INFORMATICS

Informatics, or information science, involves collecting, processing, interpreting, and integrating vast data sets within a contextual framework to advance science and make more precise and improved recommendations. All informatics applications are computer-based, and as computational capacities increase, researchers face the ever-growing challenge of developing ways to effectively filter and interpret increasing amounts of data into meaningful and useful information.

Informatics experts with the University of Florida Institute of Food and Agricultural Sciences (UF/IFAS) integrate computer science and statistical techniques to analyze patterns, tendencies, associations, and causality.

UF/IFAS researchers collaborate with scientists in a multitude of disciplines when analyzing massive data sets of genomic information to improve animal reproduction, epidemiological data to predict disease outbreaks, and ecological systems information to structure food webs to better understand species interactions. Scientists' overarching goal is to use "big data" to better meet the ever-increasing needs of stakeholders for more reliable, useful, and easily accessible information.
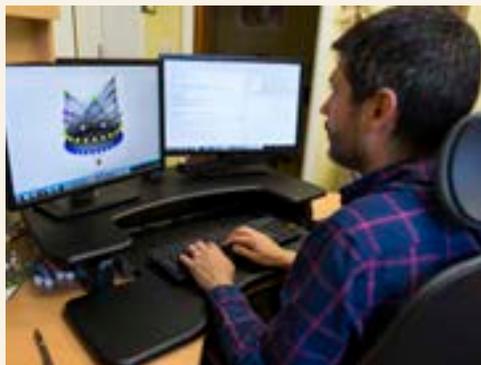
NATURAL RESOURCES

AGRICULTURE

HUMAN SYSTEMS

Ongoing Research

DEVELOPING FOOD WEBS

Ecologists provide recommendations based on existing knowledge, but the databases they use are growing at an exponential pace. Ben Baiser, an assistant professor in the UF/IFAS Department of Wildlife Ecology and Conservation, joined a team of international researchers to address the limited usability of these datasets by building new, synthetic datasets. Using research from more than 1,000 research articles, his team defined the structure and function of "food webs," an important way to quantify relationships and interdependencies between species. The researchers then generated qualitative predictions showing likely global hotspots for species interactions. Development of this reproducible pipeline for creating ecological food-web datasets will allow researchers to explore biodiversity at the rate and spatial scale necessary to address complex ecological questions.

IDENTIFYING MOLECULAR NETWORKS

An ongoing challenge in animal genomics is identifying causal relationships, not just associations, between how genes and molecules affect specific traits in animals of economic importance. Francisco Peñagaricano, an assistant professor with the UF/IFAS Department of Animal Sciences, is working on data integration of observable traits, genetic markers, gene activity, and metabolites from multi-generation animal populations in an attempt to reconstruct molecular causal networks underlying the multiple layers of "omics" data. His team is using this approach to determine the causal structural determinants underlying muscularity, fat composition, and meat quality in livestock, giving breeders the knowledge to optimize management practices and breeding strategies. Further development could lead to predicting with certainty these complex traits using network information before selecting for breeding stock.

AUTOMATING GENE COUNTS

Gene activity, which is the extent to which a gene is turned "on" or "off," controls many plant traits, and sequencing instruments are used to count the thousands of RNA molecules in the genes that are "on." Improved accuracy of the gene counts, which is essential to assessing genetic composition, requires ongoing experimental design. Ana Conesa, a professor with the UF/IFAS Department of Microbiology and Cell Science, helped developed the NOISeq R-package to facilitate gene counts. Her team then analyzed and compared this advanced bioinformatics tool with leading competing software packages and demonstrated that it outperformed the others by reducing bias and improving accuracy. Life science researchers are increasingly using this new tool for more accurate predictions of how gene activity controls plant traits such as yield, disease resistance, and product quality.

Research with Impact

INTEGRATING DIVERSE DATA

Recent developments in bioinformatics have enabled UF/IFAS plant breeders to make more precise genetic adjustments in blueberry varieties. Salvador Gezan, an assistant professor in the UF/IFAS School of Forest Resources and Conservation, collaborated with a multidisciplinary team of geneticists, food nutritionists, and biochemists to compare how genetic and environmental factors influence the perceived flavor of blueberries, a subjective trait. The team identified traits that were heritable and stable across growth environments, and were most appealing to volunteers who participated in sensory evaluation panels. Gezan used a multivariate statistical approach to combine the data and identify the biochemical compounds and fruit characteristics associated with consumer preference, which is an indicator of improved marketability. These traits have now become breeding targets for the development of superior blueberry varieties.

LINKING DNA SEQUENCES

Genomic information is used to enhance the efficiency of breeding and selection in beef cattle, and it is especially useful for crossbreeding Angus cattle (that have high-quality ribeye areas) with Brahman cattle (that have high heat tolerance) to harness the optimal characteristics from each. Mauricio Elzo, a professor with the UF/IFAS Department of Animal Sciences, and his colleagues capitalized on this information by linking a series of ultrasound traits with measured and imputed DNA-sequence variation data, and then used the entire dataset to evaluate animals for specific heritable traits such as ribeye area, back-fat thickness, percentage of intramuscular fat, and weight. This statistical procedure allows breeders to identify ideal parents that will create genetically valuable offspring before the cattle reach reproductive age, improving time and economic efficiency.
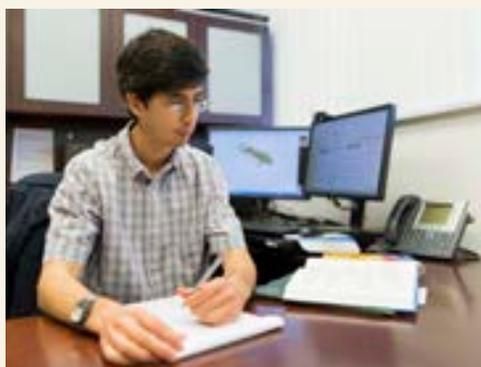
REDUCING MISCLASSIFICATION ERRORS

Malaria epidemiologists seek to identify disease risk factors so they can better recommend preventative measures, but many are unaware that the standard statistical model used for this purpose fails to account for imperfect diagnostic test results and misclassifications. Denis Valle, an assistant professor with the UF/IFAS School of Forest Resources and Conservation, and his team improved malaria analyses by developing advanced analytical tools that were more powerful and improved confidence of the identified risk factors. Using malaria data from the western Brazilian Amazon, one of their new models detected an influence of long-term exposure to the disease on malaria detection and infection probability, a significant factor overlooked by other methods. Valle's new data analysis approach provides critical guidance for malaria prevention and control methods.

Unit Leaders

Eric C. Hellgren
Chair and Professor
Wildlife Ecology and Conservation
352-846-0552
hellgren@ufl.edu

Geoffrey E. Dahl
Chair and Professor
Animal Sciences
352-392-1981 ext. 221
gdahl@ufl.edu

Eric W. Triplett
Chair and Professor
Microbiology and Cell Science
352-392-1906
ewt@ufl.edu

Terrell T. "Red" Baker
Director and Professor
School of Forest Resources and Conservation
352-846-0850 • ttredbaker@ufl.edu

Researcher Contacts

DEVELOPING FOOD WEBS
Benjamin Baiser,
Assistant Professor
Wildlife Ecology and Conservation
352-392-1947
bbaiser@ufl.edu

IDENTIFYING MOLECULAR NETWORKS
Francisco Peñagaricano,
Assistant Professor
Animal Sciences
352-392-1981, ext. 231
fpenagaricano@ufl.edu

AUTOMATING GENE COUNTS
Ana Conesa, Professor
Microbiology and Cell Science
352-273-8230
aconesa@ufl.edu

INTEGRATING DIVERSE DATA
Salvador A. Gezan,
Assistant Professor
School of Forest Resources and Conservation
352-846-0133
sgezan@ufl.edu

LINKING DNA SEQUENCES
Mauricio Elzo, Professor
Animal Sciences
352-392-7564
maelzo@ufl.edu

REDUCING MISCLASSIFICATION ERRORS
Denis Valle,
Assistant Professor
School of Forest Resources and Conservation
352-392-3806
drvalle@ufl.edu