Matt Ralston
E-mail: professional.bio.coder@gmail.com
Phone: 610-470-8028
Website: https://matthewralston.github.io
Github and Social Media: https://matthewralston.github.io

SUMMARY

Results-oriented laboratory scientist with 6 years of wet lab research, 5 years in R&D, and 13 years of exp. in software engineering. M.Sc. (3.97) in bioinformatic data science, computer science, mathematics, statistics, and B.Sc. in Biochemistry with molecular bio and oncology research. Expert in Python, R, Linux, HPC, and cloud computing (AWS). Developer and maintainer of kmerdb, a kmer database and alignment-free algorithm library for Illumina/Sanger sequencing (WGS/WES/RNAseq) non-alignment sequence similarity and probabilities, minimizer selection, and sequence alignment heuristics.

Published in the areas of RNA-sequencing, oncology, molecular biology, gene expression analysis, bioinformatics, anaerobic fermentation, and biofuels. Broad experience with natural science research both in and outside of the laboratory, everything from analytical chemistry to Illumina library preparation and data analysis, to algorithm development and statistics.

During the pandemic, I transitioned from the pharma giant Bristol Myers Squibb (where I won a LevTech Innovation Award, 2015-2020) and Bayer Crop Sciences (2020) to Newark DE, attended Data Science classes at UDel, improved my portfolio (Multivariate, ML/AI, Data-Science, genomics) and I've done a mixture of Doordash and website consulting for University of Delaware professors to make ends meet. I've proved resilient and honed a broad range of skills in Computer Science, Data Science, Probability, Linear Algebra, Statistics, ML/AI, Web Development and more. Now, I'm looking for a dynamic team to leverage my background in natural sciences, maths, and comp-sci to address stakeholder needs alongside exciting opportunities for interpersonal and business skill development.

EDUCATION

University of Delaware   2012-2015
Masters (M.S.) in Bioinformatics and Computational Biology

- Thesis: *Assembling Improved Gene Annotations in Clostridium acetobutylicum with RNA Sequencing.*

- 3.97 GPA

University of Delaware   2008-2012
Bachelors (B.S.) in Biochemistry

- ● Minor in Biochemical Engineering

- ● Merit Scholarships

- ● Undergraduate Research Fellow

AT A GLANCE

2008-2015: research assistant, 2015-2020: pharmaceutical industry, 2021-2024 (freelance and web design): developing LLC IP, open source software development, and restaurant delivery

SKILLS

Business and research:

Agile Development, Scrum, **GitHub, Electronic Laboratory Notebooks, ELN, GLP, SOP,** Atlassian Jira/Confluence, Requirements gathering, **Project management,** value propositions, kanban, sprints, prioritization, Cost-benefit analysis, budgeting, business slidedecks, presentation, Public Speaking, amortization, impact forecasting, stakeholder framing, abstracts, Technical writing, modular writing, methodology, Business narratives, referencing, citations, "Living" documentation (i.e. Quarto/Rmd, Markdown, Jupyter) aka literate programming, scientific notebooks, SOP writing, data justification, Powerpoint, **MS Office** (Word, Excel, Powerpoint, Access, OneNote), spreadsheets, Canva, Figma, MacOS/OSX, **Linux, Cloud computing, AWS/GCP,** High-performance computing, HPC. Research, Discovery, R&D, and much more.

COMPUTER SKILLS:

Programming:

Python, R, Bash, Rust, Linux, web-development, REST-APIs, Django, FastAPI, Flask, SQLAlchemy, Alembic, SQLite, libSQL, MySQL, Oracle SQL, PostgreSQL, **numpy,** pandas, Jupyter notebooks, Rstudio, Rshiny, Bioconductor, **Rmarkdown, Quarto, Markdown,** static compiled languages (rustc/cargo, zig, Typescript, Scala/Java/Clojure, C extensions in Python and R), RAII, CUDA, Cython, anaconda, LaTeX/Pandoc, Statistics, Data Science, Hypothesis testing, bonferonni, t-test, f-test, **ANOVA, Linear regression,** OLS, discrete/continuous probability, multivariate statistical analysis, PCA, UMAP,

XGBoost, random forests, clustering, classification, distribution fitting, **Amazon Web Services, AWS,** Google Cloud Platform, GCP, Azure, **Docker,** k8s, High-Performance Computing, HPC, Javascript/NodeJS, Ruby, Emacs, d3.js, Julia, Perl, awk, grep/sed, LaTeX**,** Matlab, HTML, HTML5, CSS3, technical writing, literate programming, Systems Administration, sysadmin, systemd, tmux, Nginx, Apache, web servers, SGE/UGE, bash, parallel, time, recurrence relations, algorithms

Data Science:

Statistics: p-values, Student t-test, Chi-Square test, Fisher's Exact Test, Hypothesis testing, **ANOVA / GLM/ regression,** linear algebra, matrices, limma, distribution fitting, R/Rstudio/Rshiny, Bioconductor, discrete math, discrete/continuous probability, multivariate models, model analysis, multivariate statistical analysis, regression, least-squares estimation, PCA, clustering, classification, normalization, regularization, variance reduction, linear regression, naive Bayes, random forest, XGBoost, DESeq2,

Dimensionality reduction: PCA/SVD, Uniform Manifold Approximation and Projection (UMAP), t-SNE, canonical correlation analysis,

Optimizers: Gradient descent, simplex, particle swarm, simulated annealing, self-organizing maps, others.

Maching learning & AI: perceptron, SVM, PCA, UMAP, t-SNE, Deep Neural Networks, Recurrent Neural Networks, and Artificial Intelligence (AI), and formal OLS-family assumptions and caveats, SAS programming, R programming, Python programming

Bioinformatics:

Sequence Alignment, VCF, **SAM/BAM, Bowtie2**, Tophat, STAR, HISAT2, BWA, blastn, blastp, **BLAST,** Blat, clustal, multiple-sequence alignment, **Genome Assembly, Transcriptome Assembly,** Metagenomics, Trinity, Velvet, Prokka, BUSCO, checkM, Cufflinks, FastQC, **Samtools, PicardTools, Bedtools,** Cuffdiff, DEseq2, Limma, Circos, Amazon Web Services, SASS, SaaS development

LABORATORY SKILLS:

Illumina NGS, library preparation, Whole Genome Sequencing, Whole Exome Sequencing, **WGS**, **RNA-seq, WES,** qRT-PCR, RT-PCR, SDS-PAGE, Western blot, Northern blot, **HPLC, GC-MS, LC-MS**, Nucleic acid extraction, enzymatic treatment, FACS, Light microscopy, Fluorescence microscopy,

immunohistochemistry, IHC, immunofluorescence, IF, UV-Vis, Fluorometry, Nanodrop, Human cell culture, anaerobic cell culture, anaerobic fermentation, bacterial cell culture.

WORK EXPERIENCE (COMPUTATIONAL)

Bristol Myers Squibb   2015-2019

Research Scientist II – Genomics and Computer Assisted Drug Design (CADD)
- Galaxy bioinformatics server administrator, sequencer automation, Amazon Web Services (AWS) integration, NGS WGS/WES/RNAseq QC pipeline and large sample (20-500+) summary reports, and user support (Comp Bio)
- Developed expertise in expression technologies (Affymetrix, Illumina RNAseq) from sequence alignment/assembly,contamination detection, linear regression and ANOVA modeling, sysadmin (CompBio)
- Built a large deduplicated dataset version control system, provided git expertise that won a LevTech innovation award. (Comp. Bio)
- Built services for pan-organization chemical database mirroring and ETL, structural and UUID deduplication of pharmacophore libraries. (CADD)
- Expanded a multi-platform theoretical chemistry framework (Schrodinger, Openeye, ChemAxon) to enumerate all protomers, tautomers, and rotamers/conformers (2D->3D enumeration) (CADD)
- Parameter optimization and scalability investigation of the framework across multiple chemotypes. Built SMARTS filters for eliminating chemically implausible substructures from the enumerations.
- Increased skills with AWS, Flask/FastAPI full-stack development, Linux, git, bash, NodeJS, Python, R/Biocondustor, PostgreSQL and Oracle, software engineering, and education.

Bayer Crop Sciences   June-December 2020

Genomics Pipeline Developer – Microbial Genomics and Data Science (GDS)
- Used AWS CLI, developer console, and custom in-house tools Amazon workflow language/CDL and Nextflow to manage resources related to

pipeline scale-up, fault tolerance, proper interfaces and co-development with the RDBMS team.

● Upgraded a pipeline from 5 modules to 24 modules in 6 months, adding substantial new bioinformatics capabilities, including biosynthetic gene discovery, phylogenetic annotations, genome completeness metrics, and other functional genomics annotations.

● Developed ETL and database specifications for dataset generation and fault tolerance in cloud (AWS) deployments.

Terra Informatics LLC   June 2022 -June 2025

Principal Consultant and sole proprietor
  ● Django, FastAPI, HTML5/CSS3, Javascript, PostgreSQL/MariaDB/SQLite, Kubernetes(k8s), Docker, bioinformatics tools (bowtie2, bwa-mem, fastqc, limma, R/bioconductor, ANOVA) and much more.
  ● Currently seeking Delaware Small Business Development Council facilitated 7(a) loans to develop the consultancy in Newark DE for U. Delaware faculty and local businesses.
  ● Building IP backbone for a consultancy offering. Pipeline-as-a-Service (microSaaS), web development, and data science services.
  ● Working with publicly available and synthetic datasets to characterize genome similarities and produce Kubeflow/Airflow pipelines for cloud-agnostic burst computing on AWS, GCP, and other cloud providers.
  ● Currently benchmarking methods for RNAseq service glue code and infrastructural related priorities.
  ● Developed and maintained a 48CPU server with conformant kubenetes, 38-component pipeline for multi-cloud and cloud-agnostic deployment (~$1k – 5k per website, $10k+ for data science services).

WORK EXPERIENCE (LAB)
     Note: see "Skills" section for analytical chemistry, mol bio, and bioinformatics skills

Papoutsakis Laboratory   2012-2015
Graduate Research Assistant
  ● Utilized batch fermentation in bioreactors to produce bacterial cultures for sampling and nucleic acid extraction
  ● Developed a laboratory protocol to produce strand-specific Illumina RNA-seq libraries from high quality RNA.

- Ensured RNA quality by designing and implementing quality control analytical checkpoints.
- Constructed a bioinformatics pipeline to process, align, and assess 1.5 billion paired-end Illumina reads (RNAseq)
- Produced and corrected a transcriptome assembly from a range of culture conditions.
- Supported a team of scientists and engineers with computational analyses.

Center for Translational Cancer Research   2010-2012
Undergraduate Research Fellow
- Characterized the apoptotic response of colon cancer cell lines to statin treatment in human cell culture with Annexin V assay.
- Investigated differential BMP-pathway signaling in fresh and FFPE colon tissue with immunohistochemistry, fluorescence microscopy, Western blot.
- Assisted researchers with fluorescence assisted cell sorting (FACS) and flow cytometry.
- Presented BMP-pathway IHC micrographs and culture kinetics in response to statins in fellowship poster session.

AWARDS

Bristol Myers Squibb   2018
  LevTech Innovation Award: Project Utilities for Reproducible Research
- Co-developer of a large dataset version control system (git for large files) that included a CLI and Electron application for storage, Access Control Lists, deduplication, and encryption for clinical and discovery datasets.
- Worked with Mark Russo (Assoc. Director) to release the framework as part of the "Checkmate" series of clinical trials for NSCLC combination therapies or Opdivo (nivolumab) and Yervoy(ipilimumab)

Fraunhoffer and NSF Integrative Graduate Education and Research   2013
  Center for Manufacturing Innovation - **Winning Proposal**
- During an IGERT course with an industry component, a winning proposal was delivered to Fraunhoffer CMB for a revision to their GMP vaccine production system featuring TThe proposal included an innovative technological process to improve product purity and reduce production overhead.

Helen Graham Cancer Center   2011-2012
  Research Fellowship
- While at the CTCR, I utilized various molecular techniques (Western blot, immunofluorescence, RT-PCR) to study apoptotic and differentiating effects of lovastatin and simvastatin on colon cancer cell lines.

University of Delaware   2008-2010
    Merit Scholarship and Honors Program

PUBLICATIONS

- Ralston, Matthew T., and Eleftherios T. Papoutsakis. "RNAseq-based transcriptome assembly of *Clostridium acetobutylicum* for functional genome annotation and discovery." AIChE Journal 64.12 (2018): 4271-4280.

- "The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in *C. acetobutylicum*", contributing author. BMC Genomics. 14.1 (2013): 849.

- "Overexpression of the *L. plantarum* peptidoglycan biosynthesis murA2 gene increases the tolerance of E. coli to alcohols and enhances ethanol production," contributing author. Applied Microbiology and Biotechnology. 98.19 (2014): 8399-8411.

- "Complex and extensive post-transcriptional regulation revealed by integrative proteomic and transcriptomic analysis of metabolite stress response in *Clostridium acetobutylicum,*" contributing author. Biotechnology for biofuels 8.1 (2015): 81.

LEADERSHIP
Terra Informatics LLC   2022
Sole Proprietor and Principal Consultant

- Building an IP backbone structure for RNA-Seq, DNA-Seq, metagenomics, and more. Using conformant kubernetes(k8s) to deploy the pipeline on multi-cloud. A MicroSaaS offering primarily for academic researchers in the Ammon Pinizzotto Engineering and Bioinformatics departments. Targeting collaborations in the incubation community amongst the notorious Chemical Engineering dept of the Univ. of Delaware. Sequencing is needed more than ever.

- Working on sequence alignment algorithm and minimizer selection methods for alignment, and de Brujin gralh de novo assembly, sequence alignment heuristics, Smith Waterman from-scratch alignment methods in Cython and Rust,

Bioinformatics Student Association   2013-2014
President

- Organized and presented in a UNIX and high performance computing (HPC) workshop.

- Secured funding for this and other student development workshops.

IGERT Business, Ethics, and Communication in Life Sciences   2013
Project Manager

- Won a proposal for a revision to Fraunhoffer CMB's GMP vaccine production process.

National Honor Society, A.I. chapter   2008
Vice President

- Coordinated volunteer activities

A.I. Marching Band   2008
Vice President

- Coordinated field show, trip to Rose Bowl

PROJECTS

Please see my GitHub or primary website to understand more about my research projects and my professional portfolio.
https://matthewralston.github.io/portfolio

Consultancy Django application   2022
Principal Consultant – Terra Informatics LLC

Vanilla Django + HTML5 templating, 4 apps with PostgreSQL, marketing pages, login functionality, Stripe checkout, HTML, ticket access to microSaaS pipeline offerings (RNASeq, metagenomics). Dashboards and run logging. 38 component, conformant kubernetes SaaS offering with Stripe checkout.

K-mer Database (.kdb)   2019-2024
Software Engineer and Researcher
https://github.com/MatthewRalston/kmerdb
https://matthewralston.github.io/kmerdb

kmerdb is a Python CLI designed for k-mer counting and k-mer graph edge-lists. It addresses the 'k-mer' problem (substrings of length k) in a simple and performant manner. It stores the k-mer counts in a columnar format (input checksums, total and unique k-mer counts, nullomers, mononucleotide counts) with a YAML formatted metadata header in the first block of a bgzf formatted file. I characterized the sensitivity of the analyses to errors from simulated data,

examining the relationship between k-mer profile distances and phylogenetic distance.

Not Very Humerus   2018-2020
Blog Author
https://matthewralston.github.io/blog

A simple blog on beginner programming topics, industry trends, personal coding projects, and current events. Some poetry and fan fiction sprinkled in.

Kmer.js   2018-2019
Software Developer
https://github.com/MatthewRalston/kmer.js

A very very simple in-memory k-mer array.

Liganator v2.0   2019
Software Developer – Bristol Myers Squibb

Developed and planned benchmarking for a utility to explore different vendors, parameters, and more regarding the process of preparing a 2D chemical structure for all its physicochemically identical isoforms, as well as rotational perturbations, in preparation for ligand docking studies. A product designed for protein biochemists.

Project Utilities for Reproducible Research   2016
Software Developer – Bristol Myers Squibb

A NodeJS REST-API and accompanying CLI client for a reproducibility database. It was similar in principle to git-lfs. It's pretty easy to imagine what we did if you have a technical background, we used certain checksums to identify a chunk of data, and then stored metadata associated with it, to make a hard drive on a selected server a reproducibility hub.

AWARDS
IGERT Business, Ethics, and Communication in Life Sciences   2013
IGERT SBE2 Faculty

● Team of students won tuition for a semester by providing a proposal for a
   revision to Fraunhoffer CMB's GMP vaccine production process.
Undergraduate Research Fellowship   2010
Undergraduate Research Program

- Proposed a fractional-factorial investigation of statin treatment on colon cancer cell lines.
- Presented findings in a Research Symposium

2nd Place – Sound of Music event   2008
Delaware Science Olympiad

- Engineered a guitar and PVC chime set
- Wrote original melodies, harmonies

VOLUNTEER
Carisma Peru   2014
Clinical assistant

- Assisted physicians in Peru with clinical laboratory sampling, spirometry, blood pressure measurements in ObGyn and E.R. settings.