

James H. Bullard
263 Arkansas St.
San Francisco, CA 94107

bullard@stat.berkeley.edu
510.402.8704
www.stat.berkeley.edu/~bullard

EDUCATION

2006 – 2009 **University of California, Berkeley, CA**
PhD Biostatistics, Designated Emphasis in Computational and Genomic Biology
GPA: 3.93

Relevant Coursework: Applications of Parallel Computing, Statistical Learning Theory, Statistical Phylogenetics, Censored Longitudinal Data Analysis, Statistical Consulting

2004 – 2006 **University of California, Berkeley, CA**
MA Biostatistics
GPA: 3.87

Relevant Coursework: Advanced Introduction to Statistics, Applied Statistical Genetics, Computational Methods in Biology, Probabilistic Models in Computational Biology, Advanced Categorical Data Analysis, Statistical Models I and II, Survival Analysis

1997 – 2000 **Northwestern University, Evanston, IL**
BA Computing and Information Systems

Cumulative GPA: 3.68, Major GPA: 3.94

Relevant Coursework: Operating Systems, Compilers, Databases, Natural Language Processing, Building Internet Services, Enterprise Java Beans, Experimental Mathematics, Game Theory, Probability Theory.

PROFESSIONAL EXPERIENCE

January 2010 – Current, **Menlo Park, CA**
Senior Scientist, Biostatistics, Pacific Biosciences

- Developed R/C tool-chain for in-house statistical analyses. Built open-source *h5r* package for interfacing HDF5 files, *pbh5* package for management and analysis of Pacific Biosciences data formats, and *pbgalaxy*, a plotting toolkit and plugin for Galaxy bioinformatics server
- Designed and implemented custom sorting and indexing scheme for fast interval queries using Python and C
- Designed, planned, and analyzed numerous large-scale experiments for diverse sets of internal and external groups

September 2006 – September 2007, **UC Berkeley, CA**
Graduate Student Researcher, Tager Group

- Developed Deletion/Substitution/Addition statistical learning / prediction algorithm
- Developed *modelTools* R package written in C and R
- Rolled-out Subversion version control software for best-practice software development

December 2004 – September 2005, **UC Berkeley, CA**

Graduate Student Researcher, Center for Infectious Disease Preparedness

- Taught intensive relational database management course for epidemiologists
- Built relational database management online course website; system administration, content creation, course instructor

February 2004 – April 2004, **Buenos Aires, Argentina**

Software Engineer, SONY

- Designed and implemented Java-based *Time Sheet Software*; employed technologies: Hibernate, Struts, and Velocity
- Built image navigation software for photo album management, using SWT and JFace

August 2003 – November 2003, **Temuco, Chile**

Systems Consultant, Conectu Inc.

- Designed and managed release architecture and network topology for Anaconda Web, core product of Conectu Inc.
- Improved operations efficiency by implementing modern programming practices and tools such as, CVS, Bug Track, and project management software

April 1998 – August 2002, **Chicago, IL**

Software Engineer/Architect, Perceptual Robotics Inc.

- Architected J2EE web application using UML and OOA/OOD principles
- Responsible for creating a unified build and release process for distribution on Linux, Solaris, Windows NT/2000, WinCE, and Palm OS
- Designed and implemented multithreaded configuration system using Java and XML for J2EE web applications
- Developed code-generation libraries for build-time creation of RDBMS backed Java Objects using Python
- Responsible for improving programmer efficiency using source control, bug tracking, and unit testing software
- Designed and managed relational schema for varying database platforms, such as Oracle, DB2, and InstantDB
- Developed distributed-load-generation tool in Scheme for quality control of server scalability
- Designed multithreaded-connection-pooling software for high-traffic web applications

April 1998 – June 1999, **Evanston, IL**

Software Developer, Neodesic Corporation.

- Built 'Information Workbench' software for NASA SBIR (Small Business Innovation Research) grant in Visual Basic; composed of 1.) GUI for managing and creating GIS workflow plans, and 2.) Task processing software for dealing with parsed results from natural language queries
- Developed networking protocol for communication between Visual Basic clients and LISP natural language servers
- Programmed Java/J2EE framework for web based natural language 'Ask an Expert System'

TEACHING EXPERIENCE

August 10th, **CSB 2009, Stanford, CA**

Instructor, Tutorial on the Analysis of Next-Generation Sequencing using R/Bioconductor

July 27th, **BioC 2009, Seattle, Washington**
Co-instructor, *Tutorial on GenomeGraphs and biomaRt*

Fall 2008, **UC Berkeley**
Graduate Student Instructor, PH240C *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine*

July 2008, **UC Berkeley**
Main instructor, *R/Statistics* week-long intensive course

January 2008, **Center for Genomic Sciences, University of Cuernavaca**
Main instructor, *R/Bioconductor* week-long intensive course

Fall 2006, **UC Berkeley**
Graduate Student Instructor, PH240C *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine*

August 2005, **Center for Infectious Disease Preparedness, UC Berkeley**
Main instructor, week-long intensive course: *Design and Management of Relational Databases and SQL*

PUBLICATIONS

Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments

James H. Bullard, Elizabeth A. Purdom, Kasper D. Hansen, and Sandrine Dudoit
BMC Bioinformatics, 2010

Polygenic and Directional Regulatory Evolution across Pathways in *Saccharomyces*

James H. Bullard, Yulia Mostovoy, Sandrine Dudoit, and Rachel B. Brem
Proc. Natl. Acad. Sci., 2010

Statistical Inference in mRNA-Seq: Exploratory Data Analysis and Differential Expression

James H. Bullard, Elizabeth A. Purdom, Kasper D. Hansen, Steffen Durinck, and Sandrine Dudoit
Working Paper 247, UC Berkeley Division of Biostatistics Working Paper Series, 2009

GenomeGraphs: Integrated Genomic Data Visualization with R

Steffen Durinck, **James H. Bullard** (joint first author), Paul T. Spellman, and Sandrine Dudoit
BMC Bioinformatics, 2009

Novel Low Abundance and Transient RNAs in Yeast Revealed by Tiling Microarrays and Ultra High-Throughput Sequencing Are Not Conserved Across Closely Related Yeast Species

Albert Lee, Kasper D. Hansen, **James H. Bullard** (joint first author), Sandrine Dudoit, Gavin Sherlock
PLoS Genetics, 2008

aroma.affymetrix: A Generic Framework in R for Analyzing Small to Very Large Affymetrix Data Sets in Bounded Memory

Henrik Bengtsson, Ken Simpson, **James H. Bullard**, Kasper D. Hansen

Berkeley Statistics Technical Report, 2008

Book review of “R Graphics”

James H. Bullard,

Journal of Statistical Software, 2007

Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany

Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, **Bullard J**, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK

New England Journal of Medicine, 2011

The origin of the Haitian cholera outbreak strain

Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, **Bullard J**, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK

New England Journal of Medicine, 2011

SOFTWARE

h5r: R package for reading hdf5 files

Genominator: R package for analysis and management of high-throughput sequencing data

Affxparser: R/Bioconductor package for parsing Affymetrix data files

GenomeGraphs: R/Bioconductor visualization package for genomic data

Panjo: MPI based parallel implementation of neighbor-joining algorithm via distributed memory

Newick Parser: Fast C parser for Newick tree format written using Flex and Bison

DSA/modelTools: R/C packages for data-adaptive regression modeling

GRANTS AND FELLOWSHIPS

NIH Genomics Training Grant

2006 – 2009. UC Berkeley

Reshetko Family Scholarship

2008, UC Berkeley

Bioconductor Scholarship
2007, 2008

2nd Prize Retreat Poster
Designated Emphasis in Computational Biology, UC Berkeley Retreat 2007, Tahoe, CA

INVITED TALKS

Working with Pacific Biosciences Data, September 8th 2011
Statistics and Genomics Seminar, Berkeley, CA

An overview of PacBio Data and Applications, July 20th 2010
Statistical Genomics in Biomedical Research, Banff, CA

Expression Analysis using mRNA-Seq, July 23rd 2009
Weill Medical College Department of Pathology, Cornell NY, NY

mRNA-Seq for Transcriptomics, June 14th 2009
WNAR, Portland OR

Finding Novel RNSs Using High-throughput Sequencing Data, October 20th 2008
Designated Emphasis in Computational and Genomic Biology Retreat, Walnut Creek, CA

The PhyloChip: An Application in Metagenomics, August 3rd 2007
El Aleph, Valencia, ES

COMPUTER SKILLS

Languages: R, Java, C/C++, Python, Bash, Scheme, Scala, Lisp, XML, SQL, HTML

Methodologies: OOA/OOD, UML, MVC, Scrum, Extreme Programming, Waterfall Approach, Design Patterns

Operating Systems: Linux, Solaris X86/Sparc, Windows NT/2000/XP, OSX, PalmOS, WinCE

RDBMS: SQLite, MySQL, Oracle, InstantDB, PostgreSQL

Tools/Software: ANT, Make, CVS, SVN, Tomcat, Apache, Emacs, VI

Languages:

English: native speaker

Spanish: fluent

References:

Available upon request.