

Konstantinos Krampis, PhD
Associate Professor, Dept. of Biological Sciences, Hunter College
Director of Bioinformatics, Center for Translational and Basic Research
City University of New York (CUNY)

PROFESSIONAL EXPERIENCE

Hunter College, CUNY, Associate Professor

04/2014 - present

Principal Investigator (PI) of NIH award (\$485K) for building core bioinformatics infrastructure and genomic sequencing facility at Hunter College. Designed a high-performance compute cluster, coordinated hardware acquisition and installation, implemented data production pipeline from the genome sequencer to the cluster. Established a 5-member group performing research on large-scale data analysis infrastructures using cloud computing.

J. Craig Venter Institute, Assistant Professor, Bioinformatics

05/2011 - 04/2014

Principal Investigator (PI) of National Institutes of Health award (\$1.5 million) for deploying next-generation sequencing data analysis pipelines for assembly, annotation and visualization of genomes on Amazon EC2. Directed development of a hybrid bioinformatics infrastructure, consisting of a local cluster and cloud computing. Defined project budget, execution plan, coordinated server and storage acquisition and installation, managing a team of three software developers, one system administrator and two graduate students.

Co-PI of a National Science Foundation award (\$2.5million), for building a scalable data analysis web portal to serve the international *Arabidopsis* genomics community. Developed project execution plan (100+ pages) with infrastructure and software implementation strategy across two USA and one UK institutes. Execution plan also included design, budget, evaluation and acquisition of IT infrastructure within the project, defining agile software development approach, education and outreach plan.

J. Craig Venter Institute, Bioinformatics Cloud Infrastructure Lead

05/2009 – 05/2011

Pioneered application of cloud computing and virtualization technologies in the field of bioinformatics, by founding the Cloud BioLinux consortium in 2009 and implementing the first Amazon Machine Image (AMI) with genome assembly, RNAseq, ChipSeq, and other sequencing data analysis software. Scientific publication from the project was top 3 in yearly downloads for *BMC Bioinformatics*. Established Cloud BioLinux as open source project with hundreds of active developers worldwide, used by the Federal Drug Administration and National Institutes of Health for analysis of the 1000 Human Genome dataset.

Developed data analysis pipeline using the Hadoop/MapReduce distributed computing framework for generating non-redundant sequencing datasets in large-scale genomic databases, reducing compute time from 36 to 2 hours. Studied genome assembly algorithm designs and evaluated performance at different levels of sequencing coverage and cost. Supervised graduate students' research on comparison of traditional versus parallel genome assemblers. Extensive experience with Sun Grid Engine scheduler and MPI-BLAST for high throughput sequence alignment.

Virginia Bioinformatics Institute, Bioinformatics Engineer

08/2003 – 05/2009

Designed statistical algorithms for *in silico* discovery of sequence polymorphisms using microarray hybridization intensities, discovering 5,000 new SNPs and constructing a high-density genetic map. Implemented Terabyte-scale relational database for storing the gene expression data, optimized with indexing scheme to reduce query response times. Developed RDF/XML database for storing metabolic graphs and algorithms using semantic web queries (SPARQL) for computing graph intersections and identifying novel biological pathway modules.

Coded microarray data analysis pipelines and performed NCBI web service programming using a range of scripting languages (Perl, Ruby, Python). Implemented algorithms in procedural SQL language, performed Linux and relational database administration, and gained experience with object-oriented programming and Java.

University of Athens, Department of Molecular Biology, Research Associate

08/2000 – 06/2003

Studied localization of *Arabidopsis* HSP90 proteins in cellular compartments and their developmental role during plant embryo desiccation and seed formation, using fusions of GFP and splice variants of the genes.

EDUCATION

Virginia Tech, Blacksburg, VA: Ph.D., Bioinformatics and Computational Biology, 08/2003 - 05/2009

University of Athens, Athens, Greece: B.S., M.S., Molecular Biology, 09/1997 – 06/2003

RESEARCH FUNDING

PI, Bioinformatics Core, Center for Translational and Basic Research, Hunter College, CUNY “*Development of a Bioinformatics Computing Core and Sequencing Facility at Hunter College*”, 2 years, amount: \$485,424

Co-PI, National Science Foundation - Advances in Biological Informatics (ABI, submitted October 2012, awarded May 2013), “*Development of a next-generation Arabidopsis Informatics Portal*”, 3 years, amount: \$2,434,910

PI, National Institutes of Health - NIAID (submitted October 2010, awarded June 2011), “*Large-Scale Prokaryotic and Viral Next-Gen Sequencing Data Pipelines Using Cloud Computing*”, 3 years, amount: \$1,661,217

AWARDS

Bioinformatics Open Source Conference – Best Audience Voted Presentation Award (2011)

Virginia Tech Graduate Association Executive Board Elected Member (2008)

Virginia Tech Outstanding Interdisciplinary Doctoral Student Award (2008)

Horace E. Alphin Doctoral Tuition Scholarship (2008)

Virginia Tech Graduate Student Assembly Research Grant (2007)

James and Ina Mae Doctoral Tuition Scholarship (2004, 2006)

KEY SKILLS

Teaching skills:

- In-depth knowledge of next-generation sequencing, bioinformatics and computer science subject areas
- Ability to assemble highly-structured course syllabi, that gradually introduce concepts to students
- Delivery of hands-on, interactive student experience, such as coding in class lead by instructor

Technical skills:

- Expert coder in scripting languages, object oriented programming
- Unix / Linux and relational database administration, advanced SQL and procedural PL/SQL
- Online database web service programming (NCBI, EBI, both SOAP and REST APIs)
- Distributed computing frameworks (Hadoop/MapReduce, MPI, Sun Grid Engine)
- In-depth knowledge of cloud computing application using Amazon EC2 and Eucalyptus / OpenStack

WORKSHOPS AND CONFERENCE PRESENTATIONS

Cloud computing technologies genome sequence data analysis, and the Cloud BioLinux platform, presented at:

- *Cloud Computing in Bioinformatics Symposium 2012, Center For Genomic Regulation, Barcelona, Spain*
- *Translational Bioinformatics On The Cloud Conference 2012, New Brunswick, NJ*
- *NIAID Bioinformatics Festival 2012, National Institutes of Health, Bethesda, MD*
- *Genomics Standards Consortium Conference 2011, European Bioinformatics Institute, Hinxton, UK*
- *Cloud Computing Track (Session Chair), Next-Gen Sequencing Data Management 2010, Providence, RI*

“South Africa National Research Network Workshop - Cloud BioLinux for High Performance Data Analysis”, *Center of High Performance Computing, Johannesburg, South Africa (December 8 – 11, 2011)*

“Cloud BioLinux for Bioinformatics Open Source Applications” *ISMB 2011, Vienna, Austria (June 27 – 30, 2011)*

“Empowering Genomics in South Africa Workshop - Cloud Computing Solutions for Genome Sequence Data Analysis”, *National Institutes of Health, University of Limpopo, South Africa (May 28 – June 03, 2011)*

“Safe In the Clouds Workshop: Biologically - Inspired Approaches to Security and Resilience, Scale-Free Metabolic Networks”, *Institute of Human and Machine Cognition, Ocala, FL, USA (April 08 – 10, 2011)*

“From monolithic software to real distributed bioinformatics applications on the cloud”, *Amazon Web Services Workshop on Genomics and Cloud Computing 2010, Seattle, WA*

“Biological pathway graph integration using SPARQL queries with RDF data”, *ISMB / ECCB 2007 International Conference, Vienna, Austria*

“myGrid and ToolBus/PathPort web service interoperability”, *ISMB 2006 International Conference, Fortaleza, Brazil*

SELECTED RESEARCH PUBLICATIONS

Shamsaddini A., Yang P., Johnson E., **Krampis K.**, Simonyan V. and Mazumder M. (2014) *BMC Genomics* 15:918 “Census-based rapid and accurate metagenome taxonomic profiling”.

Cole C, **Krampis K.**, Karagiannis K, Almeida JS, Faison WJ, Motwani M, Wan Q, Golikov A, Pan Y, Simonyan V, Mazumder R. (2014) *BMC Bioinformatics* 15:28 “Non-synonymous variations in cancer and their effects on the human proteome: workflow for NGS data biocuration and proteome-wide analysis of TCGA data.”.

Krampis K., Booth T., Chapman B., Tiwari B., Field D. and Nelson K.E. (2012) *BMC Bioinformatics* 13:42. “Cloud Biolinux: pre-configured and on-demand computing for the genomics community”.

Madupu R., Rogers Yu-H., Rusch D., Miller J., **Krampis K.** and Nelson K.E. (2012) “Microbiomes”. Book Chapter in *Encyclopedia of Molecular Cell Biology and Molecular Medicine*. Published by Wiley-VCH Verlag GmbH & Co. Germany.

Miller J., Rusch D., **Krampis K.**, Tovichgrechko A., Sutton G., Yooseph S. and Nelson K.E. (2012) “Bioinformatics for Genomes and Metagenomes in Microbial Ecology Studies”. Book Chapter in *Infectious Micro-ecology: Theory and Applications*. Published by Wiley-VCH Verlag GmbH & Co. Germany.

Quirino B.F., Barreto C.C., Pappas G.J., Zengler K., **Krampis K.** and Krüger R.H. (2012) “Genomes and Post-Genome Technology” Book Chapter in *Prokaryotes*. Published by Springer GmbH & Co. Germany.

Wang H., Waller L., Tripathy S., St. Martin S.K., Zhou L., **Krampis K.** et al. (2010) *The Plant Genome* 3(1): 23-40. “Discovery of genes underlying soybean QTLs conferring partial resistance to *P. sojae*”.

Zhou L., Mideros S.X., Bao L., Hanlon R., Arredondo F., Tripathy S., **Krampis K.** et al. (2009) *BMC Genomics* 26(10): 49. “Infection and genotype remodel the entire Soybean transcriptome”.

Prassinis C., Haralampidis K., Milioni D., Samakovli D., **Krampis K.** and Hatzopoulos P. (2008) *Plant Molecular Biology* 67(4): 323. “Complexity of Hsp90 in organelle targeting”.

Tyler B.M., Jiang R.H.Y., Zhou L., Tripathy S., Dou D., Torto-Alalibo T., Li H., Mao, Y., Liu B., Vega-Sanchez M., Mideros S.X., Hanlon R., Smith, B.M., **Krampis, K.** et al. (2008) “Functional genomics and bioinformatics of the *Phytophthora sojae* soybean interaction”. Book Chapter in *Genomics of Disease*. Published by Springer Science US, New York.

Krampis, K. (2007) “Semantic Web Bioinformatics: distilling knowledge by mining heterogeneous data”. *VBI e-connections* 2(3): 2.

Tyler B.M., Tripathy S.,..., **Krampis K.**, et al. (2006) *Science* 313(5791): 1261. “*Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis”.

Krampis K., Tyler B.M. and Boore J.L. (2006) *Molecular Plant Microbe Interaction* 19(12): 1329. “Extensive variation in nuclear mitochondrial DNA content between the genomes of *P. sojae* and *P. ramorum*”.

Prasinis C., **Krampis K.**, Samakovli D. and Hatzopoulos P. (2005) *Journal of Experimental Botany* 56(412): 633. “Tight regulation of expression of two *Arabidopsis* cytosolic Hsp90 genes during embryo development”.

JOURNAL REVIEW BOARDS

BMC Bioinformatics; BMC Research Notes; In Silico Biology; Future Generation Computer Systems; BioMed Research; British Journal of Mathem. and Computer Science; Journal of Bioinformatics & Computational Biology