

RAFFAELLA RIZZI – Curriculum Vitae Didattico e Scientifico

Informazioni personali

Luogo di nascita: Galliate (NO)

Data di nascita: 15 Luglio 1970

Nazionalità: italiana

Email: raffaella.rizzi@unimib.it

Formazione

Laurea in Ingegneria Civile presso il Politecnico di Milano (luglio 1997).

Titolo della tesi: “Esperimenti di ricostruzione tridimensionale di oggetti mediante accoppiamento di contorni estratti da immagini digitali”

Relatore: prof. Gianfranco Forlani

Dottore di Ricerca in Scienze Geodetiche e Topografiche (XIII ciclo) presso il Politecnico di Milano (ottobre 2000).

Titolo della tesi: “Sviluppo di metodologie per l’analisi di immagini di microarray”

Tutore: prof. Luigi Mussio (Politecnico di Milano)

Cotutore: dott. Luciano Milanese (CNR-ITB, Segrate, MI)

Posizione accademica

Da marzo 2021 RR è professore associato presso il Dipartimento di Informatica Sistemistica e Comunicazione dell’Università degli Studi di Milano-Bicocca.

Da dicembre 2008 a febbraio 2021 RR è stata ricercatore a tempo indeterminato presso il Dipartimento di Informatica Sistemistica e Comunicazione dell’Università degli Studi di Milano-Bicocca, dove collabora principalmente con il laboratorio BIAS (Bioinformatica e Algoritmica Sperimentale) focalizzato sul disegno e l’implementazione di algoritmi efficienti per risolvere problemi computazionali. Nel 2018 ha ottenuto l’abilitazione scientifica nazionale per il ruolo di professore di seconda fascia nel settore concorsuale 01/B1 – Informatica; fa parte del collegio del XXXVI dottorato in Informatica del Dipartimento di Informatica, Sistemistica e Comunicazione.

RR è stata beneficiaria del FFABR 2017 - Ricercatori (Finanziamento delle attività base di ricerca). È membro della Commissione Orari del Dipartimento di Informatica Sistemistica e Comunicazione e collabora con la Commissione per l’orientamento nell’organizzazione dell’evento annuale NERD (Non É Roba Per Donne) in collaborazione con IBM al fine di promuovere lo studio dell’Informatica in ambito femminile.

Da febbraio 2007 a dicembre 2008 RR ha ottenuto un contratto di ricercatore di chiara fama sul tema *Progettazione, sviluppo e realizzazione di un software per l’analisi computazionale dello splicing alternativo di geni eucariotici* (responsabile: prof. Giulio Pavesi) presso il Dipartimento di Bioscienze dell’Università degli Studi di Milano.

Da aprile 2006 a gennaio 2007 RR ha avuto un incarico di collaborazione presso il Dipartimento di Informatica Sistemistica e Comunicazione dell’Università degli Studi di Milano-Bicocca.

Da aprile 2005 a marzo 2006 RR è stata assegnista di ricerca presso il Dipartimento di Biotecnologie e Bioscienze dell'Università degli Studi di Milano-Bicocca (responsabile: prof. Piercarlo Fantucci).

Da aprile 2001 a marzo 2005 RR è stata assegnista di ricerca sul tema *Metodi per l'analisi e il confronto di sequenze biologiche* (responsabile: prof. Paola Bonizzoni) presso il Dipartimento di Informatica Sistemistica e Comunicazione dell'Università degli Studi di Milano-Bicocca.

Attività didattica

- **1999/2000:**
 - docente delle esercitazioni per l'insegnamento di *Statistica* - corso di laurea in Pianificazione Territoriale Urbanistica e Ambientale (Vecchio Ordinamento), I Facoltà di Architettura del Politecnico di Milano (docente: prof. Federica Migliaccio).
- **2002/2003:**
 - docente delle esercitazioni per l'insegnamento di *Algoritmi e strutture dati (elementi)* - corso di laurea in Informatica (I livello, vecchio ordinamento), Università degli Studi di Milano-Bicocca.
- **2003/2004:**
 - docente delle esercitazioni per l'insegnamento di *Algoritmi e strutture dati (elementi)* - corso di laurea in Informatica (I livello, vecchio ordinamento), Università degli Studi di Milano-Bicocca.
- **2004/2005:**
 - docente del laboratorio per l'insegnamento di *Algoritmi della Bioinformatica* - Master di I livello in Bioinformatica, Università degli Studi di Milano-Bicocca.
- **2005/2006:**
 - docente del laboratorio per l'insegnamento di *Informatica* - corso di laurea in Biotecnologie (I livello, vecchio ordinamento), Università degli Studi di Milano-Bicocca.
- **2006/2007:**
 - docente dell'insegnamento di *Informatica* (8 CFU) - corso di laurea in Biotecnologie (I livello, vecchio ordinamento), Università degli Studi di Milano-Bicocca.
- **2007/2008:**
 - docente dell'insegnamento di *Informatica* (8 CFU) - corso di laurea in Biotecnologie (I livello, vecchio ordinamento), Università degli Studi di Milano-Bicocca.
 - docente del laboratorio per l'insegnamento di *Bioinformatica: tecniche di base I MOD* (2 CFU) - corso di laurea specialistica in Bioinformatica, Università degli Studi di Milano-Bicocca.
- **2008/2009:**
 - docente del laboratorio per l'insegnamento di *Bioinformatica: tecniche di base I MOD* (2 CFU) - corso di laurea Specialistica in Bioinformatica, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Fondamenti dell'Informatica, modulo Complessità* (1 CFU) - corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
- **2009/2010:**
 - docente per l'insegnamento di *Informatica, moduli di Programmazione e di Basi di Dati* (6 CFU) - corso di laurea Triennale in Biotecnologie, Università degli Studi di Milano-Bicocca.
 - docente del laboratorio per l'insegnamento di *Bioinformatica: Tecniche di Base* (4 CFU) - corso di laurea Magistrale in Bioinformatica, Università degli Studi di Milano-Bicocca.

- docente per l'insegnamento di *Programmazione* (4 CFU) - corso di laurea Magistrale in Bioinformatica, Università degli Studi di Milano-Bicocca.
- **2010/2011:**
 - docente del laboratorio per l'insegnamento di *Bioinformatica: Tecniche di Base* (4 CFU) - corso di laurea Magistrale in Bioinformatica, Università degli Studi di Milano-Bicocca.
 - docente per l'insegnamento di *Programmazione* (4 CFU) - corso di laurea Magistrale in Bioinformatica, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Modelli e Sistemi, modulo Disegno ed Analisi di Algoritmi* (1 CFU) - corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
- **2011/2012:**
 - docente titolare dell'insegnamento di *Informatica* (8 CFU) - corso di laurea Triennale in Biotecnologie, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Metodologie Bioinformatiche* (1 CFU) - corso di laurea Magistrale in Biotecnologie Industriali, Università degli Studi di Milano-Bicocca.
- **2012/2013:**
 - docente titolare dell'insegnamento di *Informatica* (8 CFU) – corso di laurea Triennale in Biotecnologie, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Metodologie Bioinformatiche* (1 CFU) - corso di laurea Magistrale in Biotecnologie Industriali, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Bioinformatica* (2 CFU) – corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
- **2013/2014 e 2014/2015:**
 - docente titolare dell'insegnamento di *Informatica, modulo di Programmazione* (4 CFU) - corso di laurea Triennale in Biotecnologie, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Metodologie Bioinformatiche* (1 CFU) - corso di laurea Magistrale in Biotecnologie Industriali, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Bioinformatica* (1 CFU) – corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
 - docente del laboratorio per l'insegnamento di *Elementi di Bioinformatica* (2 CFU) . corso di laurea Triennale in Informatica, Università degli Studi di Milano-Bicocca.
- **2015/2016:**
 - docente delle esercitazioni per l'insegnamento di *Metodologie Bioinformatiche* (1 CFU) - corso di laurea Magistrale in Biotecnologie Industriali, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Bioinformatica* (1 CFU) – corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
 - docente del laboratorio per l'insegnamento di *Elementi di Bioinformatica* (2 CFU) . corso di laurea Triennale in Informatica, Università degli Studi di Milano-Bicocca.
- **2016/2017, 2017/2018, 2018/2019 e 2019/2020:**
 - docente delle esercitazioni per l'insegnamento di *Modelli e Computazione, modulo di Teoria della Computazione* (2 CFU) - corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Bioinformatica* (1 CFU) – corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.

- docente del laboratorio per l'insegnamento di *Elementi di Bioinformatica* (2 CFU) . corso di laurea Triennale in Informatica, Università degli Studi di Milano-Bicocca.
- **2020/2021:**
 - docente delle esercitazioni per l'insegnamento di *Modelli e Computazione, modulo di Teoria della Computazione* (2 CFU) - corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Bioinformatica* (1 CFU) – corso di laurea Magistrale in Informatica, Università degli Studi di Milano-Bicocca.
 - docente del laboratorio per l'insegnamento di *Elementi di Bioinformatica* (2 CFU) . corso di laurea Triennale in Informatica, Università degli Studi di Milano-Bicocca.
 - docente delle esercitazioni per l'insegnamento di *Analisi e Progetto di Algoritmi* (2 CFU) - corso di laurea Triennale in Informatica, Università degli Studi di Milano-Bicocca.

Partecipazione a collegi di dottorato e supervisione di tesi di dottorato

- Partecipazione al collegio dei docenti del XXXVI ciclo di Dottorato di Ricerca in Informatica del Dipartimento di Informatica Sistemistica e Comunicazione dell'Università degli Studi di Milano-Bicocca.
- Collaborazione alla supervisione della tesi dottorato di Stefano Beretta dal titolo "Algorithms for Next Generation Sequencing Data Analysis". Ciclo XXV del corso di dottorato in Informatica del Dipartimento di Informatica Sistemistica e Comunicazione dell'Università di Milano-Bicocca. Tale lavoro di tesi ha portato allo sviluppo di un metodo per fornire una rappresentazione compatta della struttura dei trascritti di un gene ottenuta a partire da dati di Next-Generation Sequencing (NGS).
- Collaborazione alla supervisione della tesi dottorato di Marco Previtali dal titolo "Self-indexing for de novo assembly" del ciclo XXIX del corso di dottorato in Informatica del Dipartimento di Informatica Sistemistica e Comunicazione dell'Università di Milano-Bicocca. Tale lavoro di tesi ha portato allo sviluppo di un software (LightStringGraph - LSG) per la costruzione in memoria esterna di uno string-graph di assemblaggio di un genoma.
- Co-supervisione della tesi dottorato di Luca Denti del ciclo XXXII del corso di dottorato in Informatica del Dipartimento di Informatica Sistemistica e Comunicazione dell'Università di Milano-Bicocca. Tale lavoro di tesi a portato allo sviluppo di un software (ASGAL) per rilevare eventi di splicing, nuovi rispetto a un'annotazione (fornita da splicing graph), mediante allineamento a grafo di reads NGS (Next-Generation Sequencing).

Insegnamento in corsi di dottorato e summer schools

- Docente del corso di dottorato "Hands on - Translational Genomics and Bioinformatics" (2017) nell'ambito del dottorato in Medicina Molecolare, Immunologia e Oncologia di Base e Applicate presso l'Università Vita-Salute del San Raffaele.
- Docente del corso di dottorato "Advanced Techniques for Combinatorial Algorithms (2018)" nell'ambito del dottorato in Informatica dell'Università degli Studi di Milano-Bicocca (pagina web: <https://algolab.eu/2018/01/07/advanced-techniques-for-combinatorial-algorithms-2018/>).
- Docente del corso di dottorato "Advanced Techniques for Combinatorial Algorithms (2020)" nell'ambito del dottorato in Informatica dell'Università degli Studi di Milano-Bicocca (pagina web: <https://www.unimib.it/node/17929>).
- Docente alla ISCD Summer School 2017 "Scientific Trends at the Interfaces Bioinformatics –

Visual Data Analysis", 17 luglio-11 agosto, 2017(Roscoff Marine Station, France) organizzata da Alessandra Carbone (CNRS - Université Pierre et Marie Curie, Parigi), Pascal Frey e Elodie Laine (Sorbonne Université, Parigi).

Supervisione di Tesi di Laurea

RR svolge attività di coordinamento e supervisione per stage e tesi di laurea svolti in collaborazione con il Laboratorio BIAS (Bioinformatica e Algoritmica Sperimentale) del Dipartimento di Informatica, Sistemistica e Comunicazione dell'Università di Milano-Bicocca (ex laboratorio AlgoLab).

Ha seguito 27 stage di laurea triennale e le seguenti tesi di laurea magistrale di cui RR è stata relatore/correlatore:

- *Metodi computazionali per l'analisi della tessuto-specificità degli splicing alternativi*, corso di laurea Specialistica in Bioinformatica, 2007-2008, studente: Simona Giudici.
- *Metodi algoritmici per l'inferenza di variazioni geniche mediante i dati di short read e paired-end*, corso di laurea Magistrale in Informatica, 2009-2010, studente: Fabrizio Ferrari.
- *Algoritmi per la predizione di eventi di splicing alternativo da dati NGS (Next-Generation Sequencing)*, corso di laurea Magistrale in Informatica, 2012-2013, studente: Marco Previtali.
- *Un ambiente per l'analisi automatica di dati di sequenziamento NGS per inferire eventi di splicing alternativo*, corso di laurea Magistrale in Informatica, 2014-2015, studente: Mirco Carucci.
- *A new algorithmic approach for the approximate alignment of a string to a graph*, corso di laurea Magistrale in Informatica, 2015-2016, studente: Luca Denti.
- *Mapping-free classification of NGS data with Bloom filter*, corso di laurea Magistrale in Informatica, 2017-2018, studente: Tamara Ceccato.
- *Signature insight of sequencing data by the Lyndon factorization*, corso di laurea Magistrale in Informatica, 2018-2019, studente: Matteo Costantini.
- *Splice junction classification via deep learning*, corso di laurea Magistrale in Data Science, 2019-2020, studente: Anastasia Marzi
- *Lyndon signatures-based feature embedding for biological sequence mining*, corso di laurea Magistrale in Data Science, 2019-2020, studente: Alessia Petescia.

Attività di Ricerca

L'attività di ricerca di RR riguarda principalmente il settore degli algoritmi in Bioinformatica con particolare riguardo al disegno e alla sperimentazione di metodi per analisi e confronto di sequenze biologiche di DNA e RNA. RR si è anche occupata (durante il periodo di dottorato) dei seguenti temi: analisi di dati di espressione genica, di immagini di espressione da microarray e banche dati genomiche.

Attualmente si occupa della risoluzione algoritmica di problemi computazionali derivanti dallo studio del problema della predizione della struttura di un gene e dei diversi trascritti alternativi che un gene può esprimere, di strutture di indicizzazione per assemblare genomi e trascritti, di

filogenesi, e di determinare in maniera efficiente gli overlaps all'interno di un set di reads di sequenziamento, tramite fingerprints basate su parole di Lyndon.

Predizione della struttura e dei trascritti alternativi di un gene. In tale ambito RR si è occupata di disegnare un metodo per predire la struttura in esoni-introni di un gene e i suoi potenziali trascritti alternativi (espressi in virtù del meccanismo dello splicing alternativo), attraverso l'allineamento (*spliced alignment*) di un set di frammenti di trascritto (ESTs o Next-Generation (NGS) reads) a un genoma. Tale metodo è stato implementato nel *tool* ASPIC. Tale attività si è collocata all'interno del progetto ASPIC (Alternative Splicing Prediction), in collaborazione con il prof. Graziano Pesole del Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica dell'Università degli Studi di Bari e con il CASPUR di Roma. Il progetto ASPIC ha condotto alla realizzazione di un ambiente web per il processamento, mediante il *tool* ASPIC, di geni di diverse specie e alla realizzazione di **ASPicDB** (<http://srv00.recas.ba.infn.it/ASPicDB/>) un database pensato per dare accesso ai pattern di alternative splicing in geni di diverse specie.

RR si è in seguito occupata del disegno, dell'implementazione e della sperimentazione di un algoritmo, basato su suffix tree e sul concetto di maximal embedding tra due sequenze, che ha lo scopo di migliorare il metodo ASPIC nella determinazione dello spliced alignment tra un trascritto e una sequenza genomica, al fine di determinare le regioni codificanti (esoni) di un gene. Tale metodo è stato implementato nel *tool* **Pintron** (<https://github.com/AlgoLab/Pintron>) per la predizione della struttura in esoni-introni di un gene.

Agreement di allineamenti spliced. La determinazione dell'allineamento di un frammento di trascritto rispetto ad una sequenza genomica è fondamentale per risolvere il problema della determinazione della struttura genica e dello splicing alternativo. Una data sequenza di trascritto può però dare origine a diversi allineamenti rispetto ad una stessa genomica, soprattutto se la genomica è molto ampia. RR si è occupata quindi di risolvere il seguente problema computazionale MFA (Minimum Factorization Agreement): dato un set C di allineamenti *spliced* di un set S di frammenti di trascritto rispetto ad una data genomica (un frammento può avere anche più di un allineamento), estrarre un subset C' di C tale che ogni frammento abbia un solo allineamento in C' , e C' sia in accordo con una struttura genica comune.

RR si è occupata anche del disegno di un nuovo algoritmo basato su grafo, per assemblare frammenti di trascritto (ESTs in particolare) in trascritti completi (isoforme *full-length*). Tale algoritmo viene applicato per predire i potenziali trascritti espressi da uno specifico gene.

Clustering di sequenze EST. Il *clustering* di sequenze EST (frammenti di trascritto) è di fondamentale importanza per identificare geni sul DNA genomico, per rilevare variazioni genetiche, per studiare l'espressione genica e per ridurre la ridondanza che, molto spesso, i dati di sequenziamento (frammenti di trascritto) contengono.

In tale contesto RR si è occupata di disegnare un metodo di EST clustering a partire dagli spliced alignments rispetto ad una sequenza genomica, con lo scopo di identificare i geni contenuti nella genomica stessa e di predire gli eventi di splicing alternativo e i trascritti completi.

Ricostruzione dei trascritti *full-length* a partire da frammenti NGS, in assenza di un genoma di riferimento. Le nuove tecnologie di sequenziamento di molecole biologiche (Next-Generation Sequencing, NGS) hanno avuto un forte impatto in Bioinformatica, dal momento che consentono di ottenere una grande mole di dati ad un prezzo relativamente basso. In tale contesto, RR ha affrontato e investigato il problema di predizione della struttura di un gene e dello splicing alternativo in relazione ai trascritti *full-length* ricostruiti a partire da RNA-Seq reads (dati NGS ottenuti da tecnologia RNA-Seq), in assenza di un genoma di riferimento. Viene quindi presentato un nuovo modello di struttura genica, chiamato *Isoform Graph*, ed un approccio per ricostruirlo a partire da un dataset di RNA-Seq reads.

Indicizzazione di collezioni di stringhe per l'assemblaggio di genomi. In tale ambito RR si è occupata del problema di costruire la Burrows-Wheeler Transform (BWT) e l'FM-index in memoria esterna per una collezione di stringhe. BWT e FM-index sono infatti due importanti strutture di indicizzazione che facilitano l'assemblaggio di genomi a partire da un set di reads NGS (Next-Generation Sequencing). E inoltre RR ha contribuito al disegno dei metodi LightStringGraph (LSG) and FastStringGraph (FSG) per costruire lo *string graph* di un set di reads genomici NGS al fine di assemblare il genoma di origine.

Burrows-Wheeler Transform (BWT), FM-index e Longest Common Prefix (LCP) Array di una collezione di stringhe sono strutture di indicizzazione fondamentali per assemblare genomi a partire da un set di reads NGS (Next-Generation Sequencing). La caratteristica principale di un set di reads NGS è quella di occupare una notevole memoria (dati di grandi dimensioni), e ciò richiede spesso di disegnare e implementare algoritmi in memoria esterna.

In tale ambito, RR si è occupata del problema di costruire (per una collezione di stringhe) Burrows-Wheeler Transform (BWT) e LCP Array in memoria esterna, mediante due approcci diversi: (1) di forward extension di Q-intervalli e (2) backward extension di Q-intervalli.

Rilevamento di eventi di alternative splicing. Rilevare eventi di splicing che sono nuovi rispetto a un'annotazione presente in banca dati è uno dei compiti fondamentali della trascrittomica in quanto permette di caratterizzare gli eventi che si manifestano all'interno di un particolare campione biologico. RR si è occupata del problema di estrarre da un campione di frammenti di trascritto (RNA-Seq reads) gli eventi di splicing che sono nuovi rispetto a un'annotazione (*splicing graph*), allineando i frammenti al grafo di splicing. Un metodo che risolve tale problema è stato disegnato e implementato nel *tool* ASGAL.

Uso di un Bloom Filter per assegnare degli RNA-Seq reads ai geni di origine. Un problema importante in Bioinformatica è quello di assegnare i frammenti, ottenuti per sequenziamento di trascritti, al loro gene di origine. RR ha collaborato allo sviluppo di SHARK che è un alignment-free tool basato su Bloom Filter, per tenere solo quei frammenti che hanno origine da geni rilevanti analisi downstream.

Determinazione degli overlaps in un set di stringhe usando fingerprints basate su parole di Lyndon. Determinare gli overlap (regioni comuni) in un insieme di stringhe è un compito fondamentale per assemblare frammenti di genoma o di trascrittoma e ricostruire la sequenza originale. Attualmente RR sta lavorando su una nozione di fingerprint di una stringa basata su fattorizzazione in parole di Lyndon al fine di sostituire la stringa stessa per ricercare gli overlaps.

Collaborazioni di ricerca

- Da luglio 1997 a gennaio 1998: collaborazione scientifica con il prof. Luigi Mussio (Dipartimento di Ingegneria Idraulica Ambientale e del Rilevamento (sezione Rilevamento) del Politecnico di Milano) sulla ricostruzione 3D di manufatti a partire da immagini digitali.
- Da gennaio 1998 a marzo 2001: collaborazione con la sezione di Bioinformatica (responsabile: Dott. Luciano Milanese) dell'Istituto di Tecnologie Biomediche Avanzate del CNR di Segrate (Milano) finalizzata allo svolgimento della tesi di dottorato dal titolo "Sviluppo di metodologie per l'analisi di immagini di microarray".
- Da aprile 2001: partecipazione alle attività del laboratorio di ricerca in Bioinformatica e Calcolo Naturale del Dipartimento di Informatica, Sistemistica e Comunicazione, Università di Milano-Bicocca.
- Da gennaio 2003: collaborazione scientifica con il prof. Graziano Pesole (Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari - CNR di Bari, Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica - Università degli Studi di Bari). La

collaborazione ha portato alla pubblicazione di 6 pubblicazioni su rivista (indicizzate su Scopus) e alla realizzazione del *tool* Pintron (di predizione della struttura genica e di ricostruzione di trascritti; pagina web: <http://pintron.algolab.eu/>) e del database ASPicDB (pagina web: <http://srv00.recas.ba.infn.it/ASPicDB/>) per rendere accessibili le predizioni di Pintron.

- Da ottobre 2005 a ottobre 2007: partecipazione alle attività del progetto FIRB 2003 "Bioinformatica per la Genomica e la Proteomica". Il progetto ha portato a 4 pubblicazioni indicizzate su Scopus.
- Da gennaio 2007: collaborazione scientifica con il prof. Eduardo Eyras (Australian National University) sul tema dello splicing alternativo (predizione della struttura genica e di ricostruzione di trascritti) che ha portato allo sviluppo del database ASPicDB (pagina web: <http://srv00.recas.ba.infn.it/ASPicDB/>).
- Da gennaio 2013: partecipazione alle attività di ricerca del Laboratorio " -Bioinformatics and Experimental Algorithmics", Dipartimento di Informatica, Sistemistica e Comunicazione, Università di Milano-Bicocca (attuale laboratorio BIAS – Bioinformatica e Algoritmica Sperimentale).
- Da giugno 2014: collaborazione con il prof. Massimiliano Pagni (Università degli Studi di Milano e Istituto Nazionale Genetica Molecolare – INGM) in relazione alla partecipazione alle attività del progetto Fondazione Cariplo 2013-0955 (Modulation of anti cancer immune response by regulatory non-coding RNAs).
- Da novembre 2016: collaborazione con il gruppo di ricerca del prof. Tobias Marschall (Max-Planck-Institut für Informatik, Saarbrücken). La collaborazione ha portato alla stesura di un progetto MCSA-ITN accettato per finanziamento.
- Da novembre 2016: collaborazione con il gruppo di ricerca del prof. Alexander Schoenhuth (CWI, Amsterdam) La collaborazione ha portato alla stesura di un progetto MCSA-ITN e di un progetto MCSA-RISE, entrambi poi accettati per finanziamento.
- RR ha avviato assieme al gruppo di BIAS (Bioinformatica e Algoritmica Sperimentale) collaborazioni all'interno di un progetto H2020 che coinvolge Università di Bielefeld (Germania), Università Comenio di Bratislava (Repubblica Slovacca), Cornell University (US), Università di Tokyo (Giappone), Simon Fraser University (Canada), Pennsylvania State University (US) e Stichting Nederlandse Wetenschappelijk Onderzoek Instituten (Olanda), e coinvolge inoltre i seguenti enti di ricerca: Geneton, s.r.o. (Repubblica Slovacca), Institut Pasteur (Francia) e Illumina Cambridge LTD (Regno Unito)

Partecipazione e organizzazione di eventi

- Relatore al Meeting del Gruppo di Cooperazione Bioinformatica, Certosa di Pontignano 17-18 febbraio 2001. Titolo dell'intervento: "Microarray imaging data pre-processing".
- Relatore al "Third International Workshop, WABI 2003, Budapest, Hungary, September 15-20, 2003". Titolo dell'intervento "A Method to detect gene structure and alternative splice sites by agreeing ESTs to a genomic sequence".
- Membro del comitato organizzatore del Workshop "Alternative Splicing in Animals and Plants", Dipartimento di Informatica Sistemistica e Comunicazione, Università degli studi di Milano-Bicocca, 3 ottobre 2008.
- Relatore al Workshop "Alternative Splicing in Animals and Plants", 3 ottobre 2008, Dipartimento di Informatica Sistemistica e Comunicazione, Università degli studi di Milano-Bicocca. Titolo dell'intervento: "A suffix-tree based method to predict the exon-intron structure of a gene from a set of EST sequences".
- Membro del comitato organizzatore del Workshop "Colloquium on Unconventional Models of Computation". 1-day workshop all'interno di ICTCS'09 (Italian Conference on Theoretical

- Computer Science), Cremona, 28 settembre 2009.
- Relatore al Workshop "Sysbiohealth Symposium, Milano, 25-27 Novembre 2009". Titolo dell'intervento: "EST clustering for gene identification".
 - Relatore a "1st Post-EURASNET Symposium Regulation of Gene Expression through RNA Splicing", International Centre for Theoretical Physics, 24-27 marzo 2012, Trieste. Titolo "PintronNext: a Fast Method for Detecting the Gene Structure due to Alternative Splicing via ESTs, mRNAs, and RNA-Seq Data".
 - Membro del comitato organizzatore della conferenza "Computability in Europe (CiE) 2013: The Nature of Computation", Milano, 1-5 luglio 2013.
 - Relatore a "1° Bioinformatics, Computational Biology and System Biology Lombardy Day", MILANO, 27 novembre 2013. Titolo "ASPicDB-next: an NGS development of a database of alternative splicing events".
 - Membro del comitato organizzatore e scientifico del Workshop "Graph Assembly Algorithms for Omics Data", Dipartimento di Informatica Sistemistica e Comunicazione, Università degli studi di Milano-Bicocca, 18 novembre 2016.
 - Membro del comitato organizzatore e scientifico del Workshop "Workshop on Combinatorial Algorithms in Bioinformatics", Dipartimento di Informatica Sistemistica e Comunicazione, Università degli studi di Milano-Bicocca, 26 gennaio 2018. Pagina web: <https://algolab.eu/2017/12/19/workshop-on-combinatorial-algorithm-in-bioinformatics/>.
 - Membro del comitato organizzatore e scientifico del "First Workshop On Pangenomics Algorithms for Omics Data". June 21st, 2019. Univ. Milano-Bicocca.
 - Membro del comitato di programma della "46th International Conference on Current Trends in Theory and Practice of Computer Science" (SOFSEM 2020).
 - Membro del comitato organizzatore del 7° workshop "Data Structure in Bioinformatics" DSB21, 11-12 febbraio 2021, evento online.

Responsabilità e partecipazione a progetti

- Responsabile dei seguenti Fondi d'Ateneo dal titolo:
 - FA 2017 - Disegno e sperimentazione di approcci algoritmici per problemi computazionali in Bioinformatica
 - FA 2018 - Approcci algoritmici innovativi per la Bioinformatica
 - FA 2019 - Algoritmi e strutture dati efficienti per problemi combinatori
- Beneficiaria del FFABR 2017 - Ricercatori (Finanziamento delle attività base di ricerca)
- Responsabile di un fondo associato al progetto SPAC3 (servizi smart della nuova Pubblica Amministrazione per la Citizen-Centricity in Cloud) - Regione Lombardia.
- Partecipazione al progetto PANGAIA (Pan-genome Graph Algorithms and Data Integration) - Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE). Partecipazione al progetto in qualità di Experienced Researcher quale supervisore di ESR (Early Stage researchers) e nell'organizzazione di eventi e workshops, nonché di docente delle scuole programmate nel progetto.

Attività di Revisione

RR svolge attività di revisione per le seguenti riviste: *Bioinformatics*, *Genomics* e *BMC Bioinformatics*, e per le seguenti conferenze di carattere internazionale: CIE (Computability in Europe), WABI (Workshop on Algorithms in Bioinformatics), ICCABS (IEEE International Conference

on Computational Advances in Bio and Medical Sciences) ISBRA (International Symposium on Bioinformatics Research and Applications) BIBM (IEEE International Conference on Bioinformatics and Biomedicine) ACM-BCB (ACM Conference on Bioinformatics, Computational Biology, and Health Informatics) IWOCA (International Workshop on Combinatorial Algorithms), ICCS (International Conference on Computational Science), SPIRE (International Symposium on String Processing and Information Retrieval).

RR è stata reviewer del progetto internazionale "Modeling Alternative Splicing and its Structural Impact during eVolution" MASSIV (Agence Nationale de la Recherche, ANR).

Attività di Servizio

2015/201, 2019/2020, 2020/2021: organizzazione dell'evento "NERD? Non è roba per donne" in collaborazione con IBM.

Dal 2013 al 2020: partecipazione alla Commissione Orari del Dipartimento di Informatica Sistemistica e Comunicazione dell'Università di Milano –Bicocca.

2020: partecipazione alla commissione giudicatrice dell'esame di stato per la professione di Ingegnere dell'Informazione (senior: 16 luglio, 16 novembre; junior: 24 luglio, 23 novembre).

2020/2021: laboratorio disciplinare "Informatica e analisi del DNA" nell'ambito delle iniziative del Piano delle Lauree Scientifiche (PLS) dell'Università di Milano –Bicocca, con l'obiettivo di introdurre gli studenti delle scuole superiori alla programmazione in linguaggio Python e alla sua applicazione nel contesto della analisi di dati biologici. Le attività svolte nel corso del laboratorio sono state riconosciute come attività PCTO.

Pubblicazioni su riviste indicizzate da scopus

[1] P. Bonizzoni, G. Pesole e R. Rizzi. A Method to detect gene structure and alternative splice sites by agreeing ESTs to a genomic sequence. *Lecture Notes in Computer Science (WABI 2003)*, 2812:63-77.

[2] P. Bonizzoni, R. Rizzi, G. Pesole. ASPIC: a novel method to predict the exon-intron structure of a gene that is optimally compatible to a set of transcript sequences. *BMC Bioinformatics* (2005), 6(1):244.

[3] P. Bonizzoni, R. Rizzi, G. Pesole. Computational methods for alternative splicing prediction. *Briefings in Functional Genomics and Proteomics* (2006), 5:46-51.

[4] T. Castrignanò, R. Rizzi, I. G. Talamo, P. D'Onorio De Meo, A. Anselmo, P. Bonizzoni, G. Pesole. ASPIC: a web resource for alternative splicing prediction and transcript isoforms characterization. *Nucleic Acids Research* (2006), 34:W440-3.

[5] P. Bonizzoni, G. Della Vedova, R. Dondi, G. Fertin, R. Rizzi, S. Vialette. Exemplar Longest Common Subsequence. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2007), 4(4):535-543.

[6] T. Castrignanò, M. D'Antonio, et al. ASPicDB: a database resource for alternative splicing analysis. *Bioinformatics* (2008), 24(10):1300-1304.

- [7] P. Bonizzoni, G. Mauri, G. Pesole, E. Picardi, Y. Pirola, R. Rizzi. Detecting alternative gene structures from spliced ESTs: a computational approach. *Journal of Computational Biology* (2009), 16(1):43-66.
- [8] P. Bonizzoni, G. Della Vedova, R. Dondi, Y. Pirola, R. Rizzi. Minimum factorization agreement of spliced ESTs. *Lecture Notes in Computer Science* (WABI 2009), 5724:1-12.
- [9] P. L. Martelli, M. D'Antonio, P. Bonizzoni, T. Castrignanò, A. M. D'Erchia, P. D'Onorio De Meo, P. Fariselli, M. Finelli, F. L., M. Mangiulli, F. Mignone, G. Pavesi, E. Picardi, R. Rizzi, I. Rossi, A. Valletti, A. Zauli, F. Zambelli, R. Casadio, G. Pesole. ASPicDB: a database of annotated transcript and protein variants generated by alternative splicing. *Nucleic Acids Research* (2011), 39:D80-5.
- [10] P. Bonizzoni, G. Della Vedova, Y. Pirola, R. Rizzi. Plntron: a fast method for gene structure prediction. Proceedings of the 1st IEEE International Conference on Computational Advances in Bio and Medical Sciences ICCABS 2011, 5729935:33-39.
- [11] Y. Pirola, R. Rizzi, E. Picardi, G. Pesole, G. Della Vedova, P. Bonizzoni. Plntron: a fast method for detecting the gene structure due to alternative splicing via maximal pairings of a pattern and a text. *BMC Bioinformatics* (2012), 13(Suppl 5):S2.
- [12] S. Beretta, P. Bonizzoni, G. Della Vedova, R. Rizzi. Reconstructing isoform graphs from RNA-Seq data. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2012, 6392734:499-502.
- [13] S. Beretta, P. Bonizzoni, G. Della Vedova, Y. Pirola, R. Rizzi. (2014). Modeling Alternative Splicing Variants from RNA-Seq Data with Isoform Graphs. *Journal of Computational Biology* (2014), 21:16-40.
- [14] P. Bonizzoni, G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. Constructing String Graphs in External Memory. *Lecture Notes in Computer Science* (WABI 2014), 8701:311-325.
- [15] P. Bonizzoni, G. Della Vedova, G. Pesole, E. Picardi, Y. Pirola, R. Rizzi. (2015) Transcriptome assembly and alternative splicing analysis. *Methods in Molecular Biology* (2015), 1269:173-188.
- [16] P. Bonizzoni, G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. FSG: Fast string graph construction for de novo assembly of reads data. *Lecture Notes in Computer Science* (ISBRA 2016), 9683:27-39.
- [17] Palorini, R., Votta, G., Pirola, Y., De Vitto, H., De Palma, S., Airoidi, C., et al. Protein Kinase A Activation Promotes Cancer Cell Resistance to Glucose Starvation and Anoikis. *PLoS Genetics* (2016), 12(3): e1005931.
- [18] P. Bonizzoni, G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. LSG: An External-Memory Tool to Compute String Graphs for Next-Generation Sequencing Data Assembly. *Journal of Computational Biology* (2016), vol. 23, p. 137-149.
- [19] S. Beretta, P. Bonizzoni, L. Denti, M. Previtali, R. Rizzi. Mapping RNA-seq data to a transcript graph via approximate pattern matching to a hypertext. *Lecture Notes in Computer Science* (ALCoB 2017), 10252:49-61.

- [20] P. Bonizzoni, A.P. Carrieri, G. Della Vedova, R. Rizzi, G. Trucco. Species-driven persistent phylogeny. *Fundamenta Informaticae* (2017), 154:47-63.
- [21] G. Della Vedova, M. Patterson, R. Rizzi, M. Soto. Character-based phylogeny construction and its application to evolution. *Lecture Notes in Computer Science* (CiE 2017), 10307:3-13.
- [22] P. Bonizzoni, A.P. Carrieri, G. Della Vedova, R. Rizzi, G. Trucco. A colored graph approach to perfect phylogeny with persistent characters. *Theoretical Computer Science* (2017), 658:60-73.
- [23] P. Bonizzoni, G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. An External-Memory Algorithm for String Graph Construction. *Algorithmica* (2017), 78:394-424.
- [24] P. Bonizzoni, G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. FSG: Fast String Graph Construction for de Novo Assembly. *Journal of Computational Biology* (2017), vol. 24, p. 953-968.
- [25] P. Bonizzoni, G. Della Vedova, S. Nicosia, Y. Pirola, M. Previtali R. Rizzi. Divide and Conquer computation of the multi-string BWT and LCP array. *Lecture Notes in Computer Science* (CiE 2018), 10936:107-117.
- [26] L. Denti, R. Rizzi, S. Beretta, G. Della Vedova, M. Previtali, P. Bonizzoni. ASGAL: aligning RNA-Seq data to a splicing graph to detect novel alternative splicing events. *BMC Bioinformatics* (2018), 19(1):444.
- [27] R. Rizzi, S. Beretta, M. Patterson, Y. Pirola, M. Previtali, G. Della Vedova, P. Bonizzoni. Overlap graphs and de Bruijn graphs: data structures for de novo genome assembly in the big data era. *Quantitative Biology* (2019), 7(4):278-292.
- [28] P. Bonizzoni, G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. Multithread Multistring Burrows-Wheeler Transform and Longest Common Prefix Array. *Journal of Computational Biology* (2019), 26(9):948-961.
- [29] P. Bonizzoni., G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. Computing the multi-string BWT and LCP array in External Memory. *Theoretical Computer Science* (2021). <https://doi.org/10.1016/j.tcs.2020.11.041>. ISSN = 0304-3975.
- [30] Luca Denti, Yuri Pirola, Marco Previtali, Tamara Ceccato, Gianluca Della Vedova, Raffaella Rizzi, Paola Bonizzoni, Shark: fishing relevant reads in an RNA-Seq sample, *Bioinformatics* (2021), 37(4):464-472. <https://doi.org/10.1093/bioinformatics/btaa779>

Partecipazione a conferenze e a workshops internazionali

- [31] L. Milanesi, R. Rizzi. Microarray Imaging Data Reader SPOTVIEW. Proceedings of the Third International Conference on Bioinformatics of Genome Regulation and Structure BGRS 2002, Novosibirsk 14-20 Luglio 2002, Vol. 2, p. 196-198.
- [32] S. Beretta, R. Rizzi, G. Della Vedova, P. Bonizzoni. Alternative Splicing from RNA-Seq Data without the Genome. AS-SIG-11, Alternative Splicing - Special Interest Group Meeting, Vienna, 14-16 Luglio 2011.
- [33] L. Denti, R. Rizzi, S. Beretta, G. Della Vedova, M. Previtali, P. Bonizzoni. ASGAL: Aligning

RNA-Seq Data to a Splicing Graph to Detect Novel Alternative Splicing Events. RECOMB 2018, Parigi (Francia), 21-24 aprile 2018 (Best Poster Award).

Partecipazione a conferenze e a workshops nazionali

[34] G. Forlani, C. Nardinocchi, R. Rizzi. 3D Linear Feature Extraction and Matching. International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, Part 6W4, Milano 1998, p. 169-174.

[35] L. Milanesi, R. Rizzi. Gene Functional Mapping. International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, Part 6W7, Milano 1999, p. 321-328.

[36] L. Milanesi, I. Rogozin, R. Rizzi. Application of ESTs mapping to improve gene prediction methods. Gruppo di Cooperazione Bioinformatica, Certosa di Pontignano, 3-4 Marzo 2000.

[37] L. Milanesi, R. Rizzi. Bioinformatics applications to human genome analysis. Proceedings of XIII International Conference "Computing in Clinical Laboratories", Milano 21-23 Settembre 2000.

[38] R. Rizzi, M. Falchi, L. Milanesi. Microarray imaging data pre-processing. Gruppo di Cooperazione Bioinformatica, Certosa di Pontignano, 17-18 febbraio 2001.

[39] R. Rizzi, L. Milanesi, L. Mussio. Un Tool per la Lettura delle Immagini di Espressione da Microarray. Atti della Conferenza La qualit` nell'informazione Geografica, V Conferenza Nazionale dell'ASITA, Rimini 9-12 Ottobre 2001, Vol. 2, p. 1285-1290.

[40] R. Rizzi, L. Milanesi, L. Mussio. Pre-processamento di Dati di Espressione da Microarray. Atti della Conferenza La qualit` nell'informazione Geografica, V Conferenza Nazionale dell'ASITA, Rimini 9-12 Ottobre 2001, Vol. 2, p. 1291-1296.

[41] L. Milanesi, R. Rizzi. Microarray Imaging Data Pre-processing. Proceedings of the IEEE-EMBS Special Topic Conference on Molecular, Cellular and Tissue Engineering, Genova 6-9 Giugno 2002, p. 182-183.

[42] R. Rizzi, L. Milanesi, L. Mussio. L'Integrazione di Banche Dati Genomiche e Proteiche. Atti della Conferenza Geomatica per l'ambiente, il territorio e il patrimonio culturale, VI Conferenza Nazionale dell'ASITA, Perugia 5-8 Novembre 2002, Vol. 2, p. 1803-1808.

[43] L. Milanesi, R. Rizzi. Microarray Image Data Analysis SPOTVIEW. NETTAB 2003, Bologna 27-28 Novembre 2003.

[44] P. Bonizzoni, R. Dondi, R. Rizzi, G. Pesole. ASPIC: a novel method to predict alternative splicing. BITS 2004, Bioinformatics Italian Society Meeting, Padova, 26-27 Marzo 2004.

[45] T. Castrignanò, M. D'Antonio, P. D'Onorio De Meo, N. Sanna, R. Rizzi, F. Mignone, P. Bonizzoni, G. Pesole. HPC for Aspic: a distributed web resource for alternative splicing prediction and transcript isoform characterization. BITS 2007, The 2007 Annual Meeting of the Italian Bioinformatics Society (BITS), Napoli, 26-28 Aprile 2007.

[46] T. Castrignanò, M. D'Antonio, P. D'Onorio De Meo, N. Anselmo, A. D'Erchia, F. Licciulli, M. Mangiulli, F. Mignone, G. Pavesi, E. Picardi, R. Rizzi, P. Bonizzoni, G. Pesole. Automated data

retrieval of alternative splicing sites and transcript isoforms in eukaryotic organisms. BITS 2009, Sixth Annual Meeting Bioinformatics Italian Society, Genova, 18-20 Marzo 2009.

[47] P. Bonizzoni, G. Della Vedova, F. Mignone, G. Mauri, Y. Pirola, R. Rizzi. EST clustering for gene identification. AS-SIG-09, Alternative Splicing - Special Interest Group Meeting, Stoccolma 27-28 Giugno 2009. Lo stesso lavoro è stato presentato al workshop: *Sysbiohealth Symposium*, Milano, 25-27 Novembre 2009.

[48] S. Beretta, P. Bonizzoni, G. Della Vedova, R. Rizzi. Alternative Splicing from RNA-seq Data without the Genome. 3° NGS Workshop, Bari, 12-14 October 2011.

[49] S. Beretta, P. Bonizzoni, G. Della Vedova, R. Rizzi. NGS workshop Bari. Alternative Splicing from RNA-seq data without the genome. Third Next Generation Sequencing Workshop, Bari, 12-14 Ottobre 2011.

[50] Y. Pirola, R. Rizzi, S. Beretta, E. Picardi, G. Pesole, G. Della Vedova, P. Bonizzoni. PlntronNext: a fast method for detecting the gene structure due to alternative splicing via ESTs, mRNAs, and RNA-Seq data. 1st Post-EURASNET Symposium on "Regulation of Gene Expression through RNA Splicing", Trieste, 24-27 Marzo 2012.

[51] M. D'Antonio, P.L. Martelli, P. Bonizzoni, T. Castrignanò, G. Della Vedova, P. D'Onorio De Meo, P. Fariselli, M. Giulietti, F. Licciulli, F. Mignone, G. Pavesi, Y. Pirola, F. Piva, E. Picardi, R. Rizzi, F. Zambelli, R. Casadio, G. Pesole. ASPicDB: a database of annotated transcript and protein variants generated by alternative splicing. 1st Post-EURASNET Symposium on "Regulation of Gene Expression through RNA Splicing", Trieste, 24-27 Marzo 2012.

[52] S. Beretta, P. Bonizzoni, G. Della Vedova, Y. Pirola, M. Previtali, R. Rizzi. ASPicDB-next: an NGS development of a database of alternative splicing events. 1° Bioinformatics, Computational Biology and System Biology Lombardy Day, Milano, 27 novembre 2013.