Combinatorial Problems in Physical Mapping

Richard M. Karp*

karp@cs.washington.edu

Department of Computer Science and Engineering, University of
Washington Seattle, WA 98195, U.S.A.

1 Abstract

Physical mapping is the process of determining the locations of landmarks along a large DNA
molecule. Physical maps are an essential tool for homing in on interesting genes, and also facil-
itate the process of sequencing large DNA molecules. Physical mapping is usually carried out
by constructing a clone library covering the DNA molecule of interest, performing experiments
to obtain "fingerprint data" about each clone, and then inferring the arrangement of clones
along the molecule with the help of this data. This inference process requires the solution of
large-scale combinatorial problems. The speaker will describe the combinatorial problems that
arise in each of the following approaches to physical mapping:

1. Sequence-tagged site (STS) mapping, in which the goal is to determine the locations of
   a set of sites, each of which the unique location of an associated sequence of nucleotides.
   STS mapping can be based either on a clone library or on a panel of radiation hybrids.

2. Oligonucleotide mapping, in which hybridization experiments are performed to determine
   which oligonucleotides from a given set occur on each clone;

3. Multiple-complete-digest mapping, in which each clone is completely digested with several
   different restriction enzymes, and the fragment sizes are measured. The goal is not only
   to determine the arrangement of the clones, but also to construct a restriction map giving
   the locations of all the measured restriction fragments. The speaker will also discuss a
   recently developed approach in which physical mapping is bypassed, and instead end-
   sequence data for each clone is used to guide the sequencing process.

The main emphasis in the talk will be on multiple-complete-digest mapping, an approach which is being pursued actively at the University of Washington. Algorithms will be presented for the following two problems:

(1) Given multiple complete digest data for each clone in a library of 105 or 106 clones, find those pairs of clones whose patterns of fragment lengths are highly similar. Such clones are likely to overlap.

(2) Given a proposed left-to-right ordering of the clones, construct the corresponding restriction map.