CSC-IBM International Joint Research Program (PhD, Post-doc, Junior Faculty)

Professor/Department/University	Project Title	Project Description	Technical/Knowledge Background	Qualifications
Climate Change, Agriculture and Digital Adaptations				
Art Degaetano, Earth and Atmospheric Sciences, Cornell University (http://www.eas.cornell.edu/people/profile.cfm?netid=atd2_eng)	(1) Spatial and Temporal Scales of Climate Data Required to Make Sound Agricultural Decisions	The project I have in mind would focus on the spatial and temporal scales of climate data required to make sound agricultural decisions. Often the perceptions of users is that the finer the spatial scale of climate data the more useful (or accurate) it is. Through this research I would hope to isolate the spatial resolution below which finer resolution information is indistinguishable from random noise. This will be shown based on the accuracy of the data themselves, as well as differences in the manage practices that would have been applied based on that data. An ultimate goal would be to assess the value of finer resolution data relative to the savings realized by the use of the data when applied to a management decision. This assessment could be conducted both on real time weather data and/or future climate projections.	Physical Science (not necessarily Atmospheric Science, but that would be preferred). At least some background computer programming and statistics.	Python/MatLab programming proficiency. Moderate-high ability in written and spoken English. Familiarity with field measurement technology (preferably weather stations).
Johannes Lehmann, School of Integrative Plant Sciences, Soil and Crop Sciences Section, Cornell University (https://scs.cals.cornell.edu/people/johannes-lehmann)	(2) Soil Carbon Stabilization for Promoting Food Security and Climate Change Mitigation	More carbon is stored in soil than in the biosphere and atmosphere combined, and small changes in soil carbon storage therefore has profound influences on atmospheric carbon dioxide and global warming. Yet, our prediction models for climate change contain mathematical models that are based on erroneous assumptions of the nature of soil organic matter. Similarly important is organic matter in securing soil health functions and food security. Through ultra-high resolution	The postdoc should either have technical background in modeling (either spatial modeling using GIS, spatial statistics, use of remote sensing information; or process modeling either life cycle assessments or systems dynamics models) or analyses using spectroscopy (NEXAFS, XPS, NMR, Raman, FTIR) and microscopy (SEM, TEM,	Knowledge about soil and earth system science is desirable.

		spectroscopy and isotope experiments, basic mechanisms of carbon-mineral and aggregation formation can be investigated that will feed into next-generation mathematical models. These models need then be verified through wide geographic analyses of regional and continental carbon behavior. Special interest pertains to the properties and behavior of fire-derived pyrogenic carbon in its effects on soil carbon dynamics and global carbon cycles.	NanoSIMS) and sample preparation (microtoming, FIB)	
Peter Hess, Biological and Environmental Engineering, Cornell University (https://bee.cals.cornell.edu/people/peter-hess)	(3) Land-Atmosphere Biogeochemical Nitrogen Cycling: Climate and Air Pollution	The research addresses fundamental questions regarding the interaction between the biogeochemistry of the nitrogen cycle and the resulting impacts on climate and air-quality. It addresses how the nitrogen cycle will change in the future as the climate changes, emissions change and agricultural changes. The research involves simulating various aspects of the nitrogen cycle and agriculture within an Earth System model.	Someone with quantitative training in a physical science (physics, atmospheric science) or mathematics. At any rate knowledge of mathematics and hopefully some statistics is necessary. Also good knowledge of English is important.	More specifically skill in computing is necessary. Knowledge of large scale computing, big data, atmospheric science, biogeochemistry, and Earth System modeling would be a plus.
Harold van Es, School f Integrative Plant Science, Soil and Crop Sciences Section, Cornell University (https://scs.cals.cornell.edu/people/harold-van-es)	(4) Climate Change, Soil Health and Cropping Systems	My program is working on important aspects related to climate change, agriculture, and digital solutions. We have developed an assessment and management framework for soil health that allows cropping systems to be more productive and resilient to climate change. We have also developed computational technology that allows for more efficient use of nitrogen fertilizer using weather, soil, and crop management inputs, which allows for greater production efficiencies as well as reduced impacts of greenhouse gas emissions and water contaminants.	In terms of the technical background, I would be looking for a person who has a strong background in soil science and understands agricultural systems. In addition, the person needs to have a quantitative interest in terms of the application of digital technologies to soils and agronomy.	I would be looking for someone who has strong background in statistics (including familiarity with R software) and has some programming skills.
David Wolfe, School of Integrative Plant Science, Horticulture Section, Cornell University (https://hort.cals.cornell.edu/people/david-wolfe)	(5) Agricultural Soil and Water Management for Changing Climate	I currently have research on-going that might succinctly be described as "Agricultural Soil and Water Management for Changing Climate". I have a postdoc working on NYS water resources-agriculture-climate change, and there is more to be done there than 1 postdoc can accomplish, so that would be one research	→	- English proficiency - Excel - crop production, agroecology, research experience (particularly maize, and/or vegetable crops) - soil management, soil

		area. I'm also working with The Nature Conservancy on various research concepts associated with "climate smart soils", soil carbon sequestration and soil health etc.. Another area is policy interests at international level- developing mechanisms for agriculture, forestry, other land uses (AFOLU) to be part of COP negotiations and address national and international mitigation goals. Planning a major side event for the next COP (in Bonn, Germany about a year from now).		sampling and analyses experience, particularly in relation to soil organic matter, soil carbon sequestration, and soil health *Other qualifications of interest:* - water management in crop production - role of agriculture, forestry, other land uses (AFOLU) in climate change mitigation and policy
Natalie Mahowald, Earth and Atmospheric Sciences, Cornell University (http://www.eas.cornell.edu/people/profile.cfm?netId=nmm63)	(6) Aerosol-Climate-Biogeochemistry Interactions	Recent studies have highlighted the role of natural and anthropogenic aerosols and their impacts on biogeochemistry, especially the carbon cycle and their resulting impact on climate. Indirect changes in aerosols can also result from land use, including changing emissions from wild fires, desert dust or forests as well as direct emissions from agriculture, which can impact climate and biogeochemistry. Our group focuses on addresses these poorly understood processes.	Requires atmospheric science, physics, chemistry or engineering background	Good computer skills
Toby Ault, Earth and Atmospheric Sciences, Cornell University (http://www.eas.cornell.edu/people/profile.cfm?netid=tra38_eng)	(7) Quantification of Megadrought Risk	Megadroughts are prolonged periods of aridity unlike anything seen during the historical period, and they have been linked to the demise of several preindustrial civilizations. Mounting evidence suggests that the risks of such events during climate change is increasing due to rising temperatures and dynamic circulation changes throughout many of the world's subtropical dry zones. This project will work towards quantifying global megadrought risk on near-term (decadal and multidecadal) time horizons using a combination of statistical techniques and new numerical model simulations. The postdoc will assist in both the data analysis and climate modeling aspects of this work.	In general, the postdoc should be comfortable working in a linux environment and should know at least one of the following (or related) interpreted languages: Matlab, Python, R, IDL, NCL, or equivalent. Proficiency in any of these languages is largely transferable, so knowing one (or a related one not listed here), would be sufficient. In terms of background: interest/experience with chaotic dynamical systems, multivariate statistics, or applied physics	Basic familiarity with Fortran would be a plus, but is not essential as this is an esoteric language still only widely used within the atmospheric sciences. Knowledge of C++ would likewise be advantageous for getting started, but isn't essential.

			would make for a strong post-doc.	
Ying Sun, School of Integrative Plant Science, Soil and Crop Sciences Section, Cornell University (https://scs.cals.cornell.edu/people/ying-sun)	(8) The Application of Chlorophyll Fluorescence for Crop Stress Monitoring and Yield Prediction	Chlorophyll Fluorescence (F) is a direct probe of photosynthesis and has the potential to be applied for crop stress monitoring and yield prediction. However, the quantitative relationship of F with plant physiology and crop yield likely differs among crop cultivars and varies with environmental conditions, and management practices. This project aims to develop a predictive understanding of F dynamics and use the understanding gained to guide practical applications of F measurements from different observational platforms in crop stress management and yield prediction.	The postdoc is expected to have general background in ecology-, agriculture-, computational-related fields.	Preference would be given if the postdoc have skills in machine learning technique or/and field measurement.
Michael Gore, School of Integrative Plant Science, Plant Breeding and Genetics Section, Cornell University (https://plbrgen.cals.cornell.edu/people/michael-gore)	(9) Deep-Learning for High-Throughput Plant Phenotyping	We will develop an unmanned aircraft system (UAS) platform to collect images over experimental crops, and aim to develop deep learning algorithms to identify plant pathologies and morphologies at an accuracy that is on par with human experts. The UAS will consult human experts in ambiguous cases and gradually learn to make decisions autonomously.	This postdoctoral associate position involves the phenotyping of foliar diseases in maize with several complementary ground- and aerial-based methods in the field. The postdoc will computationally process collected images along with geospatial information and apply deep learning algorithms for reliable identification of foliar diseases. The ideal candidate will have expertise in remote sensing, image processing, deep learning, and statistical genetics. Responsibilities will include research in the collection and processing of geospatial and image data, statistical dissection, prediction and validation of disease phenotypes, and training scientists and students. The position will involve close collaboration with a dynamic	A Ph.D. in remote sensing, statistics, computer programming or related discipline with at least 2 years of intensive training in statistical methods. Programming (R/Java/Python/Julia), image (ImageJ/Agisoft/Pix4D) analysis skills, development and/or application of deep learning algorithms, and working knowledge of remote sensing, geospatial, and statistical approaches. Excellent interpersonal and communication skills with a strong publication record in the field of remote sensing and statistical genetics.

			team of robotics engineers, computer scientists, statistical geneticists, and plant pathologists.	
Mark Sorrells, School of Integrative Plant Science, Plant Breeding and Genetics Section, Cornell University (https://plbrgen.cals.cornell.edu/people/mark-sorrells)	(10) Incorporating High Throughput Phenotypes and Environmental Covariates in Genomic Prediction Models to Accelerate Genetic Gain in response to Climate Change	This project would use correlated phenotypes from repeated aerial imaging of breeding research plots and environmental parameters as components of genomic prediction models to increase prediction accuracy. We have data sets and computing facilities to use in these analyses and can provide the expertise for training in the necessary methods. We have published more than a dozen peer-reviewed articles on genomic selection methods including one using environmental covariates that enables the prediction of performance in untested environments expected with climate change.	The person should have deep knowledge of statistics and quantitative analysis. A good working knowledge of genetics and breeding would be desirable.	Ability to program in R and manage large datasets, aerial image analysis of plants.
Marc Fuchs, School of Integrative Plant Science, Plant Pathology and Plant-Microbe Biology Section, Cornell University (https://pppmb.cals.cornell.edu/people/marc-fuchs)	(11) Improving Our Understanding of Grapevine Red Blotch-Associated Virus	Among the recently described plant viruses is grapevine red blotch-associated virus, a monopartite single-stranded DNA virus. Little is known about the interaction of this virus with its natural host and treehopper vector. Using cutting-edge microbiological and molecular techniques, we will advance our knowledge of the virus interface with its host and vector.	I am anticipating the postdoc to have excellent knowledge of plant pathology and experience in molecular biology. More importantly, enthusiasm and dedication to excellence are expected.	
Christine Smart, School of Integrative Plant Science, Plant Pathology and Plant-Microbe Biology Section, Cornell University (https://pppmb.cals.cornell.edu/people/christine-smart)	(12) Understanding Genetic Diversity in Pathogens of Vegetable Crops	Plant diseases such as cucurbit downy mildew, Phytophthora blight, late blight of tomato and potato, and tomato leaf mold are having a major economic impact on vegetable production. My lab studies pathogen diversity to identify effectors that are present in unique pathogen populations, and also to track pathogens geographically. Be identifying the effector complement of each pathogen population, it is possible to determine the plant resistance genes that will be effective against each population, which will enhance disease control.	The postdoc should have a working knowledge of plant pathology (preferably a PhD in plant pathology or microbiology), understanding of plant biology and molecular biology.	Special qualification include ability/desire to work both in the field and in the lab. Some knowledge of bioinformatics.

Dan Buckley, School of Integrative Plant Science, Soil and Crop Sciences Section, Cornell University (https://scs.cals.cornell.edu/people/daniel-buckley)	(13) Metagenomic and Isotopic Techniques in Microbiome Identification	Soil microbiomes provide ecological services which underlie the sustainability of both agricultural and ecological systems, and yet the vast majority of soil microorganisms remain poorly characterized. We are using a suite of metagenomic and isotopic techniques to identify dominant members of soil microbiomes and to characterize their impacts on soil health, plant productivity, and on the terrestrial carbon cycle.	Candidates should be familiar with the analysis of high-throughput DNA sequencing data generated from microorganisms or microbial communities. They should have familiarity working with either amplicon sequence data, comparative genomic data, or metagenomic data. Successful applicants will have a background in microbial ecology, environmental microbiology or allied field and experience in bioinformatic analysis of DNA sequence data. Experience with R and Python, and experience working with computer scripts necessary for bioinformatics is required.	The post-doc should have a Ph.D. in Microbiology or an allied field with specific research experience that relates to either the genomics, ecology, physiology, or evolutionary biology of environmental microorganisms (i.e. non-pathogenic microbes).
Lailiang Cheng, School of Integrative Plant Science, Horticulture Section, Cornell University (https://hort.cals.cornell.edu/people/lailiang-cheng)	(14) Sugar Metabolism and Accumulation in Fleshy Fruits	The objective is to understand the molecular mechanisms that underlie sugar metabolism and accumulation in fleshy fruits such as apple, with particular emphasis on sugar transporters at the cell membrane and tonoplast, with the ultimate goal of manipulating sugar levels in these fruits for quality improvements. In addition to routine molecular approaches, functional characterization of sugar transporters via electrophysiology techniques such as patch clamp will be used.	Good skills in gene cloning and functional characterization in yeast, cell culture and whole plants, and experience in working with fruit crops are preferred.	A Ph.D in plant physiology, biochemistry, molecular biology, horticultural science or crop science is required.
Lailiang Cheng, School of Integrative Plant Science, Horticulture Section, Cornell University (https://hort.cals.cornell.edu/people/lailiang-cheng)	(15) Physiological and Molecular Mechanisms for Cellular Calcium Partitioning in Fleshy Fruits	The objective is to understand the physiological and molecular mechanisms that regulate calcium uptake into fleshy fruits such as apple and cellular calcium partitioning among subcellular compartments in relation to fruit quality, particularly calcium-related physiological disorders. This project will have the potential to significantly reduce Ca deficiency-related disorders and improve	Good skills in gene cloning and functional characterization in yeast, cell culture and whole plants, and experience in working with fruit crops are preferred.	A Ph.D in plant physiology, biochemistry, molecular biology, horticultural science or crop science is required.

		fruit quality. A variety of experimental techniques including transcriptome analysis, cellular calcium imaging, and electrophysiology methods will be used.		
Todd Walter, Biological and Environmental Engineering, Cornell University https://bee.cals.cornell.edu/people/m-todd-walter	(16) Climate Change Impacts on Agricultural Water Resources Management	This past summer much of New York State experienced a severe drought, which resulted in considerable crop loss. Ironically, in the previous year there was too much rain early in the growing season and many crops drowned and had to be replanted. This project would consist of three activities: (1) use the hydro-meteorological record (e.g., rainfall, snow, stream discharge, etc.) across New York State (NYS) in combination with stochastic and/or simulation models to determine the historical frequency and spatial distribution of droughts/floods that result in crop failures; (2) apply stress tests to the previous analysis to determine how sensitive NYS's agricultural systems are to changes in weather extreme magnitudes and frequencies; and (3) test the capacity of management decisions (e.g., adopt irrigation, increase soil water holding capacity with carbon amendments, etc.) to potentially mitigate water-related crop risks. This project would interface with those being led by Drs. DeGaetano, Ault, van Es, and Ault.		
Todd Walter, Biological and Environmental Engineering, Cornell University https://bee.cals.cornell.edu/people/m-todd-walter	(17) Climate Change Impacts on Water Quality in Agricultural Watersheds	My research group and colleagues at Cornell have been working on strategies for mitigating nonpoint source pollution from agricultural runoff based on watershed hydrology. We have developed a prototype model and web interface that predicts runoff-generating locations throughout a watershed; currently we have set up this model for the Owasco Lake watershed in central New York. This project would expand the scope of the model statewide and apply the historical weather records in order to identify the frequency, duration, and seasonality of		

		storm runoff generation from rural lands statewide and then apply a suite of potential future climates to assess how these patterns are likely to change. One tangible output from this project will be a user-friendly, web-based mapping interface that stakeholders can use to determine runoff risks at daily-to-annual scales, including forecasted risks.		
Todd Walter, Biological and Environmental Engineering, Cornell University https://bee.cals.cornell.edu/people/m-todd-walter	(18) Climate-Change Impacts on Flood Risk	This project will be part of our on-going efforts to assess flood hazards and risks across New York State (NYS). One of the challenges in making large-scale assessments like this is a lack of data; both data required to runoff watershed models and data to assess our confidence in the models. We have proposed an approach that allows us to stochastically describe the flood generating weather features (e.g., storm intensity, duration, snowmelt, etc.) and utilize the existing stream discharge record to assess our models' ability to predict "flood risks;" in other words, we are not concerned about predicting a particular discharge on a particular day but only whether or not there is a flood or no flood on any given day. By reducing our dependence on simulation models in this way, we can assess our overall confidence in our flood risk predictions, which may suffer due to any number of factors, e.g., length of records, uncertainty in model parameters, etc. We can also incorporate "soft data" like human observations or recollections of floods to reduce our prediction uncertainty. The researcher working on this project will apply this approach across the state to estimate both flood risks and patterns of uncertainty in our risk estimates. The latter will be especially useful in locating areas where additional monitoring or other environmental data are critically needed to improve our confidence in flood risk		

		predictions.		
Larry Smart, School of Integrative Plant Science, Horticulture Section, Cornell University https://hort.cals.cornell.edu/people/lawrence-smart	(19) Genetic Basis for Hybrid Vigor in Shrub Willow Bioenergy Crops	Little is known about the genetic basis for hybrid vigor, even though it occurs in many crops - even in shrub willow. We have developed a diverse set of segregating trait mapping populations that can be used to correlate phenotypic variation in key traits with genotypic polymorphisms. Identification and characterization of the QTL and genes involved in important biomass yield traits will lead to the development of tools for early selection of genetically improved cultivars and improved knowledge of the underlying genetic factors controlling complex traits.	They should have a strong background in plant molecular biology, plant genetics, genomics, and bioinformatics. This should include gene cloning, nucleic acid isolation from plants, gene expression analysis, analysis of RNA-Seq and genomic sequencing data.	In addition to the background knowledge above, they should be qualified to execute statistical analysis, especially of genomic data in R.
Food Safety				
Sam Nugen, Food Science, Cornell University (https://foodscience.cals.cornell.edu/people/sam-r-nugen)	(20) Genetic Engineering of Bacteriophages for the Rapid Detection of E. coli	Genetic engineering of bacteriophages for the rapid detection of E. coli: Bacteriophage tail fibers can be genetically engineered to tailor a bacterial host range. The phage can then be used to separate, concentrate, and detect bacteria from a complex matrix. The assay can be designed to low-cost and deliver rapid results.	The candidate should have good basic microbiological skills and some experience in cloning and molecular biology. Additionally, the ideal candidate will have: • Excellent written and communication skills • Ability to conduct independent research • Ability to lead a team consisting of graduate and undergraduate students	Candidates with a Ph.D. in microbiology, food safety, molecular genetics, or similar fields are strongly encouraged to apply.
Sam Alcaine, Food Science, Cornell University (https://foodscience.cals.cornell.edu/people/sam-alcaine)	(21) Enabling Realtime Detection of Adulterants in Fermented Dairy Products Using Metagenomic Analysis	Description: Changes in milk quality and adulterants (residual antibiotics, melamine, etc) will impact how bacterial populations perform during fermentation. By analyzing the metagenomic expression patterns of bacterial cultures during yogurt fermentation in the presence and absence of adulterants, we could identify early warnings genes that could help us flag these food safety and quality issues early.	The candidate should have good basic microbiological skills and some experience in cloning, transcriptome analysis, and general molecular biology techniques. Experience in dairy fermentations and working with lactic acid bacteria and/or bacteriophage is preferred, but not required	• Excellent written and communication skills • The ability to conduct independent research • The ability to lead a team consisting of graduate (MS, PhD) and undergraduate students
Sam Alcaine, Food Science, Cornell University (https://foodscience.cals.cornell.edu/people/sam-alcaine)	(22) Metagenomic Illumination of Protective Cultures	Protective bacterial cultures are increasingly being used in dairy and other	The candidate should have good basic microbiological skills and	• Excellent written and communication skills

		food applications as alternatives to chemical preservatives to inhibit the growth of pathogens and spoilage organisms. Little is known about the mechanisms of this inhibition. This project would leverage metagenomic analysis to identify potential genes necessary for inhibition by comparing closely related lactic acid bacteria strains that are and are not inhibitory of eukaryotic spoilage organisms.	some experience in cloning, transcriptome analysis, and general molecular biology techniques. Experience in dairy fermentations and working with lactic acid bacteria and/or bacteriophage is preferred, but not required	• The ability to conduct independent research • The ability to lead a team consisting of graduate (MS, PhD) and undergraduate student
Sam Alcaine, Food Science, Cornell University (https://foodscience.cals.cornell.edu/people/sam-alcaine)	(23) Visualization of Bacterial Contamination Patterns in Food	We know that pathogenic bacteria regularly contaminate food, but know little about the routes (surface, internalization, etc). This project will involve engineering bacteriophage to express visual reporters that can be seen, as a color bloom, on the surface or within a food (model food will be cheese) if the pathogen of interest is present. This will potentially allow investigators to identify routes of contamination in food processing plants and improve food safety.	The candidate should have good basic microbiological skills and some experience in cloning, transcriptome analysis, and general molecular biology techniques. Experience in dairy fermentations and working with lactic acid bacteria and/or bacteriophage is preferred, but not required	• Excellent written and communication skills • The ability to conduct independent research • The ability to lead a team consisting of graduate (MS, PhD) and undergraduate students
Martin Wiedmann, Food Science, Cornell University (https://foodscience.cals.cornell.edu/people/martin-wiedmann)	(24) Development and Implementation of Food Safety Genomics Tools	This project will involve the development or implementation of bioinformatics and/or genomics tools that can be used to improve food safety and reduce microbial food spoilage. For example, scholars may (i) develop and implement new approaches that can be used to determine genetic signatures that can be used to predict the source of a foodborne contaminant; (ii) develop approaches that can be used to identify abnormalities in raw materials based on metagenomic signatures; or (iii) perform whole genome sequencing (WGS) and analyze WGS data to characterize pathogen or spoilage organism transmission in food processing plants.	Expertise in bioinformatics, molecular biology, genome analyses, GIS data analyses or related fields.	Good written and oral English communication skills, including expertise in publishing peer-review papers in English journals, are also required.
Wind Energy				
Sara Pryor, Earth and Atmospheric Sciences, Cornell University (http://www.eas.cornell.edu/people/profile.cfm?netid=sp2279_eng)	(25) Improved Understanding of, and Simulation of, the Causes of Intra-Annual to Inter-Annual	The feasibility of wind energy installations at given locations are dictated in large part by the wind resource (‘power in the wind’).	Experience with numerical modeling and high-performance computing (i.e. Running models	PhD in Atmospheric Science and/or Mechanical Engineering. Strong analytical

	Variability in Wind Resources and Operating Conditions	Variability of, and uncertainty in, that resource increase project risk (and uncertainty in electrical power production). We seek to improve prediction of the wind farm lifetime resource and operating conditions by developing and optimizing efficient numerical (computational) tools.	such as WRF on high performance computing platforms, cloud or conventional). It would be desirable to have also experience with model performance methods and of course wind energy resource assessments.	skills and if an Mech Eng graduate at least one course in atmospheric boundary layers (or similar).
Lindsay Anderson, Biological and Environmental Engineering, Cornell University (https://bee.cals.cornell.edu/people/catherine-anderson)	(26) Hybrid Statistical Optimization Methods for Stochastic Resources in Power Systems	Incorporating data analytic approaches with stochastic optimization, to accelerate operational decisions on large power networks with significant renewable resources.	Technical background should include some experience in formulating and solving optimization problems. Solutions are computational, so some scientific computing background is important. IN my lab we work with Matlab, Python and (occasionally) GAMS	No specific qualifications, other that a PhD in a quantitative field – could be math, engineering, operations research, statistics. Some background in energy applications is helpful, but mathematically strong researchers can generally pick that up fairly quickly.
Lindsay Anderson, Biological and Environmental Engineering, Cornell University (https://bee.cals.cornell.edu/people/catherine-anderson)	(27) Bi-level Optimization Under Uncertainty to Incorporate Utility Scale Renewables with Demand Side Flexility and Distributed Resources	This project will investigate bi-level optimization algorithms to model interactions between transmission and distribution level components to develop synergistic operational strategies.	Technical background should include some experience in formulating and solving optimization problems. Solutions are computational, so some scientific computing background is important. IN my lab we work with Matlab, Python and (occasionally) GAMS	No specific qualifications, other that a PhD in a quantitative field – could be math, engineering, operations research, statistics. Some background in energy applications is helpful, but mathematically strong researchers can generally pick that up fairly quickly.
Healthcare				
Philip S.Li, Center for Male Reproductive Medicine ＆Microsurgery, Cornell University (http://urology.weillcornell.org/philip-s-li)	(28) Male Infertility Microsurgical Big Data Research for Male Reproductive Medicine and Microsurgery	The project focuses on Male Infertility Microsurgical Data Research for Male Reproductive Medicine and Microsurgery. Microsurgical data analytics training is essential for clinical audiologists specializing in male infertility. Success in clinical microsurgery depends on practice in the laboratory and data analytics with technology. Microsurgery for male infertility is among the most technically and mentally challenging of microsurgical procedures, which generates big amount of unstructured and structured data for	Domain knowledge in Male Reproductive Medicine and Microsurgery; proficiency in SPSS, SAS, or R. Strong skills in data cleaning, data mining, data analysis and MIM HPC tools.	Clinical / Resident doctor with Master degree, PhD candidate, Post-doc.

		advanced research and clinical usage. Most male infertility micro-procedures are performed under 10 to 25 power magnification and required all image to be stored and processed in high performance computing. In contrast to conventional post surgery evaluation, tones of surgical procedures are as dependent on technical perfection and technology application. Coordination, dexterity and steadiness of the microsurgery can only be developed with extensive practice in the laboratory with strong technical skills, especially with data mining and analytics background.		
Open Power Technology				
Liang Huang, EECS, Oregon State University (http://web.engr.oregonstate.edu/~huanlian/)	(29) Natural language processing, computational biology, machine learning	Project Background: Simultaneous speech-to-speech translation is just like real-time interpreters: you have to start translation before the source sentence ends, and gradually translates as more input is available. Example usage: United Nations. Project Goals: Build the first simultaneous speech-to-speech translation framework and software using deep learning and reinforcement learning. Publish 2 top conference papers.	required: Strong in algorithms design and analysis, esp. dynamic programming Strong coding experience (Python, C/C++, Java). Experience in deep learning toolkits (Theano, Torch, Tensorflow, etc.) Experience in machine translation. recommend: Experience in machine learning (esp. reinforcement learning)	PhD Candidate; Post-doc
Stephen Ramsey, Department of Biomedical Sciences, Oregon State University (http://eecs.oregonstate.edu/people/Ramsey-Stephen)	(30) Computational biology; machine learning in biology/biomedicine	Project Background: Worldwide DNA sequencing capacity is almost an exabase pair (1018 bp) per year, and growing exponentially. Human population genetics datasets are being generated on a massive scale (multiple "Million Human Genomes" sequencing projects are underway worldwide) that will enable mapping the molecular basis of complex traits and human diseases with unprecedented precision. At present, a lack of scalable computational methods–and an integrated cloud-based system making such methods available to researchers–to mine and extract knowledge from discoveries of trait associations within "noncoding" regions of the genome are lagging far behind data	CS background with emphasis on machine-learning and big data. Interest in biology or genetics. Some background in biology preferred. Skill Set: python/numpy/scipy/scikit-learn; SQL; DB2; bash; Linux shell tools; experience with cloud computing & virtualization. Also a plus: Theano or similar deep learning framework (e.g., Tensorflow)	PhD Candidate; Post-doc

		generation. Our lab's expertise is in developing machine learning methods to identify and functionally characterize human genetic variants using large-scale datasets from human population genetic studies. Project Goals: We propose to develop and deploy an IBM DB2 Cloud-based system that would accelerate and advance human population genetics studies. More specifically the system would enable life scientists to search for, rank, and view evidence for candidate causal genetic variants within regions of the genome that have been implicated in genetic association studies for traits of interest (for example, risk of heart attack or stroke). The system would incorporate advances in machine learning–that have originated in our lab–for discriminating functional from nonfunctional genetic variants. The Y-100 scholar would work with our team of three researchers (the PI and three computer science graduate students from China), and would specifically work on enabling the system that we are building to leverage IBM DB2 to enable efficient querying of large-scale population genetics datasets (10-100 billion rows). Industry: Public Health (Computational biology; genomics; machine learning; population genetics; cloud computing)		
Brett Tyler; Chris Sullivan, Center for Genome Research and Biocomputing, Oregon State University (http://bpp.oregonstate.edu/tyler)	(31) Genome sequence alignment and assembly is a major bottleneck. New advances in hardware architecture allow for changes in how data can be processed. This project will re-write genome assembly algorithms to run on the IBM Power8 with NVIDIA NVLink GPUs using the CAPI interconnect.	Project Background: The skyrocketing amounts of genomic data generated by modern DNA sequence technologies are creating major data processing bottlenecks. The most important processing tools use sequence alignment and assembly, for example to assemble full genomes. Currently, these tools use many cores, require much memory and take weeks to run. This project will use IBM POWER8 technology with GPUs to attack this bottleneck. The CGRB has purchased a new	Knowledge of CUDA and GPU technologies. Skill Set: Ability to work on a Linux based operating system, IBM Power8 processors, IBM CAPI, IBM 822LC Server, NVIDIA NVLink P100 Pascal using.	PhD candidate, Post-doc, Junior professor

		IBM S822LC server with two 10-core Power8 processors, two NVLink P100 Pascal GPUs and 1TB of local RAM. The CGRB plans to use the new hardware with the CAPI interface to connect the GPU and in order to change the way genome sequence data can be processed. The CGRB has spent the last year working with the IBM to compile and help port scientific software to the Power8 processors with NVIDIA K80 GPUs (https://www.ibm.com/blogs/systems/ibm-power8-and-osu-advance-genomics-research- through-porting/). These older systems with card based GPUs work great for data on the CPU and simulations on the GPU. The new machine with GPUs using the CAPI interface will accelerate the way data can be processed on the GPU. Project Goals: The main goal of this project is to port a tool that uses De Bruijn graph theory to assemble genomes to the IBM Power8 with CAPI NVLink GPU processing using CUDA. The final genome assembly tool will take advantage of the multiple cores within the GPU to dramatically decrease processing time. To take full advantage of the CUDA Cores in the GPU the tool will move data through the CAPI interface to interact with system memory to reduce needed GPU memory footprint. Finally we will use of the Power8 cores to manage the data moving onto and off of the GPU to ensure throughput. Industry: Genome assembly tools are used throughout biomedical research and increasingly for genome-informed personalized health care. The industry aims to go from patient sample to assembled genome sequence within 24 hours. This can only be achieved by adapting assembly tools onto new hardware technologies.		
Analytics and Visualization				

Chen Li , School of Information and Computer Science, University of California-Irvine (http://www.ics.uci.edu/faculty/profiles/view_faculty.php?ucinetid=chenli) Mike Carey, School of Information and Computer Science, University of California-Irvine (http://www.ics.uci.edu/faculty/profiles/view_faculty.php?ucinetid=mjcarey) [1] http://cloudberry.ics.uci.edu [2] http://cloudberry.ics.uci.edu/demos/twittermap/	(32) Cloudberry: Interactive Analytics and Visualization of Large-Scale Fast Data	We are developing a new, general-purpose open source system called "Cloudberry" [1] to support interactive data analytics and data visualization over large amounts of fast data. The Cloudberry system aims to provide several unique and important capabilities: Scalability, Interactivity, Visualization, Currency. A system on 500 million tweets with live data being ingested is available at [2]. We will study various open challenges in this exciting research direction, including: (1) A domain-independent middleware layer to translate frontend Restful requests to queries to the backend AsterixDB; (2) Cache module and replacement policies at the middleware; (3) Intelligent query slicing to reduce initial query responsive time and return results progressively; (4) Making the middleware distributed across multiple machines; (5) Improving the AsterixDB LSM storage and indexing to reduce computational cost per query; and (6) Supporting continuous queries to reduce costs of compilation and deployment.	Expertise in data management, hands-on programming skills	Post-doc; Junior Professor
Sharad Mehrotra, School of Information and Computer Science, University of California-Irvine (http://www.ics.uci.edu/faculty/profiles/view_faculty.php?ucinetid=smehrotr)	(33) Innovation in Data Cleaning	The key insight on which this proposal is based is that big data analytics in streaming, real-time, and interactive settings requires a paradigm shift in how data cleaning is performed Proposed research will explore 2 new innovations to help advance data cleaning for Big Data analysis. The first explores a progressive approach to ER to support progressive analysis. The proposed research will explore an approach wherein progressiveness is pervasive spanning all the phases of cleaning specially in scenarios when cleaning is based on complex logic possibly requiring dynamic acquisition of additional contextual information. The second is the analysis-aware cleaning that is developed for structured queries (e.g., Hive and SQL) for both one-time and	Expertise in data management	Post-doc; Junior Professor

		continuous query scenarios that are issued on top of static and streaming data. The project will exploit a concrete context to guide the research exploration – viz., line analysis of social media data.		