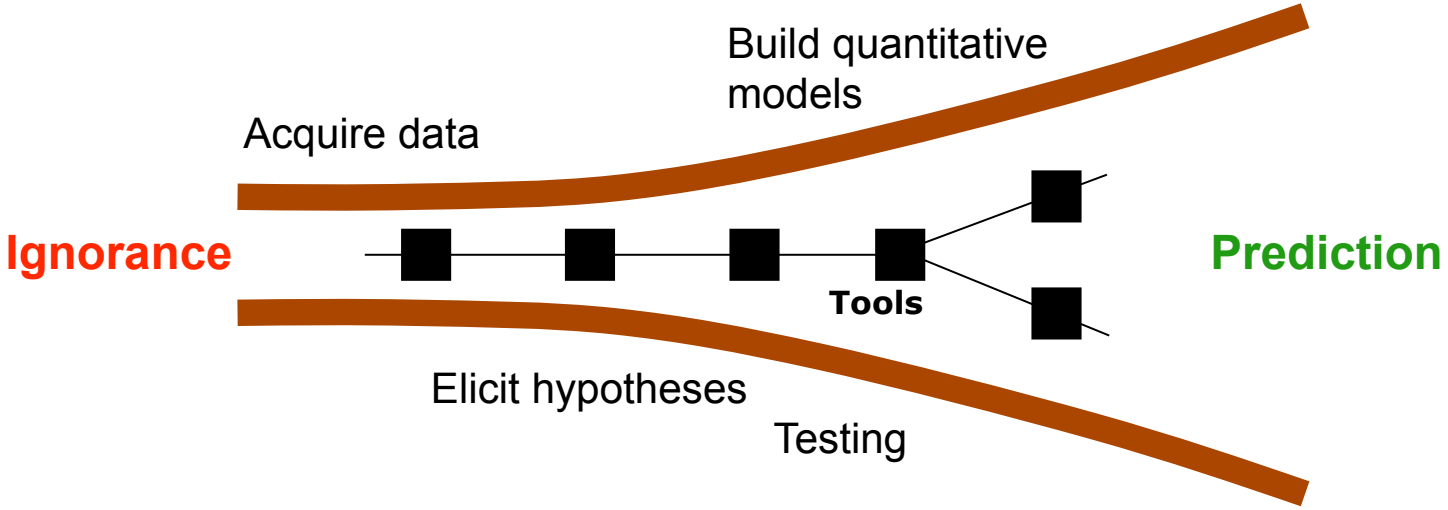
What is a “solution” to the G2P problem?

- Is it a database of facts about genotypes & phenotypes?
- Is it a systems biology simulation, ecophysiological, or QTL model?
- Is it a machine learning approach that starts with the first and ends with the second?
- Or something qualitatively different?

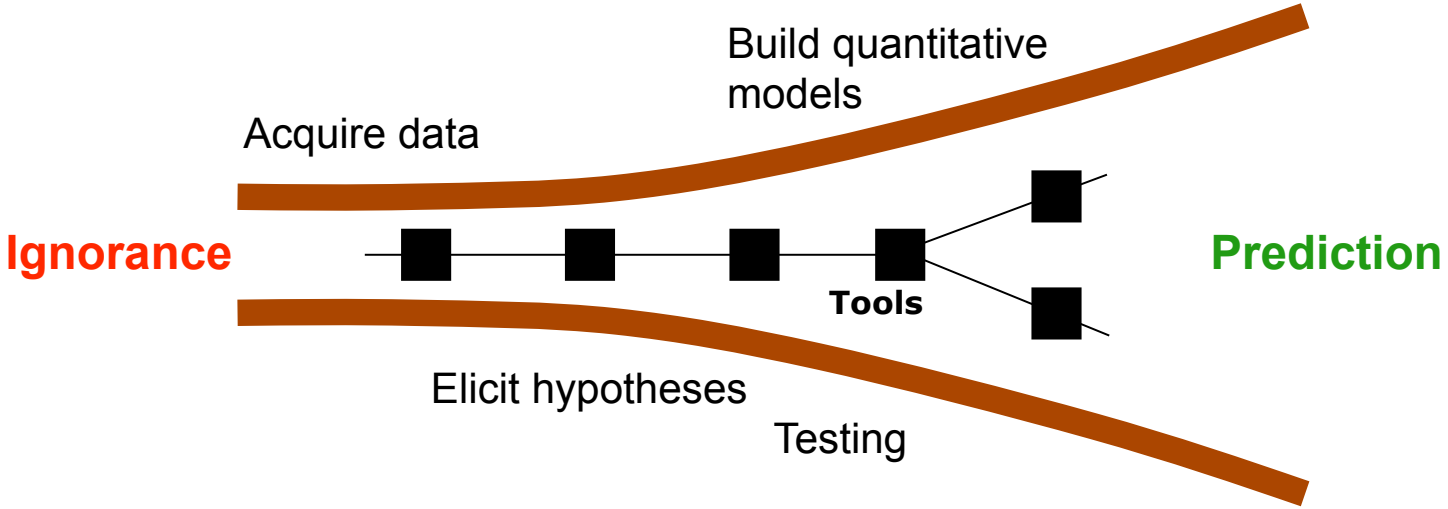
Solving the G2P problem means developing a methodology…

…that lets one start with some species & trait that one knows very little about and end with the ability to quantitatively predict trait scores for target genotype/environment combinations.

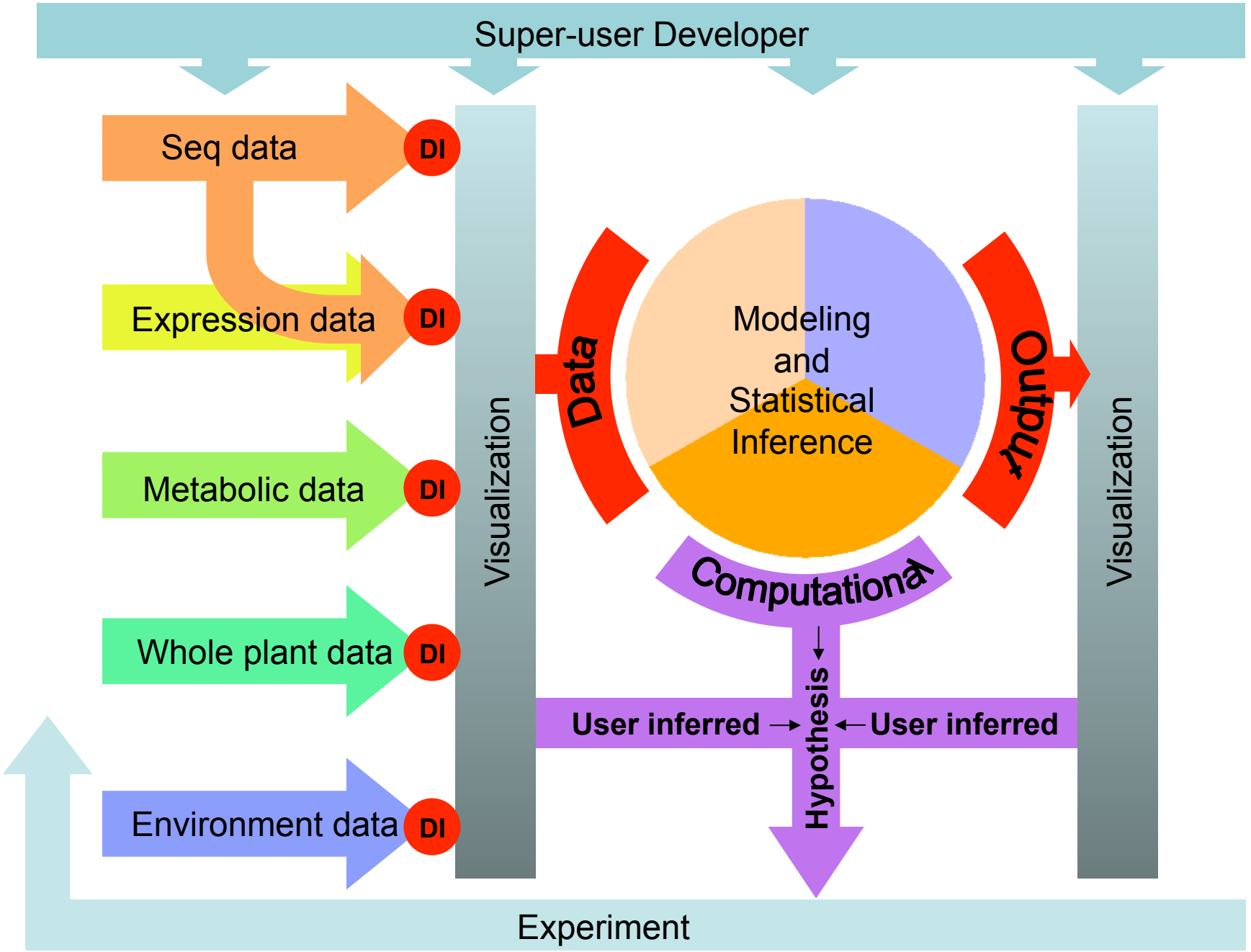


Solving the G2P problem means developing a methodology…

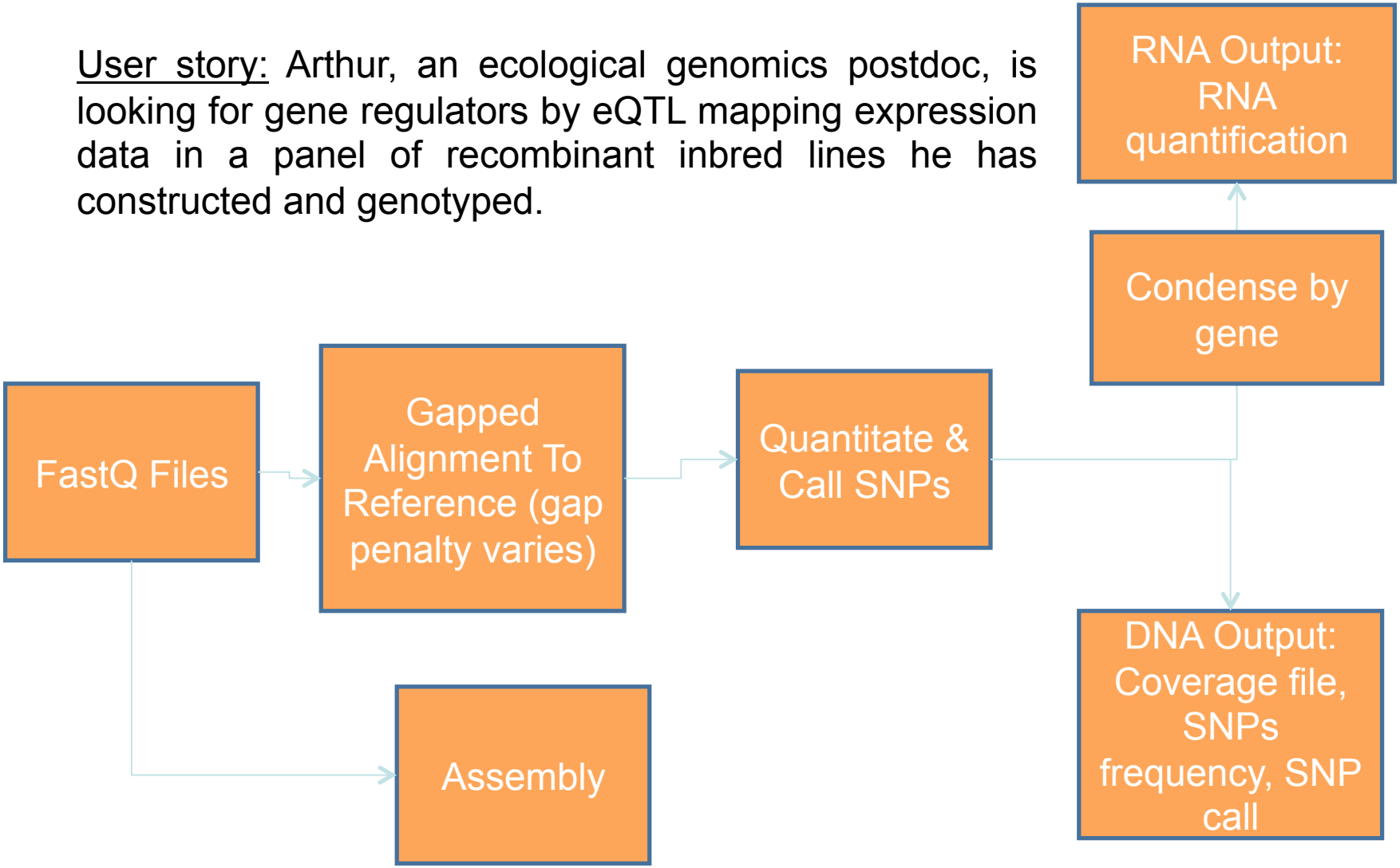
…that lets one start with some species & trait that one knows very little about and end with the ability to quantitatively predict trait scores for target genotype/environment combinations.

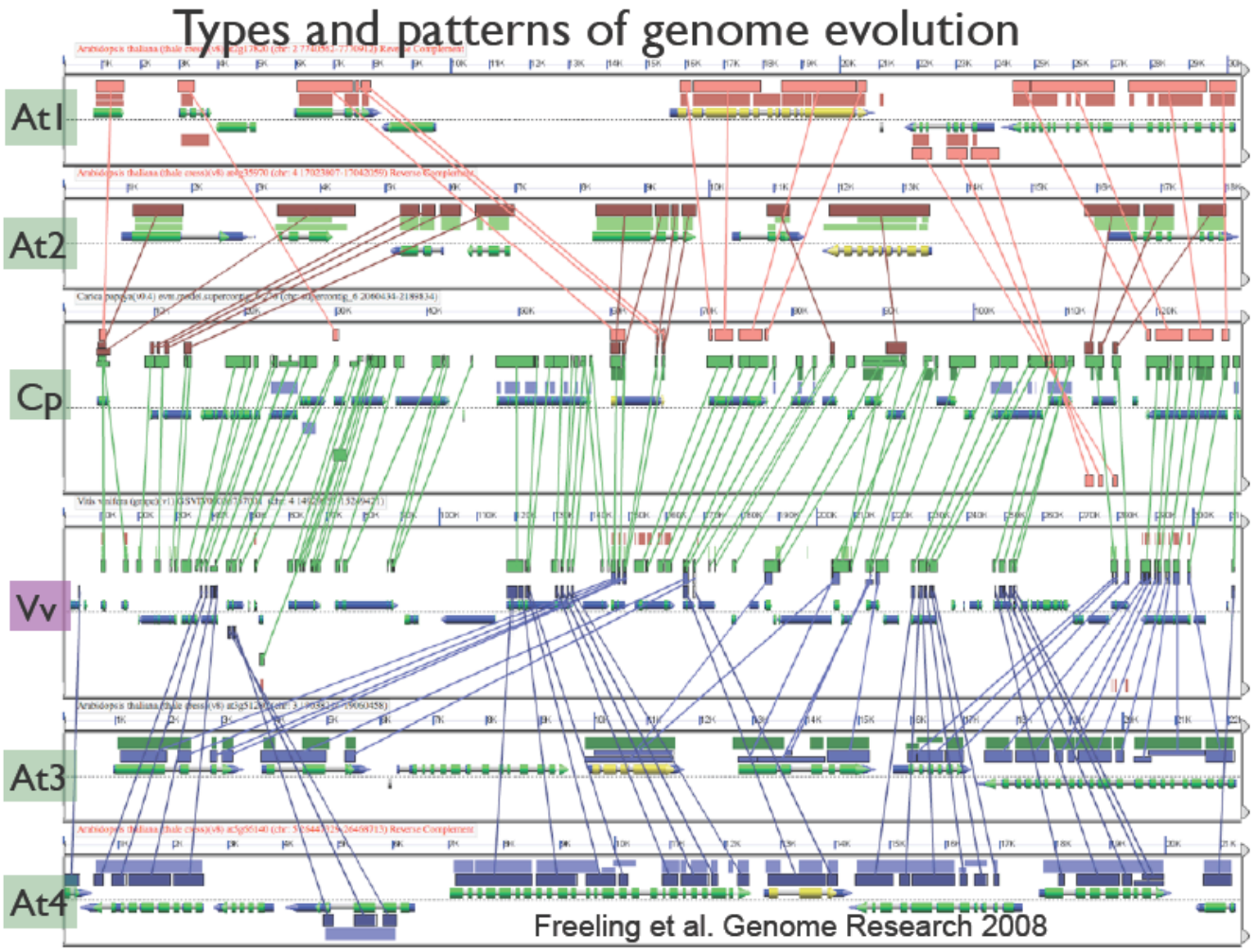


To work, such a methodology must be cyber-enabled


Super-user Developer
Seq data
DI
Expression data
DI
Metabolic data
DI
Whole plant data
DI
Environment data
DI
Visualization
Data
Modeling
and
Statistical
Inference
Output
Computational
Hypothesis
User inferred
User inferred
Visualization
Experiment

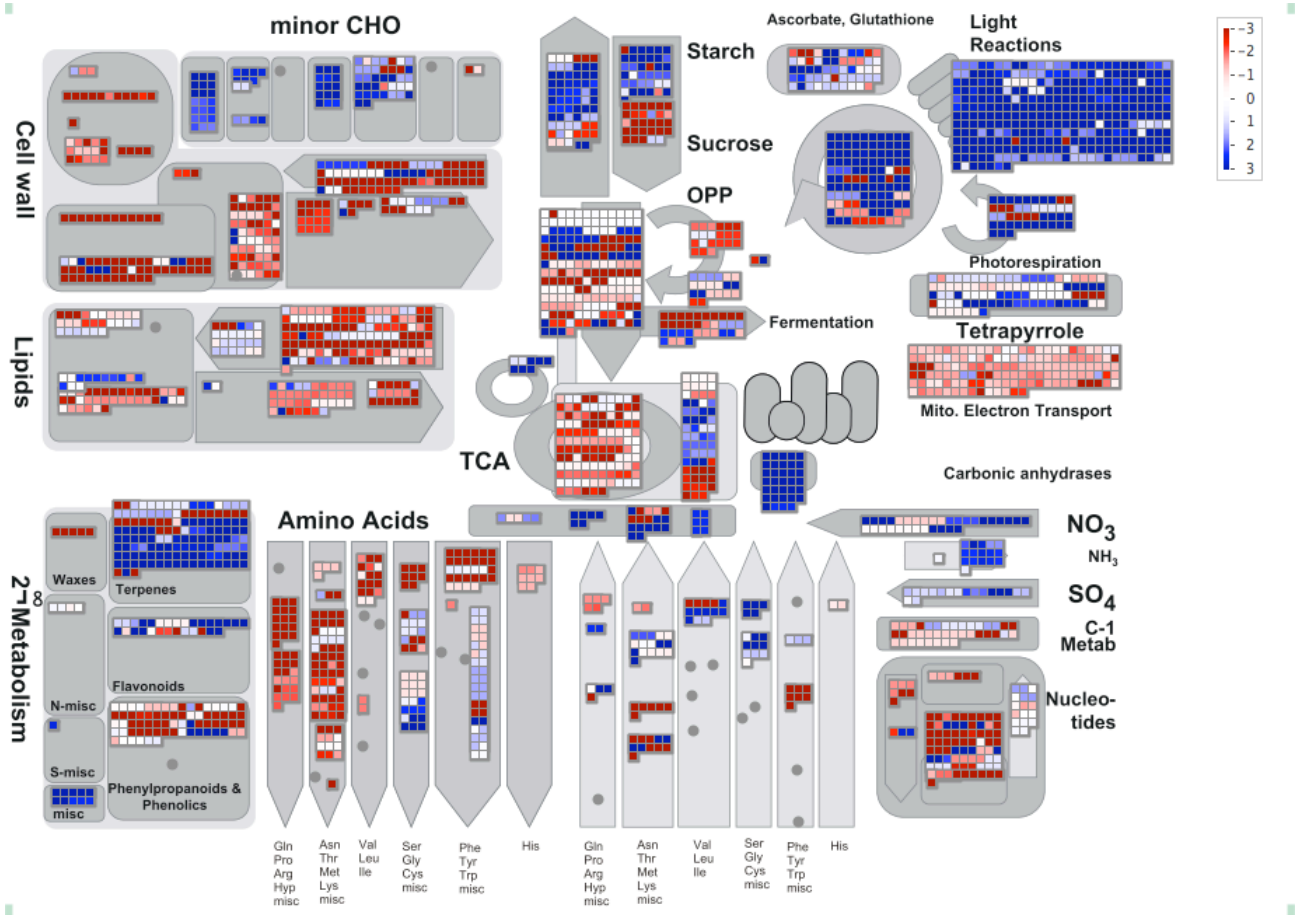
User story: Arthur, an ecological genomics postdoc, is looking for gene regulators by eQTL mapping expression data in a panel of recombinant inbred lines he has constructed and genotyped.





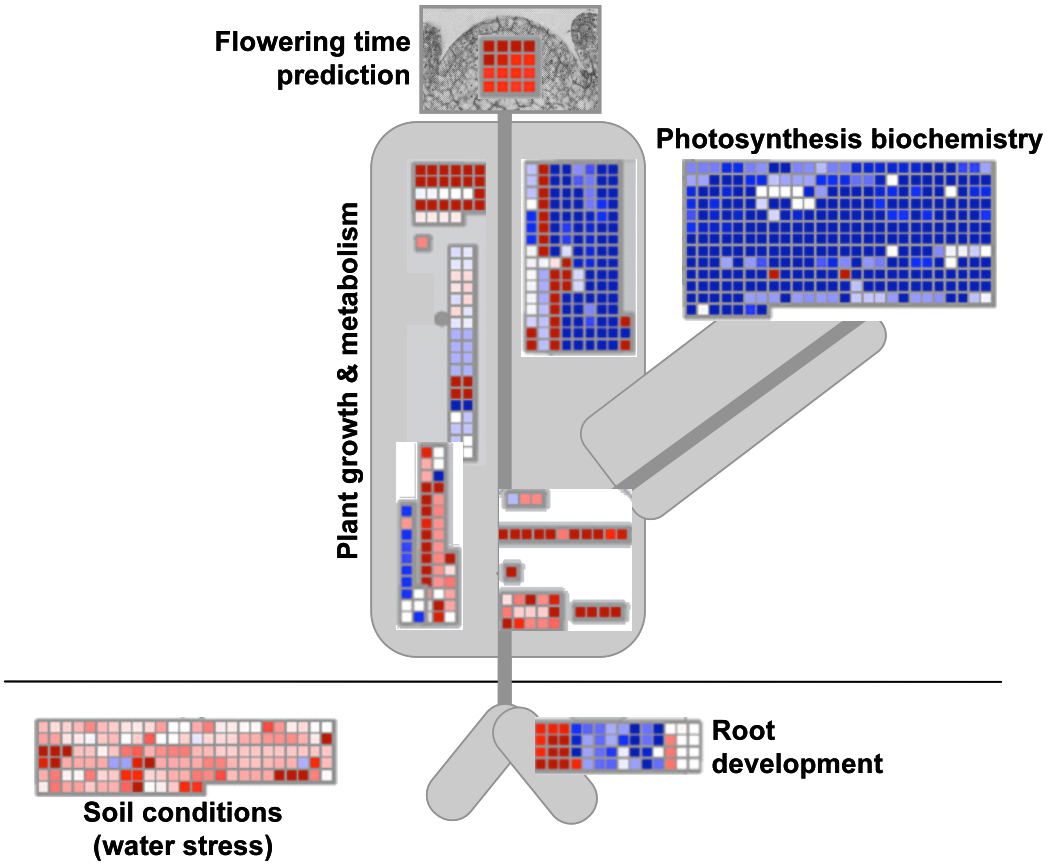
User story: Joachim, a bioinformatics postdoc, is seeking clues to a stubborn problem in genome evolution by comparing the alignments of related genes in several different species and ecotypes.

MapMan



User story: Marisha, a plant physiology PhD student, is looking for patterns in time series of gene expression data she has painted over a set of metabolic pathways.

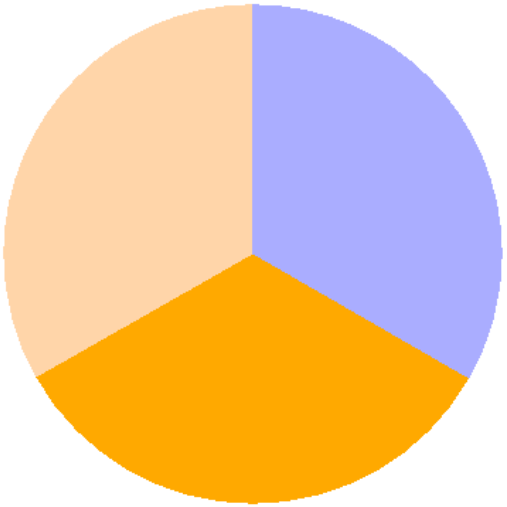
ModMan?



User story: Sue & Bill, high school science lab partners, are using a model to predict the development of the *Brachypodium* they are rearing from seeds supplied by iPlant.

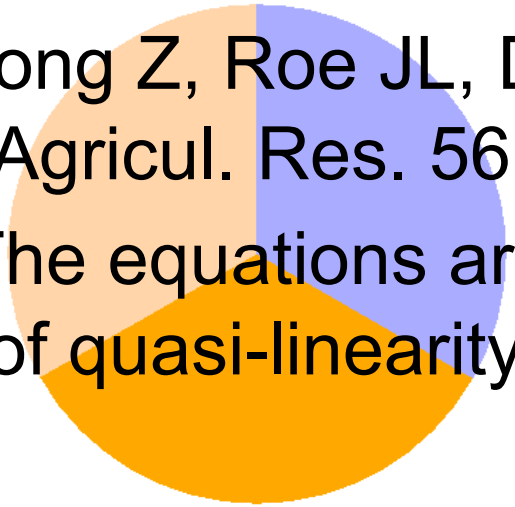
Linking Kinetic & Statistical Models

- This story starts in late 2002…



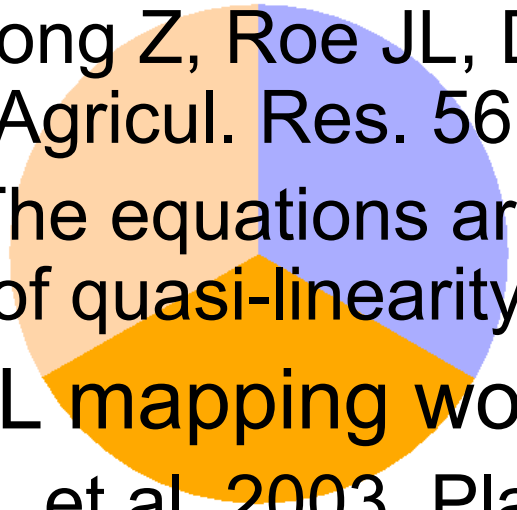
Linking Kinetic & Statistical Models

- This story starts in late 2002…
- By 2005 we had enough ideas to publish
  - Welch SM, Dong Z, Roe JL, Das S. 2005. Australian J. Agricul. Res. 56:919-36.
  - Short story: The equations are nonlinear, but hide regions of quasi-linearity within them.

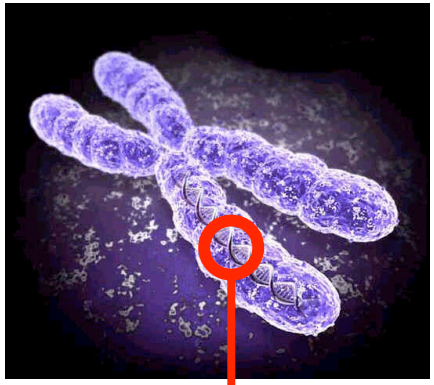


Linking Kinetic & Statistical Models

- This story starts in late 2002…
- By 2005 we had enough ideas to publish
  - Welch SM, Dong Z, Roe JL, Das S. 2005. Australian J. Agricul. Res. 56:919-36.
  - Short story: The equations are nonlinear, but hide regions of quasi-linearity within them.
- Why does QTL mapping work?
  - Reymond, M., et al. 2003. Plant Physiology, 131:664-75.



A Parametric Correspondence Principle



Polymorphic locus

Alleles: A1…An

Phenotypes: Y1….Yn

Model

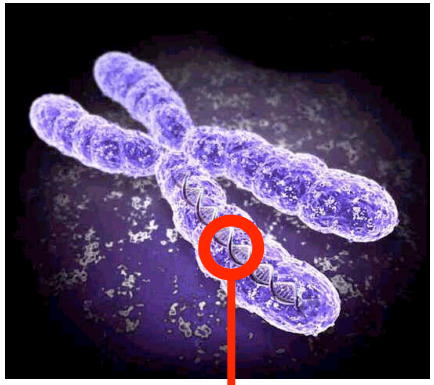
Some numerical parameter

Range of values in a set of mapping lines

2.71…3.14

y1….yn : predictions

A Parametric Correspondence Principle



Polymorphic locus

Alleles: A1…An

Model

Some numerical parameter

Range of values in a set of mapping lines

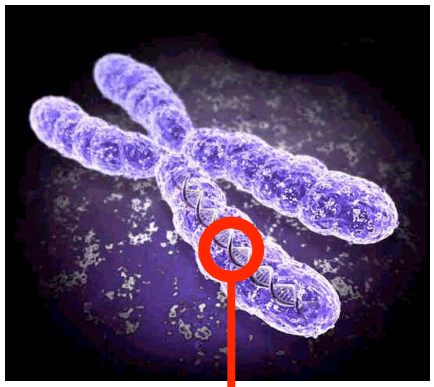
2.71…3.14

Phenotypes: Y1….Yn

A correspondence here…

y1….yn : predictions

A Parametric Correspondence Principle



Polymorphic locus

Alleles: A1…An

…induces a correspondence here

Model

Some numerical parameter

Range of values in a set of mapping lines

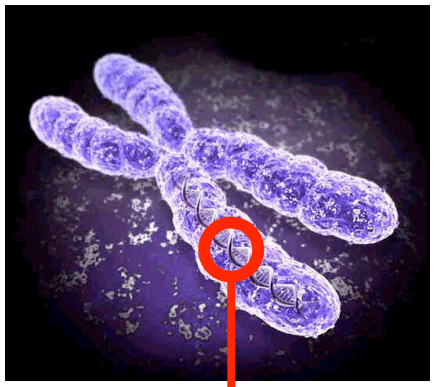
2.71…3.14

Phenotypes: Y1….Yn

A correspondence here…

y1….yn : predictions

A Parametric Correspondence Principle



Polymorphic locus

Detectable by a statistical association test

Model

Some numerical parameter

Range of values in a set of mapping lines

2.71…3.14

Alleles: A1…An

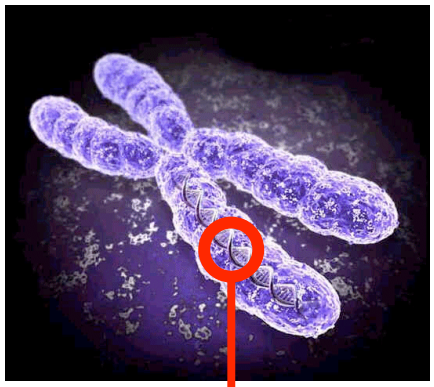
…induces a correspondence here

Phenotypes: Y1….Yn

A correspondence here…

y1….yn : predictions

A Parametric Correspondence Principle



Polymorphic locus

No matter what the model's mathematical form !!!

Detectable by a statistical association test

Model

Some numerical parameter

Range of values in a set of mapping lines

2.71…3.14

Alleles: A1…An

…induces a correspondence here

Phenotypes: Y1….Yn

A correspondence here…

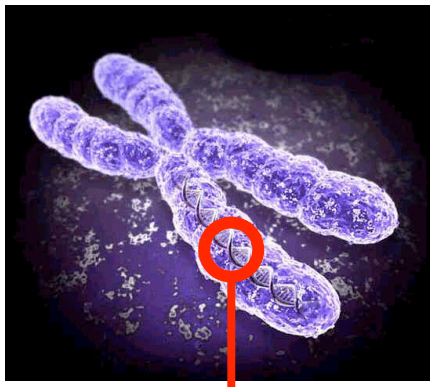
y1….yn : predictions

Models	Prediction	Assoc. Tests
• QTL, GWA	• General Linear Models	• ?
• Network	• Differential equations, etc,	• ?
• Ecophys	• Mathematical mess	• ?

Models	Prediction	Assoc. Tests
• QTL, GWA	• General Linear Models	• General Linear Models
• Network	• Differential equations, etc,	• ?
• Ecophys	• Mathematical mess	• ?

Models	Prediction	Assoc. Tests
• QTL, GWA	• General Linear Models	• General Linear Models
• Network	• Differential equations, etc,	• General Linear Models
• Ecophys	• Mathematical mess	• General Linear Models

A Parametric Correspondence Principle



Polymorphic locus

A1…An

Linear Estimation
GLM association test

Phenotypes: Y1….Yn

Model

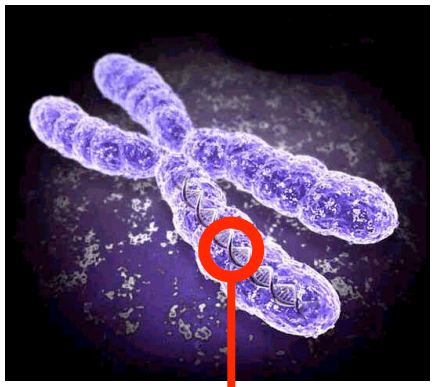
Some numerical parameter

Range of values in a set of mapping lines

2.71…3.14

y1….yn : predictions

A Parametric Correspondence Principle



Polymorphic locus

A1…An

Nonlinear Estimation
GLM association test

Model

Some numerical parameter

Range of values in a set of mapping lines

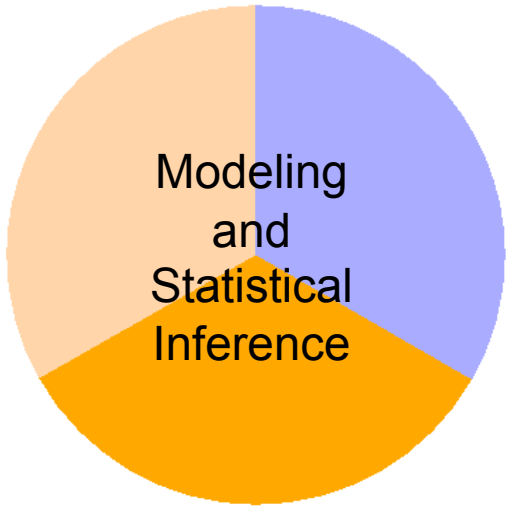
2.71…3.14

Phenotypes: Y1….Yn

y1….yn : predictions

Enables & Empowers

- Gene discovery
- Network inference
- Progeny phenotype prediction
- Etc.



Junk used to
make pictures