

Workshop Focuses on DNA Sequence Annotation

By Richard Mural, Life Sciences Division, Oak Ridge National Laboratory

Introduction

Automatic annotation of large amounts of genomic DNA sequence clearly is and will continue to be a formidable challenge. This problem will be addressed properly only by developing very efficient computational tools for initial sequence annotation, treating the annotations as hypotheses, and testing and verifying them in the laboratory. Additionally, if the generated annotations are to be of maximum usefulness, results must be stored in an easily retrievable and queryable form in well-curated databases. The "If you sequence it, the community will annotate it" approach is unlikely to produce desired results, and new paradigms and possibly new organizational models will need to be implemented to present genomic sequence in its most useful form.

Annotation Meeting

The Fifth International Conference on Intelligent Systems for Molecular Biology held June 21-25, 1997, in Porto Carras, Greece, ended with a workshop on Automatic Annotation of Genome Sequence Data. Eight workshop speakers addressed three basic questions:

What are the challenges in automatic annotation?

What are the best technologies for doing this job?

What is the best division of labor between biology and computer science?

Introductory remarks by session chairman Chris Sander [European Molecular Biology Laboratory European Bioinformatics Institute (EBI)] made clear that no one yet has the experience to know the right way to proceed with automatic annotation. Richard Durbin (Sanger Centre) stressed an often-repeated theme that proper annotation will require wet-laboratory work as well as computational annotation. He also stressed the need for curated databases. Michael Ashburner (EBI) discussed his experience in annotating *Drosophila* sequences and the need for hierarchical controlled vocabularies. He suggested the possibility of an annotation database that would be separate from but seamlessly linked to the sequence databases.

Three other speakers addressed general problems in genomic-sequence annotation: Antoine Danchin (Institut Pasteur) discussed *Bacillus subtilis*, Terry Gaasterland (Argonne National Laboratory) described annotating microbial genomes, and Chris Overton (University of Pennsylvania) shared experiences from a project to annotate genomic sequence from human chromosome 22. Four other speakers discussed annotation efforts and tools being developed in the bioinformatics industry. [Richard Mural, muralrj@ornl.gov]