Murray Patterson

Curriculum Vitae

Education

2006—2012   **Doctor of Philosophy**, *Department of Computer Science, University of British Columbia*, Vancouver, Canada, *GPA − 4.0/4*.
Advisors: Ján Maňuch, Cedric Chauve and Arvind Gupta
Thesis title: "Variants of the consecutive-ones property motivated by the reconstruction of ancestral species"

2004—2006   **Master of Science**, *School of Computing Science, Simon Fraser University*, Burnaby, British Columbia, Canada, *GPA − 3.9/4*.
Advisors: Eugenia Ternovska and Arvind Gupta
Thesis title: "Grounding for model expansion in k-guarded formulas"

1999—2003   **Bachelor of Computer Science (Honours)**, *Jodrey School of Computer Science, Acadia University*, Wolfville, Nova Scotia, Canada, *GPA − 3.8/4 (top 5%)*.
Advisors: Raymond J. Spiteri and Jim Diamond
Thesis title: "Implementing Runge-Kutta solvers in Java"

Work Experience

2020—   **Assistant Professor**, *Department of Computer Science, Georgia State University*, Atlanta, Georgia, USA.
Teaching both undergraduate and graduate courses in computer science, as well as carrying out my research program in bioinformatics

2019—2020   **Visiting Assistant Professor**, *Department of Computer Science and Engineering, Fairfield University*, Fairfield, Connecticut, USA.
Teaching both undergraduate and graduate courses in computer science, as well as carrying out my research program in bioinformatics

2016—2019   **Postdoctoral Researcher**, *Experimental Algorithmics Lab (AlgoLab), Informatics Department, Università degli Studi di Milano-Bicocca*, Milano, Italy.
Advisors: Gianluca Della Vedova and Paola Bonizzoni
Performing research on haplotype assembly, cancer phylogenetics and their application

2014—2016   **Postdoctoral Researcher**, *Laboratoire de Biométrie et Biologie Évolutive (LBBE), Université Claude Bernard Lyon 1*, Villeurbanne, France.
Advisors: Daniel Kahn and Vincent Daubin
Studying the co-evolution of (metabolic) function, and how it can be used to help the construction of phylogenetic trees

1 Park Place, Room 636 – Atlanta, Georgia 30303 (USA)
☎ *+1 475 330 0978* • ✉ *mpatterson30@gsu.edu*
🌐 *murraypatterson.github.io*   *1/9*

2013—2014 **ERCIM Postdoctoral Research Fellow**, *Life Sciences Group, Centrum Wiskunde Informatica (CWI)*, Amsterdam, The Netherlands.
Advisors: Alexander Schönhuth and Gunnar W. Klau
Improving the state-of-the-art haplotype assembly methods for next-generation genome sequencing data
Research funded by ERCIM Marie Curie ABCDE Postdoctoral Fellowship

2012—2013 **ERCIM Postdoctoral Research Fellow**, *INRIA Rhône-Alpes and the Laboratoire de Biométrie et Biologie Évolutive, Université Claude Bernard Lyon 1*, Villeurbanne, France.
Advisor: Eric Tannier
Performing a study of the co-evolution of genes, and how it can be used to help the construction of phylogenetic trees
Research funded by ERCIM Marie Curie ABCDE Postdoctoral Fellowship

Contracts

2022 **Research Agreement**, *AI Learning Alliance*, Facebook (Meta), Inc., Amount: $50,000 USD (2 years).
PI: Anu Bourgeois
To support work related to providing education at the frontier of R&D and to increase representation of historically marginalized researchers and engineers in machine learning to prepare learners for jobs in the artificial intelligence field

2021 **Collaboration Agreement**, *Hanyang University ERICA*, Industry-University Cooperation Foundation, Amount: $25,000 USD (1 year).
PI: Zhipeng Cai
For the development of online modules in computer science and entrepreneurship as part of a global education program

2018 **Project Collaboration**, *Experimental Algorithmics Lab, Informatics Department, Università degli Studi di Milano-Bicocca*, Milano, Italy, Amount: 5,000 € (2 months).
Advisor: Raffaella Rizzi
The development of efficient algorithms for comparing evolutionary histories of cancerous tumours as inferred by the latest methods

2017 **Project Collaboration**, *Experimental Algorithmics Lab, Informatics Department, Università degli Studi di Milano-Bicocca*, Milano, Italy, Amount: 5,000 € (2 months).
Advisor: Gianluca Della Vedova
Studying phylogenetic models and how they can be used to model the evolution of tumours. This project resulted in a publication at CiE 2017

Teaching Experience

Spring 2022 **Georgia State University**.
○ CSC 4330 / 6330 — a senior year / masters-level programming language concepts course

Fall 2021 **Georgia State University**.
○ CSC 4710 / 6710 — a senior year / masters-level course on database systems

Spring 2021 **Georgia State University**.
○ CSC 4330 / 6330 — a senior year / masters-level programming language concepts course

1 Park Place, Room 636 – Atlanta, Georgia 30303 (USA)
📞 *+1 475 330 0978* • ✉ *mpatterson30@gsu.edu*
🌐 *murraypatterson.github.io* *2/9*

Fall 2020	**Georgia State University**.
	○ CSC 4330 / 6330 — a senior year / masters-level programming language concepts course
Spring 2020	**Fairfield University**.
	○ CS 131 — an introductory programming course in Java, and
	○ CS 354 — a junior year theory of programming languages course using Haskell and Prolog
Fall 2019	**Fairfield University**.
	○ CS 131 — an introductory programming course in Java
	○ CS 231 — an intermediate (sophomore year) programming course, focusing on the object-oriented features of Java, and
	○ SW 402 — a senior-year / masters-level course on database management systems
2015	**Instructor**, *Department of Mathematics, Université Claude Bernard Lyon 1*, Villeurbanne, France.
	Section head: Raymond Lobry
	Teaching a section of mathematics for the life sciences to first-year university students

Advising

Spring 2022	PhD students Zahra Tayebi, Sarwan Ali and Prakash Chourasia, and Master's students Babatunde Bello and Supriya Arun
Fall 2021	PhD students Zahra Tayebi, Sarwan Ali and Prakash Chourasia, Master's students Babatunde Bello and Kaustubh Khandai, and Honors student Yijing Zhou. Sarwan Ali won an MBD fellowship, funding the remainder of his study. Yijing Zhou graduated with honors thesis titled: "Towards a complete workflow for inferring tumor phylogeny from raw single-cell sequencing reads"
Spring 2021	PhD students Zahra Tayebi and Sarwan Ali
Fall 2020	PhD student Zahra Tayebi

Volunteer Experience

2017—2019	**Research Advisor**, *Lycée Stendhal*, Milan, Italy.
	Advising and guiding students on an after-school mathematics project which is part of the larger European program: Math en Jeans. There is an annual conference where the students go to present their work on the year's project.

Grants

2021	**Global Education Online Module Design**, *Hanyang University ERICA*, Amount: $30,195 USD (3 years).
	PI: Zhipeng Cai
	For the development of online modules in data science and business intelligence, which could be offered as a course or training program

Awards

2011	The Marie Curie Alain Bensoussan (ABCDE) Postdoctoral Fellowship of the European Research Consortium for Informatics and Mathematics (ERCIM) Award of 60,300 € and 5000 € for travel, in order to fully fund 2 years of postdoctoral study in the European ERCIM member countries	
2006	The Postgraduate Scholarship (PGS) of the Natural Sciences and Engineering Research Council (NSERC) of Canada Award of $63,000 CAD in order to fully fund 3 years of postgraduate (PhD) research	
2006	Simon Fraser University Computing Science Graduate Fellowship Award of $3000 CAD towards funding postgraduate research	
2005	Simon Fraser University Computing Science Graduate Fellowship Award of $3000 CAD towards funding postgraduate research	
2002	The Undergraduate Student Research Award (USRA) of the Natural Sciences and Engineering Research Council (NSERC) of Canada Award of $6400 CAD in order to fully fund a summer of undergraduate research	

Technical Skills

advanced	C/C++, Python, LATEX	*7+ years experience*
intermediate	Java, Unix tools, Shell programming	*4+ years experience*
basic	Matlab, R, Prolog, Haskell	*2+ years experience*

Languages

English	maternal language	
French	second language	*conversational and technical fluency*
Italian	intermediate	*comprehension and basic conversational skills*
Spanish	intermediate	*comprehension and basic conversational skills*
中文	basic	*basic words and phrases*

Publications

Journal Articles

1. 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*, 2022. To appear, preprint at `http://arxiv.org/abs/2211.08267`.

2. Sarwan Ali, Babatunde Bello, Zahra Tayebi, and Murray Patterson. Characterizing SARS-CoV-2 spike sequences based on geographical location. *Journal of Computational Biology*, 2022. To appear, preprint at `https://arxiv.org/abs/2110.00809`.

1 Park Place, Room 636 – Atlanta, Georgia 30303 (USA)
📞 *+1 475 330 0978* • ✉ *mpatterson30@gsu.edu*
🌐 *murraypatterson.github.io* *4/9*

3. Sarwan Ali, Bikram Sahoo, Muhammad Asad Khan, Alexander Zelikovsky, Imdad Ullah Khan*, and Murray Patterson*. Efficient approximate kernel based spike sequence classification. *Transactions on Computational Biology and Bioinformatics*, 2022. To appear, preprint at `https://arxiv.org/abs/2209.04952`.

4. Kaustubh Khandai, Cristian Navarro-Martinez, Brendan Smith, Rebecca Buonopane, S. Ashley Byun*, and Murray Patterson*. Determining significant correlation between pairs of extant characters in a small parsimony framework. *Journal of Computational Biology*, 29(10):1132–1154, 2022. doi:10.1089/cmb.2022.0141.

5. Sarwan Ali, Babatunde Bello, Prakash Chourasia, Ria Thazhe Punathil, Yijing Zhou, and Murray Patterson. PWM2Vec: An efficient embedding approach for viral host specification from coronavirus spike sequences. *MDPI Biology*, 11(3):418, 2022. doi:10.3390/biology11030418.

6. Sarwan Ali, Yijing Zhou, and Murray Patterson. Efficient analysis of COVID-19 clinical data using machine learning models. *Medical & Biological Engineering & Computing*, 60(7):1881–1896, 2021. doi:10.1007/s11517-022-02570-8.

7. Zahra Tayebi, Sarwan Ali, and Murray Patterson. Robust representation and efficient feature selection allows for effective clustering of SARS-CoV-2 variants. *Algorithms*, 14(12):348, 2021. doi:10.3390/a14120348.

8. Sarwan Ali, Simone Ciccolella, Lorenzo Lucarella, Gianluca Della Vedova, and Murray Patterson. Simpler and faster development of tumor phylogeny pipelines. *Journal of Computational Biology*, 28(11):1142–1155, 2021. doi:10.1089/cmb.2021.0271.

9. Andrew Melnyk, Fatemeh Mohebbi, Sergey Knyazev, Bikram Sahoo, Roya Hosseini, Pavel Skums, Alex Zelikovsky*, and Murray Patterson*. From alpha to zeta: Identifying variants and subtypes of SARS-CoV-2 via clustering. *Journal of Computational Biology*, 28(11):1113–1129, 2021. doi:10.1089/cmb.2021.0302.

10. 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*, 25(11):4068–4078, 2021. doi:10.1109/JBHI.2021.3081380.

11. Simone Ciccolella, Mauricio Soto, Murray D. Patterson, Gianluca Della Vedova, Iman Hajirasouliha, and Paola Bonizzoni. Gpps: An ILP-based approach for inferring cancer progression with mutation losses from single cell data. *BMC Bioinformatics*, 21(1):413, 2020. doi:10.1186/s12859-020-03736-7.

12. Simone Ciccolella, Camir Ricketts, Mauricio Soto Gomez, Murray Patterson, Dana Silverbush, Paola Bonizzoni, Iman Hajirasouliha, and Gianluca Della Vedova. Inferring cancer progression from single-cell sequencing while allowing mutation losses. *Bioinformatics*, 37(3):326–333, 2020. doi:10.1093/bioinformatics/btaa722.

13. Raffaella Rizzi, Stefano Beretta, Murray Patterson, Yuri Pirola, Marco Previtali, Gianluca Della Vedova, and Paola Bonizzoni. Overlap graphs and de Bruijn graphs: data structures for de novo genome assembly in the big data era. *Quantitative Biology*, 7(4):278–292, 2019. doi:10.1007/s40484-019-0181-x.

14. Stefano Beretta*, Murray Patterson*, Simone Zaccaria, Gianluca Della Vedova, and Paola Bonizzoni. HapCHAT: Adaptive haplotype assembly for efficiently leveraging high coverage in long reads. *BMC Bioinformatics*, 19(1):252, 2018. doi:10.1186/s12859-018-2253-8.

15. Wandrille Duchemin, Yoann Anselmetti, Murray Patterson, Yann Ponty, Sèverine Bérard, Cedric Chauve, Celine Scornavacca, Vincent Daubin, and Eric Tannier. DeCoSTAR: Reconstructing the ancestral organization of genes or genomes using reconciled phylogenies. *Genome Biology and Evolution*, 9(5):1312–1319, 2017. doi:10.1093/gbe/evx069.

16. Andrea Bracciali, Marco Aldinucci, Murray Patterson, Tobias Marschall, Nadia Pisanti, Ivan Merelli, and Massimo Torquati. PWhatsHap: Efficient haplotyping for future generation sequencing. *BMC Bioinformatics*, 17(11):342, 2016. doi:10.1186/s12859-016-1170-y.

17. Murray Patterson*, Tobias Marschall*, Nadia Pisanti, Leo van Iersel, Leen Stougie, Gunnar W. Klau[†], and Alexander Schoenhuth[†]. WhatsHap: Weighted haplotype assembly for future-generation sequencing reads. *Journal of Computational Biology*, 22(6):498–509, 2015. doi:10.1089/cmb.2014.0157.

18. Murray Patterson, Gergely Szöllősi, Vincent Daubin, and Eric Tannier. Lateral gene transfer, rearrangement, reconciliation. *BMC Bioinformatics*, 14(15):S4, 2013. doi:10.1186/1471-2105-14-S15-S4.

19. Mohammed El-Kebir*, Tobias Marschall*, Inken Wohlers*, Murray Patterson, Jaap Heringa, Alexander Schoenhuth, and Gunnar W. Klau. Mapping proteins in the presence of paralogs using units of coevolution. *BMC Bioinformatics*, 14(15):S18, 2013. doi:10.1186/1471-2105-14-S15-S18.

20. Ján Maňuch*, Murray Patterson*, Roland Wittler*, Cedric Chauve, and Eric Tannier. Linearization of ancestral multichromosomal genomes. *BMC Bioinformatics*, 13(19):S11, 2012. doi:10.1186/1471-2105-13-S19-S11.

21. Ján Maňuch*, Murray Patterson*, and Cedric Chauve. Hardness results on the gapped consecutive-ones property problem. *Discrete Applied Mathematics*, 160(18):2760–2768, 2012. doi:10.1016/j.dam.2012.03.019.

22. Ján Maňuch* and Murray Patterson*. The complexity of the gapped consecutive-ones property problem for matrices of bounded maximum degree. *Journal of Computational Biology*, 18(9):1243–1253, 2011. doi:10.1089/cmb.2011.0128.

23. Roland Wittler, Ján Maňuch*, Murray Patterson*, and Jens Stoye. Consistency of sequence-based gene clusters. *Journal of Computational Biology*, 18(9):1023–1039, 2011. doi:10.1089/cmb.2011.0083.

Conference Proceedings

24. Prakash Chourasia, Sarwan Ali, and Murray Patterson. Informative initialization and kernel selection improves t-SNE for biological sequences. In *the 2022 IEEE International Conference on Big Data (IEEE BigData, Osaka, Japan, 2022)*. 2022. To appear, preprint at `https://arxiv.org/abs/2211.09263`.

1 Park Place, Room 636 – Atlanta, Georgia 30303 (USA)
📞 *+1 475 330 0978* • ✉ *mpatterson30@gsu.edu*
🌐 *murraypatterson.github.io*   *6/9*

25. Sarwan Ali, Taslim Murad, and Murray Patterson. PSSM2Vec: A compact alignment-free embedding approach for coronavirus spike sequence classification. In *the 29th International Conference on Neural Information Processing (ICONIP, Indore, India, 2022)*, Springer CCIS. 2022. To appear.

26. Taslim Murad*, Prakash Chourasia*, Sarwan Ali*, and Murray Patterson. Hashing2Vec: Fast embedding generation for SARS-CoV-2 spike sequence classification. In *the 14th Asian Conference on Machine Learning (ACML, Hyderabad, India, 2022)*, PMLR. 2022. To appear.

27. 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 *the 11th International Conference on Computational Advances in Bio and medical Sciences (ICCABS, Virtual Conference, 2021)*, volume 13254 of Springer LNBI, 133–148. 2022. doi:10.1007/978-3-031-17531-2_11.

28. Sarwan Ali, Taslim Murad, Prakash Chourasia, and Murray Patterson. Spike2Signal: Classifying coronavirus spike sequences with deep learning. In *the 8th IEEE International Conference on Big Data Service and Applications (IEEE BDS, San Francisco Bay Area, USA, 2022)*, 81–88. 2022. doi:10.1109/BigDataService55688.2022.00020.

29. Sarwan Ali, Bikram Sahoo, Naimat Ullah, Alex Zelikovsky, Murray Patterson*, and Imdad Ullah Khan*. A k-mer based approach for SARS-CoV-2 variant identification. In *the 17th International Symposium on Bioinformatics Research and Applications (ISBRA, Shenzhen, China, 2021)*, volume 13064 of LNCS, 153–164. 2021. doi:10.1007/978-3-030-91415-8_14.

30. Brendan Smith, Cristian Navarro-Martinez, Rebecca Buonopane, S. Ashley Byun*, and Murray Patterson*. Correlated evolution in the small parsimony framework. In *the 17th International Symposium on Bioinformatics Research and Applications (ISBRA, Shenzhen, China, 2021)*, volume 13064 of LNCS, 608–619. 2021. doi:10.1007/978-3-030-91415-8_51.

31. Sarwan Ali and Murray Patterson. Spike2Vec: An efficient and scalable embedding approach for COVID-19 spike sequences. In *the 2021 IEEE International Conference on Big Data (IEEE BigData, Virtual Conference, 2021)*, 1533–1530. 2021. doi:10.1109/BigData52589.2021.9671848.

32. Sarwan Ali, Tamkanat-E-Ali, Muhammad Asad Khan, Imdad Ullah Khan, and Murray Patterson. Effective and scalable clustering of SARS-CoV-2 sequences. In *the 5th International Conference on Big Data Research (ICBDR, Tokyo, Japan, 2021)*, 42–49. 2021. doi:10.1145/3505745.3505752.

33. Andrew Melnyk, Fatemeh Mohebbi, Sergey Knyazev, Bikram Sahoo, Roya Hosseini, Pavel Skums, Alex Zelikovsky*, and Murray Patterson*. Clustering based identification of SARS-CoV-2 subtypes. In *the 10th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS, Virtual Conference, 2020)*, volume 12686 of LNCS, 127–141. 2021. doi:10.1007/978-3-030-79290-9_11.

34. Simone Ciccolella*, Murray Patterson*, Paola Bonizzoni, and Gianluca Della Vedova. Effective clustering for single cell sequencing cancer data. In *the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB, Niagara Falls, NY, USA, 2019)*, ACM, 437–446. 2019. doi:10.1145/3307339.3342149.

35. Giulia Bernardini*, Paola Bonizzoni*, Gianluca Della Vedova*, and Murray Patterson*. A rearrangement distance for fully-labelled trees. In *the 30th Annual Symposium on Combinatorial Pattern Matching (CPM, Pisa, Italy, 2019)*, volume 128 of LIPIcs, 1–15. 2019. doi:10.4230/LIPIcs.CPM.2019.28.

36. Gianluca Della Vedova*, Murray Patterson*, Raffaella Rizzi*, and Mauricio Soto*. Character-based phylogeny construction and its application to tumor evolution. In *the 13th Conference on Computability in Europe (CiE, Turku, Finland, 2017)*, volume 10307 of LNCS, 3–13. 2017. doi:10.1007/978-3-319-58741-7_1.

37. Marco Aldinucci, Andrea Bracciali, Tobias Marschall, Murray Patterson, Nadia Pisanti, and Massimo Torquati. High-performance haplotype assembly. In *the 11th International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB, Cambridge, UK, 2014)*, volume 8623 of LNCS, 245–258. 2015. doi:10.1007/978-3-319-24462-4_21.

38. Murray Patterson*, Tobias Marschall*, Nadia Pisanti, Leo van Iersel, Leen Stougie, Gunnar W. Klau[†], and Alexander Schoenhuth[†]. WhatsHap: Haplotype assembly for future-generation sequencing reads. In *the 18th Annual International Conference on Research in Computational Molecular Biology (RECOMB, Pittsburgh PA, USA, 2014)*, volume 8394 of LNCS, 237–249. 2014. doi:10.1007/978-3-319-05269-4_19.

39. Cedric Chauve*, Murray Patterson*, and Ashok Rajaraman*. Hypergraph covering problems motivated by genome assembly questions. In *the International Workshop on Combinatorial Algorithms (IWOCA, Rouen, France, 2013)*, volume 8288 of LNCS, 428–432. 2013. doi:10.1007/978-3-642-45278-9_37.

40. Cedric Chauve*, Ján Maňuch*, Murray Patterson*, and Roland Wittler*. Tractability results for the consecutive-ones property with multiplicity. In *the 22nd Annual Symposium on Combinatorial Pattern Matching (CPM, Palermo, Italy, 2011)*, volume 6661 of LNCS, 90–103. 2011. doi:10.1007/978-3-642-21458-5_10.

41. Ján Maňuch*, Murray Patterson*, and Arvind Gupta. Towards a characterization of the generalized cladistic character compatibility problem for non-branching character trees. In *the 7th International Symposium on Bioinformatics Research and Applications (ISBRA, Changsha, China, 2011)*, volume 6674 of LNCS, 440–451. 2011.

42. Ján Maňuch*, Murray Patterson*, Sheung-Hung Poon*, and Chris Thachuk*. Complexity of finding non-planar rectilinear drawings of graphs. In *the 18th International Symposium on Graph Drawing (GD, Konstanz, Germany, 2010)*, volume 6502 of LNCS, 305–316. 2010. doi:10.1007/978-3-642-18469-7_28.

43. Cedric Chauve*, Ján Maňuch*, and Murray Patterson*. On the gapped consecutive-ones property. In *the European Conference on Combinatorics, Graph Theory and Applications (EuroComb, Bordeaux, France, 2009)*, volume 34 of ENDM, 121–125. 2009. doi:10.1016/j.endm.2009.07.020.

44. Ján Maňuch*, Murray Patterson*, and Arvind Gupta. On the generalised character compatibility problem for non-branching character trees. In *the 15th Annual International Computing and Combinatorics Conference (COCOON, Niagara Falls, USA, 2009)*, volume 5609 of LNCS, 268–276. 2009. doi:10.1007/978-3-642-02882-3_27.

45. Murray Patterson, Yongmei Liu, Eugenia Ternovska, and Arvind Gupta. Grounding for model expansion in k-guarded formulas with inductive definitions. In *the 20th International Joint Conferences on Artificial Intelligence (IJCAI, Hyderabad, India, 2007)*, 161–166. 2007.

note In any publication listing containing a *, this indicates that all authors whose names are followed by a * have contributed equally to this publication. In order to be consistent, this means that for papers that list the names of authors alphabetically (which is a common practice amongst computer science and theoretical bioinformatics papers), there is a * following every name.