W. Augustine Dunn, III Ph.D.

4 Lester PL, #3
Jamaica Plain MA, 02130
770-312-9544
w.gus.dunn@gmail.com
gusdunn.com

Technical Expertise

EXPERT
Git, Python, Snakemake, Bash, Regular Expressions, Jupyter Notebooks, Advanced Data Visualization, Reproducible/Shareable Anlysis pipelines, Amazon Web Services, Computational Environment Provisioning, HPC job schedulers (SLURM, SGE & PBS), YAML, HTML, CSS, XML, LaTeX, pandoc, markdown, reStructuredText, Python software packaging & templating, Software Documentation with Sphinx, Gnu Parallel, Linux, Basic System Administration

INTERMEDIATE
Perl, R, Makefiles/build-systems, JSON, vim, unit testing, Mercurial, Bazaar, Subversion

BASIC
MySQL, PostgreSQL, SQlite, Lua, Tcl, Apache, javascript, MongoDB

PYTHON LIBS
pandas, scikit-learn, scipy, numpy, statsmodels, pyMC3, matplotlib/pyplot, seaborn, ggplot, rpy2, networkx, pybedtools, pysam, cookiecutter, jupyter, flask

MISC
OSX, Windows, Linux, MS Word, MS Excel, Photoshop/Gimp, Illustrator/Inkscape

WET-LAB
RNA-seq, ddRAD-seq, proteomics, broad range of molecular biology & protein biochemistry techniques

Authored Software

bespoke pipelines
Repeatable, failure tolerant, concurrent analysis solutions with built-in reporting.

table-enforcer
A python package to facilitate the iterative process of developing and using schema-like representations of DataFrames with pandas.

blacktie
An object oriented python pipeline that simplifies & streamlines the running of complex tophat/cufflinks-based RNA-seq experiments to a single command plus configuration file: prioritizing repeatability & usability. Downloaded from `https://pypi.python.org/pypi/blacktie` over 9,000 times.

gFunc
A python-based integrative analysis framework using network graphs to combine multidimensional data-types from disparate "Omics" sources for creating/exploiting functional-genomic gene sets across multiple species.

spartan
A bioinformatics package, providing the essentials to get a variety of computational jobs done quickly without flourish when that is all that is needed.

Experience

2018–present — **Computational Scientist**, *Computational Sciences*, Repertoire Immune Medicines, Cambridge, MA.
Support development of immune regulating medicines through novel single cell assays to decode the immune synapse.
Highlights:
- *Reduced load on production pipeline devs by establishing and maintaining a cloud-based HPC cluster that empowers our Tech-Dev team to run their own analyses in scalable, state-of-the-art computing environments.*
- *Developed tool to simulate diploid individuals drawn from the U.S. population preserving racial sub-structure of HLA allele haplotype frequencies to aid in subject recruitment and indication selection.*
- *Developed command line tools to access and evaluate pipeline data in S3 locations or local storage.*
- *Curated and integrated multiple large-scale cancer mutation database resources into a purpose oriented dataset to aid selection of cancer epitope targets for future IND submission.*
- *Developed multiple web-apps to make custom analysis tools accessible to experimentalists.*
- *Acted as computational liaison to multiple lab project groups to facilitate experimental design and analysis, as well as communicate implications of project results.*

2016–2018 — **Postdoctoral Associate**, *Div. of GI/Nutrition - Research*, Boston Children's Hospital, Boston, MA.
Genomic and clinical analysis of Very Early Onset Inflammatory Bowel Disease (VEOIBD) and prediction of treatment outcomes using Machine Learning.
Highlights:
- *Lead data management and sharing for international consortium.*

| 2014–2016 | **Postdoctoral Associate**, *Dept. of Ecology & Evolutionary Biology*, Yale University, New Haven, CT. |

Characterization of gene-flow & genotype/phenotype relationships in tsetse fly populations in Uganda.

Highlights:
- *Supervised month-long field expedition collecting tsetse flies in northern Uganda.*
- *Maximized sensitivity while balancing specificity by implementing Bayesian filtering method to set thresholds for "interesting" SNP-pairs in distributions of distance-binned linkage disequilibrium values with ddRAD-based population genomics results.*
- *Designed & implementing relational database to track/curate samples with web front-end.*
- *Established & maintained a shared environment modules-based software library for the lab group on super-computing cluster.*
- *Substantially increased the potential user-base of* pyrad *(open-source RAD-seq analysis pipeline) by re-factoring project to allow one-step installation with the conventional* pip install pyrad *form.*

| 2008-2014 | **Doctoral Research**, *Anthony James Lab*, University of California – Irvine, Irvine, CA. |

Studied post-bloodmeal gene expression in midguts of four mosquito species in support of designing synthetic DNA elements for engineering disease transmission-refractory phenotypes.

Highlights:
- *Awarded national NIH/NLM bioinformatics training fellowship (see Honors & Awards) providing three years of funding, formal training in computational methods, & experience communicating effectively between biologists and computer scientists.*
- *Applied data mining methods such as clustering, hidden Markov models, expectation maximization, & dimension reduction to large-scale multidimentional transcriptomics data.*
- *Integrated & applied multiple genome-scale data-types to complex phenotypes of expression profiles in species separated by 250 MY to generate testable hypotheses about underlying transcriptional control mechanisms.*
- *Set up/maintained local Galaxy-based site along with custom workflows & bespoke analysis tools to provide data-sharing & provenance for transcriptomic analyses.*
- *Established numerous & productive multidisciplinary collaborative relationships which generated 11 publications from 2008-2014.*

| 2009 | **Master Consultant**, *KDH Research & Communication*, Atlanta, GA. |

Reviewed & provided *pro bono* feedback & corrections on materials for "*Genetics for Kids*" modules, a supplement for communicating genetics in public school curricula.

| 2006–2007 | **Research Laboratory Technician**, *Geomicrobiology Group*, University of Southern California, Los Angeles, CA. |

Studied role of microorganisms in mediating rock, mineral, & organic matter transformations on progressively older mid-oceanic basalt samples.

Highlights:
- *Set up a new lab & interfaced with construction contractors while supervisor was at sea.*
- *Established the lab's capacity to perform qPCR for the purpose of comparing bacterial population structures on aged basalt.*

| 2003–2006 | **Laboratory Technician II to Research Professional I**, *Dept. of Cellular Biology*, University of Georgia, Athens, GA. |

Characterized the expression & annotated the gene models of cuticular protein genes in *Anopheles gambiae*.

Highlights:
- *Established the lab's capacity to perform qRT-PCR.*
- *Trained three other labs in the theory, protocols, & analysis of qRT-PCR.*
- *Wrote several bespoke Perl tools to aid in primer design & a web-based primer database to manage the hundreds of real-time primer-pairs.*

Education

2008–2014	**PhD, Biological Sciences**, *University of California – Irvine*, Irvine, CA.
TOPIC	*Comparative Transcriptomics of Blood-feeding in Midguts of Three Disease-vector Mosquito Species*
ADVISOR	Professor Anthony James (Molecular Biology & Biochemistry)
CO-ADVISOR	Professor Xiaohui Xie (Department of Computer Science)
1998–2003	**BS, Biology (*emphasis in molecular genetics*)**, *University of Georgia*, Athens, GA.
MENTOR	Professor Judith Willis

Selected Coursework

| UC IRVINE | Representations & Algorithms for Molecular Biology, Probabilistic Modeling of Biological Data, Quantitative Methods in Biology |

Honors & Awards

2011 The Pacific-Southwest Regional Center of Excellence for Biodefense & Emerging Infectious Diseases Annual Meeting Travel Award.

2010 President of UCI's IGB Biomedical Informatics Training fellows

2009-2012 Biomedical Informatics Training fellow (NIH/NLM 5T15LM007443)

Selected Publications

Bonizzoni, Mariangela et al. "Probing functional polymorphisms in the dengue vector, Aedes aegypti". In: *BMC genomics* 14.1 (2013), p. 739.

Bonizzoni, Mariangela et al. "Complex Modulation of the *Aedes aegypti* Transcriptome in Response to Dengue Virus Infection". In: *PLoS ONE* 7.11 (Nov. 2012), e50512.

Bonizzoni, Mariangela et al. "Comparative Transcriptome Analyses of Deltamethrin-Resistant and -Susceptible *Anopheles gambiae* Mosquitoes from Kenya by RNA-Seq". In: *PLoS ONE* 7.9 (Sept. 2012), e44607.

Bonizzoni, Mariangela et al. "Strain variation in the transcriptome of the dengue fever vector, Aedes aegypti". In: *G3: Genes| Genomes| Genetics* 2.1 (2012), pp. 103–114.

Bonizzoni, Mariangela et al. "RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, Aedes aegypti". In: *BMC genomics* 12.1 (2011), p. 82.

Dissanayake, Sumudu N et al. "aeGEPUCI: a database of gene expression in the dengue vector mosquito, Aedes aegypti". In: *BMC research notes* 3.1 (2010), p. 248.

Sieglaff, Douglas H et al. "Comparative genomics allows the discovery of cis-regulatory elements in mosquitoes". In: *Proceedings of the National Academy of Sciences* 106.9 (2009), pp. 3053–3058.