

# W. Augustine Dunn, III Ph.D.

417 Newport Rd

Hull MA, 02045

☎ 770-312-9544

✉ [w.gus.dunn@gmail.com](mailto:w.gus.dunn@gmail.com)

🌐 [gusdunn.com](http://gusdunn.com)

## Technical Expertise

EXPERT	Python, Snakemake, Bash, Regular Expressions, Advanced Data Visualization, Reproducible/Shareable Analysis pipelines, Computational Environment Provisioning, Jupyter notebooks, SGE & PBS HPC schedulers, YAML, HTML, CSS, XML, $\LaTeX$ , pandoc, markdown, reStructuredText, Git, 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
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

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 <u>repeatability</u> & <u>usability</u> . Downloaded from <a href="https://pypi.python.org/pypi/blacktie">https://pypi.python.org/pypi/blacktie</a> over 9,000 times.
gFunc	A python-based <u>integrative analysis framework</u> 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.
bespoke pipelines	Repeatable, failure tolerant, concurrent analysis solutions with built-in reporting.

## Experience

2016–present	<b>Postdoctoral Associate</b> , <i>Div. of GI/Nutrition - Research</i> , 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. <b>Highlights:</b> <ul style="list-style-type: none"><li>◦ <i>Lead data management and sharing for international consortium.</i></li></ul>
2014–2016	<b>Postdoctoral Associate</b> , <i>Dept. of Ecology &amp; Evolutionary Biology</i> , Yale University, New Haven, CT. Characterization of gene-flow & genotype/phenotype relationships in tsetse fly populations in Uganda. <b>Highlights:</b> <ul style="list-style-type: none"><li>◦ <i>Supervised month-long field expedition collecting tsetse flies in northern Uganda.</i></li><li>◦ <i>Maximized sensitivity while balancing specificity by implementing <u>Bayesian filtering method</u> to set thresholds for “interesting” SNP-pairs in distributions of distance-binned linkage disequilibrium values with <u>ddRAD</u>-based population genomics results.</i></li><li>◦ <i>Designed &amp; implementing relational database to track/curate samples with web front-end.</i></li><li>◦ <i>Established &amp; maintained a shared environment modules-based software library for the lab group on super-computing cluster.</i></li><li>◦ <i>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 <code>pip install pyrad</code> form.</i></li></ul>

- 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:**
- o 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.
  - o Applied data mining methods such as clustering, hidden Markov models, expectation maximization, & dimension reduction to large-scale multidimensional transcriptomics data.
  - o 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.
  - o Set up/maintained local Galaxy-based site along with custom workflows & bespoke analysis tools to provide data-sharing & provenance for transcriptomic analyses.
  - o 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:**
- o Set up a new lab & interfaced with construction contractors while supervisor was at sea.
  - o 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:**
- o Established the lab’s capacity to perform qRT-PCR.
  - o Trained three other labs in the theory, protocols, & analysis of qRT-PCR.
  - o 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)