

CAMERON NUGENT PHD

EDUCATION & ACADEMIC ACHIEVEMENTS

University of Guelph
PhD, Department of Integrative Biology
Graduated April 2019

Queen's University
Bachelor of Science, Honours with Distinction
Biology Specialization
Graduated June 2014
Dean's Honour List 2012, 2013, 2014

WORK & RESEARCH EXPERIENCE

Postdoctoral Researcher, Fisheries and Oceans Canada, Halifax, NS October 2021 – present

- **Bioinformatics:** Created custom data analysis pipelines for processing low-depth whole genome re-sequencing data for over 900 samples. The pipeline allocated available resources to facilitate efficient, parallelized computation and led to the successful identification of over 5 million novel genetic markers for the fish species cunner (*Tautoglabrus adspersus*).
- **Machine Learning:** Used supervised and unsupervised machine learning algorithms to leverage molecular markers and environmental variables to characterize the structure of wild populations and make predictions about which populations are most susceptible to the effects of climate change.
- **Data science and software engineering:** Designed custom programs to clean, visualize, and statistically analyze large datasets in the programming languages Python, R, and C++.
- **Cloud-based big data analysis:** Made scientific discoveries through the management and analysis of terabytes of data on remote access computing clusters. Use of Microsoft Azure cloud computing environment and resources. Use of Slurm workload manager for job submission and management.
- **Software and data management:** Use of Git and GitHub for version control and software management to share code across team members and environments while ensuring repeatability.

Data Scientist, SomaDetect Inc., Remote July 2020 – October 2021

- **Image Classification:** Designed, tested, and deployed a quality control algorithm for selection of high-quality image data for subsequent biological predictions. Design involved extraction of a variety of image feature sets and testing of numerous model architectures. The optimal model (convolutional neural network) was deployed on sensors (edge), where it significantly improved collected data quality and reduced data transfer and storage costs.
- **Time Series Analysis:** Designed an image-time-series flow detection algorithm (bidirectional LSTM) capable of identifying when milk was flowing through sensors with high accuracy.
- **Statistics and Data Visualization:** Developed statistical tests to validate performance of a variety of machine learning models and facilitate the biological interpretation of results.
- **Software development:** Packaged machine learning models into Python libraries for edge deployment on company's sensors.

- **Scientific communication:** Responsible for meeting with external stakeholders including farmers, venture capitalists, and researchers to clearly communicate the methods and efficacy of the machine learning algorithms used in the company's products.
- Worked as a member of a team employing Agile management, and project management tools including Jira, Confluence, and Trello.
- Maintained and documented software using Git and Bitbucket for version control.

Postdoctoral Researcher, Centre for Biodiversity Genomics, Guelph, ON May 2019 – July 2020

- **Data science and machine learning:** Designed an alignment-free taxonomic classifier for DNA sequence data.
 - The program uses a custom Python class to extract features from DNA sequence fragments of unidentified origin and assign taxonomic classifications with greater than 99.5% accuracy through evaluation with a neural network (tensorflow).
 - Classifier design involved testing machine learning algorithms (KNN, xgboost, SVM, random forest, and neural networks), feature sets, and hyperparameter combinations to produce an optimal sequence classifier.
 - Designed and documented both a command line interface and a Python package to make the method available to a wide variety of users.
 - Available on GitHub: github.com/CNuge/alfie and PyPi: pypi.org/project/alfie/1.0/
- **Software design:** Development of novel software tools that utilize supervised machine learning models for error identification and correction in DNA barcode data.
- **R package development:** Created the R packages coil and debar.
 - **coil** is a software tool that allows users to frame and translate DNA barcode data of varying quality. coil operates as a binary classifier, identifying the presence or absence of insertion and deletion errors with greater than 97% accuracy. Available on CRAN: <https://CRAN.R-project.org/package=coil>.
 - **debar** is a denoiser for DNA barcode data. The program corrects greater than 95% of insertion and deletion errors and enables the generation of more accurate DNA barcodes and consensus sequences. Available on CRAN: <https://CRAN.R-project.org/package=debar>.
- **Communication and Teaching:** Authored scientific publications to communicate research findings to a wide audience. Lead tutorials where I teach programming and statistics techniques to graduate students and faculty.

PhD Candidate, University of Guelph

September 2014 – April 2019

Thesis: [The Arctic charr \(*Salvelinus alpinus*\) genome: An analysis of evolutionary change after whole genome duplication.](#)

- **Marker Discovery:** Identified over 30,000 novel genetic markers (SNPs) for Arctic charr using custom data processing pipelines and Bash scripts on remote-access servers.
- **Creation of Arctic charr linkage map:** Assembled a 4,508 SNP marker linkage map of the genome, which characterized all Arctic charr chromosomes and identified homologs in related species. These results described evolutionary patterns that follow whole genome duplications.
- **Design of genotyping array for Arctic charr:** Created and utilized Bash scripts, MySQL queries and custom Python programs on remote-access servers to validate and select molecular markers for designing an Affymetrix Axiom custom genotyping array based on biological criteria, technical requirements and the needs of collaborators with a variety of goals.
- **Genome-wide association studies:** Determined the genetic basis of size and growth rate in farmed Arctic charr populations using linear mixed effect regression models and genome-wide association study (GWAS) methodologies.
- **Data science and programming:** Designed custom programs to clean, visualize and statistically analyze large DNA sequence datasets in the programming languages Python, R, and Go.

- **Oral, written and graphical communication:** Communicated findings to colleagues and the public through scientific publications and award-winning conference presentations.
- **Collaboration:** Collaborated with researchers from other universities (Simon Fraser University, University of Victoria) and government agencies (Ontario Ministry of Agriculture, Food and Rural Affairs) to complete large, long term projects.

PUBLICATIONS

- Nugent CM**, Elliott TA, Ratnasingham S, Hebert PDN, Adamowicz SJ. 2021. Debar: A sequence-by-sequence denoiser for COI-5P DNA barcode data. *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.13384.
- Kess T, Dempson JB, Lehnert SJ, Layton KK, Einfeldt A, Bentzen P, Salisbury SJ, Messmer AM, Duffy S, Ruzzante DE, **Nugent CM**, Ferguson MM, Leong JS, Koop BF, O'Connell MF, Bradbury IR. 2021. Genomic basis of deep-water adaptation in Arctic Charr (*Salvelinus alpinus*) morphs. *Molecular Ecology*, 30(18):4415-32. DOI: 10.1111/mec.16033.
- Layton KK, Snelgrove PV, Dempson JB, Kess T, Lehnert SJ, Bentzen P, Duffy SJ, Messmer AM, Stanley RR, DiBacco C, Salisbury SJ, DiBacco C, Salisbury SJ, Ruzzante DE, **Nugent CM**, Ferguson MM, Leong JS, Koop BF, Bradbury IR. 2021. Genomic evidence of past and future climate-linked loss in a migratory Arctic fish. *Nature Climate Change*, 11(2):158-65. DOI: 10.1038/s41558-020-00959-7.
- Nugent CM**, Adamowicz SJ. 2020. Alignment-free classification of COI DNA barcode data with the Python package Alfie. *Metabarcoding and Metagenomics*, e55815. DOI: 10.3897/mbmg.4.55815.
- Salisbury SJ, McCracken GR, Perry R, Keefe D, Layton KK, Kess T, **Nugent CM**, Leong JS, Bradbury IR, Koop BF, Ferguson MM. 2020. Limited genetic parallelism underlies recent, repeated incipient speciation in geographically proximate populations of an Arctic fish (*Salvelinus alpinus*). *Molecular Ecology*, 29(22):4280-94. DOI: 10.1111/mec.15634.
- Nugent CM**, Elliott TA, Ratnasingham S, Adamowicz SJ. 2020. coil: an R package for cytochrome C oxidase I (COI) DNA barcode data cleaning, translation and error evaluation. *Genome*, 63(6):291-305. DOI: 10.1139/gen-2019-0206.
- Layton KKS, Dempson JB, Snelgrove PV, Duffy SJ, Messmer AM, Paterson IG, Jeffery NW, Kess T, Horne JB, Salisbury SJ, Ruzzante DE, Bentzen P, Côté D, **Nugent CM**, Ferguson MM, Leong JS, Koop BF, Bradbury IR. 2020. Resolving fine-scale population structure and fishery exploitation using sequenced microsatellites in a northern fish. *Evolutionary Applications*, 13(5), 1055-1068. DOI: 10.1111/eva.12922.
- Nugent CM**, Christensen KA, Easton AA, Rondeau EB, Ouellet-Fagg CL, Davidson WS, Koop BF, Danzmann RG, Ferguson MM. 2019. Design and characterization of an 87K SNP genotyping array for Arctic charr (*Salvelinus alpinus*). *PLoS One*, 14(4) e0215008. DOI: 10.1371/journal.pone.0215008.
- Khan R, **Nugent CM**, Scaffidi MA, Gimpaya N, Grover SC. 2019. Association of biologic prescribing for inflammatory bowel disease with industry payments to physicians. *JAMA Internal Medicine*, 179(10) 1424-1425. DOI: 10.1001/jamainternmed.2019.0999.

Nugent CM, Easton AA, Norman JD, Ferguson MM, Danzmann RG. 2017. A SNP based linkage map of the Arctic Charr (*Salvelinus alpinus*) genome provides insights into the diploidization process after whole genome duplication. *G3: Genes, Genomes, Genetics*, 7(2) 543-556. DOI: 10.1534/g3.116.038026.

ADDITIONAL WORK EXPERIENCE & PERSONAL PROJECTS

Collaborative analysis of drug company payment influences June 2018 – July 2019

- Collaborated with several gastroenterologists from St. Michael's Hospital (Toronto) on a research project that resulted in a publication in the journal: JAMA Internal Medicine.
- Responsible for cleaning and statistically analyzing several years of Medicare prescription and drug company payment records in order to correlate payments made from drug companies to physicians, with the tendencies of physicians to prescribe certain medications.
- Wrote custom R and Python scripts to search and clean large datasets, conduct statistical analyses, and visualize results.

Software Engineer, Visa Run Inc. July – August 2017 (contract)

- **Software design:** Created a custom Python web crawler for market research and customer outreach, which was used to search Instagram and identify high-quality collaborators and promoters to help grow awareness of company's products.
- **Collaboration:** Worked closely with marketing team to tailor software to their specific business and computational needs.

Zillow's Home Value Prediction Challenge link: tinyurl.com/zillowcomp

- Competed in Kaggle data science prediction challenge.
- Finished in the top 7% of worldwide competitors by designing an ensemble solution composed of gradient boosting and neural network machine learning techniques.

Kaggle.com Expert designation link: tinyurl.com/NugentKernels

- Achieved expert ranking in the Kernels (top rank: 110th out of 141,728) and Datasets (top rank: 25th out of 13,272) categories.
- Construction of award-winning data science tutorials and scripts for tasks including: neural network design, data visualization, data augmentation and feature engineering.

BOLD-CLI link: github.com/CNuge/BOLD-CLI

- A command line interface I created for the fast and efficient retrieval of data from the Barcode of Life Data System (<https://www.boldsystems.org/>).

go-fasta link: github.com/CNuge/go-fasta

- A Go program I wrote for obtaining, manipulating, and summarizing DNA sequence files from a public database (NCBI).

SKILLS

Programming languages: Python, R, Bash, C++, Go, SQL

Data scientist, software engineer, and bioinformatician

- Expert in the use of Python and R for statistical analysis, data cleaning, and data visualization.
- Expert in the application of machine learning algorithms, especially for biological problems.
- Experienced in the analysis of image and time series data.

- Designed user-facing programs in Python, R, and Go for DNA sequence analysis and data manipulation.
- Extensive experience cleaning and analyzing large datasets in a command line environment.
- Experienced in use of remote access cloud computing clusters to conduct large-scale data analysis.
- Packaging and deployment of code for efficient reuse.
- GitHub page: github.com/CNuge.

Additional programming skills

- Use of Git/GitHub for software design and project management.
- Experienced Linux user.
- Development, testing, and maintenance of R and Python packages.
- Data visualization (ggplot), web scraping, data cleaning, distributed systems (PySpark).

Bioinformatics specialties

- Linkage mapping, genome-wide association studies (GWAS), genotyping array design and use, genome assembly, biological application of machine learning algorithms, DNA sequence alignment and comparison, DNA barcoding, analysis of high-throughput sequencing data.
- Fast learner with the ability to adapt to and effectively implement new bioinformatics techniques.

Biologist

- Experienced in experimental protocols using a wide range of genetic techniques, including: DNA extractions, PCR, polyacrylamide gels, and agarose gels.
- Designed and executed experiments involving measurement of mass, length and sexual maturity of thousands of Arctic charr in an aquaculture setting.
- Management of personnel in execution of experiments in a field setting.

GUEST LECTURES & CONFERENCE PRESENTATIONS

2021

- Selected information on job searching, resumes, and technical interviews for bioinformaticians. Invited speaker for BINF*6999 Bioinformatics Master's Project class. University of Guelph, Canada.

2020

- Information on resumes and technical interviews. Invited speaker for BINF*6999 Bioinformatics Master's Project class. University of Guelph, Canada.
- Improving the accuracy of biodiversity analyses from high-throughput sequence data with the R package debar. Invited speaker for TorBUG: Toronto Bioinformatics Users Group. Toronto, Canada (remote).

2019

- Designing R packages: moving from scripting to software development. Guest lecture for BINF*6210 Software Tools for Biological Data Analysis. University of Guelph, Canada.

2018

- The development of fast growing, late maturing and salinity tolerant strains of Arctic charr. Atlantic Innovation Fund project meeting. Coastal Zones Research Institute, Shippagan, Canada.
- The Arctic charr genome: evolutionary change after whole genome duplication. Invited speaker Centre for Biodiversity Genomics, Guelph, Canada.

- Design and characterization of an 87K SNP genotyping array for Arctic charr (*Salvelinus alpinus*). 9th International Charr Symposium. Duluth, USA.
- How can two chromosomes look like one? Graduate Student Seminar Series. University of Guelph, Canada.

2017

- Biological applications of machine learning algorithms. Evelyn Pielou Discussion Group presentation. University of Guelph, Canada.
- Effective use of R in undergraduate research projects. Guest lecture for IBIO*4600 Integrative Marine and Freshwater Research. University of Guelph, Canada.
- Genome duplication and subsequent evolution in vertebrates. Guest lecture for ZOO*4910 Integrative Vertebrate Biology. University of Guelph, Canada.
- Gathering stock data with Python - changing methods following the demise of Yahoo Finance. PyCon Canada 2017. Montreal, Canada.

2016

- The genomic architecture of Arctic charr: using a SNP-based linkage map to characterize evolution after whole genome duplication. 3rd International Conference on Integrative Salmonid Biology. Puerto Varas, Chile.

2015

- The genomic architecture of Arctic charr: using genotype by sequencing to produce a dense SNP linkage map. 8th International Charr Symposium. Tromsø, Norway.

AWARDS

- Best student talk, 8th International Charr Symposium, Tromsø, Norway 2015.
- Kaggle.com Kernels Award Winner, January 26th 2018. “Geospatial Feature Engineering and Visualization” by Cam Nugent tinyurl.com/KernelAward.

TEACHING & MENTORSHIP EXPERIENCE

Mentor to Bioinformatics MSc. students, University of Guelph May 2019 – July 2020

- Provided programming and technical support to five Bioinformatics MSc students during weekly or fortnightly meetings depending on student needs.
- Gave feedback on research proposals and project design.
- Helped students with programming, development of data analysis workflows, statistical analysis, data visualization and interpretation of results.

Teaching Assistant, University of Guelph September 2014 – April 2019

Courses: Biostatistics, Genomics, Evolution, Discovering Biodiversity, Marine and Freshwater Research

- Ran hands-on R programming tutorials, lead lab and seminar sessions, ran exam review sessions, wrote weekly quizzes, provided feedback on experimental design and led field excursions with students.

Undergraduate Mentor, University of Guelph

September 2017 – April 2018

- Mentored two students in their fourth year of undergraduate studies through the Integrative Biology Undergraduate Mentorship Program (iBUMP).
- Helped students to design and execute research projects.
- Helped students explore post-graduate research project options and prepare job applications.

PROFESSIONAL SERVICE

- Reviewer: Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics.
- Reviewer: Great Lakes Fishery Commission's Fishery Research Program – 2020 funding cycle.
- Member of the Society for Molecular Biology & Evolution (Member Number: 00038676).

EXTRA-CURRICULAR INVOLVEMENT & INTERESTS

- Speed River Cycling Club member, 2016-2020.
- University of Guelph Varsity Squash team member, 2017-2018.
- University of Guelph Varsity Ultimate Frisbee team member, 2014-2017.
- Crash Ultimate Frisbee (Waterloo), 2014 Canadian mixed division national champions.
- Queen's University Varsity Ultimate Frisbee team member, 2013 – 2014.