

Biological evolution and innovation a complex systems approach

Overview

- Milestones of evolutionary biology
- The problem of innovation
- Metabolism: the earliest evolving systems
- Genotype networks in metabolism
- Genotype networks in other biological systems proteins, RNA, regulation circuits
- Where do genotype networks come from
- Reconciling neutralism and selectionism
- Robustness and innovability
- Evolvable and adaptable technologies

1859: Charles Darwin and «The origin of species»

Natural selection as the cause of life's diversity

All life originated from a common ancestor

Two main areas of ignorance:

"The **laws governing inheritance** are for the most part unknown."

"I have hitherto sometimes spoken as if the variations... had been due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation."

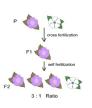


1866: Gregor Mendel

Solved part of the inheritance problem

Showed that traits can behave like particles when inherited

Forgotten and rediscovered in 1905 by de Vries





1906: Hugo de Vries

Natural selection may explain the survival of the fittest, but it cannot explain the arrival of the fittest.

Hugo de Vries



1909: Wilhelm Johannsen

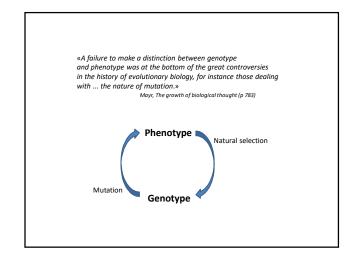
Danish botanist who coined three key concepts

Mendel's particulate units of inheritance (still completely unclear what genes were)

GenotypeIn modern terms: all genetic material

Phenotype
All other observable traits





1930-1940s: The modern synthesis

Named after a 1942 book by Julian Huxley

Main characters: R.A. Fisher, Sewall Wright, J.B.S. Haldane

Married three elements

The importance of natural selection Mendel's laws of particulate inheritance Population thinking

Created population genetics

Quantitative and mathematical theory that predicts how selection changes allele frequencies in a population

Allele frequency: the proportion of individuals that carry that allele







1930-1940s: The modern synthesis

Powerful because of two key abstractions:

1. simple phenotypes, often represented only as a scalar quantity (fitness) Evolution as "change in allele frequency within a gene pool" (Mayr)

2. a simple relationship between genotype and phenotype







1983: The neutral theory of molecular evolution

New alleles created by mutation can be neutral – they do not affect

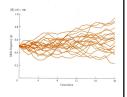
The fate of such mutations is determined by genetic drift – the random sampling of alleles from generation to generation.

Theory developed since the 1920s by Wright and Haldane.

Developed further with greater mathematical sophistication by Motoo Kimura

Genetic drift can be a strong force of evolutionary change in small populations.





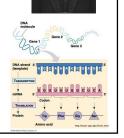
1944-: The molecular era

1944: Oswald Avery discovers that DNA is the material basis of inheritance

1953: James Watson and Francis Crick find the chemical structure of DNA

1977: Frederick Sanger describes the first efficient

1995: The genome era begins with the sequencing of Hemophilus influenzae



1966-: Neutralism and selectionism

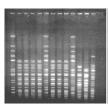
1966: Lewontin and Hubby describe great amounts of *genetic variation* in fruit flies species.

"39% of loci in the genome are polymorphic over the whole [Drosophila pseudoobscura] species. The average population is polymorphic for 30% of all loci."

Populations of many other organisms, even living fossils, also contain lots of genetic variation.

Is most of this variation neutral or adaptive, that is, influenced by selection?





J. L. Hubby and R. C. Lewontin, "A Molecular Approach to the Study of Ge

1966-: Neutralism and selectionism

Ernst Mayr, a selectionist

"It is altogether unlikely that two genes would have identical selective values under all the conditions under which they may coexist in a population... ... cases of neutral polymorphisms do not exist."

Motoo Kimura, a neutralist

"Selective elimination of definitely deleterious mutants and random fixation of selectively neutral ... mutants occur far more frequently in evolution than positive Darwinian selection of definitely advantageous mutants."





$Neutralism \ and \ selection is min \ a \ \underline{broader} sense$

"non-adaptationism and adaptionism"

 $\label{prop:continuous} What \, role \, might \, non-adaptive \, traits \, play \, in \, evolution?$

Gould and Lewontin (1979):

"We fault the adaptationist programme for its failure to distinguish current utility from reasons for origin; ... for its unwillingness to consider alternatives to adaptive stories; ... and for its failure to consider adequately such competing themes as random fixation of alleles,... and current utility as an epiphenomenon of non-adaptive structures."





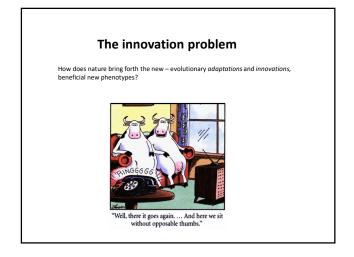
Today: Population genomics

Theory of the modern synthesis combined with genome sequencing

Great descriptive and explanatory power in many areas human history conservation genomics evolutionary origins of diseases (lactose intolerance etc.)

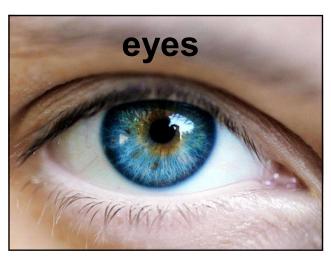
Focuses on data from genotypes

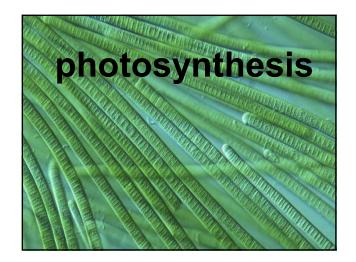
Still limited by (tacit) simplifications regading phenotypic complexity







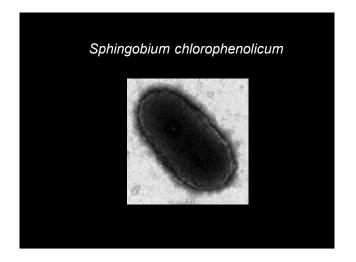


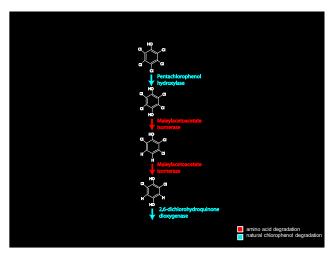




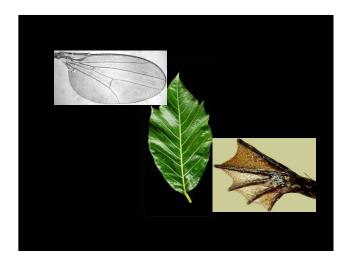
1. metabolic innovation

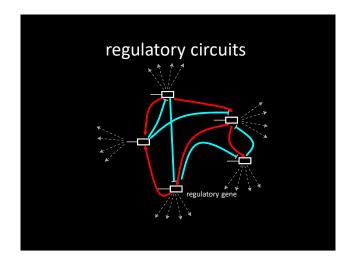


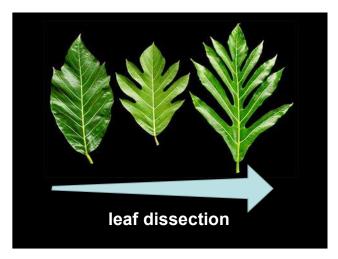


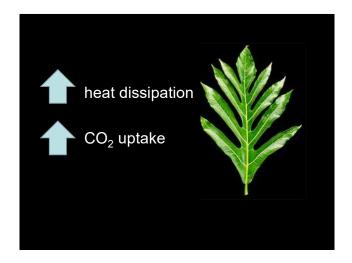


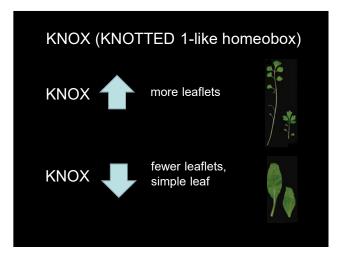
2. regulatory innovation





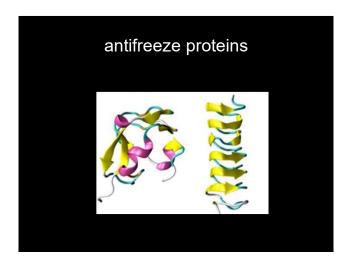


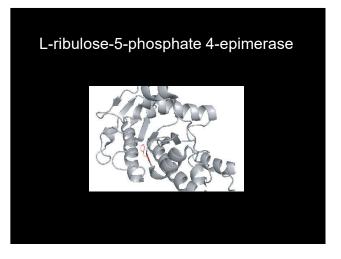


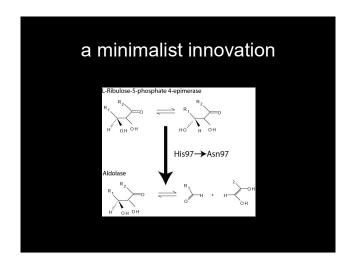


3. new macromolecules





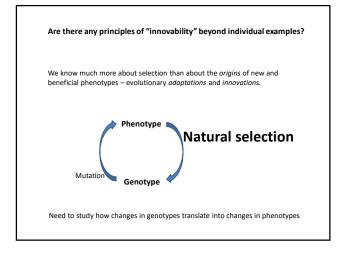


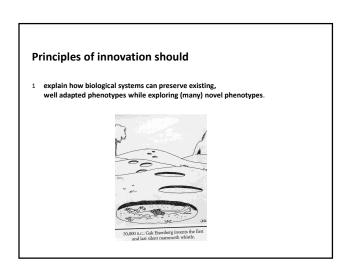


Most or all innovations involve

novel metabolic pathways novel regulation novel molecules

or (entangled) combinations thereof.





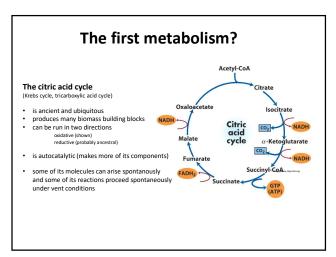
Principles of innovation should

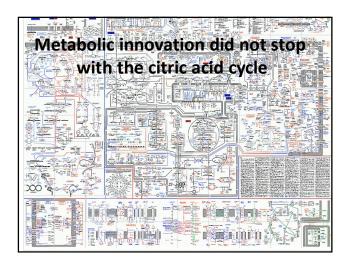
- 1 explain how biological systems can preserve existing, well adapted phenotypes while exploring (many) novel phenotypes.
- 2. unify innovations that involve different levels of biological organization.
- 3. be able to capture the combinatorial nature of innovation.
- 4. enable us to study the role of environments in promoting innovations.
- 5. explain how the same problem can have multiple different solutions
- 6. be applicable to technological systems.



Innovation in complex metabolic systems







A metabolism is a network of chemical reactions whose two main functions are to produce

chemical energy

(for maintenance of cell functions and for biosyntheses)

molecular building blocks for biosyntheses

These reactions are catalyzed by enzymes that are encoded by genes.

The metabolism of a whole organism comprises many chemical reactions

Organism	Reactions	Molecules
E. coli S. cerevisiae A. thaliana	2077 1412 1567	1039 713 1748
B. aphidicola	263	240

Metabolic genotype

The part of a genome that encodes metabolic enzymes

Less unwieldy:

An organism's set of (enzyme-catalyzed) metabolic reactions

These can be written as follows

A metabolic genotype



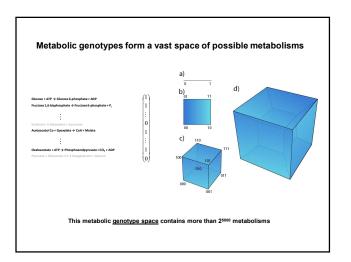
The known "universe" of metabolic reactions (>5000 reactions

The universal library

The universe (which others call the Library) is composed of an indefinite and perhaps infinite number of hexagonal galleries...there are five shelves for each of the hexagon's walls; each shelf contains thirty-five books...The Library exists ab aeterno...there are no two identical books. The Library is total and ... its shelves register all the possible combinations of the twenty-odd orthographical symbols.

Jorge Luis Borges The library of babel



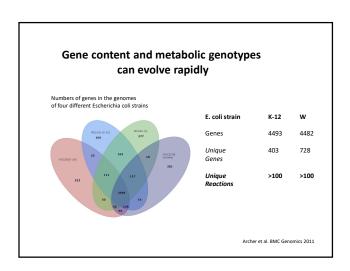


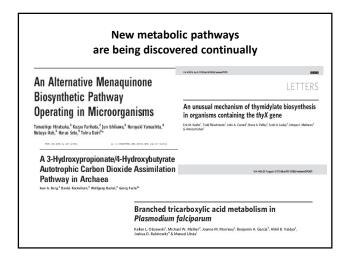
Metabolic genotypes form a vast space of possible metabolisms

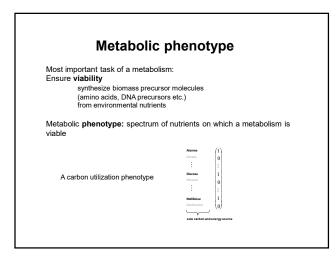
Two genotypes are <u>neighbors</u> if they differ in a single reaction.

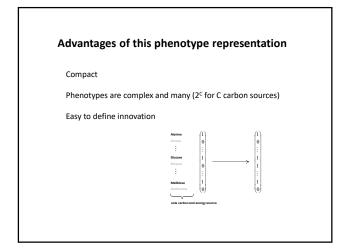
A genotype's neighborhood comprises all of its neighbors.

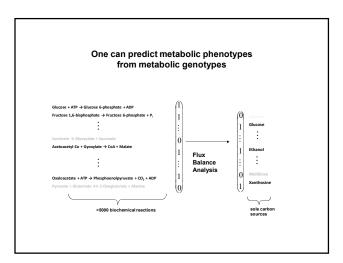
The <u>genotype distance D</u> indicates the fraction of reactions at which two networks differ.











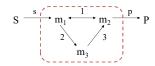
Flux balance analysis needs

- a list of chemical reactions known to be catalyzed by enzymes in a given organism
- Information about nutrients in the chemical environment of a cell and their uptake rate (usually in mol/g dry weight [DW] and hour)

Flux balance analysis computes

- $1. \, \underline{allowable} \, metabolic \, fluxes \, through \, a \, metabolic \, network \, (fluxes \, that \, do \, not \, violate \, the \, law \, of \, mass \, conservation)$
- 2. within the set of allowable fluxes, those that have desirable properties (e.g., maximal rate of biomass production, maximal biomass yield per unit of carbon source.)

A simple chemical reaction network



Two external metabolites

1 substrate (nutrient) S 1 product P

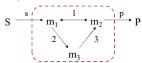
Two transport reactions (s,p)

- s determines the uptake rate of S
- p determines the rate at which P is excreted

Three internal metabolites (m_i)

Three internal reactions (1,2,3)

A simple chemical reaction network



Metabolite concentrations m_i change according to the equations

$$\frac{dm_1}{dt} = v_s - v_1 - v_2$$

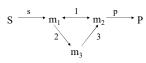
$$\frac{dm_2}{dt} = v_1 + v_3 - v_\rho$$

$$\frac{dm_3}{dt} = \mathbf{S}\vec{v}$$

$$\mathbf{S} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix}$$

v_i metabolic flux through reaction i

Stoichiometry matrix

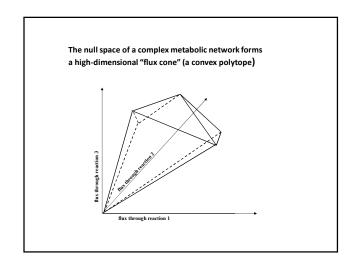


FBA assumes that metabolism is in a steady state where the concentrations of metabolites no longer change

$$\frac{d\vec{m}}{dt} = 0$$

$$\mathbf{S}\vec{v} = 0$$

The solutions of these equations are the allowable metabolic fluxes. They form the so-called <u>null space of S</u>



Several important properties of a metabolic network can be expressed as a weighted sums of fluxes

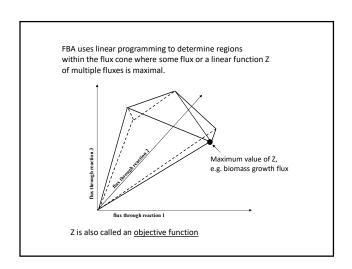
$$Z(\vec{v}) = \sum_{i=1}^{m} c_i v_i$$

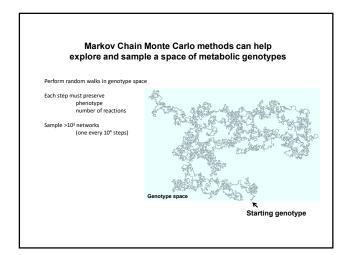
Example:

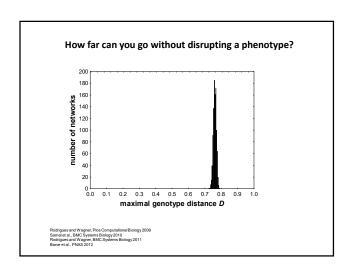
In the biomass growth reaction,

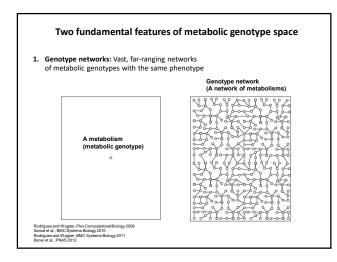
 \mathbf{v}_i is the rate at which essential biochemical precursor i is produced by a metabolic network.

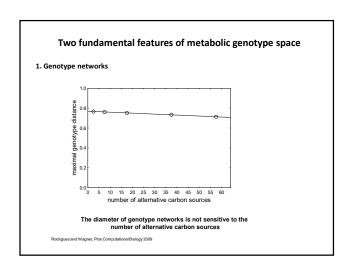
 c_i is a constant that reflects the relative contribution of precursor i to biomass (can be estimated from the biomass composition of a cell.)

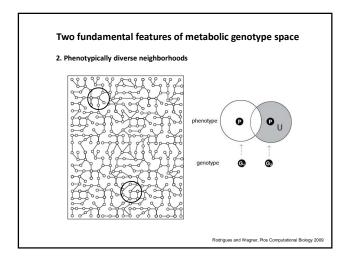


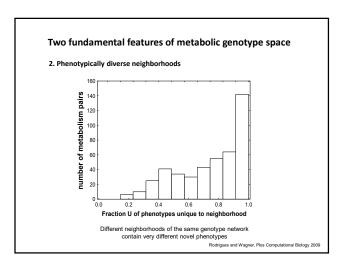




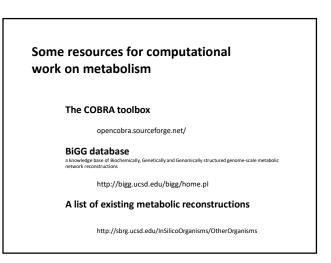








Two fundamental features of metabolic genotype space. Metabolisms form vast phenotype-preserving genotype networks that extend far through genotype space Different neighborhoods of these metabolisms contain very different novel phenotypes



Gene regulatory circuits create gene expression patterns that are the basis of many adaptations and innovations

Genotype: circuit topology, a pattern of regulatory interactions

Phenotype: the activity or expression pattern of circuit genes

Gene regulatory circuits form vast genotype spaces of circuits that vary in their topologies

Neighbors: circuits that differ in exactly one (cis)-regulatory interaction

Circuit (genotype) distance: fraction of regulatory interactions in which two circuits differ.

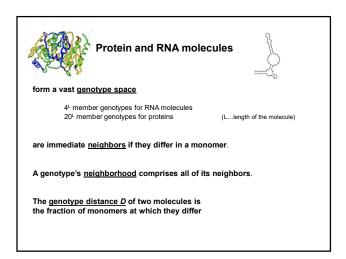
Wagner, Evolution 1986
Cilied, Marin, and Wagner, PNAS 2007
Cilied, Marin, Cil

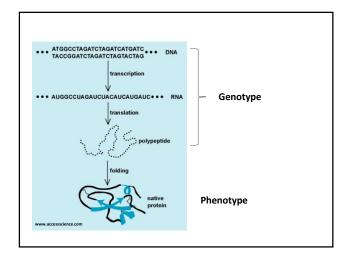
Regulatory circuits with the same gene activity phenotype form vast genotype networks that extend far through genotype space

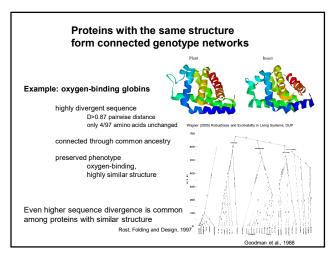
Different neighborhoods of these networks contain very different novel phenotypes

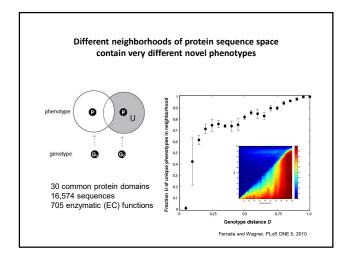
Output

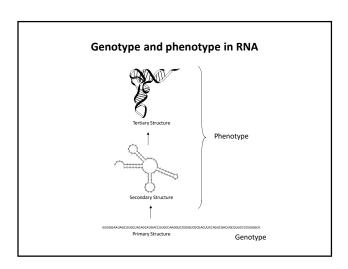
Outpu









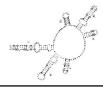


RNA secondary structures are often themselves functionally important phenotypes

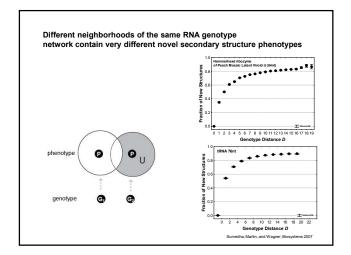
Internal ribosomal entry sites (IRES) in the genome of Hepatitis C virus and other flaviviridae $\,$

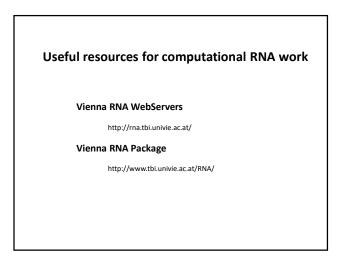
TAR region of HIV-1, necessary for viral RNA replication

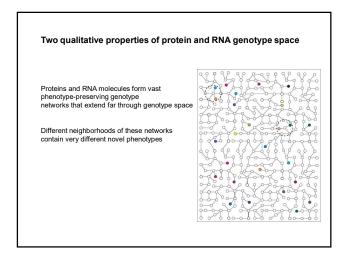
Rev responsive element (RRE) region of HIV-1, interaction with the viral Rev protein influences splicing patterns.

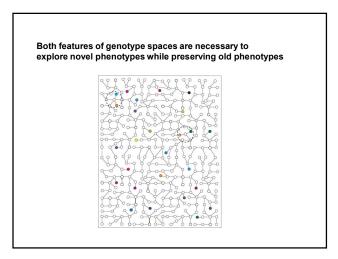


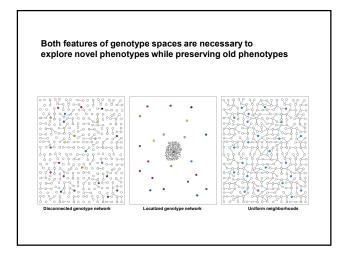
RNA sequences with the same secondary structure form vast connected genotype networks that span genotype space Discovered by Schuster and collaborators (Proc. Roy. Soc.1994) and called 'neutral networks' Maximal genotype distance of RNA sequence with the same structure after 5,000 single structure-preserving nucleotide changes.

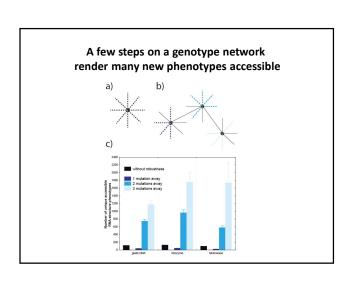


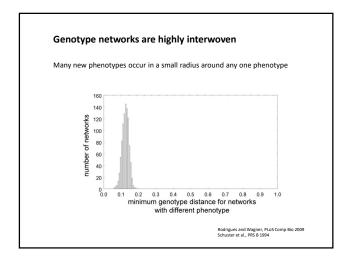












Summary (so far)

Three very different biological systems

metabolic networks regulatory circuits protein and RNA molecules

that are involved in most evolutionary adaptations and innovations

share two fundamental features

vast genotype networks that preserve phenotype and extend far through genotype space

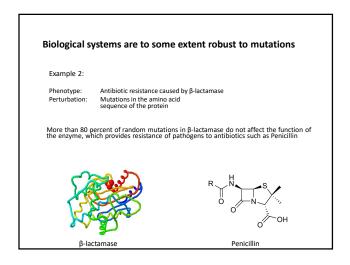
different neighborhoods of any one genotype network contain many diverse phenotypes

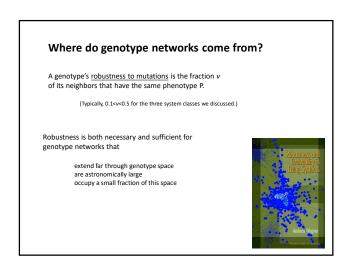
Both features are required for the exploration of novel phenotypes while preserving old phenotypes

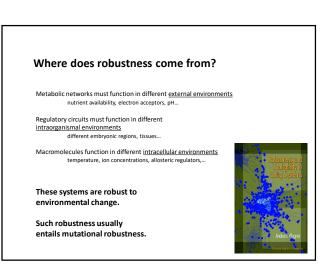
The origins of genotype networks

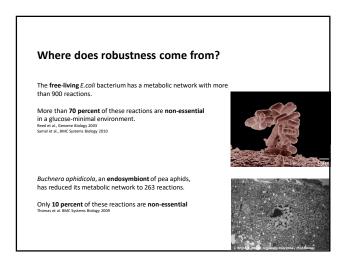
Robustness is a system's ability to preserve a system <u>property</u> (phenotype) in response to <u>perturbations</u>

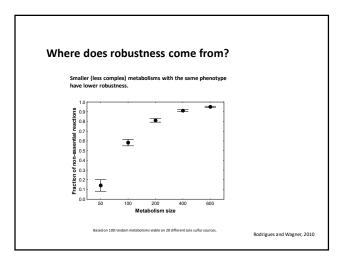
Biological systems are to some extent robust to mutations Example 1: Phenotype: Rate of cell growth and division Perturbation: Eliminations of genes from a genome Smith et al. Nature 275, 464 (1997)

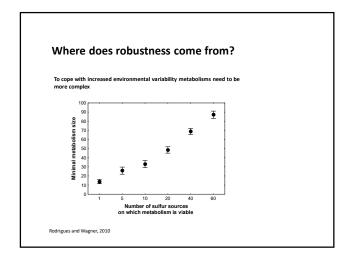






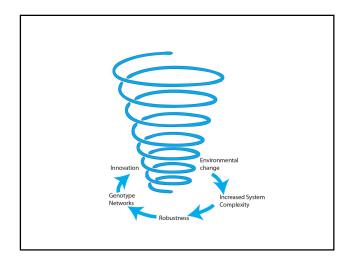






Greater environmental variability requires greater system complexity

Greater complexity entails robustness to genetic change $\underline{\text{in any}}$ $\underline{\text{one environment}}$



Summary of the origins of genotype networks

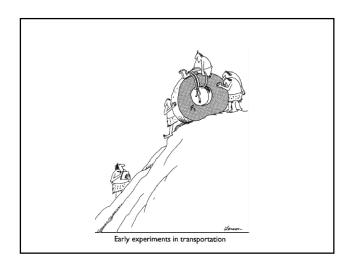
Biological systems are to some extent robust to mutations

 $\label{lem:mutational} \mbox{Mutational robustness is correlated with robustness to environmental change}$

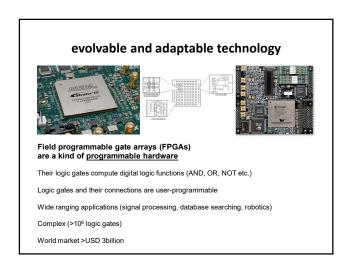
 $\label{prop:control} \mbox{Much of this robustness originates from adaptation to changing environments}$

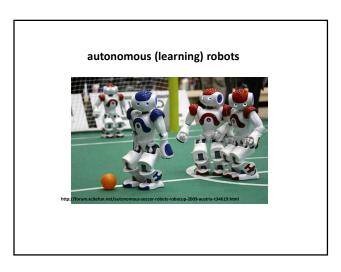
A robust genotype has many neighbors with the same phenotype $\,$

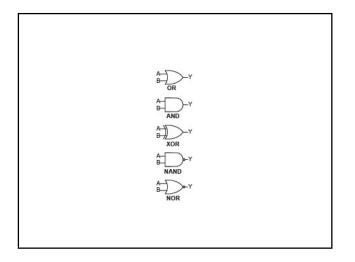
Mutational robustness is both necessary and sufficient for the existence of genotype networks

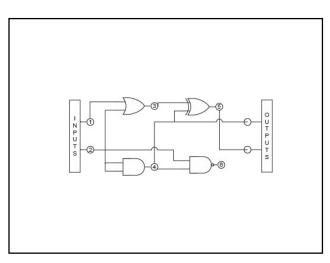


digital electronics



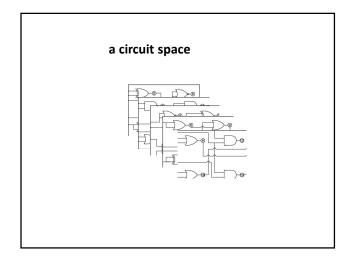


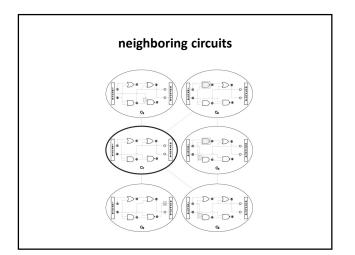


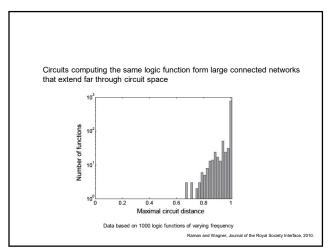


Boolean logic

 $011010101111010101 \rightarrow 1111010101010110101$







Two maximally different circuits that both compute the circular shift function.

They are two among many different solutions to the same problem.

