Gianluca Della Vedova | CV

✉ gianluca.dellavedova@unimib.it

🌐 https://www.unimib.it/gianluca-della-vedova • ⦿ gdv

ⓘ 0000-0001-5584-3089

Current position

Associate professor with tenure (SSD INF/01 — Informatica)
Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano – Bicocca

Positions

Associate Professor	**Univ. Milano – Bicocca**
Dept. of Informatics, Systems, and Communications	*10/2012–today*
Associate Professor	**Univ. Milano – Bicocca**
School of Statistics	*10/2005–09/2012*
Assistant Professor	**Univ.Milano – Bicocca**
School of Statistics	*05/2001–09/2005*

Habilitation

Abilitazione a Professore di I fascia, settore concorsuale 01/B1
Italian habilitation to be Full professor of Computer Science   *08/2018–today*

Education

Ph.D. in Computer Science   **Università degli Studi di Milano**
1996–2001

Thesis: "Multiple Sequence Alignment and Phylogenetic Reconstruction: Theory and Methods in Biological Data Analysis"

M.Sc. in Computer Science   **Università degli Studi di Milano**
1990–1995

Thesis: "Sequential and Parallel Algorithms for Graph Decomposition"

Funding

Leading

2020–2023: H2020-MSCA-RISE-2019 (Univ. Milano – Bicocca amount: 197 800€) Pangenome Graph Algorithms and Data Integration (PANGAIA) — Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange programme.

This project is coordinated by a co-leader of my research lab and it involves a post-doc I supervise (Simone Ciccolella), another post-doc a co-supervise, and 2 PhD students (that I co-supervise).

The main goal of this project is to foster collaborations between different research groups in computational pangenomics. I have been one of the main participants in planning and writing the proposal. I am **leading WP5** Communication and Dissemination.

The project involves Universität Bielefeld (Germany), Centrum Wiskunde & Informatica (the Netherlands), The Institut Pasteur (France), Comenius University Bratislava (Slovakia), Geneton s.r.o. (Slovakia), Cornell University (USA), University of Tokyo (Japan), Simon Fraser University (Canada), Pennsylvania State University (USA), University of California (USA). I have developed an especially strong collaboration with Prof. Iman Hajirasouliha at Cornell University on tumor phylogeny inference and metagenomics where also Simone Ciccolella is actively working.

2021–2024: H2020-MSCA-ITN (Univ. Milano – Bicocca amount: 261 499,68€) ALgorithms for PAngenome Computational Analysis (ALPACA) — Horizon 2020 Marie Skłodowska-Curie Innovative Training Network.

As part of the activities of this project I am **co-supervising** a PhD student. The project involves Universität Bielefeld (Germany), Centrum Wiskunde & Informatica (the Netherlands), The Institut Pasteur (France), Comenius University Bratislava (Slovakia), Geneton s.r.o. (Slovakia), University of Helsinki (Finland), European Molecular Biology Laboratory — European Bioinformatics Institute, Heinrich-Heine-Universität Düsseldorf (Germany), Università di Pisa (Italy), CNRS (France), University of Cambridge (UK), INRIA (France), Pendulum Therapeutics, Inc. (USA), Vrije Universiteit Amsterdam (the Netherlands), Oxford Nanopore Technologies plc (UK), DNANexus (US), Sorbonne University (France), Deinove (France), Finnish Red Cross (Finland), BaseClear B.V. (the Netherlands), French Alternative Energies and Atomic Energy Commission (France), Cornell University (USA), University of Tokyo (Japan), Simon Fraser University (Canada), Pennsylvania State University (USA), University of California (USA).

2013–2016: Modulation of anti-cancer immune response by regulatory non-coding RNAs — Fondazione Cariplo 2013. I have been in **charge of the bioinformatics WP**. The project had two research units: the National Institute of Molecular Genetics and the Univ. Milano – Bicocca.

During this project, I have been the **scientific supervisor of two 1-year post-docs** and I have co-supervised a third postdoc.

2011–2014: MIUR/Regione Lombardia 2011 (Univ. Milano – Bicocca amount: 199 991€). Piattaforma di Analisi TRaslazionale Integrata (PATRI). I have been **in charge of all bioinformatics aspects**.

This project has led to a 1-year postdoc (23 000€) and 3 short-term research contracts (each between 10 000€ and 12 000€). I have co-supervised the postdoc and I have been the only supervisors of all short-term contracts.

2016: Fondo di Ateneo 2016 (12 490€). Modelli computazionali e algoritmi: aspetti teorici e sperimentali, con applicazioni alla Bioinformatica. **PI**.

2015: Fondo di Ateneo 2015 (10 980€). Algoritmi combinatori e modelli di calcolo: aspetti teorici e applicazioni in Bioinformatica. **PI**.

2014: Fondo di Ateneo 2014 (12 186€). Algoritmi e modelli computazionali: aspetti teorici e applicazioni nelle scienze della vita. **PI**.

2013: Fondo di Ateneo 2013 (9 337€). Metodi algoritmici e modelli: aspetti teorici e applicazioni in bioinformatica. **PI**.

2011: Fondo di Ateneo 2011 (4 055€). Tecniche algoritmiche avanzate in Biologia Computazionale. **PI**.

2006: Grandi Attrezzature 2006 (40 000€). Laboratorio Virtuale Statistico-Territoriale — Virtual Statistical–Territorial Laboratory. This was a joint project between the Statistics Department and the Dept. of Sociology and I have been one of the two people **in charge of the entire project** (and the only one from the Statistics Department). It has resulted in the acquisition of 2 servers that are used to provide services to research and teaching activities on territorial sociology and computational statistics. **co-PI**.

2005–2008: MIUR/PRIN 2005, Potenzialità e ottimizzazione delle banche dati automatizzate in epidemiologia. **In charge of all bioinformatics aspects**.

As a Participant

2013–2016: Regione Lombardia. SPAC3 — Servizi smart della nuova Pubblica amministrazione per la Citizen-Centricity in cloud.

2011–2014: MIUR/PRIN 2011 Automi e Linguaggi Formali: Aspetti Matematici e Applicativi

2003: MIUR/FIRB 2003. Bioinformatica per la Genomica e la Proteomica.

2000–2001: NSF CCR-9988353, ITR-0085910.

1999–2001: MURST COFIN 98 "Bioinformatica e ricerca genomica".

1994–1995: MURST 40% "Algoritmi e strutture di calcolo".

1994–1995: ESPRIT-BRA ASMICS 2 n. 6317.

Teaching

I have started teaching when I have been hired as an Assistant professor at the School of Statistics in 2001. I was the first Computer Science faculty of the school, therefore I had to design new courses tailored to students with a good mathematical background, but with no Computer Science experience. During my career I had to design a course several times, partly due to the evolution of the background of the students and to changes in the overall degree in Statistics and Statistical Economics, partly because I had to teach also to students majoring in Computer Science or in Data Science.

My teaching philosophy is based on active learning. This requires important and continuous interactions between teacher and students, requiring the development of problems that students have to attack, as a group or solo. Designing an active learning course is usually innovative, since existing courses are mostly designed for more traditional approaches, where students have a more passive role and acquiring problem solving skills is not a main focus.

Ph.D. courses taught

2016, 2018, 2020: *Advanced Algorithms*, PhD in Computer Science, Univ. Milano – Bicocca. I have designed the course, since it had never been taught before.

2021: *Graph Theory and Algorithms*, PhD in Computer Science, Univ. Milano – Bicocca. I have co-designed the course, since it had never been taught before.

Courses taught — in charge of the course

2022–today: *Large-Scale Graph Algorithms*, M.Sc. in Computer Science, Univ. Milano – Bicocca (6 ECTS). I have designed the course, since it had never been taught before.

2021–today: *Laboratorio di Informatica (Computer Science Lab)*, B.Sc. in Statistics, B.Sc.

in Statistical Economics, Univ. Milano – Bicocca (3 ECTS). I had fully redesigned the course.

2017-today: *Foundations of Computer Science*, M.Sc. in Data Science, Univ. Milano – Bicocca. (6 ECTS). I have designed the course, since it had never been taught before.

2014-today: *Elementi di Bioinformatica (Elements of Bioinformatics)*, B.Sc. in Computer Science, Univ. Milano – Bicocca (8 ECTS). I have designed the course, since it had never been taught before.

2001–2020: *Basi di Dati (Databases)*, B.Sc. in Statistics, B.Sc. in Statistical Economics, Univ. Milano – Bicocca (6 ECTS).

2007-2022: *Bioinformatica (Bioinformatics)*, M.Sc. in Biostatistics, Univ. Milano – Bicocca (6 ECTS). I have designed the course, since it had never been taught before.

2010–2013: *Algoritmi su stringhe (Text Algorithms)*, B.Sc. in Computer Science, Univ. Milano – Bicocca. I have designed the course, since it had never been taught before.

2008: *Strumenti informatici per la statistica M (Computational Tools for Statistics)*, M.Sc. in Biostatistics, blended e-learning, Univ. Milano – Bicocca. I have designed the course, since it had never been taught before.

2007–2009: *Informatica Applicata S (Applied Computer Science)*, M.Sc. in Biostatistics (2 ECTS), Univ. Milano – Bicocca. I have designed the course, since it had never been taught before.

2001–2008: *Laboratorio Statistico-Informatico (Computational Statistics Lab)*, all B.Sc. programs of the School of Statistics (6 ECTS), Univ. Milano – Bicocca. I have designed the course, since it had never been taught before.

2001–2008: *Programmazione e Basi Dati (Programming and Databases)*, all B.Sc. programs of the School of Statistics (6 ECTS), Univ. Milano – Bicocca. I have designed the course, since it had never been taught before. The course has changed name to "Basi di Dati". It was the first database course in Italy designed for students in Statistics.

2008: *Fondamenti di Informatica (Fundamentals of Computer Science)*, M.Sc. in Biostatistics, Univ. Milano – Bicocca. I have designed the course, since it had never been taught before.

2006: *Laboratorio di Informatica (Computer Science Lab)*, all B.Sc. programs of the School of Statistics (6 ECTS), Univ. Milano – Bicocca. I have designed the course, since it had never been taught before.

Postgraduate courses

2021: *Beyond Genome Assembly*, Master "qOmics: quantitative methods for Omics Data", Univ. Milano – Bicocca and Univ. Pavia.

2012, 2014: *Data Base e Sistemi Informativi (Databases and Information Systems)*, Master di primo livello in Amministratore di Sistema per la Diagnostica per Immagini (Master in Image Diagnostics Systems Administrator), Univ. Milano – Bicocca.

2007, 2010, 2011: *Fondamenti di Informatica (Fundamentals of Computer Science)*, Master di primo livello in Amministratore di Sistema per la Diagnostica per Immagini (Master in Image Diagnostics Systems Administrator), Univ. Milano – Bicocca.

2003: *Fondamenti di Informatica e Elementi di Programmazione (Fundamentals of Computer Science and Elements of Programming)*, Master di primo livello in Bioinformatica (Master in Bioinformatics), Univ. Milano – Bicocca.

2001–2002: *Sistemi Informatici e Elementi di Programmazione (Introduction to Program-*

ming), Master di primo livello in Bioinformatica (Master in Bioinformatics), Univ. Milano – Bicocca.

Supervision

Postdocs

I have been **scientific supervisor** of the following postdoctoral positions:

2022–today: Simone Ciccolella. 2-year postdoc position on *Computational Approaches for tumor and viral phylogenies*.

2018–2019: Murray Patterson. 2-year postdoc position on *Haplotype assembly from sequencing reads*. Currently Assistant Professor (Tenure-track) at Georgia State University.

2015: Hassan Mahmoud Mohamed Ramadan Mohamed. 1-year postdoc position on *Methodology for treatment and data analysis of NGS data for the detection of alternate splicing events*.

Moreover, I have co-supervised the following postdocs:

○ Luca Denti

○ Marco Previtali

○ Stefano Beretta

PhD students supervision

Supervisor of

2021–2024 (expected): Jorge Avila Cartes — PhD in Computer Science

2017–2021: Simone Ciccolella, "*Algorithms for cancer phylogeny inference*" — PhD in Computer Science. Currently Post-doc at Università di Milano – Bicocca.

2013–2017: Marco Previtali, "*Self-indexing for de novo assembly*" — PhD in Computer Science. Currently at Bloomberg

2008–2012: Stefano Beretta, "*Algorithms for Next Generation Sequencing Data Analysis*" — PhD in Computer Science. Currently at San Raffaele Telethon Institute for Gene Therapy

2001–2004: Riccardo Dondi, "*Computational Problems in the Study of Genomic Variations*" — PhD in Computer Science. Currently Associate Professor at Università di Bergamo

Moreover, I have actively contributed to the training of the following PhD students:

○ Giulia Bernardini — PhD in Computer Science. Currently Assistant Professor at Univ. Trieste.

○ Simone Zaccaria — PhD in Computer Science. Currently Group Leader presso Department of Oncology, Univ. College London

○ Anna Paola Carrieri — PhD in Computer Science. Currently Research Staff Member presso IBM Research Lab, UK

M.Sc. students supervision

I have supervised or co-supervised the thesis of more than 10 M.Sc. students in Computer Science, Statistical Economics, Biostatistics, Data Science at University of Milano–Bicocca and at the University of Milan. I am currently supervising the thesis of a M.Sc. in Bioinformatics at the Pwani University Biosciences Research Centre, Kilifi (Kenya), as the result of a long-standing collaboration with Pjotr Prins (University of Tennessee).

I have supervised the activities related to the **Exchange mobility EXTRA UE** program of:

- ○ Ramesh Rajaby ha spent 6 months with the research group led by prof. Jesper Jansson (Kyoto University, Japan). Ramesh Rajaby is currently postdoc at National University of Singapore.
- ○ Simone Ciccolella has spent 3 months with the research group led by prof. Iman Hajirasouliha (Weill Cornell Medicine, New York). Simone Ciccolella is currently a postdoc under my supervision.

B.Sc. students supervision

I have supervised the final projects of more than 50 B.Sc. students in Computer Science, Statistics, and Statistical Economics.

Awards

Best Poster Awards

Recomb 2018: Luca Denti, Raffaella Rizzi, Stefano Beretta, Gianluca Della Vedova, Marco Previtali, and Paola Bonizzoni. *ASGAL: Aligning RNA-Seq Data to a Splicing Graph to Detect Novel Alternative Splicing Events*

Professional activities

Steering Committee Member

2018-today: **Steering Committee member** of Computability in Europe (CiE). During the same period I have also been the Executive Officer of the Steering Committee.

Program Committee Chair

2023: **Program Committee Chair** of Computability in Europe (CiE) 2023. This international conference will take place in Batumi (Georgia) in July 24-28 and is the flagship conference of the Association Computability in Europe that has more than 1000 members and focuses on computability science in all its multidisciplinary facets. The conference proceedings will be published in the LNCS series.

2021: **Chair** of the Data Structure in Bioinformatics Workshop (DSB 2021), February 11-12, 2021. The workshop had 19 research talks given from international scholars.

2021: **Organizer**, of Pangenome Bio Hacking (PGBH) 2021 (co-organized with Erik Garrison, Enza Colonna, Pjotr Prins). This online conference has taken place on December 9-10, 2021, and we have invited 9 speakers from Europe and the US. It has been the first world-wide conference for researchers who have a focus on pangenomics free and open source software development.

2020: **Organizer**, of the Computability in Europe special session "*Large Scale Bioinformatics and Computational Sciences*" (co-organized with Iman Hajirasouliha, Weill-Cornell Medicine, USA). We have selected and invited four speakers who have given an invited talk.

PhD schools Board

2022: **Supervisory Board Member**, PhD Summer School "Introduction to Pangenomics", held at the Lake Como School of Advanced Studies, July 4-8, 2022. The school has been attended by 25 international PhD students and had 12 teachers from Europe and the US. The Supervisory Board had 3 members who have decided the full program of the school.

Program Committee Member

Program Committee Member:
○ Computability in Europe (CiE) 2013, 2019, 2020, 2022
○ Workshop on Algorithms in Bioinformatics (WABI) 2020, 2022
○ ISCB European Conference on Computational Biology (ECCB 2019)
○ Combinatorial Pattern Matching (CPM 2019)
○ Symposium on String Processing and Information Retrieval (SPIRE 2017)
○ Bioinformatics Open Source Conference (BOSC) from 2016 to 2021.

2020: **Editor** of the proceedings of the conference Computability in Europe 2020. The proceedings have been published in the LNCS 12098 volume, Springer-Verlag.

Editorial Board Member

2016–2019: **Editorial Board** member of the journal "Advances in Bioinformatics".

2004: **special issue editor**, Journal of Computer Science and Technology.

PhD defence committee member

2018: Evaluation committee member, PhD thesis in Computer Science, Mattia Gastaldello, Univ. Roma la Sapienza.

2018: Evaluation committee member, PhD thesis in Computer Science, Luca Ferrari, Univ. Milano.

Competition evaluation member

2017: Evaluation committee member, competition for an Assistant Professor position (valutazione comparativa per RTDb), Univ. Milano

2006: Evaluation committee member, competition for an Assistant Professor position (valutazione comparativa per ricercatore universitario), Univ. Milano

Incoming visits

2018: I have been in charge of hosting Prof. Vladimir Filipovic (Associate Professor at University of Belgrade, Faculty of Mathematics, Department of Computer Science) during his sabbatical from 20/01/2018 to 30/09/2018. During his visit, we have collaborated on the development of metaheuristics in Bioinformatics

Reviewer Service

I have reviewed papers for the following journals:
○ ACM/IEEE Transactions on Computational Biology and Bioinformatics
○ Algorithmica
○ Algorithms
○ Bioinformatics
○ Briefings in Bioinformatics
○ Graphs and Combinatorics
○ Information Processing Letters
○ Journal of Computational Biology
○ INFORMS J. Computing
○ Journal of Computer Science and Technology
○ Theoretical Computer Science
○ Theory of Computing Systems

University Service

2021-today: **Scientific Committee Member** Postgraduate Master Programme "quantitative methods for Omics Data" (EQF level 8), University of Milano – Bicocca

2020-today: **Representative of the Università degli Studi di Milano – Bicocca** Joint Research Unit ELIXIR IIB General Assembly, Italian Node of Elixir Europe, the main intergovernative Europan organization in Bioinformatics.

2019-today: **Deputy coordinator of the PhD program** in Computer Science, Univ. Milano – Bicocca

2018-today: **Quality Assurance**, M.Sc. in Data Science.

2019-today: **President of the Teaching Committee (Commissione Didattica)**, M.Sc. in Data Science.

2018-today: **PI** of Università degli Studi di Milano – Bicocca of the MoU (Convenzione quadro) with Istituto Nazionale di Genetica Molecolare (National Institute of Molecular Genetics).

2020-today: **Scientific Committee Member**, Bicocca Ambiente Società Economia, by decree of the rector.

2018-today: Member of the **Students-Professors Committee (Commissione Paritetica Docenti-Studenti)**, Dipartimento di Informatica, Sistemistica e Comunicazione

2013-today: **Board Member**, PhD program in Computer Science, Univ. Milano – Bicocca

2011-today: **Executive Committee (Comitato direttivo)** Member, Centro di Produzione Multimediale di Ateneo (Multimedia Production Center), by decree of the Academic Senate.

2015–2020: Local Technical Coordinator, Univ. Milano – Bicocca, Elixir IIB

2016-today: **In charge of Elixir IIB training activities** in Univ. Milano – Bicocca. I have organized course on "Genome Assembly and Annotation", "Data Carpentry Workshop", "Software Carpentry Workshop", "Docker Advanced", and "Exome analysis with Galaxy".

2002–2012: **Representative of the Computer Science Area**, School of Statistics.

2004–2012: **Representative of the School of Statistics**, University Committee on Computer Science.

2010–2012: **Delegate of the Dean on e-learning** School of Statistics.

2007–2012: **In charge of the Computer Labs** School of Statistics.

2010–2012: Member of the e-learning committee, School of Statistics.

2002–2004: **Representative of the Computer Science Area**, School of Statistics Committee in charge of designing the B.Sc. program in Statistics (Statistica e Gestione delle Informazioni). The program is still active.

2004: **Representative** of the School of Statistics for the short course "Information Technology For Problem Solving (IT4PS)", organized by the CRUI Foundation.

2003: **In charge** of the course 167388 "Laboratorio Complementare di Informatica per Statistici (Elective Computer Science Lab for Statistics)", part of the FSE 156165 project "Progetto Quadro Università degli Studi di Milano – Bicocca". The course focused on intermediate Computer Science skills for statistics students. The course has been fully funded by the European Commission.

Speaker

GGI Seminar Series, March 2022: Invited by Department of Genetics, Genomics and Informatics, University of Tennessee. Title of the talk: "Computational Aspects of Models of Evolution"

invited speaker: Computability in Europe 2017 special session "Algorithmics for Biology"

speaker at ISMB 2001: I have presented a paper as first author — and the only author of my University — to the conference Intelligent Systems for Molecular Biology 2001: the main Bioinformatics scientific conference (**CORE: A, Microsoft Academic: A+**)

Research activity

I have published more than **50 journal papers** with more than **1800 citations** and **h-index 20** (from Google Scholar). According to Google Scholar, 7 of my papers have at least 100 citations each, and I have 166 coauthors.

From 2016 to 2019 I have been the founder and the leader of the Research Lab "AlgoLab — Experimental Algorithmics Lab", at the Università di Milano – Bicocca. The research lab has several international collaborations and it focuses on algorithm design, implementation, and experimental analysis on large datasets. In fact, the research activities regards both methodological and experimental aspects, leading to the development of software tools to attack bioinformatics problems. In 2019 the lab has merged with the Bioinformatics Lab, resulting in the Bionformatics and Experimental Algorithmics (BIAS) Lab which I am effectively co-leading with Paola Bonizzoni.

My personal research activity has focused on the development of combinatorial algorithms in Bioinformatics, with a strong attention on foundational aspects. Moreover, I have paid special attention to algorithm implementation and validation, from both a theoretical and an experimental point of view.

Bioinformatics algorithmics has gained relevance with the establishment of the Human Genome project whose main goal is to determine the influence that biological sequences (such as DNA, RNA) have on living beings. Completing the sequencing of the human genome was only the first fundamental step; finding which proteins are expressed by each gene and determining the interaction among the various DNA sites are among the most important open problems. The huge quantity of data to analyze (remember that human DNA contains 3 billions nucleotides) makes computers and efficient algorithms a needed cornerstone of the field. My research has been mainly devoted to the design of such efficient algorithms and can be detailed in a few subfields.

Sequence comparison

The central dogma in Computational Biology states that sequence homology leads to functional homologies (that is similarity among the effects performed by such sequences), therefore it is of the utmost importance to have some computational tools for comparing sequences. In this direction the multiple sequence alignment problem has been formally introduced in (Altschul, Lipman SIAM J. Appl. Math. 1989), even though it had already been studied previously, and some of its variant are intractable. My focus is on the study of the computational and approximation complexity and on the development of efficient algorithms for the classical notions of longest common subsequence and shortest common supersequence and for sequence alignment.

Most recently, my focus is on computational pangenomics, where a large set of genomes is considered and represented as a (pangeome) graph. In this case, my main contribution has been the introduction of a formal notion of the computational problem of pangenome graph construction, which in turn allows to analyze existing and new algorithms for the problem. Moreover, I studied efficient and practical sequence-to-graph alignment algorithms that fully exploit the topology of the graphs to infer recombination events — this problem was not previously attacked with approaches based on sequence alignment.

Journal papers
[1] [4] [36] [39] [43] [52] [53]

Software
○ RecGraph `https://github.com/AlgoLab/RecGraph`

Collaborations
○ Universitaet Bielefeld (Alexander Schoenhuth)
○ University of California at Santa Cruz (Jouni Siren)

Phylogeny reconstruction and comparison

Reconstructing phylogenies is another problem that has great relevance in Computational Biology, as a phylogeny is an intuitive representation of a common evolutionary history of a set of extant species. In this setting I have studied the quartet-based reconstruction technique (a quartet is the optimal solution over four species), developing new algorithms to clean some of the error that inherent in the use of such technique.

More recently, I have studied the algorithmic implications of new models of evolution, especially for tumor and viral phylogeny. In fact, most of the literature focuses on a very restrictive model of evolution where mutations can only be acquired. Cancer evolution is too complex for that model, in fact losses of entire regions of the genome is common. Unfortunately, this makes the computational problem much harder, since the solution space explodes. My contributions on these aspects spans from mathematical exploration of combinatorial properties of phylogenies to efficient and practical algorithms to infer phylogenies, leading their design, development, and analysis.

The phylogeny comparison problem is fundamental when you have to compare the results of various experiments on the same set of species. I have extensively studied the problem In (Amir e Keselman, SIAM J. Comp. 1997) some formulations of the problem have been introduced. My initial contributions have been on the approximability of the maximum common isomorphic subtree, and on efficient algorithms to reconcile gene and species trees. Most recently, I have shifted my focus on practical similarity measures between tumor phylogenies. In fact, there are several tools for inferring tumor phylogenies, but the research community needs a measure to understand which of the phylogenies are more similar then others.

Journal papers
[3] [5] [8] [9] [10] [16] [19] [24] [40] [48] [51]

Software
○ SASC `https://github.com/sciccolella/sasc`
○ MP3 `https://github.com/AlgoLab/mp3treesim`
○ ggpf `https://github.com/AlgoLab/gppf`
○ ggps `https://github.com/AlgoLab/gpps`

Collaborations
○ Georgia State University (Murray Patterson)
○ Weill Cornell Medicine (Iman Hajirasouliha)

Clustering

The problem of classifying data in similar sets is one of the most important problems in Computer Science; it is common to have a similarity measure between pairs of elements and to aim at computing a partition of the elements so that elements in a common sets are similar while elements in different sets are not similar. In this field I have studied the correlation clustering problem on weighted graphs, which has the important property that the number of clusters is not fixed a priori, but depends on the dataset analyzed. My contribution has been to develop the connections with another version of clustering, called consensus clustering, showing that an interesting restriction of the problem is NP-complete and providing two polynomial-time approximation schemes for a different formulation of the problem.

Moreover I have studied a different clustering problem, closely related to the analysis of microarray data. In this case, data are represented as vectors on $\{0, 1, N\}$ alphabet, where N stands for missing or undecided data. His contributions consist of proving that some restrictions of the problem are APX-hard and designing a polynomial-time constant-factor approximation algorithm. Moreover, I have designed an efficient algorithm for a different restriction of the problem.

Another research topic related to Clustering is the k-anonymity problem, where we want to cluster the rows of a matrix so that each cluster has identical entries, except for k columns, where I have designed some algorithms that are efficient for small values of k. This formulation has originally appeared in data privacy, but I am investigating its applicability to pangenome privacy.

Tumor phylogeny inference methods are especially computationally expensive. For this reason, I have developed some clustering techniques that are tailored for clustering single-cell tumoral data. These new approaches are based on community detection, and I have proved experimentally that they result in faster and more precise phylogenies.

Journal papers

[7] [27] [29] [33] [34] [38]

Software

○ plastic `https://github.com/plastic-phy/plastic`
○ CouGaR-g `https://github.com/AlgoLab/CouGaR-g`
○ MALVIRUS `https://algolab.github.io/MALVIRUS/`
○ celluloid `https://github.com/AlgoLab/celluloid`

Collaborations

○ Georgia State University (Murray Patterson)

Alternative Splicing

Alternative splicing is the biological mechanism that allows a gene to encode and produce more than one protein and it is correlated to the onset of several diseases. In this field I have focused on efficient algorithms for detecting novel (that is not previously known) alternative splicing events.

I have supervised the development of two approaches to infer novel alternative splicing events, based on the alignment of reads against a transcriptome. In the second approach, the alignment is against a graph transcriptome which has been one of the very first papers introducing graph-based representation of a set of transcripts. I have supervised the algorithm design and the experimental analysis.

The sheer amount of data requires some efficient filtering. I have developed a fast and accurate filter, based on Bloom filters, that is able to extract from a sample on the much smaller reads

that originate from a given gene, allowing to use a more precise, albeit slower, downstream tool. I have developed the algorithm and supervised the experimental analysis.

Software

○ PIntron `https://github.com/AlgoLab/PIntron`

○ ASGAL `https://github.com/AlgoLab/galig`

○ Shark

Main collaborations

○ CNR (Graziano Pesole)

Journal papers

[4] [14] [22] [31]

DNA microarray design

The introduction of DNA microarray have greatly increased the throughput of experimental data in Molecular Biology. Such technology (Drmanac et. al. Science 1991) has given relevance to some computational problems on the optimal microarray synthesis or on the experimental data analysis.

More precisely, classifying microbial communities can be performed only exploiting microarrays, as the microbial external aspect is hard to study, due to their extremely small size. A fundamental computational problem in this field is computing the minimum set of substrings that are able to distinguish a set of strings (Probe Selection). The importance of this problem is due to the fact that it formalizes the search for the cheapest experiment obtaining the desired result.

I have proposed, implemented and analyzed some algorithms, while supervising the use of the implementation on some biological data that were previously impossible to analyze. This effort has led to a new protocol for microbial communities analysis.

Journal papers

[32] [46] [47] [49]

Software

○ ProbeSelection

Main collaborations

○ University of California at Riverside (Tao Jiang, James Borneman)

Haplotyping

Several species, including human beings, are diploid, that is each chromosome consists of two distinct copies called haplotypes. Current technological limitations do not allow to cheaply compute those haplotypes, but only genotypes (that is the two nucleotides that are in the same position in those haplotypes). Since it is important to know the actual haplotypes, a number of related computational problems have been recently introduced; those problems compose the field of haplotyping.

I have designed and analyzed an algorithm to complete haplotypes on incomplete data and under the coalescent model (such models forbids some otherwise possible recombinations), where the objective function is the entropy of the solution. The algorithm belongs to the class of Kernighan-Lin heuristics and it has been empirically and favorably compared to the greedy algorithm that was previously routinely employed.

Moreover I have studied the xor-genotyping problem, where the input data contains even less information that in most other haplotyping problems, as it is known only the positions where the two haplotypes differ, and not their contents. This formulation is a faithful model of the

results that can be obtained with a recent and economically viable technology. I have obtained some preliminary results on the computational complexity of the problem, and he has designed some efficient algorithms.

Another problem that I have studied is single-individual haplotype assembly, where we want to determine which variants a person actually has. My contribution on this problem has been the development of a software tool (HapChat) and the supervision of its experimental analysis. This approach has been incorporated into one of the most widely used tools for this problem. Most recently, we have extended this idea to viral data, where we want to quickly and effectively identify the viral strain in a sample. In this case I have supervised the entire work.

Journal papers

[2] [13] [30] [35] [41] [45]

Software

○ HapChat `https://github.com/AlgoLab/HapCHAT`

Graph Algorithms

Graph theory is one of the most important research fields that are common to Computer Science and Discrete Mathematics, as graphs are a mathematical device that is suitable for natural modeling of various real-world problems. One of the techniques that has been widely employed for designing efficient graph algorithm consists of decomposing the graph and then solving the problem on the smaller parts for finally recombining the partial solutions.

I have developed some efficient algorithms for computing the modular decomposition on hypergraphs and k-structures.

Journal papers

[50]

Text Algorithms

The sheer size of genomic data makes important to develop data structures to represent with a small amount of memory while querying them efficiently. The main such data structure is the Burrows-Wheeler Transform (BWT), and I have developed some algorithms to efficiently compute the BWT together with another auxiliary data structure (the Longest Common Prefix array) and for building the string graph (a data structure that represents how some strings overlap with each other). Moreover, I have supervised the implementation and the experimental analysis of those algorithms.

Journal papers

[6] [11] [12] [17] [18] [20]

Software

○ bwt-lcp-parallel `https://github.com/AlgoLab/bwt-lcp-parallel`
○ bwt-lcp-em `https://github.com/AlgoLab/bwt-lcp-em`
○ FSG `https://github.com/AlgoLab/FastStringGraph`
○ LSG `https://github.com/AlgoLab/LightStringGraph`

Research and Technology Transfer

Research and consulting contracts

2003–2004: Consultant on the computational aspects of the project INTERREG IIIB (2000–2006) W.E.S.T. WOMEN EAST SMUGGLING TRAFFICKING (WP.2.2). **Proponent: Fondazione Ismu – Iniziative e Studi sulla Multietnicità**. I have been the only computer

scientist involved in the project.

2005: Consultant on the computational aspects of the project "Indagine Finalizzata all'Analisi degli Effetti Prodotti dai Processi di Regolarizzazione dei Lavoratori Extracomunitari, con Particolare Riferimento al Mercato del Lavoro e all'Integrazione Sociale nelle Regioni Ob. 1", funded under Misura I.2 FESR "Adeguamento del Sistema di Controllo Tecnologico del Territorio", PON Sicurezza 2000/2006. **Proponent: Fondazione Ismu − Iniziative e Studi sulla Multietnicità**. I have been the only computer scientist involved in the project.

Publications

Journal articles..

[1] Jasmijn A. Baaijens, Paola Bonizzoni, Christina Boucher, Gianluca Della Vedova, Yuri Pirola, Raffaella Rizzi, and Jouni Sirén. Computational graph pangenomics: a tutorial on data structures and their applications. *Natural Computing*, March 2022.

[2] Simone Ciccolella, Luca Denti, Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, and Marco Previtali. MALVIRUS: an integrated application for viral variant analysis. *BMC Bioinformatics*, 22(15):625, April 2021.

[3] Sarwan Ali, Simone Ciccolella, Lorenzo Lucarella, Gianluca Della Vedova, and Murray Patterson. Simpler and faster development of tumor phylogeny pipelines. *Journal of Computational Biology*, 28(11):1142–1155, 2021.

[4] Luca Denti, Yuri Pirola, Marco Previtali, Tamara Ceccato, Gianluca Della Vedova, Raffaella Rizzi, and Paola Bonizzoni. Shark: fishing relevant reads in an rna-seq sample. *Bioinformatics*, 37(4):464–472, 2021.

[5] Simone Ciccolella, Camir Ricketts, Mauricio Soto Gomez, Murray Patterson, Dana Silverbush, Paola Bonizzoni, Iman Hajirasouliha, and Gianluca Della Vedova. Inferring cancer progression from single-cell sequencing while allowing mutation losses. *Bioinformatics*, 37(3):326–333, 2021.

[6] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. Computing the multi-string BWT and LCP array in external memory. *Theoretical Computer Science*, 862:42–58, 2021.

[7] Simone Ciccolella, Murray Patterson, Paola Bonizzoni, and Gianluca Della Vedova. Effective clustering for single cell sequencing cancer data. *IEEE J. Biomed. Health Informatics*, 25(11):4068–4078, 2021.

[8] Simone Ciccolella, Giulia Bernardini, Luca Denti, Paola Bonizzoni, Marco Previtali, and Gianluca Della Vedova. Triplet-based similarity score for fully multilabeled trees with poly-occurring labels. *Bioinformatics*, 37(2):178–184, 2021.

[9] Simone Ciccolella, Mauricio Soto Gomez, Murray D. Patterson, Gianluca Della Vedova, Iman Hajirasouliha, and Paola Bonizzoni. gpps: an ILP-based approach for inferring cancer progression with mutation losses from single cell data. *BMC Bioinformatics*, 21-S(1):413, 2020.

[10] Paola Bonizzoni, Simone Ciccolella, Gianluca Della Vedova, and Mauricio Soto Gomez. Does relaxing the infinite sites assumption give better tumor phylogenies? an ILP-based comparative approach. *IEEE ACM Trans. on Computational Biology and Bioinformatics*, 16(5):1410–1423, 2019.

[11] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. Multithread multistring Burrows-Wheeler transform and longest common prefix array. *Journal of Computational Biology*, 26(9):948–961, 2019.

[12] Raffaella Rizzi, Stefano Beretta, Murray Patterson, Yuri Pirola, Marco Previtali, Gianluca Della Vedova, and Paola Bonizzoni. Overlap graphs and de Bruijn graphs: data structures for de novo genome assembly in the big data era. *Quantitative Biology*, 7(4):278–292, 2019.

[13] Stefano Beretta, Murray Patterson, Simone Zaccaria, Gianluca Della Vedova, and Paola Bonizzoni. HapCHAT: adaptive haplotype assembly for efficiently leveraging high coverage in long reads. *BMC Bioinformatics*, 19(1):252:1–252:19, 2018.

[14] Luca Denti, Raffaella Rizzi, Stefano Beretta, Gianluca Della Vedova, Marco Previtali, and Paola Bonizzoni. ASGAL: aligning rna-seq data to a splicing graph to detect novel alternative splicing events. *BMC Bioinformatics*, 19(1):444:1–444:21, 2018.

[15] G. Ukmar, G. E. M. Melloni, L. Raddrizzani, P. Rossi, S. Di Bella, M. R. Pirchio, M. Vescovi, A. Leone, M. Callari, M. Cesarini, A. Somaschini, Gianluca Della Vedova, M. G. Daidone, M. Pettenella, A. Isacchi, and R. Bosotti. PATRI, a Genomics Data Integration Tool for Biomarker Discovery. *BioMed Research International*, 2018.

[16] Paola Bonizzoni, Anna Paola Carrieri, Gianluca Della Vedova, Raffaella Rizzi, and Gabriella Trucco. A colored graph approach to perfect phylogeny with persistent characters. *Theoretical Computer Science*, 658:60–73, 2017.

[17] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. An external-memory algorithm for string graph construction. *Algorithmica*, 78(2):394–424, 2017.

[18] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. FSG: fast string graph construction for de novo assembly. *Journal of Computational Biology*, 24(10):953–968, 2017.

[19] Paola Bonizzoni, Anna Paola Carrieri, Gianluca Della Vedova, Raffaella Rizzi, and Gabriella Trucco. Species-driven persistent phylogeny. *Fundamenta Informaticae*, 154(1-4):47–63, 2017.

[20] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. LSG: an external-memory tool to compute string graphs for next-generation sequencing data assembly. *Journal of Computational Biology*, 23(3):137–149, 2016.

[21] Jon C. Ison, Kristoffer Rapacki, Hervé Ménager, Matús Kalas, Emil Rydza, Piotr Chmura, Christian Anthon, Niall Beard, Karel Berka, Dan M. Bolser, Tim Booth, Anthony Bretaudeau, Jan Brezovsky, Rita Casadio, Gianni Cesareni, Frederik Coppens, Michael Cornell, Gianmauro Cuccuru, Kristian Davidsen, Gianluca Della Vedova, Tunca Dogan, Olivia

Doppelt-Azeroual, Laura Emery, Elisabeth Gasteiger, Thomas Gatter, Tatyana Goldberg, Marie Grosjean, Björn A. Grüning, Manuela Helmer-Citterich, Hans Ienasescu, Vassilios Ioannidis, Martin Closter Jespersen, Rafael C. Jimenez, Nick S. Juty, Peter Juvan, Maximilian Koch, Camille Laibe, Jing-Woei Li, Luana Licata, Fabien Mareuil, Ivan Micetic, Rune Møllegaard Friborg, Sébastien Moretti, Chris Morris, Steffen Möller, Aleksandra Nenadic, Hedi Peterson, Giuseppe Profiti, Peter M. Rice, Paolo Romano, Paola Roncaglia, Rabie Saidi, Andrea Schafferhans, Veit Schwämmle, Callum Smith, Maria Maddalena Sperotto, Heinz Stockinger, Radka Svobodová Vareková, Silvio C. E. Tosatto, Victor de la Torre, Paolo Uva, Allegra Via, Guy Yachdav, Federico Zambelli, Gert Vriend, Burkhard Rost, Helen E. Parkinson, Peter Løngreen, and Søren Brunak. Tools and data services registry: a community effort to document bioinformatics resources. *Nucleic Acids Research*, 44(Database-Issue):38–47, 2016.

[22] Stefano Beretta, Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, and Raffaella Rizzi. Modeling alternative splicing variants from RNA-Seq data with isoform graphs. *Journal of Computational Biology*, 21(1):16–40, 2014.

[23] Arianna Ghirardi, Mauro Di Bari, Antonella Zambon, Lorenza Scotti, Gianluca Della Vedova, Francesco Lapi, Francesco Cipriani, Achille P. Caputi, Alberto Vaccheri, Dario Gregori, Rosaria Gesuita, Annarita Vestri, Tommaso Staniscia, Giampiero Mazzaglia, and Giovanni Corrao. Effectiveness of oral bisphosphonates for primary prevention of osteoporotic fractures. *European Journal of Clinical Pharmacology*, 70(9):1129–1137, September 2014.

[24] Paola Bonizzoni, Anna Paola Carrieri, Gianluca Della Vedova, and Gabriella Trucco. Explaining evolution via constrained persistent perfect phylogeny. *BMC Genomics*, 15(Suppl 6):S10, October 2014.

[25] A. Ghirardi, L. Scotti, G.D. Vedova, L.C. D'Oro, F. Lapi, F. Cipriani, A.P. Caputi, A. Vaccheri, D. Gregori, R. Gesuita, A. Vestri, T. Staniscia, G. Mazzaglia, G. Corrao, E. Sessa, V. Arcoraci, A. Parodi, A. Zambon, N. Montanaro, C. Piccinni, C. Suzzi, A. Puccini, M. Sturkenboom, P. Geppetti, L. Sati, Bari Di, F. Forlan, F. Carle, A. Menna, and M. Valenti. Oral bisphosphonates do not increase the risk of severe upper gastrointestinal complications: A nested case-control study. *BMC Gastroenterology*, 14(1), 2014.

[26] Giovanni Corrao, Arianna Ghirardi, Giulia Segafredo, Antonella Zambon, Gianluca Della Vedova, Francesco Lapi, Francesco Cipriani, Achille Caputi, Alberto Vaccheri, Dario Gregori, Rosaria Gesuita, Annarita Vestri, Tommaso Staniscia, Giampiero Mazzaglia, Mauro Di Bari, and on behalf of the BEST investigators. User-only design to assess drug effectiveness in clinical practice: application to bisphosphonates and secondary prevention of fractures. *Pharmacoepidemiology and Drug Safety*, pages n/a–n/a, June 2014.

[27] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Yuri Pirola. Parameterized complexity of k-anonymity: hardness and tractability. *Journal of Combinatorial Optimization*, 26(1):19–43, 2013.

[28] Arianna Ghirardi, Lorenza Scotti, Antonella Zambon, Gianluca Della Vedova, Luca Cavalieri D'oro, Francesco Lapi, Francesco Cipriani, Achille P. Caputi, Alberto Vaccheri, Dario Gregori, Rosaria Gesuita, Annarita Vestri, Tommaso Staniscia, Giampiero Mazzaglia, Giovanni Corrao, and on behalf of the BEST Investigators. Risk of Severe Upper Gastrointestinal Complications among Oral Bisphosphonate Users. *PLoS ONE*, 8(12):e73159, December 2013.

[29] Paola Bonizzoni, Gianluca Della Vedova, and Riccardo Dondi. A randomized PTAS for the minimum consensus clustering with a fixed number of clusters. *Theoretical Computer Science*, 429:36–45, 2012.

[30] Yuri Pirola, Gianluca Della Vedova, Stefano Biffani, Alessandra Stella, and Paola Bonizzoni. A fast and practical approach to genotype phasing and imputation on a pedigree with erroneous and incomplete information. *IEEE ACM Trans. on Computational Biology and Bioinformatics*, 9(6):1582–1594, 2012.

[31] Yuri Pirola, Raffaella Rizzi, Ernesto Picardi, Graziano Pesole, Gianluca Della Vedova, and Paola Bonizzoni. Pintron: a fast method for detecting the gene structure due to alternative splicing via maximal pairings of a pattern and a text. *BMC Bioinformatics*, 13(S-5):S2, 2012.

[32] Paul M. Ruegger, Gianluca Della Vedova, Tao Jiang, and James Borneman. Improving probe set selection for microbial community analysis by leveraging taxonomic information of training sequences. *BMC Bioinformatics*, 12:394, 2011.

[33] Paola Bonizzoni, Gianluca Della Vedova, and Riccardo Dondi. Anonymizing binary and small tables is hard to approximate. *Journal of Combinatorial Optimization*, 22(1):97–119, 2011.

[34] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Giancarlo Mauri. Fingerprint clustering with bounded number of missing values. *Algorithmica*, 58(2):282–303, 2010.

[35] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, Yuri Pirola, and Romeo Rizzi. Pure parsimony xor haplotyping. *IEEE ACM Trans. on Computational Biology and Bioinformatics*, 7(4):598–610, 2010.

[36] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Yuri Pirola. Variants of constrained longest common subsequence. *Information Processing Letters*, 110(20):877–881, 2010.

[37] Gianluca Della Vedova, Riccardo Dondi, Tao Jiang, Giulio Pavesi, Yuri Pirola, and Lusheng Wang. Beyond evolutionary trees. *Natural Computing*, 9(2):421–435, 2010.

[38] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Tao Jiang. On the approximation of correlation clustering and consensus clustering. *Journal of Computer and Systems Science*, 74(5):671–696, 2008.

[39] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, Guillaume Fertin, Raffaella Rizzi, and Stéphane Vialette. Exemplar longest common subsequence. *IEEE ACM Trans. on Computational Biology and Bioinformatics*, 4(4):535–543, 2007.

[40] Paola Bonizzoni, Gianluca Della Vedova, and Riccardo Dondi. Reconciling a gene tree to a species tree under the duplication cost model. *Theoretical Computer Science*, 347(1-2):36–53, 2005.

[41] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Lorenzo Mariani. Experimental analysis of a new algorithm for partial haplotype completion. *International Journal of Bioinformatics Research and Applications*, 1(4):461–473, 2005.

[42] G. Cattaneo, Gianluca Della Vedova, A. Leporati, and R Leporini. Towards a Theory of Conservative Computing. *International Journal of Theoretical Physics*, 44:861–873, January 2005.

[43] Winfried Just and Gianluca Della Vedova. Multiple sequence alignment as a facility-location problem. *INFORMS Journal on Computing*, 16(4):430–440, 2004.

[44] Gianluca Della Vedova and Riccardo Dondi. A library of efficient bioinformatics algorithms. *Applied Bioinformatics*, 2(2):117–121, 2003.

[45] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Jing Li. The haplotyping problem: An overview of computational models and solutions. *Journal of Computer Science and Technology*, 18(6):675–688, 2003.

[46] L. Valinsky, G. Della Vedova, A. J. Scupham, S. Alvey, A. Figueroa, B. Yin, R. J. Hartin, M. Chrobak, D. E. Crowley, T. Jiang, and J. Borneman. Analysis of Bacterial Community Composition by Oligonucleotide Fingerprinting of rRNA Genes. *Applied and Environmental Microbiology*, 68(7):3243–3250, July 2002.

[47] L. Valinsky, Gianluca Della Vedova, T. Jiang, and J Borneman. Oligonucleotide Fingerprinting of rRNA Genes for Analysis of Fungal Community Composition. *Applied and Environmental Microbiology*, 68(12):5999–6004, January 2002.

[48] Gianluca Della Vedova and Todd Wareham. Optimal algorithms for local vertex quartet cleaning. *Bioinformatics*, 18(10):1297–1304, 2002.

[49] J. Borneman, M. Chrobak, Gianluca Della Vedova, A. Figueroa, and T Jiang. Probe selection Algorithms with applications in the analysis of microbial communities. *Bioinformatics*, 17(Suppl 1):S39–S48, January 2001.

[50] Paola Bonizzoni and Gianluca Della Vedova. An algorithm for the modular decomposition of hypergraphs. *Journal of Algorithms*, 32(2):65–86, 1999.

[51] Paola Bonizzoni, Gianluca Della Vedova, and Giancarlo Mauri. Approximating the maximum isomorphic agreement subtree is hard. *International Journal on the Foundations of Computer Science*, 11(4):579–590, 2000.

[52] Paola Bonizzoni, Gianluca Della Vedova, and Giancarlo Mauri. Experimenting an approximation algorithm for the LCS. *Discrete Applied Mathematics*, 110(1):13–24, 2001.

[53] Paola Bonizzoni and Gianluca Della Vedova. The complexity of multiple sequence alignment with SP-score that is a metric. *Theoretical Computer Science*, 259(1-2):63–79, 2001.

Conference proceedings

[1] Prakash Chourasia, Sarwan Ali, Simone Ciccolella, Gianluca Della Vedova, and Murray Patterson. Clustering SARS-CoV-2 Variants from Raw High-Throughput Sequencing Reads Data. In Mukul S. Bansal, Ion Măndoiu, Marmar Moussa, Murray Patterson, Sanguthevar Rajasekaran, Pavel Skums, and Alexander Zelikovsky, editors, *Computational Advances in Bio and Medical Sciences*, volume 13254 of *LNCS*, pages 133–148. Springer International Publishing, 2022.

[2] Simone Ciccolella, Murray D. Patterson, Paola Bonizzoni, and Gianluca Della Vedova. Effective clustering for single cell sequencing cancer data. In Xinghua Mindy Shi, Michael Buck, Jian Ma, and Pierangelo Veltri, editors, *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, BCB 2019, Niagara Falls, NY, USA, September 7-10, 2019*, pages 437–446. ACM, 2019.

[3] Giulia Bernardini, Paola Bonizzoni, Gianluca Della Vedova, and Murray Patterson. A rearrangement distance for fully-labelled trees. In Nadia Pisanti and Solon P. Pissis, editors, *30th Annual Symposium on Combinatorial Pattern Matching, CPM 2019, June 18-20, 2019, Pisa, Italy*, volume 128 of *LIPIcs*, pages 28:1–28:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[4] Paola Bonizzoni, Gianluca Della Vedova, Serena Nicosia, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. Divide and conquer computation of the multi-string BWT and LCP array. In Florin Manea, Russell G. Miller, and Dirk Nowotka, editors, *Sailing Routes in the World of Computation - 14th Conference on Computability in Europe, CiE 2018, Kiel, Germany, July 30 - August 3, 2018, Proceedings*, volume 10936 of *Lecture Notes in Computer Science*, pages 107–117. Springer, 2018.

[5] Simone Ciccolella, Mauricio Soto Gomez, Murray Patterson, Gianluca Della Vedova, Iman Hajirasouliha, and Paola Bonizzoni. gpps: an ILP-based approach for inferring cancer progression with mutation losses from single cell data. In *8th IEEE International Conference on Computational Advances in Bio and Medical Sciences, ICCABS 2018, Las Vegas, NV, USA, October 18-20, 2018*, page 1. IEEE Computer Society, 2018.

[6] Paola Bonizzoni, Simone Ciccolella, Gianluca Della Vedova, and Mauricio Soto Gomez. Beyond perfect phylogeny: Multisample phylogeny reconstruction via ILP. In Nurit Haspel, Lenore J. Cowen, Amarda Shehu, Tamer Kahveci, and Giuseppe Pozzi, editors, *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB 2017, Boston, MA, USA, August 20-23, 2017*, pages 1–10. ACM, 2017.

[7] Gianluca Della Vedova, Murray Patterson, Raffaella Rizzi, and Mauricio Soto Gomez. Character-based phylogeny construction and its application to tumor evolution. In Jarkko Kari, Florin Manea, and Ion Petre, editors, *Unveiling Dynamics and Complexity - 13th Conference on Computability in Europe, CiE 2017, Turku, Finland, June 12-16, 2017, Proceedings*, volume 10307 of *Lecture Notes in Computer Science*, pages 3–13. Springer, 2017.

[8] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. FSG: fast string graph construction for de novo assembly of reads data. In Anu G. Bourgeois, Pavel Skums, Xiang Wan, and Alex Zelikovsky, editors, *Bioinformatics Research and Applications - 12th International Symposium, ISBRA 2016, Minsk, Belarus, June 5-8, 2016, Proceedings*, volume 9683 of *Lecture Notes in Computer Science*, pages 27–39. Springer, 2016.

[9] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. Constructing string graphs in external memory. In Daniel G. Brown and Burkhard Morgenstern, editors, *Algorithms in Bioinformatics - 14th International Workshop, WABI 2014, Wroclaw, Poland, September 8-10, 2014. Proceedings*, volume 8701 of *Lecture Notes in Computer Science*, pages 311–325. Springer, 2014.

[10] Yuri Pirola, Gianluca Della Vedova, Paola Bonizzoni, Alessandra Stella, and Filippo Biscarini. Haplotype-based prediction of gene alleles using pedigrees and SNP genotypes. In Jing Gao, editor, *ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics. ACM-BCB 2013, Washington, DC, USA, September 22-25, 2013*, page 33. ACM, 2013.

[11] Stefano Beretta, Paola Bonizzoni, Raffaella Rizzi, and Gianluca Della Vedova. Reconstructing isoform graphs from rna-seq data. In *2012 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2012, Philadelphia, PA, USA, October 4-7, 2012*, pages 1–4. IEEE Computer Society, 2012.

[12] Yuri Pirola, Gianluca Della Vedova, Stefano Biffani, Alessandra Stella, and Paola Bonizzoni. A fast and practical approach to genotype phasing and imputation on a pedigree with erroneous and incomplete information. In Sorin Istrail, Ion I. Mandoiu, Mihai Pop, Sanguthevar Rajasekaran, and John L. Spouge, editors, *IEEE 2nd International Conference on Computational Advances in Bio and Medical Sciences, ICCABS 2012, Las Vegas, NV, USA, February 23-25, 2012*, pages 1–6. IEEE Computer Society, 2012.

[13] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, and Raffaella Rizzi. Pintron: A fast method for gene structure prediction via maximal pairings of a pattern and a text. In Ion I. Mandoiu, Satoru Miyano, Teresa M. Przytycka, and Sanguthevar Rajasekaran, editors, *IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences, ICCABS 2011, Orlando, FL, USA, February 3-5, 2011*, pages 33–39. IEEE Computer Society, 2011.

[14] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Yuri Pirola. Parameterized complexity of k-anonymity: Hardness and tractability. In Costas S. Iliopoulos and William F. Smyth, editors, *Combinatorial Algorithms - 21st International Workshop, IWOCA 2010, London, UK, July 26-28, 2010, Revised Selected Papers*, volume 6460 of *Lecture Notes in Computer Science*, pages 242–255. Springer, 2010.

[15] Paola Bonizzoni, Gianluca Della Vedova, and Riccardo Dondi. The k-anonymity problem is hard. In Miroslaw Kutylowski, Witold Charatonik, and Maciej Gebala, editors, *Fundamentals of Computation Theory, 17th International Symposium, FCT 2009, Wroclaw, Poland, September 2-4, 2009. Proceedings*, volume 5699 of *Lecture Notes in Computer Science*, pages 26–37. Springer, 2009.

[16] Paola Bonizzoni, Gianluca Della Vedova, and Riccardo Dondi. A PTAS for the minimum consensus clustering problem with a fixed number of clusters. In Alessandra Cherubini, Mario Coppo, and Giuseppe Persiano, editors, *Theoretical Computer Science, 11th Italian Conference, ICTCS 2009, Cremona, Italy, September 28-30, 2009, Proceedings*, pages 55–58, 2009.

[17] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, Yuri Pirola, and Romeo Rizzi. Pure parsimony xor haplotyping. In Ion I. Mandoiu, Giri Narasimhan, and Yanqing Zhang, editors, *Bioinformatics Research and Applications, 5th International Symposium, ISBRA 2009, Fort Lauderdale, FL, USA, May 13-16, 2009, Proceedings*, volume 5542 of *Lecture Notes in Computer Science*, pages 186–197. Springer, 2009.

[18] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, Yuri Pirola, and Raffaella Rizzi. Minimum factorization agreement of spliced ESTs. In Steven Salzberg and Tandy J. Warnow, editors, *Algorithms in Bioinformatics, 9th International Workshop, WABI 2009, Philadelphia, PA, USA, September 12-13, 2009. Proceedings*, volume 5724 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2009.

[19] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Giancarlo Mauri. Fingerprint clustering with bounded number of missing values. In Moshe Lewenstein and Gabriel Valiente, editors, *Combinatorial Pattern Matching, 17th Annual Symposium, CPM 2006, Barcelona, Spain, July 5-7, 2006, Proceedings*, volume 4009 of *Lecture Notes in Computer Science*, pages 106–116. Springer, 2006.

[20] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, Guillaume Fertin, and Stéphane Vialette. Exemplar longest common subsequence. In Vassil N. Alexandrov, G. Dick van Albada, Peter M. A. Sloot, and Jack J. Dongarra, editors, *Computational Science - ICCS 2006, 6th International Conference, Reading, UK, May 28-31, 2006, Proceedings, Part II*, volume 3992 of *Lecture Notes in Computer Science*, pages 622–629. Springer, 2006.

[21] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Lorenzo Mariani. Experimental analysis of a new algorithm for partial haplotype completion. In Vaidy S. Sunderam, G. Dick van Albada, Peter M. A. Sloot, and Jack J. Dongarra, editors, *Computational Science - ICCS 2005, 5th International Conference, Atlanta, GA, USA, May 22-25, 2005, Proceedings, Part II*, volume 3515 of *Lecture Notes in Computer Science*, pages 952–959. Springer, 2005.

[22] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Tao Jiang. Correlation clustering and consensus clustering. In Xiaotie Deng and Ding-Zhu Du, editors, *Algorithms and Computation, 16th International Symposium, ISAAC 2005, Sanya, Hainan, China, December 19-21, 2005, Proceedings*, volume 3827 of *Lecture Notes in Computer Science*, pages 226–235. Springer, 2005.

[23] Sergio Pozzi, Gianluca Della Vedova, and Giancarlo Mauri. An explicit upper bound for the approximation ratio of the maximum gene regulatory network problem. In Vincent Danos and Vincent Schächter, editors, *Computational Methods in Systems Biology, International Conference, CMSB 2004, Paris, France, May 26-28, 2004, Revised Selected Papers*, volume 3082 of *Lecture Notes in Computer Science*, pages 1–8. Springer, 2004.

[24] Paola Bonizzoni, Gianluca Della Vedova, and Riccardo Dondi. Reconciling gene trees to a species tree. In Rossella Petreschi, Giuseppe Persiano, and Riccardo Silvestri, editors, *Algorithms and Complexity, 5th Italian Conference, CIAC 2003, Rome, Italy, May 28-30, 2003, Proceedings*, volume 2653 of *Lecture Notes in Computer Science*, pages 120–131. Springer, 2003.

[25] Gianluca Della Vedova and Todd Wareham. Optimal algorithms for local vertex quartet cleaning. In Gary B. Lamont, Hisham Haddad, George A. Papadopoulos, and Brajendra Panda, editors, *Proceedings of the 2002 ACM Symposium on Applied Computing (SAC), March 10-14, 2002, Madrid, Spain*, pages 173–177. ACM, 2002.

[26] Gianluca Della Vedova, Tao Jiang, Jing Li, and Jianjun Wen. Approximating minimum quartet inconsistency (abstract). In David Eppstein, editor, *Proceedings of the Thirteenth*

Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA, pages 894–895. ACM/SIAM, 2002.

[27] James Borneman, Marek Chrobak, Gianluca Della Vedova, Andres Figueroa, and Tao Jiang. Probe selection algorithms with applications in the analysis of microbial communities. In *Proceedings of the Ninth International Conference on Intelligent Systems for Molecular Biology, July 21-25, 2001, Copenhagen, Denmark*, pages 39–48, 2001.

[28] Paolo Barone, Paola Bonizzoni, Gianluca Della Vedova, and Giancarlo Mauri. An approximation algorithm for the shortest common supersequence problem: an experimental analysis. In Gary B. Lamont, editor, *Proceedings of the 2001 ACM Symposium on Applied Computing (SAC), March 11-14, 2001, Las Vegas, NV, USA*, pages 56–60. ACM, 2001.

[29] Paola Bonizzoni, Gianluca Della Vedova, and Giancarlo Mauri. Approximating the maximum isomorphic agreement subtree is hard. In Raffaele Giancarlo and David Sankoff, editors, *Combinatorial Pattern Matching, 11th Annual Symposium, CPM 2000, Montreal, Canada, June 21-23, 2000, Proceedings*, volume 1848 of *Lecture Notes in Computer Science*, pages 119–128. Springer, 2000.

[30] Winfried Just and Gianluca Della Vedova. Multiple sequence alignment as a facility location problem. In *Proceedings of the Prague Stringology Club Workshop 2000, Prague, Czech Republic, September 2, 2000*, pages 60–70. Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University, 2000.

[31] Paola Bonizzoni, Massimo D'Alessandro, Gianluca Della Vedova, and Giancarlo Mauri. Experimenting an approximation algorithm for the lcs. In *Algorithms and Experiments (ALEX98)*, pages 96–102, 1998.

[32] Paola Bonizzoni and Gianluca Della Vedova. Modular decomposition of hypergraphs. In Manfred Nagl, editor, *Graph-Theoretic Concepts in Computer Science, 21st International Workshop, WG '95, Aachen, Germany, June 20-22, 1995, Proceedings*, volume 1017 of *Lecture Notes in Computer Science*, pages 303–317. Springer, 1995.

Book chapters

[1] Paola Bonizzoni, Gianluca Della Vedova, Graziano Pesole, Ernesto Picardi, Yuri Pirola, and Raffaella Rizzi. Transcriptome Assembly and Alternative Splicing Analysis. In Ernesto Picardi, editor, *RNA Bioinformatics*, number 1269 in Methods in Molecular Biology, pages 173–188. Springer New York, January 2015.

[2] Paola Bonizzoni, Anna Paola Carrieri, Gianluca Della Vedova, Riccardo Dondi, and Teresa M. Przytycka. *When and How the Perfect Phylogeny Model Explains Evolution*, pages 67–83. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[3] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Jing Li. *The Haplotyping Problem: An Overview of Computational Models and Solutions*, pages 151–181. Artech House, Boston, USA, 2008.

[4] L Valinsky, AJ Scupham, Gianluca Della Vedova, Z Liu, A Figueroa, K Jampachaisri, B Yin, E Bent, R Mancini-Jones, J Press, et al. *Oligonucleotide fingerprinting of ribosomal RNA genes (OFRG)*, pages 569–586. 2004.

[5] Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, and Giancarlo Mauri. The Comparison of Phylogenetic Networks: Algorithms and Complexity. In Ion I. Măndoiu and Alexander Zelikovsky, editors, *Bioinformatics Algorithms*, pages 143–173. John Wiley & Sons, Inc., Hoboken, NJ, USA, August 2007.

Preprints...

[1] Stefano Beretta, Yuri Pirola, Valeria Ranzani, Grazisa Rossetti, Raoul Bonnal, Raffaella Rizzi, Gianluca Della Vedova, Massimiliano Pagani, and Paola Bonizzoni. A tool for the comparison of transcript differential expression analysis pipelines. jul 2016.

Milano, November 10, 2022

Gianluca Della Vedova