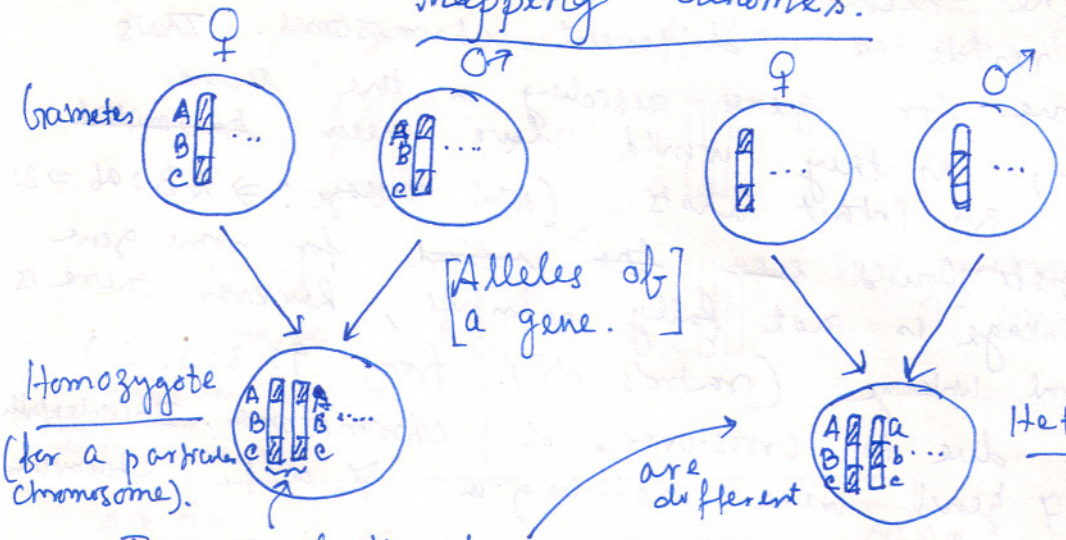
Mapping Genomes.



Two copies of the chromosome are same ⇒

Alleles : ~~Alternative~~ / Polymorphic ~~copy~~ existence of a gene — responsible for ~~diff~~ variation in phenotypes.

The genetic map: records the order of genes & the approx. dist. between them on their respective chromosome.

Mendel's observation: (Experimentation on pea plants, observing seven different characteristics).

(a) Alleles segregate randomly: (Mendel's first law)

$$Aa \times Aa$$

$$\Downarrow$$

phenotypes: $\overset{A}{AA}$ $\overset{A}{Aa}$ $\overset{A}{Aa}$ $\overset{a}{aa}$ ⇒ 3 : 1

⇒ Dominant & recessive alleles. However, it may be incomplete dominance (in between dominant & recessive phenotype), or codominance (both present equally, e.g. red flower crossed with white flower yields pink flower.

(b) Pairs of alleles segregate independently: (Mendel's)

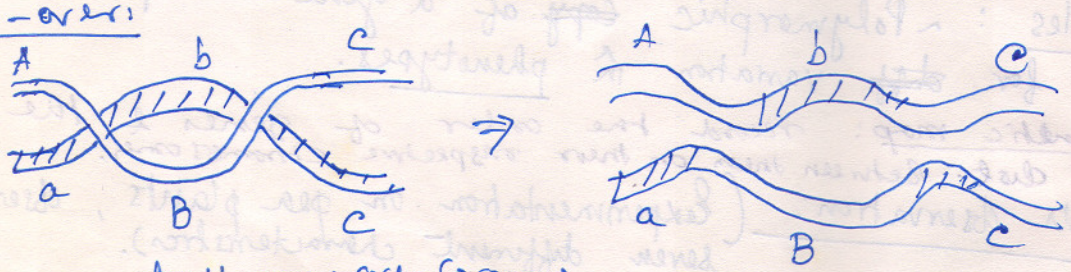
Crossing of two dihetero zygotes ⇒ double heterozygote $Aa\,Bb \times Aa\,Bb$

Phenotypes: AB Ab aB ab

⇒ 9 : 3 : 3 : 1

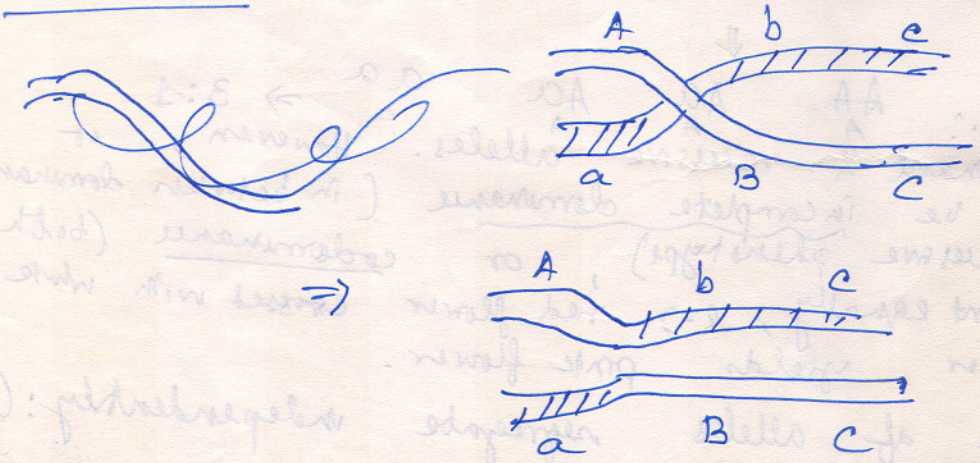
However, the second law is true if a pair of genes reside in different chromosomes. This is not true for genes residing in the same chromosome, as they would have been inherited as intact units (total linkage: ⇒ AB : ab ⇒ 3 : 1). Biologists found even the second for some gene total linkage is not fully maintained, however there is a partial linkage (ratio's diff. from 9 : 3 : 3 : 1) which is due to cross-over of chromosomes. Incidentally Mendel's 7 genes were residing in 7 diff. chromosomes of pea-plant.

Cross-over:



double cross over (more)
(less probable)

Single Cross over.

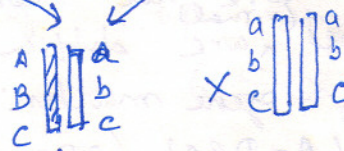


⇒ Closer the genes greater is the prob. for being inherited both in the same form (phenotype characteristics).

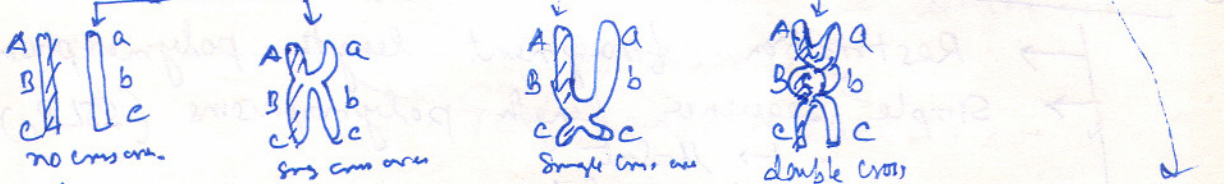
⇒ Independent segregation indicates genes are in different chromosom.

Recombination frequencies from a test-cross.

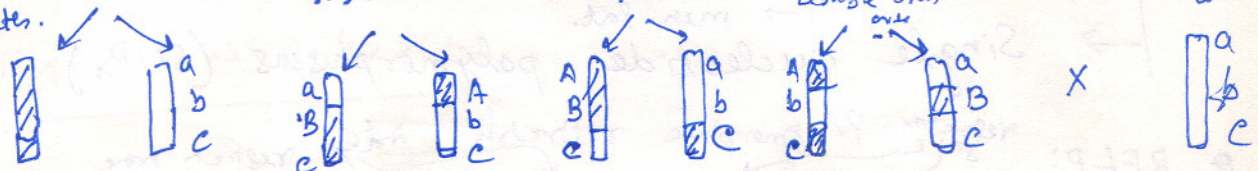
Pure genotypes are used for breeding hybrid genotypes.



Gametes due to no-crossover, single crossover & double crossover



Gametes.



Phenotype ABC abc aBc Abc ABc abC AbC aBC

[only crossovers are reflected in phenotype due to recessive characteristics of the pure gamete]

Let $n(\cdot) \Rightarrow$ no. of a phenotype.

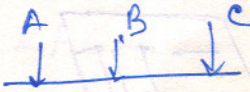
Recombination Frequency: Frequency of separation (in %) of two pure genes in this test cross. Further the gene, greater is the separation. $\Rightarrow$ (no. cross over)

Unit of recomb. freq. is centimorgans.

A centimorgan corresponds to a rf of 1% one in hundred

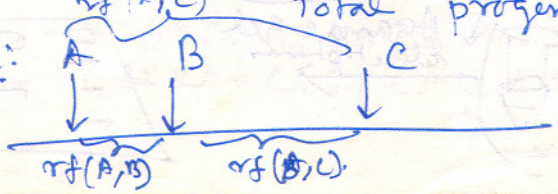
$rf(A,B) < rf(A,C)$

$rf(B,C) < rf(A,C)$

$\Rightarrow$ 

$$rf(A,B) = \frac{n\{aBC, AbC, AbC, aBC\}}{\text{Total progenies observed.}}$$

These are progenies in which A & B are not observed together (separated)

Partial linkage: 

[genetic mapping method]

However, there are hotspots in chromosomes for cross-over, making certain recombinations favoured.

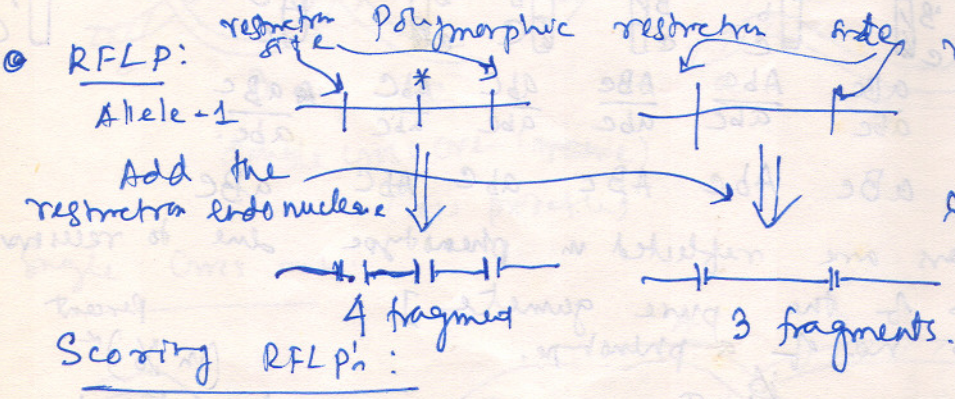
(#) Other types of markers for genetic mapping:

Multiallelic genes are not common and making observations on phenotypes (visuals) are difficult in many cases.
Some other examples of gene marker:
blood groups, HLA-DRB1 (290 alleles), HLA-B (400)

(@) DNA markers:

- Restriction fragment length polymorphisms (RFLP)
- Simple sequence length polymorphisms (SSLPs)
  - Micro-Sat.
  - mini-Sat.
- Single nucleotide polymorphisms (SNPs)

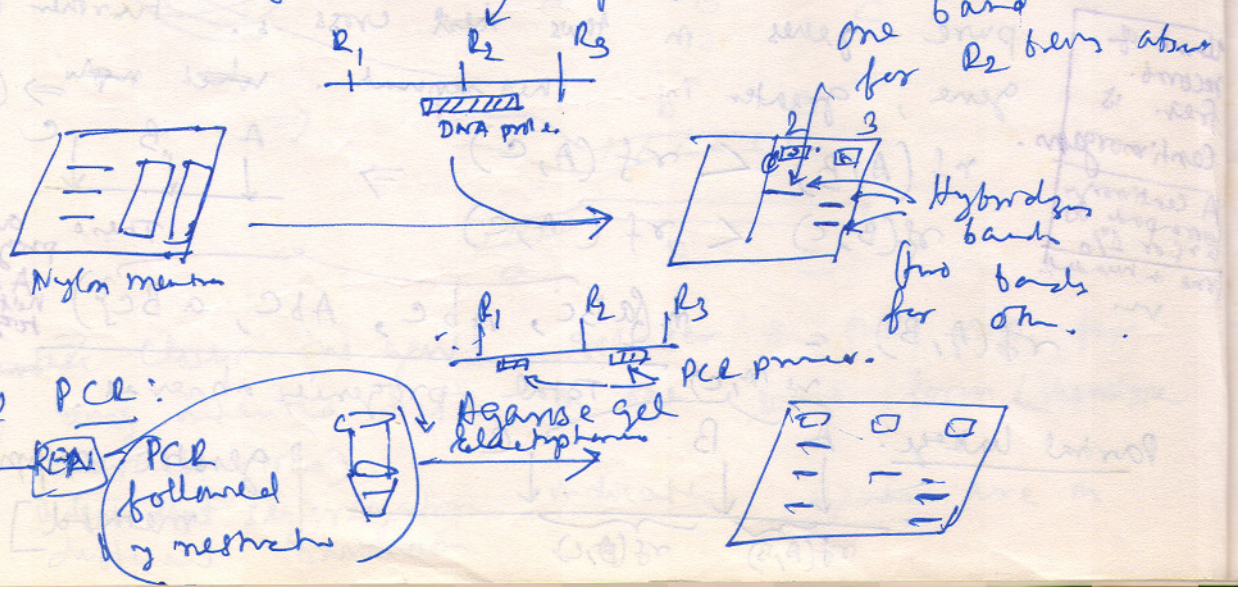
(@) RFLP:



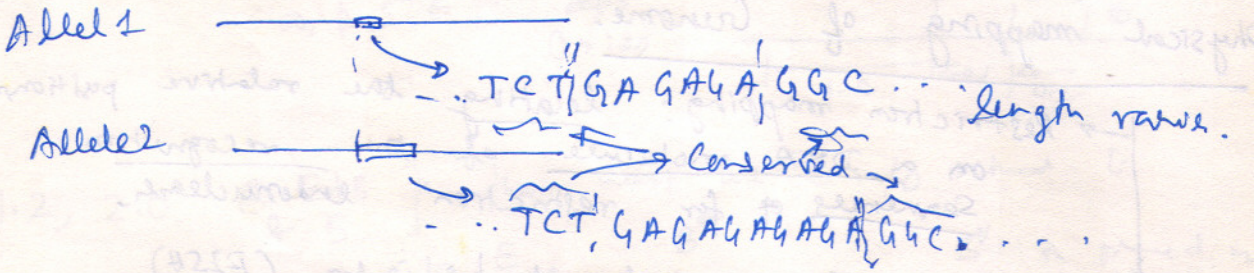
$\approx 10^5$ RFLP sites in human genome.

Scoring RFLPs:

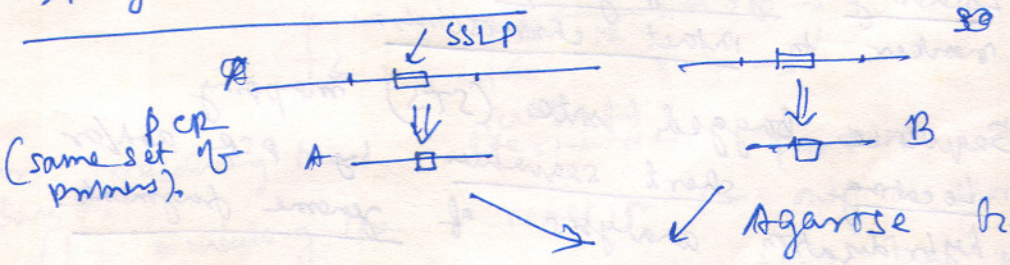
(A) Southern hybridization:



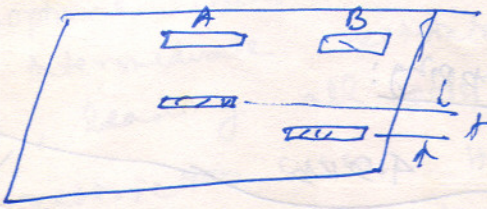
- **SSLP's:**
  - Mini satellite or Variable no of tandem repeats (VNTR). ⇒ usually seen in Telomere (not suitable for genome mapping as it does not provide genome wide marker distribution)
  - Microsatellites are more suitable: typically 10–30 copies of 4 bp length ≲ 300 bp have suitable for PCR. ⇒ 6.5×10^5 μ satellite.



Typing SSLP by PCR:



↙ Agarose Gel Electrophoresis



- **SNP:** Usually SNP. have two alleles (due to mutation & natural selection)
  There are 1.42×10^6 SNP in human genome of which 100,000 results into RFLP.
- Scoring: highly stringent hybridization (not allows a single mismatch of a base pair).

Oligonucleotide hybridized strategies

5′ - - CTGTCGTCAGT [C] TTTAGTT .. 3′
— GACAGCAGTCA [G] AAATCAA - -
↑ SNP.

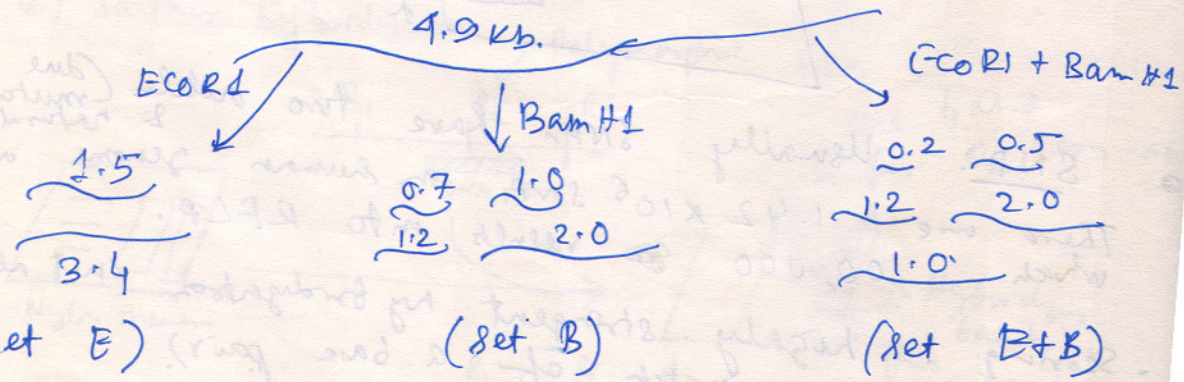
stringent hybridization takes place just below the melting temp. ($T_m - 5°C < T < T_m$)

T_m = (4 × no. of G & C nucleotides + 2 × no. of A & T nucleotides) for 15–30 nucleotides in length.

(#) Physical mapping of Genome:

- → Restriction mapping: locating the relative positions on a DNA molecule of the recognition sequences of restriction endonucleases.
- → Fluorescent in situ hybridization (FISH) Locating hybridizing probes containing the marker to intact chromosomes.
- → Sequence tagged site (STS) mapping Locating short sequences by PCR and/or hybridization analysis of genome fragments.

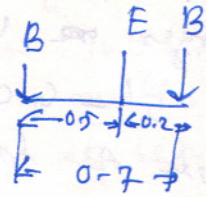
(*) Restriction mapping:



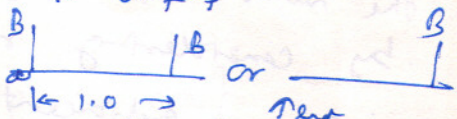
Interpretation of the double restriction:

Fragments.

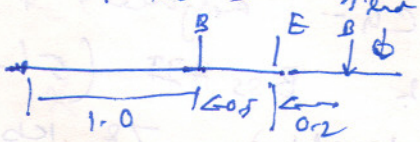
0.2, 0.5 → As, 0.7 = 0.2 + 0.5 ∈ B



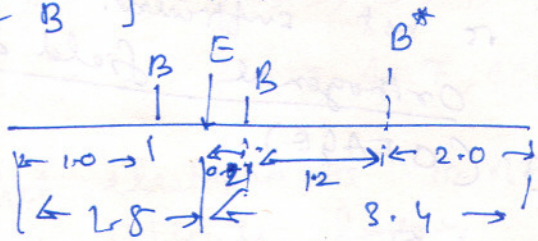
1.0 ⇒ 1.0 ∈ B ⇒



as 0.5 ∈ E ⇒ (only 1 by)

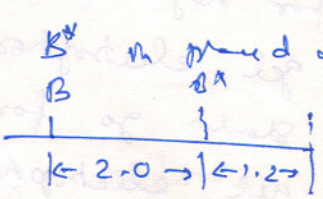


1.2, 2.0 ∈ E & B } they must be w/o mutual [illegible]



(alt. 1)

or



(alt. 2)

For resolving this (Sam H1) alternatives partial restriction (by incubating the reaction for only a short time or using a suboptimal incubation temp. In partial restriction intermediate restriction sites may not be fragmented, leading all possible lengths between a pair of restriction sites. However, to remove some of them, only lengths from ends of the starting DNA molecule could be ~~cleaved~~ filtered by attaching a radioactive or other type of marker at the ends & screening the labeled fragment bg from agarose gel.)

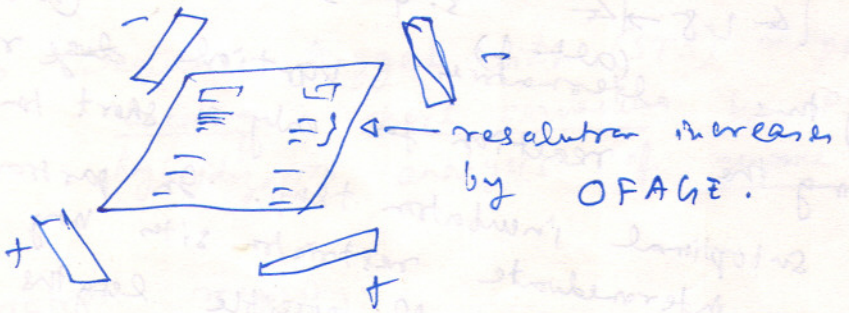
4.9 Kbs.
Run H1

1.0, 1.7, 3.7, 0.7, 2.7, 3.9, 4.9, 2.0, 3.2, 1.2

Presence of 2.7 indicates (alt. 2) is the mapped configuration.

Restriction mapping is applicable for smaller molecules. However by considering restriction site of larger length (of recog. seq. e.g. ApaI 5'-GGCTCTTC-3' & SgfI 5'-GCGATCGC-3', the number of fragmentation could be minimized, or by considering rare restriction sites (containing -CG- adjacent pair in human DNA, eg. SmaI (5'-CCCGGG-3') & BssHII (5'-GCGCGC-3') (once in every 10000 bp, 78 Kb & 390 Kb respectively).

For higher molecular weight, resolution of ordinary gel electrophoresis is not sufficient. Hence one should go for Orthogonal field alternation gel electrophoresis (OFAGE).



(III) Computational problem

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a ~~multiset~~ set of n integers (allowing duplicate elements) in an increasing order & ΔX denotes the multiset (of all $\binom{n}{2}$ pairwise distances between points in X i.e.

$$\Delta X = \{x_j - x_i : 1 \le i \le j \le n\}$$

Given ΔX compute X.

e.g. $\Delta X = \{2, 2, 3, 3, 4, 5, 6, 7, 8, 10\}$

Compute $X = \{0, 2, 4, 7, 10\}$

N.B. In Gel electrophoresis technique of finding multiplicity of length is difficult, however with lot of work it is possible to determine experimentally.

Let $\Delta X = L$

Let $M = \max\{\Delta X\}$

Exhaustive search technique:

Brute Force Algo.:

$x_1 = 0$; $x_n = M$

$\therefore$ for every $(n-2)$ integers in ascending order (between 0 & M) check if $\Delta X = L$ is obtained

$\therefore$ $\binom{M-1}{n-2}$ possible ~~rods~~ elements to be checked $\Rightarrow \simeq O(M^{n-2})$.

Another Brute Force Algo.

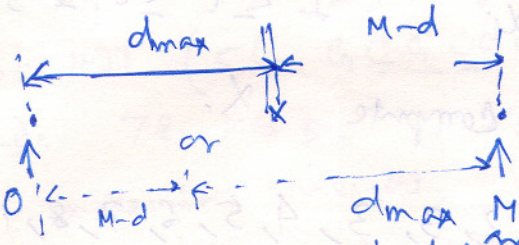
$x_1 = 0$, $x_n = M$

chose x_i's only from ΔX specified $\Delta X = L$ & perform the same.

No. of elements: $\binom{|L|}{n-2} \approx O(|L|^{n-2})$.

or $|L| = \frac{n(n-1)}{2} \Rightarrow O(n^{2n-4})$

Practical Algo.



$X_0 = \{0, M\}$

$L_0 = \{\}$

$= L - \{0, M\}$

Iteratively place the element with highest d_f (or $M - d_f$) in the set L_i w.r.t. either (i) '0' or (ii) 'M'. Compute distances for each case with X_i, if they are elements of L_i, include the element to d_f & delete those distances from L_i. However, if it fails for both the cases, then you need to backtrack & choose the other option (locally works in some cases).

Worst Case: (for one particular case)

$T(n) = 2T(n-1) + O(n) \Rightarrow$ exponential

$\Rightarrow \Theta(n^2 \log n)$

Though on the average it works fine.

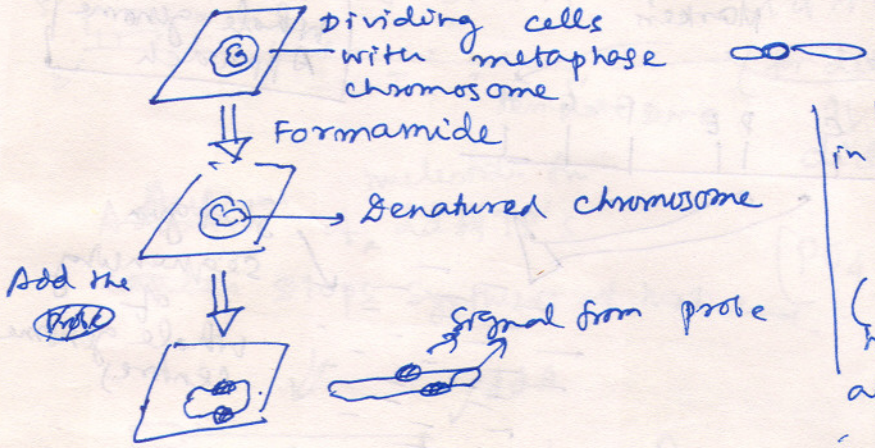
— Polynomial problem solved by M. Nivat et. al (2002)

Q. 1. Suppose adjacent distance & distance from end points are given? How do you modify the algo.?

* Direct examination of DNA molecules for restriction sites:

→ Optical mapping through Gel Stretching / Molecular Combing (Producing DNA glass slides)

(*) Fluorescent in situ hybridization (FISH)

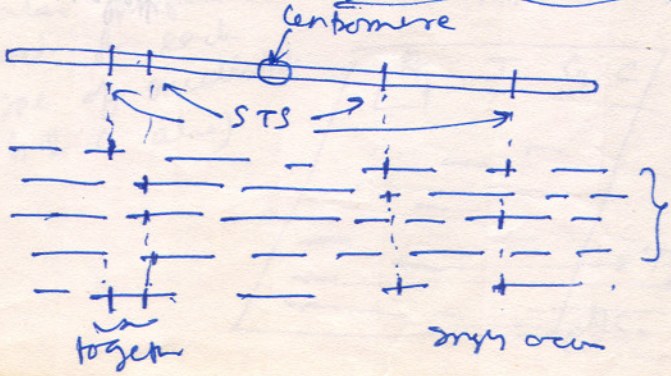


Repeat sequences in the probe are now blocked, before hybridization. (Otherwise probe may hybridize randomly at repeat sequences).

(*) Sequence tagged site (STS) mapping.

A sequence tagged site (STS) is simply a short DNA sequence, generally between 100 and 500 bp in length, that is easily recognizable and occurs only once in the chromosome or genome being studied.

To map a set of STS's, a collection of overlapping DNA fragments from a single chromosome or from the entire genome is needed.



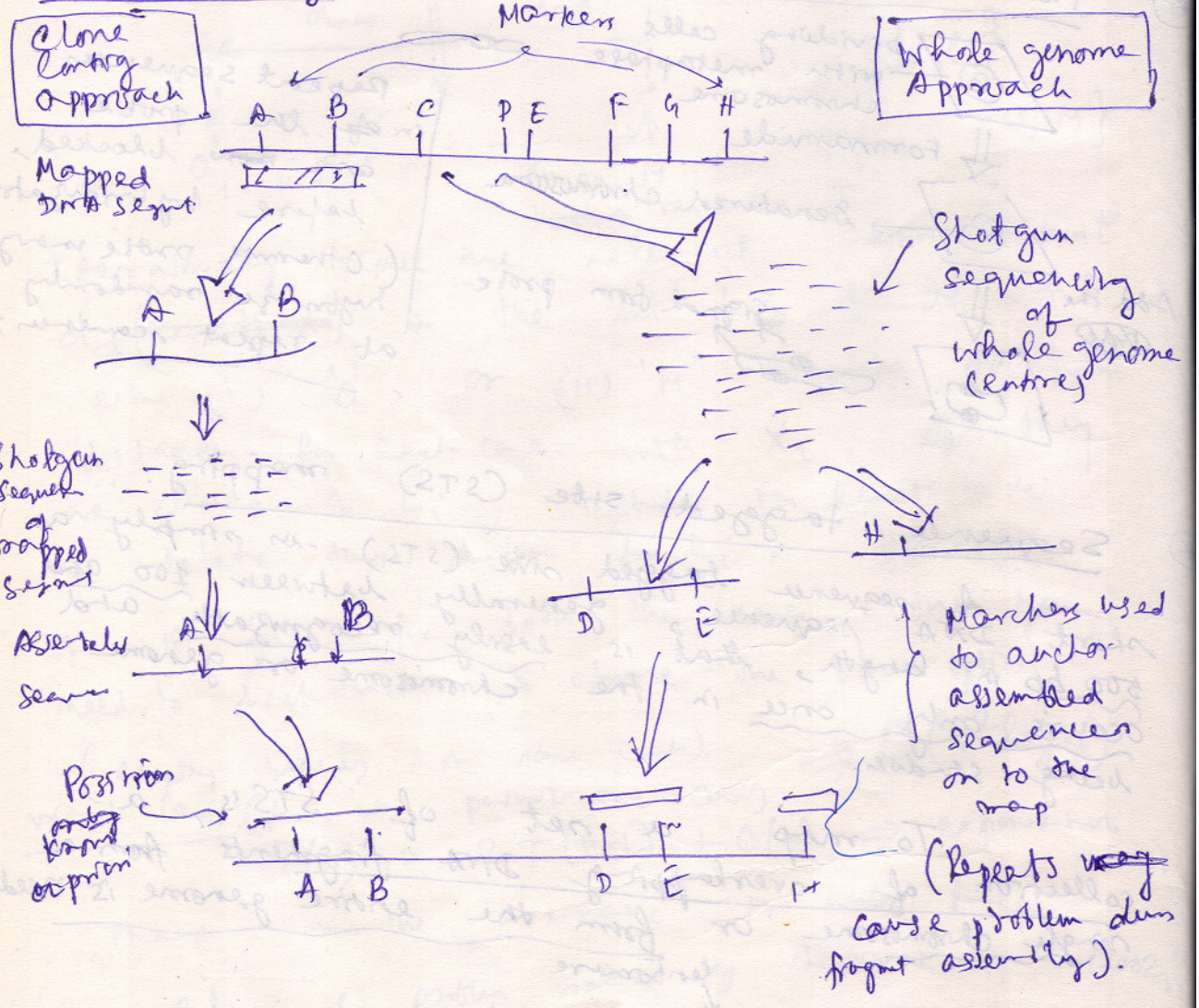
Use recombinant bacterial principle (for genetic mapping),

STS → Expressed Sequence Tags (ESTs)
→ SSLPs
→ Random genomic sequences

Expressed sequence tags (EST)

from cDNA, (fragments of mRNA).

DNA sequencing:



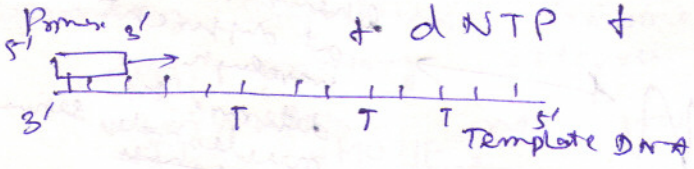
(Repeats ~~may~~ cause problem during fragment assembly).

Methodology of Sequencing Short segments

→ Chain termination method (Sanger et al 1977)

→ Chemical degradation method (Maxam & Gilbert, 1977)

Chain termination method:



+ dNTP + ddATP

(for detection T nucleotide)

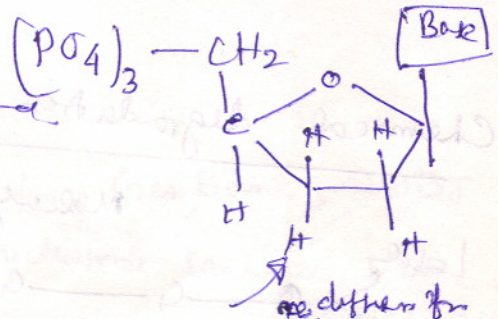
↓

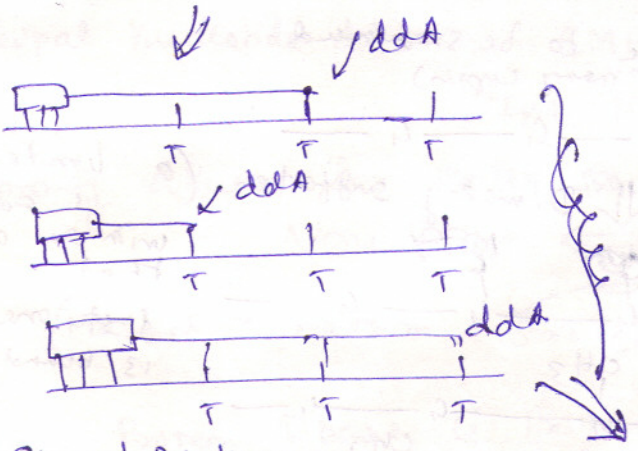
dideoxy nucleotide

nucleotide for

A ~~Primer~~ + ddNTP's

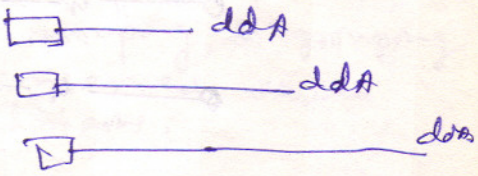
~~Reaction~~ stops Synthesis at next end





Strand synthesis terminates at ddA

The A family

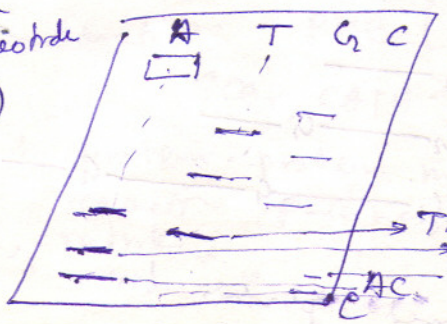


↓

Similar perform reaction for each type of nucleotide (T, G & C also)

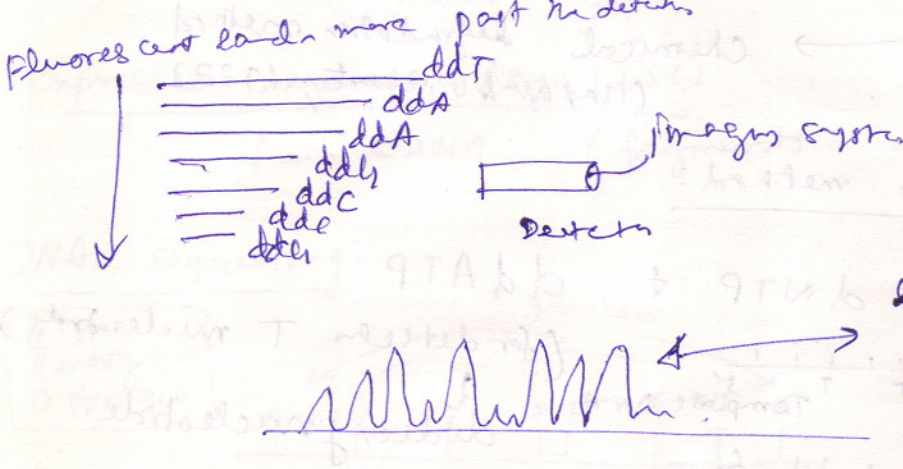
The results Autoradiogram

(Polyacrylamide Gel Electrophoresis)



Reads of strand

Fluorescent Labelers of ddNTP's (each type with different fluorescence label).



Chemical degradation sequencing method.