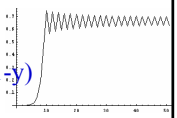
## Integrate 1: Minimal & exponential Systems (last week)

Life & computers : Self-assembly
Math: be aware of assumptions & approximations

Catalysis & Replication
Differential equations: dy/dt=ky(1-y)



Mutation & the single molecule: Noise is overcome
Directed graphs & pedigrees
Bell curve statistics: Binomial, Poisson, Normal
Selection & optimality

1

## Integrate 1: Minimal & exponential Systems (last week)

Big questions: How can we define (& design) "General biology" (self-replicating systems)?

How soon might the exponential of IT and/or Biotech overtake human intelligence?

DNA single molecule stochastics is a given.
How does life reduce the noise by dozens of logs?

2

## Integrate 2: Optimal BioSystems

•Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

3

## From atoms to (bio)molecules

| | | |
|---|---|---|
| $H_2O$ | $H_2$ , $O_2$ | $H^+$ , $OH^-$ |
| $CH_4$ | $C_{60}$ | $CO_3^-$ |
| $NH_3$ | $N_2$ | $NO_3^-$ |
| $H_2S$ | $S_n$ | $SO_4^{--}$ $Mg^{++}$ |
| $PH_3$ | | $K^+PO_4^{--}$ $Na^+$ |
| **Gas** | **Elemental** | **Salt** |

4

## Purify

**Elements, molecules, assemblies, organelles, cells, organisms**

chromatography

Clonal growth

## Purified history

**Pre 1970s: Column/gel purification revolution**

**Mid-1970s: Recombinant DNA brings clonal (single-step) purity.**

**1984-2002: Sequencing genomes & automation aids return to whole systems.**

6

## Integrate 2: Optimal BioSystems

Elements & Purification
•Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

7

## Systems Engineering to Systems Biology

**1950 Norbert Wiener, Cybernetics**

**1951 Alan Turing reaction-diffusion equations**

**…**

**2003 Price ND, et al. Trends Biotechnol. 21(4):162-9. Genome-scale microbial in silico models: the constraints-based approach.**

**2003 Wolf DM, Arkin AP. Curr Opin Microbiol. 6(2):125-34. Motifs, modules and games in bacteria.**

**2002 Segre, D, et al. Analysis of optimality in natural and perturbed metabolic networks . PNAS 99: 15112-7.**

http://www.wolframscience.com/reference/notes/1003g

8

## Why Genomes & Systems?

#0. Why sequence the genome(s)? To allow #1,2,3 below.

#1. Why map variation?
#2. Why obtain a complete set of human RNAs, proteins & regulatory elements?
#3. Why understand comparative genomics and how genomes evolved? To allow #4 below.

#4. Why quantitative biosystem models of molecular interactions with multiple levels (atoms to cells to organisms & populations)?
To share information. Construction is a test of understanding & to make **useful products**.

9

## Grand (& useful) Challenges

**A) From atoms to evolving minigenome-cells.**
- Improve *in vitro* macromolecular synthesis.
- Conceptually link atomic (mutational) changes to population evolution (via molecular & systems modeling).
- Novel polymers for smart-materials, mirror-enzymes & drug selection.

**B) From cells to tissues.**
- Model combinations of external signals & genome-programming on expression.
- Manipulate stem-cell fate & stability.
- Engineer reduction of mutation & cancerous proliferation.
- Programmed cells to replace or augment (low toxicity) drugs.

**C) From tissues to physio- & eco- systems**
- Programming of cell and tissue morphology.
- Quantitate robustness & evolvability.
- Engineer sensor-effector feedback networks where macro-morphologies determine the functions; past (Darwinian) or future (prosthetic).

10

## Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
•Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

11

## Number of component types (guesses)

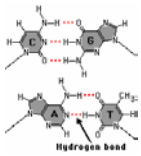
| | M.gen | Worm | Human |
|---|---|---|---|
| Bases | .58M | >97M | 3000M |
| DNAs | 1 | 7 | 25 |
| Genes | .48k | 19k | 21k |
| RNAs | .4k | >30k | .2-3M |
| Proteins | .6k | >50k | .3-10M |
| Cells | 1 | 959 | $10^{14}$ |

http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v409/n6822/full/409860a0_fs.html
http://www.bio-itworld.com/news/071503_report2902.html

12

## From monomers to polymers

Complementary surfaces
Watson-Crick base pair
(Nature April 25, 1953)

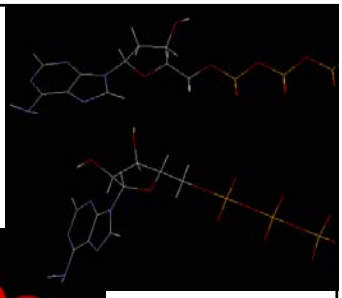


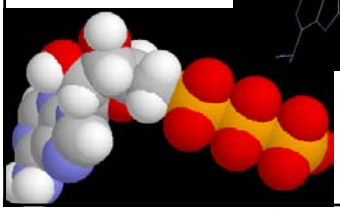

13

## Nucleotides

dATP

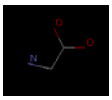
rATP



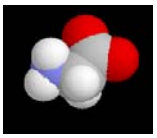


14

## The simplest amino acid component of proteins



**Glycine**
**Gly**
**G**



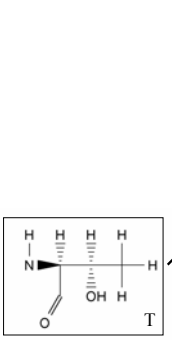
```
config(glycine,[
    substituent(aminoacid_L_backbone),
    substituent(hyd),
    linkage(from(aminoacid_L_backbone,car(1)),
        to(hyd,hyd(1)),
        nil,single)]).
```

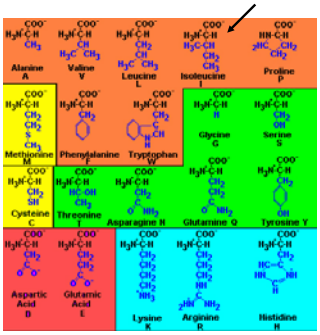
Klotho

Smiles String: [CH2]([NH3+])[C](=[O])[O-]

15

## 20 Amino acids of 280






www.people.virginia.edu/~rjh9u/aminacid.html
www-nbrf.georgetown.edu/pirwww/search/textresid.html

16

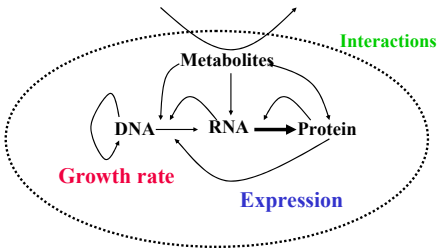
## Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
•Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

17

## Continuity of Life & Central Dogma

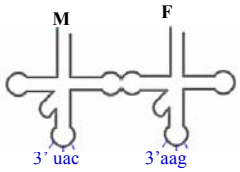
Self-assembly, Catalysis, Replication, Mutation, Selection
Regulatory & Metabolic Networks

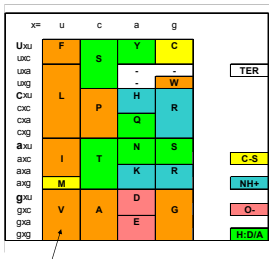



**Polymers:** Initiate, Elongate, Terminate, Fold, Modify, Localize, Degrade

18

# "The" Genetic Code

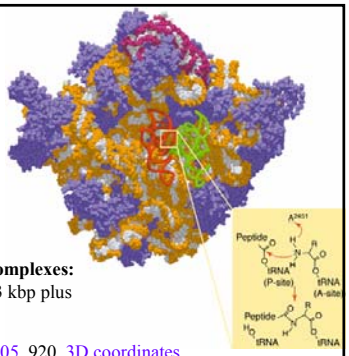



'Silent' codon changes

19

# Translation
# t-,m-,r-RNA



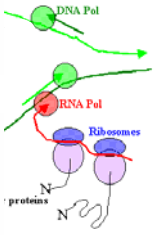

**Large macromolecular complexes:**
Ribosome: 3 RNAs (over 3 kbp plus over 50 different proteins)

Science (2000) 289: 878, 905, 920, 3D coordinates.
The ribosome is a ribozyme.

20

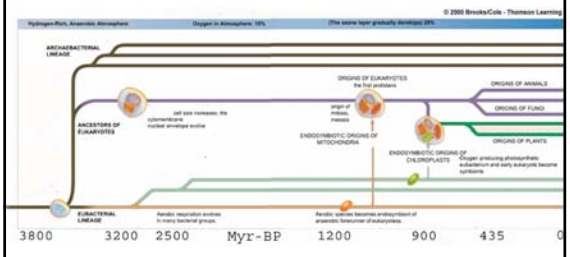
# Perl Dogma (EditPlus)

```
4  #### The Central Dogma ####
5  ############################
6
7  ### Genome
8  $DNA_seq = "ATGACCCTACTAGATCATCTATGATAGCTCAT";
9
10 ### Transcription
11 $RNA_seq = $DNA_seq;
12 $RNA_seq =~ s/T/U/gi;
13 print "$RNA_seq\n";
14
15 ### Translation
16 $position = 0;
17 while (substr $RNA_seq,$position,3) {
18       $codon = substr $RNA_seq,$position,3;
19       print translate_codon($codon);
20       $position = $position + 3;
21 }
22 sub translate_codon {
23       if ($_[0] =~ /GC./i) {return Ala;}
24       if ($_[0] =~ /UGC|UGU/i) {return Cys;}
```




21

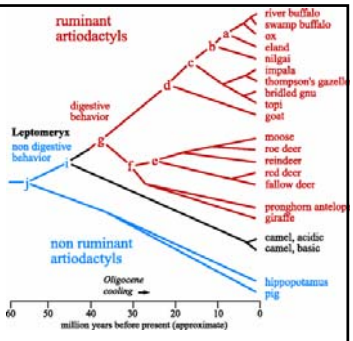
# Continuity & Diversity of life




22

http://www.colorado.edu/epob/epob1030cornwall/fall_2001/Evolution_of_Life.gif

# Evidence




Radiochemical dating:
Initial atoms remaining
$f = 1 - \exp(-kt)$ (endpt=0)
Molecular: endpt codon bias $b$.
Conserved silent codons $f_2 = b + (1 - b)\exp(-kt)$,

Benner, et al. (2002) Science 296:864
Planetary Biology--Paleontological, Geological, & Molecular Histories of Life.

23

# How many living species?

5000 bacterial species per gram of soil (<70% DNA bp identity)
Millions of non-microbial species (& dropping)
Whole genomes: 100 done since 1995, 700 in the pipeline! (ref)
Sequence any: 16234 (in 1995) to **79961** species (in 2000) NCBI

& Why study more than one species?
Comparisons allow discrimination of subtle functional constraints.

24

## Genetic codes (ncbi)

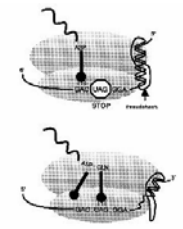
```
1. "Standard Code"
Base1  = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG
Base2  = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG
Base3  = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
  AAs  = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M---------------M----------------------------
2. The Vertebrate Mitochondrial Code
  AAs  = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSS**VVVVAAAADDEEGGGG
Starts = --------------------------------MMMM---------------M------------
3. The Yeast Mitochondrial Code
  AAs  = FFLLSSSSYY**CCWWTTTTPPPPHHQQRRRRIIMMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ----------------------------------MM----------------------------
11. The Bacterial "Code"
  AAs  = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M------------MMMM---------------M------------
14. The Flatworm Mitochondrial Code
  AAs  = FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts = ----------------------------------M----------------------------
22. Scenedesmus obliquus mitochondrial Code
  AAs  = FFLLSS*SYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ----------------------------------M----------------------------
```



## Translational reprogramming

Gesteland, R. F. and J. F. Atkins. 1996. Recoding - Dynamic reprogramming of translation (1996). Ann. Rev.Biochem 65:741-768

Herbst KL, et al. 1994
PNAS 91:12525-9
A mutation in ribosomal protein L9 affects ribosomal hopping during translation of gene 60 from bacteriophage T4.
**"Ribosomes hop over a 50-nt coding gap during translation..."**

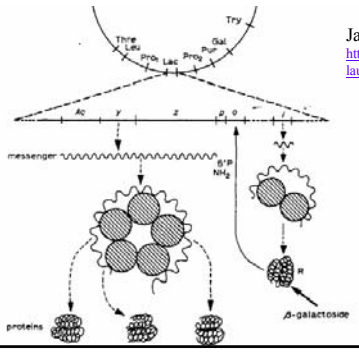



## Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
•Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection



## Qualitative Models & Evidence




Jacob & Monod
http://www.nobel.se/medicine/laureates/1965/jacob-lecture.html



## Qualitative biological statements (beliefs) and evidence

- metabolism
- cryptic genes
- information transfer
- regulation
  - type of regulation
  - genetic unit regulated
  - trigger
  - trigger modulation
- transport
- cell processes
- cell structure
- location of gene products
- extrachromosomal
- DNA sites

Riley, GeneProtEC

MIPS functions



## Gene Ontology (nature of being)

The objective of GO is to provide controlled vocabularies for the description of the molecular function, biological process and cellular component of gene products.
...
Many aspects of biology are not included (domain structure, 3D structure, evolution, expression, etc.)... small molecules (Klotho or LIGAND )

## Gene Ontology

GO

• **Molecular function**
What a gene product can do without specifying where or when. (e.g. broad "enzyme" ; narrower "adenylate cyclase“)

• **Biological process**
>1 distinct steps, time, transformation (broad: "signal transduction." narrower: "cAMP biosynthesis.")

• **Cellular component**
part of some larger object, (e.g. ribosome)_

31

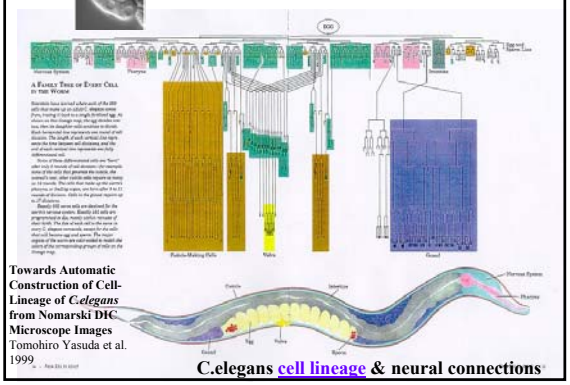
## Evidence for facts

GO

IMP inferred from mutant phenotype
IGI genetic interaction
IPI physical interaction
ISS sequence similarity
IDA direct assay
IEP expression pattern
IEA electronic annotation
TAS traceable author statement
NAS non-traceable author statement

32

## Direct observation

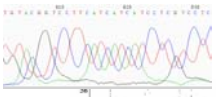


**Towards Automatic Construction of Cell-Lineage of *C.elegans* from Nomarski DIC Microscope Images**
Tomohiro Yasuda et al. 1999

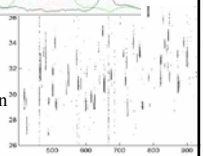
**C.elegans cell lineage & neural connections**

## Sources of Data for Qualitative Models:

Capillary electrophoresis (DNA Sequencing) : 0.4Mb/day



Chromatography-Mass Spectrometry (eg. peptide LC-ESI-MS) : 20Mb/day



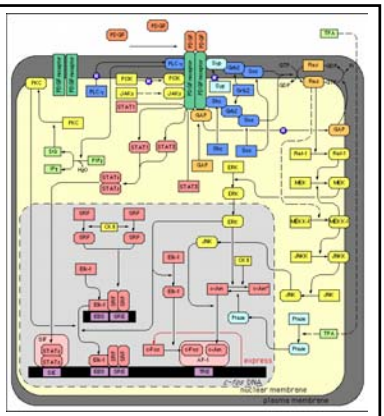

Microarray scanners (eg. RNA): 300 Mb/day

Other microscopy (e.g. subcell, cell, tissue networks)

34

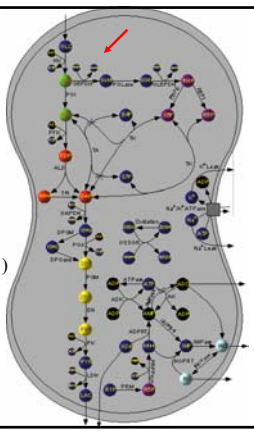
## Signaling PAthway Database SPAD

>500 bio-databases
How are the data & models entered?



## Dynamic simulation of the human red blood cell metabolic network.

Jamshidi, et al(2001) Bioinformatics 17: 286-287.



Dominant alleles affecting variety of RBC proteins, malaria, drug-hemolysis, etc. Rare individually, common as a group.

36

# Enzyme Kinetic Expressions

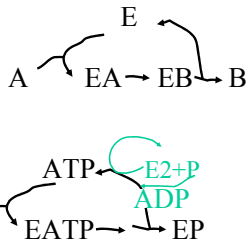
Phosphofructokinase

$$v_{PFK} = \frac{v_{max}^{PFK}}{N_{PFK}} \left( \frac{F6P / K_{F6P}^{PFK}}{1 + F6P / K_{F6P}^{PFK}} \right) \left( \frac{Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}}{1 + Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}} \right)$$

$$N_{PFK} = 1 + L_0^{PFK} \frac{\left(1 + ATP_{free} / K_{ATP}^{PFK}\right)^4 \left(1 + Mg / K_{Mg}^{PFK}\right)^4}{\left(1 + AMP / K_{AMP}^{PFK}\right)^4 \left(1 + F6P / K_{F6P}^{PFK}\right)^4}$$

37

# How do enzymes & substrates formally differ?

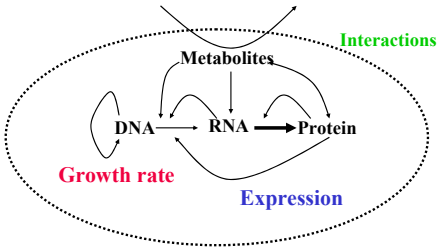



Catalysts increase the rate (&specificity) without being consumed.

38

# Continuity of Life & Central Dogma

Self-assembly, Catalysis, Replication, Mutation, Selection
Regulatory & Metabolic Networks




**Polymers:** **Initiate, Elongate, Terminate, Fold, Modify, Localize, Degrade**

39

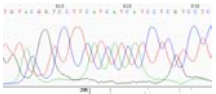
# Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
•Functional Genomics & Quantitative models
Mutations & Selection

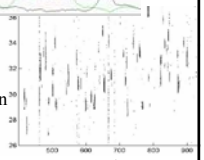
40

# Sources of Data for Quantitative Models:

Capillary electrophoresis
(DNA Sequencing) :
0.4Mb/day



Chromatography-Mass Spectrometry
(eg. peptide LC-ESI-MS) :
20Mb/day




Microarray scanners (eg. RNA):
300 Mb/day mpg

Other microscopy (e.g. subcell, cell, tissue networks)

41

# Structural Genomics

(the challenge of distant homologs)

? ↓ ?

# Functional Genomics

(quantitative ligand interactions)

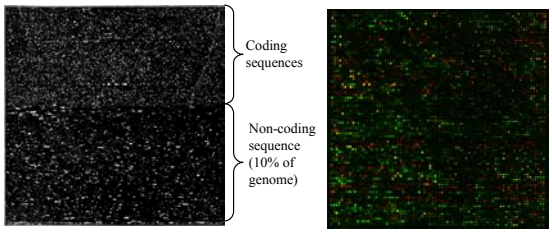
**100% Sequence Identity:**
1. Enolase Enzyme
2. Major Eye Lens Protein

**100% Sequence Identity:**
1. Thioredoxin Redox
2. DNA Polymerase Processivity

42

## mRNA expression data

& protein binding & mutant growth …




Affymetrix *E. coli* oligonucleotide array

Spotted microarray mpg

43

## What is functional genomics?

**Function** (1): Effects of a mutation on fitness (reproduction) summed over typical environments.
Function (2): Kinetic/structural mechanisms.
Function (3): Utility for engineering relative to a non-reproductive objective function.

**Proof** : Given the assumptions, the odds are that the hypothesis is wrong less than 5% of the time, keeping in mind (often hidden) multiple hypotheses.

**Is the hypothesis suggested by one large dataset already answered in another dataset?**

44

## Genomics Attitude

**Whole systems:**
Less individual gene- or hypothesis-driven experiments;
Automation from cells to data to **model** as a proof of protocol.

**Quality of data:** DNA sequencing raw error: 0.01% to 10%.
Consensus of 5 to 10 error: 0.01% (1e-4)

**Completion:** No holes, i.e. regions with data of quality less than a goal (typically set by cost or needs of subsequent projects).

**Impossible**: The cost is higher than reasonable for a given a time-frame and quality assuming no technology breakthroughs. Cost of computing vs. experimental "wet-computers".

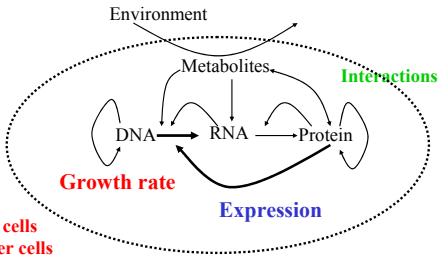
45

## Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
•Mutations & Selection

46

## Mutations and selection

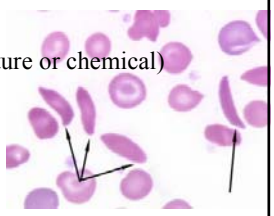



**stem cells**
**cancer cells**
**viruses**
**organisms**

47

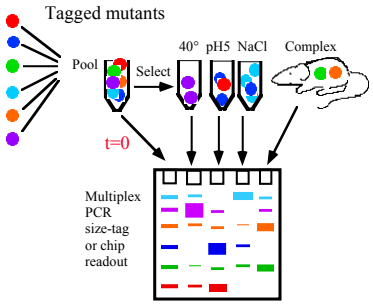
## Types of Mutants

Null: PKU
Dosage: Trisomy 21
Conditional (e.g. temperature or chemical)
Gain of function: HbS
Altered ligand specificity



48

## Multiplex Competitive Growth Experiments




49

## Ratio of strains over environments, **e** , times, $\mathbf{t_e}$ , selection coefficients, $\mathbf{s_e}$, $R = R_o \exp[-\Sigma s_e t_e]$

80% of 34 random yeast insertions have s<-0.3% or s>0.3% t=160 generations, e=1 (rich media); ~50% for t=15, e=7. Should allow comparisons with population allele models.

**Multiplex competitive growth experiments:**
Thatcher, et al. (1998) PNAS 95:253.
Badarinarayana, et al. (2001) Nature Biotech.19: 1060.
Smith V, et al. (1995) PNAS 92:6479.
Shoemaker D, et al. (1996) Nat Genet 14:450.

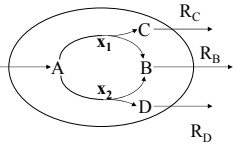
50

## Steady-state flux optima

**Flux Balance Constraints:**

$R_A < 1$ molecule/sec (external)
$R_A = R_B$ (because no net increase)
$x_1 + x_2 < 1$ (mass conservation)
$x_1 > 0$ (positive rates)
$x_2 > 0$

$Z = 3R_D + R_C$




51

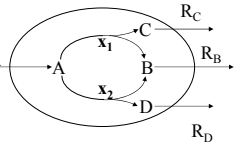
## Steady-state flux optima

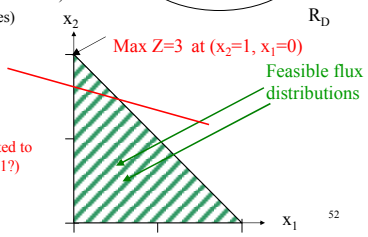
**Flux Balance Constraints:**

$R_A < 1$ molecule/sec (external)
$R_A = R_B$ (because no net increase)
$x_1 + x_2 < 1$ (mass conservation)
$x_1 > 0$ (positive rates)
$x_2 > 0$

$Z = 3R_D + R_C$

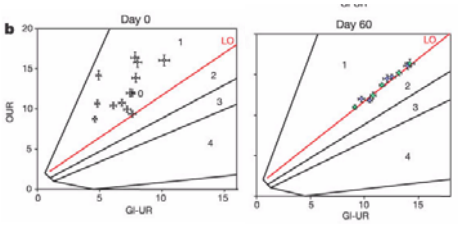
(But what if we really wanted to select for a fixed ratio of 3:1?)




52

## Non-optimal evolves to optimal




Ibarra et al. Nature. 2002 Nov 14;420(6912):186-9. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth.

53

## Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

54