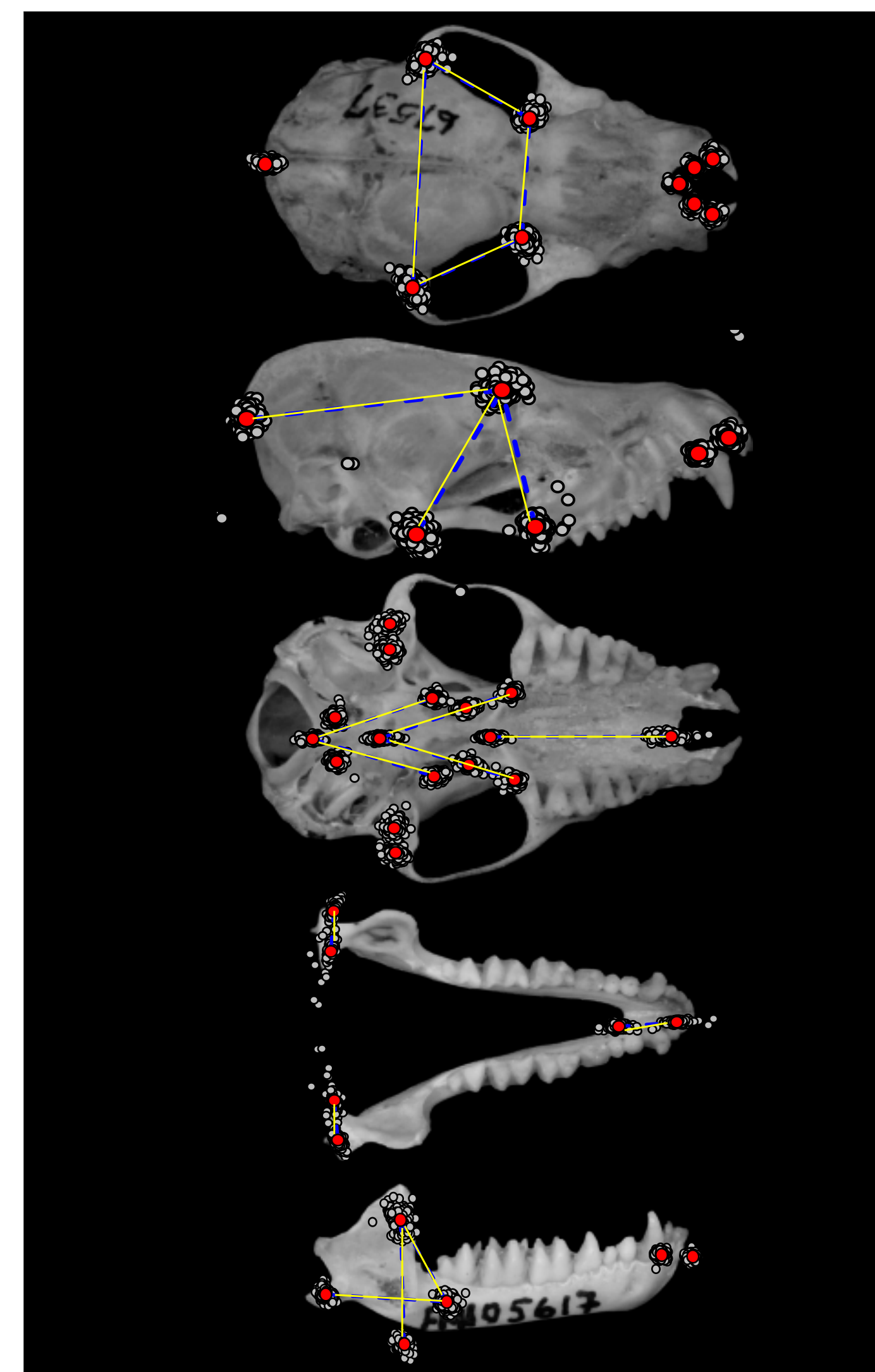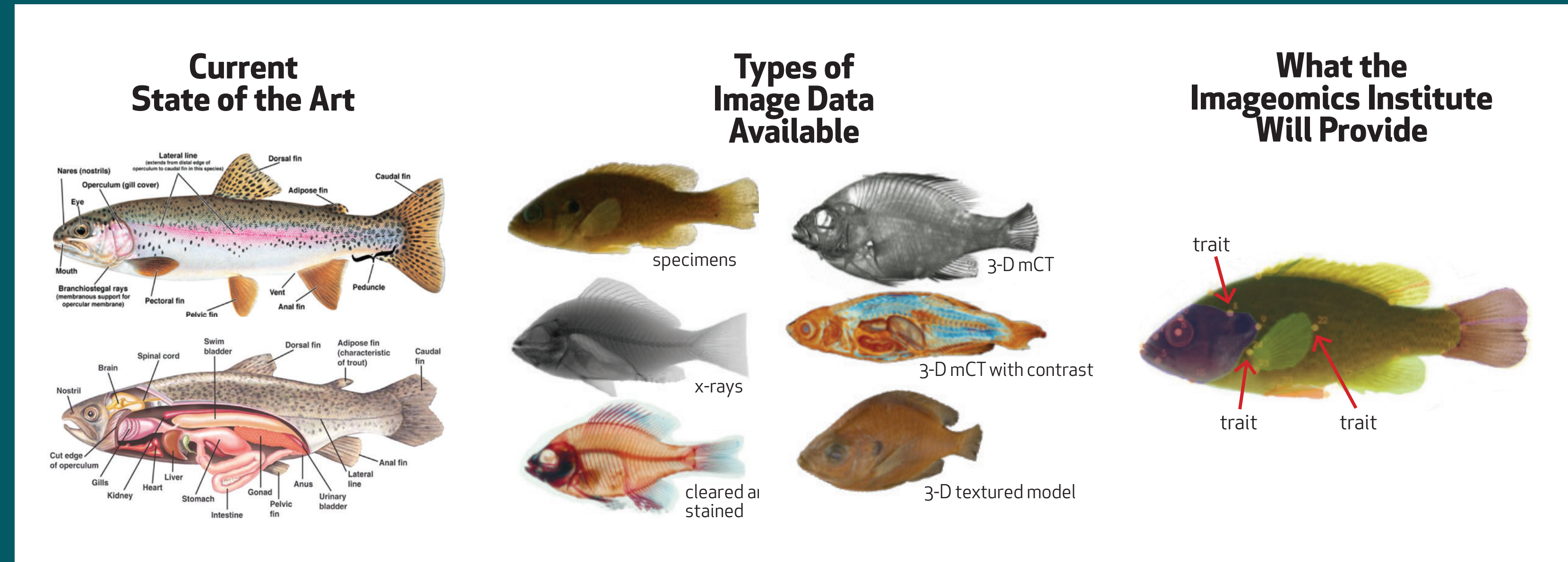# Introducing *imageomics*

imageomics | brand-new noun

i-'mi-jə-'ō-miks

A new scientific field in which machine learning tools built around biological knowledge bases are used by biologists to analyze image data in order to characterize patterns and gain insights into traits and relationships at individual, population and species scales—insights that then get incorporated into the algorithms that run the tools.
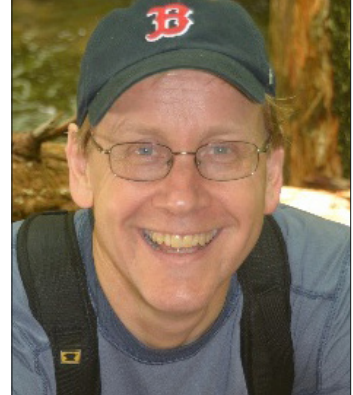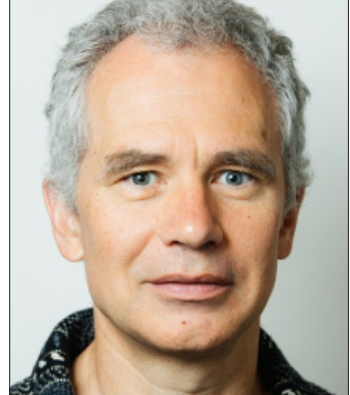
Imageomics Institute | brand-new entity

'i-mi-jə-'ō-miks 'in(t)-stə-tüt

An NSF Harnessing the Data Revolution Institute established in September 2021 that will create machine learning tools using publicly funded collections of digital image data from national centers, field stations, museums and individual laboratories that will enable scientists to study how function follows form in all areas of biology and expand public understanding of the rules of Life on Earth and how it evolves.
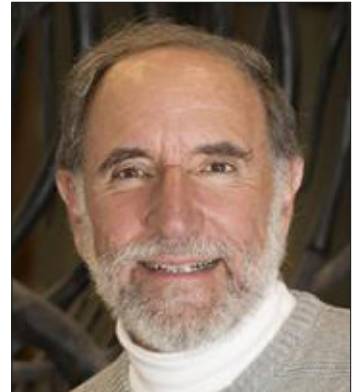

Current State of the Art

Types of Image Data Available

specimens

x-rays

cleared and stained

3-D mCT

3-D mCT with contrast

3-D textured model

What the Imageomics Institute Will Provide
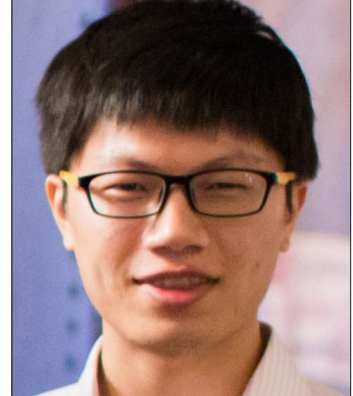
trait

Team

Executive Leadership Team

Institute Director
Tanya Berger-Wolf
Translational Data Analytics Institute; Computer Science and Engineering; Ecology, Evolution and Organismal Biology; The Ohio State University

Education and Community
Leanna House
Statistics, Virginia Tech

Research Convergence
Henry L. Bart Jr.
Ecology and Evolutionary Biology, Tulane University

Research Convergence
Charles Stewart
Computer Science, Rensselaer Polytechnic Institute

HDR Ecosystem Coordination
Paula Mabee
National Ecological Observatory Network – Battelle

Informatics and Infrastructure
Hilmar Lapp
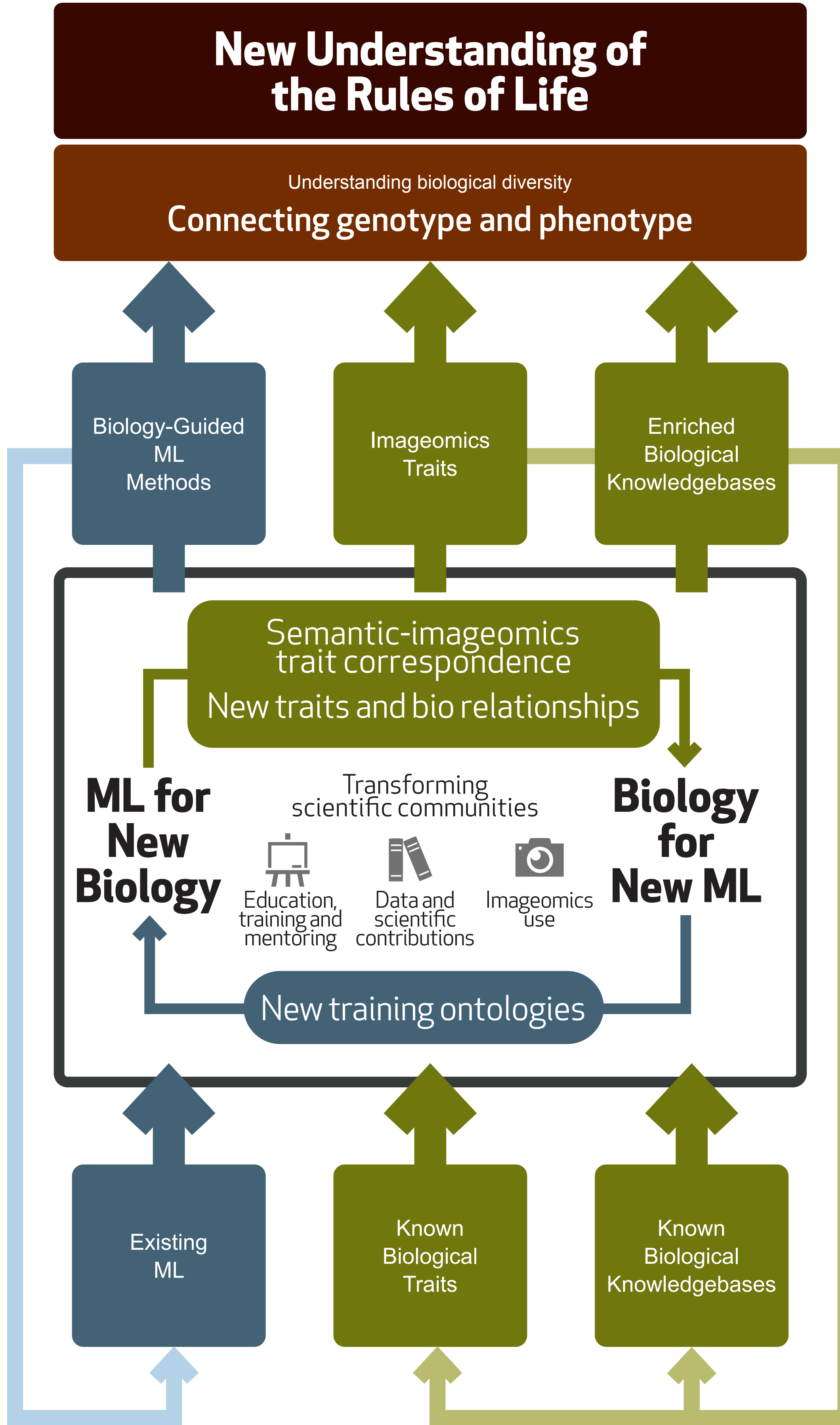Duke Center for Genomics and Computational Biology, Duke University

Convergence Workgroup Leads

Knowledge-Guided ML
Anuj Karpatne
Computer Science, Virginia Tech

Education and Participatory Science
Daniel Rubenstein
Ecology and Evolutionary Biology, Princeton University

Trait-Based Biology
Josef Uyeda
Biological Sciences, Virginia Tech

Data
Nico Franz
Life Sciences, Arizona State University

Structured Knowledge
Yu Su
Computer Science and Engineering, The Ohio State University

Investigators

Yasin Bakış
Museum of Natural History, Tulane University

Jim Balhoff
Renaissance Computing Institute, University of North Carolina – Chapel Hill

Bryan Carstens
Evolution, Ecology and Organismal Biology, The Ohio State University

Wei-Lun Chao
Computer Science and Engineering, The Ohio State University

Wasila Dahdul
University of California – Irvine Libraries

A. Murat Maga
University of Washington – Seattle Children's Research Institute

Collaborators

Conservation International
Cornell University
Ecole Polytechnic Institute and State University
Florida State University
Grevy's Zebra Trust
iNaturalist
Indiana University
Lincoln Park Zoo
Max Planck Institute
Microsoft AI for Earth
Microsoft Research
Midwest Big Data Hub
Natural History Museum, London
NOAA
Ohio Supercomputer Center
Open Traits Network
Open Tree of Life
Oregon Health and Science University
Smithsonian Institution
South Big Data Hub
University of Florida
University of Illinois – Urbana-Champaign
University of Minnesota
University of Pennsylvania
University of Washington
U.S. Fish and Wildlife Service
Wild Me

New Understanding of the Rules of Life

Understanding biological diversity
Connecting genotype and phenotype

Biology-Guided ML Methods

Imageomics Traits

Enriched Biological Knowledgebases

Semantic-imageomics trait correspondence
New traits and bio relationships

ML for New Biology

Transforming scientific communities

Education, training and mentoring

Data and scientific contributions

Imageomics use

Biology for New ML

New training ontologies

Existing ML

Known Biological Traits

Known Biological Knowledgebases

Granular

Control

Dentary lateral view

Approach

➤ **Knowledge from well-developed biological ontologies** (phylogenies, morphology, anatomy, ethograms, etc.), leveraging the **structure they impose**,

➤ together with **multiple imaging formats** (photos, videos, radiographs, CT scans and other 3-D images) and other data (textual descriptions, genetic sequences),

➤ to fuel the development of the next generation of **Knowledge-Guided Machine Learning models** that are more interpretable, transferable, robust, and label- and sample-specific,

➤ expanding the role of machine learning in addressing the most challenging biology problems, **forming a virtuous cycle.**

Broader Impacts

➤ Open, accessible digital resources for professional and community scientists

➤ Cyber, information and model-development infrastructure

➤ Virtual and in-person research and training environment

➤ Educational resources for next generations

➤ Inclusive communities of practice

➤ Public participation in democratized science