BLAST / Korf, I.,

ISBN 9780596002992

Using NCBI-BLAST 4 --
Alternate Output Formats 12 --
Alternate Alignment Views 13 --

Part II. Theory --

2. Biological Sequences 19 --
The Central Dogma of Molecular Biology 19 --
Evolution 27 --
Genomes and Genes 35 --
Biological Sequences and Similarity 38 --

3. Sequence Alignment 40 --
Global Alignment: Needleman-Wunsch 40 --
Local Alignment: Smith-Waterman 46 --
Dynamic Programming 50 --
Algorithmic Complexity 50 --
Global Versus Local 50 --
Variations 51 --

4. Sequence Similarity 55 --
Introduction to Information Theory 55 --
Amino Acid Similarity 57 --
Scoring Matrices 59 --
Target Frequencies, lambda, and H 60 --
Sequence Similarity 64 --
Karlin-Altschul Statistics 65 --
Sum Statistics and Sum Scores 67 --

Part III. Practice --

5. BLAST 75 --
The Five BLAST Programs 75 --
The BLAST Algorithm 76 --

6. Anatomy of a BLAST Report 88 --
Basic Structure 88 --
Alignments 90 --

7. A BLAST Statistics Tutorial 96 --
Basic BLAST Statistics 96 --
Using Statistics to Understand BLAST Results 109 --
Where Did My Oligo Go? 109 --

8. 20 Tips to Improve Your BLAST Searches 116 --
8.1 Don't Use the Default Parameters 116 --
8.2 Treat BLAST Searches as Scientific Experiments 116 --
8.3 Perform Controls, Especially in the Twilight Zone 117 --
8.4 View BLAST Reports Graphically 118 --
8.5 Use the Karlin-Altschul Equation to Design Experiments 119 --
8.6 When Troubleshooting, Read the Footer First 119 --
8.7 Know When to Use Complexity Filters 120 --
8.8 Mask Repeats in Genomic DNA 121 --

8.9 Segment Large Genomic Sequences 121 --
8.10 Be Skeptical of Hypothetical Proteins 123 --
8.11 Expect Contaminants in EST Databases 123 --
8.12 Use Caution When Searching Raw Sequencing Reads 124 --
8.13 Look for Stop Codons and Frame-Shifts to find Pseudo-Genes 124 --
8.14 Consider Using Ungapped Alignment for BLASTX, TBLASTN, and TBLASTX 124 --
8.15 Look for Gaps in Coverage as a Sign of Missed Exons 126 --
8.16 Parse BLAST Reports with Bioperl 126 --
8.17 Perform Pilot Experiments 128 --
8.18 Examine Statistical Outliers 128 --
8.19 Use links and topcomboN to Make Sense of Alignment Groups 128 --
8.20 How to Lie with BLAST Statistics 128 --

9. BLAST Protocols 130 --
BLASTN Protocols 131 --
BLASTP Protocols 144 --
BLASTX Protocols 147 --
TBLASTN Protocols 152 --
TBLASTX Protocols 155 --
Part IV. Industrial-Strength BLAST --

10. Installation and Command-Line Tutorial 161 --
NCBI-BLAST Installation 161 --
WU-BLAST Installation 166 --
Command-Line Tutorial 170 --
Editing Scoring Matrices 186 --

11. BLAST Databases 188 --
FASTA Files 188 --
BLAST Databases 193 --
Sequence Databases 198 --
Sequence Database Management Strategies 206 --

12. Hardware and Software Optimizations 213 --
The Persistence of Memory 213 --
CPUs and Computer Architecture 215 --
Compute Clusters 216 --
Distributed Resource Management 218 --
Software Tricks 220 --
Optimized NCBI-BLAST 224 --
Part V. BLAST Reference --

13. NCBI-BLAST Reference 229 --
Usage Statements 229 --
Command-Line Syntax 229 --
blastall Parameters 230 --
formatdb Parameters 240 --
fastacmd Parameters 242 --
megablast Parameters 245 --
bl2seq Parameters 252 --
blastpgp Parameters (PSI-BLAST and PHI-BLAST) 256 --
blastclust Parameters 264 --

14. WU-BLAST Reference 267 --
Usage Statements 268 --
Command-Line Syntax 268 --

WU-BLAST Parameters 269 --
xdformat Parameters 281 --
xdget Parameters 285 --

A. NCBI Display Formats 291 --
B. Nucleotide Scoring Schemes 299 --
C. NCBI-BLAST Scoring Schemes 302 --
D. blast-imager.pl 305 --
E. blast2table.pl 309.