

Matteo Comin

Curriculum Vitae

Personal Data:

Name:	Matteo Comin
Data of birth:	02/20/1977
Place of birth:	Venice, Italy
Citizenship:	Italian
Military status:	Honorable discharge
E-mail:	comin@dei.unipd.it
Office:	+39 049 8277935
Fax:	+39 049 8277799

Research Summary:

Dr. Comin research interests focus on the area of algorithms for computational biology. During his activity he has been research intern at IBM T.J.Watson Research Center twice where he developed several motif discovery systems for biological sequences, now part of their Bioinformatics and Pattern Discovery Toolbox. He is interested in computational methods for protein structural comparison and protein-protein docking prediction, and also in developing algorithms for next-generation sequencing. He has been a visiting researcher at University of Purdue (US) once and at Universitat Politcnica de Catalunya, Barcelona (Spain) three times. Some of his results have been considered by IBM Research of great importance and three of his ideas are now U.S. patent. In 2007 he received from the University of Padova the C. Offelli Award for best young researcher. Since 2007 he is a tenured Assistant Professor at the University of Padova.

Education:

- 2007 - Ph.D. in Computer Science, Department of Information Engineering, University of Padova.
- 2004 - 2006. Doctoral Student, Ph.D program in Computer Science, Department of Information Engineering, University of Padova (XIX cycle).
Thesis title: “Motif Discovery and Data Compression: Models, Methods and Applications”.
Supervisor: prof. Alberto Apostolico.
- 2003 - Qualifying for engineering profession.
- 2003 - Laurea degree in Computer Science, Department of Information Engineering University of Padova. Supervisor: Prof. Alberto Apostolico.
- 1996 - High School Diploma in Computer Science, C. Zuccante Institute, Mestre (VE).

Professional Experience:

- 2012 September - October. Visiting Scientist at the Barcelona Supercomputing Center and Universitat Politcnica de Catalunya, Barcelona.
- 2011 Octobr - December. Visiting Scientist at the Barcelona Supercomputing Center and Universitat Politcnica de Catalunya, Barcelona.
- 2010 August - September. Visiting Scientist at the Barcelona Supercomputing Center and UPC Barcelona Tech.
- 2007 October - present. Ricercatore (Tenured Assistant Professor), Department of Information Engineering, University of Padova.
- 2007 March - September. Research Fellow at the Department of Information Engineering, University of Padova.
- 2006 June - October. Research intern in the group “Genographic”, supervisor Dr. Ajay Royyuru, IBM T.J. Watson Research Center, Yorktown Heights, New York, U.S..
- 2006 April. Member of the Organizing Committee of the “Tenth Annual International Conference on Research in Computational Molecular Biology, RECOMB 2006”, April 2-5, 2006, Venice, Italy.
- 2004 September - December. Research intern in the group “Bioinformatics and Pattern Discovery”, supervisor Dr. Isidore Rigoutsos, IBM T.J. Watson Research Center, Yorktown Heights, New York, U.S..
- 2004 - 2006. Doctoral Student, Ph.D program in Computer Science, Department of Information Engineering, University of Padova.

- 2003 April - December. Research Collaborator, F.I.R.B. Project “Enabling platforms for high-performance computational grids”, Supervisor: prof. A. Apostolico, MI.U.R., 2003-2005.
- 2002 October - November. Research Fellowship: Research on Algorithms and Data Compression, Computer Sciences Department, Purdue University, Indiana, US.

Funded Projects (involved as Research Assistant or PI):

- P.R.I.N.: “Compositional approaches for characterization and mining of omics-data”, MI.U.R. 2013-2016, PI C. Pizzi.
- Ateneo Project “Pattern discovery and computational problems for whole genomes comparison”, University of Padova 2012-14, PI M. Comin.
- Cariparo Research Project, “Inference of transcriptional regulatory pathways from integrative analysis of gene and protein data”, Fondazione Cariparo 2008-2010, PI Prof. G. Toffolo.
- Ateneo Project “Computational assessment of Protein-Protein interaction networks: target prediction and validation as guide for modern system biology”, University of Padova 2008-10, PI Prof. C. Guerra.
- Italy-Israel F.I.R.B Project “Pattern Discovery Algorithms in Discrete Structures, with Applications to Bioinformatics”, MI.U.R. 2006-2009, PI Prof. R. Giancarlo.
- P.R.I.N.: “Data mining methods for e-business applications”, MI.U.R. 2005-2006.
- F.I.R.B.: “Enabling platforms for high-performance computational grids”, Ministry of University and Research (MI.U.R.) 2003-2005.
- F.I.R.B.: “Bioinformatics for genomics and proteomics”, MI.U.R. 2003-2005.

Teaching:

- A.Y. 2012/2013 - Responsible of the course “Algorithms for Bioinformatics”, Dept. of Information Engineering, University of Padova.
- Since A.Y. 2008/2009 - now; responsible of the course “Computer Architectures”, Dept. of Information Engineering, University of Padova.
- A.Y. 2007/2008 - Lecturer within the course “Three-Dimensional Data Elaboration”, Prof. C. Guerra, Dept. of Information Engineering, University of Padova.
- A.Y. 2007/2008 - Lecturer within the course “Computer Networks”, Prof. C. Guerra, Dept. of Information Engineering, University of Padova.
- A.Y. 2005/2006 - Lecturer within the course “Distributed Systems”, Prof. C. Ferrari, Dept. of Information Engineering, University of Padova.

- A.Y. 2005/2006 - Lecturer within the course “Three-Dimensional Data Elaboration”, Prof. C. Guerra, Dept. of Information Engineering, University of Padova.
- A.Y. 2005/2006 - Lecturer within the course “Computer Networks”, Prof. C. Guerra, Dept. of Information Engineering, University of Padova.
- A.Y. 2004/2005 - Lecturer within the course “Bioinformatics and Computational Biology”, Prof. A. Apostolico, Dept. of Information Engineering, University of Padova.

Other University Services:

- Since 2009 he is member of the ”Didactic Commission” for the Laurea degree in Information Engineering, University of Padova.
- 2009-2012 member of the Ph.D school in Information Engineering, University of Padova.
- Advisor of more than twenty undergraduate/master thesis.

Invited Seminars (excluding conferences talks):

- 2012 - “Whole genome comparison using non-overlapping patterns”, Dagstuhl seminar in “Structure Discovery in Biology: Motifs, Networks and Phylogenies”.
- 2011 - “Phylogenetic methods for biological sequences”, Lipari Summer School: “Biological Sequence Analysis and High Throughput Technologies”.
- 2010 - “Recent Challenges in Computational Biology”, Barcelona Supercomputing Center, UPC Barcelona Tech.
- 2010 - “Classification of Protein Sequences”, Dagstuhl seminar in “Structure Discovery in Biology: Motifs, Networks and Phylogenies”.
- 2010 - “Phylogenetic methods for biological sequences”, Lipari Summer School: Statistical and Machine Learning Methods in Computational Biology
- 2009 - “RNA-protein interaction”, Lipari Summer School in RNAs : structure, function and therapy.
- 2007 - “From Conventional Biology to Computational Biology”, University of Verona.
- 2007 - “Some Results on Computational Biology”, *C. Offelli Award* for best young researcher, Dept. of Information Engineering, University of Padova.
- 2006 - “Mining Regularities: The Experience with Biological Data”, University of Trieste.
- 2005 - “Conservative Extraction of Over-Represented Extensible Motifs”, presented at the DIMACS Workshop on Detecting and Processing Regularities in High Throughput Biological Data, Rutgers University, New Jersey, US.

- 2005 - “Conservative Extraction of Over-Represented Extensible Motifs”, IBM T.J. Watson Research Center, New York, U.S..
- 2005 - “Pattern Discovery and Data Compression: Models, Methods and Tools”, Dept. of Information Engineering, University of Padova.
- 2004 - “From Motif Discovery to Off-line Data Compression”, research meeting within the “International School on Advanced BioMedicine and BioInformatics” Lipari Island, June 2004.

Former students/Ph.D. Advisor :

- Research Fellow Michele Schimd (2013-2014).
- Ph.D. student Melanie Grieb (2010-2013), University of Ulm (external PhD Advisor).
- Ph.D. student Davide Verzotto (2009-2011), now at Singapore Genomic Institute.

Program Committees:

- 14th Workshop on Algorithms in Bioinformatics (WABI) 2014.
- Biological Knowledge Discovery and Data Mining (BIOKDD) 2014.
- IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) 2013.
- 9th International Symposium on Bioinformatics Research and Applications (ISBRA) 2013.
- International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI) 2013.
- International Symposium on Network Analysis and Mining for Health Informatics, Biomedicine and Bioinformatics (Net-HI-BI-BI) 2013.
- IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) 2012.
- 8th International Symposium on Bioinformatics Research and Applications (ISBRA) 2012.
- 23rd Genome Informatics Workshop (GIW) 2012.
- IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) 2011.
- Biological Knowledge Discovery and Data Mining (BIOKDD) 2011.
- Biological Knowledge Discovery and Data Mining (BIOKDD) 2010.
- 21st Annual Symposium on Combinatorial Pattern Matching (CPM) 2010.

- 9th Network Tools and Applications in Biology (NETTAB) 2009.

Other Committees:

- Member of the Organizing Committee of the “The 9th IAPR conference on Pattern Recognition in Bioinformatics, PRIB 2014”, August 21-23, 2014, Stockholm, Sweden.
- Member of the Organizing Committee of the “Tenth Annual International Conference on Research in Computational Molecular Biology, RECOMB 2006”, April 2-5, 2006, Venice, Italy.

Editorial Boards:

- Guest co-Editor: Springer series of Lecture Notes in Bioinformatics. Proceedings of “The 9th IAPR conference on Pattern Recognition in Bioinformatics, PRIB 2014”, 2014.
- Guest co-Editor: BMC Bioinformatics, Special Issue of PRIB 2014.

Reviewer for:

Wiley - book proposals.

Projects:

The French National Research Agency (ANR), Italian Ministry of Instruction, University and Research (MIUR).

International Journals:

Briefings in Bioinformatics, BMC Algorithm for Molecular Biology (AMB), BMC Bioinformatics, Journal of Bioinformatics Research and Applications (IJBRA), ACM/IEEE Transaction in Computational Biology and Bioinformatics (TCBB), Journal of Computational Biology (JCB), Parallel Processing Letters (PPL), ACM Journal of Experimental Algorithmics (JEA), Theoretical Computer Science (TCS), Journal of Foundation of Computer Science (JFCS), Journal of Visual Communication and Image Representation (JVCI).

International Conferences:

Research in Computational Molecular Biology (RECOMB), Workshop on Algorithms in Bioinformatics (WABI), The Asia-Pacific Bioinformatics Conference (APBC), Computational Systems Bioinformatics (CSB), Combinatorial Pattern Matching (CPM), International Workshop in Bioinformatics and Research Applications (IWBRA), IEEE International Conference on Data Mining (ICDM), European Conference on Parallel Computing (EuroPar).

Awards and Fellowships:

- 2010, 2011 and 2012 HPC-Europa visiting fellowship, to support the research visits at the Barcelona Supercomputing Center.

- 2007 *C. Offelli Award* for best young researcher, Department of Information Engineering, University of Padova.
- 2005 International Society of Computational Biology Travel Fellowship to attend the “13th Annual International Conference on Intelligent Systems for Molecular Biology” (ISMB 2005), Detroit, USA.
- 2005 U.S. National Science Foundation Travel Fellowship to attend the workshop “DIMACS Workshop on Detecting and Processing Regularities in High Throughput Biological Data”, DIMACS Center, Rutgers University, New Jersey, USA.
- 2004 U.S. National Science Foundation Travel Fellowship to attend the workshop “DIMACS Working Group on The Burrows - Wheeler Transform: Ten Years Later”, DIMACS Center, Rutgers University, New Jersey, USA.

Patents:

Inventor of the following IBM patents:

- “Method and structure for lossy compression of continuous data with extensible motifs”, US Patent 7259701, 2007.
- “Methods and systems for conservative extraction of over-represented extensible motifs”, US Patent 7,865,313, 2011.
- “Method and apparatus for detection consensus motifs in data sequences”, United States Patent Application 20080167850, filed.

PUBLICATIONS

Books and Book chapters:

1. M. Comin, L. Kall, E. Marchiori, A. Ngom, J. C. Rajapakse. Proceedings of “The 9th IAPR conference on Pattern Recognition in Bioinformatics, PRIB 2014”. Springer series of Lecture Notes in Bioinformatics, 2014.
2. M. Comin, L. Kall, E. Marchiori, A. Ngom, J. C. Rajapakse. Special Issue of PRIB 2014. BMC Bioinformatics, 2014.
3. M. Comin, D. Verzotto, “Alignment-Free Measures for Whole-Genome Comparison”, in “Pattern Recognition in Computational Molecular Biology: Techniques and Approaches”, Edited by M. Elloumi, C. S. Iliopoulos, J.T.L. Wang and A.Y. Zomaya. Wiley, 2014.
4. M. Comin, D. Verzotto, “Comparing ranking and filtering motifs with character classes: application to biological sequence analysis”, in “Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing of Biological Data”, Edited by M. Elloumi and A. Y. Zomaya Wiley, 2013.
5. M. Comin, C. Guerra, “Protein Structure Comparison, High Performance Computing”, in “Encyclopedia of System Biology” Edited by W. Dubitzky, O. Wolkenhauer, K.H. Cho and H. Yokota. Springer, 2013.
6. M. Comin, D. Verzotto, “Remote Homology Detection of Protein Sequences”, in Dagstuhl Proceedings 10231 “Structure Discovery in Biology: Motifs, Networks and Phylogenies”, 2010: 2741.
7. M. Comin, C. Guerra, G. Zanotti, “Invariant geometric properties of secondary structure elements in proteins”, in “Biological Data Mining” edited by J. Chen and S. Lonardi, Chapman & Hall/CRC, 2009; Chapter 2, pp. 27-49.
8. M. Comin, L. Parida, “Detection of Subtle Variations as Consensus Motifs”. in “SAIL String Algorithms, Information and Learning” edited by R. Giancarlo and S. Lonardi, special issue of *Theoretical Computer Science*, Vol. 395 No. 2-3 (2008), pp. 158-170.
9. A. Apostolico, M. Comin, L. Parida, “Bridging Lossy and Lossless Compression by Motif Pattern Discovery”, General Theory of Information Transfer and Combinatorics, *Lecture Notes in Computer Science*, Vol. 4123 (2006), pp. 793-813.
10. M. Cannataro, M. Comin, C. Ferrari, C. Guerra, A. Guzzo, P. Veltri, “Modelling a Protein Structure Comparison Application on the Grid using PROTEUS”, SAG 2004, *Lecture Notes in Computer Science* Vol. 3458 (2005), pp. 75-85.

International Journals:

11. M. Comin, M. Farreras, “Parallel Continuous Flow: A Parallel Suffix Tree Construction Tool for Whole Genomes”, Accepted for publication in *Journal of Computational Biology* 2014. (IF: 1.6)

12. M. Comin, D. Verzotto, "Filtering Degenerate Motifs with Application to Protein Sequence Analysis", *Algorithms* 2013, Vol. 6, no. 2: pp. 352-370.
13. A. Apostolico, M. Comin, A. Dress, L. Parida, "Ultrametric Networks: A New Tool For Phylogenetic Analysis", In *BMC Algorithms for Molecular Biology* 2013, 8:7. (IF: 1.61)
14. M. Comin, D. Verzotto, "Alignment-Free Phylogeny of Whole Genomes using Underlying Subwords", In *BMC Algorithms for Molecular Biology* 2012, 7:34. (IF: 1.61)
15. M. Comin, D. Verzotto, "The Irredundant Class Method for Remote Homology Detection of Protein Sequences", in *Journal of Computational Biology* 2011, 18(12): 1819-1829. (2011 IF: 1.86)
16. A. Apostolico, M. Comin, L. Parida, "VARUN: Discovering Extensible Motifs under Saturation Constraints", in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, October-December 2010 (vol. 7 no. 4) pp. 752-762. (2010 IF: 2.1)
17. M. Comin, F. Dellaert, C. Guerra, "Binding Balls: Fast detection of Binding Sites using a property of Spherical Fourier Transform", *Journal of Computational Biology*, Vol. 16 No. 11: pp. 1577-1591, (2009) . (2009 IF: 1.7)
18. M. Comin, C. Guerra, G. Zanotti, "Mining Over-Represented 3DPatterns of Secondary Structures in Proteins", *Journal of Bioinformatics and Computational Biology*, Vol. 6 , No. 06 (2008), pp. 1067 - 1087. (2008 IF: 1.3)
19. G. Ciriello, M. Comin, C. Guerra, "Algorithmic Re-Structuring and Data Replication for Protein Structure Comparison on a GRID", *Future Generation Computer System* (23) Special Issue on Life Science Grids for Biomedicine and Bioinformatics, (2007), pp. 391-397. (IF: 1.86)
20. A. Apostolico, M. Comin, L. Parida, "Mining, Compressing and Classifying with Extensible Motifs", *BioMed Central Algorithms for Molecular Biology*, Vol. 1:4, (2006). (Highly Accessed, IF 1.61)
21. A. Apostolico, M. Comin, L. Parida, "Bridging lossy and lossless compression by motif pattern discovery", (extended abstract), *Electronic Notes in Discrete Mathematics*, Vol. 21 (2005), pp. 219-225.
22. M. Comin, C. Guerra, G. Zanotti, "PROuST: a comparison method of three-dimensional structures of proteins using indexing techniques", *Journal of Computational Biology*, Mary Ann Liebert Inc. , Vol. 11, No. 6 (2004), pp. 1061-1072. (2004 IF: 3.78)
23. M. Comin, C. Ferrari, C. Guerra, "Grid deployment of bioinformatics applications: a case study in protein similarity determination", *Parallel Processing Letters*, World Scientific Publishing Company, Vol. 14 No. 2 (2004), pp. 163-176.

International Conferences:

24. M. Comin, D. Verzotto, “Beyond fixed-resolution alignment-free measures for mammalian enhancers sequence comparison”, accepted for presentation at The Twelfth Asia Pacific Bioinformatics Conference 2014. Proceedings in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2014. (IF: 1.7)
25. M. Comin, D. Verzotto, “Reducing the space of degenerate patterns in protein remote homology detection”, Proceedings of *24rd International Workshop on Database and Expert Systems Applications*, BIOKDD 2013, pp. 76-80.
26. M. Comin, M. Farreras, “Efficient Parallel Construction of Suffix Trees for Genomes Larger than Main Memory”, Proceedings of *the 20th European MPI* 2013, pp. 211-216.
27. M. Antonello, M. Comin, “Fast Computation of Entropic Profiles for the Detection of Conservation in Genomes”, Proceedings of *Pattern Recognition in Bioinformatics* 2013, Lecture Notes in Bioinformatics (LNBI) 2013, 7986, pp. 277–288. (accepted 39%)
28. M. Comin, D. Verzotto, “Whole-Genome Phylogeny by Virtue of Unic Subwords”, Proceedings of *23rd International Workshop on Database and Expert Systems Applications*, BIOKDD 2012. pp. 190-194.
29. M. Comin, D. Verzotto, “Classification of Protein Sequences by means of Irredundant Patterns”, Proceedings of Asia-Pacific Bioinformatics Conference, 2010. *BMC Bioinformatics*, 11(Suppl 1):S16, 2010. (accepted 28%, IF: 3.02)
30. M. Comin, C. Guerra, G. Zanotti, “Mining Over-Represented 3D Patterns of Secondary Structures in Proteins”, Proceedings of *7th International Workshop on Data Mining in Bioinformatics (BIOKDD '07)*, pp 19-26. (accepted 25%)
31. M. Comin, L. Parida, “Subtle Motifs Discovery for Detection of DNA Regulatory Sites”, Proceeding of *Asian-Pacific Bioinformatics Conference*, APBC 2007, pp. 27-36. (accepted 17%)
27. G. Ciriello, M. Comin, C. Guerra, “Application Re-Structuring and Data Management on a GRID Environment: a Case Study for Bioinformatics”, Proceedings of *IEEE International Workshop on High Performance Computational Biology*, 2006:9.
28. G. Ciriello and M. Comin, “High-Performance Protein Structure Comparison”, Proceedings of the *Tenth Annual International Conference on Research in Computational Molecular Biology*, RECOMB (2006).
29. A. Apostolico, M. Comin, L. Parida, “Conservative Extraction of Over-represented Extensible Motifs”, Proceeding of the 13th Annual International conference on Intelligent Systems for Molecular Biology (ISMB 2005) Supplement of *Bioinformatics* 2005: pp. 9-18. (accepted 13%, impact factor 6.02)
30. A. Apostolico, M. Comin, L. Parida, “Off-line Compression by Extensible Motifs”, Proceedings of *IEEE Data Compression Conference*, Computer Society Press, pp. 450, (2005).

31. A. Apostolico, M. Comin, L. Parida, “Motifs in Ziv-Lempel-Welch Clef”, Proceedings of *IEEE Data Compression Conference*, Computer Society Press, pp. 72-81, (2004). (accepted 20%)

Other Manuscripts:

32. M. Comin, “Motif Discovery and Data Compression: Models, Methods and Applications”. Ph.D Thesis, Ph.D Program in Computer Science, Department of Information Engineering, University of Padova (XIX cycle). (supervisor: Prof. Alberto Apostolico)
33. M. Comin, “Motivi Strutturali e Compressione Fuori Linea”. Laurea Thesis, Computer Science, Department of Information Engineering, University of Padova.