Circumventing the Parameter Values Bottleneck: Addressing the Challenge by Development of Phenotype-Centric Modeling Strategies

Michael A. Savageau
Microbiology & Molecular Genetics
University of California, Davis

masavageau@ucdavis.edu

BIRS2021 October 1, 2021

The Take-Home Message Regarding Phenotype-Centric Modeling

It is based on linear algebra in a log space and avoids dense sampling and numerical simulation

This strategy is especially useful at the early stage of investigations when little is known

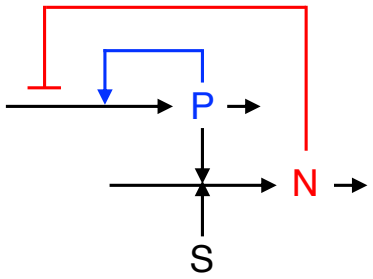
It provides an efficient “Fail-Early” method of hypothesis testing

For a given model architecture, the strategy starts without kinetic parameter values and ends with predicted values for the realization of specific phenotypes

Properties of specific phenotypes as well as relationships among phenotypes are related mechanistically to genotype and environment

What Is Meant By Architecture And What Can It Tell Us About Phenotypes?

Mechanism Architecture



- Who are the players?
- Who is talking to whom?
- How are they doing it?
- What are they saying?

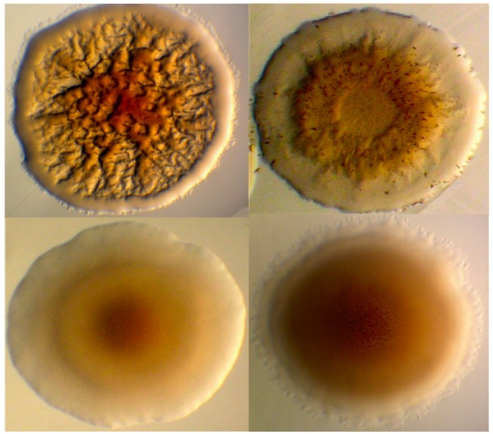
Speak Up! or Shut Up!

?

Phenotypes

What Can Phenotypes In Turn Tell Us About Their Population Dynamics And Evolution?

Phenotypes



?

Population Dynamics

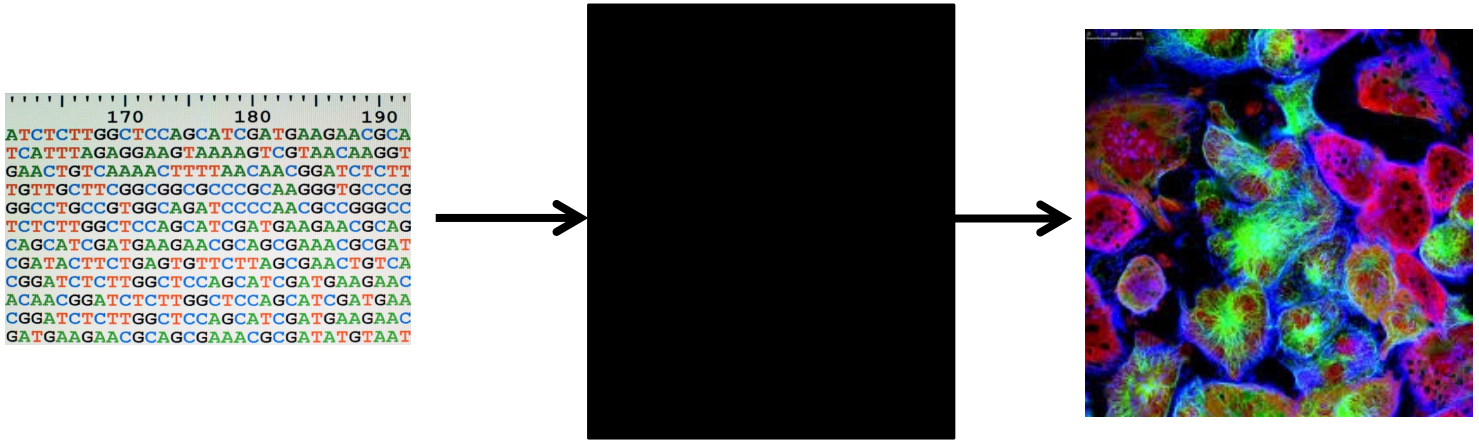
$$\frac{dN_i}{dt} = \sum_{j \neq i}^{n} m_{ji}\mu_j N_j - \sum_{j \neq i}^{n} m_{ij}\mu_i N_i + \mu_i N_i$$

- Growth rate constants &
- Mutation rate constants

I Will Address These Two Questions In A Three-Part Presentation

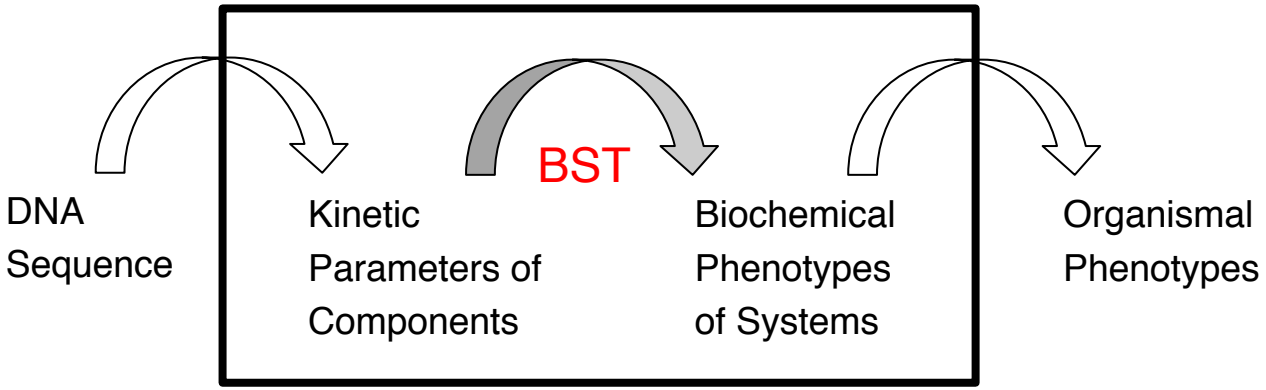
- Overview of Biochemical Systems Theory, Phenotypes, Design Principles and Modeling Strategy
- Derivation of Phenotype-Specific Mutation Rates
- Simple Example of Population Dynamics and Evolution

Underlying Mechanisms Are Key To The Problem



The inability to relate genotype and environment to phenotypes exhibited by biological systems is one of the ‘Grand Challenges’ in biology (Brenner, 2000)

Mechanisms Provide One Of Three Critical Mappings Between Genotype And Phenotype



Biochemical Systems Theory

- Scope includes mechanistic models governed by rate laws
- Rate laws are the power functions of chemical kinetics and the rational functions of biochemical kinetics
- These functions and conserved quantities are integrated into a network by means of Kirchhoff's Node Law
- The result is a system of Differential-Algebraic equations

Generalized Mass Action Equations

Without loss of generality, the Differential-Algebraic equations consisting of power-law and rational functions can be recast into Generalized Mass Action equations consisting only of sums and products of power-law functions.

$$\frac{dX_i}{dt} = \sum_{k=1}^{Pi} \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} - \sum_{k=1}^{Qi} \beta_{ik} \prod_{j=1}^{n+m} X_j^{h_{ijk}}, \qquad i = 1,\ldots,n_c$$

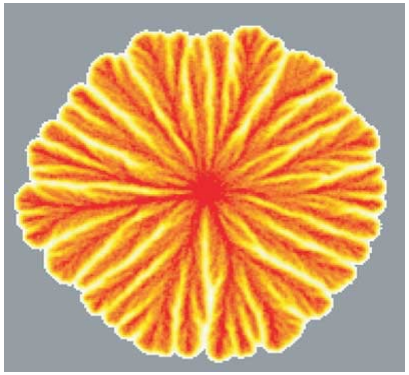
$$0 = \sum_{k=1}^{Pi} \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} - \sum_{k=1}^{Qi} \beta_{ik} \prod_{j=1}^{n+m} X_j^{h_{ijk}}, \qquad i = n_c+1,\ldots,n$$

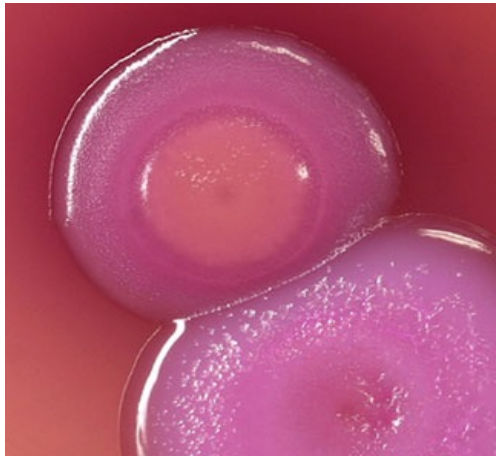
Savageau, M.A. and Voit, E.O. (1987) Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. Math. Biosci. **87**, 83-115.

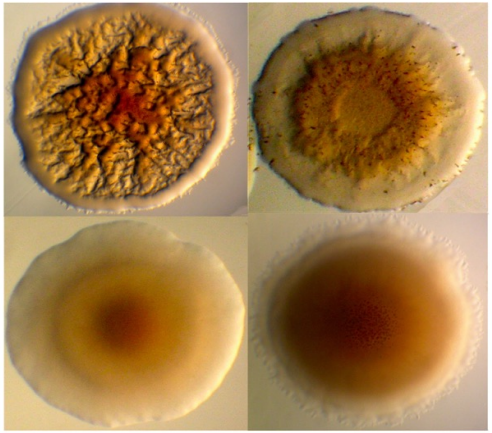
Comparable Quantitative Concepts Lacking For A Deep Understanding Of Genotype To Phenotype

- Genotype has a well-defined generic definition: genome sequence
- Phenotype has no comparable generic definition: ad hoc and descriptive

Microbial Phenotypes

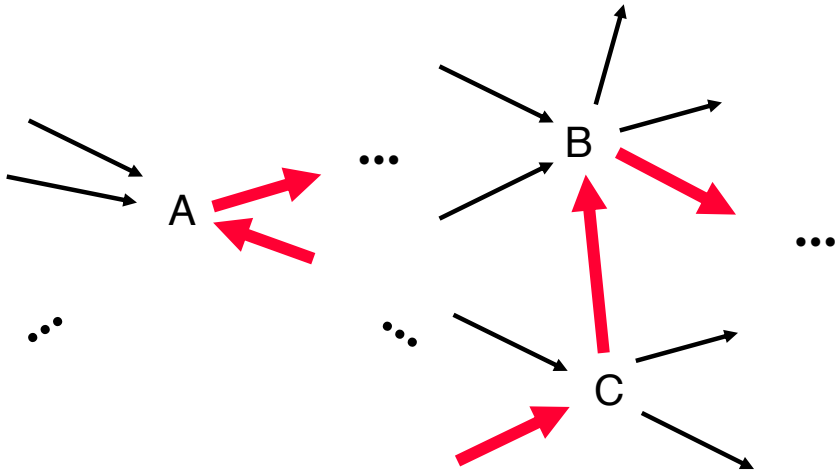






Without A Generic Definition Of Phenotype,
You Cannot “Predict” Phenotypes That
You Have Not Already Seen!

Definition Of Phenotypes:
Valid Combination Of Dominant Processes
Involving All Concentrations and Fluxes



(First conceptually; later more mathematical detail in the concrete example)

Test For Validity Involves S-System Equations And Inequalities

$$\frac{dX_k}{dt} = \sum_{p=1}^{c} \alpha_{kp} \prod_{j=1}^{n+m} X_j^{g_{kjp}} - \sum_{p=1}^{r} \beta_{kp} \prod_{j=1}^{n+m} X_j^{h_{kjp}}$$

$$\frac{dX_k}{dt} = \alpha_{kq} \prod_{j=1}^{n+m} X_j^{g_{kjp}} - \beta_{kq} \prod_{j=1}^{n+m} X_j^{h_{kjp}}$$

- Solution of linear equations in logarithmic coordinates
- Satisfying linear inequalities in logarithmic coordinates
- Rigorously define linear hyper-planes for phenotype boundaries (polytopes)

Definitions Of Phenotype Based On Biochemical Systems Theory

- A *phenotype* is the set, or sets, of concentrations and fluxes corresponding to a valid combination of dominant processes functioning within an intact system
- A *qualitatively distinct phenotype* is the characteristic phenotype that exists throughout a region of validity (polytope) in parameter space
- A *phenotypic repertoire* is the collection of qualitatively distinct phenotypes integrated into a space-filling structure -- the System Design Space of parameter values

Savageau, et al. (2009) Phenotypes and Tolerances in the Design Space of Biochemical Systems. *Proc Natl Acad Sci U S A*. **106**(16), 6435-6440.

Biochemical Phenotypes Characterized By

- Mechanisms: Portions of the system's mechanisms being exercised
- Equations: S-system equations
- Geometry: Boundaries, volumes & robustness in design space
- Design: System design principles
- Behavior: Qualitative and quantitative

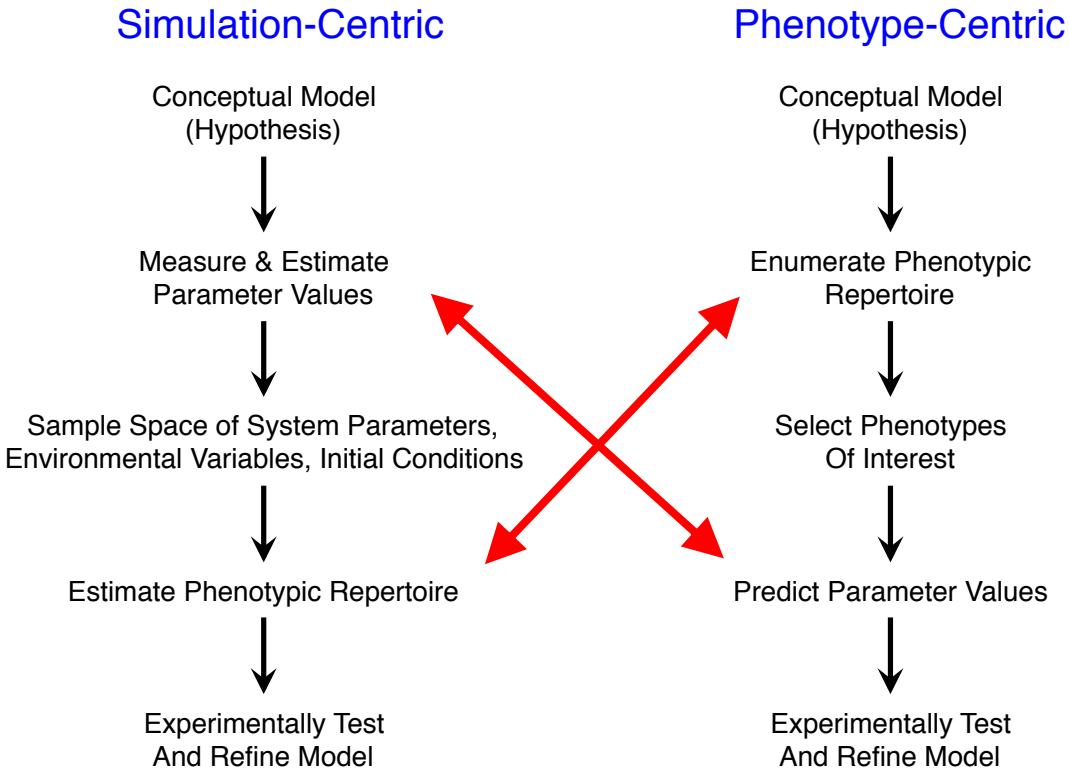
Design Space Toolbox (DST) For Automated Prediction Of Phenotype Characteristics

The Docker images used by DST3 are freely available at https://hub.docker.com/r/savageau/dst3

- Repertoire dependent only on Architecture
- Parameter values for the realization of each phenotype
- Concentrations & Fluxes in steady state
- Global tolerances for each parameter
- Signal amplification measured by input-output gain factors
- Eigenvalues for local dynamics
- Mutation rates based on polytope volumes and centroids (as we shall see)

Valderrama Gómez, M.A, et al. (2020) Mechanistic Modeling of Biochemical Systems without A Priori Parameter Values Using the Design Space Toolbox v.3.0, *iScience* **23**, 1-19.

Toolbox Enables A Very Different Modeling Strategy

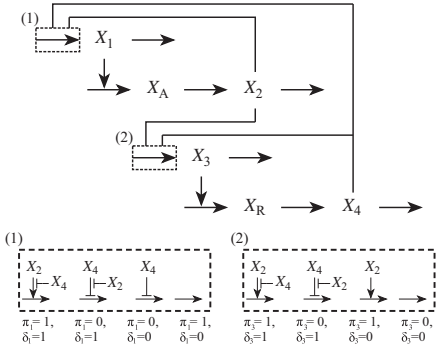


Valderrama Gómez, et al. (2018) Phenotype-centric modeling for elucidation of biological design principles. *J. Theoret. Biol.* **455**, 281-292.

First: The Phenotypic Repertoire Identifies
(Without Kinetic Parameter Values)

- "Physiological" phenotypes
  - Normal
  - Cyclic (rapid equilibrium)
  - Co-dominant
- Pathological phenotypes
  - Exploding
  - Imploding
- Phenotypes of interest by filtering the repertoire
  - Multi-modality
  - Oscillations
  - Complex logic functions
  - "Fail-Early" hypothesis testing

We examined 16 alternative logic functions without specifying kinetic parameter values and identified the number of phenotypes capable of generating oscillations in each case



$$\frac{dX_1}{dt} = \alpha_1 \left[\frac{\rho_1^{-\pi_1} + \delta_1 \left(\frac{X_2}{K_2}\right)^{g_{12}} + \rho_1^{-1} \left(\frac{X_4}{K_{4A}}\right)^{g_{14}}}{1 + \delta_1 \left(\frac{X_2}{K_2}\right)^{g_{12}} + \left(\frac{X_4}{K_{4A}}\right)^{g_{14}}} \right] - \beta_1 X_1 - \mu X_1$$

$$\frac{dX_A}{dt} = \alpha_A X_1 - \beta_A X_A - \mu X_A$$

$$\frac{dX_2}{dt} = \beta_A X_A - \beta_2 X_2 - \mu X_2$$

$$\frac{dX_3}{dt} = \alpha_3 \left[\frac{\rho_3^{-\pi_3} + \left(\frac{X_2}{K_{2R}}\right)^{g_{32}} + \delta_3 \rho_3^{-1} \left(\frac{X_4}{K_4}\right)^{g_{34}}}{1 + \left(\frac{X_2}{K_{2R}}\right)^{g_{32}} + \delta_3 \left(\frac{X_4}{K_4}\right)^{g_{34}}} \right] - \beta_3 X_3 - \mu X_3$$

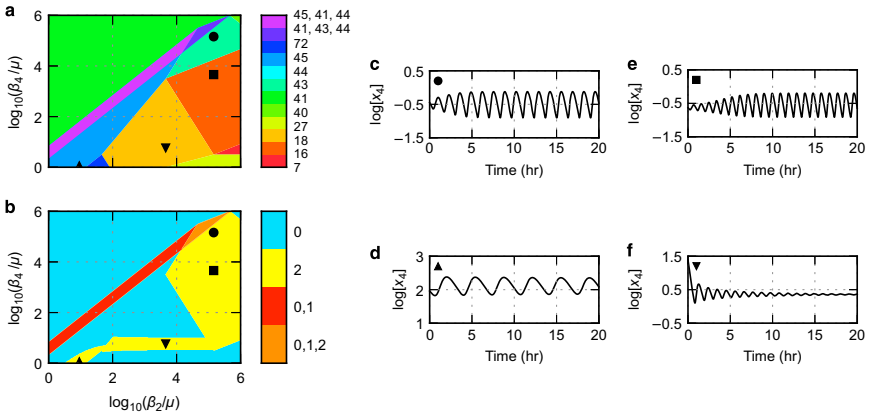
$$\frac{dX_R}{dt} = \alpha_R X_3 - \beta_R X_R - \mu X_R$$

$$\frac{dX_4}{dt} = \beta_R X_R - \beta_4 X_4 - \mu X_4$$

Table 2. Summary of global properties for the 16 designs in the general class of two-gene circuits

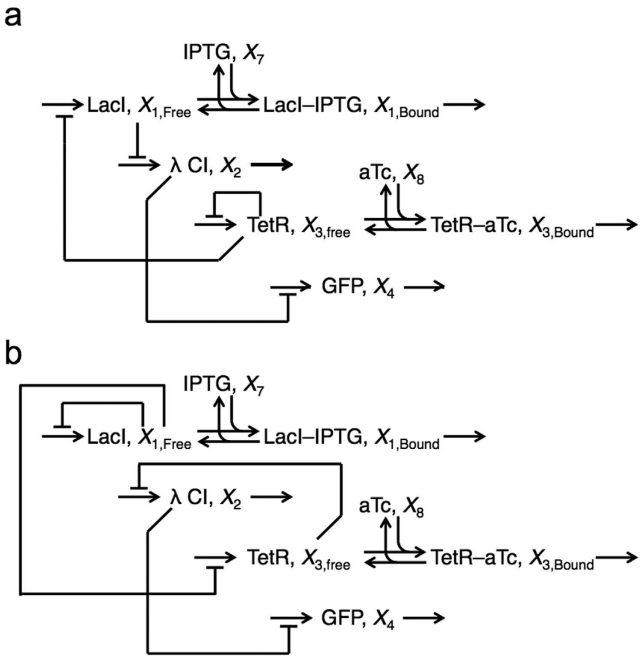
Design identifier	*Indices for the mode of control*[a]	*Phenotypic fraction*[b]	*No. of oscillatory phenotypes*
D.1	$\pi_1=0, \delta_1=0, \pi_3=0, \delta_3=0$	6/16	0
D.2	$\pi_1=0, \delta_1=1, \pi_3=0, \delta_3=0$	10/36	0
D.3	$\pi_1=0, \delta_1=0, \pi_3=0, \delta_3=1$	15/36	1
D.4	$\pi_1=0, \delta_1=1, \pi_3=0, \delta_3=1$	25/81	2
D.5	$\pi_1=1, \delta_1=0, \pi_3=0, \delta_3=0$	4/16	0
D.6	$\pi_1=1, \delta_1=1, \pi_3=0, \delta_3=0$	10/36	0
D.7	$\pi_1=1, \delta_1=0, \pi_3=0, \delta_3=1$	10/36	0
D.8	$\pi_1=1, \delta_1=1, \pi_3=0, \delta_3=1$	25/81	1
D.9	$\pi_1=0, \delta_1=0, \pi_3=1, \delta_3=0$	9/16	1
D.10	$\pi_1=0, \delta_1=1, \pi_3=1, \delta_3=0$	15/36	2
D.11	$\pi_1=0, \delta_1=0, \pi_3=1, \delta_3=1$	15/36	2
D.12	$\pi_1=0, \delta_1=1, \pi_3=1, \delta_3=1$	25/81	4
D.13	$\pi_1=1, \delta_1=0, \pi_3=1, \delta_3=0$	6/16	0
D.14	$\pi_1=1, \delta_1=1, \pi_3=1, \delta_3=0$	15/36	1
D.15	$\pi_1=1, \delta_1=0, \pi_3=1, \delta_3=1$	10/36	0
D.16	$\pi_1=1, \delta_1=1, \pi_3=1, \delta_3=1$	25/81	2

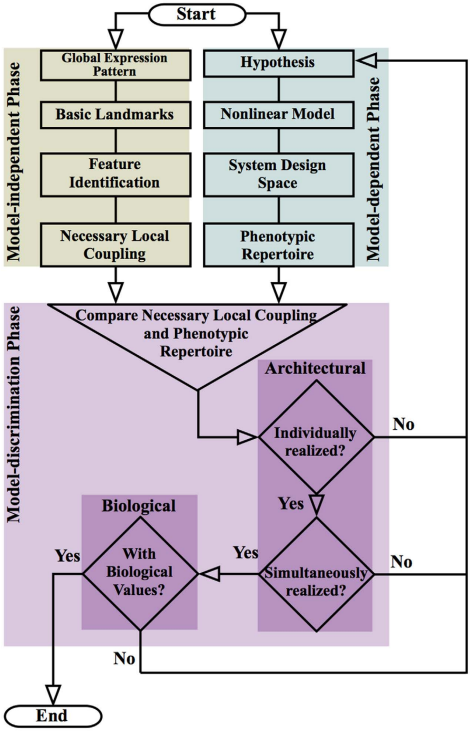
[a]The meaning of the π and δ symbols is described in the caption of Figure 1.
[b]The phenotypic fraction is shown as the number of valid phenotypes divided by the maximum number of potential phenotypes.



Lomnitz & Savageau (2015) Elucidating the genotype–phenotype map by automatic enumeration and analysis of the phenotypic repertoire. npj Systems Biology and Applications 1, 15003; doi:10.1038/npjsba.2015.3

We examine 40 alternative hypotheses and showed that 5 have the potential to reproduce the experimental data, and only one can do so with biologically relevant parameter values.



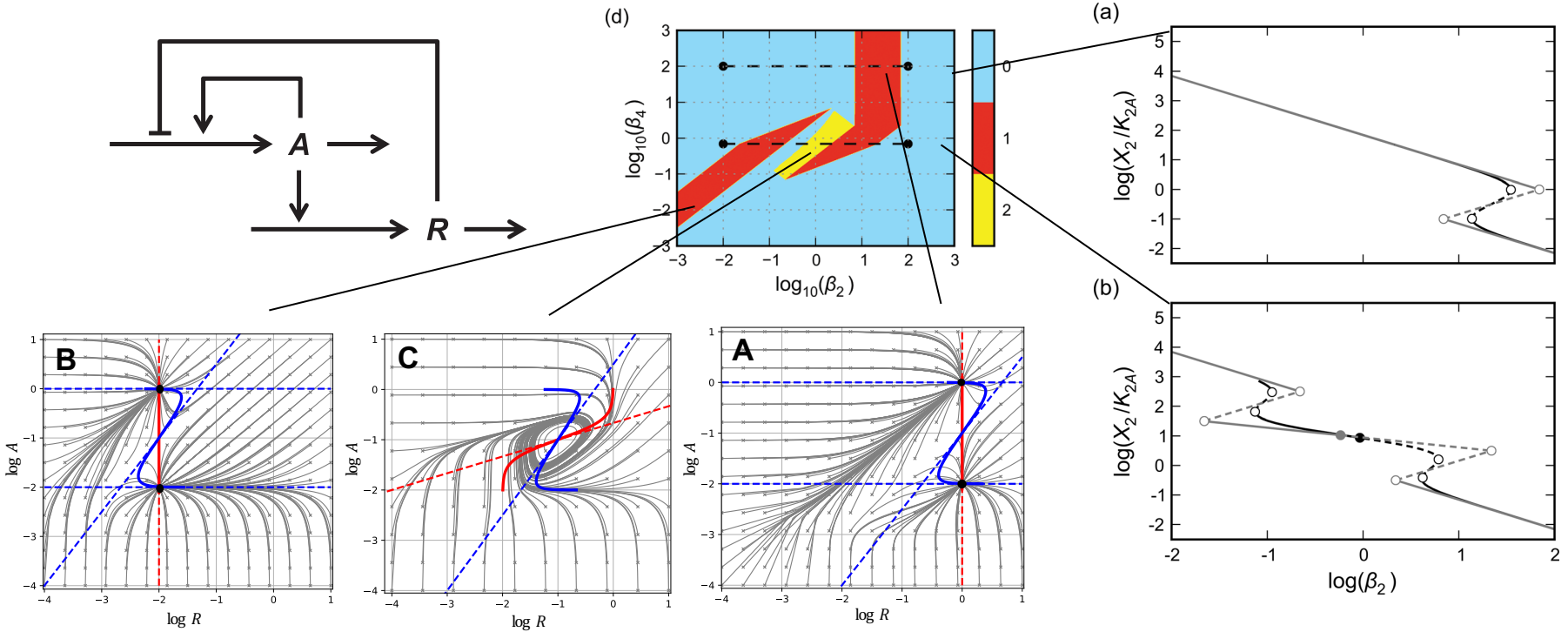


Lomnitz & Savageau (2015) Rapid Discrimination Among Putative Mechanistic Models of Biochemical Systems. Scientific Reports 6:32375 DOI: 10.1038/srep32375

Second: Properties Of Specific Phenotypes Are Predicted Without Sampling Parameter Values Or Numerical Simulation

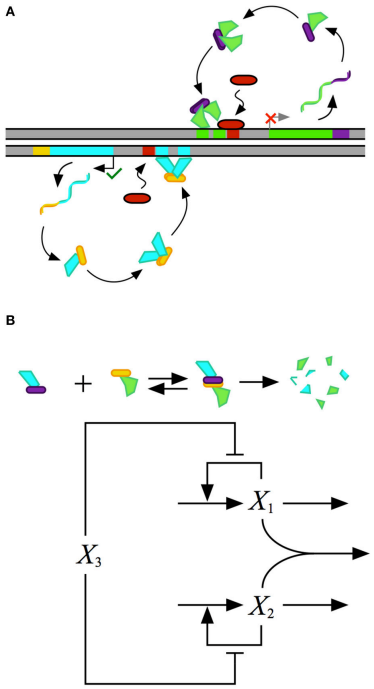
- Steady State Solution
- Input-output Amplification Factors
- Eigenvalues
- Kinetic Parameter Values
- Boundaries
  - Global Robustness
  - Bifurcations
  - Design Principles

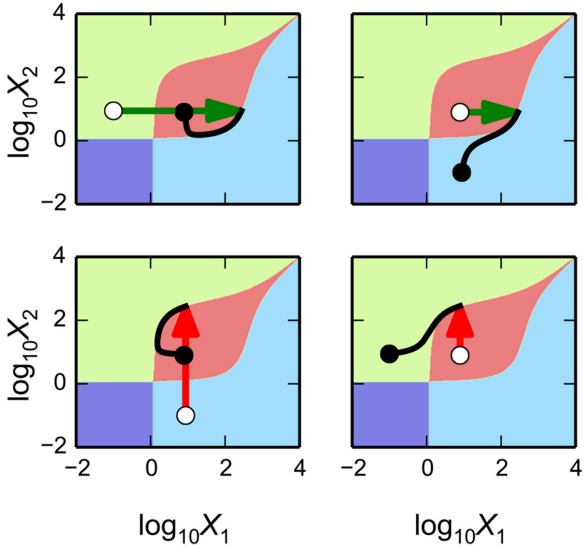
Phenotypes in Parameter Space and State Space

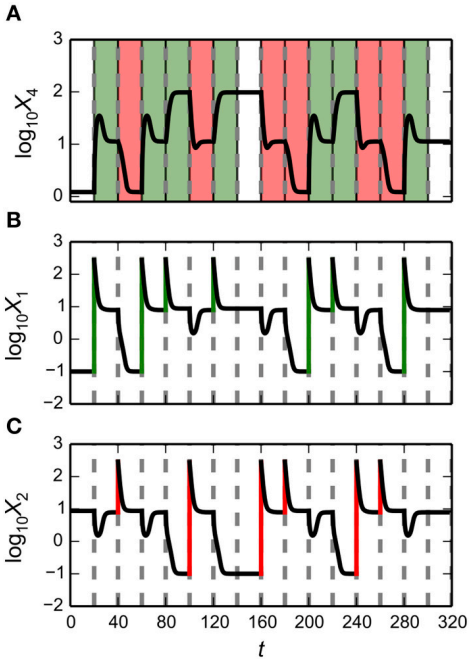


Valderrama-Gómez, et al. (2018) Phenotype-centric modeling for elucidation of biological design principles. Journal of Theoretical Biology 455, 281–292.

A Synthetic Three State Counter







Lomnitz & Savageau (2016) Design Space Toolbox V2: Automated Software Enabling a Novel Phenotype-Centric Modeling Strategy for Natural and Synthetic Biological Systems. Front. Genet. 7:118. doi: 10.3389/fgene.2016.00118

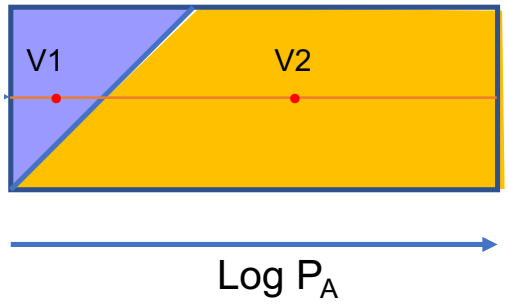
Three-Part Presentation

- Overview of Biochemical Systems Theory, Phenotypes, Design Principles and Modeling Strategy
- Derivation of Phenotype-Specific Mutation Rates
- Simple Example of Population Dynamics and Evolution

Phenotype-Specific Mutation Rates Based On Four Factors

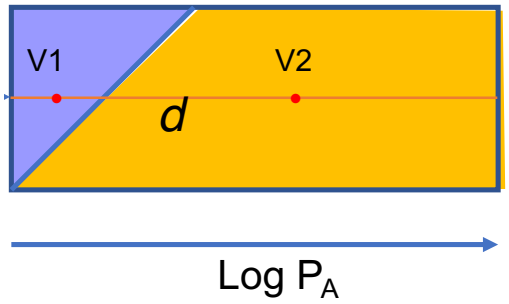
- Volume of the recipient phenotype in system design space
- Distance between centroids of the phenotype volumes
- Size scale favoring small distances between phenotypes
- Directional bias between donor and recipient phenotypes

Mutation Probability Is Proportional To Volume Of The Recipient Phenotype



- Volumes are rigorously determined by vertex enumeration methods and by maximal shared bounding box methods
- Phenotype volumes: #2 > #1
- Mutations from large to small volumes are rare
- Mutations from small to large volumes are frequent

Mutation Probability Varies With Size-Scale Factor λ Over The Distance Between Donor And Recipient Phenotypes



- Large scale mutations are rare; small scale mutations are frequent
- Average separation between phenotype #1 and #2 is the distance between their centroids, *d*
- Mutation probability: ~ exp[-*ld*/λ]
- Size-scale factor λ is estimated based on experimental data for the LAC repressor of *E. coli.*

Mutation Probability Varies With The Direction of Parameter Change Between Donor And Recipient Phenotypes

- The probability is larger when the parameter increase is in the direction of increased entropy, ~ exp*[-ldl/(λδ)]*
- The probability is smaller when the parameter increase is in the direction of decreased entropy, ~ exp*[-ldl*(δ/λ)]*
- Directional bias parameter δ is estimated based on experimental data for the LAC repressor of *E. coli.*

Phenotype-Specific Mutation Rate Is Determined In Four Steps

- The mechanistic parameter contribution K_{ij}
- Multiplied by the recipient volume contribution V_j
- Normalized to give the probability distribution
- Multiplied by the general mutation rate constant, *m*

$$K_{ij} \sim \exp(-\left|\log C_i - \log C_j\right| / \delta\lambda)$$

$$K_{ji} \sim \exp(-\left|\log C_j - \log C_i\right| \delta / \lambda)$$

$$K_{ij}V_j$$

$$k_{ij} = K_{ij}V_j \Big/ \left(\sum_{j=1}^{n_j} K_{ij}V_j \right) \qquad \sum_{j=1}^{n_j} k_{ij} = 1$$

$$m_{ij} = mk_{ij}$$

Population Dynamic Equations For Phenotypes In Steady-State Exponential Growth

$$\frac{dN_i}{dt} = \sum_{\substack{j=1 \\ j\neq i}}^{n} mk_{ji}\mu_j N_j - \sum_{\substack{j=1 \\ j\neq i}}^{n} mk_{ij}\mu_i N_i + \mu_i N_i \qquad i = 1,\cdots,n \qquad\qquad N_i \to \infty$$

$$\frac{dR_i}{dt} = \sum_{\substack{j=1 \\ j\neq i}}^{n} mk_{ji}\mu_j R_j - \sum_{\substack{j=1 \\ j\neq i}}^{n} mk_{ij}\mu_i R_i + \mu_i R_i - R_i\left(\sum_{j=1}^{n}\mu_j R_j\right) \qquad\qquad R_i \to R_i^{SS}$$

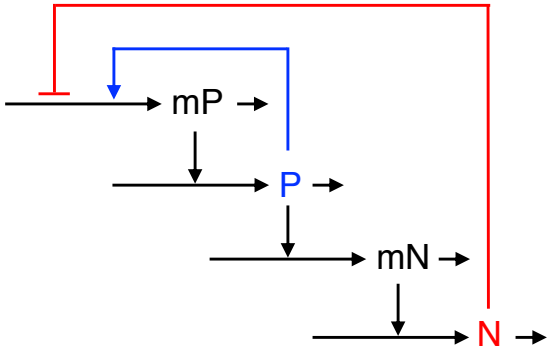
$$R_i = N_i \Big/ \sum_{j=1}^{n} N_j = \textit{Relative Frequency} \qquad\qquad \mu_i = \textit{Exponential Growth Rate}$$

Three-Part Presentation

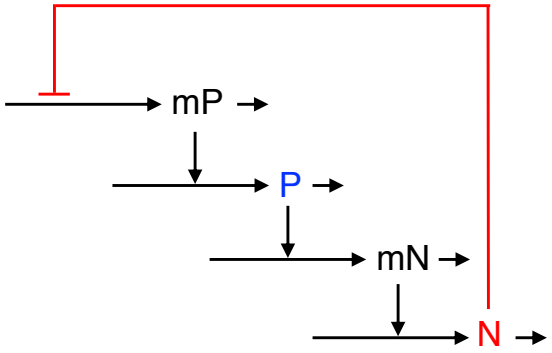
- Overview of Biochemical Systems Theory, Phenotypes, Design Principles and Modeling Strategy
- Derivation of Phenotype-Specific Mutation Rates
- Simple Example of Population Dynamics and Evolution

Simplified Example Illustrating The Strategy

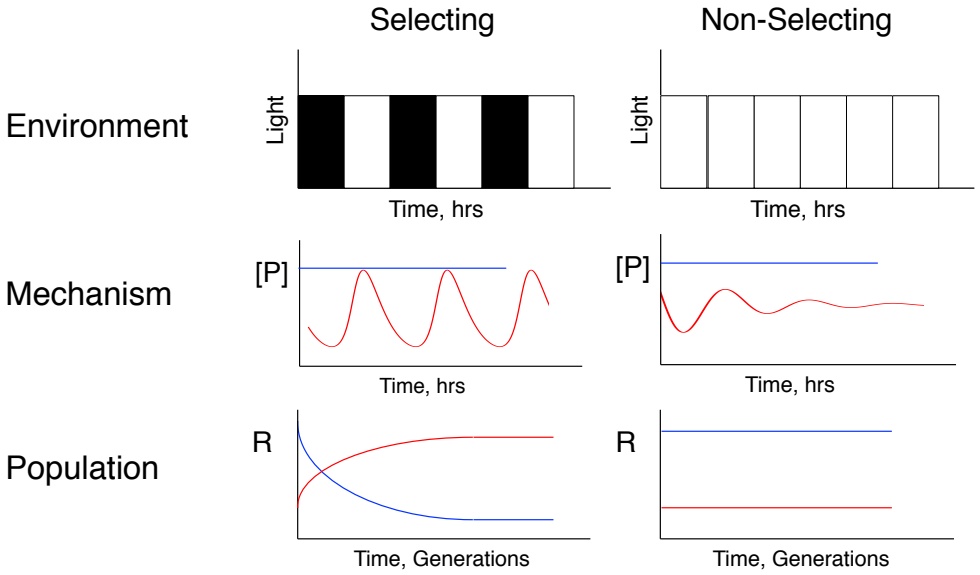
Circadian clock module found in nearly all modern organisms



Primordial precursor to a circadian clock



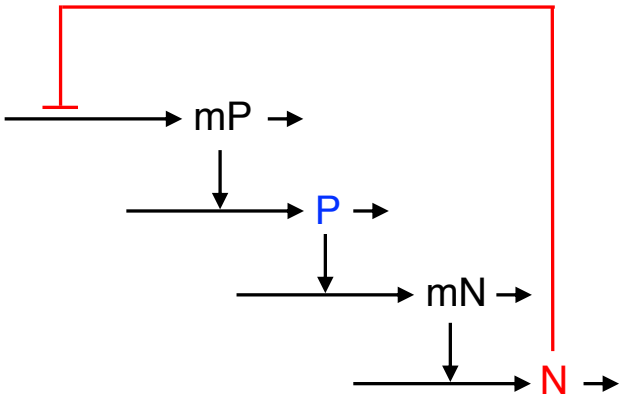
Oscillatory And Non-Oscillatory Clock Phenotypes Responding To Selecting And Non-Selecting Conditions



Woelfle MA, et al (2004) The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Curr Biol*. **14**:1481–1486.

Mechanism Predicting Phenotypes

Biochemical Kinetic Equations



$$\frac{dmP}{dt} = \frac{\alpha_{mP\max} + \alpha_{mP\min}\left(\frac{N}{K_N}\right)^n}{1+\left(\frac{N}{K_N}\right)^n} - \beta_{mP} mP$$

$$\frac{dP}{dt} = \alpha_P mP - \beta_P P$$

$$\frac{dmN}{dt} = \frac{\alpha_{mN\min} + \alpha_{mN\max}\left(\frac{P}{K_P}\right)^p}{1+\left(\frac{P}{K_P}\right)^p} - \beta_{mN} mN$$

$$\frac{dN}{dt} = \alpha_N mN - \beta_N N$$

4 variables and 12 kinetic parameters

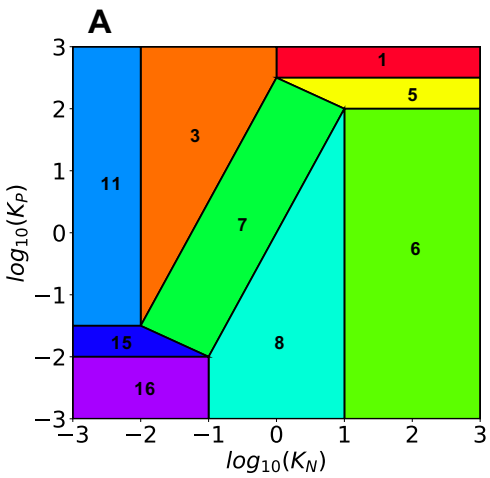
Phenotypic Repertoire Showing That Only Phenotype #7 Is Capable Of Oscillation

Phenotype Number	Phenotype Signature	Eigenvalues with Positive real part	Complex Conjugate Eigenvalues
1	11 11 11 11 11 11	0	-
3	11 11 11 11 21 11	0	-
5	11 11 21 11 11 11	0	-
6	11 11 21 11 11 21	0	-
7	11 11 21 11 21 11	0	+
8	11 11 21 11 21 21	0	-
11	21 11 11 11 21 11	0	-
15	21 11 21 11 21 11	0	-
16	21 11 21 11 21 21	0	-

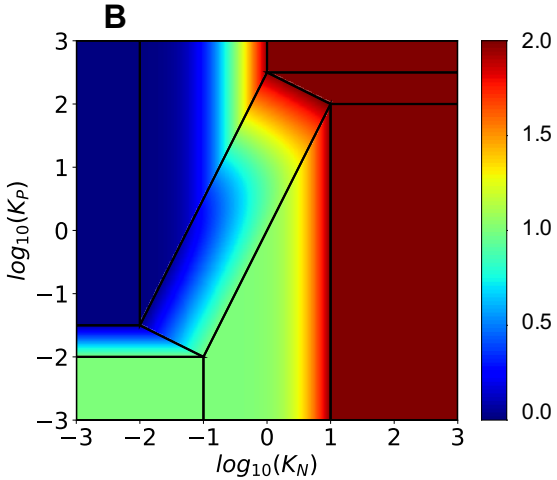
Predicted Parameter Values For The Realization Of Oscillatory Phenotype #7

Parameters	Value
KN	0.316
KP	1.78
amNmax	10.0
amNmin	1.00
amPmax	10000
amPmin	1.00
aN	1.00
aP	0.01
bmN	1.00
bmP	1.00
bN	1.00
bP	1.00

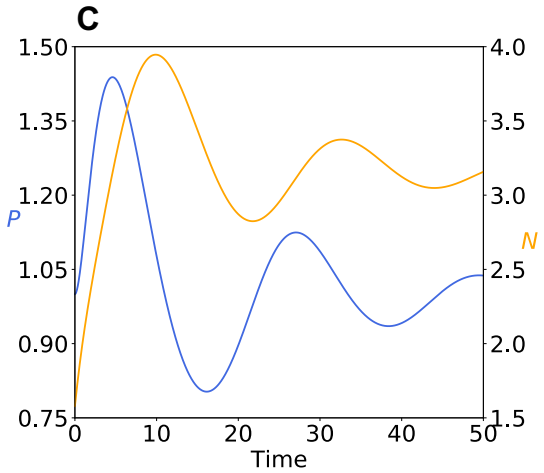
Automatically Predicted Characteristics Within Phenotypes In System Design Space



Phenotype Volumes

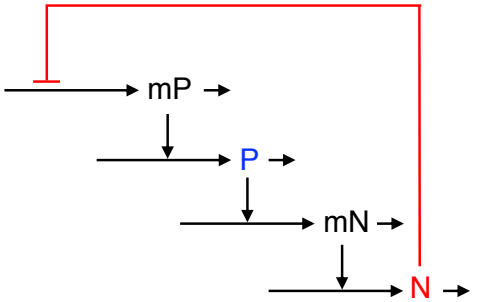


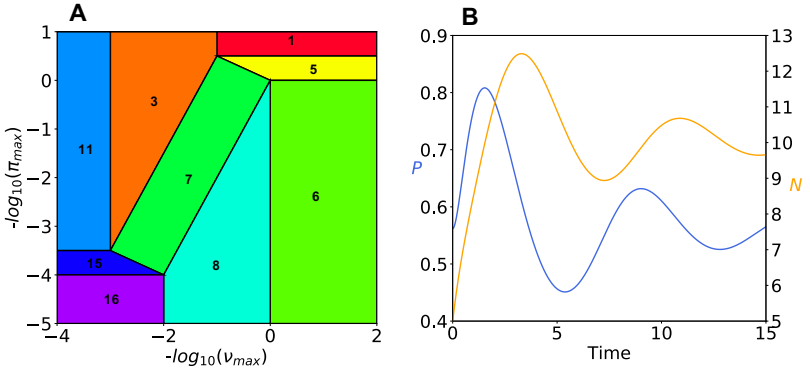
Protein Burden



Oscillatory Dynamics

Rescaled Molecular Model With 2 Fixed Parameters (ρ_N, ρ_P) Reveals A 10-Parameter Invariant System Design Space And The Full Design Principle For Oscillation





Dimensionless Parameter Groups

$$\pi_{\max} = \frac{1}{K_P}\left(\frac{\alpha_{mP\max}}{\beta_{mP}}\right)\left(\frac{\alpha_P}{\beta_P}\right) \qquad \nu_{\max} = \frac{1}{K_N}\left(\frac{\alpha_{mN\max}}{\beta_{mN}}\right)\left(\frac{\alpha_N}{\beta_N}\right)$$

System Design Principle

$$1 < [\pi_{\max}]^{-4/5}[\nu_{\max}]^{-2/5} < \rho_P$$

$$1 < [\pi_{\max}]^{+2/5}[\nu_{\max}]^{-4/5} < \rho_N$$

Phenotypes Predicting Equilibrium Distributions And Population Dynamics

Estimation Of Size Scale λ And Directional Bias δ Parameters Base On Experimental Data For The LAC Repressor Of *E. coli*

Proteins	Substitution Effect Map			Evolution Alignment Map	
	Wild type	Constitutive	Supper-Repressed	Non-Conserved	Conserved
LAC Repressor	67%	31%	2%	--	--
LAC Family	--	--	--	61%	39%
Clock N (λ=0.6, δ=1.85)	67%	31%	2%	--	--

Markiewicz, P, et al. 1994. Genetic Studies of the *lac* repressor XIV: Analysis of 4000 altered *Escherichia coli lac* repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. J. Mol. Biol. 240: 421-433.

Alternative Hypotheses Concerning Neutral and Protein Burden Fitness Effects In The Non-Selecting Condition

Phenotype, n	Growth Rate Differences Relative to Phenotype #7 in Non-Selecting Conditions, %		
	Non-Selecting		Selecting
	No Fitness Effects	Fitness Effects	
1	0	-2.43E-04	-2.43E-04
3	0	3.22E-05	3.22E-05
5	0	-2.47E-04	-2.47E-04
6	0	-2.64E-04	-2.64E-04
7	0	0	4.00E-04
8	0	-3.07E-05	-3.07E-05
11	0	4.12E-05	4.12E-05
15	0	1.15E-05	1.15E-05
16	0	-2.40E-05	-2.40E-05

Oscillation →

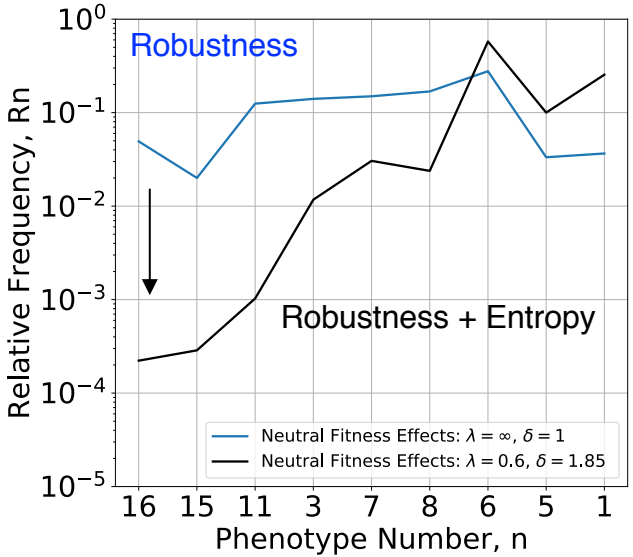
Neutral

Protein Burden

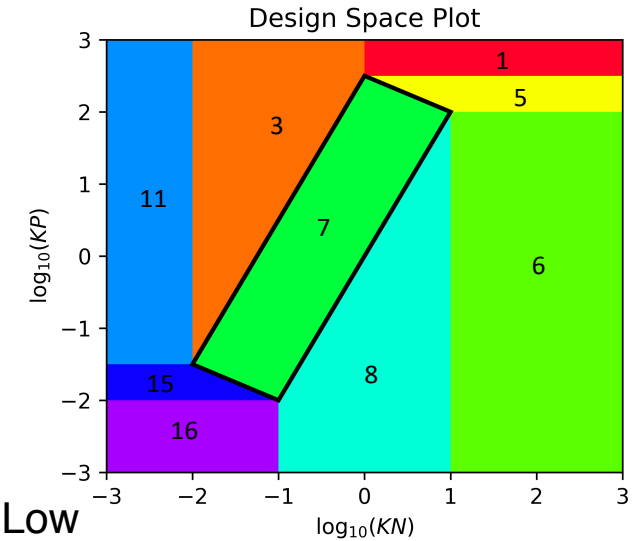
Selecting

Two Contributions to the Equilibrium Distribution Of Phenotype Diversity

Neutral Fitness Effects



High

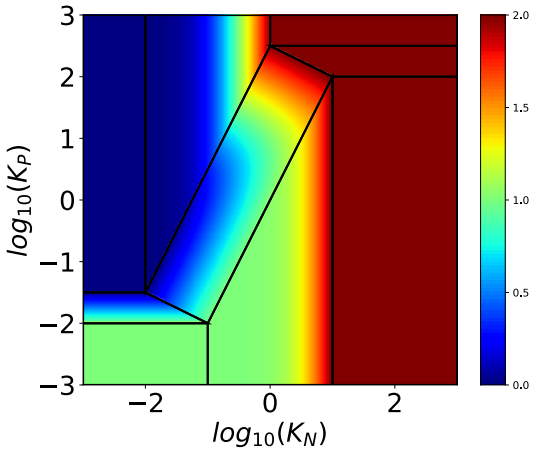


Low

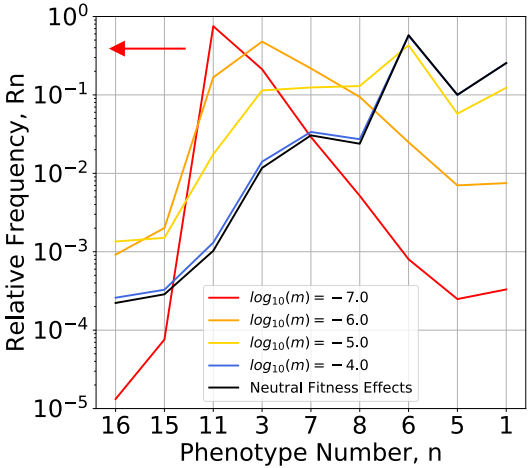
The distribution due to robustness (volume) alone is shifted downward to create a gradient from low to high entropy independent of mutation rate

Various Contributions to the Equilibrium Distribution Of Phenotype Diversity

Protein Burden Fitness Effects

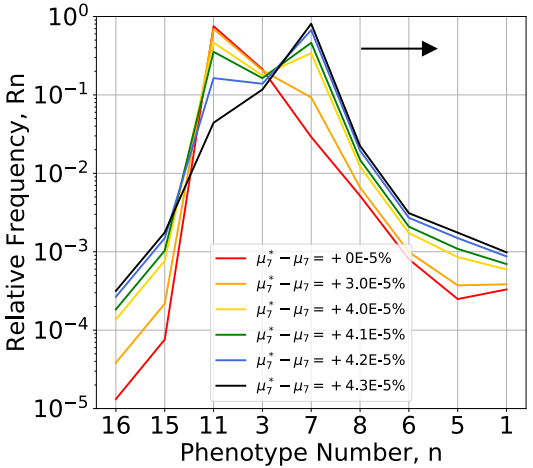


Robustness, Entropy & Mutation Rate



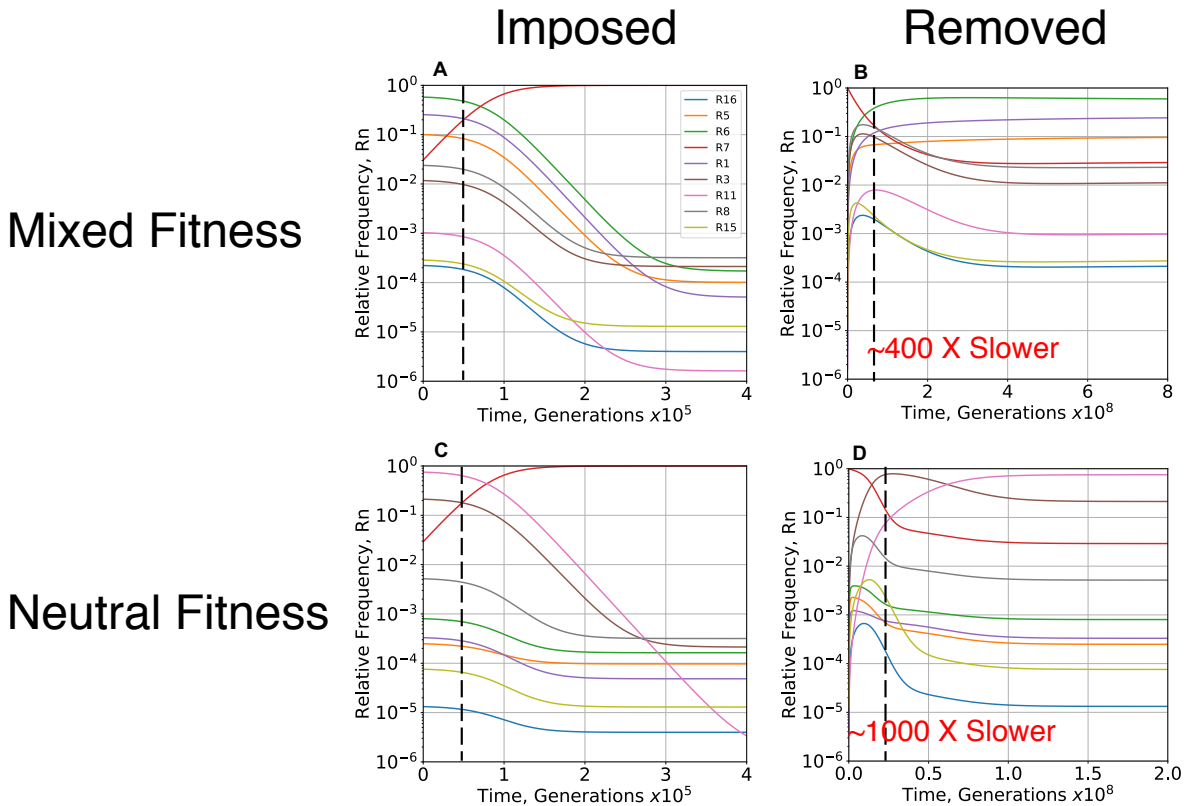
Under non-selecting conditions, distributions are shifted to lower entropy phenotypes depending on mutation rate

Robustness, Entropy & Selection

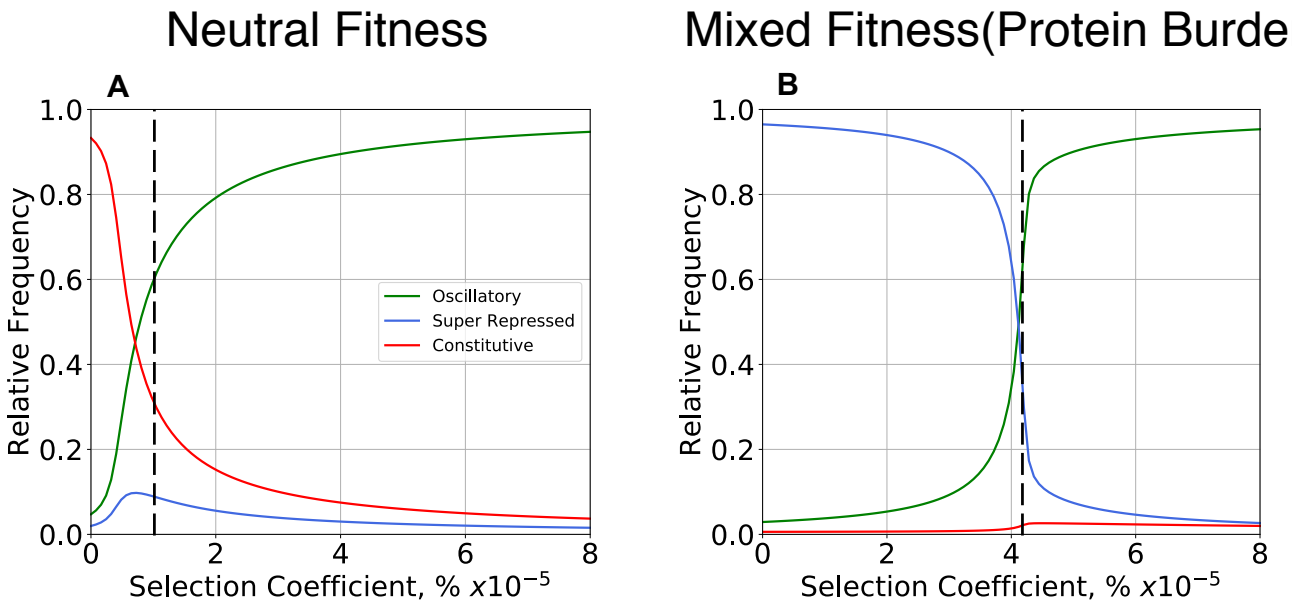


Under selecting conditions, distributions are shifted to the selected phenotype at the expense of lower entropy phenotypes

Changes In Phenotype Hierarchy And Time Scale Following Imposition And Removal Of Selection For Oscillatory Phenotype #7



Back Calculating Selection Coefficients From The Predicted Distribution Of The Qualitatively-Distinct Phenotypes Based On Constructed Mutants



Only the neutral hypothesis is consistent with the LAC experimental data

Key Attributes Of The Phenotype-Centric Modeling Strategy And The Design Space Toolbox For Its Automated Application

It is based on linear algebra in a log space and avoids dense sampling and numerical simulation

It is especially useful at the early stage of investigations when little is known

It provides an efficient “Fail-Early” method of hypothesis testing

It starts without requiring parameter values and ends with predicted values for the realization of specific phenotypes

It predicts systemic properties of specific phenotypes, as well as relationships among phenotypes, that are related mechanistically to genotype and environment

Acknowledgements

Biochemical and S-system Theory

Eberhard Voit
Masahiro Okamoto
Albert Sorribas
Fumihide Shiraishi
William Hlavacek
Rui Alves

Definitions of Phenotype and Global Tolerance

Armindo Salvador
Pedro Coelho
Dean Tolla

System Design Space and Toolbox

Rick Fasani
Jason Lomnitz
Miguel Valderrama-Gómez

Support:

US NIH-NIGMS
US NSF