Simone Ciccolella

CV

✉ *simone.ciccolella@unimib.it*
↳ *unimib.it/simone-ciccolella*
⊕ *sciccolella*
0000-0002-6469-4887

Current position

PostDoc, University fellowship, Assegno di ricerca A2 (SSD INF/01 — Informatica)
Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano – Bicocca

Positions

03/2022–today	**Dept. of Informatics, Systems, and Communications**, *PostDoc, Assegno A2*, Univ. Milano – Bicocca
10/2021–02/2022	**Dept. of Informatics, Systems, and Communications**, *Research assistant, Borsa di ricerca*, Univ. Milano – Bicocca

Education

2018–2021	**Ph.D.**, *Computer Science*, Università degli Studi di Milano – Bicocca
Thesis: "Practical algorithms for Computational Phylogenetics"
Grade: Approved cum laude
2016–2018	**M.Sc.**, *Computer Science*, Università degli Studi di Milano – Bicocca
Thesis: "Inferring cancer progression from Single Cell Sequencing data while allowing loss of mutations"
Grade: 110/110 summa cum laude
2013–2016	**B.Sc.**, *Computer Science*, Università degli Studi di Milano – Bicocca
Thesis: "A linear Programming Approach to reconstruct cancer phylogenies from multi-sample sequencing data"
Grade: 110/110 summa cum laude

Research Visits

02/2025–04/2025	**Visiting Experienced Researcher**, *hosted by Prof. Kunihiko Sadakane*, University of Tokyo, Data structure and algorithms for multi-string indexing
Tokyo, Japan
07/2024–09/2024	**Visiting Experienced Researcher**, *hosted by Prof. Kunihiko Sadakane*, University of Tokyo, Succinct data structure and algorithms for pangenome graphs
Tokyo, Japan
08/2023–10/2023	**Visiting Experienced Researcher**, *hosted by Prof. Tomáš Szemes*, Geneton s.r.o., Definition of a novel pipeline for phylogenetic analysis of endangered species of Apollo Butterfly
Bratislava, Slovakia

02/2023– 02/2023	**Visiting Experienced Researcher**, *hosted by Prof. Kunihiko Sadakane*, University of Tokyo, Discussions on succinct data structure and algorithms for pangenome graphs Tokyo, Japan
09/2021– 10/2021	**Visiting Experienced Researcher**, *hosted by Prof. Tomáš Szemes*, Geneton s.r.o., Efficient algorithm for computational comparative pangenomic Bratislava, Slovakia
09/2021– 10/2021	**Visiting Ph.D. Student**, *hosted by Prof. Tomáš Szemes*, Geneton s.r.o., Classification of Copy Number Variation from multi-temporal sequecing data Bratislava, Slovakia
08/2020– 12/2020	**Visiting Ph.D. Student**, *hosted by Prof. Tomáš Szemes*, Geneton s.r.o., Development of algorithmic methods for data analysis of multi temporal samples of cancer patients Bratislava, Slovakia
08/2017– 10/2017	**Visiting M.Sc. Student**, *hosted by Prof. Iman Hajirasouliha*, Weill Cornell Medicine, Cornell University, Reconstruction of cancer progressions form Single Cell Sequencing data New York, NY, USA

Research activity

Starting from my Bachelor Thesis, I started collaborating with the Research Lab "AlgoLab — Experimental Algorithmics Lab", at the Università di Milano – Bicocca; mostly on problems related to cancer phylogeny inference. My collaboration continued during my Master Degree and I later became a Ph.D. student for the group. I am currently a PostDoc Researcher with the "BIAS Lab", which is the newly reformed AlgoLab group.

During my career I was able to visit different Universities around the world, such as Cornell University in the USA and the University of Tokyo; furthermore I visited the Bioinformatics company Geneton in Slovakia. These experiences allowed me to experience how different groups work and to incorporate many aspects of management.

My personal research activity has focused on the development of combinatorial and heuristic algorithms in Bioinformatics, with a strong focus on praticality and efficiency. I have paid special attention to the implementation, validation and experimentation on simulated and, especially, real data.

Bioinformatics algorithmics has gained more and more relevance over the last decade and has established itself as a key problem in Science, whose results are used for many real-world applications, such as the understanding of Biology, the development of new drugs and therapies, the study and control of biohazards and viral outbreaks and will continue to gain importance over the coming years. The large adoption and usage of DNA and RNA sequencing has produced enormous amount of data – that continues to increase on a daily basis – and for that reason it is necessary to develop and employ extremely efficient algorithms and, more recently, the use of Machine-Learning techniques. Therefore, my research has been mainly been devoted to the design of such efficient methods and recently to the exploitation of Machine and Deep Learning to solve key Bioinformatics problems, which I detail below in different subfields.

Cancer Phylogeny reconstruction and comparison

Inferring phylogenies is one of the fundamental problems in Bioinformatics whose importance has only exploded in the last decade, when the availability of genetic data extracted from tumoral biopsies has allowed to study the accumulation of somatic mutation during cancer progression.

Given the importance of the problem, several methods have been developed in recent years to infer cancer history from next-generation sequencing data. However, a large majority of them rely on the Infinite Sites Assumption (ISA), which postulates that each mutation is acquired exactly once, but never lost. Due to the aggressive nature of the disease, where large portions of a genome can be deleted, such assumption may not hold in general. Still, it is largely used since it greatly reduces computational complexity and running times.

The main focus of my research is to (i) overcome limitations present in the current literature regarding the models of evolution and (ii) provide practical and efficient algorithms to solve this computational problem. I attacked the problems by first exploring different evolutionary models that could substitute the simpler model commonly used and overcome its limitations. This work used *Integer Linear Programming* approaches and comparative methods to assess the different theoretical models [C3, J10, J1] [software: `gppf`, `PhISCS`].

Between all the ones tested a model has been selected – the Dollo-k Phylogeny – for further development, for which I produced two efficient open source tools that where the firsts to introduce such model of evolution. The first method employed a an ILP formulation to first find an optimal region of the search space followed by an Hill-Climbing algorithm to perform a local search [J9] [software: `gpps`]. While proving the feasibility of the problem the tool was limited by the computational time required by the ILP solver, thus I proposed an improved method that employs a Simulated Annealing algorithm to perform a maximum-likelihood tree search [J5] [software: `SASC`]. Other meta-heuristic approaches have been explored on the same topic and showed promising results worth more investication [J11].

The large availability of many tools for inferring tumor phylogenies means that it is easy to have several different phylogenies from the same tumor dataset, motivating the search for methods that are able to compare and cluster those phylogenies — a large cluster with several highly similar phylogenies is likely to be more reliable, since the underlying evolution is confirmed by multiple methods. In this direction I defined and proposed a novel similarity metric which was the first to allow multiple occurrences of the same label in the tree, which is a critical key factor for cancer progressions; moreover an exact algorithm for computing this measure has been proposed, proved to be correct and implemented as open-source program [J4] [software: `MP3`].

Publications

[C3] [J5] [J4] [J9] [J10] [J1] [J11]

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`
- PhISCS `https://github.com/sfu-compbio/PhISCS`

International Collaborations

- Prof. Murray Patterson. Georgia State University, Computer Science Department. Atlanta, GA, USA.
- Prof. Iman Hajirasouliha. Weill Cornell Medicine, Institute of Precision Medicine, Institute

of Computational Biomedicine. New York City, NY, USA.
- Prof. Tomáš Szemes. Geneton s.r.o. . Bratislava, Slovakia.

Clustering and classification

The problem of the classification of data into similar categories is a key problem in Computer Science and has large ramifications in Bioinformatics; for example the advent of single cell sequencing (SCS) technologies provide a level of resolution that makes it indispensable for inferring phylogenies from a sequenced tumor; however a drawback of SCS is elevated false negative and missing value rates, resulting in a large space of possible solutions, which in turn makes it difficult, sometimes infeasible using current approaches and tools. One possible solution is to reduce the size of an SCS instance – usually represented as a matrix of presence, absence, and uncertainty of the mutations found in the different sequenced cells – and to infer the tree from this reduced-size instance. With this goal I presented a new clustering procedure aimed at clustering such categorical vector, or matrix data. By accurate experimental evaluation of the method it is found that this novel clustering method not only decreased the runtime of different state-of-the-art phylogeny inference methods, but can also improve the quality of the inferred phylogenies, at no disadvantage for the downstream analysis [J8] [software: `celluloid`]. Furthermore I worked on the integration of the previous works to produce a complete pipeline with the goal of providing a simple to use and user-friendly interface to users that want to analyze cancer data, including preprocessing of the data, evolutionary inference and comparison of phylogenies [J7, C5] [software: `plastic`].

Similarly to clustering another interesting problem I am working on is the classification of organisms according to some property, typically phylogenetic assignment. In this direction I worked on the analysis and classification of viruses, and in particular of SARS-CoV-2 variants. With this goal I worked on the development of a pipeline and web-interface for the analysis of genetic variants in the virus combining classical SNV genotyping, specifically modified to work on this type of data [J6] [software: `MALVIRUS`].

On the other hand I also applied Artificial Intelligence techniques, especially Deep Neural Networks, to classify strains of SARS-CoV-2 by using embedding of k-mer counting into images on which a Recurrent Neural Network was trained of, reaching high levels of performance [J2] [software: `CouGaR-g`]. Embedding of reads also play a crucial role in machine learning techniques and it is not a trivial task due to the nature of the data itself that is not easily representable as a numerical nor categorical feature, to this goal I collaborated to the creation of embeddings and their applications [J3, C4].

Other applications of Artificial Intelligence are on the subject of Alternative Slicing classification, for which I have developed a method that leverages novel image representations of different events from sequencing reads and developed a deep neural network model to accurately classify such events even in a noisy environment. Crucially, I also created a manually curated dataset of AS events that can be used to evaluate the performance of future studies [J15][software: `deepSpecas`].

Publications

[J3] [J7] [J2] [J8] [J6] [C4] [C5] [J15]

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`
- deepSpecas `https://github.com/sciccolella/deepSpecas`

International Collaborations
- Prof. Murray Patterson. Georgia State University, Computer Science Department. Atlanta, GA, USA.

Computational Pangenomics

In the framework of computational biology, a pangenome graph aims to represent the multiple pairwise comparison of genomic sequences. It has been proposed as a new paradigm that replaces the traditional view of a reference genome as a linear sequence, to a sequence labeled graph able to summarize the main variations among a population of evolutionary-related sequences or genomes.

The construction of such pangenome graphs is a fundamental task in pangenomics. A natural theoretical question is how to formalize the computational problem of building an optimal pangenome graph, making explicit the underlying optimization criterion and the set of feasible solutions. I developed a novel graph-building approach by leveraging the notion of maximal block in a Multiple Sequence Alignment, re-framing the pangenome graph construction problem as an exact cover problem and then propose an Integer Linear Programming formulation that allows to study the most natural objective functions for building a graph. With an extensive experimental evaluation we proved that a customized construction of a pangenome graph based on selecting objective functions has a direct impact on the resulting graphs, paving the way to novel practical approaches to graph representations [J12].

Integrating genetic variability with transcriptome diversity can improve our understanding of alternative splicing, a regulation mechanism which allows a single gene to code for multiple proteins. However, many unanswered questions are limiting our comprehension of the relationship between genetic and trancriptomic variations. I have developed the first approach based on pangenome graphs for the detection and differential quantification of alternative splicing events; proving that pangenome graphs can be effectively used to perform such an analysis by comparing with state-of-the-art approaches [J14][software: `pantas`].

Another key problem in computational pangenomics is the Sequence-to-graph alignment in which despite its relevance, some basic questions, such as what is a *good* sequence-to-graph alignment and what are its main desired features, have not been fully investigated from a theoretical viewpoint. I worked on an extension of the notion of sequence-to-graph alignment to a pangenomic graphs that incorporates a recombination, so that they are explicitly represented and evaluated in an alignment; In the paper we presented a dynamic programming approach for the special case where there is at most a recombination and experimentally proved that the method can accurately aligns recombinant bacterial sequences [J13].

Publications [J12] [J13] [J14]

Software
- pangeblocks `https://github.com/AlgoLab/pangeblocks`
- pantas `https://github.com/AlgoLab/pantas`
- RecGraph `https://github.com/AlgoLab/RecGraph`

International Collaborations
- Prof. Kunihiko Sadakane. University of Tokyo. Tokyo, Japan.

Leading roles in EU funded research projects

2024–2027 **HORIZON-WIDERA-2023-ACCESS-02-01** Fostering Excellence in Advanced Genomics and Proteomics Research at Comenius University in Bratislava (FORGENOM II). (Univ. Milano – Bicocca amount: 386 050€)

- I am one of five research staff member of University of Milano – Bicocca of this project whose main goal is to improve the scientific excellence of the bioinformatics research activities in Slovakia.
- I am **leader of the Work Package** "Training coordination & content II - research management" with the objective to develop a coherent training plan (initial and updated version) following the needs of research management staff; to organise and deliver training events in a timely and efficient manner; to ensure the training overall quality, the level of participation and the degree of satisfaction of the participants.
- I am collaborating to the organization of events related to the project. In details FORGENOM II has two hackathons, one PhD School and two Training events for administrative staff, that are organized by the Unit at the University of Milano-Bicocca. Currently I co-organized one hackathon and I am co-organizing the second hackathon in conjunction with the PhD School, that will be delivered in summer 2025. Lastly I am contributing to the definition and delivery of the Trainings, by selecting and guiding key personnel from University of Milano-Bicocca to produce courses for the involved project partners.

Contributions to EU funded research projects

2020–2025 **H2020-MSCA-RISE-2019** Pan-genome Graph Algorithms and Data Integration (PANGAIA) — Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange programme. (Univ. Milano – Bicocca amount: 197 800€)

- I contributed to the definition of the strategy and organization of the website and the communications of the project and its dissemination [`https://www.pangenome.eu/`]. I am currently the social media manager for the PANGAIA project and have received three paid contracts for the ongoing management, as part of my duty and contributions I have designed and updated the website and social media communications.
- I was a seconded research staff at *Geneton s.r.o., Bratislava, Slovakia* to collaborate with Prof. Tomáš Szemes (founder and Chief Scientific Officer of Geneton s.r.o. — one of major Slovak genomic companies employing about 20 researcher — and Genomics core lab Manager at Comenius University Science Park) with whom I collaborated on different projects regarding discovery and cathegorization of somatic cancer variants in blood circulating DNA samples; *University of Tokyo, Tokyo, Japan* with Prof. Kunihiko Sadakane, with whom I recently started collaborating on succint data structure and efficient algorithms for indexing and query pangenome graphs.
- I organized a hackathon on the analysis and comparison of graph construction and alignment during the scheduled PhD School *"Evolution in Computational Pangenomics"* in Como, Italy. 8-12 September, 2024. The hackathon was attended by about 20 PhD students from all Europe.

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

○ I am contributing to the supervision of PhD student *Jorge Avila Cartes*, with whom I co-authored many publications on topics regarding Deep Learning models for SARS-CoV-2 clade classification and combinatorial algorithms for pangemone graph construction and sequence-to-graph alignment.

○ I organized a hackathon on the subject of sequence-to-graph alignment during the annual ALPACA meeting at *Centrum Wiskunde & Informatica* in Amsterdam, Netherlands. 20–24 November, 2023. The hackathon was attended by about 20 PhD students from all Europe.

International collaborations

Georgia State University – Atlanta, GA, USA

I have been collaborating with Prof. Murray Patterson, since 2018 when I started my PhD and Dr. Patterson was a PostDoc at University of Milano – Bicocca. We have continued to collaborate on various projects regarding cancer phylogenomics, evolutionary studies of viruses, clustering and machine learning techniques for biological data, mostly regarding Single Cell sequencing data.

Geneton s.r.o. – Bratislava, Slovakia

I have been collaborating with Prof. Tomáš Szemes, Dr. Jaroslav Budis (head of the bioinformatics research and development team) since 2020 on the application of algorithms to the study of cancer genomics, in particular in discovery of somatic mutations from circulating DNA in blood samples. Furthermore I am collaborating in the phylogenetic analysis of endangered butterfly species and their impact on the environment.

Weill Cornell Medicine – New York City, NY, USA

I have been collaborating with Prof. Iman Hajirasouilha since 2017 on the design and implementation of novel and efficient algorithms for cancer phylogeny reconstructions as well as novel evolutionary models to overcome the current limitations of the available approaches.

University of Tokyo – Tokyo, Japan

I have been collaborating with Prof. Kunihiko Sadakane with goal of developing succinct data structures and efficient algorithms for strings and graphs. We worked on a graph indexing for pangenome graphs and are currently collaborating on block decomposition using string-based indices.

Invited Speaker

01/2025 International Conference on Computational Advances in Bio and Medical Sciences (ICCABS). *CASCODA workshop.* Atlanta, Georgia, USA. January 12-14, 2025. Title: *"Cancer progression inference from single-cell data with recurrences and mutation losses"*.

07/2024 Computability in Europe. Special session *Computational methods in evolutionary biology*. Amsterdam, Netherlands. July 04-12, 2024. Title: *"Phylogenetics in bioinformatics: importance and applications"*.

Conference Speaker

02/2023 Computational Pangenomics – Shonan Meeting No. 197. Shonan, Japan. February 19-24, 2023. Title: *"Open problems in cancer phylogenetics"*.

05/2020 RECOMB-Computational Cancer Biology (CCB). Verona, Italy (Virtual). May 10-13, 2020. Title: *"Triplet-based similarity score for fully multi-labeled trees with poly-occurring labels"*.

10/2018 IEEE 8th International Conference on Computational Advances in Bio and medical Sciences (ICCABS). Las Vegas, NV, USA. October 18-20, 2018 Title: *"gpps: An ILP-based approach for inferring cancer progression with mutation losses from single cell data"*.

07/2018 SCANGEN: Single-cell cancer genomics, special session of ISMB (the main Conference in BioInformatics). Chicago, IL, USA. July 06-10, 2018. Title: *"Inferring Cancer Progression from Single Cell Sequencing while allowing loss of mutations"*.

05/2018 Cancer Development And Complexity PhD School. (CDAC) Como, Italy. May 22-25, 2018. Title: *"Inferring Cancer Progression from Single Cell Sequencing while allowing loss of mutations"*.

04/2018 RECOMB-Computational Cancer Biology (CCB). Paris, France. Apr 21-24, 2018. Title: *"Inferring Cancer Progression from Single Cell Sequencing while allowing loss of mutations"*.

08/2017 8th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB). Boston, MA, USA. August 20-23, 2017. Title: *"Beyond Perfect Phylogeny: Multisample Phylogeny Reconstruction via ILP"*.

Workshops and Schools Speaker

01/2025 Pangenome Bio Hacking 2025 Workshop. Online International Workshop. Milan, Italy. January 22-24, 2025. Title: *"Modelling and combinatorial approaches for Pangenomics"*. Focus on combinatorial modeling and algorithms.

09/2024 PhD Shool on evolution in Computational Pangenomics. Lake Como School of Advanced, Como, Italy. September 8-12, 2024. Title: *"Hackathon on sequence-to-graph alignment"*. Focus on the comparative analysis of the methods.

11/2023 ALPACA-PANGAIA annual workshop. Centrum Wiskunde & Informatica (CWI), Amsterdam, Netherlands. November 20–24, 2023. Title: *"Hackathon on sequence-to-graph alignment"*. Focus on the algorithmic aspects of the methods.

Teaching

Teacher - (Affidamento di incarico personale)

2024 *Foundation of Computer Science*, M.Sc. in Data Science, Univ. Milano – Bicocca. *Total: 18 hours of teaching.*

2024 *Genome Assembly & Beyond Genome Assembly*, Master (Post-graduate course) "qOmics: quantitative methods for Omics Data", Univ. Milano – Bicocca. *Total: 24 hours of teaching.*

2023 *Elementi di Bioinformatica (Introduction to Bioinformatics)*, B.Sc. in Computer Science, Univ. Milano – Bicocca. *Total: 20 hours of teaching.*

2023 *Foundation of Computer Science*, M.Sc. in Data Science, Univ. Milano – Bicocca. *Total: 18 hours of teaching.*

2023 *Laboratorio di programmazione I (Programming Lab I)*, B.Sc. in Physics, Univ. Milano – Bicocca. *Total: 24 hours of teaching.*

2022 *Foundation of Computer Science*, M.Sc. in Data Science, Univ. Milano – Bicocca. *Total: 18 hours of teaching.*

2022 *Laboratorio di programmazione I (Programming Lab I)*, B.Sc. in Physics, Univ. Milano – Bicocca. *Total: 24 hours of teaching.*

2021 *Beyond Genome Assembly*, Master (Post-graduate course) "qOmics: quantitative methods for Omics Data", Univ. Milano – Bicocca and Univ. Pavia. *Total: 10 hours of teaching.*

2021 *Foundation of Computer Science*, M.Sc. in Data Science, Univ. Milano – Bicocca. *Total: 18 hours of teaching.*

2020 *Laboratorio di programmazione I (Programming Lab I)*, B.Sc. in Physics, Univ. Milano – Bicocca. *Total: 24 hours of teaching.*

Teaching assistant (Tutor)

2023 *Modelli e computazione (Models and computation)*, M.Sc. in Computer Science, Univ. Milano – Bicocca.

2019 *Analisi e Progetto di Algoritmi (Analysis and Development of Algorithms)*, B.Sc. in Computer Science, Univ. Milano – Bicocca.

PhD Students supervision activities

Supervision of PhD Students entails the supervision of all the scientific activities of the candidates.

○ [2024-2027] **Co-supervisor** of Davide Cesare Monti, PhD in Computer Science, University of Milano – Bicocca. *Davide's main focus has been on the developement of sequence-to-pangenome graph alignment algorithms. In particular, the possibility of including the detection of recombination events in the alignment step. He has also been recently working on a new heuristic approach to solve the haplotype-aware co-linear chaining problem.*

○ [2022-2025] **Co-supervisor** of Davide Cozzi, PhD in Computer Science, University of Milano – Bicocca. *Davide's research activity is focused on the development of space-efficient algorithms for the analysis of genomic variants, through the use of succinct data structures. In addition, he collaborated on the development of methodologies for the analysis of alternative splicing events. Lastly, he worked on secure algorithms for performing pattern matching on text using the multi-party computation framework.*

- [2021-2024] As PostDoc of the laboratory I am contributing to the supervision of Jorge Avila Cartes, PhD in Computer Science, University of Milano – Bicocca, H2020 ITN Student. *I am contributing to the supervision of the scientific activities of Jorge including Deep Learning application of classification of genomic data as well as the more theoretical aspects of algorithms and data structures.*
- [2023-2026] As PostDoc of the laboratory I am contributing to the supervision of Mattia Sgrò, PhD in Computer Science, University of Milano – Bicocca. *Mattia has been working on algorithmic approaches for the construction of indices from multiple strings and has recently working on plasmid genomic assembly.*

Student Supervision

Following is the list the thesis I had co-supervised, the names of the students have been redacted for privacy reasons.

Students supervision - M.Sc. in Data Science

- [2024] Disegni ed implementazione di euristiche sul pangenoma
- [2024] Metodi per apprendimento di predizione ACGM
- [2024] Classificazione di campioni virali
- [2023] Metodi per apprendimento di predizione di patogenicità
- [2022] New approaches for SARS-CoV-2 strain identification

Students supervision - B.Sc. in Computer Science

- [2025] Sviluppo di algoritmi di riconciliazioni di alberi di filogenesi tumorali
- [2024] Sviluppo di framework software per pipile di filogenesi tumorale
- [2024] Sviluppo di un sistema di package manager per linguaggio C
- [2023] Tree Embedding on Hyperbolic Space
- [2021] Pipeline di integrazione per inferenza e analisi di filogenesi tumorale
- [2020] Implementazione di misure tra alberi tumorali
- [2020] Clusterizzazione di dati tumorali da single-cell
- [2019] Inferenza di alberi tumorali tramite Particle Swarm Optimization
- [2019] Applicazione web per il confronto di progressioni di cellule tumorali
- [2019] Strategie euristiche per il clustering di dati tumorali da single cell

Program Committee Member

- IEEE International Conference on Bioinformatics & Biomedicine (IEEE BIBM), 2025
- CASCODA workshop of International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), 2025
- IEEE International Conference on Bioinformatics & Biomedicine (IEEE BIBM), 2024
- International Conference of Artificial Intelligence in Medicine (AIME), 2024
- IEEE International Conference on Bioinformatics & Biomedicine (IEEE BIBM), 2023

Reviewer Service

I have reviewed papers for the following journals:

- BMC Bioinformatics, BioMed Central, ISSN: 1471-2105. – *Scimago Ranking: Q1*
- BioData Mining, BioMed Central, ISSN: 1756-0381 – *Scimago Ranking: Q2*
- Computational and Structural Biotechnology Journal, Elsevier, ISSN: 2001-0370 – *Scimago Ranking: Q1*
- Royal Society Open Science Journal, Royal Society, ISSN: 2054-5703 – *Scimago Ranking: Q1*

I have reviewed papers for the following conferences:

- RECOMB – Research in Computational Molecular Biology – *CORE ranking: A*
- ISMB – International Society for Molecular Biology – *CORE ranking: A*
- APBC – Asia Pacific Bioinformatics Conference – *CORE ranking: B*
- AIME – Artificial Intelligence in MEdicine – *CORE ranking: B*
- IEEE BIBM – International Conference on Bioinformatics and Biomedicine – *CORE ranking: B*
- WABI – Workshop on Algorithms in Bioinformatics – *CORE ranking: B*

Conference/School organization

July 8-14, 2025
Pangenomics and Cancer genomics PhD School and Hackathon TBD. Italy. *I am the leading organizer of the both the School and the Hackathon; my main involvements are related to the selection of invited speakers and of the participants; to the dissemination of the event and the social activities related to the events; to the practical organization of the events, from selecting the venue to the schedule.*

June 17-19, 2025
36th Annual Symposium on Combinatorial Pattern Matching (CPM 2025) Milan. Italy. *Member of the Organizing committee, mostly concerning local organization of the conference.*

January 22-24, 2025
Pangenome Bio Hacking 2025 Workshop Online International Workshop. Milan, Italy. *I contributed to the organization of the workshop and selection of invited speakers, to the dissemination of the event and the social activities related to the workshop.*

September 8-12, 2024
Evolution in Computational Pangenomics Lake Como School of Advanced Studies. Como, Italy. *I contributed to the organization of the school and selection of invited speakers and participants, to the dissemination of the event and the social activities related to the meeting.*

February 19-24, 2023
Computational Pangenomics Shonan Meeting No. 197. Shonan, Japan. *I contributed to the selection of the participants, I managed the dissemination of the meeting and created the website containing the information for the meeting.*

July 4-8, 2022
Introduction to Pan-genomics Lake Como School of Advanced Studies. Como, Italy. *I contributed to the local organization of the school, to the management of the participants and to organization of the social activities related to the meeting.*

University Service

2022-today **Member of the Council of the Research Fellow for the University (Membro della Giunta Rappresentati Assegnisti di Ateneo)** University of Milano – Bicocca. *I participate and collaborate to the discussions and meeting between the representatives and the main university boards of organization.*

2022-today **Representative of the PostDocs (Rappresentante Assegnisti)** Council of the Department (Consiglio Di Dipartimento), Department of Computer-Science University of Milano – Bicocca. *I participate and collaborate the meetings regarding the organization of the Department in term of lessons and research.*

2018-2020 **Representative of the PhD Students (Rappresentante Dottorandi)** Council of the Department (Consiglio Di Dipartimento), Department of Computer-Science University of Milano – Bicocca. *I participated and collaborated the meetings regarding the organization of the Department in term of lessons and research.*

2016-2017 Co-President of the **Students-Professors Committee (Commissione Paritetica Docenti-Studenti)**, Dipartimento di Informatica, Sistemistica e Comunicazione. *I was part of the commission evaluating the courses offered by the Department in which I collaborated to finding issues and propose solutions to better the delivery from the students perspective.*

2016-2019 **Representative of the Students (Rappresentante Studenti)** Council of the Department (Consiglio Di Dipartimento), Department of Computer-Science University of Milano – Bicocca. *I participated and collaborated the meetings regarding the organization of the Department in term of lessons.*

Publications

Journal articles

[J1] Salem Malikic, Farid Rashidi Mehrabadi, Simone Ciccolella, Md. Khaledur Rahman, Camir Ricketts, Ehsan Haghshenas, Daniel Seidman, Faraz Hach, Iman Hajirasouliha, and S. Cenk Sahinalp. PhISCS: a combinatorial approach for subperfect tumor phylogeny reconstruction via integrative use of single-cell and bulk sequencing data. *Genome Research (Scimago Ranking Q1)*, 29(11):1860–1877, October 2019.

[J2] Jorge Avila Cartes, Santosh Anand, Simone Ciccolella, Paola Bonizzoni, and Gianluca Della Vedova. Accurate and fast clade assignment via deep learning and frequency chaos game representation. *GigaScience, (Scimago Ranking Q1)*, 12, December 2022.

[J3] Prakash Chourasia, Sarwan Ali, Simone Ciccolella, Gianluca Della Vedova, and Murray Patterson. Reads2vec: Efficient embedding of raw high-throughput sequencing reads data. *Journal of Computational Biology (Scimago Ranking Q2)*, 30(4):469–491, 2023.

[J4] 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 (Scimago Ranking Q1)*, 37(2):178–184, 2021.

[J5]   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 (Scimago Ranking Q1)*, 37(3):326–333, 2021.

[J6]   Simone Ciccolella, Luca Denti, Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, and Marco Previtali. MALVIRUS: an integrated application for viral variant analysis. *BMC Bioinformatics (Scimago Ranking Q1)*, 22-S(15):625, 2021.

[J7]   Sarwan Ali, Simone Ciccolella, Lorenzo Lucarella, Gianluca Della Vedova, and Murray Patterson. Simpler and faster development of tumor phylogeny pipelines. *Journal of Computational Biology (Scimago Ranking Q2)*, 28(11):1142–1155, 2021.

[J8]   Simone Ciccolella, Murray Patterson, Paola Bonizzoni, and Gianluca Della Vedova. Effective clustering for single cell sequencing cancer data. *IEEE Journal of Biomedical and Health Informatics (Scimago Ranking Q1)*, 25(11):4068–4078, 2021.

[J9]   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 (Scimago Ranking Q1)*, 21-S(1):413, 2020.

[J10] 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 Transactions on Computational Biology and Bioinformatics (Scimago Ranking Q2)*, 16(5):1410–1423, 2019.

[J11] Simone Ciccolella, Gianluca Della Vedova, Vladimir Filipovic, and Mauricio Soto Gomez. Three metaheuristic approaches for tumor phylogeny inference: An experimental comparison. *Algorithms (Scimago Ranking Q2)*, 16(7):333, 2023.

[J12] Jorge Avila Cartes, Paola Bonizzoni, Simone Ciccolella, Gianluca Della Vedova, and Luca Denti. Pangeblocks: customized construction of pangenome graphs via maximal blocks. *BMC Bioinform. (Scimago Ranking Q1)*, 25(1):344, 2024.

[J13] Jorge Avila Cartes, Paola Bonizzoni, Simone Ciccolella, Gianluca Della Vedova, Luca Denti, Xavier Didelot, Davide Cesare Monti, and Yuri Pirola. Recgraph: recombination-aware alignment of sequences to variation graphs. *Bioinformatics (Scimago Ranking Q1)*, 40(5), April 2024.

[J14] Simone Ciccolella, Davide Cozzi, Gianluca Della Vedova, Stephen Njuguna Kuria, Paola Bonizzoni, and Luca Denti. Differential quantification of alternative splicing events on spliced pangenome graphs. *PLOS Computational Biology (Scimago Ranking Q1)*, 20(12):e1012665, December 2024.

[J15] Simone Ciccolella, Luca Denti, Jorge Avila Cartes, Gianluca Della Vedova, Yuri Pirola, Raffaella Rizzi, and Paola Bonizzoni. Differential analysis of alternative

splicing events in gene regions using residual neural networks. *Neural Computing and Applications (Scimago Ranking Q1)*, January 2025.

Conference proceedings

[C1] 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.

[C2] 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.

[C3] 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.

[C4] 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 I. Mandoiu, Marmar Moussa, Murray Patterson, Sanguthevar Rajasekaran, Pavel Skums, and Alexander Zelikovsky, editors, *Computational Advances in Bio and Medical Sciences: 11th International Conference, ICCABS 2021, Virtual Event, December 16-18, 2021, Revised Selected Papers*, volume 13254 of *Lecture Notes in Computer Science*, pages 133–148. Springer International Publishing, 2021.

[C5] Akshay Juyal, Zahra Tayebi, Alexander Zelikovsky, Mauricio Soto-Gomez, Simone Ciccolella, Gianluca Della Vedova, and Murray Patterson. Plastic: An easy to use and modular tool for benchmarking tumor phylogeny reconstruction pipelines. In Mukul S. Bansal, Wei Chen, Yury Khudyakov, Ion I. Măndoiu, Marmar R. Moussa, Murray Patterson, Sanguthevar Rajasekaran, Pavel Skums, Sharma V. Thankachan, and Alexander Zelikovsky, editors, *Computational Advances in Bio and Medical Sciences*, pages 82–93, Cham, 2025. Springer Nature Switzerland.

Milano, May 19, 2025
Simone Ciccolella