
Contents

Preface	ix
Part 1. Fundamental Concepts	1
Chapter 1. Introduction	3
1.1. Biological sequences: DNA/RNA/proteins	3
1.2. From DNA to RNA and proteins	4
1.3. RNA sequence, structure and function	5
1.4. Protein sequence, structure and function	7
1.5. Sequence evolution	9
1.6. MSA: basic concepts	10
1.6.1. Pairwise versus multiple alignment	11
1.6.2. Local versus global alignment	11
1.7. Multiple sequence alignment applications	12
1.7.1. Phylogenetic studies	12
1.7.2. Comparative genomics	14
1.7.3. Gene prediction and validation	15
1.7.4. Protein structure and function	17
1.7.5. RNA structure and function	20
1.7.6. Interaction networks	22
1.7.7. Genetics	23
1.7.8. Drug discovery and design	24

Part 2. Traditional Multiple Sequence Alignment Methods	27
Chapter 2. Heuristic Sequence Alignment Methods	29
2.1. Optimal sequence alignment	30
2.1.1. Dynamic programming	30
2.1.2. Alignment parameters	32
2.1.3. Limitations	34
2.2. Progressive multiple alignment	35
2.3. Iterative alignment	37
2.4. Consistency-based alignment	39
2.5. Cooperative alignment strategies	40
Chapter 3. Statistical Alignment Approaches	43
3.1. Probabilistic models of sequence evolution	44
3.2. Profile HMM-based alignment	45
3.3. Simulated annealing	46
3.4. Genetic algorithms	48
Chapter 4. Multiple Alignment Quality Control	53
4.1. Objective scoring functions	54
4.2. Determination of reliable regions	57
4.3. Estimation of homology	59
Chapter 5. Benchmarking	63
5.1. Criteria for benchmark construction	64
5.1.1. Defining the benchmark test cases	64
5.1.2. Defining the correct alignment	65
5.1.3. Defining the alignment score	65
5.2. Multiple alignment benchmarks	66
5.2.1. BAliBASE	66
5.2.2. OxBench	67
5.2.3. PREFAB	68
5.2.4. SABmark	69
5.2.5. BRALiBASE	69
5.2.6. Phylogeny-based benchmark	70
5.3. Comparison of multiple alignment benchmarks	71

Part 3. Large-scale Multiple Sequence Alignment Methods	73
Chapter 6. Whole Genome Alignment	75
6.1. Pairwise genome alignment	76
6.2. Progressive methods for multiple genome alignment	78
6.3. Graph-based methods for multiple genome alignment	79
6.4. Meta-aligners for multiple genome alignment	81
6.5. Accuracy measures for genome alignment methods	82
6.6. Benchmarking genome alignment	84
Chapter 7. Multiple Alignment of Thousands of Sequences	87
7.1. Extension of the progressive alignment approach	89
7.2. Meta-aligners for large numbers of sequences	90
7.3. Extending “seed” alignments	91
7.4. Benchmarking large numbers of sequences	93
Chapter 8. Future Perspectives: High-Performance Computing	97
8.1. Coarse-grain parallelism: grid computing	99
8.2. Fine-grain parallelism: GPGPU	101
8.3. MSA in the cloud	104
Bibliography	109
Index	133