

J. Gregory Caporaso
1298 S. Knoles Drive, PO Box 4073
Building 56, 3rd Floor
Flagstaff, AZ 86011-4073, USA
(928) 523-5845 (office)
(928) 523-4015 (fax)
gregcaporaso@gmail.com
www.caporaso.us

Education

Ph.D., May 2009

Dissertation title: *Extracting signal from noise in biological data: Evaluations and applications of text mining and sequence coevolution.*

Mentor: Lawrence Hunter, Ph.D.

Program in Biophysics and Structural Biology

Department of Biochemistry and Molecular Genetics

University of Colorado Denver, Aurora, CO (Formerly U of C Health Sciences Center.)

Bachelor of Arts in Biochemistry, Minor in Chemistry, August 2004

University of Colorado at Boulder, College of Arts and Sciences, Boulder, CO

Bachelor of Science in Computer Science, May 2001

University of Colorado at Boulder, College of Engineering and Applied Science, Boulder, CO

Current Positions

Assistant Professor, 2011 – Present

Department of Computer Science, Northern Arizona University, Flagstaff, AZ, USA.

Assistant Professor, 2012 – Present

Institute for Genomics and Systems Biology, Argonne National Laboratory, Argonne, IL, USA.

Research Affiliate, 2012 – Present

Microbial Genetics and Genomics Center, Northern Arizona University, Flagstaff, AZ, USA.

Awards and Honors

Amazon Web Services Research Grant

January 2011, June 2011; co-PI: Rob Knight

\$62,500 in credit (total for initial grant and two renewals) to deploy Bioinformatics software on the Amazon Elastic Compute Cloud.

Lewis and Clark Fund for Exploration and Field Research in Astrobiology

May, 2011

\$5000 award for hyper-arid soil microbial ecology project in the Atacama Desert, Chile.

Amazon Web Services Education Grant

July 2011; co-PI: John Spear

\$2000 in credit to support student computational work on the Amazon Elastic Compute Cloud at the 2011 International Geobiology Course.

NLM Informatics Pre-Doctoral Training Fellowship

July 2007 - July 2009

Provided full funding, including tuition and stipend, for the final two years of my pre-doctoral training.
Grant number: T15LM009451.

Travel award for Argonne Soils Workshop

Provided registration and travel costs to attend the 2010 Argonne National Labs Soils Workshop, Sept. 2010.

Outstanding Dissertation Award Finalist

University of Colorado Denver, May 2009.

Outstanding Research Award, 23rd Annual Student Research Forum, University of Colorado Denver
Sequence co-occurrence and covariation suggest a model of the Type VI Secretion System, Jan. 2009.

NSF Travel Fellowship, Dec. 2008

Provided registration and travel costs to attend the 2008 Rocky Mountain Bioinformatics Conference.

Outstanding Research Award, 22nd Annual Student Research Forum, University of Colorado Denver
A Comparison of Coevolution Detection Methods on Protein Alpha Helices, Jan. 2008.

Best Presentation, Rocky Mountain Bioinformatics Conference

A Sentence Recognizer for Mutant Protein Structure Studies: Toward Intelligent Systems for the Management of Structural Biology Data, Dec. 2005.

Previous Positions

Research associate, March 2009 – July 2011

Rob Knight Laboratory, Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO.

Research assistant, February 2005 - 2009

Lawrence Hunter Laboratory, Computational Bioscience Program, Department of Pharmacology, University of Colorado Denver, Aurora, CO.

Research assistant (as rotation student), December 2004 - February 2005.

Mark Duncan Laboratory, Biomolecular Structure Program, University of Colorado Health Sciences Center, Aurora, CO.

Research assistant (as rotation student), August 2004 - November 2004.

Robert Garcea Laboratory, Biomolecular Structure Program, University of Colorado Health Sciences Center, Aurora, CO.

Research assistant, January 2003 - August 2004.

Michael Yarus and Rob Knight Laboratories, Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Boulder, CO.

Contract Web Developer, 1999 - 2002.

American Business English Internet School, Boulder, CO.

Software Engineer, 2001 - 2002.

Freshwater Software (acquired by Mercury Interactive), Boulder, CO.

Student Intern, 2000 - 2001.

Freshwater Software, Boulder, CO.

Web Developer, System Administrator, 1998 - 2000.

Web Communications, University of Colorado at Boulder, Boulder, CO.

Teaching

CS486: Senior Capstone, Spring 2012. Mentored a team of five Computer Science seniors as they developed a software package for the biogeographical analysis of microbial communities. These tools were contributed to the QIIME software package in May of 2012 under an open source license.

CS399/Bio399: Introduction to Bioinformatics, Northern Arizona University, Spring 2012.

Designed and taught an undergraduate bioinformatics course for computer science and biology majors.

CS299/Bio299: Introduction to Bioinformatics, Northern Arizona University, Fall 2011.

Designed and taught an undergraduate bioinformatics course for computer science and biology majors.

Quantitative Insights Into Microbial Ecology (QIIME) Workshop

University of Arizona, May 2012, Tucson, AZ, USA. (Eight-hour course.)

Northern Arizona University, January 2012, Flagstaff, AZ, USA. (Five-hour course)

International Geobiology Course, July 2011, Catalina Island, CA, USA. (Seven-day course)

Workshop on Comparative Genomics, July 2011.

Co-taught for two days on the PyCogent and QIIME software packages.

Extracting ecological signal from noise workshop. September 2010.

Co-taught a three-day course on computational tools for microbial ecology.

Molecular and Computational Biology Research School, Bergen, Norway.

Guest Lecturer, Fall semester, 2010; Fall semester 2009; Spring 2007.

Bioinformatics 7711/2, University of Colorado Denver, Aurora, CO.

Guest Lecturer, Spring semester, 2009.

CHEM/MCDB 4621/5621; CSCI 4317/5317, University of Colorado at Boulder, Boulder, CO.

High School Mathematics Tutor (IMP Program), December 2008 - June 2009.

Boulder, CO.

Curriculum Development: Mathematics assignments for undergraduate Biology students, 2003 - 2004.

Department of Human Resources, University of Colorado at Boulder, Boulder, CO.

Students Mentored

Daniel Domogala, Biology Master's student, Northern Arizona University, Spring 2012 - Present.

Jai Rideout, Computer Science Bachelor's student, Northern Arizona University, Fall 2011-Spring 2012.

Recipient of the *Most Promising Undergraduate Researcher* award at NAU in Spring of 2012. This was the first time this was awarded to a Computer Science student at NAU. Two co-publications.

William A. Walters, Molecular Biology Ph.D. student, University of Colorado at Boulder (graduate student in Rob Knight's laboratory), Spring 2009-Fall 2011. Four co-publications.

Anna Lindemann, Computer Science Bachelor's student, University of Colorado Health Science Center, Summer 2006. Two co-publications.

Invited Presentations

Ultra-high-throughput microbial ecology: software, sequencing and practice for studying tens of thousands of environments.

Genomics Standards Consortium 13, March 2012, Shenzhen, China.

Arizona State University, April 2012, Tempe, Arizona, USA.

University of Arizona, February 2012, Tucson, Arizona, USA.

University of Montana, April 2011, Bozeman, Montana, USA.

Procrustes Analysis to Compare Biological Conclusions in Studies of the Human Microbiome

Cloud Computing for the Microbiome, April 2011, Boulder, CO, USA.

Earth Microbiome Project Data Analysis Update

Earth Microbiome Project/National Ecological Observatory Joint Meeting, November 2011, Boulder, CO.

Argonne Soil Metagenomics Meeting Wrap-up Talk

3rd Argonne Soil Metagenomics Meeting, October 2011, Chicago, IL, USA.

Moving Pictures of the Human Microbiome

Infectious Disease Research Network: Microbial Community Profiling Workshop, June 2011, London, UK.

Quantitative Insights Into Microbial Ecology

Infectious Disease Research Network: Microbial Community Profiling Workshop, June 2011, London, UK.

Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.

Genomics Automation Congress, May 2010, Boston, MA, USA.

Los Alamos National Laboratories, September 2010, Los Alamos, NM, USA.

Detecting Coevolution by Disregarding Evolution? Tree-Ignorant Metrics of Coevolution Perform As Well As Tree-Aware Metrics.

Genome Biology Seminars, Australian National University, November 2009, Canberra, Australia.

Invited Interdepartmental Seminars

Intermolecular Coevolution Suggests Interactions Between Type VI Secretion System Proteins.

Bioinformatics Supergroup, University of Colorado at Boulder, April, 2009.

Detecting Coevolution by Disregarding Evolution?

Bioinformatics Supergroup, University of Colorado at Boulder, April, 2008.

Peer-reviewed publications

Human gut microbiome viewed across age and geography

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, **Caporaso JG**, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JL.
Nature, 2012.

The Biological Observation Matrix (BIOM) Format or: How I Learned To Stop Worrying and Love the Ome-ome

McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, **Caporaso JG**.
GigaScience (in press).

Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat

Harris JK*, **Caporaso JG***, Walker JJ, Spear, JR, Gold NJ, Robertson CE, Hugenholtz P, Goodrich J, McDonald D, Knights D, Marshall P, Tufo H, Knight R, Pace NR (*contributed equally).
ISME Journal (in press).

Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.
Caporaso JG, Lauber C, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Betley J, Fraser L, Bauer M, Gormley N, Smith G and Knight R.

ISME Journal (April, 2012).

**Most downloaded article from ISME Journal, April 2012. Selected for review by Faculty of 1000.*

The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites.

Ursell LK, Clemente JC, Rideout JR, Gevers D, **Caporaso JG**, Knight R.

J Allergy Clin Immunol. 2012 May;129(5):1204-8.

The under-recognized dominance of Verrucomicrobia in soil bacterial communities.

Bergmann GT, Bates ST, Eilers KG, Lauber CL, **Caporaso JG**, Walters WA, Knight R, Fierer N.

Soil Biol Biochem. 2011 Jul;43(7):1450-1455.

Sequencing our way towards understanding global eukaryotic biodiversity.

Bik HM, Porazinska DL, Creer S, **Caporaso JG**, Knight R, Thomas WK.

Trends Ecol Evol. 2012 Apr;27(4):233-43.

Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys.

Werner JJ, Zhou D, **Caporaso JG**, Knight R, Angenent LT.

ISME J. 2011

Using QIIME to analyze 16S rRNA gene sequences from microbial communities.

Kuczynski J, Stombaugh J, Walters WA, González A, **Caporaso JG**, Knight R.

Curr Protoc Bioinformatics. 2011

The Western English Channel contains a persistent microbial seed bank.

Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA.

ISME J. 2012 Jun;6(6):1089-93.

TopharyExplorer: visualizing large phylogenetic trees with environmental metadata.

Pirrung M, Kennedy R, **Caporaso JG**, Stombaugh J, Wendel D, Knight R.

Bioinformatics. 2011

Defining seasonal marine microbial community dynamics.

Gilbert JA, Steele JA, **Caporaso JG**, Steinbrück L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I, Somerfield P, Fuhrman JA, Field D.

ISME J. 2011 Aug 18

Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys.

Werner JJ, Koren O, Hugenholtz P, Desantis TZ, Walters WA, **Caporaso JG**, Angenent LT, Knight R, Ley RE.

ISME J. 2011 Jun 30.

Moving Pictures of the Human Microbiome

Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R

Genome Biol. 2011 May 30;12(5):R50.

New insight into the diversity of life's building blocks: evenness, not variance.

Caporaso JG, Knight R.

Astrobiology. 2011 Apr;11(3):197-8

PrimerProspector: de novo design and taxonomic analysis of barcoded PCR primers

Walters WA*, **Caporaso JG***, Lauber CL, Berg-Lyons D, Fierer N Knight R (*contributed equally)

Bioinformatics. 2011 Apr 15;27(8):1159-61.

Host-Associated and Free-Living Phage Communities Differ Profoundly in Phylogenetic Composition

Caporaso JG, Knight R, Kelley ST

PLoS One. 2011 Feb 24;6(2):e16900.

Bacterial communities associated with the lichen symbiosis

Bates ST, Cropsey GW, **Caporaso JG**, Knight R, Fierer N

Appl Environ Microbiol. 2011 Feb;77(4):1309-14

Examining the global distribution of dominant archaeal populations in soil

Bates ST, Berg-Lyons D, **Caporaso JG**, Walters WA, Knight R, Fierer N

ISME J. 2010

Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes

Chu H, Fierer N, Lauber CL, **Caporaso JG**, Knight R, Grogan P

Environmental Microbiology, 2010 Nov;12(11):2998-3006.

Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R

Proc Natl Acad Sci U S A. 2010

Soil bacterial and fungal communities across a pH gradient in an arable soil

Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, **Caporaso JG**, Knight R, Fierer N

International Society for Microbial Ecology Journal, 2010 May.

QIIME allows analysis of high-throughput community sequencing data

Caporaso JG*, Kuczynski J*, Stombaugh J*, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JL, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (*contributed equally).

Nature Methods 2010 May, 7(5):335-6.

PyNAST: a flexible tool for aligning sequences to a template alignment

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R

Bioinformatics 2010, 26(2):266-7.

Detecting Coevolution by Disregarding Evolution? Tree-Ignorant Metrics of Coevolution Perform As Well As Tree-Aware Metrics

Caporaso JG, Smit S, Easton BC, Hunter L, Huttley GA, Knight R

BMC Evolutionary Biology 2008, 8:327.

An Integrated Approach to Concept Recognition in Biomedical Text

Baumgartner WA Jr., Lu Z, Johnson HL, **Caporaso JG**, Paquette J, Lindemann A, White EK, Medvedeva O, Cohen KB, Hunter L

Genome Biology, 2008, Suppl 2:S9.

Intrinsic Evaluation of Text Mining Tools May Not Predict Performance on Realistic Tasks

Caporaso JG, Deshpande N, Fink JL, Bourne PE, Cohen KB, Hunter L
Pacific Symposium on Biocomputing 13:640-651(2008).

Rapid Pattern Development for Concept Recognition Systems: Application to Point Mutations

Caporaso JG, Baumgartner WA Jr., Randolph DA, Cohen KB, Hunter L
Journal of Bioinformatics and Computational Biology, 2007 Dec;5(6):1233-59.

PyCogent: A Toolkit for Making Sense from Sequence

Knight R, Maxwell P, Birmingham A, Carnes J, **Caporaso JG**, Easton BC, Eaton M, Hamady M, Lindsay H, Liu Z, Lozupone C, McDonald D, Robeson M, Sammut R, Smit S, Wakefield MJ, Widmann J, Wikman S, Wilson S, Ying H, Huttley GA
Genome Biology 2007, 8:R171.

MutationFinder: A High-Performance System for Extracting Point Mutation Mentions from Text

Caporaso JG, Baumgartner WA Jr, Randolph DA, Cohen KB, Hunter L
Bioinformatics, 2007 23(14):1862-1865.

Error Minimization and Coding Triplet/Binding Site Associations are Independent Features of the Canonical Genetic Code

Caporaso JG, Yarus M, and Knight R
Journal of Molecular Evolution, 2005.

Origins of the Genetic Code: The Escaped Triplet Theory

Yarus,M, **Caporaso JG**, and Knight R
Annual Review of Biochemistry, 2005.

Invited Publications

Concept Recognition, Information Retrieval, and Machine Learning in Genomics Question Answering

Caporaso JG, Baumgartner WA Jr., Kim H, Lu Z, Johnson HL, Medvedeva O, Lindemann A, Fox LM, White EK, Cohen KB, and Hunter L
TREC 2006 Conference Proceedings, January, 2007.

Concept recognition and the TREC Genomics tasks

Caporaso JG, Baumgartner WA Jr, Cohen KB, Johnson HL, Paquette J, Hunter L
TREC 2005 Conference Proceedings, January, 2006.

Conference Presentations

Ultra-high-throughput microbial ecology: software, sequencing and practice for studying tens of thousands of environments.

American Society for Microbiology (Arizona regional meeting), Tempe, AZ, USA, April, 2012.

American Society for Microbiology (Colorado regional meeting), Boulder, CO, USA, April, 2011.

Recovery of biological soil crust-like microbial communities in previously submerged soils of Glen Canyon.

11th Biennial Conference of Research on the Colorado Plateau, Flagstaff, AZ, USA, September, 2011.

Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample

13th International Symposium on Microbial Ecology (poster presentation), Seattle, WA, USA, August, 2010.

Intermolecular Coevolution Suggests Interactions Between Type VI Secretion System Proteins

Rocky Mountain Bioinformatics Conference (oral and poster presentation), December, 2008.

Rapid Pattern Development for Concept Recognition Systems.

DMAP meeting, April, 2008, Aurora, CO.

Intrinsic Evaluation of Text Mining Tools May Not Predict Performance on Realistic Tasks
Pacific Symposium on Biocomputing (oral presentation), January, 2008.

A Comparison of Coevolution Detection Methods on Protein Alpha Helices
Pacific Symposium on Biocomputing (poster presentation), January, 2008.
Rocky Mountain Bioinformatics Conference (oral and poster presentation), December, 2007.

Recognizing Point Mutations in Text: Techniques and Applications in Biomedical Text Mining
Pacific Symposium on Biocomputing (poster presentation), January, 2007.

Concept Recognition, Information Retrieval, and Machine Learning in Genomics Question Answering
Rocky Mountain Bioinformatics Conference (oral and poster presentation), December, 2006.
Text REtrieval Conference (TREC) 2006 (poster presentation), November, 2006.

A Sentence Recognizer for Mutant Protein Structure Studies: Toward Intelligent Systems for the Management of Structural Biology Data.
Rocky Mountain Bioinformatics Conference (oral and poster presentation), December, 2005.

Concept Recognition and the TREC Genomics Tasks
Text REtrieval Conference (TREC) 2005 (poster presentation), November, 2005.

Stereochemical and Adaptive Factors in Genetic Code Evolution
Evolution 2004 (poster presentation), July, 2004.

Open-source Software Projects

Lead developer, QIIME (Quantitative Insights Into Microbial Ecology)
<http://www.qiime.org> <http://blog.qiime.org> <http://forum.qiime.org>

Developer, PyCogent (Python COmparative GENomics Toolkit)
<http://www.pycogent.org>

Developer, PI-CRUST (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States)
<http://picrust.sourceforge.net>

Developer, BIOM (Biological Observation Matrix) project and file format
<http://biom-format.org>
**A candidate project of the Genomics Standards Consortium, and an Earth Microbiome Project standard.*

Lead developer, PyNAST (Python Nearest Alignment Space Termination tool)
<http://pynast.sourceforge.net> <http://pynast.wordpress.com>

Developer, PrimerProspector
<http://pprospector.sourceforge.net>

Lead developer, MutationFinder (no longer under active development)
<http://mutationfinder.sourceforge.net>

Professional Society Memberships

International Society for Microbial Ecology, since 2010.
International Society for Computational Biology, 2005-2009.

American Association for the Advancement of Science, 2006-2009.

References

Rob Knight, Associate Professor
Howard Hughes Medical Institute
Department of Chemistry and Biochemistry, University of Colorado at Boulder.
rob.knight@colorado.edu; (303) 492-1984

Lawrence Hunter, Professor
Computational Bioscience Program, University of Colorado Denver.
Department of Pharmacology, University of Colorado Denver.
larry.hunter@ucdenver.edu; (303) 724-3574

Noah Fierer, Assistant Professor
Department of Ecology & Evolutionary Biology, University of Colorado at Boulder
Cooperative Institute for Research in Environmental Sciences
noah.fierer@colorado.edu; (303) 492-5615