

Map Assembly

- Physical mapping can be understood in terms of the following analogy:
 1. Several copies of a book cut by scissors into thousand of pieces.
(Break the DNA molecule into small pieces.)
 2. Each copy is cut in an individual way such that a piece from one copy may overlap a piece from another copy.
 3. For each piece and each word from a list of key words (fingerprinting) we are told whether the piece contains the key word.
 4. Given this data (clone library), we wish to determine the patterns of overlaps of the pieces.

Continued

- After a clone library is constructed, we are aiming at reconstructing the relative placements of the clones along the DNA molecule. Some information may have lost, so, the reconstruction starts with the **fingerprinting of the clones**.
- In order to get the job done, we can build a mathematical model through the idea of interval graphs.

Definition (Interval graph)

An interval graph $G = (V, E)$ is the graph where $V(G)$ is a set of closed intervals in real line and two intervals are adjacent if and only if their corresponding intervals have non-empty intersection.

Interval Graphs

Lemma 1

Every interval graph G is triangulated, that is, every cycle of length larger than 3 has a chord. (**Chordal graphs**)

Proof.

Suppose not. Let $C = (u_1, u_2, \dots, u_k)$ be a cycle of length larger than 3 which has no chords. Thus $k > 3$. We may assume that $u_i = [a_i, b_i]$, $i = 1, 2, \dots, k$ where $a_1 \leq a_2 \leq a_3 \leq \dots \leq a_k$. Since u_1 is adjacent with u_k , $a_k \leq b_1$. Now, if $b_1 \leq b_j$ for some $j = 2, \dots, k-2$, then $a_k \leq b_1 \leq b_j$, this implies that u_k is adjacent with u_j which is a chord, a contradiction. On the other hand, $b_1 > b_j$ for all $j = 2, \dots, k-2$, and thus $b_1 > b_j \geq a_j$, implies u_1 is adjacent to u_j for all $j = 2, \dots, k-2$. Again, a contradiction. 

Comparability Graphs

Definition (Transitive orientation property)

An undirected graph $G = (V, E)$ has a transitive orientation property if G has an orientation on E such that the resulting digraph $D_G(V, A)$ satisfies for all vertices a, b and c in V , (a, b) and (b, c) are arcs of D_G implying (a, c) is an arc of D_G .

Note that if G has no triangles, then G is transitively orientable.

Definition (Comparability graphs)

An undirected graph that is transitively orientable is called a comparability graph.

Important Fact

Lemma 2

The complement of an interval graph is a comparability graph.

Proof.

Let G' be the complement of an interval graph G . Then, $[a, b]$ and $[a', b']$ of $V(G')$ are adjacent in G' if and only if they are mutually disjoint. Now, orient $[a, b]$ to $[a', b']$ provided $b < a'$. So, in case that $[a', b']$ is oriented to $[a'', b'']$, $b' < a''$. This implies that $[a, b]$ and $[a'', b'']$ are mutually disjoint, moreover, $b < a''$, that is $[a, b]$ is oriented to $[a'', b'']$ as well. 

Characterization of Interval Graphs

Theorem (Gilmore and Hoffman, 1964)

Let G be an undirected graph. The following statements are equivalent.

1. G is an interval graph.
2. G is a triangulated graph and its complement is a comparability graph.
3. The maximal cliques of G can be linearly ordered such that for each vertex v in G , the maximal cliques containing v occur consecutively.

Proof. (Omitted.)

Recognition of Interval Graphs

- Since the maximal cliques of G can be linearly ordered, the recognition of interval graphs turns out to be “**simple**” in the sense of **algorithm**.
- Also, we can use the idea of **recognition of triangulated graphs or comparability graphs**.



Restriction Fragment Fingerprints

- The simplest case of mapping with restriction fragment fingerprints is “**Single Complete Digest**” mapping (**SCD** mapping). This idea was proposed by Olson et al. in 1986.
- In this case, the fingerprint of a clone is a **multiset of the sizes of its restriction fragments** in a digest by a restriction enzyme.
- An SCD map is a placement of clones and restriction fragments consistent with the given SCD data (Gillett et al. 1995).

SCD Mapping Problem

Find a most compact map that is consistent with SCD data. (Using the minimum number of restriction fragments.)

Example

$A = \{5, 7, 7\}$, $B = \{5, 6, 7, 8\}$, $C = \{5, 6, 7, 8\}$ and $D = \{5, 6, 9\}$.

Solutions.

1. $\langle 6, 8, 7, 5, 7, 8, 6, 5, 9 \rangle$ $\langle 6, 8, 7, 5, 7, 8, 6, 5, 9 \rangle$

$\langle 6, 8, 7, 5, 7, 8, 6, 5, 9 \rangle$ $\langle 6, 8, 7, 5, 7, 8, 6, 5, 9 \rangle$

This solution has 9 restriction fragments.

2. $\langle 7, 5, 7, 8, 6, 5, 9 \rangle$ is a better solution with 7 restriction fragments.

() The problem of finding the most compact map for SCD problem is NP-hard.*

An Approach of Solving SCD Problem

- Jiang and Karp (1998) formulated SCD mapping with clone ordering as a constrained path cover problem on a special multistage graph.

Definition (Labeled Multistage Graph)

Let $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$ be an instance of SCD mapping where S_i is a multiset representing the fingerprint of the i -th clone ordering by the left endpoints. A labeled multistage graph G (called a clone-fragment graph) consists of n stages, with the i -th stage containing $|S_i|$ vertices. At stage i , G has a vertex for each element x of S_i with label x . Two vertices are adjacent if they are at adjacent stages and have identical label.

Example of Labeled Multistage Graph

$S_1 = \{5, 7, 7\}$, $S_2 = \{5, 6, 7, 8\}$, $S_3 = \{5, 6, 7, 8\}$ and $S_4 = \{5, 6, 9\}$.

The graph obtained is as following. (There is a P_4 for 5's, P_3 for 6's, S_3 for 7's, P_2 for 8's and a singleton for 9.

5		7	7	
5	6	7	8	
5	6	7	8	
5	6			9

Path Cover

Definition (Path Cover)

A path cover of a graph G is a collection of paths on G such that every vertex of G is contained in exactly one path.

(*) The solution 2 in page 9 corresponds to a path cover of its labeled multistage graph which has 7 paths and they are 7, 2x5, 3x7, 2x8, 2x6, 2x5 and 9. “3x7” is a path with three vertices labeled 7.

(*) The solution 1 in page 9 can be obtained by changing the order of S_1 and S_2 . (The star graph with three edges can not be covered by two disjoint paths of length 2.)

Conclusions

- The idea of conflicting path cover was forbidden since each S_i must be recognized as a path, for example $7 - 5 - 7$.
- The cover we find must be conflict-free.
- This is known as a consistent cover.
- *In 1998, Jiang and Karp obtained a 2-approximation algorithm for SCD mapping with a given ordering of clones.*
- *In order to get the correct map, we need signals in DNA.*