

1. Record Nr.	UNINA9910484594803321
Titolo	Algorithms in Bioinformatics : 15th International Workshop, WABI 2015, Atlanta, GA, USA, September 10-12, 2015, Proceedings // edited by Mihai Pop, H�el�ene Touzet
Pubbl/distr/stampa	Berlin, Heidelberg : , : Springer Berlin Heidelberg : , : Imprint : Springer, , 2015
ISBN	3-662-48221-5
Edizione	[1st ed. 2015.]
Descrizione fisica	1 online resource (XX, 328 p. 85 illus.)
Collana	Lecture Notes in Bioinformatics, , 2366-6331 ; ; 9289
Disciplina	572.80285
Soggetti	Bioinformatics Algorithms Computer science - Mathematics Discrete mathematics Data mining Proteins Computational and Systems Biology Discrete Mathematics in Computer Science Data Mining and Knowledge Discovery
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Intro -- Preface -- Organization -- Abstracts -- Reference-free Compression of High Throughput Sequencing Data with a Probabilistic de Bruijn Graph -- Genome Scaffolding with PE-Contaminated Mate-Pair Libraries -- Network Properties of the Ensemble of RNA Structures -- Contents -- BicNET: Efficient Biclustering of Biological Networks to Unravel Non-Trivial Modules -- 1 Introduction -- 2 Background -- 3 Solution -- 3.1 Network Modules with Flexible Coherency -- 3.2 BicNET: Efficient Biclustering of Biological Networks -- 4 Results and Discussion -- 4.1 Results on Synthetic Data -- 4.2 Results on Real Data -- 5 Conclusions and Future Work -- References -- Simultaneous Optimization of both Node and Edge Conservation in Network Alignment via WAVE -- 1 Introduction -- 1.1 Motivation -- 1.2 Related Work -- 1.3 Our Contributions and Significance -- 2 Methods -- 2.1

Data -- 2.2 Combining Topological and Sequence Information Within NCF -- 2.3 Evaluation of Alignment Quality -- 2.4 Our Methodology -- 3 Results and Discussion -- 3.1 Comparison of Edge-Weighted and Edge-Unweighted Versions of WAVE -- 3.2 Comparison of Different Parameter Values Within WAVE -- 3.3 Comparison of Five NCF-AS Methods -- 3.4 Comparison of WAVE with Very Recent Methods -- 4 Concluding Remarks -- A Appendix -- A.1 Appendix Figures -- References -- The Topological Profile of a Model of Protein Network Evolution Can Direct Model Improvement -- 1 Introduction -- 2 Methods -- 2.1 The Evolutionary Model -- 2.2 The Empirical Network -- 2.3 Topological Measures Assessed -- 3 Results -- 3.1 Success of Driving Properties Towards the Empirical Topology -- 3.2 Non-Optimized Topological Properties that Trend Towards Empirical Values -- 3.3 Variability of Topological Measures -- 3.4 Correlation Among Topological Characteristics -- 3.5 Improving the Evolutionary Model -- 4 Conclusion.

References -- Algorithms for Regular Tree Grammar Network Search and Their Application to Mining Human-Viral Infection Patterns -- 1 Introduction -- 2 Regular Tree Grammars and Curry-Encoding -- 3 The RTG Network Parse-and-Search Problem -- 4 Algorithms for Solving Problem 2 -- 4.1 Stage 1: Hypergraph Construction and Its Optimal Derivation -- 4.2 Stage 2: Computing K-Best Scoring Trees -- 5 Results -- 6 Discussion -- References -- Orthology Relation and Gene Tree Correction: Complexity Results -- 1 Introduction -- 2 Trees and Orthology Relations -- 2.1 Evolution of a Gene Family -- 2.2 Relation Graph -- 3 Relation Correction Problems -- 3.1 The Minimum Edge-Removal Consistency Problem -- 3.2 The Minimum Node-Removal Consistency Problem -- 4 Gene Tree Correction Problems -- 4.1 The Maximum Homology Correction Problem -- 4.2 The Maximum Clade Correction Problem -- 5 Algorithmic Avenues -- 6 Conclusion -- References -- Finding a Perfect Phylogeny from Mixed Tumor Samples -- 1 Introduction -- 2 Problem Formulation -- 3 A Characterization of Row-Conflict Graphs -- 4 Complexity Results -- 5 An Algorithm for the Case When No Column Is Contained in both Columns of a Pair of Conflicting Columns -- 6 Discussion -- References -- A Sub-quadratic Time and Space Complexity Solution for the Dated Tree Reconciliation Problem for Select Tree Topologies -- 1 Background -- 2 Methodology -- 2.1 The Number of Mapping Sites Required for Each Parasite Node -- 2.2 Space Complexity Reduction -- 2.3 Time Complexity Reduction -- 3 Results and Discussion -- 3.1 Analysis of Space Complexity Improvements (Synthetic Data) -- 3.2 Analysis of Time Complexity Improvements (Synthetic Data) -- 3.3 Time and Space Complexity Improvements for Biological Data -- References -- Maximum Parsimony Analysis of Gene Copy Number Changes -- 1 Introduction -- 2 Methods.

2.1 The Rectilinear Steiner Minimum Tree (RSMT) Problem -- 2.2 The Maximum Parsimony Tree (MPT) Problem -- 2.3 From MPT to RSMT -- 3 Experimental Results -- 3.1 Real Cancer Datasets -- 3.2 Simulated Cancer Data -- 4 Discussion -- References -- Multiple-Ancestor Localization for Recently Admixed Individuals -- 1 Introduction -- 2 Materials and Methods -- 2.1 Estimation of Spatial Parameters from Training Data -- 2.2 Localization of Individuals of 1-Generation Admixture -- 2.3 A Spatial Model for Individuals of g-generations Admixture -- 2.4 Simulation Setup -- 2.5 Method Comparison -- 3 Results -- 3.1 Localization of Simulated 1-Generation Admixed Individuals -- 3.2 Localization of Real Europeans from the POPRES Dataset -- 3.3 Localization of Simulated g-generation Admixed Individuals -- 4 Discussion -- References -- Association Mapping for

Compound Heterozygous Traits Using Phenotypic Distance and Integer Programming -- 1 Biological Background and CH-Model -- 1.1 A Formal Model of a CH-trait at a Single Gene -- 2 The Phenotypic Distance Problem -- 2.1 An ILP Formulation for the Phenotypic Distance Problem -- 3 Simulated Data -- 4 The Most Striking and Positive Empirical Results -- 4.1 Speeding Up the Computations for Non-causal Genes and Permuted Data -- References -- Semi-nonparametric Modeling of Topological Domain Formation from Epigenetic Data -- 1 Introduction -- 1.1 Additional Related Work -- 2 The nTDP Model -- 2.1 The Likelihood Function -- 2.2 Nonparametric Form of the Effect Functions -- 3 Optimal Algorithms for Training and Inference -- 3.1 Training -- 3.2 Training Extensions -- 3.3 Inferring Domains Using the Trained Model -- 4 Results -- 4.1 Experimental Setup -- 4.2 nTDP Finds a Small Subset of Modifications Predictive of TADs -- 4.3 Predicting TADs from Histone Marks in Human -- 4.4 Multiscale Analysis of the Predicted TADs.

5 Conclusion -- References -- Scribe: Ultra-Accurate Error-Correction of Pooled Sequenced Reads -- 1 Introduction -- 2 Preliminaries -- 2.1 Pooling Design -- 2.2 Read Decoding -- 3 Methods -- 3.1 Indexing k-mers -- 3.2 Identification and Correction of Sequencing Errors -- 4 Experimental Results -- 4.1 Results on Synthetic Reads for the Rice Genome -- 4.2 Results on Real Reads from the Barley Genome -- References -- Jabba: Hybrid Error Correction for Long Sequencing Reads Using Maximal Exact Matches -- 1 Introduction -- 1.1 Related Work -- 2 Methods -- 2.1 Overview -- 2.2 Correction of the de Bruijn Graph -- 2.3 Aligning Reads to a de Bruijn Graph -- 3 Results Concerning Maximal Exact Matches -- 3.1 Coverage by Exact Regions -- 3.2 Occurrence of Exact Regions -- 3.3 Applications -- 4 Results -- 4.1 Data -- 4.2 Parameters -- 4.3 Evaluation Metrics -- 4.4 Evaluation -- References -- Optimizing Read Reversals for Sequence Compression -- 1 Introduction -- 2 Problem Taxonomy -- 3 Hardness Results -- 4 Approximation Algorithms -- 4.1 Approximations Based on TSP -- 4.2 Ordering with Reversals for Palindromic Value Functions -- 5 Experimental Results -- References -- Circular Sequence Comparison with q-grams -- 1 Introduction -- 2 Definitions and Properties -- 3 Algorithms -- 3.1 Algorithm hCSC: a Heuristic Algorithm -- 3.2 Algorithm saCSC: an Exact Suffix-Array-based Algorithm -- 4 Experimental Results -- 4.1 Accuracy -- 4.2 Time Performance -- 4.3 Application to Real Data -- 5 Final Remarks -- References -- Bloom Filter Trie -- A Data Structure for Pan-Genome Storage -- 1 Introduction -- 2 Existing Approaches -- 2.1 Data Structures -- 2.2 Software for Pan-Genome Storage -- 3 The Bloom Filter Trie -- 3.1 Uncompressed Container -- 3.2 Compressed Container -- 4 Operations Supported by the Bloom Filter Trie -- 4.1 Look-Up -- 4.2 Insertion.

4.3 Color Compression -- 5 Traversing Successors and Predecessors -- 6 Evaluation -- 7 Conclusion -- References -- A Filtering Approach for Alignment-Free Biosequences Comparison with Mismatches -- 1 Introduction -- 2 Preliminaries -- 2.1 Related Work -- 3 The MissMax Algorithm -- 3.1 Initial Set Up: MaxCork(1) -- 3.2 Minimum MaxCork from the Previous Step -- 3.3 Potential Candidates from the Previous Step -- 3.4 Theoretical and Practical Considerations -- 4 Preliminary Experimental Analysis -- 5 Concluding Remarks -- References -- Models and Algorithms for Genome Rearrangement with Positional Constraints -- 1 Introduction -- 1.1 Genomes as Sets of Signed Integers -- 1.2 DCJ and Sorting DCJs -- 1.3 The Minimum Weighted Rearrangements Problem -- 1.4 Positional Constraints as Colored Adjacencies -- 1.5 Locality and the Adjacency Graph -- 1.6 A Positional

Weight Function -- 2 Minimum Local Parsimonious Scenario -- 2.1
Colored Partitions -- 2.2 Even-length Paths -- 3 Conclusion --
References -- Sparse RNA Folding Revisited: Space-Efficient Minimum
Free Energy Prediction -- 1 Introduction -- 2 Time and Space Efficient
Computation of the MFE -- 2.1 Time and Space Efficient Bp-Based
Folding -- 2.2 Time and Space Efficient Calculation of the MFE -- 2.3
The Difficulty of MFE Fold Reconstruction Compared to Bp-Based
Folding -- 3 MFE Folding with Fold Reconstruction -- 3.1 Adding Trace
Arrows -- 3.2 Avoiding Trace Arrows -- 3.3 Garbage Collecting Trace
Arrows -- 3.4 Algorithm Summary -- 3.5 Complexity -- 4 Empirical
Results -- 5 Conclusion and Future Work -- References -- A Sparsified
Four-Russian Algorithm for RNA Folding -- 1 Introduction -- 2
Problem Definition and Basic Algorithm -- 2.1 Extending the Notation
and Moving Towards a Vector by Vector Computation of -- 3
Sparsification of the SSF Algorithm -- 3.1 OCT and STEP Sub-instances
of Sequence.
4 On-demand Four Russians Speedup.

Sommario/riassunto

This book constitutes the refereed proceedings of the 15th International Workshop on Algorithms in Bioinformatics, WABI 2015, held in Atlanta, GA, USA, in September 2015. The 23 full papers presented were carefully reviewed and selected from 56 submissions. The selected papers cover a wide range of topics from networks to phylogenetic studies, sequence and genome analysis, comparative genomics, and RNA structure.
