

How Can We Make Phenotype Research More Effective and Efficient?

Conceiving a Cyber Infrastructure to Enable High Throughput Studies of Biological Phenotypes

Edgar P. Spalding

Tessa Durham Brooks

Miron Livny

Department of Botany, University of Wisconsin, Madison WI

Department of Biology, Doane College, Crete NE

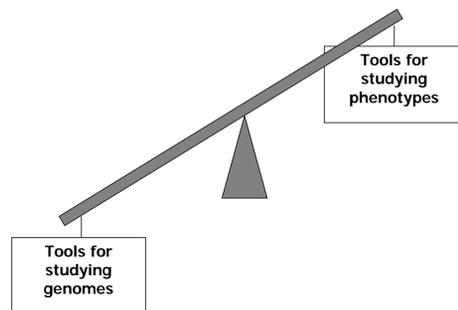
Department of Computer Sciences, University of Wisconsin, Madison WI

spalding@wisc.edu

tessa.durhambrooks@doane.edu

miron@cs.wisc.edu

A fundamental concept in biology is that the form, function, behavior, structure, or composition of an organism, its phenotype, bears a causal relationship to its genetic makeup, or genotype. A central goal in biology research is to produce a map relating the genotype to phenotype. There are typically millions of useful genotype features in an organism's DNA and the set of phenotype features is essentially unbounded because each may vary over time and depend on the environment. Biologists make piecemeal progress toward the grand genotype-to-phenotype map by focusing on phenotypes within their field of study. An agronomist would like a good map relating genotype to yield of a corn plant. A pharmacologist would like to know the genetic basis of side effects to a drug. What typically limits progress in this approach to understanding biology is the phenotype measurement. That's because the tools for characterizing genotypes (e.g. DNA sequencing) are technologically more advanced, automated, and higher throughput than technologies for measuring any phenotypic trait of an organism. Even in cases when every element of the subject organism's DNA is precisely known, progress will be slow if the phenotype is measured manually. The technologies for studying phenotypes are out of balance, technologically speaking, with those for studying genotypes.



In recent years, biologists have been devising higher throughput methods of measuring phenotypes, which would help right the imbalance shown at left. Automated image analysis is an example of an approach that can capture phenotype details in subjects studied through devices ranging in scales from microscopes to cameras mounted on a tractor. These attempts to level the technological balance between the study of genomes and phenotypes are important and necessary. But as each scientist develops methods for capturing the feature or process important to his or her particular experiment, they frequently find that the downstream steps are inadequate – making the measurement does not equal achieving an understanding in most cases. Problems with turning phenotype measurements into new understanding is limiting progress toward understanding biological systems.

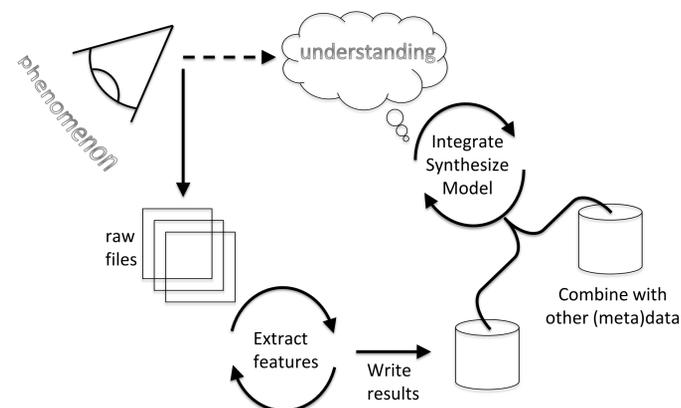
Our Conceptualization project set out to achieve an understanding of the bottlenecks in phenotype research. The primary approach was communication and community engagement. The principal methodology has been workshops, surveys, and personal communications.

Workshops brought biologists and computer scientists from the public and private sector together to distill down phenotype research across scales to its essence. One outcome is an abstraction of the process by which a machine-made observation of a phenotype is transformed into understanding of the system.



Photos from a workshop held in Madison (above) and of Miron Livny and Nirav Merchant enjoying a moment of agreement after a workshop held at the Plant and Animal Genome conference.

Understanding a Phenomenon through Observation



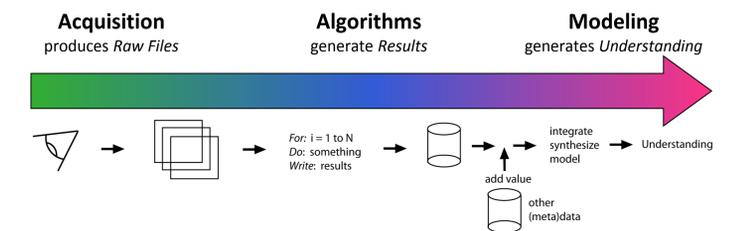
So, if phenotype research projects generally fit the above paradigm, where are the most common bottlenecks encountered?

- Dozens of survey respondents indicate the data acquisition step is rate limiting, while a very similar number indicate that data analysis is the bottleneck.
- Many participants in workshops and survey respondents indicate that image analysis tools are lacking, and too many are being created de novo, instead of taking advantage of previous efforts.
- Many people have indicated frustration with the length of time it takes to go from acquisition to result due to problems at each step of the pathway. As a result, the development trial-and-improvement cycle turns too slowly.
- Many people indicate uneasiness about how to begin making technology-based phenotype measurements due to unfamiliarity with the engineering and computing principles.

Based on these community engagement activities and many informal discussions, a framework of the problem is becoming clear. We find it helpful to recognize three domains of activity in a phenotype pipeline that relies on a software continuum to function, as outlined in the right-hand column of figures.

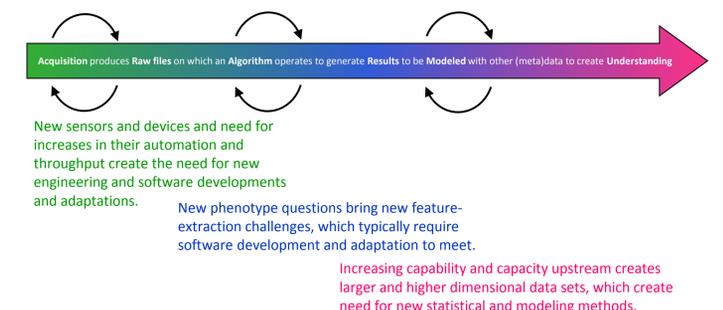
A Generalized Phenotype Pipeline

depends on a software continuum connecting three overlapping domains



An Adaptive Phenotype Pipeline

depends on a software continuum connecting three overlapping domains in which continuous development, refinement, and innovation are necessary cycles of activity

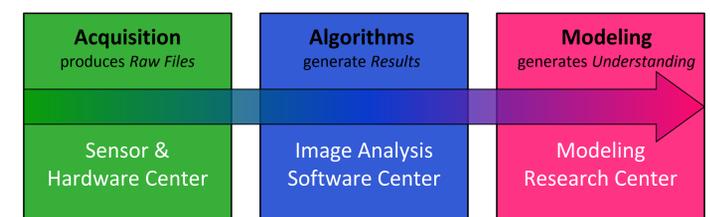


Emerging Recommendation

A framework of an actionable solution

Centers

are needed for innovation, development, and refinement of methods within each domain



Activities within each center would be driven by proposals from the biology research community. Committees would evaluate and rank the proposals and apportion resources accordingly. Products (outputs) would be made available to the community.

A Phenotype Institute

to identify bottlenecks and coordinate activities for their effective removal

