

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

25 Park Place, Room 1807 – Atlanta, Georgia 30303 (USA)

☎ +1 475 330 0978 • ✉ mpatterson30@gsu.edu

🌐 murraypatterson.github.io

1/15

- 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

- 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 2024 **Georgia State University.**
 ○ CSC 4330 / 6330 — a senior year / masters-level programming language concepts course
- Fall 2023 **Georgia State University.**
 ○ CSC 4330 / 6330 — a senior year / masters-level programming language concepts course
- Spring 2023 **Georgia State University.**
 ○ CSC / DSCI 4810 / 6810 — a senior year / master-level course on artificial intelligence
- Fall 2022 **Georgia State University.**
 ○ CSC 4330 / 6330 — a senior year / masters-level programming language concepts course
- 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

25 Park Place, Room 1807 – Atlanta, Georgia 30303 (USA)

☎ +1 475 330 0978 • ✉ mpatterson30@gsu.edu

🌐 murraypatterson.github.io

- 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

- Fall 2023 I supervised PhD students Zahra Tayebi, Sarwan Ali, Prakash Chourasia, Taslim Murad, Avais Jain and Ayotomiwa Ezekiel Adeniyi, and Master's students Raghuveer Gummadi and Eunice Olorunshola.
- Prakash Chourasia won an MBD fellowship, funding the remainder of his study.
 - Zahra Tayebi successfully completed her dissertation proposal titled: "Efficient clustering and classification of protein sequences".
 - Sarwan Ali successfully completed his dissertation proposal titled: "Designing representation learning methods for molecular sequences analysis".
 - Eunice Olorunshola successfully defended her Master's thesis titled: "Classifying different cancer types based on transcriptomics data using machine learning algorithms".
- Spring 2023 I supervised PhD students Zahra Tayebi, Sarwan Ali, Prakash Chourasia, Taslim Murad and Avais Jain, and Master's students Raghuveer Gummadi, Eunice Olorunshola, Sai Venkata Sumanth Gaddam, Kushal Khatri, Surya Subrahmanya Sai Chandra Sekhar Kolli, Samvidha Reddy Mannem, Fatma Noor, Ashwin Pilgaonkar, Ria C. Thazhe Punathil, Sainikhila Vemulapalli and Karishma Mahendra Wadhe.
- Taslim Murad successfully completed her dissertation proposal titled: "Designing methods for representation learning of biological sequences and its application in analysis tasks".
- The following Master's projects were graduated:
- Sai Gaddam: "Diabetic retinopathy detection using deep learning".
 - Kushal Khatri: "A stratified subsampler for hierarchically structured data".
 - Surya Kolli: "Recommendations using graph databases".
 - Fatma Noor: "Can community detection build a good phylogeny?".

25 Park Place, Room 1807 – Atlanta, Georgia 30303 (USA)

☎ +1 475 330 0978 • ✉ mpatterson30@gsu.edu

🌐 murraypatterson.github.io

3/15

- Ashwin Pilgaonkar: “Robinson Country Intelligence Index (RCII) – A country risk analysis tool”.
 - Ria Punathil: “CoVScan – A tool for viral host specification from coronavirus spike sequences”.
 - Sainikhila Vemulapalli: “Vehicle image classification and damage analysis using deep learning model”.
 - Karishma Wadhe: “Drive safe – A driver drowsiness alert system”.
- Fall 2022 I supervised PhD students Zahra Tayebi, Sarwan Ali, Prakash Chourasia and Taslim Murad, and Master’s students Supriya Arun, Kaustubh Khandai, Kushal Khatri, Prashanth Annam, Babitha Reddy Devireddy, Indu Munnangi, Venkata Krishna Chaitanya Nuthalapati, Dhruval Patel, Shraddha Sharma and Atharv Utkarsh Yeolekar.
- Zahra Tayebi won a B&B fellowship, funding the remainder of her study.
- The following Master’s projects were graduated:
- Supriya Arun: “Comparative analysis of Covid classifiers”.
 - Kaustubh Khandai: “Determining significant correlation between pairs of extant characters in a small parsimony framework”.
 - Prashanth Annam: “Neobook: A social medial application using Flutter and Firebase”.
 - Babitha Reddy Devireddy: “Chaos engineering using Gremlin and GC analyzer”.
 - Indu Munnangi: “My recipe book: A MERN stack web application”.
 - Venkata Krishna Chaitanya Nuthalapati: “Tsuba: Web security scanner”.
 - Dhruval Patel: “Best actor detection”.
 - Shraddha Sharma: “Sunsetting customer identification service”.
 - Atharv Utkarsh Yeolekar: “Data condensation using generative adversarial networks”.
- Spring 2022 I supervised PhD students Zahra Tayebi, Sarwan Ali, Prakash Chourasia and Taslim Murad, and Master’s students Babatunde Bello, Supriya Arun and Kaustubh Khandai.
- Babatunde Bello graduated with Master’s thesis titled: “Functional enrichment analysis of transcriptomics data of breast cancer RNA-seq”.
- 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 I supervised PhD students Zahra Tayebi and Sarwan Ali
- Fall 2020 I supervised 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: \$24,255 USD (3 years).

PI: Zhipeng Cai, Co-I: Armin Mikler, Rajshekhar Sunderraman, Pavel Skums, Murray Patterson, Rolando Estrada, Yingshu Li, Jennifer Sherer, Geoffrey Graybeal

○ For the development of online modules in data science and business intelligence, which could be offered as a course or training program.

2024 **REU site: Multidisciplinary graph data analytics**, *National Science Foundation*, Amount: \$372,720 USD (3 years).

PI: Esra Akbas, Mehmet Emin Aktas

○ To establish Georgia State University as a Research Experiences for Undergraduates (REU) site by hiring paid undergraduate summer interns to work together with faculty and graduate students on their research projects.

Contracts

2021 **AI Learning Alliance: Broadening participation**, *Meta, Inc.*, Amount: \$50,000 USD (2 years).

PI: Anu Bourgeois, Co-I: Murray Patterson

○ To support work related to providing education at the frontier of R&D and to increase the representation of historically marginalized researchers and engineers by preparing them for jobs in the artificial intelligence field.

Under Review

2023 **Unveiling large-scale trends in diet and disease from gut microbiome sequencing data**, *National Institutes of Health*, Amount: \$429,000 USD (2 years).

PI: Murray Patterson, Co-I: Alexander Kirpich

○ The objective of this project is to develop a universal analysis framework for large-scale studies on gut microbiome shotgun metagenomics sequencing data. The application of such a framework to the rapidly growing collections of such data will unveil links to diseases and disorders that previous smaller datasets could not reveal.

2023 **Tracing seeding metastasis histories from CRISPR-Cas9 sc-seq**, *National Cancer Institute*, Amount: \$383,844 USD (2 years).

PI: Murray Patterson, Co-I: Pavel Skums, External: Sagi Snir (Univ. Haifa, Israel)

25 Park Place, Room 1807 – Atlanta, Georgia 30303 (USA)

☎ +1 475 330 0978 • ✉ mpatterson30@gsu.edu

🌐 murraypatterson.github.io

- This project aims to develop a computationally efficient methodology, and resulting toolkit, for phylogenetic and phylodynamic inference from CRISPR-Cas9 single-cell lineage tracing data, with a special emphasis on scalability and accuracy. Such a tool will then need to be extensively and comprehensively benchmarked using the available lineage tracing data.
- 2023 **Data-driven support system for computer-aided antibody design**, *National Institutes of Health*, Amount: \$348,948 USD (2 years).
PI: Murray Patterson, External: Muhammad Shoaib (PI), Imdad Ullah Khan, Khurram Bashir (LUMS, Pakistan)
- Disease diagnosis, therapeutics, and biological research rely heavily on antibodies, however, antibody development is a time-consuming, expensive, and labor-intensive process requiring specialized infrastructure, trained personnel, and animal immunizations. The proposed research aims to develop an end-to-end pipeline with a user-friendly graphical interface that utilizes cutting-edge machine learning algorithms and data analysis techniques for the computer-aided design of antibodies, which will significantly reduce costs and time while ensuring high binding affinity and specificity. This advancement has the potential to accelerate the development of antibodies, leading to improved diagnostics, efficient therapies, enhanced understanding of biological systems, and ultimately better health outcomes for individuals and society as a whole.
- 2023 **Reconciling geographic resolution and statistical robustness in cancer mapping using a novel partitioning algorithm**, *National Cancer Institute*, Amount: \$429,000 USD (2 years).
PI: Murray Patterson, Co-I: Chetan Tiwari, Armin Mikler
- Disease mapping methods developed in this project will provide more accurate and reliable representations of cancer incidence, prevalence, and mortality across communities. This improvement is crucial for identifying high-risk areas and deploying targeted prevention and treatment strategies more effectively. Ultimately, our work aims to contribute significantly to cancer control efforts, potentially leading to earlier detection, better resource allocation, and improved health outcomes in populations.
- 2023 **Benchmarking tumor phylogeny reconstruction from scDNA-seq data**, *Chan Zuckerberg Initiative*, Amount: \$199,777 USD (18 months).
PI: Murray Patterson, Co-I: Eric Gaucher, External: Simone Ciccolella (PI), Gianluca Della Vedova
- To develop benchmarks and standards for existing and future tools which reconstruct tumor phylogenies from sc-DNA-seq data.

Awards and Honors

- 2023 Nomination for membership to Sigma Xi, the scientific research honor society
- 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

25 Park Place, Room 1807 – Atlanta, Georgia 30303 (USA)

☎ +1 475 330 0978 • ✉ mpatterson30@gsu.edu

🌐 murraypatterson.github.io

6/15

- 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	<i>7+ years experience</i>
intermediate	Java, Unix tools, Shell programming	<i>4+ years experience</i>
basic	Matlab, R, Prolog, Haskell	<i>2+ years experience</i>

Languages

English	maternal language	
French	second language	<i>conversational and technical fluency</i>
Italian	intermediate	<i>comprehension and basic conversational skills</i>
Spanish	intermediate	<i>comprehension and basic conversational skills</i>
中文	basic	<i>basic words and phrases</i>

Publications

Journal Articles

1. Sarwan Ali, Tamkanat E. Ali, Taslim Murad, Haris Mansoor, and Murray Patterson. Molecular sequence classification using efficient kernel based embedding. *Information Sciences*, 2024. doi:10.1016/j.ins.2024.121100.
2. Sarwan Ali, Prakash Chourasia, and Murray Patterson. From PDB files to protein features: a comparative analysis of PDB bind and STCRDAB datasets. *Medical & Biological Engineering & Computing*, 2024. doi:0.1007/s11517-024-03074-3.
3. Zahra Tayebi, Sarwan Ali, Taslim Murad, Imdadullah Khan, and Murray Patterson. PseAAC2Vec protein encoding for TCR protein sequence classification. *Computers in Biology and Medicine*, 170:107956, 2024. doi:10.1016/j.combiomed.2024.107956.
4. Sarwan Ali, Prakash Chourasia, and Murray Patterson. When protein structure embedding meets large language models. *MDPI Genes*, 15(1):25, 2023. doi:10.3390/genes15010025.
5. Sarwan Ali and Murray Patterson. Improving IOSOMAP efficiency with RKS: a comparative study with t-distributed neighbor embedding on protein sequences. *MDPI J — Multidisciplinary Scientific Journal*, 6(4):579–591, 2023. doi:10.3390/j6040038.
6. Taslim Murad*, Sarwan Ali*, Imdad Ullah Khan, and Murray Patterson. Spike2CGR: an efficient method for spike sequence classification using chaos game representation. *Machine Learning*, 112:3633–3658, 2023. doi:10.1007/s10994-023-06371-4.
7. Sarwan Ali, Prakash Chourasia, Zahra Tayebi, Babatunde Bello, and Murray Patterson. ViralVectors: compact and scalable alignment-free virome feature generation. *Medical & Biological Engineering & Computing*, 2023. doi:10.1007/s11517-023-02837-8.
8. Taslim Murad, Sarwan Ali, and Murray Patterson. Exploring the potential of GANs in biological sequence analysis. *MDPI Biology*, 12(6):854, 2023. doi:10.3390/biology12060854.
9. Bikram Sahoo*, Sarwan Ali, Pin-Yu Chen, Murray Patterson, and Alexander Zelikovsky*. Assessing the resilience of machine learning classification algorithms on SARS-CoV-2 genome sequences generated with long-read specific errors. *Biomolecules*, 13(6):934, 2023. doi:10.3390/biom13060934.
10. Danushka Bandara, Karen Exantus, Cristian Navarro-Martinez, Murray Patterson, and Ashley Byun. Identifying distinguishing acoustic features in felid vocalizations based on call type and species classification. *Acoustics Australia*, 2023. doi:10.1007/s40857-023-00298-5.
11. Sarwan Ali, Babatunde Bello, and Murray Patterson. Solvent accessibility of coronaviridae spike proteins through the lens of information gain. *MDPI J — Multidisciplinary Scientific Journal*, 6(2):236–247, 2023. doi:10.3390/j6020018.

12. Sarwan Ali, Bikram Sahoo, Alexander Zelikovsky, Pin-Yu Chen, and Murray Patterson. Benchmarking machine learning robustness in covid-19 genome sequence classification. *Scientific Reports*, 13(1):4154, 2023. doi:10.1038/s41598-023-31368-3.
13. 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. doi:10.1089/cmb.2022.0424.
14. Sarwan Ali, Babatunde Bello, Zahra Tayebi, and Murray Patterson. Characterizing SARS-CoV-2 spike sequences based on geographical location. *Journal of Computational Biology*, 2022. doi:10.1089/cmb.2022.0391.
15. 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. doi:10.1109/TCBB.2022.3206284.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.

25 Park Place, Room 1807 – Atlanta, Georgia 30303 (USA)

☎ +1 475 330 0978 • ✉ mpatterson30@gsu.edu

🌐 murraypatterson.github.io

24. 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.
25. 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.
26. 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.
27. Wandrille Duchemin, Yoann Anselmetti, Murray Patterson, Yann Ponty, S  verine B  rard, 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.
28. 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.
29. 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.
30. Murray Patterson, Gergely Sz  ll  si, Vincent Daubin, and Eric Tannier. Lateral gene transfer, rearrangement, reconciliation. *BMC Bioinformatics*, 14(15):S4, 2013. doi:10.1186/1471-2105-14-S15-S4.
31. 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.
32. J  n Ma  uch*, 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.
33. J  n Ma  uch*, 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.
34. J  n Ma  uch* 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.

35. 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

36. Taslim Murad, Sarwan Ali, Prakash Chourasia, Haris Mansoor, and Murray Patterson. Circular arc length-based kernel matrix for protein sequence classification. In *the 2023 IEEE International Conference on Big Data (IEEE BigData, Sorrento, Italy, 2023)*. 2023. To appear.
37. Sarwan Ali, Prakash Chourasia, and Murray Patterson. Expanding chemical representation with k -mers and fragment-based fingerprints for molecular fingerprinting. In *the 10th International Conference on Information Management and Big Data (SIMBig, Mexico City, Mexico, 2023)*. 2023. To appear.
38. Prakash Chourasia, Taslim Murad, Zahra Tayebi, Sarwan Ali, Imdad Ullah Khan, and Murray Patterson. Efficient classification of SARS-CoV-2 spike sequences using federated learning. In *the 10th International Conference on Information Management and Big Data (SIMBig, Mexico City, Mexico, 2023)*. 2023. To appear.
39. Zahra Tayebi, Akshay Juyal, Alex Zelikovsky, and Murray Patterson. Simulating tumor evolution from scDNA-seq as an accumulation of both SNVs and CNAs. In *the 19th International Symposium on Bioinformatics Research and Applications (ISBRA, Wrocław, Poland, 2023)*, volume 14248 of LNCS, 530–540. 2023. doi:10.1007/978-981-99-7074-2_43.
40. Sarwan Ali*, Haris Mansoor*, Prakash Chourasia*, and Murray Patterson. Hist2Vec: Kernel-based embeddings for biological sequence classification. In *the 19th International Symposium on Bioinformatics Research and Applications (ISBRA, Wrocław, Poland, 2023)*, volume 14248 of LNCS, 387–397. 2023. doi:10.1007/978-981-99-7074-2_30.
41. Sarwan Ali*, Prakash Chourasia*, and Murray Patterson. PDB2Vec: Using 3D structural information for improved protein analysis. In *the 19th International Symposium on Bioinformatics Research and Applications (ISBRA, Wrocław, Poland, 2023)*, volume 14248 of LNCS, 376–386. 2023. doi:10.1007/978-981-99-7074-2_29.
42. Prakash Chourasia*, Taslim Murad*, Sarwan Ali*, and Murray Patterson. Enhancing t-SNE performance for biological sequencing data through kernel selection. In *the 19th International Symposium on Bioinformatics Research and Applications (ISBRA, Wrocław, Poland, 2023)*, volume 14248 of LNCS, 442–452. 2023. doi:10.1007/978-981-99-7074-2_35.
43. Usama Sardar*, Sarwan Ali*, Muhammad Sohaib Ayub, Muhammad Shoaib, Khurram Bashir, Imdad Ullah Khan, and Murray Patterson. Sequence-based nanobody-antigen binding prediction. In *the 19th International Symposium on Bioinformatics Research and Applications (ISBRA, Wrocław, Poland, 2023)*, volume 14248 of LNCS, 227–240. 2023. doi:10.1007/978-981-99-7074-2_18.

44. Sarwan Ali, Pin-Yu Chen, and Murray Patterson. Unveiling the robustness of machine learning models in classifying COVID-19 spike sequences. In *the 19th International Symposium on Bioinformatics Research and Applications (ISBRA, Wrocław, Poland, 2023)*, volume 14248 of LNCS, 1–15. 2023. doi:10.1007/978-981-99-7074-2_1.
45. Sarwan Ali, Usama Sardar, Imdad Ullah Khan, and Murray Patterson. Efficient sequence embedding for SARS-CoV-2 variants classification. In *the 19th International Symposium on Bioinformatics Research and Applications (ISBRA, Wrocław, Poland, 2023)*, volume 14248 of LNCS, 16–30. 2023. doi:10.1007/978-981-99-7074-2_2.
46. Mansoor Ahmed, Usama Sardar, Sarwan Ali, Shafiq Alam, Murray Patterson, and Imdad Ullah Khan. Robust brain age estimation via regression models and MRI-derived features. In *Advances in Computational Collective Intelligence (ICCCI, Budapest, Hungary, 2023)*, volume 1864 of Communications in Computer and Information Science. 2023. doi:10.1007/978-3-031-41774-0_52.
47. Zahra Tayebi*, Sarwan Ali*, Prakash Chourasia, Taslim Murad, and Murray Patterson. T cell receptor protein sequences and sparse coding: A novel approach to cancer classification. In *the 2023 International Conference on Neural Information Processing (ICONIP, Changsha, China, 2023)*, Communications in Computer and Information Science. 2023. doi:10.1007/978-981-99-8141-0_17.
48. Sarwan Ali, Babatunde Bello, Prakash Chourasia, Ria Thazhe Punathil, Pin-Yu Chen, Imdad Ullah Khan, and Murray Patterson. Virus2Vec: Viral sequence classification using machine learning. In *the 2023 Conference on Health, Inference, and Learning (CHIL, Cambridge, MA, USA, 2023)*, volume 209 of Proceedings of Machine Learning Research, 6–18. 2023. URL: <https://proceedings.mlr.press/v209/ali23a.html>.
49. Sarwan Ali, Prakash Chourasia, and Murray Patterson. When biology has chemistry: Solubility and drug subcategory prediction using SMILES strings. In *the 11th International Conference on Learning Representations (ICLR, Kigali, Rwanda, 2023)*, Tiny Papers Track. 2023. URL: <https://openreview.net/forum?id=28si4RXwDt1>.
50. Taslim Murad, Sarwan Ali, and Murray Patterson. A new direction in membranolytic anticancer peptides classification: Combining spaced k -mers with chaos game representation. In *the International Neural Network Society Workshop on Deep Learning Innovations and Applications (INNS DLIA, Queensland, Australia, 2023)*, volume 222 of Procedia Computer Science, 666–675. 2023. doi:10.1016/j.procs.2023.08.204.
51. Prakash Chourasia*, Zahra Tayebi*, Sarwan Ali*, and Murray Patterson. Empowering pandemic response with federated learning for protein sequence data analysis. In *the 2023 International Joint Conference on Neural Networks (IJCNN, Queensland, Australia, 2023)*, 1–8. 2023. doi:10.1109/IJCNN54540.2023.10191721.
52. Sarwan Ali*, Taslim Murad*, and Murray Patterson. PCD2Vec: A Poisson correction distance based approach for viral host classification. In *the 2023 International Joint Conference on Neural Networks (IJCNN, Queensland, Australia, 2023)*, 1–8. 2023. doi:10.1109/IJCNN54540.2023.10191311.

53. Sarwan Ali, Usama Sardar, Murray Patterson, and Imdad Ullah Khan. BioSequence2Vec: Efficient embedding generation for biological sequences. In *the 27th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD, Osaka, Japan, 2023)*, volume 13936 of LNCS, 173–185. 2023. doi:10.1007/978-3-031-33377-4_14.
54. 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)*, 101–106. IEEE Computer Society, 2022. doi:10.1109/BigData55660.2022.10020217.
55. 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)*, volume 1794 of Springer CCIS, 420–432. 2023. doi:10.1007/978-981-99-1648-1_35.
56. 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)*, volume 189 of Proceedings of Machine Learning Research, 754–769. 2022. URL: <https://proceedings.mlr.press/v189/taslim23a.html>.
57. 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.
58. 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.
59. 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.
60. 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.
61. 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.

62. 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.
63. Andrew Melnyk, Fatemeh Mohebbi, Sergey Knyazev, Bikram Sahoo, Roya Hosseini, Pavel Skums, Alex Zelikovskiy*, 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.
64. 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.
65. 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.
66. 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.
67. 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.
68. 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.
69. 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.
70. 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.

71. 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.
72. 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.
73. 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.
74. 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.
75. 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.