of bacteriophage T4 RNA ligase to attach covalently a single stranded RNA anchor molecule to the developed 5' end of the mRNA. First strand cDNA synthesis can be performed using a gene-specific primer to produce a pool of cDNA encoding the anchored primer sequence. Variation of this method is referred to as ligation – anchored PCR (LA-PCR).

Q.4 Give an explanatory note on RFLP discussing its role in construction of physical map?

Ans. RFLP was the first technology that enabled the detection of poly morphism at the DNA sequence level.

Definition:-

RFLP refers to Restriction fragment length polymorphism. It is a restriction fragment whose length is variable because of the presence of a polymorphic restriction site at one or both ends.

Procedure:- Genomic DNA isolation

↓ Digestion with restriction enzyme

Fractionation on an agarose gel

↓

DNA in million of restrictions fragments
Fractionated in the gel by molecular weight

↓

DNA transferred out of the gel on to a membrane filter

Radioactive labeling nucleic acid Hybridization ↓ Incubation with cloned probe and allow for hybridization

Southern hybridization

↓ Autoradiography

RFLP pattern with positive bands.

Analysis of RFLP:-

Since RFLP of nuclear DNA cannot be directly visualized, the usual procedure is to use small pieces of chromosomal DNA as probes to detect individual restriction fragments. Using the high specificity of DNA-DNA hybridization, such probes can detect individual restriction fragments in the complex mixture of fragments of nuclear DNA present in a restriction digest.

To use this technique, a set of chromosomal DNA fragments are prepared as probes. For preparing probes DNA is isolated from the species of interest, digested with a restriction enzyme to generate relatively small fragments, usually of 0.5- 2.0 Kb. Individual restriction fragments are ligated into a bacterial plasmid and the plasmid is transformed into a bacterial cell. By growing these transformed bacteria, one obtains a large supply of a single plant DNA restriction fragments, which are suitable for use as a hybridization probe. Such a set of probes is c/d library. Southern blots are

prepared from the digested DNA as explained earlier and then probed with one of the cloned probes from the library.

Role of RFLP for the Construction of Physical Maps:-

RFLP as a molecular marker play an important role for construction of physical map. When molecular marker tightly linked to a gene of interest are identified, the next step in map-based cloning is physical mapping. Although segregation analysis yields vital information on the orientation and genetic distance between RFLPs in terms of recombination frequency (cM), it is still essential to know the actual physical distance in terms of nucleotides b/w RFLPs that flank a target gene. RFLPs that appear to be very close to a gene in terms of genetic distance may still be very far away in terms of physical distance. It is very essential to know the magnitude of the physical distance between RFLPs before initiating the next step of chromosome walking and map-based cloning.

RFLPs are used as a starting point for chromosome walking. Two RFLPs lying closest to the gene of interest yet flanking on either side are chosen chromosome walking begins by identifying genomic clone that overlap the initial RFLPs, which is a accomplished by probing a complete genomic library with radio labeled nucleic acid probes synthesized from the initial RFLPs. The newly identified genomic clones are then isolated and the ends of these clones become starting point for the next step in the chromosome walk. Thus by determining the position and physical distances b/w RFLPs physical map can be constructed.

Application of RFLP:-

1. It permits direct identification of a genotype or cultivar in any tissue at any developmental stage in an environment independent manner.
2. They are co dominant markers, enabling heterozygote to be distinguished from homozygotes.
3. Indirect selection using qualitative traits: - The greater potential use of RFLP is its role as indirect selection criteria. It is useful in conjunction with analysis of conventional markers. RFLP maps can be used to supplement regular plant breeding protocols when one wants to select for a conventional gene, but directly selecting for that gene would be expensive, difficult or time consuming.
4. It has a discriminating power that can be at the species / population (single locus probes) or individual level (multi locus probes).

5. Tagging of monogenic traits with RFLP markers: - In inbreeding species, the availability of nearly isogenic lines (NILs) facilitates detection of RFLP markers linked to monogenic disease- resistance genes.

6. Indirect selection using quantitative trait loci (QTLs) several characters of plant species among which are traits of agronomic importance are inherited quantitatively. This type of genetic variation is due to multiple factors acting collectively on the expression of a trait. These traits are c/d QTLs.

   Fundamental advances in this area of plant breeding seemed unlikely before the advent of RFLP mapping techniques. As one can use RFLP markers to simultaneously follow the segregation of all chromosome segments during a cross, the basic idea is to look for correlation b/w QTL of interest and specific chromosome segments. The RFLP markers are easily scored compared to OTL which must be scored in a conventional fashion.

   Limitations:-

   1. Conventional RFLP analysis requires relatively large amount of highly pure DNA.
   2. A constant good supply of probes that can reliably detect variation are needed.
   3. It is laborious and expensive to identify suitable marker / restriction enzyme combinations from genomic or CDNA libraries where no suitable single – locus probes are k/n to exist.
   4. RFLPs are time consuming as they are not amenable to automation.
   5. RFLP work is carried out using radioactively labeled probes and therefore required expertise in autoradiography.

Q.5 Distinguish between the following:-

RAPD and RFLP.

Ans. (a) RFLP and RAPD can be distinguished under following sub heads:-

Characteristic	RAPD	RFLP
1.) Principle	DNA amplification	Restriction digestic
2.) Detection	DNA staining	Southern blotting
3.) DNA required quantity	Low	High
4.) Primer requirement	Yes (random primer)	None
5.) Probe requirement	None	Set of specific probes
6.) Use of radioisotopes'	No	Yes
7.) Part of genome surveyed	Whole genome	Generally low copy coding region
8.) Dominant / co dominant	Dominant	Co dominant
9.) Polymorphism	Medium	Medium
10.) Automation	Yes	No
11.) Reliability	Intermediate	High
12.) Recurring cost	Low	High

Q.6 What is exon trapping? Describe the role of exon trapping in sequence analysis.

Ans. (i) **Exon trapping**:-

A method, based on cloning, for identifying the positions of exons in a DNA sequence.

Introduction:-

Exon trapping has been widely used in gene identification it is reliable and expression pattern independent protocol to identify genes from genomic DNA. It has proved to be very useful for the isolation of novel transcribed sequences very useful for the isolation of novel transcribed sequences and been used for large-scale gene identification on chromosome 21, 22 and other regions in the human genome since exon trapping does not rely on expression to identify genes, novel genes represented by trapped exons can be a substantial addition to the human gene content.

Exon trapping is also known as a versatile tissue – independent approach to detect genes in cloned DNA. In contrast to RNA- based methods, s/a cDNA selection and direct screening of CDNA libraries, exon trapping is independent of tissue-specific

gene expression. It uses cloned DNA directly to select sequences surrounded by functional splice sites.

Original exon trapping protocols have been improved with respect to speed and efficiency and improvements have been made to reduce the background consisting of cryptically spliced products and products arising from vector –vector splicing.

A major limitation is the need for sub cloning of region of interest in a vector with a capacity fro inserts typically measuring 1-2 kb. This may result into: (i) due to the small insert size after sub cloning, multiple exons will only rarely be present in one insert, resulting in exon trap clones containing only a single exon.

Consequently, many of the exon trap probes derived are small approx.80-150bp and frequently give poor signals or a high signal to noise ratio in subsequent experiments e.g. the screening of CDNA libraries or probing of northern blots. Furthermore, since the individually trapped exons' require the use of CDNA libraries in the next step to further define the gene, the initial advantage of working with an expression independent system is to a large extent lost in the subsequent step.

(ii) Due to sub cloning into plasmid-based exon trap vectors the gene (s) present are scattered into many separate disconnected pieces. Any exon thus obtained has to be aligned to reconstructs their original order. Reconstruction of the gene from individually trapped exons requires a significant amount of time and effort and implies a major loss of information originally contained within the input material prior to sub cloning.

(iii) Sub cloning disrupts the genomic context around the exons. Cloning of regions which are never transcribes. Or of intronic sequences without their naturally flanking exams often results in activation of cryptic splice sites, leading to recognition of false exams and a background of false positives. On the other hand genuine exons will be missed due to poor recognition of the host system or due to unfavorable factors resulting from the cloning (eg. Spacing of restriction sites).

(iv) Current exons trapping systems can only be used in combination with specific cell lines (eg. Cos cells since they require a system of replication in the host cell, commonly based on the SV40 origin of replication. Finally some exons of genes, with a highly tissue specific expression pattern will not be included in the nature transcripts generated in a completely different cell type.

3' Exon trapping:-

The 3' exon trapping recently described has some advantage in that it allows larger exons to be trapped, specifically identifies the end of a gene, and selects exons based on two independent criteria i.e. splicing and polyadenylation, it does not address the other limitations of small-insert exon trapping.

To overcome these limitations new large exon –trapping vectors have been designed capable of scanning 25-40 kb genomic regions for exons. The vector has a dual use : as cosmic vector for contig construction and physical mapping, and as exon trap vector for isolation of coding sequences.

Exon trapping for sequence analysis:-

Exon-trap vector

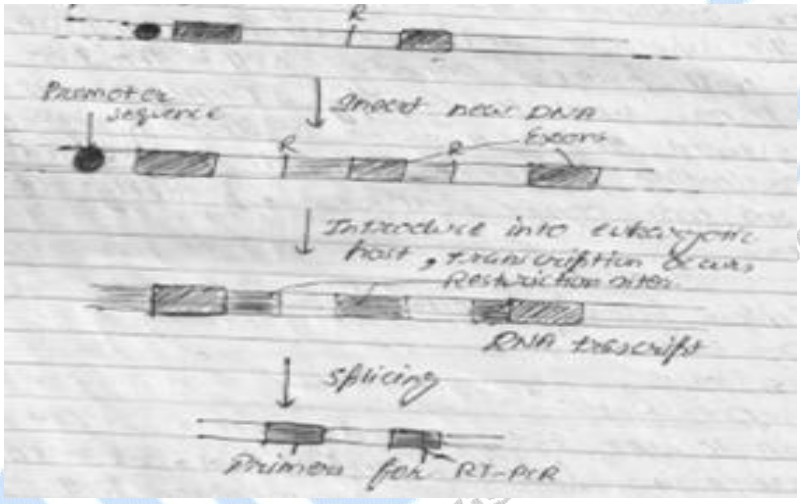


Fig: - Exon trapping for analysis of sequence. The exon trap vector consists of two sequence preceded by promoter sequences-the signals required for gene expression in a eukaryotic host. New DNA containing an unmapped exon is lighted into the vector. And the recombinant molecule introduced into the hsot cell. The resulting RNA transcript is then examined RT-PCR to identify the boundaries of the unmapped exon.

For finding genome sequences exon trapping is one of the methods. This requires a special type of vectors that contains a minigene consisting of two exons flanking on intrun sequence, the first exon being preceded by sequence signals needed to initiate transcription in a eakaryotic cell To use the vector, the piece of DNA to be studied is inserted into a restriction site located within the vectors intron region. The vector is then introduced into a suitable eukaryotic cell line, where it is transcribed and the RNA produced from it is spliced. The result is that any exon contained in the genome fragment becomes attached b/w the upstream and downstream exons from the minigene RT-PCR with primers annealing within the two indigene exons is now used to amplify a DNA fragment, which is sequenced. As the minigene sequence is already k/n, the nucleotide position at which the inserted exon starts and ends can be determined, precisely delineating this exon.

Q.7 Explain following techniques:-

(a) Dot blot

(b) Zoo blot

Ans. (a**) <u>Dot blot</u>**:-

In molecular biology and genetics, a blot is a method of transferring proteins, DNA or RNA, on to a carrier (e.g. a nitrocellulose ,PVDF or nylon membrane. In many instances, this is done after a gel electrophoresis, transferring the molecules from the gel onto the blotting membrane, and other times adding the samples directly onto the membrane. After blotting, the transferred proteins, DNA or blot DNA is first run into agarose gel , transfer to a membrane, then probe for your sequence of interest. In dot blot you simply spot your DNA cor RNA or proteins on membrane and then probe. This will show whether a particular sequence is present, but will not confirm size

<u>Dot and slot blot hybridization</u>:-

Dot or slots blot hybridization quantify the abundance of a certain RNA or DNA in the extracted nucleic acid mix without prior digestion and electrophoresis. In the procedure, the nucleic acid mix is blotted into a membrane, In dot blotting the nucleic acids are blotted as circular blots, whereas in slot blotting they are blotted into rectangular slots. The latter method allows a more precise observation of different hybridization with universal and specific oligonucleotide probes the relative abundance is calculated by dividing the amount of specific probe bound to a given sample by the amount of hybridized universal probe measured e.g. as fluorescence intensity fluorescent probes or wants per minute (radioactively labelled probes).

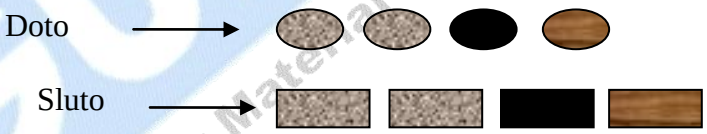


Fig: - Dot and Slots in dot and slot hybridization.

One should be aware that the data of relative DNA or RNA abundance can not directly translated in to cell numbers. Cells may have different copy numbers of various genes. Moreover, cells of different species have different ribosome contents ranging roughly b/w10^3 and 10^5 ribosome per cell. Even for one strain, cellular rRNA contents can vary significantly cat least over one order of magnitude, since they are directly correlated with growth rate. The relative rRNA abundance should, however, represent reasonable measurement of the RNA are then visualized by one or more different methods e.g. colorant sticking (sliver staining of proteins), autaradiographic. Appic visualization of radioactive labeled molecules, specific labeling of some protein or nucleic acids. It is done with antibodies or hybridization probes that bind only to some molecules of the blot and have an enzyme joined to them.

<u>Dot blot (or slot blot)</u>:-

It is a technique in molecular biology used to detect biomoleclules. It replaces northern, southern or western blot. In dot blot the bimolecular to be detected are not

separated by chromatography. Instead a mixture possibly containing the molecule to be detected applied directly on a membrane as a dot. This is then immediately follower by detection by either nuclcotides probes (northern and southern blot) or antibody (western blot.).

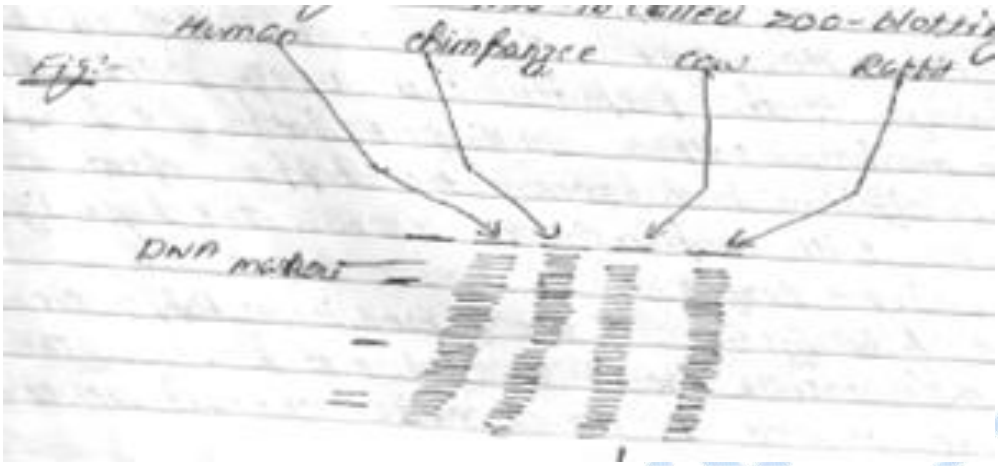
The technique offers significant savings in time as chromatography and complex blotting procedure for the chromatography gel are not required. However it offers no information on the size of the biomolecules. Furthermore, if two molecules of different sizes are detected, they will still appear as a single blot. Dot blot therefore can only confirm the presence or absence of a bimolecule or bimolecules which can be detected by the probes or the antibody.

A radioactive sample can be hybridized to it allowing the researcher to detect variation b/w samples. The DNA is quantified and equal amounts are aliquoted into tubes in excess of the number of its targets in the samples, s/a 10^4 for a (NaOH and 95^oc) and added to the wells where a vacuum sucks the water (with NaoH and NH_4OA_c) from underneath the membrane (nylon or nitro cellulose).

Dot blot analysis:-
It is a simple method of doing a southern, western or northern blot without first separating by size. In case of southern relative physiological activity of the respective population, since it is the product of the number of detected cells and the avg rRNA content. This information on the general activity of a given population should not automatically be regarded. As on indication of a specific kind of activity. Often, one population has the potential to catalyze different transformations-one genotype inlinked to several phenotypes.

(b) Zoo blot:-
A zoo blot (garden blot) is a type of southern blot that demonstrates the similarity b/w specific, usually protein – coding, DNA sequences of different species. A Zoo blot compares animal species while a garden blot compares plant species.

A zoo blot hybridization analysis avoids the problem poorly expressed and tissue specific genes by reaching not for RNAs but for related sequences in the DNAs of other organism. This approach, like homologous searching, is based on the fact that homologous genes in related organism have similar sequences, whereas the interagency DNA is usually quite different. If a DNA fragment from one species is used to probe a southern transfer of DNAs from related species, and one or more hybridization signals are obtained, then it is likely that the probe contain one or more genes. This is called zoo-blotting.



↓ Southern hybridization with human DNA probe.

Probe hybridizes to all DNA samples

<u>Zoo blotting</u>:-

The objective is to determine if a fragment of human DNA hybridizes to DNAs from related species. Samples of human. Chimpanzee. Cow and rabbit DNA are therefore prepared, restricted and electrophoreses in an agrose gel. Southern hybridization is then carried out with a human DNA fragment as the probe. A positive hybridization signal is seen with each of the animal DNA suggesting that the human DNA fragment contains an expressed gene. Note that the hybridizing restriction fragments from the cow and rabbit DNAs are small then the hybridizing fragments in the human and chimpanzee samples. This indicates that the restriction map around the transcribed sequence is different in cows and rabbits, but does not affect the conclusion that a homologous gene is present in all four species.

In order to understand the degree to which a particular gene is similar from species to species, DNA preparations from a set of species is isolated and spread over a surface. The sequence of interest is labeled and allowed to hybridize to the prepared DNA. Usually, the labeled DNA is marked with a radioactive isotope of phosphorus. The hybridization is a process that happens spontaneously DNA pairs with complementary strands. The hybridization however is not perfect.

The hybridization of two strands will happen even when the strands are similar or exact relationships b/w the DNA and other organisms, do the technique takes the advantage of non-exact hybridization. It also allows you to judge the locations of inrons and exons as the latter will be far more conserved than the former.

Q.8 Define OTL? Describe its role in plant breeding?

Ans. **QTL**- Quantitative Trait loci

Definition:-

It is a locus allows that affects a quantitative trait. The plural form of QTL is also abbreviated as QTL.

There are several characters of plant species, among which are traits of agronomic importance, are inherited quantitatively this type of genetic variation is due to multiple factors acting collectively on the expression of a trait. Each of the individual genes of such a polygenic system contributes a small positive or negative effect to the trait of interest. Clear dominance is not exhibited and the phenotype has a large environmental component.

Inheritance- They are inherited in the same way as the genes of major effects. They segregate, recombine and exhibit the linkage theory of association of marker loci with QTLs. To find QTL RFLPs are used, if a correlation b/w the two exist (the quantitative trait of interest anal specific chromosome marked by RELPs) then the chromosome segment must determining the trait must be on that chromosome segments) The difficult part in the procedure is establishing a correlation b/w the trait and specific chromosome segments. The RFLPs markers are easily swred, but the quantitative trait must be scored in a conventional fashion.

RFLP-QTL linkage relationship can be studied in self – pollinated species and among inbred lines in cross-foliated species. Parents that differ for a number of RFLPs as well as mean values of the quantitative trait are first selected. These parents are crossed and the F2 population is obtained. In the F2 population, a dromosonal segment representing linkage b/w a RFLP marker and a Qth will be present in the background of random genetic variation due to independent assortment and recombination. By growing samples of F2 population in different environments, one may also study the linkage under different environments; one may also study the linkage under different environments. In these cases if mean values of a particular quantitative traits are determined in group of plants representative alternative alleles of RFLP, a significant difference in means (of 2 groups) for quantitative traits will indicate a RFLP- QTh linkage. Individual marker alleles can be assigned breeding values circling to the realized effect fo the Qth to which they are linked. The realized effect of a QTh is a function of how large on effect the Qth has, and how highly it is linked to the marker. The selection can then be practices simultaneously for a number of markers that will have the effect of selecting for Qths with a positive effect on the quantitative trait.

QTh Mapping:-

Lander and Botstein (1989) have elaborated a method for mapping QTh using RFLP maps, the technique C/D interval mapping.

Interval mapping assesses the effects of each genome segments located b/w a pair of marker loci, rather than segments the effect of a QTh associated with a single RFLP.

When the estimated phenotypic effect of a single allele substitution at a putative QTh is k/m. Single maximum likelihood equation (MLEs) can be calculated, which maximum the provability that the observes data will have occurred. These MLEs are compared to MLEs obtained under the assumption that no QTh is linked (b=0) , and the evidence that a QTh is existing is indicated by the LOD score (log of score of oclcls)

Multiple Choice Question

Q1. Restriction mapping involves
(i)Restriction digestion of total DNA of organism
(ii) Restriction digestion of single cell DNA
(iii) partial digestion of single cell DNA
(iv)None of these (i)

Q2. FISH stands for
i) Fluorescent in situ hybridization
(ii) Fluorescent intact hybridization
(iii) Fluorescent internal hybridization
(iv)None of these (iii)

Q3. FISH uses chromosomes of which stage of cell division?
i) Metaphase
(ii) Interphase
(iii) Telophase
(iv)Prometaphase (ii)

Q4.FLPter value refers to
(i)Measure of fluorescent signal obtained by FISH relative ti its position to the end of short arm of chromosome
(ii) Measure of all signals of FISH
(iii)Measure of alternate signals of FISH
(iv) All of these (i)

Q5. The resolution of FISH can be improved by
(i)Using interphase chromosomes
(ii)By Fiber FISH
(iii)Both
(iv) None of these (iii)

Q6. STS stands for
(i)Sequence tagged sites
(ii)Simple tagged sites
(iii)Sequence tagged sequences

(iv) Similar tagged sequences (i)

Q7. Which of the following is the most powerful physical technique is
(i)STS
(ii)SNP
(iii)VNTRs
(iv)STRs (i)

Q8. Which of the following is obtained via cDNA clones?
(i)ESTs
(ii)SSLPs
(iii)STS
(iv)All of these (i)

Q9. SNPs are formed as a result of
(i)Point Mutation
(ii)Deletion
(iii)Duplication
(iv)All of these (i)

Q10. GISH stands for
(i) Genomic in situ hybridization
(ii) Genetic in situ hybridization
(iii)Gene in situ hybridization
(iv) None of these (i)

Q11. Which of the following genetic markers are present in highest number within human genome?
(i)RFLPs
(ii) Minisatellites
(iii)Microsatellites
(iv)All of these ()
!
2. QTL stands for
(i)Quantitative trait loci
(ii) Quality trait loci
(iii) Quanta trait loci
(iv)Noneof these (i)

Q13.Which of the following sequences cannot be used as sequence tagged sites?
(i)ESTs
(ii)Random genomic sequence
(iii)SSLPs

(iv)RFLPS (ii)

Q14. CpG Islands are
(i)Areas in genome which are rich in C and G
(ii) Clusters of C with scattered G
(iii) Clusters of G with scattered C
(iv) All of these

Section D

Human Genome Project & Rice Genome Project

Q.1 Describe the important features of human genome project?

Ans. **Introduction:-**

On 26 June, 2000, Successful completion of the sequencing of the Human Genome project was announced jointly by two groups- the Human Genome consortium and Celera Genomics. This was the first draft of human Genome sequence, which is over 3 billion nucleotide long, consisting of 24 chapter viz. 22 autosomes and x and y sex chromosomes.

Important features of the project:-

- ❖ Two major techniques that made the project feasible were the:-

1. Development of YAC Vectors (yeast Artificial chromosomes)
2. Fluorescence – based detection of dideoxy terminated fragments, which enabled near total automation of sequencing.

Methodology:-

The initial impetus for human genetic mapping came from the discovery of RFLP, which were the first highly polymorphic DNA markers to be recognized in animal genomes. In 1987 the first human RFLP map was established comprising 393 RFLPs and 10 additional polymorphic markers. This map was developed from analysis of 21 families, had an average marker density of one per 10mb.

One of the goals that the project set itself was a genetic map with a density of one marker per 1 Mb, although the density of one per 2-5 Mb might be the realistic limit By 1994 an international consortium had met and indeed exceeded the objective via using SSLPs and the large CEPH collection of reference families.

The 1994 map contained 5800 markers, of which over 4000 were SSLM, and had a density of one marker per 0.7 Mb.

In 19900s considerable effect for construction of clone cotig. Maps using STS screening as well as other clone fingerprinting methods was put. The major achievement of the project was publication of a clone contig map of the entire genome, consisting of 33,000 YACs containing fragments with an avg size of 0.9 Mb. Some YACs clone contained chromo. noncontiguous fragments of DNA. The project therefore turned its attention to BACs. A library of 300,000 BAC clones was generated and these clones mapped on to The genome forming a "Sequence –ready" map that would be used as the primary foundation for the sequencing phase of the

project, during which the insert from each BAC would completely sequenced by the shogun method.

Sequencing of the genome:-

Methodology:-

The substrate for sequencing was obtained by an elaborate process involving collecting DNA Samples blood from females and sperms from males form multiple anonymous donors in strict accordance to the international Review board protocol. After screening, a few of the samples (20 -50 individuals) are combined and used for library construction followed for sequencing:

I. Linear sequencing or "top down approach"

The genome is segregated into smaller fragments in a step wise manner, and when the pieces are small enough, thy are cloned into vectors and sequenced, these sections, which contained about 1,50,000 are blown apart by a computer after sequencing, these individual fragments are pieced together by backtracking until their point of origin in a given chromosome is reached. The advantage of this method is that any given fragment of a chromosome can be sequenced desperately and the whole assemble can be done independent of other regions. This was the approach adopted by the human genome sequencing consortium in its initial stages. The first step in this methods in generating genetic mapes by plotting crossover frequencies among different gene loci.

II. Shot-gun sequencing or "bottom up approach"

Chromosomes are divided into sections and then the section of genomic DNA is broken into million of small fragments and all the fragments are sequenced in an unbiased manner. Areas of overlapping DNA are method which formed larger of the degeneracy of this process; the total fragments generated typically cover the genome 10 times to original size. The overlapping of fragments is exploited in the assembly phase by computational exercise.

In either approach the actual sequencing is performed in a manner of dideoxy chain termination method proposed by Sanger in 1977. Most of the sequencing is done by specially designed high-speed sequencers, which require little human interception.

It is now known that 99.8% of the 3.2 billion base pairs between any two human is the same and 0.2% different. For every 500 nucleotides, one nucleotide varies between two individuals. This means that any two individuals differ only in 6 million locations out of 3.2 billion locations.

Q.2 Underline the important features of Rice Genome project?

Ans. **Rice Genome Project:-**

Rice (*Oryza sativa L*) is the leading genomics system among the crop plants. The sequence of the rice genome, the first cereal plant genome, was published in 2005.

Progress in sequencing and annotation of the Genome:-

The international Rice Genome sequencing project (IRGSP) has adopted the clone-by-

clone approach for sequencing rice oryza sativa sp. *Japonica nipponbare* genome sequence because it allow efficient gap-filling, avoids problems arising from distant repetitive sequences and results in the early completion of larger contiguous segments of a genome. A map- based, finished quality sequence of rice *Japonica Niponbare*, which cover 95% of the 389 Mb genome including virtually all of the euchromatin and two complete centromers, was completed in 2005. At the same time, a whole genome shotgun sequencing approach was performed for sequencing an indica variety 93-11 genome.

With the completion of the rice genome sequencing, the Rice Annotation Database (RAP-DB) was created to provide the genome sequence assembly of the IRGSP release, a manually curated annotation of the sequence, and other genomic information that could be useful for comprehensive underst ending of rice biology. The rice Annotation project, 2007, 2008 also, the Rice Genome Annotation Project of the institute for Genomic research (TIGRP) continues to improve the quality of the annotation and to update the rice genome sequence with new data.In the current release 4.0 of the annotation 42,653 non transposable elements related genes encoding 49,472 gene models were identified as a result of the detection of alternative splicing . Moreover, transposon is an important content of the rice genome, which is populated by representatives from all k/n transposon super families. Among them pack-MULEs can mobilize thousands of gene fragments, which may have an impact On rice genome, evolution. The completion and annotation of the rice genome have afforded an unprecedented opportunity for systematic studies of plant gene function. The rice genome sequence provides a complete catalog of genes that are important for improving not only rice but also other cereals, as functionally important sequences are conserved and may be identified by their similarity.

Mapping of the rice –genome:-

(G) Cytogenetic Mapping:-

Genome mapping is an appropriate method for defining the relative positions of different features of genome. This can be done in two ways- by scrutinizing the chromosome visually or by analyzing through experimentation – conventionally, cytogenetic maps are solely based on photomicroscopic examination of chromosomes. In situ hybridization is one of the most effective methods to analyse the essential characteristics of genomes. This is mainly based on the structural constituents of the genome. , It includes two techniques:-

- FISH
- GISH

The techniques had made it possible to generate a quality rice chromosome map. FISH facilitates the mapping of RFLP markers, RDNA loci, Yeast Artificial chromosome (YAC) and Bacterial artificial chromosome (BAC) Clones on rice chromosomes, while GISH provides a unique means to identify or paint a chromosome implement belonging to a specific genome.

GISH has been applied for the determination of the D genome chromosomes in rice, as well as to identify the distribution of the A, B and C genomes in rice

somatic hybrids and also to determine the genomic constitution of some rice tetraploid species.

An important modification of the FISH technique involves the use of highly decondensed nuclei or straightened. DNA fibers for FISH. This method provides a high resolution map as depicted by studies done on the molecular organization of telomere ends. And in the physical mapping of unique nucleotide sequences Or BACs. On the rice genome. But most of these experiments involved the use of mitotic metaphase chromosome. Chang etal (2001) on the other hand utilized meiotic pachytene chromosome – based karyotypes for both cytologist and physical mapping of the rice genome.

FISH and Fiber-FISH have also been used to dissect anatomical features of rice chromosome. Chromosome and they were found to be marked by a satellite repeat and a centromere-specific retrotransposn.

(b) Molecular Genetic Mapping:- Linkage study is a fundamental aspect for plant genetic several rice linkage maps based on genes for morphological and physiological traits have been constructed. In the fast half – century after the twelve linkage group corresponding to the haploid number of rice chromosomes were proposed. The linkage groups have been related to the rice chromosomes by primary trisomics and reciprocal translocations. The centromere positions and the orientation of these linkage groups have also been determined using secondary and telotrisomics, thus giving the correct orienation of the rice linkage map leading to useful resources for genetic studies.

(c) Physical Mapping:-

Physical mapping of the rice genome has received. Almost as much attention as genetic mapping. Various libraries consisting of large insert clones in vectors s/a YAC, PI – derived artificial chromosomes and BAC and the availability of anchored DNA markers from genetic maps greatly facilitated the construction of physical maps.

A rice YAC library consisting of about 7000 YAC clones with an avg insert size of 350 Kb was used for the construction of a physical map of the rice genome. The YAC clones are initially mapped on the rice chromosomes by chromosome landing using the 1383- marker genetic linkage map developed by kurata et al (1994). The information was used for developing the first – generation physical maps of individual rice chromosome.

The YAC map formed 537 contigo and islands equivalent to 216 Mb of the rice genome i.e. 52% of the genome.

The map was updated by saji's etal, who utilized 1439 markers out of the 2275 markers on the genetic map developed by Harushima et al to select 1892 clones from YAC library.

This YAC map, integrated with the genetic map, has been used for positional cloning of several genes, s/a bacterial blight resistance gene, Xa-1, rice blast resistance gene pib and the gibberellins – insensitive dwarfmutant gene, as well as for

the chromosomal assignment of 6591 ESTs to generate a rice transcript map covering 80.8% of the rice genome Additionally, it has been utilized in the IRGSP as a backbone- for construction of a sequence ready physical map using PAG & BACs.

Q.3 Write an explanatory note on "Rice Genome Project current status and Future perspective "?

Ans. The International Rice Genome Sequencing Project (IRGSP) that was started in 1998 and completed by 2005. The work was clone on *Oryza sativa* ssp. *Japonica nipponbare* genome, as it allows efficient gap-filling, avoids problems arising from distant repetitive sequences and results in the early completion of larger contiguous segments of a genome.

Current Status:-

After the completion of *Japonica nippon bare* genome map which covers 95% of the 389 Mb genome, including virtually all of the euchromatin and two complete centromers it is continuously being updated with new data.

Rice Annotation project, 2007 and Rice Annotation project 2008 and Rice Genome annotation project of the Institute for Genomic Research (TIGR) Continues to improve the quality of the annotation and to update the rice genome sequence with new data. In the current release 4.0 of the annotation, 42, 653 non-transposable element-related genes encoding 49,972 genes models were identified as a result of the detection of alternative splicing.

Rice annotation project has facilitated research in comparative genomics and functional genomics. For undertancling rice adaptation and genetic improvement comparative genomic approach is necessary to make a more integrated and detailed map that collects all kinds of genetic variations, which includes copy number variation gene loss caused by frame shift or point mutation and other specific evolutionary events.

After the release of the rice genome sequence, one of the significant challenges has been the large scale identification of gene functions. Various method and technical platforms have been employed to enable high throughput analyses and effective determination of gene functions. Two major platforms of functional genomics studies are generation of a large mutant library and isolation of full length cDNAs.

Future Perspectives:-

Rice genomics obtained a major boost in terms of the sequence drafts that were released by distinct project the data released have provided valuable information on genome structure and organization, It opens up a plethora of opportunities for research related to the life-sciences including evolutionary biology, developmental biology, biochemistry, genetics and molecular biology. The latter, in particular, would be aided with the traditional advantage of many newly developed genomic tools, molecular markers and genomic libraries. These would

be very helpful in gene cloning and functional genomics. Which are essential to gain an insight in to the working of the genome.

Along with the already existing techniques for structural and functional genomic analysis, including gene tagging and microarrays, the use of relatively never tools s/a RNA interference. Could help in the acceleration f progress in the field.
Terada et al have also unveiled the use of a direct method of gene targeting through homologous recombination in rice. The procedure could prove to be useful for obtaining Gene-targeted or knockout lines of rice or even other plants.

Use of rice genomic data has also extended to the in silico identification of regulatory elements. The approach is to identify conserved non-coding sequences based on comparisons b/w arthologous genes from different organism. The data revealed by rice genomics would be a global heritage to understand genetic events not only in this important crop but in related species as well due to the remarkable degree of syntany existing b/w them.

The explosion of information from rice genomics is also expected to have a major positive implication on rice reading especially molecular and marker assisted breeding including OTL breeding analysis.

In rice, various useful genes, such as those for quality improvement in terms of micronutrients, vitamins etc. and for resistance against pests. Diseases, herbicides and biotic stresses, have been transferred. An important e.g. in this context includes the development of "Golden Rice" rice engineered to produce Provitamin A (B-carotenes that could help in alleviating vitamin A efficiency, a major reason for vision impairment.
Recently, an International Rice Functional Genomics consortium (IRFGC) has been formed to provide a platform for sharing materials, integrate databases, seek partnerships', implements cooperative initiatives and accelerate delivery of research results to benefit rice production. Several national programs have also been developed to make use of rice genome information.

There is still a lot to be done in terms of using the knowledge of the rice-genome as a boon for further research in both the basic and the applied fields.

Q.4 Describe the various applications of Rice Genome Project?

Ans. **Introduction:-**

Rice is one of the most important cops for mankind it feeds nearly half the world's population and accounts for more than 50% of their daily calorie intake. Although, in past 30 years, world rice production has doubled due to the introduction of new high yielding varieties and improved cultivation practices, it is still inefficient to cope up with the ever-increasing global demands.

In each conditions rice genome project has opened the ways to developing high yielding new varieties and many such applications.

Rice is an excellent system for plant genomics as it represents a modest size genome of 430 Mb. It feeds more than half the population of the world.

Information related to rice genome is very useful in revealing novel features of macro- and micro-level system of rice genome with other cereals. Microarray analysis is unraveling the identity of rice gene, expressing in temporal and spatial manner and should help target candidate genes useful for improving traits of agronomic importance. Simultaneously, functional analysis of rise genome has been initiated by marker-based characterization of useful genes. And employing functional a knock outs created by mutation or gene tagging.

Applications:-

(1) Applications in comparative genomics:-

Due to the availability of the whole genome sequences of both indicia and japonica subspecies and a bundast genetic and genomic resources including mutants and wild rice species rice has become a model for comparative genome analysis.

This comparative analysis could help to detect polymorphism b/w them and also gives clues to recent genome variations with the available genome sequences and comparative analysis; DNA polymorphisms across the entire rice genome were discerned to develop molecular markers which greatly facilitates gene cloning and also molecular-assisted breeding in rice. Rice is also on important model species of poaceae. Comparative genome analysis of closely related species in cereals would be a powerful tool to identify conserved functional units and regulatory elements. Through comparison with other plant genome sequences and transcript sequences, structural and functional feature of the rice genome itself have been confirmed and improved.

(2) Functional Genomics:-

Rice functional genomics is a scientific approach that seeks to identify and define the function and interaction of genes to produce phenotypic traits, Rapid progress in rice genome sequencing has facilitated research in rice functional genomics. The rice functional genomics researches include development of technical platform, and molecular cloning and functional analysis of agronomic genes. The platform are aimed to at enabling high throughput analyses and effective determination of gene functions, which consist of 3 major components generation and characterization of a large mutant library.

II. Expression profiling of the predicted exon and expressed sequence tags (ESTs) of the entire genome.

III. And isolation of full length (DNA).

After the release of rice genome sequence, one of the significant challenges has been, the large- scale identification of gene functions. Various methods and technical platform have been employed to enable high throughput analysis and effective determination of gene function. Two major platforms of functional genomics studies are generation of a large mutant library and solution of full-length (DNA)

Cloning and functional analysis of agronomics related genes and other:-

The rice genome sequence provides a complete catalog of genes that are very important for identification of the rice genes through map-based cloning strategy. Great progresses have been made by rice researches in recent years in the cloning and functional analysis of agronomic tolerance, disease resistance, grain yield etc.

IV. MicroRNA and Gene Expression:-

Since the completion of the rice genome sequence, various technologies have been used to generate genome-wide exprtechnologies have been used to generate rice transcript one analysis including ESTs, microarrays platform, and massively parallel signature sequencing (MPSS). Till now more then 1,220,877 rice ESTs constructed from various kinds of tissues under various conditions have been sequenced and released in the BCBI-dbEST databases. These are currently the most important resources for transcript one exploration in rice; the experimental evidence greatly improved structural annotation of gene models, and also facilitated the development of whole – genome exon arrays.

Q.5 Describe the various applications of human genome project?

Ans. Introduction:-

Comparative genomics is a field where in 2 complete genome sequences are compared in order to indentify common features that, being conserved.

With human genome sequences comparative genomics has the added value that it may allow the animal versions of human disease genes to be located, having the way for studies of the genetic basis of these diseases using the animal genes as models for the human condition.

Drafts of the mouse and rat genomes were published in 2002, and the chimpanzee draft was completed in 2005. There will also be additional human genome projects aimed at building a catalog of sequence variability in different population, the results possibly enabling the ancient origins of these populations to be inferred.

By using gene sequences development of new drugs and therapies against cancer and other diseases.

Applications:-

The current and potential application of the human genome project are numerous. With the sequence of the majority of human genes now established recent work has been focused on the function of genes and how changes in the sequence relate to health and disease. The description of the genome has led to interest in other "Omes" S/a proteome and the metabolome. More knowledge detailed about genes will further help to diagnose and treat more diseases.

Same current and potential applications of genome research include:-

(a) Molecular Medicine

(b) Energy sources and environmental applications.

(c) Risk assessment

(d) Bioarchaeology, Anthropology, evolution and human migration.

(e) DNA Forensics.

(f) Agriculture, livestock breeding and bio processing.

(B) Molecular Medicine:-

Technology and resources promoted by the human gene projected are starting to have profound impacts on biomedical research and promise to revolutionize the wider spectrum of biological research and clinical medicine. Increasingly detailed genome maps have aided researches seeking genes associated with dozens of genetic conditions, including myotonic dystrophy fragile x syndrome newer fibromatosis types 1 and 2, inherited colon cancer, Alzheimer's disease and familial breast cancer.

On the horizon is a new era of molecular medicine characterized less by treating symptoms and more by looking to the most fundamental causes of disease. Rapid and more specific diagnostic tests will make possible researchers earlier treatment of countless maladies. Medical researchers also will be able to devise novel therapeutic regimes based on new closes of drugs, immune therapy techniques, disease and possible augmentation or even replacement of defective genes though gene therapy.

(b) Energy and Environmental applications:-

In 1994, taking advantage of new capabilities developed by the genome project, DOE initiated the Microbial Genome program to sequence the genomes of bacteria useful in energy production, environmental remediation, toxic waste reduction, and industrial Processing. A follow on program GTL builds on data and resources from the human genome project, the Microbial Genome programs and systems biology. GTL will accelerate understanding of dynamic living systems for solutions to DOE mission challenges in energy and the environment.

Till now only 0.01% of all microbes have been cultivated and characterized, microbial genome sequencing will help and characterized. Microbial Genome Sequencing will help lay a foundation for knowledge that will ultimately benefit human health and the environment. The economy will benefit from further industrial applications of microbial capabilities.

Microbial genomics will also help pharmaceutical researchers gain a letter understanding of how pathogenic microbes cause disease sequencing these microbes will help several vulnerabilities and identify new drug targets.

(c) Risk Assessment:-

Understanding the human genome will have on enormous impact on the ability to assess risks posed to individuals by exposure to toxic agents; Scientists know that genetic differences make some people more susceptible and others more resistant to such agents. Far more work must be done to determine the genetic basis of such variability.

(d) Bioarchaeology, Anthropology, Evolution and Human migration:-

Understanding genomics will help us understand human evolution and the common biology we share with all life. Comparative genomics b/w humans and other organism s/mice already have led to similar genes associated with diseases and traits. Further comparative studies will help determine the yet-unknown function of thousands of other genes.

Comparing the DNA sequences of entire genomes of different microbes will provide new insights about relationships among the three kingdoms of life; archaebocteria' eukaryotes and prokaryotes.

(e) DNA Forensics (Identification)

Any type of organism can be identified by examination of DNA sequence unique to that species. Identifying individuals is less precise, although when DNA sequencing technologies progress further, direct characterization of very large DNA segments, and possibly even whole genomes, will become feasible and practical and will allow precise individual identification.

To identify individuals, forensic scientists scan about 10 DNA regions that vary from person to person and use the data to create a DNA profile of that individual sometimes c/d a DNA fingerprints. There is an extremely small chance that another person has the same DNA Profile for a particular set of regions.

(f) Agriculture, livestock Breeding and Bio processing:-

Understandings' plant and animal genome will allow us to create stronger, more disease- resistant plants and animals reducing the cost of agriculture and providing consumer with more nutritious, pesticide free foods.

Alternate uses for crops S/a tobacco have been found one researcher has genetically engineered tobacco plants in his laboratoryto produce a bacterial enzyme that breaks down explosives s/a TNT and dinitroglycerin.Waste that would take centuries to breakdown in the soil can be cleaned up by simply growing these special plants in the polluted area.

Multiple Choice Question

Q1. When does Rice genome project completed?
(i)1992 (ii)1900 (iii)1998 (iv) 1953 (iii)

Q2. Which of the cereal crops has the smallest genome?
(i)Wheat (ii) Rice (iii) Barley (iv) Maize (iv)

Q3.The size of rice genome is?
(i) 430Mb (ii) 500Mb (iii)590Mb (iv) 523Mb (i)

Q4. The variety of rice used for sequencing?
(i) Japonica (ii) Aizon (iii) Peta (iv) Carolino (i)

Q5.Launching of rice genome project took place at
(i) Japan (ii) China (iii) India(iv) Singapore (iv)

Q6.IRGSP stands for
(i) International rice genome sequencing project
(ii) Indian rice genome sequencing project
(iii) International rice genome sequenator project
(iv)International rice genome sequencing process (i)

Q7. HGP stands for
(i)Human Genome Project
(ii)Heterogeneous Gene Project
(iii)Homogeneous Gene Project
(iv) None of these (i)

Q8.The aim of HGP is
(i) To identify and map 20,000-25,000 genes of humans
(ii) To determine chemical base pairs of DNA of humans
(iii) Nucleotides contained in a human haploid reference genome
(iv)All of these (iv)

Q9.HGP started in?
(i)1992 (ii)1990 (iii)1998 (iv) 1953 (ii)

Q10.NIH stands for?
(i) National Institutes of Health
(ii)National Human Genome Research Institute
(iii) National Institutes of Hospitality
(iv) None of these (i)

Q11.Working draft of human genome was announced in

(i)2003 (ii)1990 (iii)1998 (iv) 2000 (iv)

Q12. Complete draft of human genome was announced in
(i)2003 (ii)1990 (iii)1998 (iv) 2000 (i)

Q13.Mapping the human genome involves
(i) Sequencing multiple variations of each gene
(ii) To study the entire DNA found in human cells
(iii)To study heterochromatic areas
(iv)All of these (i)

Key Terms

Allele -- an alternative form of a gene; any one of several mutational forms of a gene.

Amplification -- any process by which specific DNA sequences are replicated disproportionately greater than their representation in the parent molecules.

Aneuploidy -- state of having variant chromosome number (too many or too few). (i.e. Down syndrome, Turner syndrome).

Autosome -- a nuclear chromosome other than the X- and Y-chromosomes.

Autoradiograph -- a photographic picture showing the position of radioactive substances in tissues.

Bacteriophage -- a virus whose host is a bacterium; commonly called phage.

Barr body -- the condensed single X-chromosome seen in the nuclei of somatic cells of female mammals. base pair a pair of hydrogen-bonded nitrogenous bases (one purine and one pyrimidine) that join the component strands of the DNA double helix.

Base sequence -- a partnership of organic bases found in DNA and RNA; adenine forms a base pair with thymine (or uracil) and guanine with cytosine in a double-stranded nucleic acid molecule.

cDNA -- complementary DNA produced from a RNA template by the action of RNA-dependent DNA polymerase.

Centromere -- a region of a chromosome to which spindle traction fibers attach during mitosis and meiosis; the position of the centromere determines whether the chromosome is considered an acrocentric, metacentric or telomeric chromosome.

Chromosome -- in the eukaryotic nucleus, one of the threadlike structures consisting of chromatin and carry genetic information arranged in a linear sequence.

Chromosome banding -- a technique for staining chromosomes so that bands appear in a unique pattern particular to the chromosome.

Codon -- a sequence of three nucleotides in mRNA that specifies an amino acid.

Consanguinity -- genetic relationship. Consanguineous individuals have at least one common ancestor in the preceding few generations.

Conservative change -- an amino acid change that does not affect significantly the function of the protein.

Contiguous genes -- genes physically close on a chromosome that when acting together express a phenotype.

Cosmids -- plasmid vectors designed for cloning large fragments of eukaryotic DNA; the vector is a plasmid into which phage lambda cohesive end sites have been inserted.

CpG islands -- areas of multiple CG repeats in DNA.

Crossovers -- the exchange of genetic material between two paired chromosome during meiosis.

Cytogenetics -- the study of chromosomes.

Degenerate codon -- a codon that specifies the same amino acid as another codon.

Deletion -- the loss of a segment of the genetic material from a chromosome.

Deletion mapping -- the use of overlapping deletions to localize the position of an unknown gene on a chromosome or linkage map.

DNA fingerprint technique -- a method employed to determine differences in amino acid sequences between related proteins; relies upon the presence of a simple tandem-repetitive sequences that are scattered throughout the human genome.

DNA hybridization -- a technique for selectively binding specific segments of single-stranded (ss) DNA or RNA by base pairing to complementary sequences on ssDNA molecules that are trapped on a nitrocellulose filter.

DNA probe -- any biochemical used to identify or isolate a gene, a gene product, or a protein.

DNA sequencing -- "plus and minus" or "primed synthesis" method, developed by Sanger, DNA is synthesized in vitro in such a way that it is radioactively labeled and the reaction terminates specifically at the position corresponding to a given base; the "chemical" method, ssDNA is subjected to several chemical cleavage protocols that selectively make breaks on one side of a particular base.

Dominant -- alleles that determine the phenotype displayed in a heterozygote with another (recessive) allele.

Endonuclease -- an enzyme that breaks the internal phosphodiester bonds in a DNA molecule.

Euchromatin -- the chromatin that shows the staining behavior characteristic of the majority of the chromosomal complement.

Eugenics -- the improvement of humanity by altering its genetic composition by encouraging breeding of those presumed to have desirable genes.

Exons -- portion of a gene included in the transcript of a gene and survives processing of the RNA in the cell nucleus to become part of a spliced messenger of a structural RNA in the cell cytoplasm; an exon specifies the amino acid sequence of a portion of the complete polypeptide.

FISH -- florescent in situ hybridization: a technique for uniquely identifying whole chromosomes or parts of chromosomes using florescent tagged DNA.
5' - end -- the end of a polynucleotide with a free (or phosphorylated or capped) 5' - hydroxyl group; transcription/translation begins at this end.

Gamete -- an haploid cell.gel electrophoresis the process by which nucleic acids (DNA or RNA) or proteins are separated by size according to movement of the charged molecules in an electrical field.

Gene -- a hereditary unit that occupies a certain position on a chromosome; a unit that has one or more specific effects on the phenotype, and can mutate to various allelic forms.

Gene amplification -- any process by which specific DNA sequences are replicated disproportionately greater than their representation in the parent molecules; during development, some genes become amplified in specific tissues.

Gene map -- the linear arrangement of mutable sites on a chromosome as deduced from genetic recombination experiments.

Gene therapy -- addition of a functional gene or group of genes to a cell by gene insertion to correct an hereditary disease.

Genetic counseling -- the educational process that helps individuals, couples, or families to understand genetic information and issues that may have an impact on them.

Genetic linkage map -- a chromosome map showing the relative positions of the known genes on the chromosomes of a given species.

Genetic screening -- testing groups of individuals to identify defective genes capable of causing hereditary conditions.

Genetic variation -- a phenotypic variance of a trait in a population attributed to genetic heterogeneity.

Genome -- all of the genes carried by a single gamete; the DNA content of an individual, which includes all 44 autosomes, 2 sex chromosomes, and the mitochondrial DNA.

Genotype -- genetic constitution of an organism.

Germ cell -- a sex cell or gamete (egg or spermatozoan).Haldane equation Haldane's law: the generalization that if first generation hybrids are produced between two species, but one sex is absent, rare, or sterile, that sex is the heterogamic sex.

Heterozygote -- having two alleles that are different for a given gene.

Hemophilia -- a sex-linked disease in humans in which the blood-clotting process is defective.

Heterogeneity -- the production of identical or similar phenotypes by different genetic mechanisms.

Homologous chromosomes -- chromosomes that pair during meiosis; each homologue is a duplicate of one chromosome from each parent.

Homozygote -- having identical alleles at one or more loci in homologous chromosome segments.

Housekeeping genes -- those genes expressed in all cells because they provide functions needed for sustenance of all cell types.

HUGO -- Human Genome Organization.

Huntington disease -- a disease characterized by irregular, spasmodic involuntary movements of the limbs and facial muscles, mental deterioration and death, usually within 20 years of the onset of symptoms.

Hybridization -- the pairing of a single-stranded, labeled probe (usually DNA) to its complementary sequence.

Imprinting -- a chemical modification of a gene allele which can be used to identify maternal or paternal origin of chromosome.

Incomplete penetrance -- the gene for a condition is present, but not obviously expressed in all individuals in a family with the gene.

In situ hybridization -- hybridization of a labeled probe to its complementary sequence within intact, banded chromosomes.

Introns -- a segment of DNA (between exons) that is transcribed into nuclear RNA, but are removed in the subsequent processing into mRNA.

Isochromosome -- a metacentric chromosome produced during mitosis or meiosis when the centromere splits transversely instead of longitudinally; the arms of such chromosome are equal in length and genetically identical, however, the loci are positioned in reverse sequence in the two arms.

Klinefelter syndrome -- an endocrine condition caused by a an extra X-chromosome (47,XXY); characterized by the lack of normal sexual development and testosterone, leading to infertility and adjustment problems if not detected and treated early.

Karyotype -- a set of photographed, banded chromosomes arranged in order from largest to smallest.

Lligase -- an enzyme that functions in DNA repair.

Linkage -- the greater association in inheritance of two or more nonallelic genes than is to be expected from independent assortment; genes are linked because they reside on the same chromosome.

Linkage -- analysis of pedigree the tracking of a gene through a family by following the inheritance of a (closely associated) gene or trait and a DNA marker.

Lod score -- logarithm of the odd score; a measure of the likelihood of two loci being within a measurable distance of each other.

Marker -- a gene with a known location on a chromosome and a clear-cut phenotype, used as a point of reference when mapping a new mutant.

Methylation -- addition of a methyl group (-CH3) to DNA or RNA.

mRNA -- messenger RNA; an RNA molecular that functions during translation to specify the sequence of amino acids in a nascent polypeptide.

Multifactorial -- a characteristic influenced in its expression by many factors, both genetic and environmental.

Mutation -- process by which genes undergo a structural change.

Nonsense mutation -- a mutation in which a codon is changed to a stop codon, resulting in a truncated protein product.

Northern analysis -- a technique for transferring electrophoretically resolved RNA segments from an agarose gel to a nitrocellulose filter paper sheet via capillary action.

Nucleotide -- one of the monomeric units from which DNA or RNA polymers are constructed; consists of a purine or pyrimidine base, a pentose sugar and a phosphoric acid group.

PCR -- polymerase chain reaction; a technique for copying the complementary strands of a target DNA molecule simultaneously for a series of cycles until the desired amount is obtained.

Pedigree -- a diagram of the heredity of a particular trait through many generations of a family.

Phenotype -- observable characteristics of an organism produced by the organism's genotype interacting with the environment.

Physical map -- map where the distance between markers is the actual distance, such as the number of base pairs.

Pleiotropy -- the phenomenon of variable phenotypes for a number of distinct and seemingly unrelated phenotypic effects.

Polymerase -- any enzyme that catalyzes the formation of DNA or RNA from deoxyribonucleotides or ribonucleotides.

Predisposition -- to have a tendency or inclination towards something in advance.

Primer -- nucleotides used in the polymerase chain reaction to initiate DNA synthesis at a particular location.

Probability -- the long term frequency of an event relative to all alternative events, and usually expressed as decimal fraction.

Proband -- individual in a family who brought the family to medical attention.

Probe -- single-stranded DNA labeled with radioactive isotopes or tagged in other ways for ease in identification.

Prognosis -- prediction of the course and probable outcome of a disease.

Recessive -- a gene that is phenotypically manifest in the homozygous state but is masked in the presence of a dominant allele.

Recombination -- the natural process of breaking and rejoining DNA strands to produce new combinations of genes and, thus, generate genetic variation. Gene crossover during meiosis.

Repeat sequences -- the length of a nucleotide sequence that is repeated in a tandem cluster.

RFLP -- restriction fragment length polymorphism; variations occurring within a species in the length of DNA fragments generated by a species endonuclease.

Ribosomal protein -- one of the ribonucleoprotein particles that are the sites of translation.

Sanger sequence -- "plus and minus" or "primed synthesis" method; DNA is synthesized so it is radioactively labeled and the reaction terminates specifically at the position corresponding to a given base.

Selection -- the process of determining the relative share allotted individuals of different genotypes in the propagation of a population; the selective effect of a gene can be defined by the probability that carriers of the gene will reproduce.

Sex determination -- the mechanism in a given species by which sex is determined; in many species sex is determined at fertilization by the nature of the sperm that fertilizes the egg.

Sickle cell anemia -- an hereditary, chronic form of hemolytic anemia characterized by breakdown of the red blood cells; red blood cells undergo a reversible alteration in shape when the oxygen tension of the plasma falls slightly and a sickle-like shape forms.

Southern blotting -- a technique for transferring electrophoretically resolved DNA segments from an agarose gel to a nitrocellulose filter paper sheet via capillary action; the DNA segment of interest is probed with a radioactive, complementary nucleic acid, and its position is determined by autoradiography.

Syndrome -- a recognizable pattern or group of multiple signs, symptoms or malformations that characterize a particular condition; syndromes are thought to arise from a common origin and result from more than one developmental error during fetal growth.

Trait -- any detectable phenotypic property of an organism.

Transduction -- the transfer of bacterial genetic material from one bacterium to another using a phage as a vector.

Transferase -- enzymes that catalyze the transfer of functional groups between donor and acceptor molecules.

Transcription -- the formation of an RNA molecule upon a DNA template by complementary base pairing.

Translation -- the formation of a polypeptide chain in the specific amino acid sequence directed by the genetic information carried by mRNA.

Translocation -- a chromosome aberration which results in a change in position of a chromosomal segment within the genome, but does not change the total number of genes present.

Triplet code -- a code in which a given amino acid is specified by a set of three nucleotides.

VNTR -- variable number tandem repeats; any gene whose alleles contain different numbers of tandemly repeated oligonucleotide sequences.

Vector -- a self-replicating DNA molecule that transfers a DNA segment between host cells.

Western blotting analysis -- a technique used to identify a specific protein; the probe is a radioactively labeled antibody raised against the protein in question.

YAC -- yeast artificial chromosome; a linear vector into which a large fragment of DNA can be inserted; the development of YAC's in 1987 has increased the number of nucleotides which can be cloned.

Zoo blot -- northern analysis of mRNA from different organisms.

B.Sc./M.Sc. (Part III) Examination, 2011

(FACULTY OF SCIENCE)

(Common to Three and Five Year Integrated Course)

BIOTECHNOLOGY

PAPER BT-701

GENOME ANALYSIS

Year-2011

Time.: 3 Hours

Max.M.:50 Question No. 1 is compulsory. Attempt FIVE questions in all, selecting ONE question from each section.

1. Answer the following questions briefly:

   (i) Define gene.
   (ii) What is C-value Paradox?
   (iii) What is Pedigree analysis?
   (iv) Define linkage.
   (v) What do you understand by Zoo Blot?
   (vi) Write the full form of FISH and GISH.
   (vii) What do you understand by RACE?
   (viii) What is Exon?
   (ix) Rice Genome Project was started in which year and got completed in which year?
   (x) What was the main goal of Human Genome Project? 1 x 10 = 10

Section-A

2. What do you understand about genome? Write an account on structure of prokaryotic genome. 10

   or

   Write an account on linkage analysis in *Drosophila*.

Section-B

3. What do you mean by genetic mapping? Write an account on various tools and procedure of genetic mapping of complex characters of an individual. 2 + 8

or

Write an account on multiple families and repetitive DNA in human genome. 10

Section-C

4. Write short notes on:
   (i) Chromosome Walking
   (ii) Chromosome Painting 5 + 5

or

   (i) RFLP and RAPD
   (ii) Exon trapping and Sequence analysis 5 + 5

5. Write an account on Rice Genome Project and its application. 10

or

Write an account on Human Genome Project and its application. 10

****

Bibliography

1. Genome 3 by T.A. Brown
2. Molecular biology by David Friefelder
3. The Cell : a molecular approach by Geoffrey M. Cooper
4. Encyclopedia of Genetics by Reeve
5. Cell Biology, Genetics, Molecular Biology by P.S. Verma

Websites:
1. www.kbiotech.com
2. www.google.com
3. www.bio-link.org
4. www.ranker.com
5. www.ask.com