Handling experimental noise and integrating different types of data in genomic mapping

Andrei Grigoriev[1], Richard Mott[2] and Hans Lehrach[1]

[1]Max-Planck-Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany [2]The Sanger Centre, Hixton Hall, Hixton, Cambridge CB10 1RQ, UK

High levels of noise in practical experiments of large-scale genomic mapping represent a serious problem for a reliable reconstruction of a true order of clones. In hybridisation-based mapping projects, a list of major sources of experimental noise and map non-linearity includes repetitive probes, chimeric clone inserts, deletions/rearrangements of clones, microtitre plate well-contaminants and false positive/negative hybridisation signals on filters.

We present several approaches designed to filter out the noisy data and take into account potential non-linearity of a map. These are algorithms detecting repetitive elements in probes/clones, breaking chimeric inserts into original contiguous components and eliminating random noise in the data. These algorithms are implemented as a set of interactive and non-interactive programs which allow one to minimize the need for a human intevention in resolving contradictions and finding the most consistent interpretation of hybridisation data. They will be (and some of them already are) included into our publicly available contig-building software suite (Mott et al, 1993; Grigoriev et al, 1994).

Large mapping projects typically involve a number of independent techniques aimed at establishing and verification of the order and localisation of clones. Radiation and deletion hybrid mapping, metaphase chromosome FISH, restriction digest and Alu-PCR gel fingerprinting are among the most frequently employed methods of map verification.

We discuss the ways to incorporate these heterogenuous data types into an integrated clone map, using examples of real experimental projects, such as maps of the fission yeast *Schizosaccharomises pombe* genome and the human X chromosome.

Grigoriev,A., Mott,R., and Lehrach,H. (1994) An algorithm to detect chimeric clones and random noise in genomic mapping. Genomics 22, 482-486.

Mott,R.F., Grigoriev,A.V., Maier,E., Hoheisel,J.D., and Lehrach,H. (1993) Algorithms and software tools for ordering clone libraries: appplication to the mapping of the genome of Schizosaccharomyces pombe, Nucleic Acids Research. 21, 1965-1974.