Templates, Anchors, and Matryoshkas

A Vision for Crop Systems Biology

Eberhard O. Voit
W. H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology

Crops *in silico*
Oxford, UK
June 25-28, 2017

Overview

Overall goals of crop science

Scientific and logistic issues

Challenge of complexity

Focus on mathematical modeling aspects

Describe landscape and strategies rather than solutions

No real technical details (left for other speakers)

Goal of Crop Science

Create better crops

Create crops with “better” yields

Better Crops? Not a New Idea!

Development of corn from teosinte

Successful but very slow! (9,000 years!)

Need a new approach

Crop Science Goals *versus* Evolution

Scientific community believes that:

- Evolution has increasingly improved organisms
- Organisms are optimal in comparison to older and current alternatives

If so: can we outfox nature and create “more optimal” organisms?

Microbes:

yes; have done it many times
optimize for a specific purpose
no need to withstand adverse conditions
make their surroundings optimal, expect better yields

Crops:

supposed to thrive in natural, erratic environments

New Goal of Crop Science

Create crops with “better” yields

…using computational models

Reconstruction of plants *in silicio*

Predictions of manipulation / optimizaton outcomes

Prediction of new phenotypes

Rational plant breeding toward better yields

Ideally: System Simulator

Analogy: Flight simulator

Train pilots in common tasks

Prepare pilots for unusual situations

Ideally: System Simulator

Analogy: Flight simulator

Train pilots in common tasks

Prepare pilots for unusual situations

Crop simulator

Try out ideas for crop improvement

Explore responses to perturbations

~~Scientific and Logistic Issues~~

Create community buy-in, common lexicon, data and model standards

Assemble, manage, share existing computational tools; model repositories

Create new computational models; automate model design

Establish accurate computational representations of reference and crop plants:
Arabidopsis thaliana, rice (*Oryza sativa*), maize (*Zea mays*), soybean (*Glycine max*), and cassava (*Manihot esculenta*)

Integrate crop models with other pertinent models

Risk assessments

Advance GUIs and visualization; maybe with HPC

Plant simulators

Rational, model-based plant breeding

Marshall-Colon *et al.*, 2017

Let’s Have a Great Conference!

Create community buy-in, common lexicon, data and model standards

Assemble, manage, share existing computational tools; model repositories

Create new computational models; automate model design

Establish accurate computational representations of reference and crop plants:
Arabidopsis thaliana, rice (*Oryza sativa*), maize (*Zea mays*), soybean (*Glycine max*), and cassava (*Manihot esculenta*)

Integrate crop models with other pertinent models

Risk assessments

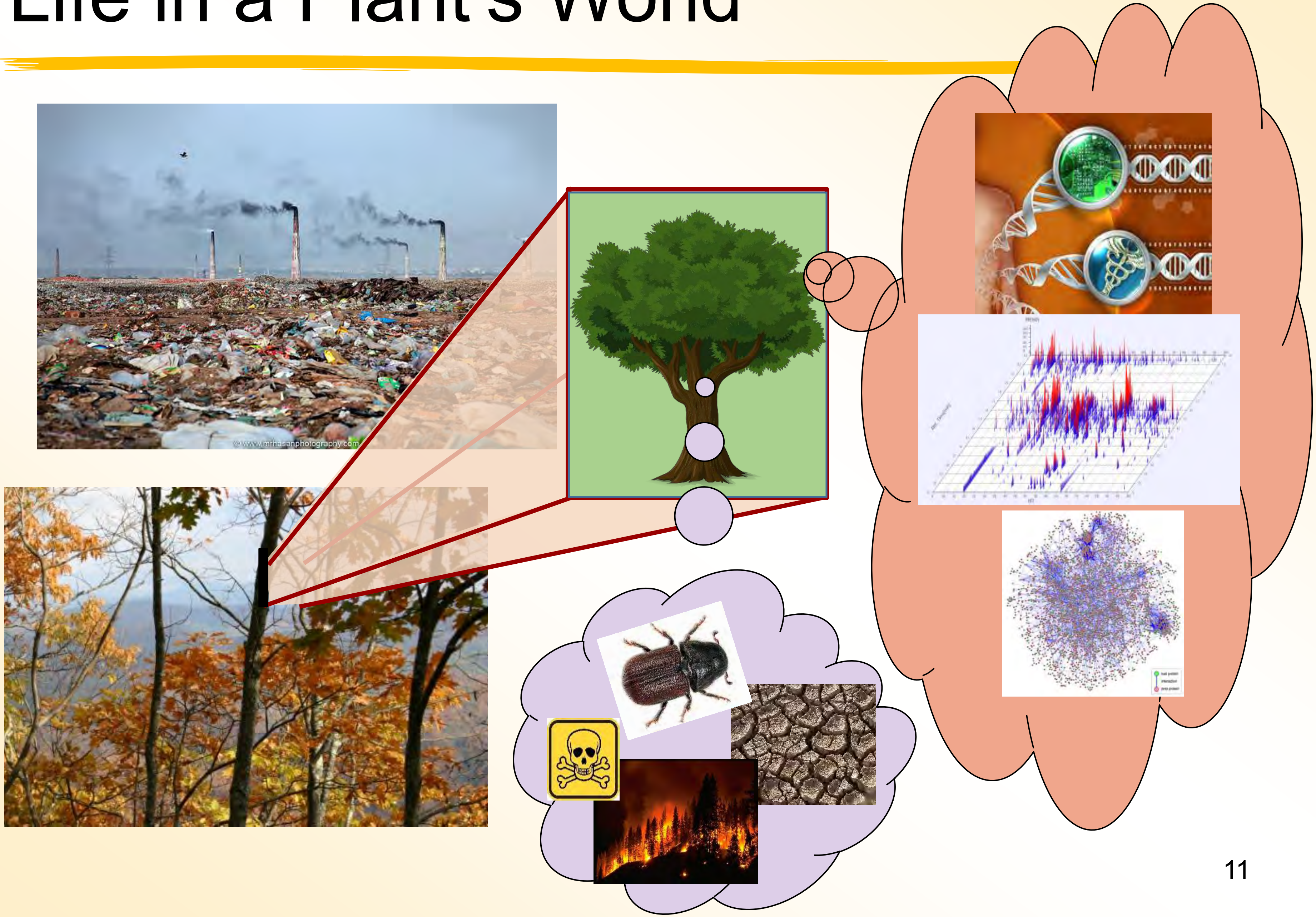
Advance GUIs and visualization; maybe with HPC

Plant simulators

Rational, model-based plant breeding

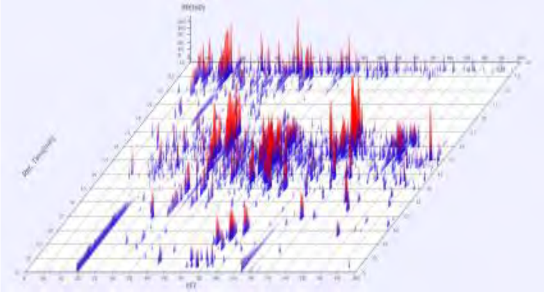
Marshall-Colon *et al.*, *Front. Plant Sci.*, 2017

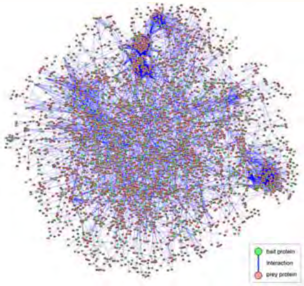
Life in a Plant’s World



Ideally, We Would…

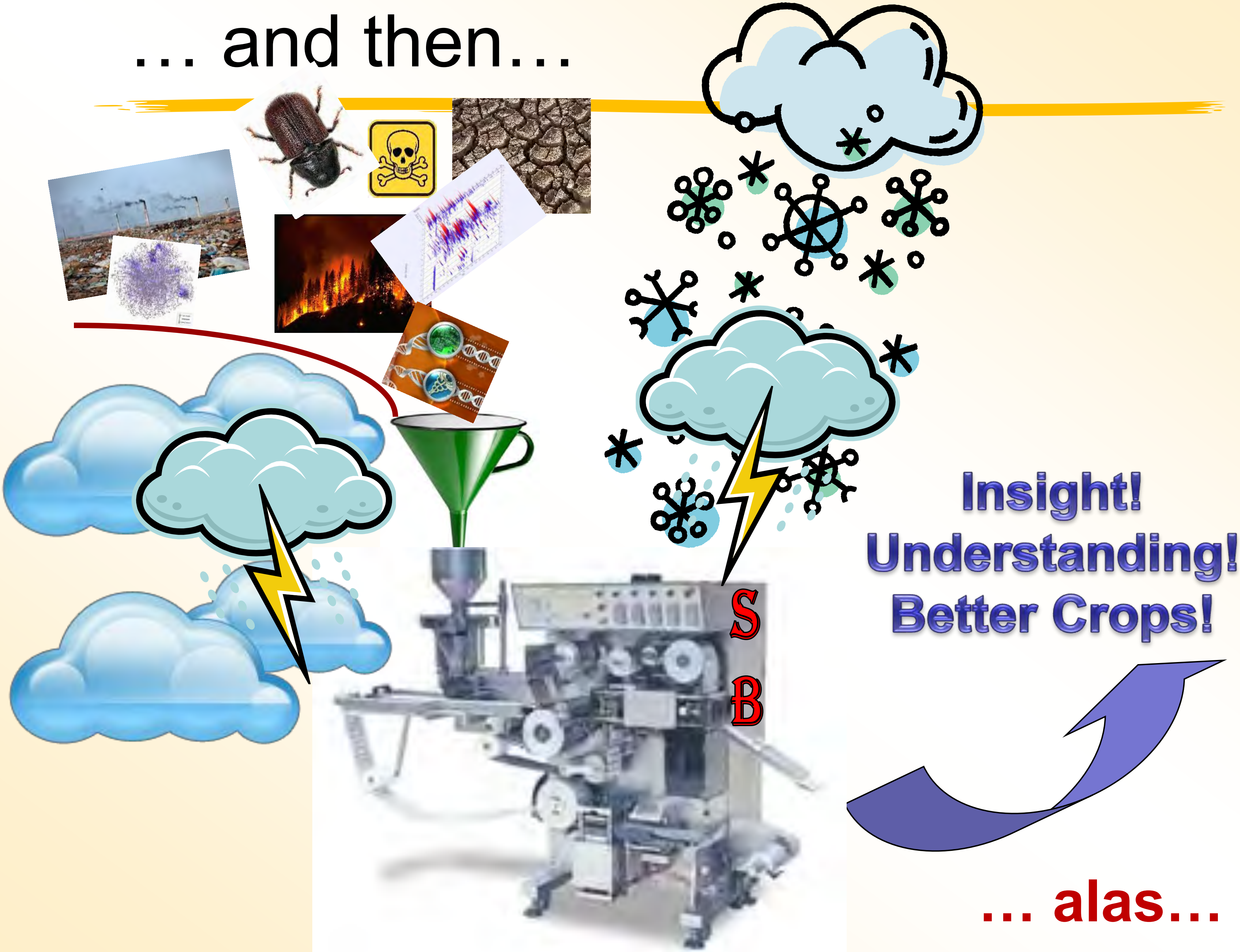
measure,
measure,
measure,
…
quantify,
quantify,
quantify,
…





… and then…

… and then…



Insight!
Understanding!
Better Crops!

… alas…

alas…We Live in the Real World

… with a lot of limitations

Do not know how to measure many aspects

Don't even know what exactly to measure

Percentage of measurable quantities almost negligible in comparison with what is affecting plant growth

Statistical analysis is great for associations, but has hard time with causality

Dynamic modeling in plant and crop science
is barely in its infancy (more like, prenatal)
e.g., nobody has a good handle on
multi-scale systems analysis

Why are we not much further along?

fact: It's Complicated!

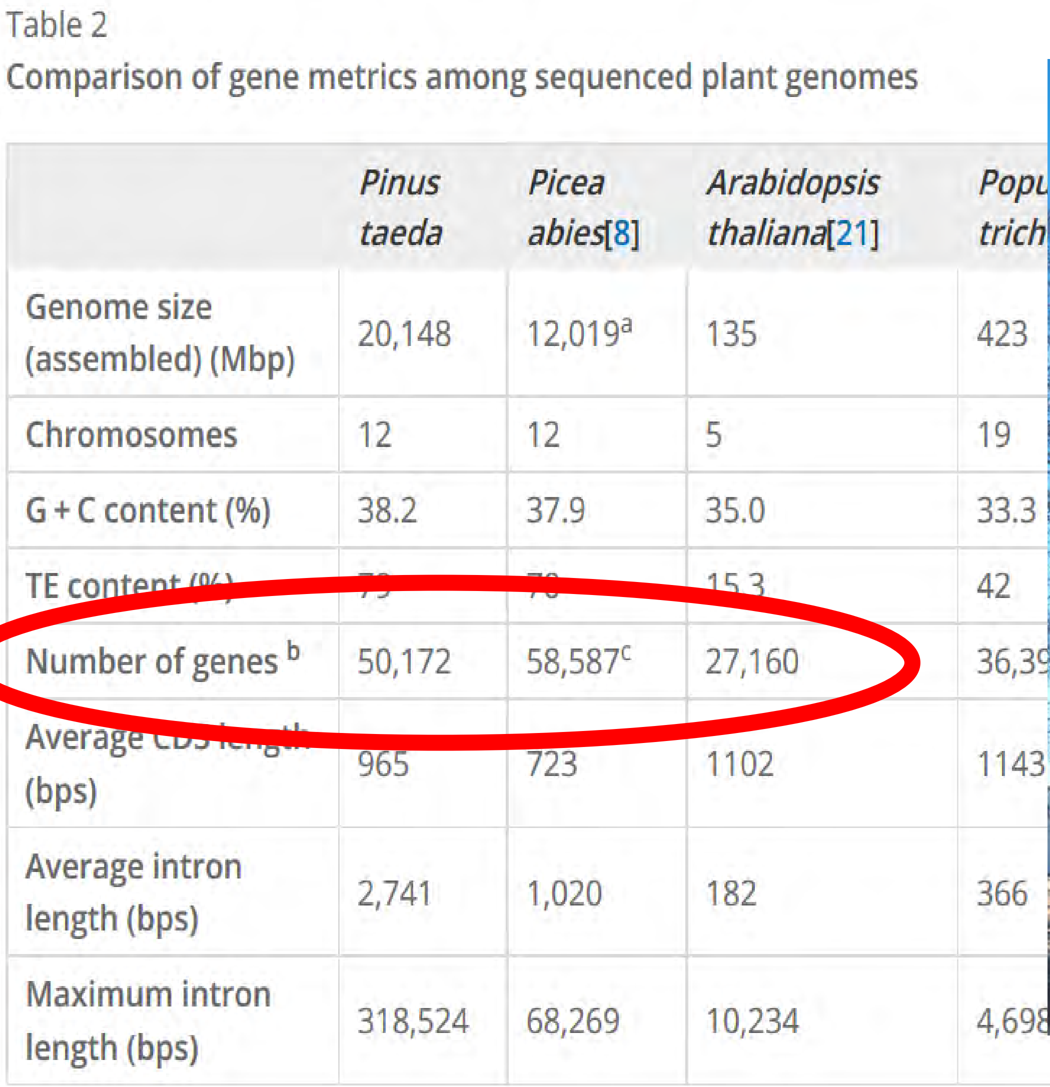


Table 2

Comparison of gene metrics among sequenced plant genomes

	Pinus taeda	*Picea abies*[8]	*Arabidopsis thaliana*[21]	*Popu trich*
Genome size (assembled) (Mbp)	20,148	12,019[a]	135	423
Chromosomes	12	12	5	19
G + C content (%)	38.2	37.9	35.0	33.3
TE content (%)	79	70	15.3	42
Number of genes [b]	50,172	58,587[c]	27,160	36,39
Average CDS length (bps)	965	723	1102	1143
Average intron length (bps)	2,741	1,020	182	366
Maximum intron length (bps)	318,524	68,269	10,234	4,698

D. Neale *et al.*, *Genome Biology* 2014 15:R59

fact: Plants are Complicated!

B10NUMB3R5

THE DATABASE OF USEFUL BIOLOGICAL NUMBERS

Estimated number of metabolites in the plant kingdom:

200,000-1,000,000

Are Large Numbers the Only Issue to be Tackled?

No, many other factors …

Other Factors …

… related to inherent *complexity of biological systems*

Vastly different temporal and spatial scales

Simultaneous operation at different levels

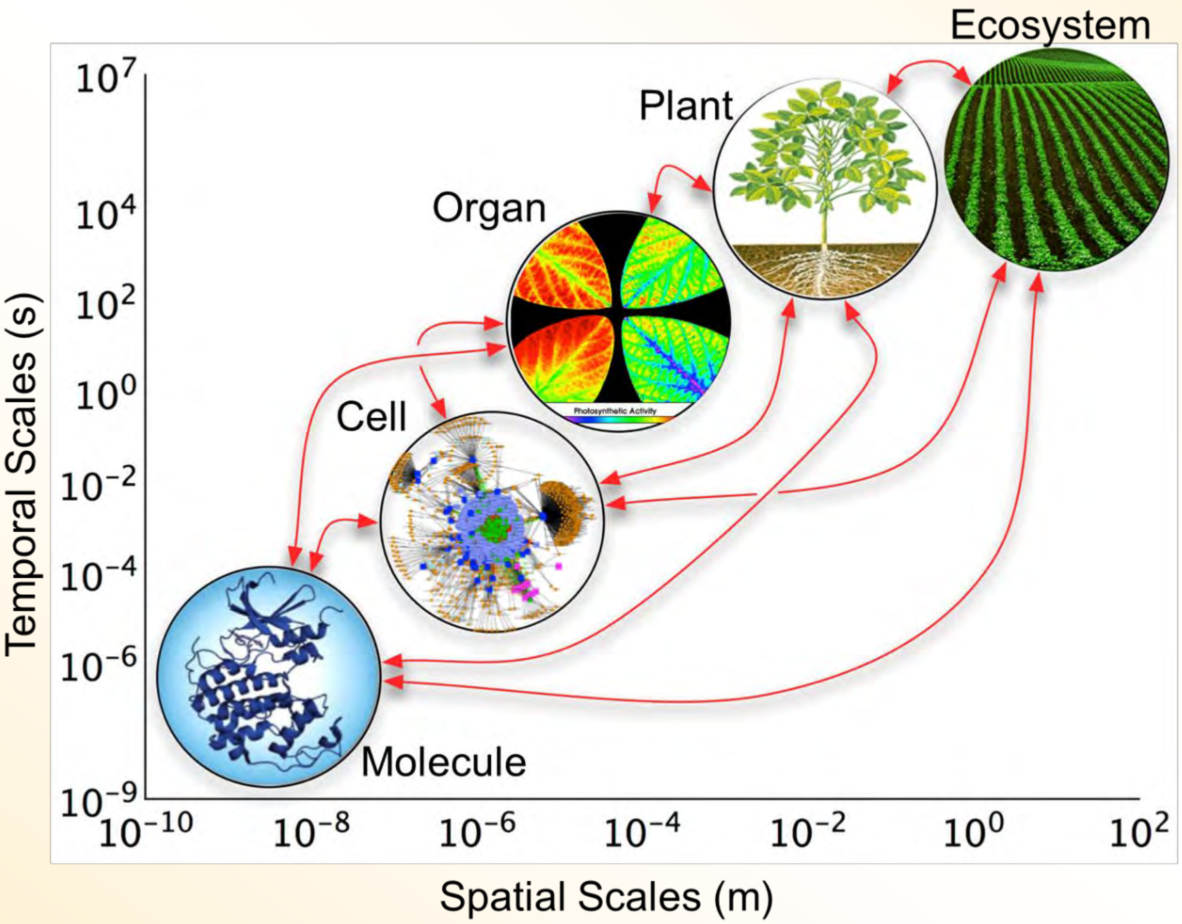
Large numbers of components

Large number of processes

Processes are nonlinear

Quantitative changes in parameters cause qualitative changes in response

Wide Range of Scales



Marshall-Colon *et al.*, *Front. Plant Sci.*, 2017

Other Factors …

… are related to inherent *complexity of biological systems*

Vastly different temporal and spatial scales

Simultaneous operation at different levels

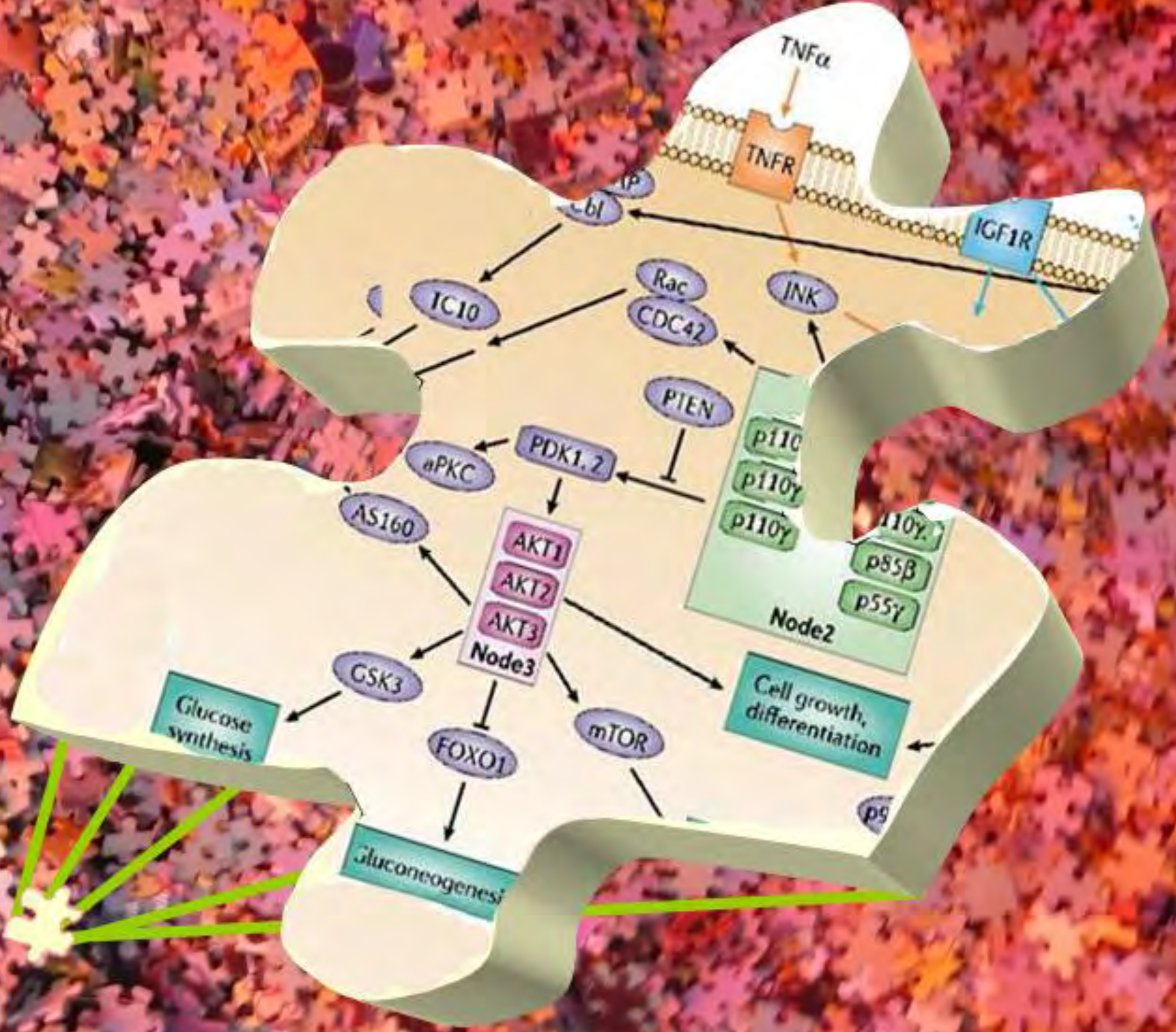
Large numbers of components

Large number of processes

Processes are nonlinear

Quantitative changes in parameters cause
qualitative changes in response

Hierarchy in Systems Biology


TNFα
TNFR
IGF1R
Cbl
TC10
Rac
CDC42
JNK
PTEN
PDK1, 2
aPKC
AS160
AKT1
AKT2
AKT3
Node3
p110
p110γ
p110γ
p110γ
p85β
p55γ
Node2
GSK3
Glucose synthesis
FOXO1
mTOR
Cell growth, differentiation
Gluconeogenesi

Other Challenges

… are related to inherent *complexity of biological systems*

Vastly different temporal and spatial scales

Simultaneous operation at different levels

Large numbers of components ✓

Large number of processes

Processes are nonlinear

Quantitative changes in parameters cause qualitative changes in response

Linear *versus* Nonlinear

Linear Scenario:

Invest $100 in company Surebet

→ be lucky → return $750

Invest $10,000 in company Surebet

→ be lucky → return $75,000

Linear *versus* Nonlinear

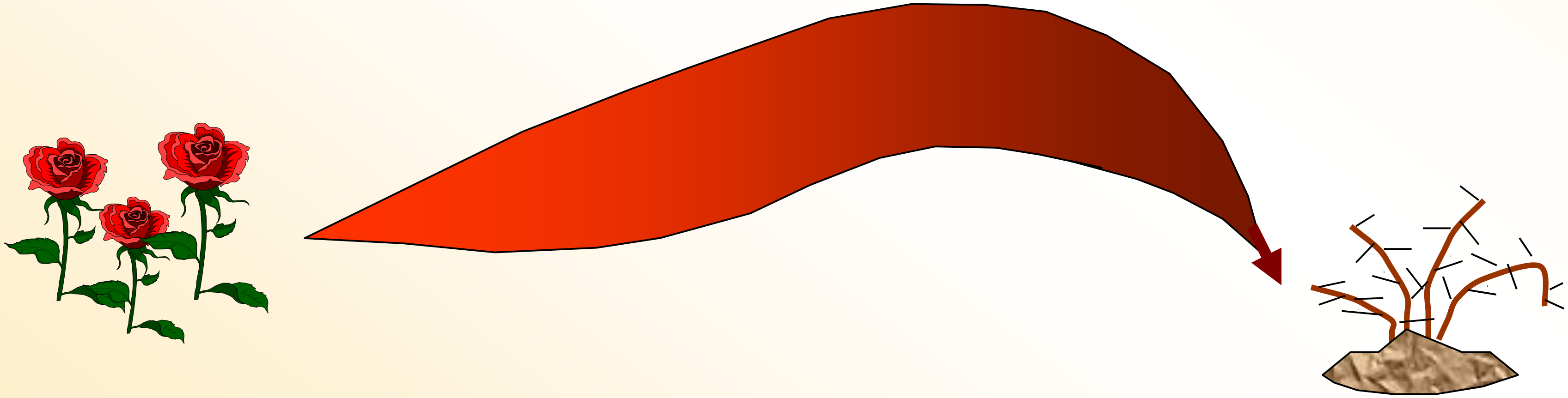
Nonlinear Scenario:

Put 2 tablespoons of fertilizer on your roses

→ be lucky → roses produces 20 blossoms

Put 200 tablespoons of fertilizer on your roses

→ roses do not produce 2,000 blossoms



Other Challenges

… are related to inherent *complexity of biological systems*

Vastly different temporal and spatial scales

Simultaneous operation at different levels

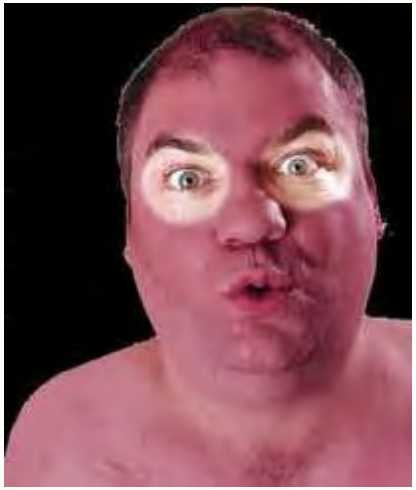
Large numbers of components

Large number of processes

Processes are nonlinear

Quantitative changes in parameters cause qualitative changes in response

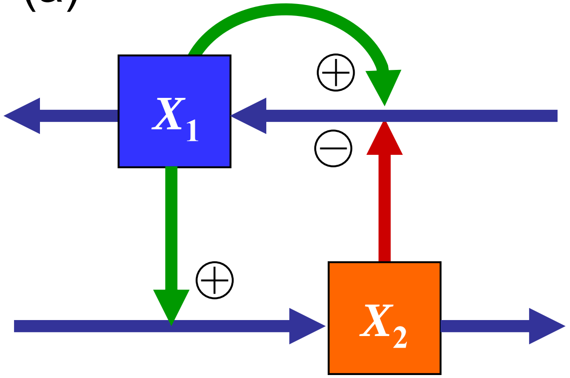
A Threshold in a Nonlinear System



malariaworld.org; forbes.com; futurity.irg; kre8ivity.blogspot.com; thelabelguy.com; sonasmedspa.com

Quantitative ~ Qualitative Changes

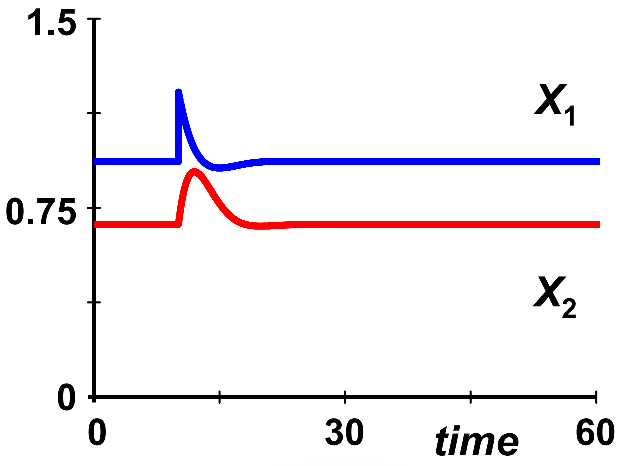
(a)



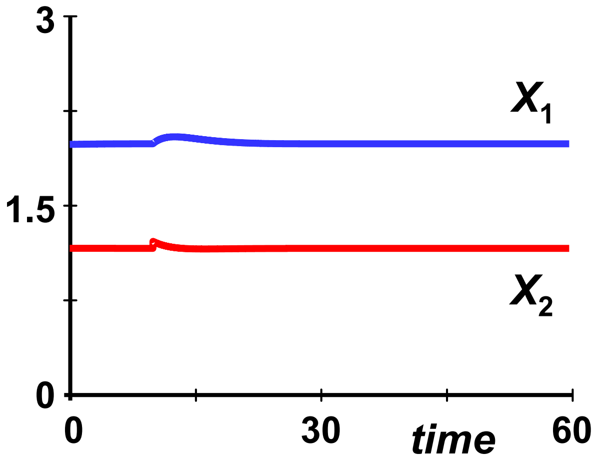
(b)

$$\dot{X}_1 = \alpha_1 X_1^{0.4} X_2^{-0.15} - X_1^{h}$$

$$\dot{X}_2 = X_1 - X_2^{0.2}$$



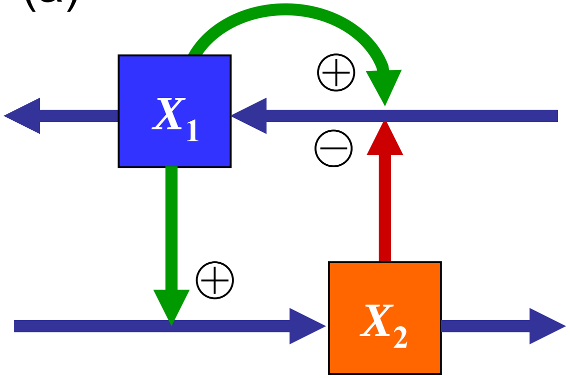
α_1 = 0.9, *h* = 1

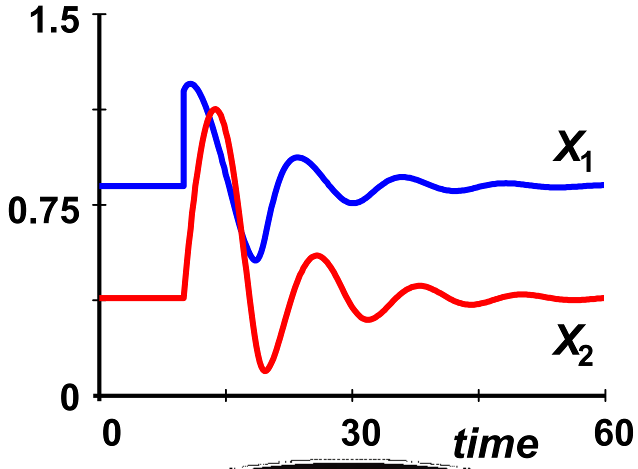


α_1 = 1.2, *h* = 1

Quantitative ~ Qualitative Changes

(a)





$\alpha_1 = 0.9, h = 0.2$

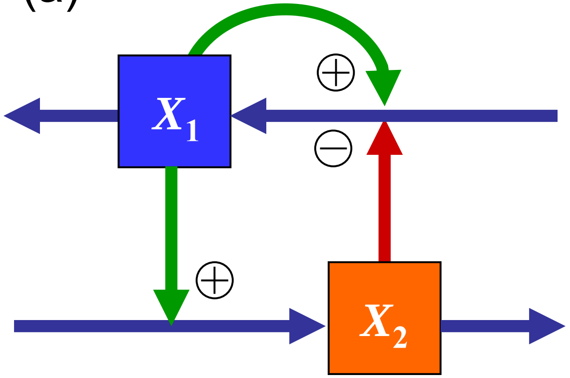
(b)

$$\dot{X}_1 = \alpha_1 X_1^{0.4} X_2^{-0.15} - X_1^{h}$$

$$\dot{X}_2 = X_1 - X_2^{0.2}$$

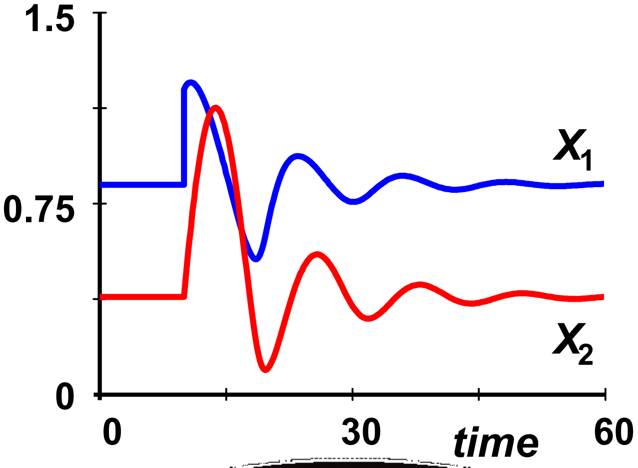
Quantitative ~ Qualitative Changes

(a)

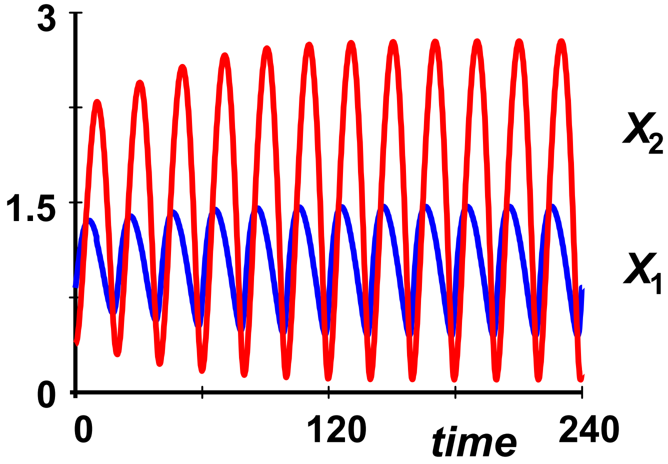


(b)

$$\dot{X}_1 = \alpha_1 X_1^{0.4} X_2^{-0.15} - X_1^{h}$$
$$\dot{X}_2 = X_1 - X_2^{0.2}$$



$\alpha_1 = 0.9,\ h = 0.2$



$\alpha_1 = 1.02,\ h = 0.2$

Is That All?

Is **reconstruction** of plants *in silicio* sufficient?

Possibly yes, but…

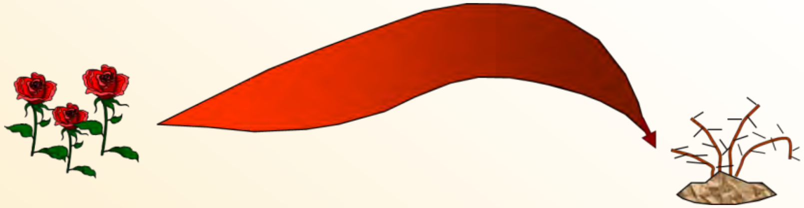
… goal of crop science is to change existing crops toward “better” crops

(crops with better yields)

This goal requires significant extensions of existing processes.

That in turn requires that process descriptions must allow extrapolations

Recall extrapolation of linear relationships!

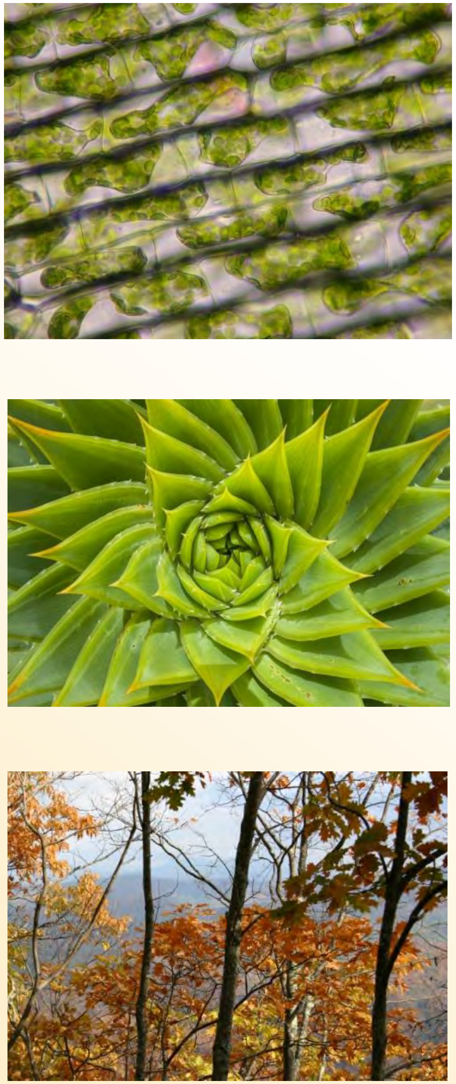


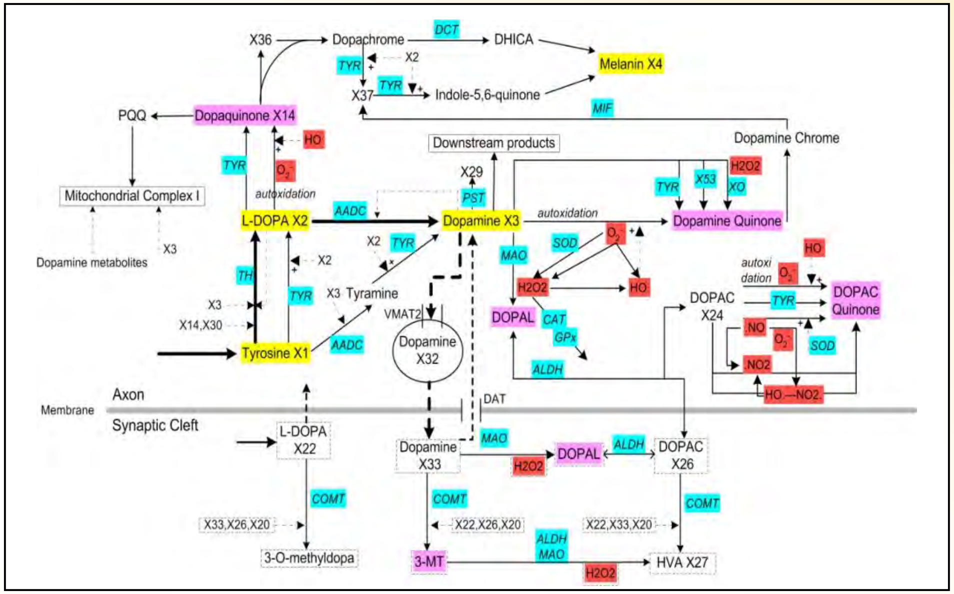
So Many Overwhelming Issues!

What to do?

ystems Biology to the Rescue!

Generic Approach of CSB



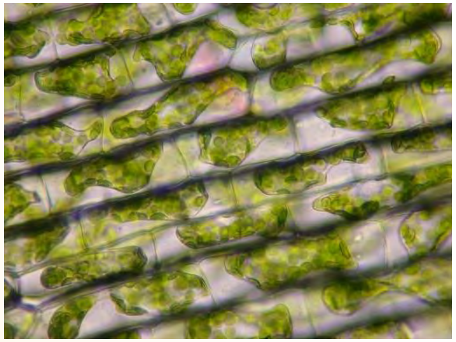


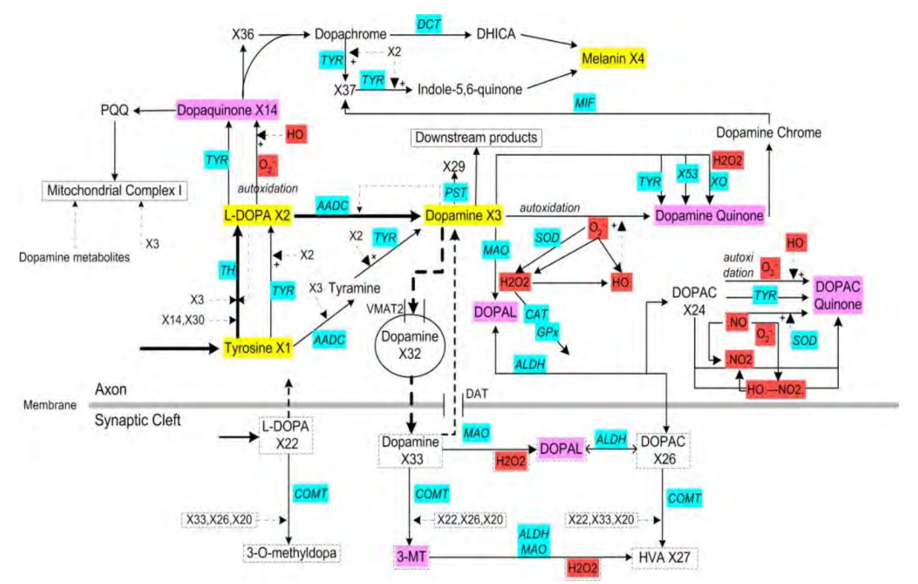
$$\frac{dX_{12}}{dt} = \gamma_{0,12} X_{24}^{f_{0,12,24}} X_{30}^{f_{0,12,30}} + \gamma_{51,12} X_{12}^{f_{51,12,12}} X_{28}^{f_{51,12,28}} X_{51}^{f_{51,12,51}} X_{52}^{f_{51,12,52}}$$

$$+ \gamma_{4,17} X_{4}^{f_{4,17,4}} X_{22}^{f_{41,17,22}} + \gamma_{6,17} X_{6}^{f_{6,17,6}} X_{22}^{f_{6,17,22}}$$

```
% model_ode
function dy = PP_ode(t,y)
...
dX12 = gam0_12 * X24^f0_12_24 * X30^f0_12_30  + ...
                gam4_17 * X4^f4_12_4 * X22^f41_17_22 +
                ...
end
```

Generic Approach of CSB



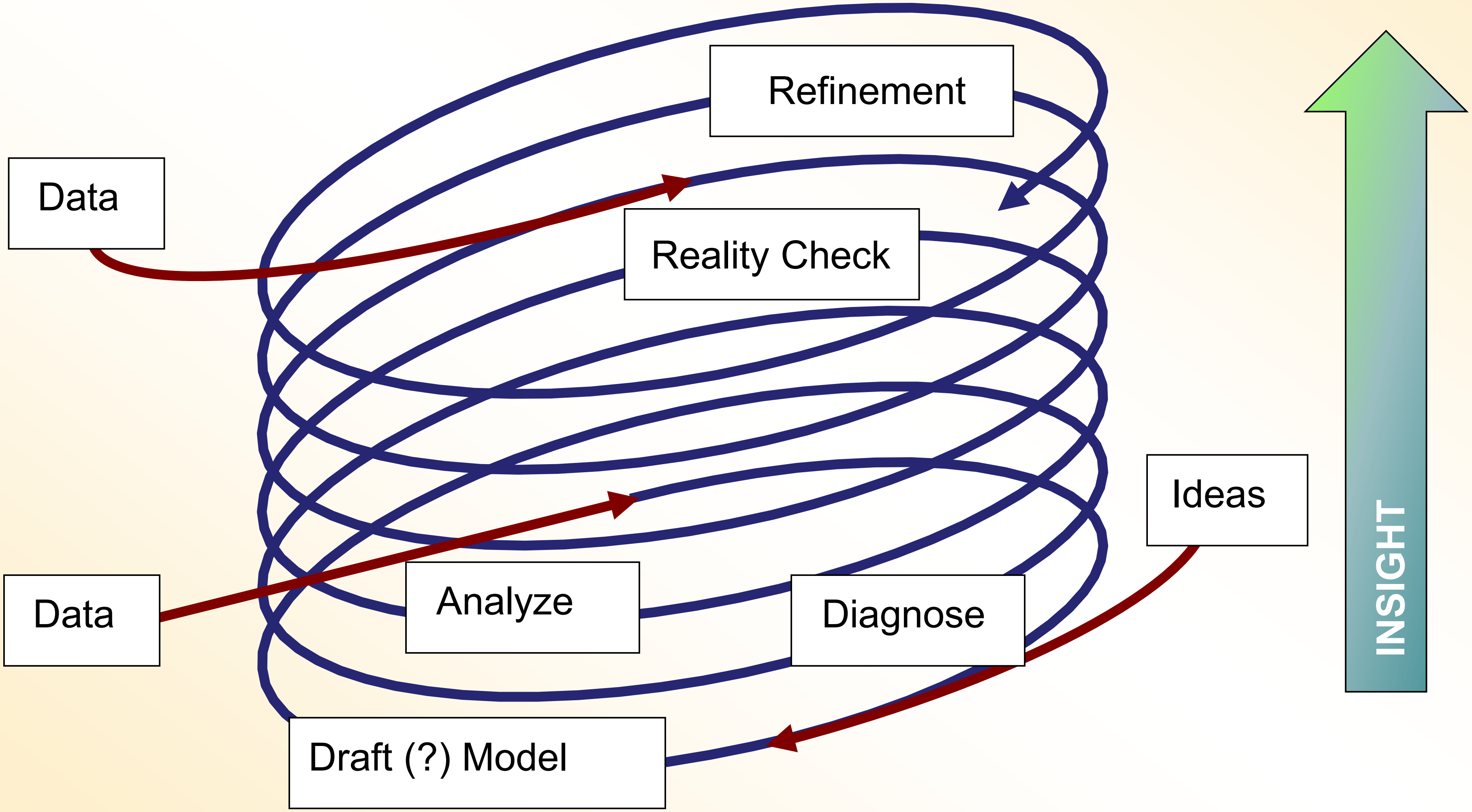


$$\frac{dX_{12}}{dt} = \gamma_{0,12} X_{24}^{f_{0,12,24}} X_{30}^{f_{0,12,30}} + \gamma_{51,12} X_{12}^{f_{51,12,12}} X_{28}^{f_{51,12,28}} X_{51}^{f_{51,12,51}} X_{52}^{f_{51,12,52}}$$

$$+ \gamma_{4,17} X_4^{f_{4,17,4}} X_{22}^{f_{41,17,22}} + \gamma_{6,17} X_6^{f_{6,17,6}} X_{22}^{f_{6,17,22}}$$

```
% model_ode
function dy = PP_ode(t,y)
…
dX12 = gam0_12 * X24^f0_12_24 * X30^f0_12_30  + …
                  gam4_17 * X4^f4_12_4 * X22^f41_17_22 +
                  …
end
```

Generic, Iterative Modeling Process


Refinement

Data

Reality Check

Ideas

Data

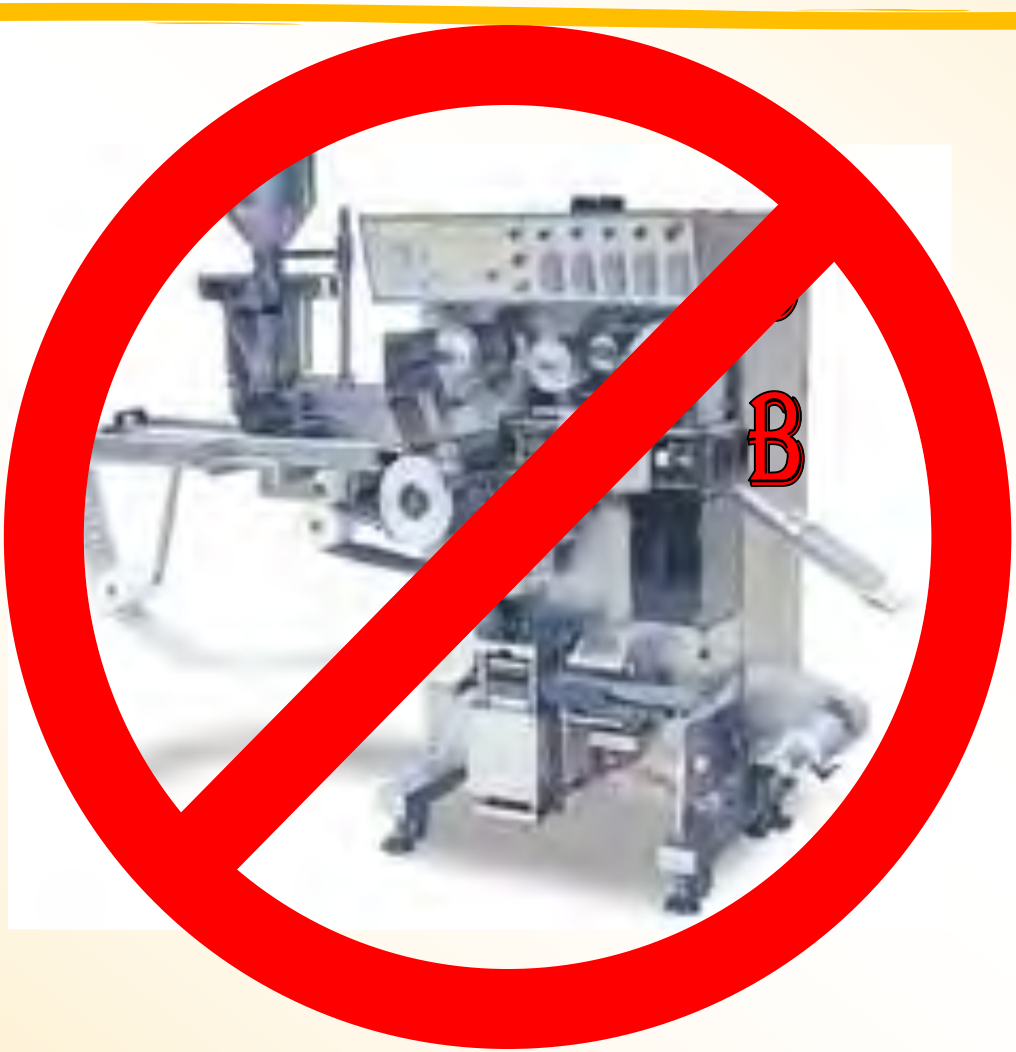
Analyze

Diagnose

Draft (?) Model

INSIGHT

Sorry, Magic Machine Not Available



What Kinds of CS and Math are Needed?

Algorithms to sift through "BigData"

- Sophisticated statistics
- Machine learning
- Artificial intelligence
- Visualization
- Clustering

Mathematics

- Linear algebra
- Calculus
- Differential equations
- Numerical methods
- Dynamical systems
- Optimization
- Modeling

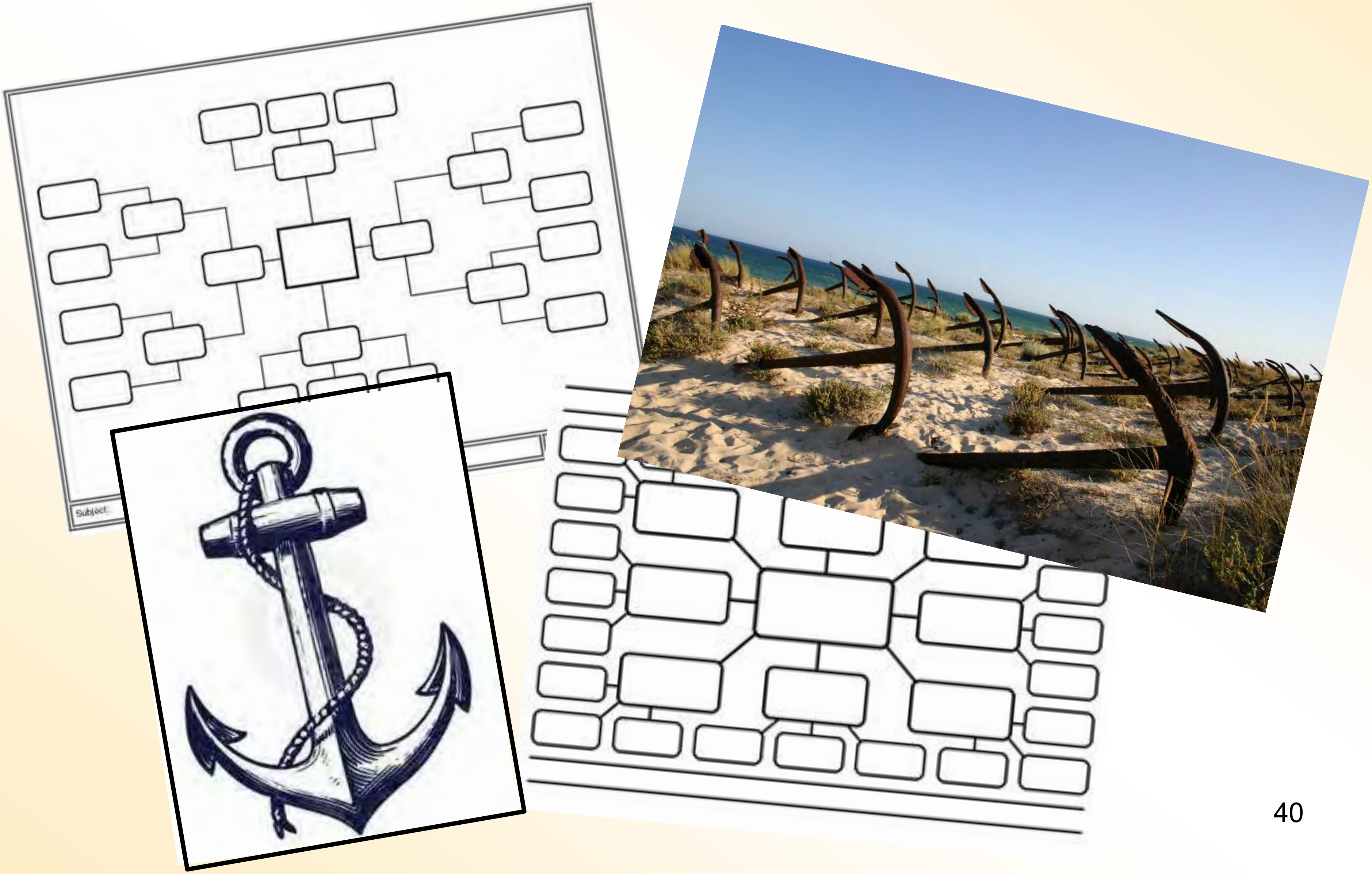
Computer Science

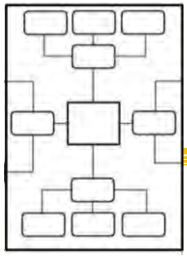
- Algorithm development
- Parallelization

Bioinformatics

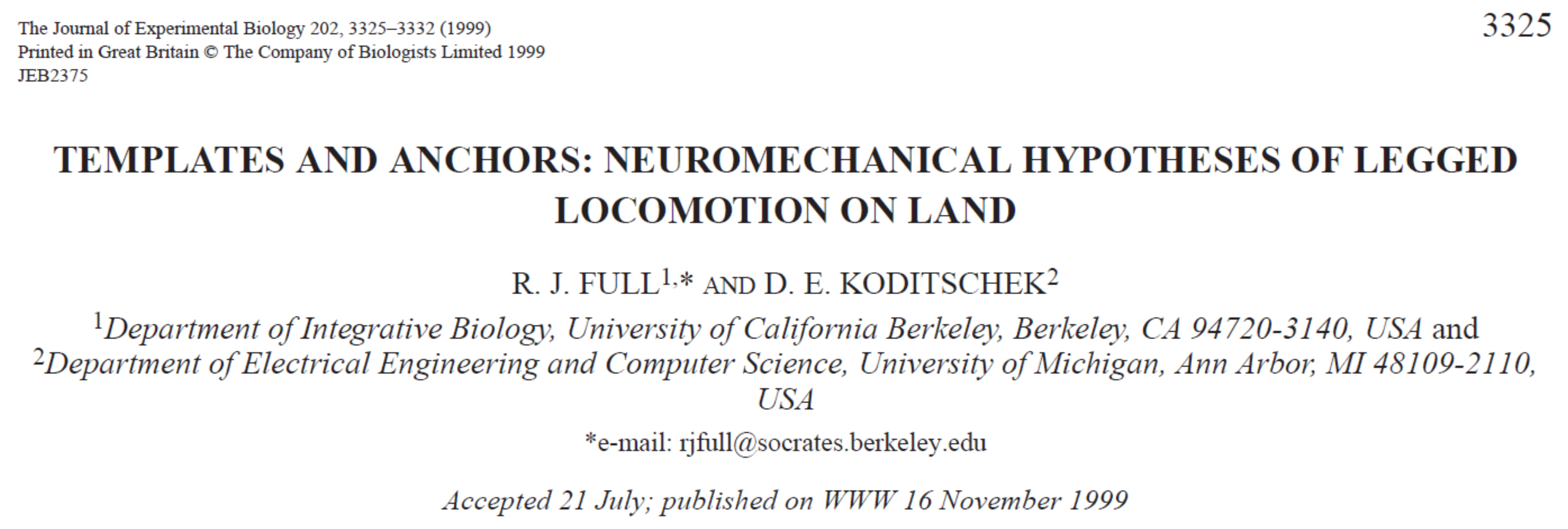
Computational Systems Biology

Templates and Anchors





Templates and Anchors



The Journal of Experimental Biology 202, 3325–3332 (1999)
Printed in Great Britain © The Company of Biologists Limited 1999
JEB2375

3325

TEMPLATES AND ANCHORS: NEUROMECHANICAL HYPOTHESES OF LEGGED LOCOMOTION ON LAND

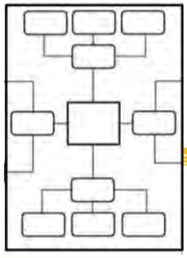
R. J. FULL[1,*] AND D. E. KODITSCHEK[2]

[1]*Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720-3140, USA* and [2]*Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2110, USA*

*e-mail: rjfull@socrates.berkeley.edu

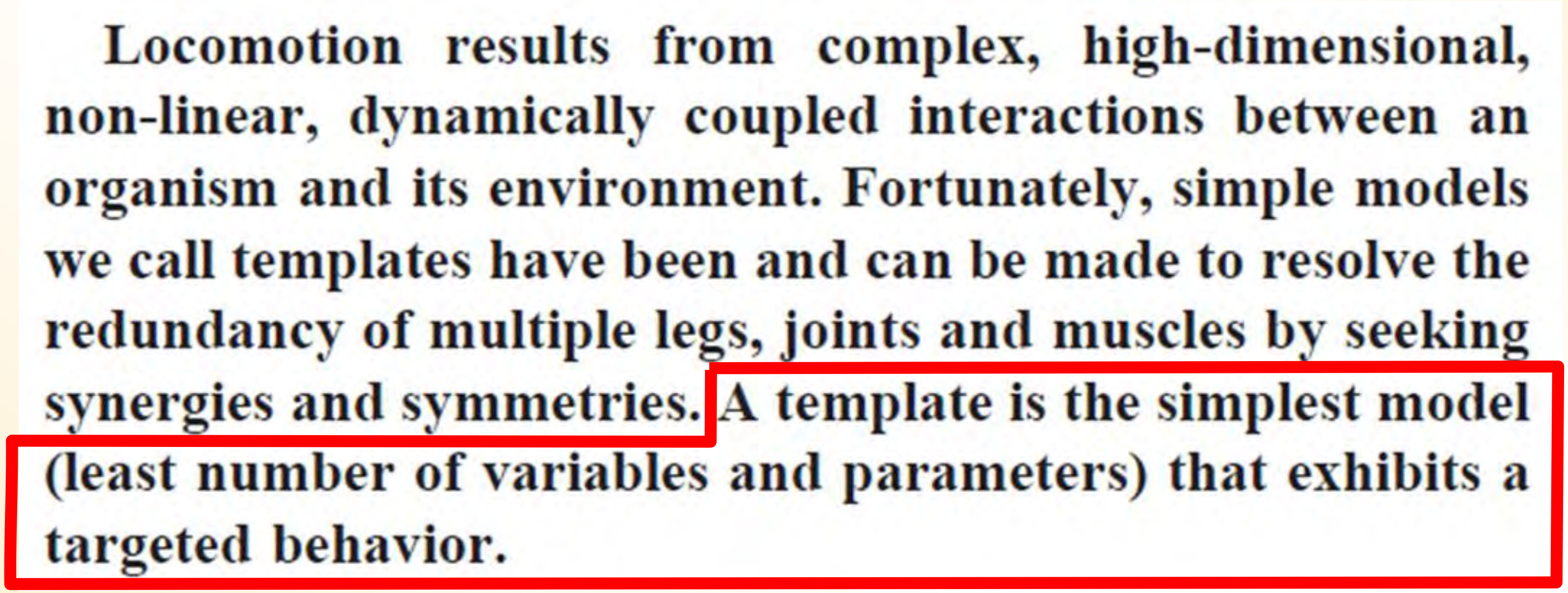
Accepted 21 July; published on WWW 16 November 1999

Strategy: Divide complex modeling task into a coarse, framework model and more detailed, finer-grained models of subsystems.

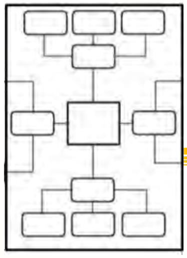


Templates and Anchors

From Abstract:

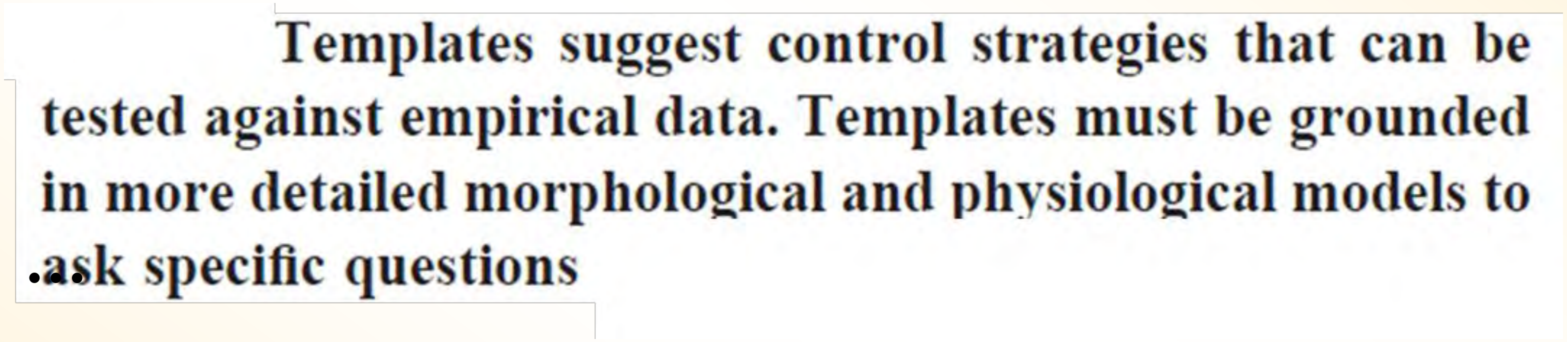


Locomotion results from complex, high-dimensional, non-linear, dynamically coupled interactions between an organism and its environment. Fortunately, simple models we call templates have been and can be made to resolve the redundancy of multiple legs, joints and muscles by seeking synergies and symmetries. A template is the simplest model (least number of variables and parameters) that exhibits a targeted behavior.

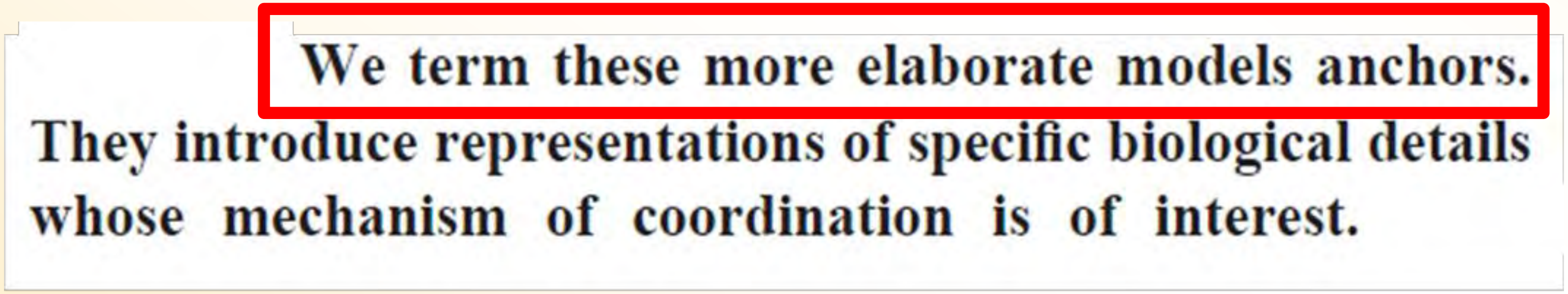


Templates and Anchors

From Abstract:



Templates suggest control strategies that can be tested against empirical data. Templates must be grounded in more detailed morphological and physiological models to …ask specific questions



We term these more elaborate models anchors. They introduce representations of specific biological details whose mechanism of coordination is of interest.

Templates, Anchors, and Matryoshkas

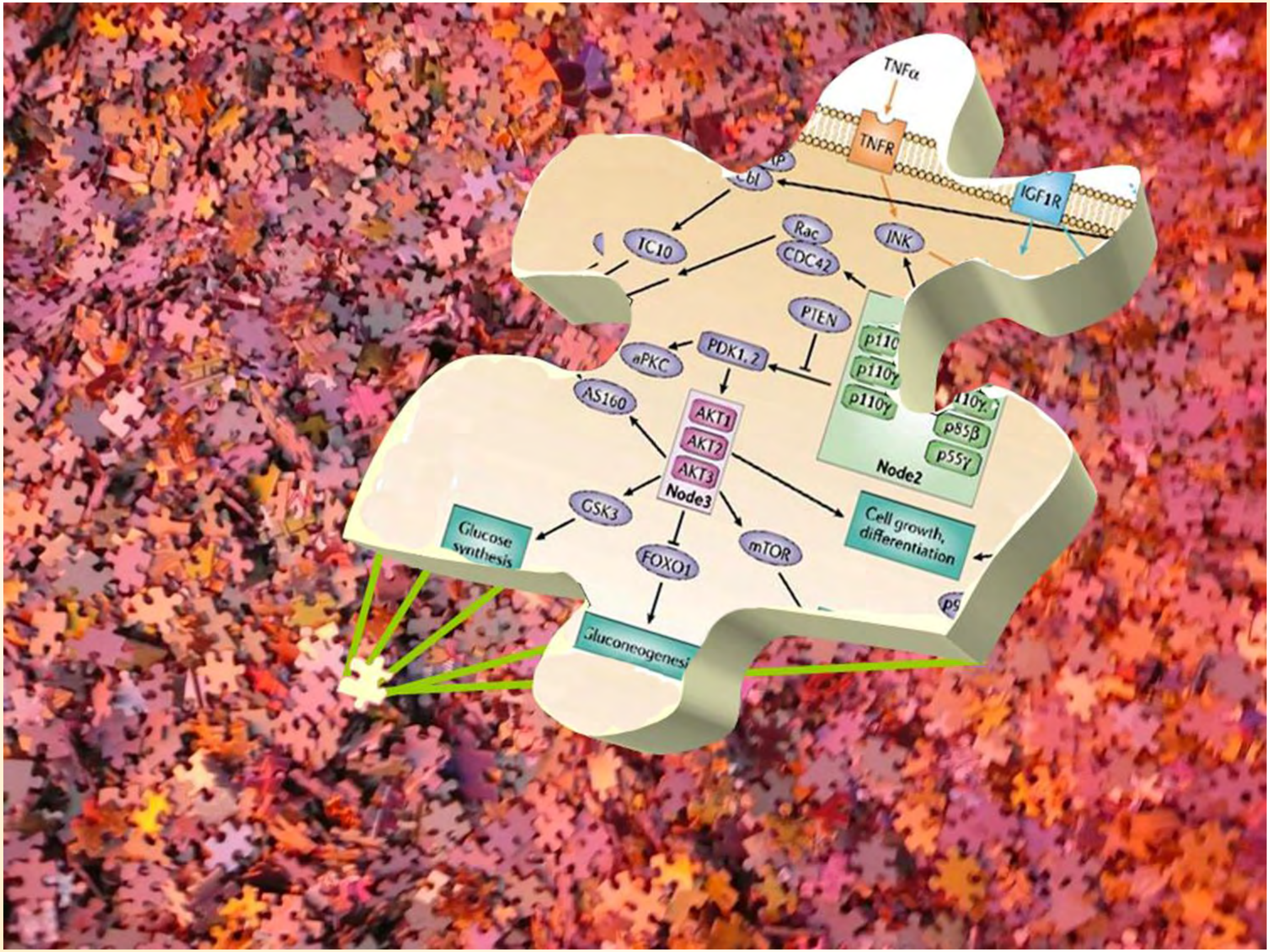
What about Matryoshkas?

First матрешка: Allegedly Sergei Malyutin and Vasily Zvyozdochkin ~1890

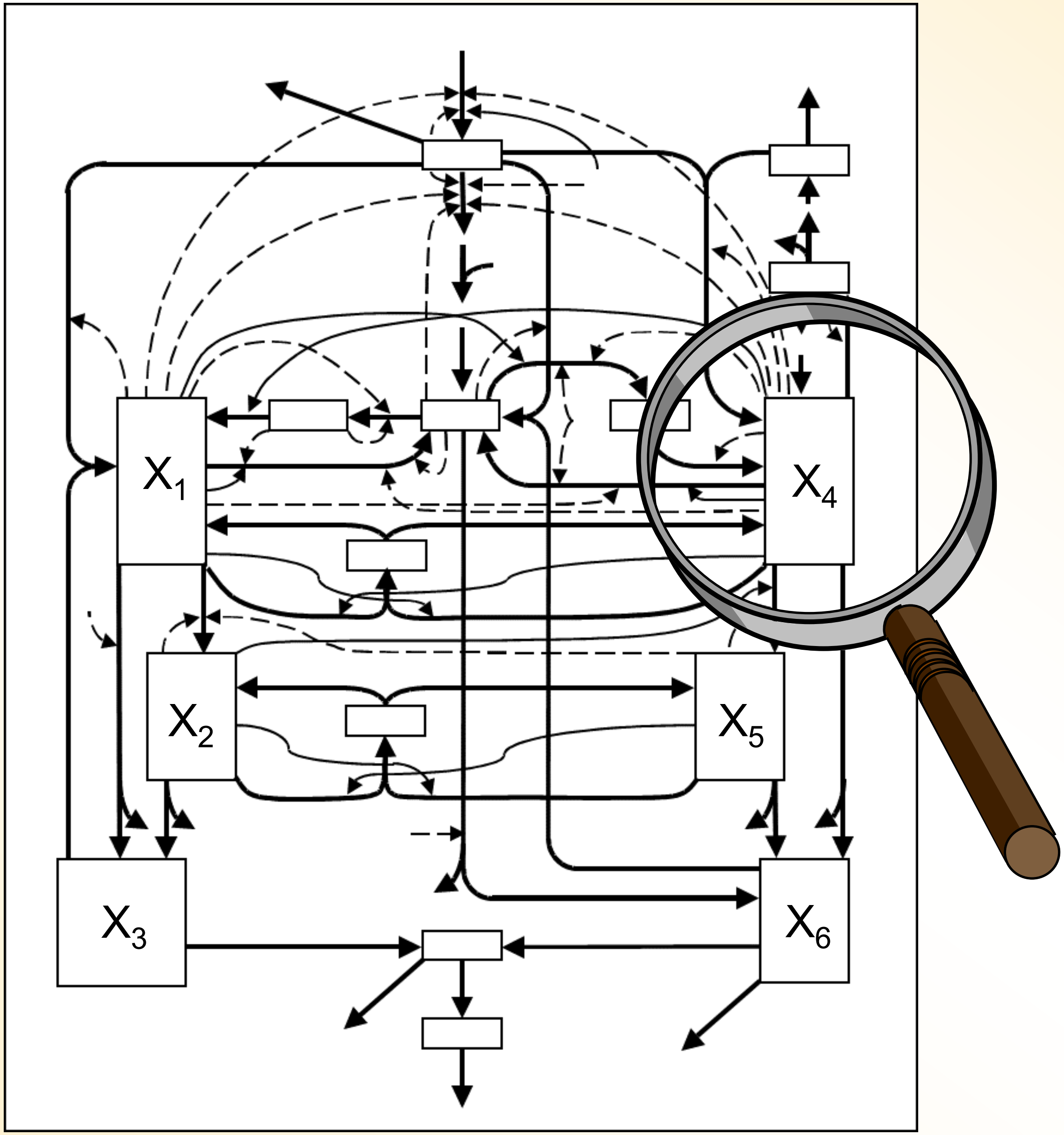
Templates, Anchors, and Matryoshkas

What about Matryoshkas? … or maybe better for crop science: Onion peeling?

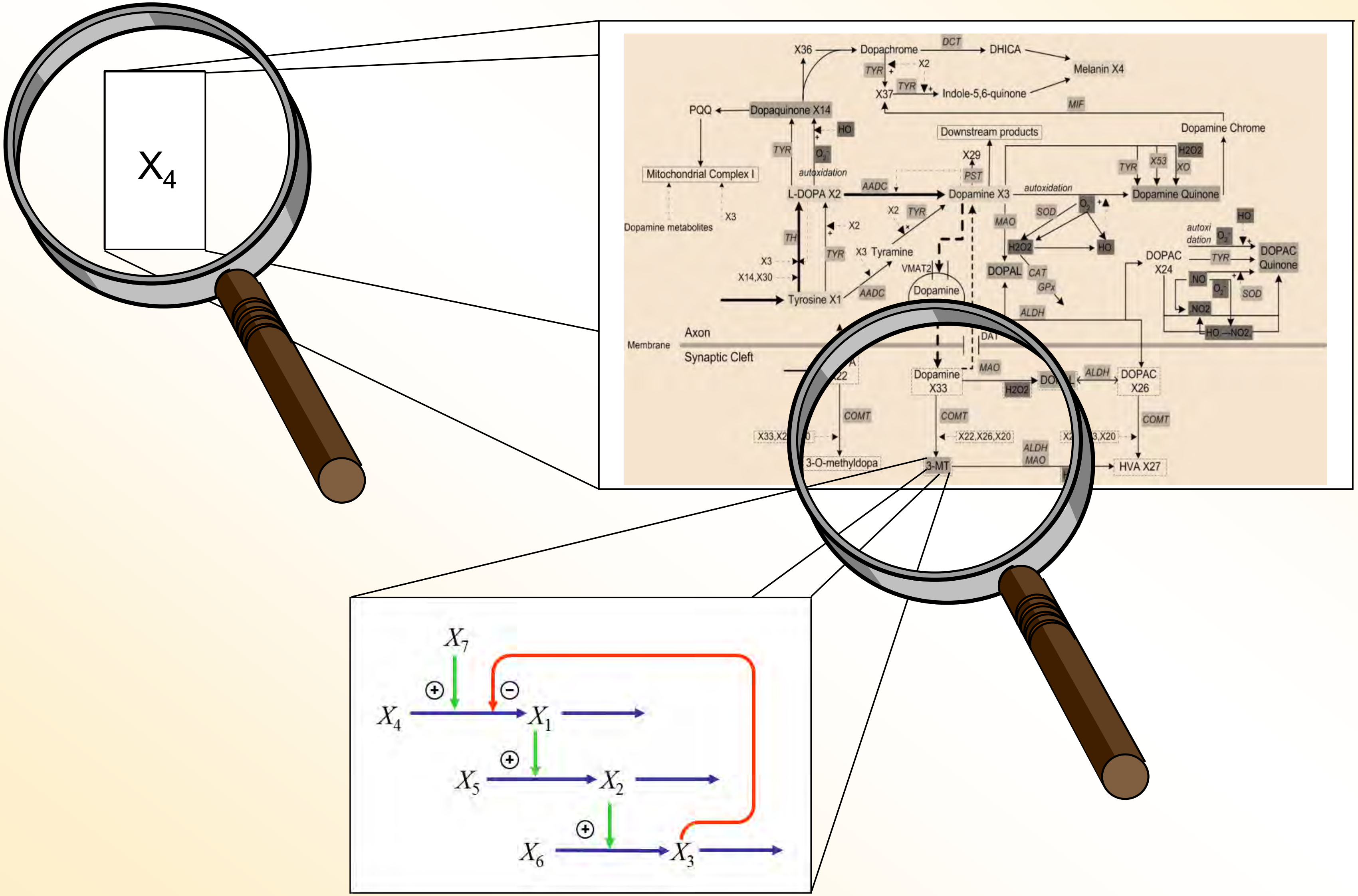
Matryoshkas: Recall Hierarchies



Matryoshkas Hidden in Most Models



Matryoshkas Hidden in Most Models



Templates, Anchors, and Matryoshkas

Example:

Discrete agent-based model as template

Template allows:

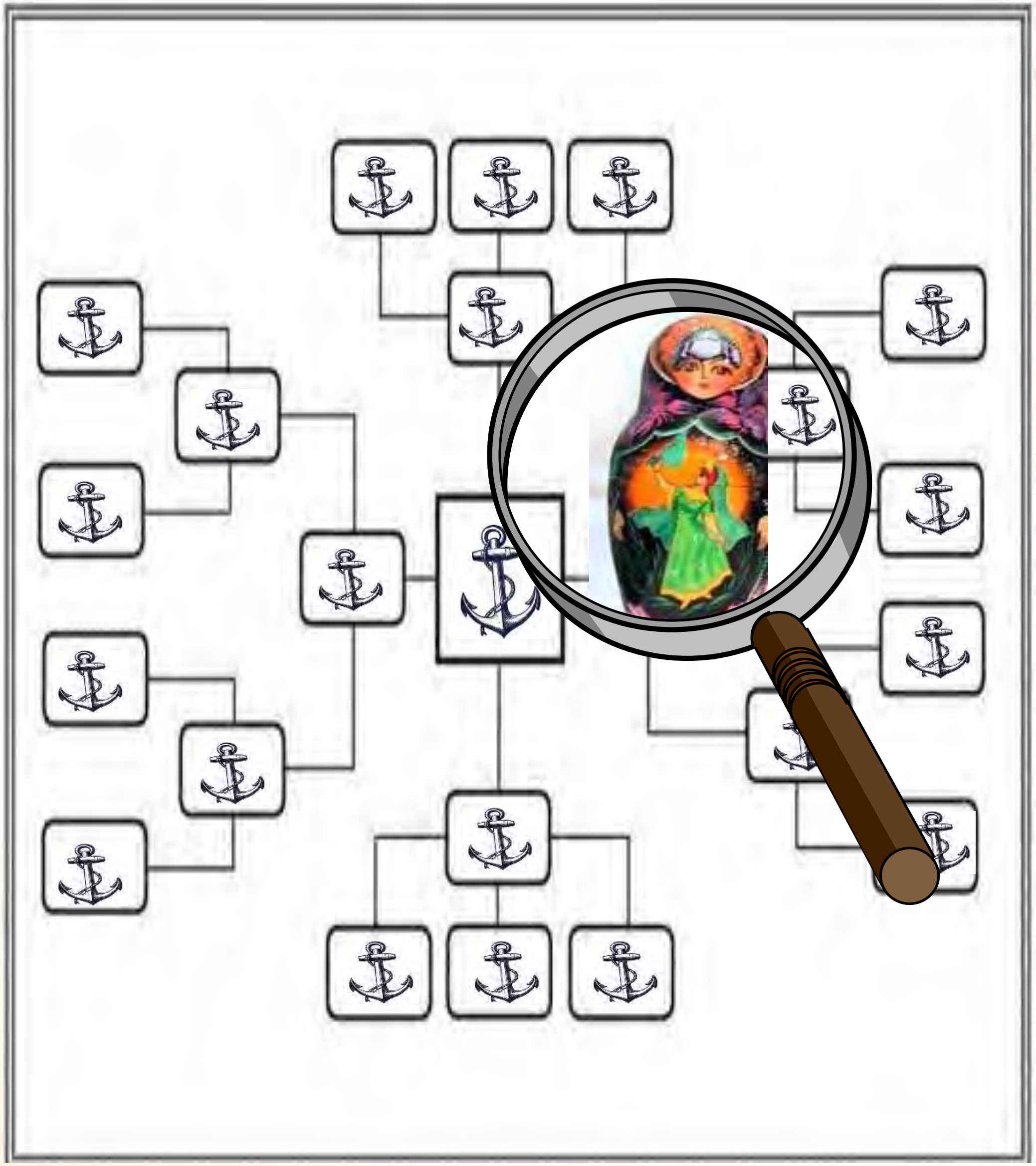
Stochasticity

Discrete events

Autonomous ODE models as anchors

Each anchor model is matryoshka

Different time scales



Templates, Anchors, and Matryoshkas in Crop Science

Vegetation response to Atmospheric and Climate Change) at the University of Illinois.

WIMOVAC is a general model applicable to a wide range of vegetation and soil types and importantly an experimental tool that managers, students and experimentalists can use easily. Its heart is a modular mathematical model of the carbon balance of vegetation and allows prediction of responses to climate change.

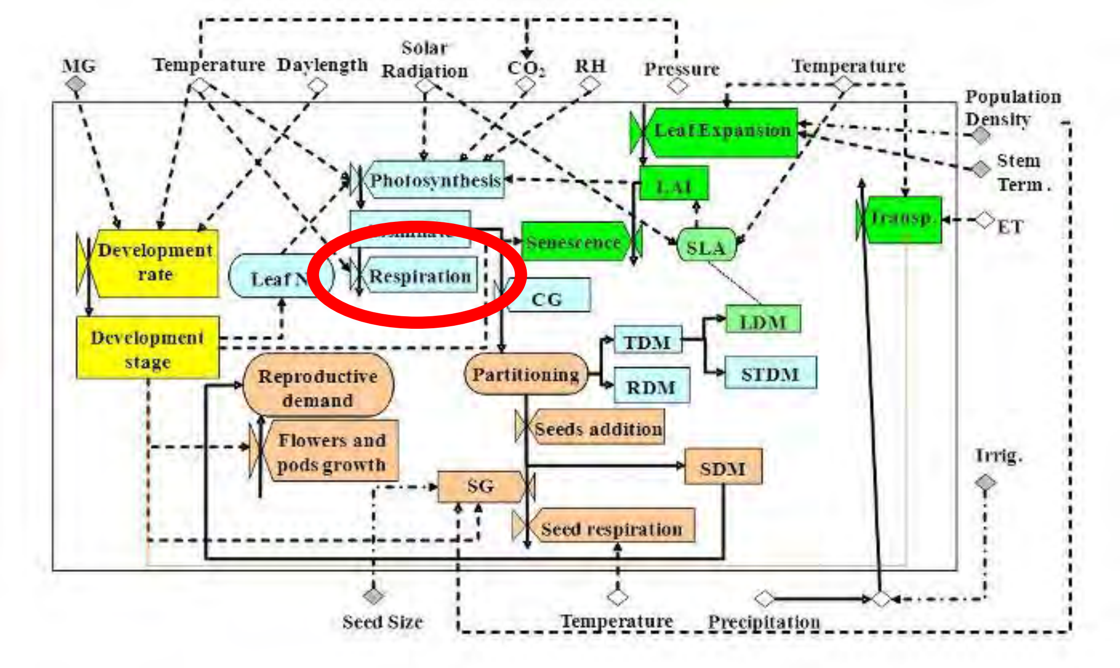
http://www.life.illinois.edu/plantbio/wimovac/

Templates, Anchors, and Matryoshkas in Crop Science

UNIVERSITY OF NEBRASKA-LINCOLN

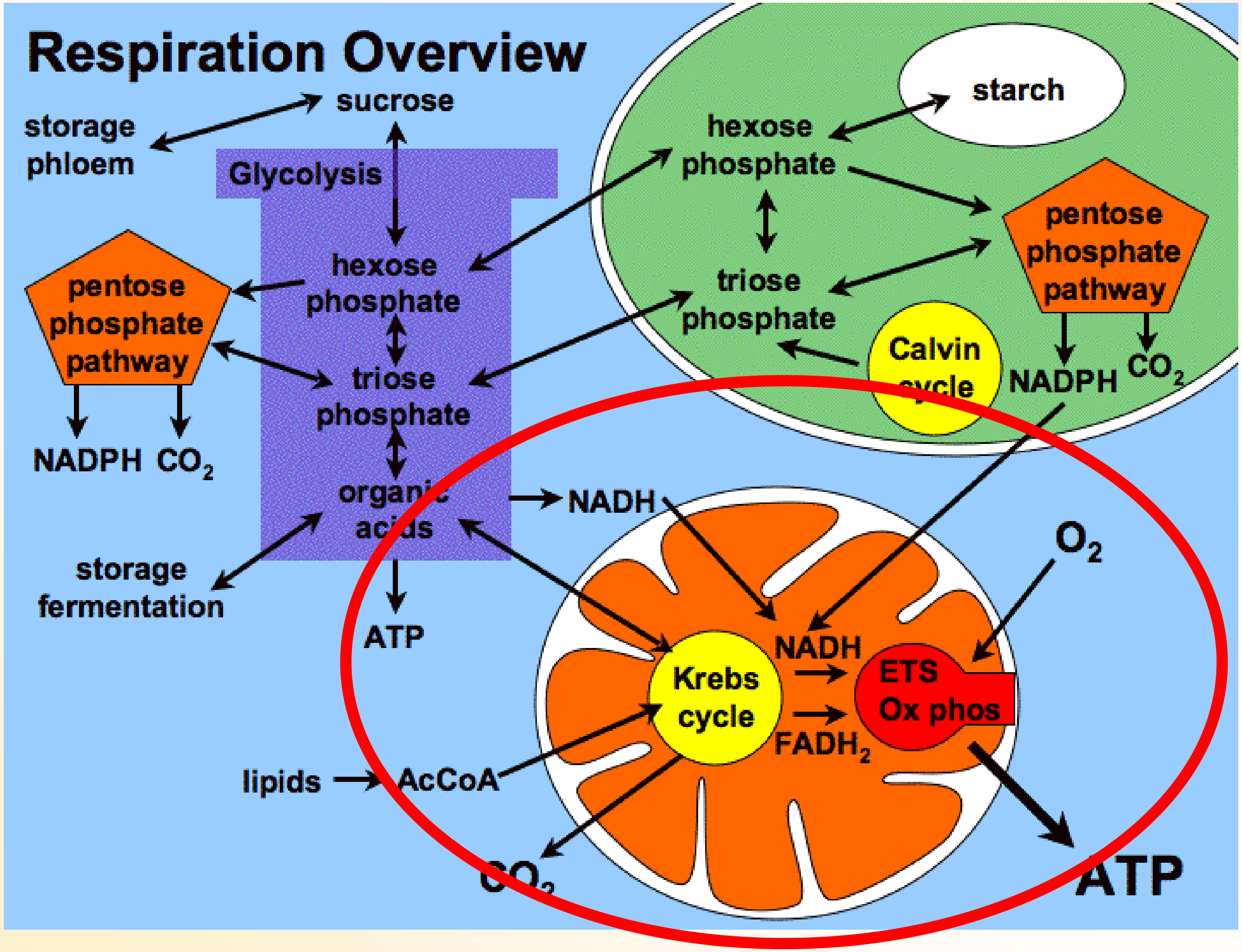
Institute of Agriculture and Natural Resources

SOYSIM - SOYBEAN GROWTH SIMULATION MODEL

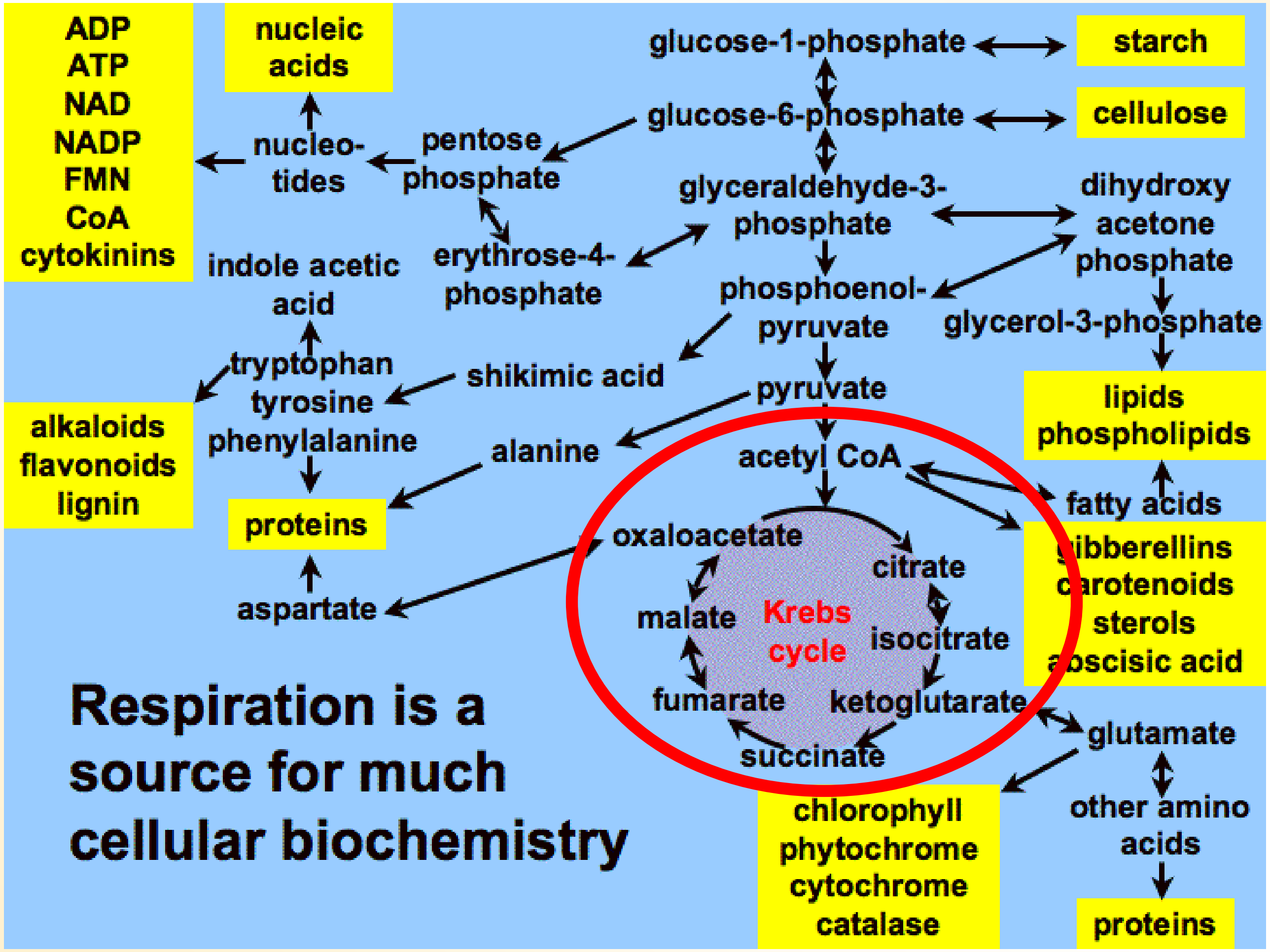


Weng *et al.*, *J. Geophys. Res. – Biogeosciences* 2008

Templates, Anchors, and Matryoshkas in Crop Science



Templates, Anchors, and Matryoshkas in Crop Science



ADP
ATP
NAD
NADP
FMN
CoA
cytokinins

nucleic acids

nucleotides

pentose phosphate

erythrose-4-phosphate

indole acetic acid

tryptophan
tyrosine
phenylalanine

alkaloids
flavonoids
lignin

shikimic acid

alanine

proteins

aspartate

glucose-1-phosphate

starch

glucose-6-phosphate

cellulose

glyceraldehyde-3-phosphate

dihydroxy acetone phosphate

glycerol-3-phosphate

phosphoenol-pyruvate

pyruvate

acetyl CoA

lipids
phospholipids

fatty acids

gibberellins
carotenoids
sterols
abscisic acid

oxaloacetate

citrate

isocitrate

ketoglutarate

succinate

fumarate

malate

Krebs cycle

glutamate

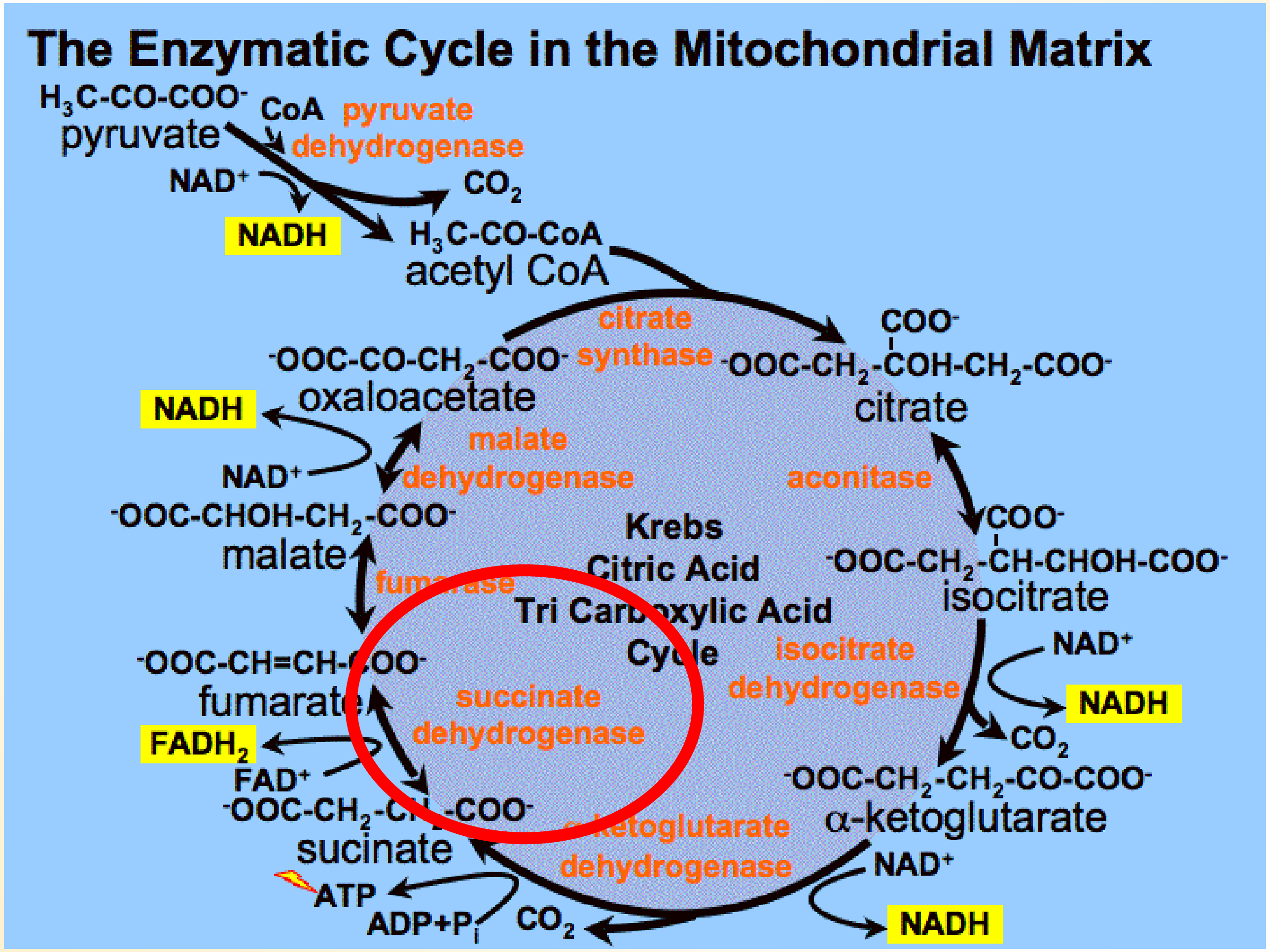
other amino acids

proteins

chlorophyll
phytochrome
cytochrome
catalase

Respiration is a source for much cellular biochemistry

Templates, Anchors, and Matryoshkas in Crop Science



The Enzymatic Cycle in the Mitochondrial Matrix

Templates, Anchors, and Matryoshkas in Crop Science

Search KEGG for succinate dehydrogenase

Database: KEGG - Search term: succinate dehydrogenase

KEGG BRITE

ko00001
KO; KEGG Orthology (KO)
ko00002
Module; KEGG modules
ko01000
Enzyme; Enzymes
ko03029
Mitochondrial biogenesis; Mitochondrial biogenesis

KEGG MODULE

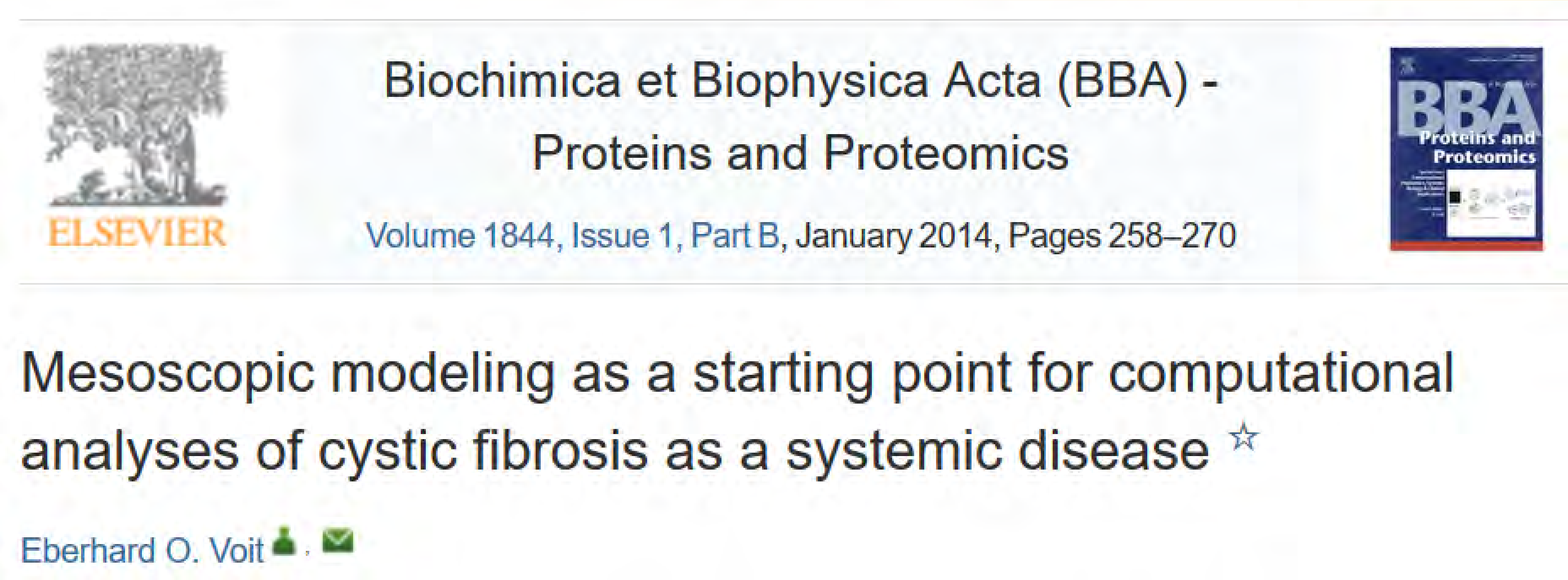
M00148
Succinate dehydrogenase (ubiquinone)
M00149
Succinate dehydrogenase, prokaryotes

KEGG ORTHOLOGY

K00135
gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde
1.2.1.79 1.2.1.20]
K00139
ALDH5A1; succinate-semialdehyde dehydrogenase [EC:1.2.1.24]

- Gene expression
- Genome modeling
- Epigenetics
- Molecular modeling

Relative Importance of Modules



ELSEVIER

Biochimica et Biophysica Acta (BBA) -
Proteins and Proteomics

Volume 1844, Issue 1, Part B, January 2014, Pages 258–270

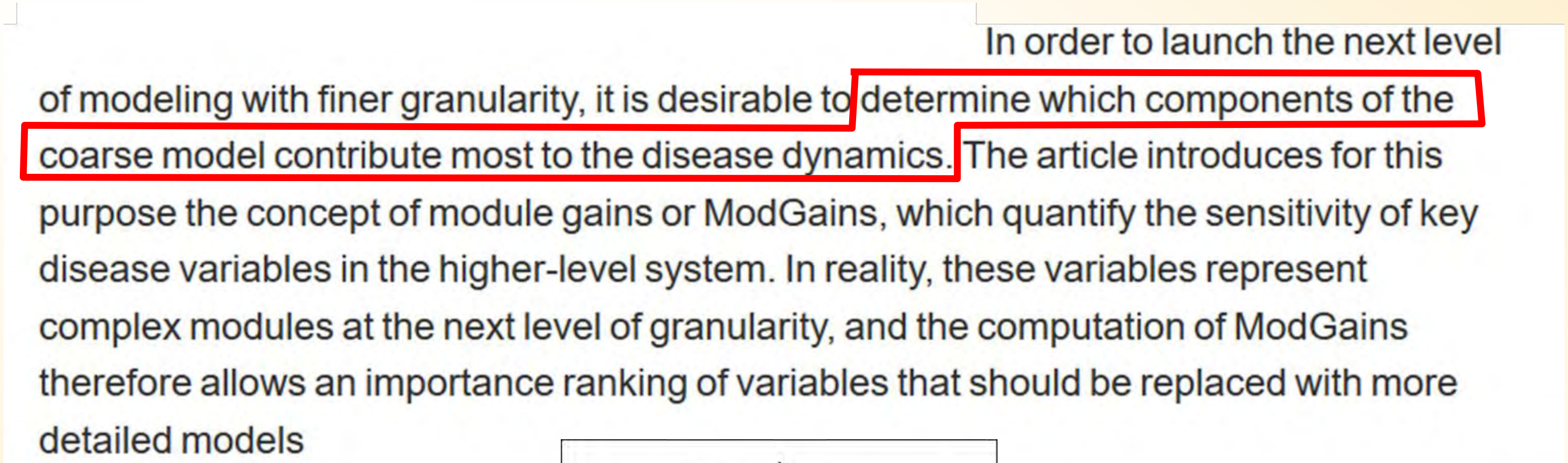
Mesoscopic modeling as a starting point for computational analyses of cystic fibrosis as a systemic disease ☆

Eberhard O. Voit

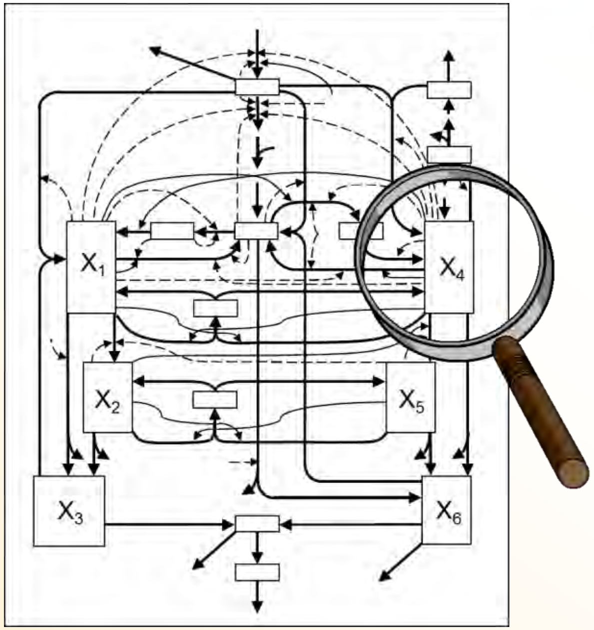
Article proposes a template and anchor model of
inflammation as basis for a disease simulator

Relative Importance of Modules

From Abstract:



In order to launch the next level of modeling with finer granularity, it is desirable to determine which components of the coarse model contribute most to the disease dynamics. The article introduces for this purpose the concept of module gains or ModGains, which quantify the sensitivity of key disease variables in the higher-level system. In reality, these variables represent complex modules at the next level of granularity, and the computation of ModGains therefore allows an importance ranking of variables that should be replaced with more detailed models

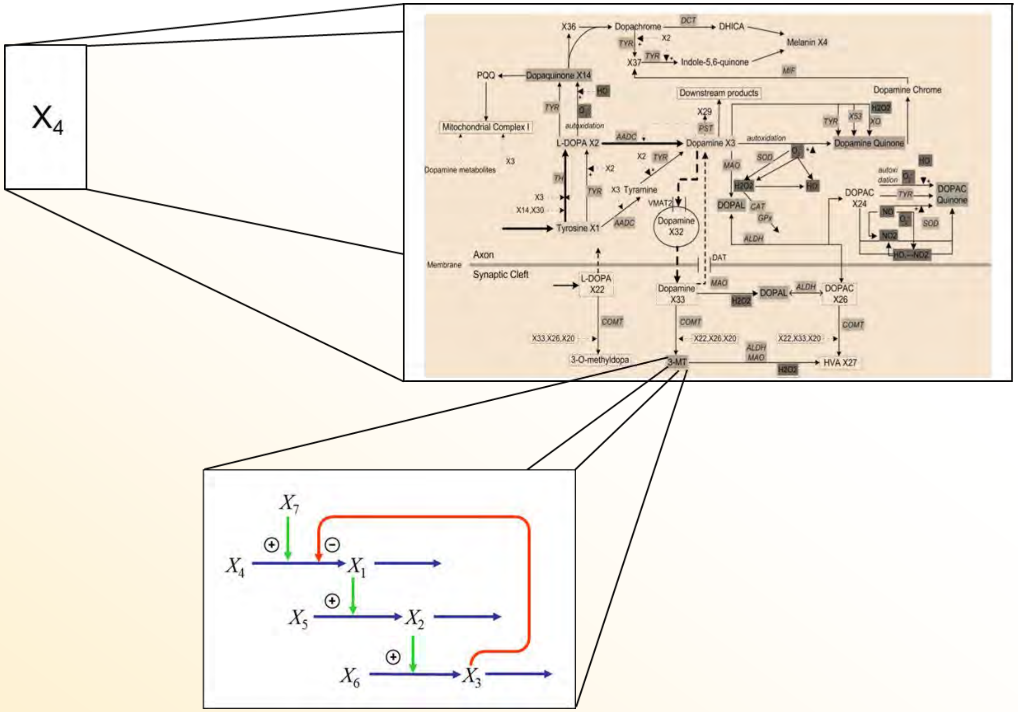


Relative Importance of Modules

From Abstract:

these variables represent complex modules at the next level of granularity, …

… The "hot-swapping" of such detailed modules for former variables is greatly facilitated by the architecture and implementation of the overarching, coarse model structure, which is here formulated with methods of biochemical systems theory (BST).



Usefulness of Modular Approach

Study modules, assuming that all outside the model is normal
or we know how it is changed

Explore responses of modules to external perturbations

Possibly replace module with context-dependent transfer function

Possibly create a catalog look-up table of quantitative I-O relationships

If this is possible:

Use look-up tables (or transfer functions) in template model

Interoperability of anchors within templates facilitated

Interoperability within hierarchies facilitated

Interoperability of Modules

Modules should be:

- self-contained
- characterized through identified input and output processes
- easily connectable within a level
- connectable across hierarchies
- possibly using the same language,
  - but may be not, as long as they can communicate

Scale-independent modeling language helps

Scale-Independent Modeling

Example for ODEs: Biochemical Systems Theory

Power-law models

Rigorous approximations

Independent of scale, all processes are represented by products of power-law functions of all variables affecting the process.

Models at a higher scale are easily expanded within the same formalism

Models at a lower scale are easily incorporated in higher-scale models

Different-Scale Plant Models in BST

Smith (1974a,b, 1975) Plant ecosystem modeling

Voit (1988) Explanation of the 3/2 rule in forests

Torsella (1991) Estimation of growth parameters in forests

Torres (1996) Magnesium flow in a tropical forest

Sands (1996) Parameter estimation in plants

Voit (1996a, b) Biomass partitioning in *Pinus taeda*

Martin (1997) Model condensation for forestry models

Kaitaniemi (2000, 2005) Resource allocation in trees

Renton (2005a, b) Functional–structural plant models

Lee (2012, 2013) Lignin biosynthesis in *Populus* and *Medicago*

Faraji (2015) Lignin biosynthesis in *Panicum*

Faraji (2017) Lignin biosynthesis in *Brachypodium*

Summary

- Biology is changing very rapidly
- Observation and description are becoming very quantitative
- Make use of this trend in modern crop science
- Challenge: complexity; to be addressed with systems biology
- Computational Systems Biology (CSB) is in its infancy
- CSB holds great promise for:
  - system simulation
  - discovery of design and operating principles
  - extrapolation and optimization
- Substantial challenges are to be overcome
- But: The future is here. Let’s solve some truly grand challenge problems!

Web-Images

http://luminexusa.org/wp-content/uploads/bfi_thumb/onion-n2fhsqcdk8a1irebz8ua3d5ne782hyz8xa8ek3jph4.jpg

https://s-media-cache-ak0.pinimg.com/736x/bf/6c/11/bf6c118adbba77d445c58187dbb652dc.jpg

https://pixabay.com/p-576847/?no_redirect

http://faculty.kutztown.edu/friehauf/science_outreach/elodia_cells_salted_002.jpg

http://media.mnn.com/assets/images/2015/07/shutterstock_121285357.jpg.838x0_q80.jpg

http://www.forestrydistributing.com/images/image/insects/mountain-pine-beetle.jpg

http://www.emergencycareforyou.org/uploadedImages/danger-poison.jpg

http://www.backincontrol.com/wp-content/uploads/2016/02/firewx1.jpg

https://upload.wikimedia.org/wikipedia/commons/thumb/e/e1/Drought.jpg/1200px-Drought.jpg

http://www.rcn-forecast.ou.edu/download/teco.php

http://soysim.unl.edu/graphics/soysim_processes_diagram.jpg

http://plantphys.info/plant_physiology/images/psnresp.gif

https://dyscover24x7.files.wordpress.com/2017/03/sergiev-posad-tale1.jpg

https://s-media-cache-ak0.pinimg.com/736x/a9/76/09/a97609066483aa641be16487569e3dc7.jpg

http://www.timvandevall.com/wp-content/uploads/2013/06/Printable-Graphic-Organizers-04.jpg

https://s-media-cache-ak0.pinimg.com/236x/b3/02/29/b3022927138bfd4aabc87dc31826f3bd.jpg

https://s-media-cache-ak0.pinimg.com/736x/21/37/8d/21378d0bedd9ca5a583c07d2077e7865.jpg

Acknowledgments

Lab Members:
Dr. An Dam
Mojdeh Faraji
Dr. Luis Fonseca
Felipe Dalvi-Garcia
Anuj Gupta
James Davis
Yun Lee
Daniel Olivenca
James Wade

Rick Dixon, UNT

Funding:
DARPA;
NSF (MCB, OCE, DEB);
NIH (GM, NIEHS, NIAID);
DoE (BESC);
Georgia Research Alliance

Information:
www.bst.bme.gatech.edu
Eberhard.voit@bme.gatech.edu