

## Topics

→ Hierarchical method of DNA Sequencing using Hybridization method

→ Algorithm for C1P Problem

There are two methods of obtaining the physical mapping of the clones using the Hierarchical method of sequencing, namely

- a) Restriction
- b) Hybridization method

Both the methods give the relative position of the clones with respect to the genome (called as physical mapping). Last class we dealt with restriction.

## Hybridization method

1. In the hybridization method, random probes of length 200-250bp are hybridized to the microarray like plates.
2. All the vectors or clones are treated with the probes and allowed to hybridize.
3. The probes to which the clones hybridize are recorded in the form of binary matrix with rows representing the clones and Probes representing the columns.

In the binary matrix, if the clone  $C_i$  hybridizes to Probe  $P_j$  then matrix  $M(i,j) = 1$  and if a mismatch occurs  $M(i,j) = 0$ .

The set of probes to which the clone hybridizes is called “Fingerprint”. Two clones overlap if they share some probes or part of the fingerprints.

In order to obtain the complete Physical mapping of the clones, we need to find the shortest string  $S$  of the probes in such a way that the substring of  $S$ , contains all set of probes for all clones. This is termed as Shortest String covering problem and it is NP-complete.

In order to solve the Shortest String covering problem we will have to permute the columns of the matrix in such a way that the columns of each row have consecutive 1's which is called as Consecutive one property (C1P).

Biological significance of C1P: As clones span a connected region of the genome - a consecutive 1's or match with the probes signify probes covering all regions of the clone without having any gaps.

Assumption for C1P :

- Error-free hybridization
- Unique probes are only present
- No two clones hybridize to same set of probes
- Every clone is detected by a probe (non-zero row in the matrix)

Algorithm for C1P problem :

Polynomial time complexity – includes 3 steps for solving the problem.

1. Separate the rows in connected components  
Row which are
  - a) Not subset of one another
  - b) Not disjoint - share atleast one probeare considered to make a connected component
2. Each connected component is considered and the columns are permuted to contain consecutive 1's. In order to do so we perform the following steps
  - a) Consider the rows in the connected component one by one
  - b) Arrange the common probes between the rows towards the center.
  - c) Mismatch probes can be either be put on the left or the right so that the C1P property obtained through the center elements is not altered.
  - d) Arrange the order of the clones in the connected component depending on the number of matches between them
    - i) Arrange in same direction if the number of matches is lesser than matches of all other combination of rows
    - ii) Arrange in opposite direction if the number of matches is greater than matches of all other combination of rows
3. Join the connected components which share at least one "FingerPrint" (one rows of same probe order) between them.

Thus the resulting matrix satisfies the C1P property and also gives the relative order of the clones for physical mapping.

With the physical mapping of clones being complete, shotgun (random fragmentation) of each clone is done and finally the whole genome sequencing is completed.