

COMPUTATIONAL SCIENCES INITIATIVE

Activity Report

A University of Nebraska Program of Excellence



2013–2015

bigdata.unl.edu

Welcome to the Computational Sciences Initiative

With unprecedented volumes of data being generated around the globe, it is clear that data science holds the promise of solving the most pressing issues of our times and the need for data scientists across all disciplines is at a critical juncture. Now more than ever, in this interconnected age of technology, data science is the key to unlocking opportunities in a vast number of fields. The Computational Sciences Initiative is a university-wide, faculty-driven program to develop research and educational resources, expertise, and opportunities in data sciences for the life sciences. We collaborate with researchers in multiple STEM disciplines – mathematics, statistics, computer science and engineering, bioinformatics, biological systems engineering, electrical engineering - to establish cross-campus linkages via data science. These linkages enable transdisciplinary research, training and consulting with partners in the life sciences from academia, industry, and government. The main objective of the initiative is to conduct and coordinate data science research and training in the life sciences, towards the advancement of knowledge of genes, complex cell

processes, and ecosystems. We also support the application of this knowledge to sustainable agriculture and improved health and well-being. Progress towards this objective can take many forms, from methods for data processing, transfer, and storage to novel analyses of large, complex datasets. We work with the Holland Computing Center and faculty across campus to facilitate hardware/software acquisition and access and providing advice on modeling and predictive analytical approaches to data. In addition the CSI facilitates the hiring of new faculty members with advanced quantitative and computational expertise. With our objective and our new doctoral program in Complex Biosystems, the initiative is positioned to have a direct impact on the life and quantitative sciences at UNL. This is an exciting time for students and faculty with interests in the future of the life sciences. We hope you will join us!

THE CSI IS SUPPORTED BY A UNIVERSITY OF NEBRASKA PROGRAM OF EXCELLENCE AWARD.

CSI Members



Jennifer Clarke, Ph.D.
**DIRECTOR, ASSOCIATE PROFESSOR FOOD SCIENCE OF
STATISTICS AND FOOD SCIENCE AND TECHNOLOGY**

B.A., Psychology, Skidmore College; B.A., Mathematics, Skidmore College; M.S., Statistics, Carnegie Mellon University, Ph. D., Statistics, The Pennsylvania State University

Dr. Clarke's research interests encompass statistical methodology (with an emphasis on high dimensional and predictive methods), statistical computation, bioinformatics/computational biology, multi-type data analysis, data mining/machine learning, and bacterial genomics/metagenomics.



Lisa Lightner
BUSINESS ASSOCIATE

Associate's Degree, Environmental Technology, SECC; B.A., Journalism, University of Nebraska-Lincoln



Dmitri Fomenko, Ph.D.
**RESEARCH ASSOCIATE, BIOCHEMISTRY & REDOX
BIOLOGY CENTER**

M.S. Biotechnology, Academy of Fine Chemical Technology, Russia; Ph.D., Molecular Biology, Institute of Molecular Genetics, Russian Academy of Sciences

Dr. Fomenko's research interests include the set of ER thiol oxidoreductases and selenoproteins with unknown function, which are associated with various disorders (cancer, mental retardation, cataract, diabetes, etc.) and the roles and mechanism of hydrogen peroxide mediated signaling.



Sanjay Anthony Babu, Ph.D.
POST-DOCTORAL RESEARCH ASSOCIATE, BIOINFORMATICS

BSc, Microbiology, University of Madras; MSc, Biotechnology, Newcastle University; PhD, Microbiology, Newcastle University

Dr. Babu's research involves using metagenomics tools to study diversity of microbes in terms of their taxonomy and function, specifically, the roles of microbes in host gut functions and their impacts on the environment.

UNL Oversight

Administration

Steve Goddard, Associate Vice Chancellor of Research; Computer Science and Engineering

Archie Clutter, Dean/Director, Agricultural Research Division; Animal Science

Rolando Flores, Chair; Director, Food Science and Technology

Bertrand Clarke, Chair; Department of Statistics

Brian Larkins, Associate Vice Chancellor

Kathy Ellis, Director, Business Center, FDST

Faculty

Stephen Scott, Professor and Vice Chair, Computer Science and Engineering

David Swanson, Director, Holland Computing Center; Research Associate Professor, Computer Science and Engineering

Stephen Kachman, Professor of Statistics

George Avalos, Professor of Mathematics

Etsuko Moriyama, Associate Professor, School of Biological Sciences; Center for Plant Science Innovation

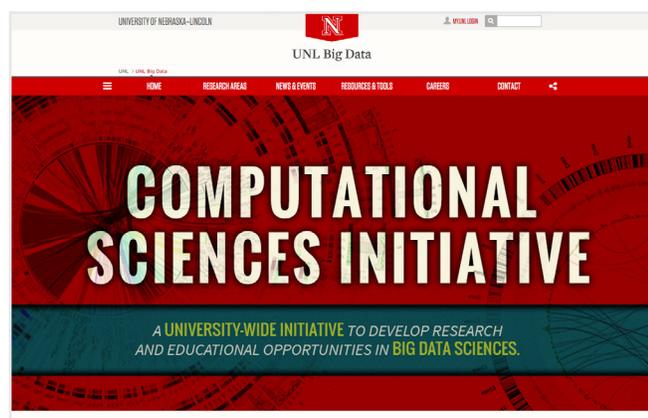
Heriberto Cerutti, Professor, Biotechnology; Professor, School of Biological Sciences

Andrew Benson, W.W. Marshall Professor, Food Science & Technology

Khalid Sayood, Professor, Electrical Engineering

CSI on the web

For current information about Computational Science Initiative, please visit our website at bigdata.unl.edu.



Partner Organizations

The CSI has developed initial partnerships with both academic and industrial organizations with shared interests in data science and Big Data in scientific research and education. These partners share their domain knowledge and expertise, and the challenges they face engaging the talent they require to continue their success: **Li-Cor, Dupont Pioneer, Valent Biosciences, Dow Agrosciences, Quantified Ag, University of North Dakota, Great Plains Network and Greater Western Library Alliance.**



Education

Complex Biosystems Degree Program

This doctoral degree program is designed to prepare students for both academic and industrial careers. It is directed toward students interested in applying quantitative statistical and computational approaches to data acquisition and analysis in multiple life sciences fields including human health and medical biochemistry, plant breeding and phenomics, microbiology and virology, and ecology, evolutionary and organismal biology. Graduates in this program will have rigorous training in the application and underlying concepts of quantitative research methods, critical literature evaluation, grantsmanship, and statistics. Through this program, students will have exposure to teaching in the life sciences and mentored laboratory research at the cutting edge of the discipline.

Microbial Interactions Specialization

Provides students with an understanding of ecological principles that drive assemblages of microbes (the microbiome) that occupy various niches in and on mammalian hosts; how microbial assemblages relate to health, performance, and disease susceptibility of the host; the host factors (immune system, host genetics) that influence microbial ecosystems; the role of diet in shaping microbial communities; the and applications of quantitative data sciences to analyze microbiome data and complex data sets.

Integrated Plant Biology Specialization

Enables students to master the skills and concepts necessary to formulate and experimentally test new hypotheses about relationships between genes, organisms, and ecosystems; demonstrate competence, proficiency and breadth in quantitative plant research; apply theories and methodologies related to plant biological systems; and conduct and provide leadership in basic and applied research in plant biology.

Pathobiology and Biomedical Science Specialization

Enables students to formulate and experimentally test new hypotheses about normal and pathogenic processes; demonstrate competence in quantitative

biology; apply theories and methodologies related to biological systems; and conduct and provide leadership in basic and applied research.

Systems Analysis Specialization

Involves students and researchers in the areas of bioinformatics and computational biology; genetics and genomics; and systems biology. We define bioinformatics as investigating questions about biology and evolution using primarily informatics and statistics; and computational biology as investigating questions about biology and evolution using primarily mathematical modeling and computational simulation. Researchers in genetics and genomics apply bioinformatics and computational techniques using data from current and novel biotechnological methods. Systems biology seeks to explain how biological units (genes, pathways, cells, tissues, organisms) function from the perspective of a biological system.

Computational Organismal Biology, Ecology and Evolution (COBEE)

Trains students with the objective of using combined computational and empirical approaches at the interface of ecology, evolution and organismal biology with the aim of modeling how processes at different scales of biology – from individual to populations to communities and ecosystems – impact each other. The goals of this specialization are to educate the next generation of leaders working to integrate organismal biology with higher levels of ecological organization; develop expertise in novel empirical and quantitative approaches for scaling biological processes from individuals to ecosystems; enable trainees to conduct research that addresses diverse and critical quantitative challenges in organismal biology, ecology and evolution.



For more information about the graduate studies program, please visit our website at bigdata.unl.edu/phd-program.

NSF–NRT Data Science in Sustainable Plant Ecosystems

The proposed National Science Foundation (NSF) National Research Traineeship (NRT) program will increase scientific understanding of sustainable, complex plant systems, and provide unique cross-disciplinary training for graduate students in the earth, agronomic, and data sciences. Students will obtain a deep understanding of how sustainability is generated from complex relationships between plants, microbes, pests and pathogens, and the advanced quantitative and computational methods for exploring and characterizing such relationships. The design of this program incorporates scalability and dissemination of educational advances as a primary goal, using modern technologies and teaching

practices to benefit the greater population of STEM graduate students. In addition to exposing NRT students to inter- and trans- disciplinary academic curricula, through internships and externships NRT students will participate in real-world applications that encourage knowledge transfer and acquisition. Through key strategic partnerships with external agencies the NRT program will prepare individuals to meet the needs of regional, national, and international groups involved in sustainable agriculture and food production. Graduates of this program will be trained and mentored in the sciences, the analytical methods and the computational approaches that can help elucidate the problems they face.

Student Involvement



Nirosh Aluthge
Ph.D. student in Food Science and Technology

His master's work involves characterizing the fecal bacterial communities of beef cattle shedding Shiga toxin-producing E. coli (STEC). His research requires the use of many different data analysis methods in data sciences in carrying out successful bioinformatics projects. His future projects will not only characterize the bacterial communities, but also determine their functional capabilities. Advisor: Samodha Fernando, ANSI.



Nate Korth
Undergraduate in Food Science and Technology, with minors in Music Technology and Mathematics, UCARE participant

Nate is working on Gluten Transfer to Gluten-Free Foods in a Restaurant Setting. His project includes several trials of traditional and gluten free pastas cooked in the same water in separate baskets testing for the presence of gluten. With this data, he will analyze gluten transfers and its significance. Advisors: Melanie Downs, FDST and Jennifer Clarke, STAT and FDST.



Kevin Beatty
Graduate student in Statistics with an undergraduate degree in Electrical Engineering

His academic interests include machine learning, data mining, and computational biology. With over 10 years of experience developing software with applications including computer vision and feature detection, time series analysis, database development, and analysis of assays in biological research, he currently works at Li-Cor Biosciences.



Regan Jones
Undergraduate in Communication Studies with minors in Statistics and Mathematics, Honors Program

Her research involves analyzing original data in order to develop evidence-based solutions to real-world challenges. As a research assistant at the University of Chicago (summer 2015) she contributed to a project that uses natural language processing to analyze the development of stock exchanges over time, as well as the implementation of the Dodd Frank Wall Street Reform and Consumer Protection Act. Regan has presented her findings to the University of Chicago Knowledge Lab, and will use some of the data to develop her B.A. thesis on the impact of media markets on the policy-making process. She has also volunteered for the Nebraska Transportation Center and assisted with data compilation.



Jessica Burow
Ph.D. student in Statistics with a minor in Sociology

As a data scientist, she uses computational analysis to develop predictive models that reveal diffusion of collective social behaviors over time. Her work enhances inferential methods to describe information flow throughout a social network with empirical sampling and simulations. The breakthrough of these findings will enable social scientists to make inferences about an entire network through representative sampling. Advisors: Jeffrey Smith, Sociology and Jennifer Clarke, STAT and FDST.

Broadening Participation, Increasing Diversity

WOMEN IN STEM-UNL

Women in STEM is an undergraduate student group supporting the recruitment, retention and success of students and faculty in science, technology, engineering and mathematics through women-centered support groups, seminars, and workshops.



BRAID: BUILDING, RECRUITING, AND INCLUSION FOR DIVERSITY

The BRAID Initiative, led by the Anita Borg Institute, is a partnership of academic institutions across the nation to increase the percentage of women and students of color majoring in computer science. UNL is proud to be a participating institution.

Our undergraduate computer science program is one of 15 across the United States that have joined this nationwide initiative to recruit more women and underrepresented minorities to the computer science field.

BROADER IMPACTS NETWORK

The CSI has partnered with the Broader Impacts Network (BIN) at the University of Missouri in exploring new ways to connect resources and researchers in data science through broader impacts. The BIN was created in response to the National Science Foundation Vision Statement and criteria for broader impacts in research, defined as the potential to benefit society and to contribute to the achievement of specific, desired societal outcomes.

University of Missouri – broaderimpacts.missouri.edu/broader-impacts/

Curricula

Informatics

The Informatics Undergraduate Minor is designed to provide UNL students with core computational skill sets and competencies that will allow them to solve problems within their chosen discipline or field. This minor will build interdisciplinary problem solving skills that are applicable and advantageous across academia and within industry. It draws on a variety of fields in the humanities, journalism, education, fine and performing arts, and computational sciences for elective credit.

Curriculum Retreat April 2014

The purpose of this retreat was to discuss the need for an undergraduate major in data science at UNL. A group of representatives from various administrative units at UNL gave brief presentations on the relevance of data science to their disciplines and how an undergraduate major could serve their students, faculty, and external stakeholders. The retreat included a discussion of similar educational offerings at other universities in the Big Ten, how a major at UNL would complement these programs, and what resources and expertise are required to design and implement such a major.

Jennifer Clarke, Director, Computational Sciences Initiative

Ellen Weissinger, Senior Vice Chancellor, UNL

Donde Plowman, Dean, College of Business Administration, UNL

Gary Kebbel, Professor, College of Journalism and Mass Communications, UNL

Archie Clutter, Dean, Agricultural Research Division, IANR, UNL

Etsuko Moriyama, Associate Professor, School of Biological Sciences, Center for Plant Science Innovation

Timothy Wei, Dean, School of Engineering, UNL

David Keck, Director, Raikes School, UNL

Keenan Amundsen, Assistant Professor, Department of Agronomy and Horticulture, College of Agricultural Sciences and Natural Resources

LeenKiat Soh, Associate Professor, Computer Science and Engineering, UNL

Matthew Dwyer, Chair, Computer Science and Engineering, UNL



For more information about the Informatics Undergraduate Minor:
bulletin.unl.edu/undergraduate/major/Informatics+Minor.

Outreach & Engagement

Online: bigdata.unl.edu

Research Areas, News & Events, Resources & Tools, and Careers



Quarterly Electronic Newsletter

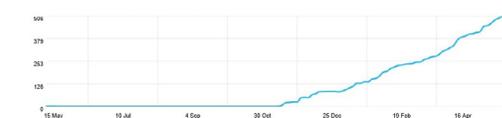
Current news, Big Red Ideas, Big News, Big Events, Big Opportunities, Big Resources



Twitter: [@UNLBIGData](https://twitter.com/UNLBIGData)



506 Followers (2013-2015)



38.3K Impressions (April-May 2015)

Seminars

Clarke, J. The Challenge of Big Data. USDA-ARS Big Data Workshop, George Washington Carver Center, Beltsville, MD. July 2015

Clarke, J. Workshop on Big Data, Annual Meeting of the Snake River Chapter of the American Statistical Association, Idaho State University, Meridian, ID. May 2015.

van der Heijden, Gerie, Dupont Pioneer. Phenotyping for transgene discovery and characterization in corn. Special Invited Seminar, UNL, March 2015.

Sinha, R., Clarke, J., and Benson, A. Functional profiling of metagenomic data using inherent structural properties of peptides from short reads. UNL Symposium on the Future of Big Data, November 2014.

Hesterberg, Tim, Google Inc. Bootstrapping for Learning Statistics. Special Seminar, Department of Statistics, UNL, November 2014

Clarke, J. From STEM to Branch (and maybe Tree). Keynote Address, Annual Meeting of UNL Women in STEM, August 2014.

Clarke, J. The New Reality of Big Data. Invited Seminar, Li-Cor, Inc., June 2014.

Clarke, J. The Computational Sciences Initiative at UNL. Joint UNL and USMARC Workshop on Livestock Metagenomics, Nebraska Innovation Campus, May 2014.

Clarke, J. A Current Perspective of Big Data and Statistical Prediction. Invited Seminar, Division of Biostatistics, University of Miami, FL, January 2014.

Clarke, J. What's the big deal with Big Data? Special Chemical Biology Seminar, Department of Chemistry, UNL, November 2013.

Cohen, Paul, University of Arizona and DARPA. The Information School: Polymathy and the Liberal Arts. Seminar, UNL, November 2013.

The Future of Big Data Symposium 2014

Hosted by the Nebraska Innovation Campus Conference Center, in November 2014, the event was welcomed by Jennifer Clarke, Director, Computational Sciences Initiative; Steve Goddard, Associate Vice Chancellor for Research; Archie Clutter, Dean, Agricultural Research Division, IANR; and Dan Hoyt, Associate Dean for Faculty, College of Arts & Sciences. This symposium brought together the UNL community and representatives from federal agencies, industry, and academia to explore the challenges and opportunities that come with our abilities to collect, store, and analyze ever growing collections of data. The event also featured a poster session and competition for graduate students and postdoctoral researchers from across the NU system whose research involves big data and data science.



Philip E. Bourne
NATIONAL INSTITUTES OF HEALTH
NIH as a Digital Enterprise



Adina Howe
ARGONNE NATIONAL LABORATORY, DOE
Riding the Data Tidal Wave in Microbiology



Natalie de Leon
UNIVERSITY OF WISCONSIN
Utilization of High-Throughput Phenotyping Tools for Plant Improvement



Iam Foster
ARGONNE NATIONAL LABORATORY, DOE
Big Data at DOE



Adam Glynn
EMORY UNIVERSITY
Challenges of Big Data in the Social Sciences



Tim Hesterberg
GOOGLE
Statistics and Big Data at Google



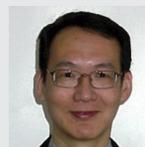
Heidi Imker
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN
Big Data in the Big Ten: Research Data at Universities



Konstantin Kashin
HARVARD
Challenges of Big Data in the Social Sciences (Co-author)



Jeffrey Gerard
THE CLIMATE CORPORATION
Data Science in Agriculture



Tsengdar Lee
NASA
The Challenges in Earth Science Informatics



Carl Lundstedt
UNIVERSITY OF NEBRASKA-LINCOLN
Big Data and High Energy Physics at the LHC



Todd Mockler
DANFORTH CENTER
High-Throughput Plant Phenotyping, Omics and Analytics



Henry Neeman
UNIVERSITY OF OKLAHOMA
Why Storage for Big Data is Hard



Valinda Scarbro Kennedy
BM
Analytics, Big Data, Cognitive Computing and More for a Better Tomorrow



Jerry Roell
JOHN DEERE
Improving Performance, Uptime and Efficiencies of Machines through the use of Data



Marc Smith
SOCIAL MEDIA RESEARCH FOUNDATION
Charting Collections of Connections in Social Media: Creating Maps and Measures with NodeXL



George Strawn
NETWORKING AND INFORMATION TECHNOLOGY RESEARCH & DEVELOPMENT
Big Data in Context



Jennifer Thoegersen
UNIVERSITY OF NEBRASKA-LINCOLN
Data Management Services at UNL Libraries



Fen Zhao
NATIONAL SCIENCE FOUNDATION
Data Science Capacity and Infrastructure: An NSF Priority

Faculty Involvement

The caliber of CSI rests on the involvement of talented and motivated members of the faculty. Over the past two years we have participated in the hiring of new faculty members with expertise in data and computation for the life sciences. We support faculty collaborations and communication through regular meetings of the Big Data Consortium, a group of faculty and students whose research involves life science data.

New Hires 2013–2015



Juan Cui
ASSISTANT PROFESSOR, COMPUTER SCIENCE AND ENGINEERING

Dr. Cui's research interests include biomedical informatics and computational system biology, specifically focusing on cancer genomics, proteomics, metabolomics and utilizing computational prediction and modeling to address important questions related to cancer diagnosis and mechanism studies.



Adam Larios
ASSISTANT PROFESSOR, MATHEMATICS

Dr. Larios' interests lie in partial differential equations, mathematical analysis, and numerical analysis with a focus on mathematical fluid dynamics.



Hasan Otu
PROFESSOR, ELECTRICAL ENGINEERING

Dr. Otu's areas of research include bioinformatics, systems biology, bayesian networks and genetic variations.



Massimiliano Pierobon
ASSISTANT PROFESSOR, COMPUTER SCIENCE AND ENGINEERING

Dr. Massimiliano's research interests are in molecular communication theory for nanonetworks, communication engineering applied to intelligent drug delivery systems, and biological circuit network engineering.



Sydney Everhart
ASSISTANT PROFESSOR, PLANT PATHOLOGY

Dr. Everhart's research interests include epidemiology, population genetics/genomics, and experimental approaches.



Tomas Helikar
ASSISTANT PROFESSOR, BIOCHEMISTRY

Dr. Helikar's research expertise is in computational systems biology, technology development, and STEM Education.



James Schnable
ASSISTANT PROFESSOR, AGRONOMY & HORTICULTURE

Dr. Schnable's research interests are comparative grain crop genomics; high temporal resolution phenotyping technologies; and gene regulation in response to abiotic stresses.



Bing Wang
ASSISTANT PROFESSOR, FOOD SCIENCE AND TECHNOLOGY

Dr. Wang's primary research interests center around human health risk assessment, epidemiology and research synthesis methodologies.



Yumou Qiu
ASSISTANT PROFESSOR, STATISTICS

Dr. Qiu's primary research focuses on high-dimensional statistical inference and its application in genetic analysis.



Qi Zhang
ASSISTANT PROFESSOR, STATISTICS

Dr. Zhang's research expertise is in statistical approaches to genomic data analysis, with a specific interest in CNVs and epigenetics.



Yuzhen Zhou
ASSISTANT PROFESSOR, STATISTICS

Dr. Zhou conducts research in novel statistical approaches for Bayesian spatial-temporal analysis, with environmental applications.

Big Data Consortium

The big data consortium is a campus-wide group of faculty, postdoctoral researchers, and students devoted to collegial exploration of computational and analytical development and research, spanning the academic fields of mathematics, statistics, computational sciences, genetics and genomics, and bioinformatics. The consortium currently has over seventy members and meets monthly to share recent developments in data science and spawn new collaborative efforts in research and education.

The ORED 2014 Interdisciplinary Research Retreat featured an area of emphasis on Big Data in research and education,

supported and organized by the Big Data Consortium in collaboration with David Swanson, Director, HCC; and Andrew K. Benson, Professor, Department of Food Science and Technology. This area featured a series of presentations by UNL faculty and their collaborators to highlight their research and experience in working with Big Data. Key speakers included Ian Fisk from Fermi National Accelerator Laboratory, William G. Thomas III from the Department of History, and Eric Lyons from the iPlant Collaborative and the University of Arizona. Presentations were followed by discussion of resources for Big Data research both locally and nationally and the skill set required to use Big Data to further scientific research.



Research

Consortium for Integrated Translational Biology

Mission Statement

Consortium for Integrated Translational Biology

The mission of the Consortium for Integrated Translational Biology (CITB) is to create a transdisciplinary environment to address the genotype to phenotype gap in plant science through the development of integrative predictive models for the selection of valuable traits that address yield, protection of yield, and quality traits across plant species used for food, feed, industrial applications and landscape resource. The CITB is directed towards facilitating translation of innovations to the field.

Modular Design

Plant Phenotyping Module

The plant phenotyping module is designed for data capture on an individual scale. Key infrastructure elements include updated greenhouse and growth chamber capacity, along with the installation and operational oversight of the Lemnatec Scanalyzer HTS and Scanalyzer 3D imaging systems. The Scanalyzer imaging systems will be complemented by coordination of hand-held devices for data capture on carbon/water flux, photosynthetic capacity, leaf area and soil parameters. Key to success of this module is computational and modeling expertise to facilitate bridging the plant genotype to phenotype gap.

Field Phenotyping Module

A unique element that distinguishes UNL from other public and private sector institutions investing in phenotyping platforms is our stellar history of canopy level image capture and analyses. UNL's School of Natural Resources (SNR) and its aligned Centers, including the Center for Advanced Land Management Information

Technologies (CALMIT), National Drought Mitigation Center (NDMC) and High Plains Climate Center (HRRCC), collectively offer a suite of expertise that will greatly aid in addressing the CITB's mission. The Field phenotyping module will help facilitate integration of genomic selection and genome wide association strategies being utilized by our breeding programs for trait predictions in collaboration with SNR areas of expertise.

End-use Phenotype module

Genomic selection for output traits including oil, protein or starch deposition, functionality and nutritional quality improvements require field phenotyping to ensure agronomic performance is not compromised. Infrastructure to end-use phenotype for such traits is critical to ensure the accuracy of the genotype to phenotype prediction model. To this end, the C-ITB will strive to create a network of expertise to address traits targeted for food, feed (for both farm animals and companion pets), and industrial applications.

Units

School of Natural Resources
 Department of Agronomy and Horticulture
 Nebraska Center for Plant Science Innovation (PSI)
 Center for Advanced Land Management (CALMIT)
 Department of Biological Systems Engineering
 Department of Statistics
 Department of Computer Science and Engineering

Center for Integrated Biomolecular Communication

The Center for Integrated Biomolecular Communication (CIBC), led by James Takacs (Chemistry) and Concetta DiRusso (Biochemistry), serves to integrate the research activities of chemists, biochemists, engineers, and bioinformaticians to address critical knowledge gaps in our understanding of how cells communicate and integrate metabolic and regulatory pathways relevant to disease development and progression. CIBC investigators bring together unique expertise in the understanding of complex diseases and in the development of novel methodologies and technologies to probe communication pathways within and between cells and tissues.

The goals of the CIBC align with those of the Centers of Biomedical Research Excellence (COBRE) program at the National Institutes of Health. The objective of the COBRE initiative is to strengthen an institution's biomedical research

infrastructure through the establishment of a thematic multi-disciplinary center and to enhance the ability of investigators to compete independently for complementary National Institutes of Health (NIH) individual research grants or other external peer-reviewed support.

(Info from: unlcms.unl.edu/cas/chemistry/integrative-biomolecular-communication/)

The CSI and its members serve as Directors of the Data Management and Analysis Core (DMAC) for the CIBC. The long-term objective of the DMAC is to establish a critical, sustainable resource for bioinformatics and analytical support, while its near-term goal is to develop and implement hardware, software, and expertise in high-throughput bio/chemical data, from multi-type data management to integrative statistical analyses.

NSF Big Data Hub

Accelerating the Big Data Innovation Ecosystem

In March 2012, the National Science Foundation announced the National Big Data Research and Development Initiative, which aims to solve some of the Nation's most pressing R&D challenges related to extracting knowledge and insights from large, complex collections of digital data. As part of the initiative, federal agencies, private industry, academia, state and local governments, non-profits, and foundations develop and participate in Big Data research and infrastructure development; education and workforce development; and multi-disciplinary collaborative teams and communities that address complex science and engineering grand challenges. To augment ongoing activities and to ignite new Big Data public-private partnerships across the Nation, the NSF Directorate for Computer and Information Science and Engineering is working to establish a National Network of Big Data Regional Innovation Hubs (BD Hubs). Four Hubs are being created across distinct geographic regions of the United States, representing the Northeast, Midwest, South, and West. Each Hub is a consortium of members that focus on key Big Data challenges and opportunities for its region of service (referred to as spokes and rings) with an aim to support the breadth of interested local stakeholders (referred to as the nodes) that would not be possible for the independent members to achieve alone. The Principal Investigator team for the Midwest hub consists of representatives from the University of Illinois Urbana-Champaign/NCSA, the University of Michigan, the University of North Dakota, the University of Indiana, and Iowa State University.

The Steering Committee consists of representatives from institutions that represent regional interests and specific subject matter areas in Big Data. Members of the Steering Committee include representatives from the University of Nebraska Lincoln (CSI), the University of Kansas, Kansas State University, the University of Missouri, the University of Iowa, the University of Cincinnati, Wayne State University, Argonne National Labs, UI Labs, and the University of Chicago.

The University of Nebraska, with interests in Big Data relevant to agriculture and natural resources, physics, earth and atmospheric sciences, and data sciences, and its stakeholders are faced with the challenge of using Big Data to increase agricultural production. This challenge involves Big Data of several types, including imaging (remote sensing, plant phenotyping), cellular/genomic (high throughput sequencing), and autogenic (real time sensing and data collection). The large amounts of available data and their complexity require that regional partners agree on standards for data sharing, data access, and data stewardship, and provide shared resources and educational opportunities to the community. As a land-grant university in the Midwest region, UNL has a long history of extension and outreach, including partnerships with other institutions for both education and research. As such, UNL and CSI are key participants in several rings and spokes of the Midwest Hub, including Digital Agriculture and Data Science. Specific industry partners involved in the early stages of Hub development in Nebraska include Li-Cor and the Nebraska Public Power District.

Bioinformatics Core

Center for Biotechnology Bioinformatics Core Research Facility

OBJECTIVES

The Bioinformatics Core Research Facility (BCRF) fosters cutting-edge, interdisciplinary, cross-campus and inter-institutional research in Bioinformatics and Computational Biology. Its efforts are focused on collaborations that are strategic to UNL, including emerging diseases, system biology, energy, environmental studies, virology, and translational medicine. The BCRF provides scientific expertise and core infrastructure support as well as education and training. Their expertise helps to integrate bioinformatics and biological sciences to facilitate efficient and effective research.

Jean-Jack Reithoven, Ph.D., Director, Bioinformatics Core
Fangrui Ma, Ph.D., Bioinformatician/Computational Biologist

Stephie Canny, M.S., Bioinformatician/Computational Biologist,
Seong-il Euyun, Ph.D., Bioinformatician/Computational Biologist

The BCRF is a core facility within the Center for Biotechnology. The Center provides state of the art technologies for research through core facilities, enriches the research environment at UNL through a weekly seminar series, provides training to faculty, students and staff through workshops and short courses, and awards student fellowships and scholarships. The Center is directed by Daniel Schachtman, Professor, Department of Agronomy & Horticulture and Center for Plant Science Innovation.

For more information, go to biotech.unl.edu.

Publications

Sinha R, Clarke J, and Benson A (2015). Alignment behavior of short peptides provide a roadmap for functional profiling of metagenomic data. *Genome Biology*, submitted.

Q&A with UNL's Director of Computational Sciences Initiative. *AMSTAT News*, Issue 453, 31-34, March 2015

Eyun S, Wang H, Pauchet Y, French-Constant RH, Benson AK, Arnubio VJ, Moriyama EN, Siegfried BD (2014) Molecular Evolution of Glycoside Hydrolase Genes in the Western Corn Rootworm (*Diabrotica virgifera virgifera*). *PLoS ONE* 9(4):e94052

Pashai A, Yi X, Xia M, Canny S, Riethoven J-J M., Moreau R. Characterization of genome-wide transcriptional changes in liver and adipose tissues of ZDF (fa/fa) rats fed R- lipoic acid by next-generation sequencing. *Physiological Genomics* Published 1 December 2013 Vol. 45 no. 23, 1136-1143 DOI: 10.1152/physiolgenomics.00138.2013

Ding Y, Virlovet L, Liu N, Riethoven JJM, Fromm ME, and Avramova ZV (2014). Dehydration stress memory genes of *Zea mays*; comparison with *Arabidopsis thaliana*. *BMC Plant Biology* 05/2014; 14(1):141. DOI: 10.1186/1471-2229-14-141. [S. Canny was acknowledged in this publication]

Tatineni, S., McMechan, A. J., Wosula, E. N., Wegulo, S. N., Graybosch, R. A., French, R., and Hein, G. L. An eriophyid mite-transmitted plant virus contains eight genomic RNA segments with unusual heterogeneity in the nucleocapsid protein. *Journal of Virology* 88: 01901-14. 2014 [S. Eyun was acknowledged in this publication]

Future

Events

The Computational Sciences Initiative will continue to develop and acquire expertise and resources for the quantitative and computational analysis of life sciences data. Priority areas are identified in cooperation with UNL oversight provided by our faculty advisory committee, the Associate Vice Chancellor for Life Sciences, the Office of Research and Economic Development, the Agricultural Research Division, the Center for Biotechnology, our external partners, student participants, and members of the Big Data Consortium.

UNL Center for Biotechnology Symposium on Systems Biology, October 20, 2015

Seminar Series on Complex Biosystems, Fall 2015/Spring 2016

Biotechnology/Life Sciences Seminar: John B. Hogenesh, Professor, Department of Pharmacology, University of Pennsylvania (September 30, 2015)

Biotechnology/Life Sciences Seminar: Dan Nettleton, Laurence H. Baker Endowed Chair Distinguished Professor, Department of Statistics, Iowa State University (October 28, 2015)

UNL Datathon, East Campus Union, Spring 2016

CSI Goals

Raise awareness and use of the Bioinformatics Core through seed grants/vouchers, workshops, and LIMS system access

Provide education and resources for reproducible research with UNL libraries and the Holland Computing Center

Support additional faculty and postdoctoral hires

Develop an undergraduate program in data sciences in cooperation with other campus units

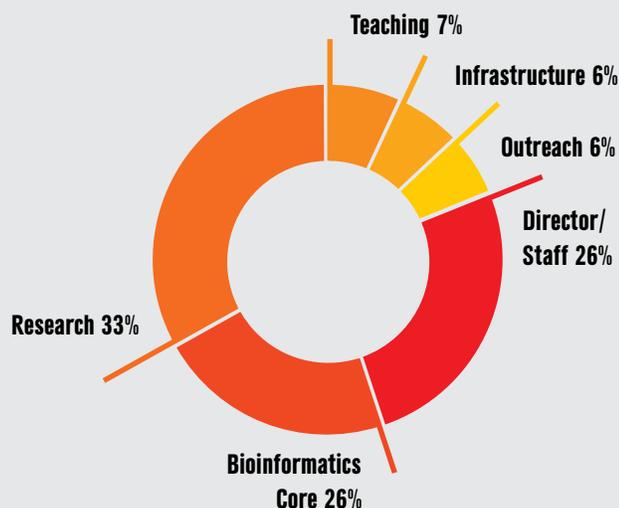
Provide life science researchers with analytical and computational expertise for external proposal development

Timeline

	Fall 2015	Spring 2016	Summer 2016	Fall 2016	Spring 2017	Summer 2017	Fall 2017	Spring 2018	Summer 2018
1. Seed grant/ voucher program	█						█		
2. LIMS system		█	█						
3. Reproducible research		█	█	█	█				
4. Additional faculty and postdoctoral hires	█		█			█			
5. Undergraduate data sciences		█	█	█	█	█	█	█	█
6. External proposal development		█			█			█	

Financials

2016 Budget



Proposals for External Funding

NIH/NIGMS Nebraska Center For Integrated Biomolecular Communication (CIBC); Core B, Data Management and Analysis (PI, J. Clarke) (pending)

USDA/AFRI Promoting Gastrointestinal Health And Reducing Subclinical Inflammation In Obese Individuals Through Intake Of Whole Wheat Products In Comparison With Fruits And Vegetables (PI, Devin Rose)

NIH R01 Site-Specific Protein Glycoform Analysis (PI, Eric Dodds) (pending)

NSF NRT DESE: Data Enabled Science and Engineering in Sustainable Plant Ecosystems (PI, J. Clarke) (pending)

USDA NIFA/AFRI Understanding how soil microbes interact with environmental conditions and affect GHG flux dynamics in managed pastures (PI, Samodha Fernando) (pending)

USDA AFRI Promoting gastrointestinal health and reducing subclinical inflammation in obese individuals by increasing whole grain intake (PI, Devin Rose) (pending)

USDA NIFA/AFRI Quantitative source characterization of bacterial contamination in lettuce production (PI, Bing Wang) (pending)

USDA NIFA/AFRI Identifying emerging viral zoonotic pathogens in swine using a novel high throughput screening method and identifying phages against zoonotic bacterial pathogens in swine for phage therapy (PI, Samodha Fernando) (pending)



Whittier Research Center
2200 Vine Street
Lincoln, Nebraska 68583
bigdata.unl.edu

