Andrew Collin Stewart

Emeryville · California

✉ andrew.c.stewart@gmail.com 📞 +202 505 3633 ⓖ andrewcstewart 🌐 andrewcstewart.com

SUMMARY

Data scientist with a strong background in applied research and development in computational biology, scientific computing, machine learning, and statistics. Passionate about improving business processes through sound experimental design and analysis.

TECHNICAL SKILLS

Expertise	Data analysis, algorithms, machine learning, statistical modeling, quality control, systems engineering, ETL, data mining, optimization, computational biology
Programming	R, Python, Javascript, C/C++, Java, SQL, Git, Docker, AWS
Software	RStudio, Jupyer, tidyverse, ggplot2, caret, pandas, numpy, scipy, scikit-learn, shiny, Tableau, Airflow, Spark, H2O, biopython, bioconductor, Stan
Machine Learning	Classification, regression, clustering, feature engineering, validation
Statistics	Bayesian inference, hierarchical modeling, hypothesis testing, power analysis, dimensionality reduction
Bioinformatics	Sequence alignment, variant calling, expression analysis, NGS

PROFESSIONAL EXPERIENCE

Twist Bioscience, Senior Data Scientist (Bioinformatics)   2015-Present

Develop and deploy predictive models to optimize manufacturing processes involving gene design and assembly, DNA fragmentation screening, and clonal colony picking. Design and analysis of experiments in coordination with process chemists to improve DNA synthesis technology. Primary architect behind data analytics platform utilizing data warehouse and ETL processes. Supervised data science interns.

Veracyte, Scientist II (R&D, Data Analysis Team)   2014-2015

Developed high dimensional classification models for cancer diagnostics. Provided statistical support for cross-run variability studies and power analysis for experimental designs. Drove transition from microarray technology to next-generation sequencing based assays through development of high performance automated data analysis pipelines.

Smithsonian Institution, Data Scientist (Office of the CIO)   2013-2014

Provided analytical support to museum and zoological research scientists in population genetics, phylogenetics, and metagenomics. Developed genomic analyses on scientific computing clusters, including ensemble machine learning approach for classification of signatures of natural selection in genetic populations. Coordinated efforts among researchers, information technology offices, and academic and commercial partners to develop scientific computing infrastructure.

Keygene, Bioinformatician (Applied Systems Biology)   2009-2012

Developed data pipeline for inference of gene regulatory networks using differential expression and network analysis to identify candidate genes for mutagenesis trials in crop engineering applications. Conducted exploratory data analysis and mining of hyperspectral imaging data from field crops to identify signatures of drought exposure as part of efforts to develop high throughput digital phenotyping. Constructed computational models to simulate plant architecture using Lindenmayer systems as a basis for studying optimal space utilization in greenhouse production.

Navy Medical Research Center, Research Assistant (Biodefense Research Directorate)   2006-2009

Applied sequence alignment, gene prediction, and clustering algorithms towards rapid identification of microbial pathogens from genomic sequencing data platforms. Developed high throughput data analytics system for collection, analysis, and reporting of pathogenomic data in SQL database used for multiple biodefense research applications.

EDUCATIONAL EXPERIENCE

Johns Hopkins University, M.S. Computer Science   2015

University of Maryland College Park, B.S. Biology   2006

TEACHING EXPERIENCE

Johns Hopkins University, Graduate Teaching Assistant   2016-2018

605.448 - Data Science
605.445 - Artificial Intelligence

VOLUNTEER EXPERIENCE

Voteocracy, Data Science Volunteer   2016

Applied network analysis techniques to identify marketing targets from Twitter data.

Bayes Impact, Data Science Intern   2015

Developed multiclass predictive model of criminal activity based on geographic and census data.

PUBLICATIONS

Chen PE, Stewart AC, Cook CP. 2010. "Genomic Characterization of the Yersinia Genus." *Genome Biology* 11 (January): R1. doi:10.1186/gb-2010-11-1-r1.

Chen PE, Zwick ME, Stewart AC. 2010. "Rapid Identification of Genetic Modifications in Bacillus Anthracis Using Whole Genome Draft Sequences Generated by 454 Pyrosequencing." *PLOS ONE* 5 (8): e12397. doi:10.1371/journal.pone.0012397.

DiLeo M, Mouron G, Stewart AC. 2011. "Gene Regulatory Network Discovery Through Data Mining: Proof of Concept Through Known Stress Pathway Recapitulation." *ACS Annual Meeting, San Antonio, TX.* https://dl.sciencesocieties.org/publications/meetings/2011am/8617/67897.

Mukhopadhyay S, Read TD, Stewart AC. 2009. "Identification of Bacillus Anthracis Spore Component Antigens Conserved Across Diverse Bacillus Cereus Sensu Lato Strains." *Molecular & Cellular Proteomics* 8 (6): 1174–91. doi:10.1074/mcp.M800403-MCP200.

Pagan M, Stewart AC, Kloos RT. 2016. "The Diagnostic Application of RNA Sequencing in Patients with Thyroid Cancer: An Analysis of 851 Variants and 133 Fusions in 524 Genes." *BMC Bioinformatics* 17 (1): S6. doi:10.1186/s12859-015-0849-9.

Stewart AC, Read TD, Osborne B. 2009. "DIYA: A Bacterial Annotation Pipeline for Any Genomics Lab." *Bioinformatics* 25 (7): 962–63. doi:10.1093/bioinformatics/btp097.

Zwick ME, Read TD, Stewart AC. 2012. "Genomic Characterization of the Bacillus Cereus Sensu Lato Species: Backdrop to the Evolution of Bacillus Anthracis." *Genome Research* 22 (8): 1512–24. doi:10.1101/gr.134437.111.

Zwick MR, Stewart AC, Kiley MP. 2008. "Genotyping of Bacillus Cereus Strains by Microarray-Based Resequencing." *PLOS ONE* 3 (7): e2513. doi:10.1371/journal.pone.0002513.