

Contents

Preface

(xiii)

1. TECHNIQUES FOR SEARCH

1

Secondary database searches
Secondary database defined
Regular expressions
Rules
Fuzzy regular expressions
Fingerprints
Blocks
Profiles
Hidden Markov Models
Remember

2. TECHNIQUES OF COMPARISON

19

Search for database
Alphabets and complexity
Algorithms and programmes
Comparing two sequences
Sub-sequences
Identity and similarity
Dayhoff Mutation Data Matrix
BLOSUM matrices
Statistical measures
The Dotplot
Local and global similarity
Global alignment
Local alignment : The Smith-Watchman algorithm
Dynamic programming
Pairwise database searching
FastA
BLAST
Gapped BLAST
Remember

3.	SEQUENCING BIOLOGICAL OBJECTS	47
	Defining Bioinformatics	
	Protein sequencing	
	Nucleic acid sequencing	
	Biological sequence/structure deficit	
	Genome projects	
	Status of the human genome project	
	Importance of bioinformation	
	Pattern recognition and prediction	
	Folding problem	
	Role of chaperones	
	Sequence analysis	
	Twilight Zone	
	Homology and analogy	
	Orthology and paralogy	
	About the devil	
	Remember	
4.	SEQUENCING DNA	69
	EMBL	
	DDBJ	
	GenBank	
	Structure of GenBank entries	
	dbEST	
	GSDB	
	Specialised genomic resources	
	SGD	
	UniGene	
	TDB	
	ACeDB	
	Remember	
5.	SEQUENCE ANALYSIS	79
	Multiple sequence alignment	
	Definition of multiple sequence alignment	
	The consensus	
	Computational complexity	
	Manual methods	
	Simultaneous methods	
	Progressive methods	
	Databases of multiple alignments	
	Searching databases with multiple alignments	
	PSI-BLAST	
	Remember	

47

6. ANALYSIS PROTOCOL

95

Protein function predicted
 Searching primary databases
 Identity searching
 Similarity searching
 Searching secondary databases
 Searching PROSITE
 Searching profile library
 Searching Pfam
 Searching PRINTS
 Searching BLOCKS databases
 Searching IDENTIFY
 Believing a result ?
 Structural and functional interpretation
 Functions of transferrins
 The look of transferrins
 SCOP
 CATH
 PDBsum
 Remember

69

7. ANALYSIS PACKAGES

113

Commercial databases
 Commercial software
 Comprehensive packages
 GCG
 EGCG
 Staden
 Lasergene
 Packages specialising in DNA analysis
 Intranet packages
 SYNERGY
 GeneMill, GeneWorld, GeneThesaurus
 Internet packages
 CINEMA
 EMBOSS
 Alfresco
 Remember

79

8. DNA SEQUENCE ANALYSED

127

Analysing DNA
 Gene structure and DNA sequences
 Untranslated regions
 Conceptual translation
 Features of DNA sequence analysis
 Detecting open reading frames

Understanding the effect of introns and exons
 DNA sequence assembly
 Issues in the interpretation of EST searches
 EST alphabet
 Insertions, deletions and frameshifts
 Splice variants
 Non-coding region ESTs
 Two approaches to gene hunting
 Positional cloning
 RNA transcript analysis
 Hierarchy of genomic information
 Expression profile of a cell
 cDNA libraries and ESTs
 Different approaches to EST analysis
 Merck/IMAGE
 Incyte
 TIGR
 TIGR Human Gene Index
 Effects of EST data on DNA databases
 EST analysis tools
 Sequence similarity search tools
 Sequence assembly tools
 Sequence clustering tools
 A practical example of EST analysis
 Remember

9. DATABASES FOR BIOLOGICAL PURPOSES

157

Protein information resources
 Primary sequence databases
 Nucleic acid sequence databases
 Protein sequence databases
 PIR
 Martinstried Institute of Protein Sequence (MIPS)
 SWISS-PROT
 Structure of SWISS-PROT entries
 TrEMBL
 NRL-3D
 An embarras de richesses
 Composite protein sequence databases
 Non-redundant data base (NRDB)
 OWL
 MIPSX
 SWISS-PROT+TrEMBL
 Embarras du choix
 Secondary databases
 Why secondary databases?

PROSITE
 Structure of PROSITE entries
 PRINTS
 BLOCKS
 Structure of BLOCKS entries
 Blocks-format PRINTS
 Profiles
 Structure of PROSITE profile entries
 Pfam
 Structure of Pfam entries
 IDENTIFY
 Composite protein pattern databases
 Structure classification databases
 SCOP
 SCOP classification
 CATH
 PDBsum
 Remember

10. WORLD WIDE WEB AND INTERNET

191

The Network
 Computer Findings
 Facilities on Internet
 World Wide Web
 Web browsers
 Lynx
 Mosaic
 Netscape Navigator
 Internet Explorer
 HTTP, HTML and URLs
 The European Molecular Biology network — EMBnet
 Hinxton Hall
 Sanger Centre
 UK MRC Human Genome Mapping Project
 European Bioinformatics Institute — EBI
 MIPS
 UCL
 Sequence Retrieval System — SRS
 National Center for Biotechnology Information NCBI
 Entrez
 Virtual tourism
 Remember

11. IMMUNOTECHNOLOGY

207

Natural and Acquired Immunity
 'Natural immunity'

Acquired immunity
 Immune responses
 Lymphoid system (B cells and T cells)
 Clonal Selection Theory
 Adjuvants and Immunosuppressants
 Structure of Immunoglobulins
 Genetics and Molecular Biology of the Production
 of Antibodies or Immunoglobulins
 Ig gene in germline and differentiated cells
 V-regions through DNA splicing
 V-C association through DNA splicing
 Allelic Exclusion
 Interferons
 Vaccine Development and Immunization
 History of vaccine development
 Nature of vaccines
 Basic Hybridoma Technology
 Improvements in hybridoma technology
 Serum free media for bulk culture of hybridoma cells
 Purification of antibodies
 Antibody Engineering using Genetic Manipulations
 Alternatives to hybridoma technology for monoclonal
 antibodies
 Bypassing hybridomas and cloning of Mab genes
 Designing and building of Mab genes
 Primary and secondary libraries for antibody genes
 Production of Human and Humanized Antibodies
 Uses of Monoclonal Antibodies
 Monoclonal antibodies in diagnosis, screening
 and therapy
 Monoclonal antibodies in vaccine production
 Monoclonal antibodies as enzymes (abzymes)
 Monoclonal antibodies for purification and
 quantitation of other molecules
 Monoclonal antibodies for cytogenetic analysis

12 MOLECULAR MANIPULATION

249

Isolation

Past attempts for isolation of ribosomal RNA genes
in *Xenopus*

Isolation of genes coding for specific proteins

Isolation of genes which are tissue specific in expression

Isolation of genes using DNA or RNA probes

Isolation of genes coding for unknown products

Isolation of novel genes by Asia Datta at JNU, New Delhi

Sequencing

Maxam and Gilbert's chemical degradation method
 Sanger's dideoxynucleotide synthetic method
 Synthesis
 Chemical synthesis of tRNA genes
 Total synthesis of a human leukocyte interferon gene
 Gene synthesis machines
 Synthesis of genes from mRNA

13. MAPPING OF GENOMES

271

Molecular Markers
 Restriction fragment length polymorphisms (RFLPs)
 Random amplified polymorphic DNAs (RAPDs)
 Minisatellites or VNTRs (variable number of tandem repeats) and microsatellites or SSRs (simple sequence repeats)
 Genetic Maps
 Approaches used for mapping of molecular markers
 Linkage and Recombination between Molecular and Phenotypic Markers
 Computer software for construction of linkage maps
 Molecular Genetic Maps in Humans (*homo sapiens*)
 Molecular Genetic Maps in Mouse
 Molecular genetic maps in other animal species
 Cytogenetic Maps
 Cytogenetic Maps Using Chromosome Specific Markers
 Cytogenetic RFLP Maps Using Radiation Hybrid (RH) Mapping
 Cytogenetic RFLP Maps Using PCR
 Physical Maps
 Physical Maps Using Chromosomal Aberrations (deletions)
 Physical Maps Using in Situ Hybridization
 Physical Maps Using Yeast Artificial Chromosomes (YAC) and ISH
 Physical Maps Using Chromosome Walking and Jumping; (or hopping)
 Physical Maps of Large DNA Fragments by Chromosome Fragmentation
 Physical Maps for Mammalian Chromosomes (including humans)
 Sequence Tagged Sites (STS) and Expressed Sequence Tags (EST)
 Integrated Genomic Maps
 Sequencing of Genomes

GLOSSARY OF KEY TERMS

301