Mahdi Belcaid

Contact Information

Information and Computer Sciences
1680 East-West Road, Room 317
Honolulu, HI 96822

Phone: 808 956-6049
E-mail: mahdi@hawaii.edu

Education

Ph.D. Information and Computer Sciences, University of Hawai'i at Mānoa 2013
Computational Methods and Algorithms for Phage Metagenomic Data Analysis

M.S. Computer Science, Université du Québec à Montréal 2006
A Novel EST Bioinformatic Resource for Studying Cold Acclimation in Wheat

Graduate Diploma in Bioinformatics, Université du Québec à Montréal 2004

B.S. Computer Science, Université du Québec à Montréal 2001

Professional Experience

Assistant Professor **2019 to Present**
Information and Computer Sciences
Hawai'i Institute of Marine Biology and
University of Hawai'i at Mānoa

Executive Director **2018 to 2019**
Hawai'i Data Science Institute
University of Hawai'i at Mānoa

Assistant Researcher **2013 to 2018**
Hawai'i Institute of Marine Biology
University of Hawai'i at Mānoa

Bioinformatics Core Director **2013 to 2018**
Pacific Center for Emerging Infectious Diseases Research (COBRE)
University of Hawai'i at Mānoa

Biomedical Informatics Core Director **2015 to 2017**
Bioscience Research Infrastructure Development for Grant
Enhancement and Success (RCMI)
University of Hawai'i at Mānoa

Bioinformatics Data Scientist **2006 to 2013**
Pacific Center for Emerging Infectious Diseases Research (COBRE)
IDeA Networks of Biomedical Research Excellence (INBRE)
University of Hawai'i at Mānoa

Research Assistant **20003 to 2005**
Laboratory of Functional Genomics of Abiotic Stress in Plants
Université du Québec à Montréal

Refereed Publications

[1] Yu, A., Cleveland, S., Stubbs, S., Belcaid, M., 2023. An LLM-based Semantic-Aware Framework for Data Wrangling, Validation and FAIR compliance. ACM Transactions on Knowledge Discovery from Data. (In preparation)

[2] Harrigan, W., Ferrell, B. D., Wommack, K. E., Polson, S. W., Schreiber, Z. D., and **Belcaid, M.**, 2023. Improvements in Viral Gene Annotation Using Large Language Models and Soft Alignments. BMC Bioinformatics (In review)

[3] **Belcaid, M.**, Leigh, J., Theriot, R., Kirshenbaum, N., Tabalba, R., Rogers, M., Johnson, A., Brown, M., Renambot, L., Long, L. and Nishimoto, A., 2023. Reflecting on the Scalable Adaptive Graphics Environment Team's 20-Year Translational Research Endeavor in Digital Collaboration Tools. Computing in Science & Engineering, 25(2), pp.50-56.

[4] **Belcaid, M.**, Martinez, A.G. and Leigh, J., 2022. Leveraging deep contrastive learning for semantic interaction. PeerJ Computer Science, 8, p.e925.

[5] Amend, A.S., Swift, S.O., Darcy, J.L., **Belcaid, M.**, Nelson, C.E., Buchanan, J., Cetraro, N., Fraiola, K.M., Frank, K., Kajihara, K. and McDermot, T.G., 2022. A ridge-to-reef ecosystem microbial census reveals environmental reservoirs for animal and plant microbiomes. Proceedings of the National Academy of Sciences, 119(33), p.e2204146119.

[6] **Belcaid, M.**, Arisdakessian, C. and Kravchenko, Y., 2022. Taming DNA clustering in massive datasets with SLYMFAST. ACM SIGAPP Applied Computing Review, 22(1), pp.15-23.

[7] Cleveland, S., Arisdakessian, C., Nelson, C., **Belcaid, M.**, Frank, K. and Jacobs, G., 2022. The C-MĀIKI gateway: a modern science platform for analyzing microbiome data. In Practice and Experience in Advanced Research Computing (pp. 1-7).

[8] Belcaid, M., Arisdakessian, C. and Kravchenko, Y., 2021, March. Efficient DNA sequence partitioning using probabilistic subsets and hypergraphs. In Proceedings of the 36th Annual ACM Symposium on Applied Computing (pp. 4-9).

[9] Kirshenbaum, N., Davidson, K., Harden, J., North, C., Kobayashi, D., Theriot, R., Tabalba Jr, R.S., Rogers, M.L., **Belcaid, M.**, Burks, A.T. and Bharadwaj, K.N., 2021. Traces of time through space: Advantages of creating complex canvases in collaborative meetings. Proceedings of the ACM on Human-Computer Interaction, 5(ISS), pp.1-20.

[10] Arisdakessian, C.G., Nigro, O.D., Steward, G.F., Poisson, G. and **Belcaid, M.**, 2021. CoCoNet: an efficient deep learning tool for viral metagenome binning. Bioinformatics, 37(18), pp.2803-2810.

[11] Bharadwaj, K., Burks, A., Johnson, A., Long, L., Renambot, L., Brown, M., Kobayashi, D., **Belcaid, M.**, Kirshenbaum, N., Tabalba, R. and Theriot, R., 2021, December. Securing Collaborative Work in Wide-band Display Environments. In 2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC) (pp. 26-34). IEEE.

[12] Jani, A.J., Bushell, J., Arisdakessian, C.G., **Belcaid, M.**, Boiano, D.M., Brown, C. and Knapp, R.A., 2021. The amphibian microbiome exhibits poor resilience following pathogen-induced disturbance. The ISME Journal, 15(6), pp.1628-1640.

[13] Casey, J.M., Ransome, E., Collins, A.G., Mahardini, A., Kurniasih, E.M., Sembiring, A., Schiettekatte, N.M., Cahyani, N.K.D., Wahyu Anggoro, A., Moore, M., Uehling, A., Belcaid, M., and others, 2021. DNA metabarcoding marker choice skews perception of marine eukaryotic biodiversity. Environmental DNA, 3(6), pp.1229-1246.

[14] Seale, A.P., Malintha, G.H.T., Celino-Brady, F.T., Head, T., **Belcaid, M.**, Yamaguchi, Y., Lerner, D.T., Baltzegar, D.A., Borski, R.J., Stoytcheva, Z.R. and Breves, J.P., 2020. Transcriptional regulation of prolactin in a euryhaline teleost: Characterisation of gene promoters through in silico and transcriptome analyses. Journal of neuroendocrinology, 32(11), p.e12905.

[15] Arisdakessian, C., Cleveland, S.B. and **Belcaid, M.**, 2020. MetaFlow— mics: scalable and reproducible nextflow pipelines for the analysis of microbiome marker data. In Practice and Experience in Advanced Research Computing (pp. 120-124).

[16] Arora, K., **Belcaid, M.**, Lantz, M.J., Taketa, R. and Nichols, R.A., 2020. Transcriptome profile of nicotinic receptor-linked sensitization of beta amyloid neurotoxicity. Scientific reports, 10(1), p.5696.

[17] Leigh, J., Kobayashi, D., Kirshenbaum, N., Wooton, T., Gonzalez, A., Renambot, L., Johnson, A., Brown, M., Burks, A., Bharadwaj, K. Nishimoto, A., ... and **Belcaid, M.**, 2019. Usage patterns of wideband display environments in e-science research, development and training. In 2019 15th International Conference on eScience (pp. 301-310).

[18] **Belcaid, M.**, Casaburi, G., McAnulty, S.J., Schmidbaur, H., Suria, A.M., Moriano-Gutierrez, S., Pankey, M.S., Oakley, T.H., Kremer, N., Koch, E.J. and Collins, A.J., 2019. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proceedings of the National Academy of Sciences, 116(8), pp.3030-3035.

[19] Moriano-Gutierrez, S., Koch, E.J., Bussan, H., Romano, K., **Belcaid, M.**, Rey, F.E., Ruby, E.G. and McFall-Ngai, M.J., 2019. Critical symbiont signals drive both local and systemic changes in diel and developmental host gene expression. Proceedings of the National Academy of Sciences, 116(16), pp.7990-7999.

[20] Harden, J., Christman, E., Kirshenbaum, N., Belcaid, M., Leigh, J. and North, C., 2023, April. "There is no reason anybody should be using 1D anymore": Design and Evaluation of 2D Jupyter Notebooks. In Graphics Interface 2023.

[21] **Belcaid, M.** and Poisson, G., 2018. Detecting anomalies in the Cytochrome C Oxidase I amplicon sequences using minimum scoring segments. ACM SIGAPP Applied Computing Review, 17(4), pp.6-14.

[22] **Belcaid, M.** and Poisson, G., 2017, April. A profile-based probabilistic approach for the detection of anomalies in the cytochrome C oxidase I amplicon sequences. In Proceedings of the Symposium on Applied Computing (pp. 11-17).

[23] SahBandar, I.N., Samonte, G., Telan, E., Siripong, N., **Belcaid, M.**, Schanzenbach, D., Leano, S., Chagan-Yasutan, H., Hattori, T., Shikuma, C.M. and Ndhlovu, L.C., 2017. Ultra-deep sequencing analysis on HIV drug-resistance-associated mutations among HIV-infected individuals: first report from the Philippines. AIDS Research and Human Retroviruses, 33(11), pp.1099-1106.

[24] Forsman, Z.H., Knapp, I.S.S., Tisthammer, K., Eaton, D.A.R., **Belcaid, M.** and Toonen, R.J., 2017. Coral hybridization or phenotypic variation? Genomic data reveal gene flow between Porites lobata and P. compressa. Molecular Phylogenetics and Evolution, 111, pp.132-148.

[25] Kumar, M., **Belcaid, M.** and Nerurkar, V.R., 2016. Identification of host genes leading to West Nile virus encephalitis in mice brain using RNA-seq analysis. Sci Rep 6: 26350.

[26] Bhattacharya, D., Agrawal, S., Aranda, M., Baumgarten, S., **Belcaid, M.**, Drake, J.L., Erwin, D., Foret, S., Gates, R.D., Gruber, D.F. and Kamel, B., 2016. Comparative genomics explains the evolutionary success of reef-forming corals. elife, 5, p.e13288.

[27] Elfekih, S., Chen, C.Y., Hsu, J.C., **Belcaid, M.** and Haymer, D., 2016. Identification and preliminary characterization of chemosensory perception-associated proteins in the melon fly Bactrocera cucurbitae using RNA-seq. Scientific reports, 6(1), p.19112.

[28] Jarvi, S.I., Bianchi, K.R., Farias, M.E., Txakeeyang, A., McFarland, T., Belcaid, M. and Asano, A., 2016. Characterization of class II β chain major histocompatibility complex genes in a family of Hawaiian honeycreepers:'amakihi (Hemignathus virens). Immunogenetics, 68, pp.461-475.

[29] **Belcaid, M.**, Kang, Y., Tuanyok, A. and Hoang, T.T., 2015. Complete genome sequence of Burkholderia cepacia strain LO6. Genome Announcements, 3(3), pp.10-1128.

[30] Beurmann, S., Videau, P., Ushijima, B., Smith, A.M., Aeby, G.S., Callahan, S.M. and **Belcaid, M.**, 2015. Complete Genome Sequence of Pseudoalteromonas sp. Strain OCN003, Isolated from Kāne'ohe Bay, O'ahu, Hawaii. Genome Announcements, 3(1), pp.e01396-14.

[31] **Belcaid, M.** and Toonen, R.J., 2015. Demystifying computer science for molecular ecologists. Molecular Ecology, 24(11), pp.2619-2640.

[32] Hoffmann, P.R., Panigada, M., Soprana, E., Terry, F., Bandar, I.S., Napolitano, A., Rose, A.H., Hoffmann, F.W., Ndhlovu, L.C., **Belcaid, M.** and Moise, L., (2015). Preclinical development of HIvax: human survivin Highly Immunogenic vaccines. *H*uman vaccines & immunotherapeutics, 11(7), pp.1585-1595.

[33] Edmunds, P.J., Pochon, X., Levitan, D.R., Yost, D.M., **Belcaid, M.**, Putnam, H.M. and Gates, R.D., 2014. Long-term changes in Symbiodinium communities in Orbicella annularis in St. John, US Virgin Islands. Marine Ecology Progress Series, 506, pp.129-144.

[34] Culley, A.I., Mueller, J.A., **Belcaid, M.**, Wood-Charlson, E.M., Poisson, G. and Steward, G.F., 2014. The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. MBio, 5(3), pp.10-1128.

[35] Jarvi, S.I., Farias, M.E., Lapointe, D.A., **Belcaid, M.** and Atkinson, C.T., 2013. Next-generation sequencing reveals cryptic mtDNA diversity of Plasmodium relictum in the Hawaiian Islands. Parasitology, 140(14), pp.1741-1750.

[36] Steward, G.F., Culley, A.I., Mueller, J.A., Wood-Charlson, E.M., **Belcaid, M.** and Poisson, G., 2013. Are we missing half of the viruses in the ocean?. The ISME journal, 7(3), pp.672-679.

[37] Owens, J.B., Urschitz, J., Stoytchev, I., Dang, N.C., Stoytcheva, Z., **Belcaid, M.**, Maragathavally, K.J., Coates, C.J., Segal, D.J. and Moisyadi, S., 2012. Chimeric piggyBac transposases for genomic targeting in human cells. Nucleic acids research, 40(14), pp.6978-6991.

[38] **Belcaid, M.**, Bergeron, A. and Poisson, G., 2011, December. The evolution of the tape measure protein: units, duplications and losses. In BMC bioinformatics (Vol. 12, No. 9, pp. 1-12). BioMed Central.

[39] **Belcaid, M.**, Bergeron, A. and Poisson, G., 2010. Mosaic graphs and comparative genomics in phage communities. Journal of Computational Biology, 17(9), pp.1315-1326.

[40] Bergeron, A., **Belcaid, M.**, Steward, G.F. and Poisson, G., 2007. Divide and conquer: enriching environmental sequencing data. PloS one, 2(9), p.e830.

[41] Houde, M., **Belcaid, M.**, Ouellet, F., Danyluk, J., Monroy, A.F., Dryanova, A., Gulick, P., Bergeron, A., Laroche, A., Links, M.G. and MacCarthy, L., 2006. Wheat EST resources for functional genomics of abiotic stress. BMC genomics, 7, pp.1-22.

Teaching

University of Hawai'i at Mānoa, Honolulu, HI

ICS 434/DATA 434: Data Science fundamentals	**S19, F20**
ICS 691D: Topics in Computer Science	**S19, F20**
ICS 438: Big Data Analytics	**F21, F22, F23**
ICS 691B: Topics in Computer Science	**S19, F20**
MBIO 740: Quantitative Methods in Biology	**S20**
ICS 675: Bioinformatics: Sequence Analysis	**S20**
MBIO 612 / OCN 682: Introduction to Data Science in R	**F21, F22, F23**
ICS 699 / MBIO 699: Directed Research	**F19, S20, F20, S21, S22, F22, S23, F23**
ICS 499: Directed Reading	(F20, S21)
Hawaii Data Science Institute Annual Workshops	**2006-2018**
Bioinformatics Core Annual Workshops	**2006-2018**

Université du Québec à Montréal, Montreal, Quebec

Introduction to Scientific Programming (INF 1105)	**F05**
Introduction to Algorithms in Bioinformatic (BIF 7000)	**S05**

Grants and Awards

Funded - Current:

CSSI Frameworks: SAGE3: Smart Amplified Group Environment for Harnessing the Data Revolution non-model. NSF-CSSI (Collaboration with the University of Illinois at Chicago).
Leigh (PI), Belcaid (co-PI), $2.5M.

- Directed evolution of a sequence-specific targeting technology for therapeutic gene delivery to the human genome. NIH-NIBIB.
  Owens (PI), Belcaid (co-PI), $3,179,046.

- RII Track-2 FEC: G2P in VOM: An Experimental and Analytical Framework for Genome to Phenome Connections in Viruses of Microbes. NSF.
  Becaid (PI), $321,376.

- RII Track 1: Change Hawaii; Harnessing the Data Revolution for Island Resilience. NSF. Jacobs (PI), Belcaid (Sub-project Co-PI), $20M Total Budget

Funding - Senior Personnel or Other:

- NSF INCLUDES Alliance: The Alliance of Students with Disabilities for Inclusion. NSF.
  STEM Faculty Mentor.

- Project Hōkūlani Hui. DOE
  Reasearch Internship Advisor

Funding - Expired:

Infections in the basement: viral interactions with microbiota inhabiting crustal basalt of the Mid-Atlantic Ridge flank 2016
NSF(Division Biological Oceanography) Steward (PI), Belcaid (coPI) $285,219.

Coral reef adaptation and acclimatization to global change 2015
NSF (Division Of Environmental Biology) Toonen (PI), Belcaid (coPI) $721,970.

Bioscience Research Infrastructure Development for Grant Enhancement and Success 2016
NIH (National Institute on Minority Health and Health Disparities)
Berry (PI), Belcaid (subproject PI), $69,330.

Computational Tools for the Analysis of COI Amplicon Data 2016
Smithsonian Institution, Belcaid (PI) $11,000.

Draft Sequencing of the *Sarcothelia edmondsoni* Genome 2014
National Center for Genome Resources, Belcaid (PI) $10,000.

Improved Cyberinfrastructure To Support Coral Reef Research And Education 2013
NSF (Div Of Biological Infrastructure) Leong (PI), Belcaid (coPI),
Jacobs (coPI), Franklin (coPI), Lemus (coPI), $220,547.

Synergistic Activities

2023–Present	Event Chair, NSA-Sponsored Aloha Datathon
2023	Member, Communication and Information Sciences DS Focus Area Committee
2023	Reviewer, National Science Foundation's ExpandAI Program
2023	Reviewer, Bioinformatics and AI Grants, Ola HAWAII
2020–Present	Member, Marine Biology Graduate Curriculum Committee
2020–Present	Member, ICS Qualified Exam Grading Committee
2020–2021	Founder and Principal Organizer, Python Professionalism Monthly Seminar Series
2022	Reviewer, Food for Thought's Thematic Research Funding Program, Canada First Resear
2021–2023	Member, Information and Computer Sciences Hiring Committee for Data Science, AI, Cyl
2020–Present	Coordinator, Undergraduate Certificate in Data Science Committee
2019–Present	Member, Marine Biology Graduate Curriculum Committee
2019–Present	Chair, Hawaii Institute of Marine Biology Information Technology Committee
2019–Present	Member, Information and Computer Sciences Curriculum and Undergraduate Committees
2019, 2023	Chair, HIMB IT Staff Position Search Committee
2019–2022	Member, HIMB Scholarship Committee
2019	Reviewer, COBRE-Hawaii Small Grants Program
2014–Present	Member, Association for Computing Machinery
2014–2023	Associate Editor, BioData Mining Journal
2013–2021	Reviewer, ACM Symposium on Applied Computing
2013–2016	Judge, Hawai'i State Science and Engineering Fair

Selected Invited Talks and Workshops

Birds of a Feather Seesion — November 15, 2022
12th Super Computing, Dallas TX
Smart Amplified Group Environment Enhanced with Artificial Intelligence for Global Collaboration.

Spring Bioinformatics Seminar Series — May 9, 2022
University of Delaware, Newark, DE
Viral Metagenomics and Deep Learning: The Perfect Pairing.

Towards a Sustainable Data and Software Cyberinfrastructure — July 25-26, 2022
2022 NSF Cyberinfrastructure For Sustained Scientific Innovation, Washington DC
Smart Amplified Group Environment Enhanced with Artificial Intelligence for Global Collaboration.

Monthly Luncheon Seminar Series. — April 16, 2019
Hawaii Information Communications Technology Association, Honolulu, HI
Fundamental Principles and Concepts in Business Data Science.

C-MAIKI Annual Symposiym — May 9, 2019
Center for MICROBIOME Analysis through Island Knowledge & Investigation, Honolulu HI
A Computational Pipeline for Microbiome Data Analysis.

Mathematical Methods and Models in Medicine — March 22-24, 2019
American Mathematical Society, Honolulu HI
A Probabilistic Approach for DNA Sequence Partitioning Using Dimensionality Reduction.

Diversity of the Indo-Pacific Network Workshop — July 25-26, 2016
University of the Philippines Diliman
Demystifying Computer Science for Molecular Ecologists.

Jiao Tong University School of Life Sciences and Biotechnology — December 2015
Shanghai Jiao Tong University
Working Group on Strategies for Metagenome Assembly.

Marine Sciences Seminar — February 2015
University of Hawai'i Hilo
State of the Art in Transcriptomics Data Analysis.

Marine Biology-Processes and Impacts — March 2014
University of Hawai'i at Mānoa
Introduction to bioinformatics. (Course + Lab Session)

The 29th Annual Edwin W. Pauley Summer Program — June 16-18, 2013
Hawai'i Institute of Marine Biology
Advancing tools for biodiversity studies.

Hawai'i Pacific University — April 2013
College of Natural Sciences Seminar, Kāneohe, HI
State of the Art of Bioinformatics Research and Applications.

First Symposium in translational Bioinformatics — June 2011
The John A. Burns School of Medicine, Honolulu, HI
Divide and Conquer: Enriching Viral Environmental Sequencing Data.

6th Annual Research Conference NCIBI & RCMI Workshop — 2011
University of Michigan Ann Arbor

Enhancing the Bioinformatics Curriculum.

Mentoring and Advising Activities

Master's or Ph.D. Advising

2021-present	Will Harrigan (Ph.D. - Marine Biology)
2021-present	Rajan Sawhney (Ph.D. - Information and Computer Sciences)
2021-present	Benjamin Strauss (Ph.D. - Information and Computer Sciences)
2021-present	Candace Edwards (M.Sc. - Information and Computer Sciences)
2021-present	Akib Sadmanee (M.Sc. - Information and Computer Sciences)
2021-present	Andy Yu (M.Sc. - Information and Computer Sciences)
2021-2023	Michael Rogers (M.Sc. - Information and Computer Sciences)
2019-2021	Nima Azbijari (M.Sc. - Information and Computer Sciences)
2019-2020	Sushil Shrestha (M.Sc. - Information and Computer Sciences)
2018-2020	Cedric Arisdakessian (Ph.D. - Information and Computer Sciences)

Master's or Ph.D. Committees

2018-Present	McLean Worsham (Ph.D. - Biology)
2018-Present	Lauren Arnold (Ph.D. - Marine Biology)
2018-Present	Blaine Billings (Ph.D. - Linguistics)
2018-Present	Rose Gallo (Ph.D. - Earth Sciences)
2018-Present	Jesse Harden (Ph.D. - Computer Science, Virginia Tech)
2018-Present	Rodrick Tabalba (Ph.D. Information and Computer Sciences)
2018-Present	Sean Swift (Ph.D. - Marine Biology)
2018-Present	McLean Worsham (Ph.D. - Biology)
2018-2020	Alberto Gonzales (Ph.D. - Information and Computer Sciences)
2015-2019	Molly Timmers (Ph.D. - Zoology)
2018-2023	Elizabeth Christman (M.Sc. Computer Science, Virginia Tech)
2018-2022	Yusuke Hatanaka (M.Sc. Information and Computer Sciences)
2018-2022	Rodrick Tabalba (M.Sc. Information and Computer Sciences)
2018-2020	Billy Troy Wooton (M.Sc. Information and Computer Sciences)
2018-2020	Nick Glazer (M.Sc. Information and Computer Sciences)
2015-2016	Monika Frazier (M.Sc. - Marine Sciences Department, UH Hilo)

Data Science Fellowship Advisor

2018	Sean Takafuji (Undergraduate student - Information & Computer Science)
2018	Charles Dickens (Undergraduate student - Electrical Engineering)
2018	Layne Fujioka (Undergraduate student - Mathematics/ Physics)
2018	Jaclyn Lee (Undergraduate student - Economics)
2018	Cedric Arisdakessian (Ph.D. student - Molecular Biosciences and Bioengineering)
2018	Yuliia Kravchenko (Ph.D. student - Mathematics)
2018	Trista McKenzie (Ph.D. student - Natural Geology & Geophysics)
2018	Charlotte Smith (Ph.D. student - Natural Resources and Environmental Management)