

How is the Landscape upon which Life Evolves Selected? Spontaneous Emergence of Modularity



feature article

Mathematical adventures in biology

Michael W. Deem

A short personal tour of biological systems reveals the flavor and variety of biological questions amenable to illumination by mathematical analysis.

Michael W. Deem is a professor of physics and astronomy and the John W. Cox Professor in Biochemical and Genetic Engineering at Rice University in Houston, Texas.

The contemplation and resolution of questions at the intersection of biology, mathematics, and physics promise to lead to a greater understanding of the natural world and to open new avenues for physics. The choice of questions in this article, most of them related to the statistical behavior of biological systems, reflects my own research interests. But it also reflects my belief that some of the most vexing issues in mathematical biology are related to the diversity, randomness, variation, and correlations in biology. With luck, physics and a good measure may shed further light on these issues.

The mathematical focus of the research questions I propose complements the public-health focus of the 14 Grand Challenges in Global Health announced in October 2003. These challenges have realigned health-related research priorities—not only those of the Bill & Melinda Gates Foundation, which initiated the identification of the challenges, but also those of other public and private research funding bodies. Several of the challenges concern the development of vaccines and drug treatments, topics that invite mathematical analysis.

The following are a few suggestions for mathematical adventures in biology. The first topic concerns the role of localization, discreteness, and fluctuations in natural biological systems. Cells are not simply well-mixed test tubes, as evidenced by the role of fluctuations in a some of life's processes. The second topic concerns a general question about the mechanism of evolution in biology. Mathematical results might quantify the probability of life as we know it and might be tested in laboratory evolution experiments. The last topic addresses a practical aspect of pathogen evolution. In this case, mathematical results might aid drug or vaccine design, or they might predict the likelihood of future epidemics or the emergence of new pathogens.

The answers to many of all of the questions suggested in the article lie within the grasp of the physics and biology communities. The answers can be tested experimentally and in the interpretation of biological data, and have direct implications for medical practice. They may lead to a more thorough understanding of biology and small likelihood will lead to an ever-expanding set of new questions.

The roll of fluctuations
From the cell to the organism, what role does fluctuation play in natural function? Cellular noise arising from the nonuniform distribution of biological molecules in the cell is now appreciated as an interesting physics problem. Figure 1, for example, shows random variations that arise in the concentration of a protein due to stochastic variation of both these quantities outside the cell and the biochemical reactions within a cell.

Many of the molecules of life, such as DNA, mRNA, and enzymes, are present at exceedingly low concentrations—often 1 to 100 molecules per cell. At such low concentrations the traditional equations of solution chemistry no longer apply and the stochastic fluctuations associated with the transport and reactions of individual molecules become apparent. At equilibrium, the thermodynamics of a small system (see the article by Carlos Bustamante, Jan Lipfert, and Felix Ritort, *PHYSICS TODAY*, July 2005, page 43), an important detail of biology is that many interesting reactions occur at nonequilibrium—that is, when equilibrium has not been established and when dynamics is important. Indeed, one of the differences between much of biology and traditional condensed-matter physics, which has provided useful techniques to the mathematics of biology, is the importance of finite-time and finite-population size.

A possible way to treat the randomness that arises from reactions at low concentrations is to write a master equation for the process and map that equation onto a stochastic quantum field theory. To date, that approach has been pursued only for the simplest reactions. An alternative, approximate treatment uses mean-field theory, but only in the regions of parameter space where at least one molecule is likely to be present; elsewhere the concentration is simply set to zero. The experiments shown in figure 1 could be described with such a theory, in combination with a description of the statistics of the random extracellular environment.

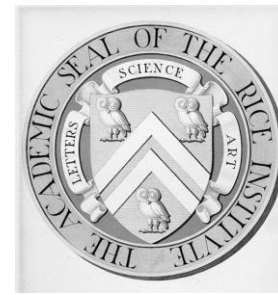
One might generally ask of biology, under what conditions does the integer constraint on the number of molecules or individuals matter? An integer constraint can induce correlations that traditional mean-field solution-chemistry theory would miss. And traditional theories of population evolution are mean field, so one might expect them to break down in the low concentrations that arise when new drug-resistant strains of bacteria evolve, as pathogens are eradicated, or when species are near extinction.

42 January 2007 Physics Today

© 2007 AMERICAN PHYSICAL SOCIETY, 0031-9228/07/0042-0042



Michael W. Deem
Rice University



NSF
NIH
DOE
DARPA

Outline

- Prevalence of modular structures
- Is modularity inevitable?
- Spontaneous emergence of modularity
- Evidence from Nature

M. W. Deem and D. J. Earl, *PNAS* **101** (2004) 11531

J. Sun and M. W. Deem, *PRL* **99** (2007) 228107

J. He, J. Sun, and M. W. Deem, submitted



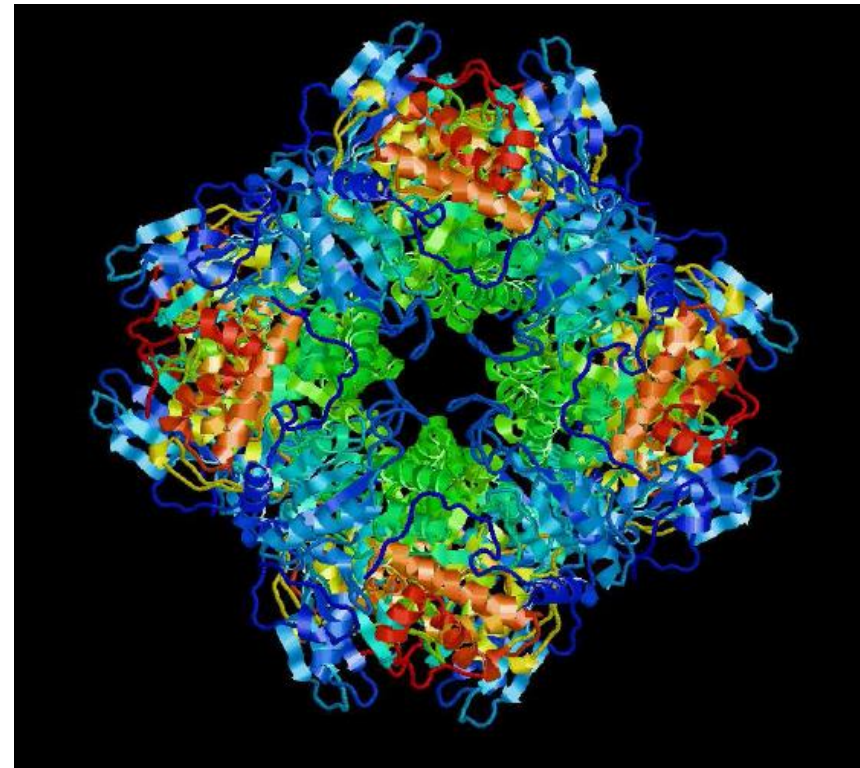
Keck Hall, Rice University

Modular Structures

- Is it true that evolution of a modular molecular structure allows life to evolve at a rapid rate, such that we can exist?
- A modular structure to the molecules of life allows for biological information to be stored in pieces
- Evolution can proceed not just by changing one base of the genetic code or movement of one atom or amino acid at a time, but rather by exchange of these functional chunks among living organisms

Proteins are Modular

- Proteins are composed of structurally-distinct, smaller modules
- Why is modularity and hierarchy so prevalent in biology?



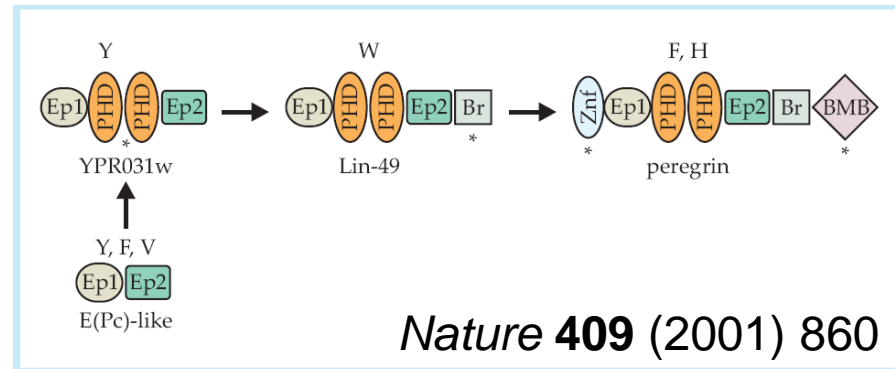
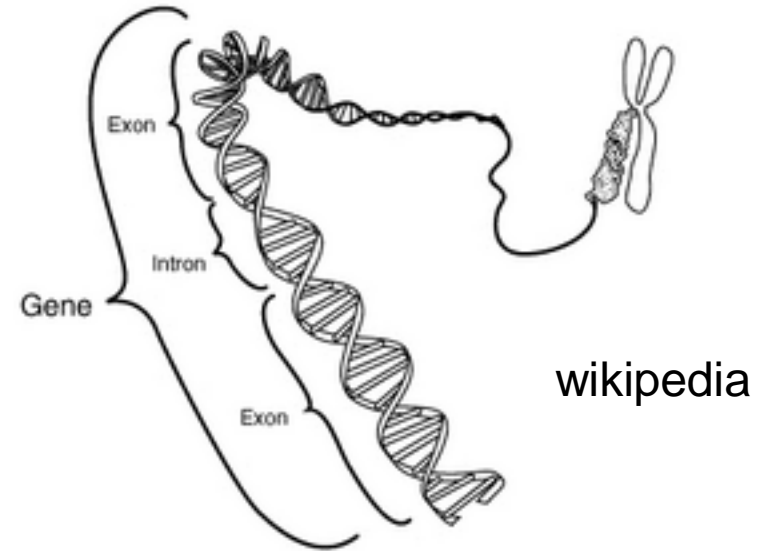
ribulose biphosphate carboxylase
1RCX

Modularity and Evolution

- Proteins are often made up of almost independent modules, which may be exchanged through evolution
- Identifiable elements at the levels of atoms, amino acids, secondary structures, domains, proteins, multi-protein complexes, pathways, organelles, cells, organs, individuals, species, and so on

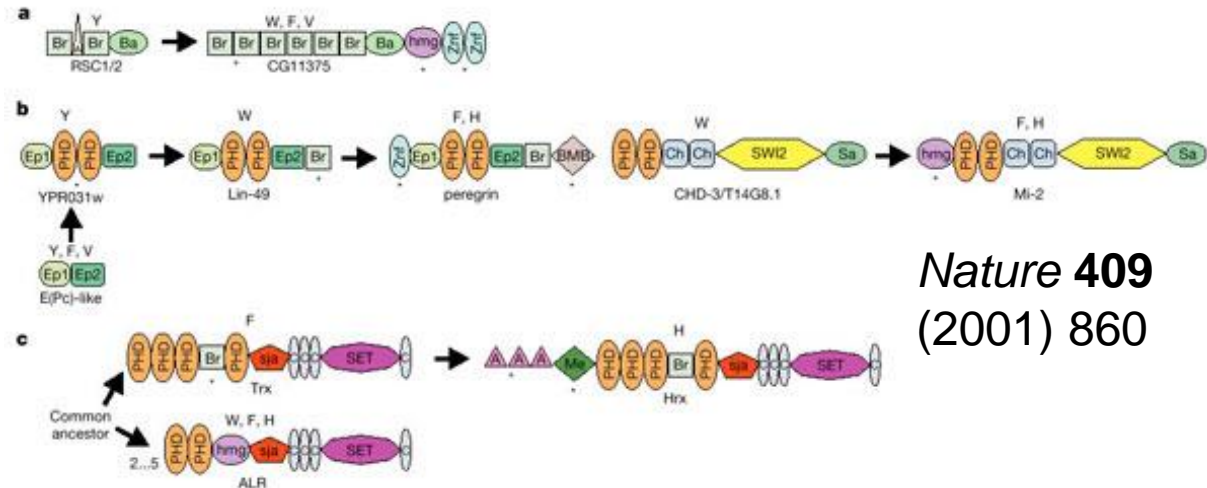
Genes are Modular

- DNA pieces that encode distinct protein modules become organized and concatenated in the course of evolution
- Evolution of E(Pc)-like protein in yeast (Y), fly (F), and vertebrate (V) to peregrin protein in fly and human (H)

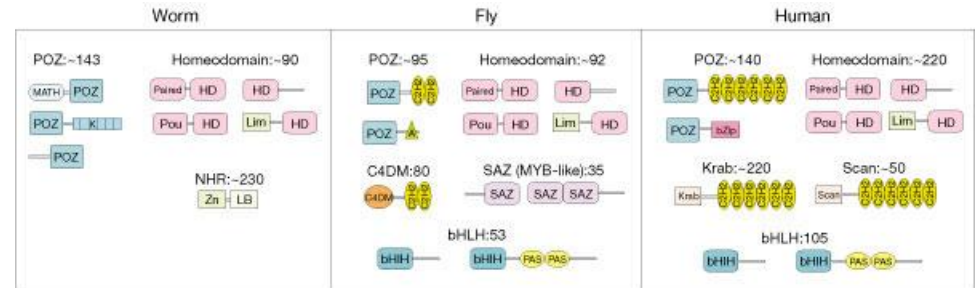


Modularity is Common

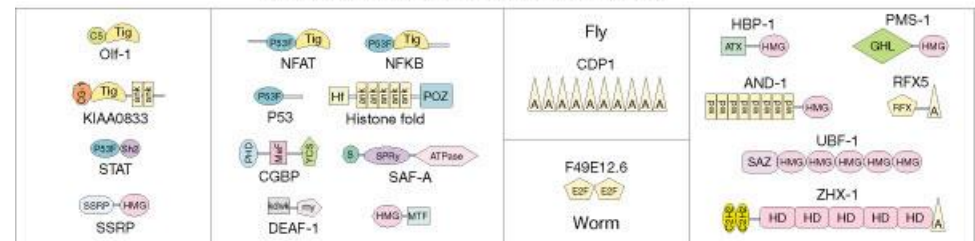
- Top) Expansion of human chromatin proteins due to insertion of modules
- Bottom) Expansion of the number and type of modular domains present in regulatory proteins
- Biological systems have evolved through the organization and concatenation of pieces of DNA that encode distinct protein modules
- Is a hierarchical organization inevitable?



Nature 409
(2001) 860



Unique and shared domain organizations in animals



Ancient architectures conserved in all animals

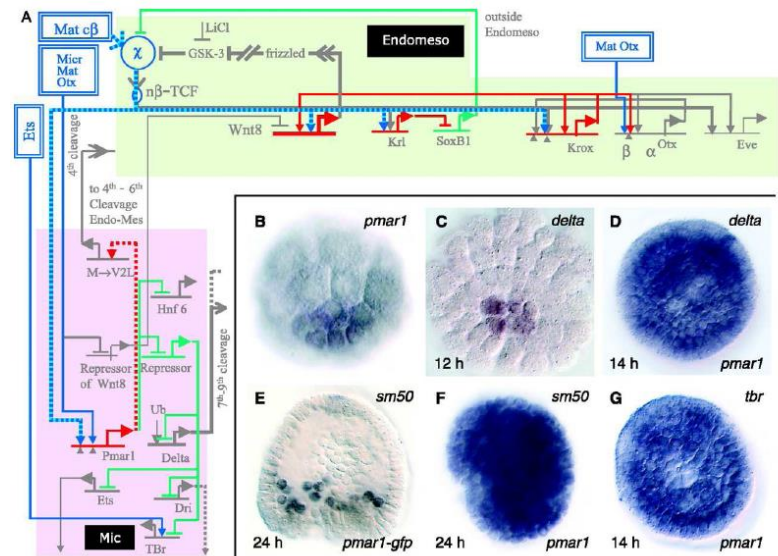
Shared by fly and human

Unique to fly or worm

Unique to human

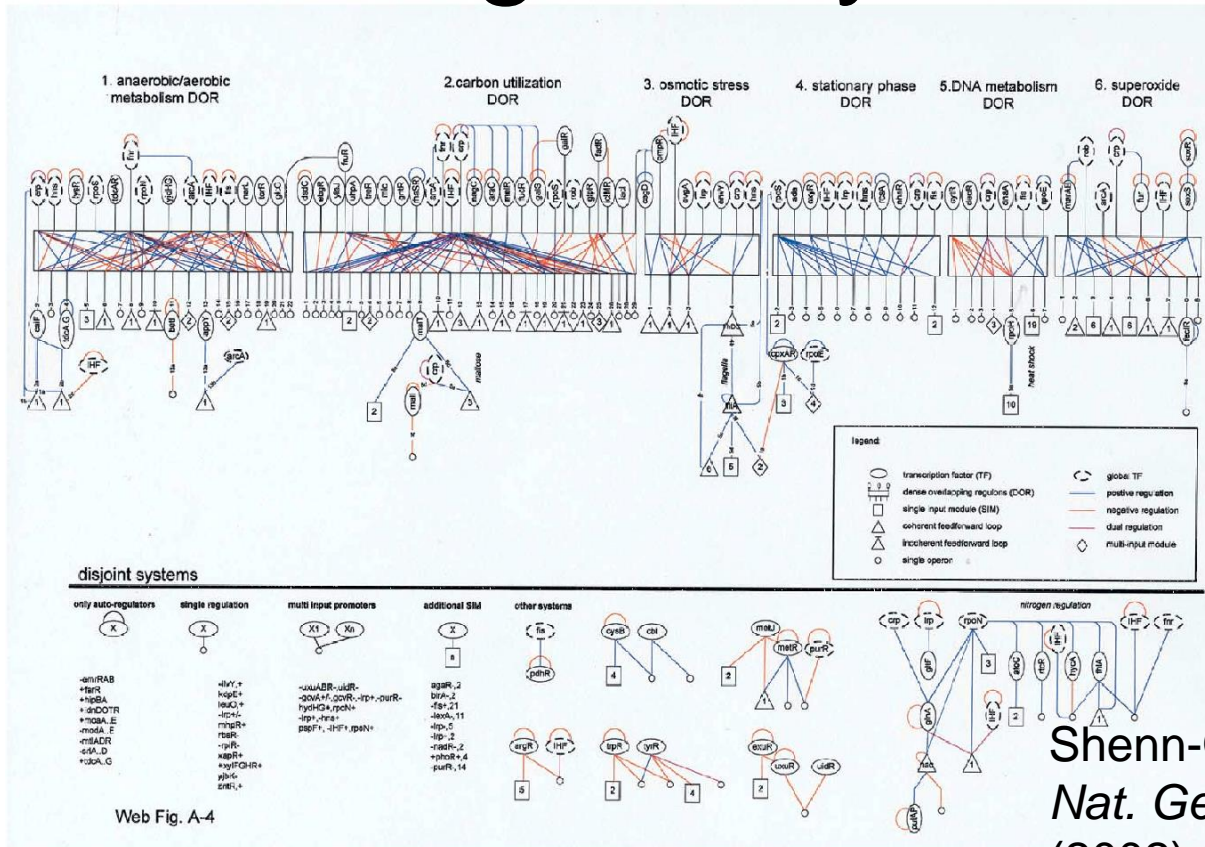
Network Modularity

- Often the topology of the interaction network is of interest to biologists
- Network often robust to changes in the detailed values of equilibrium and kinetic constants associated with the interactions between the molecules of the network
Nature **406** (2000) 188
- The characteristic structures that arise in these interaction networks also seem modular

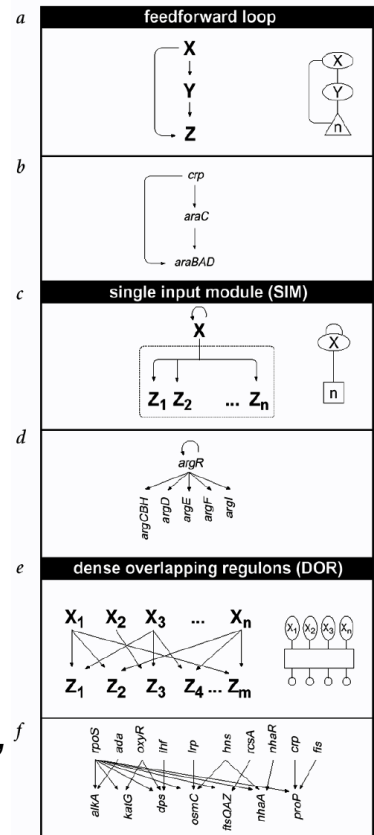


Science **295** (2002) 1669

Regulatory Networks



Shenn-Orr *et al.*,
Nat. Genet. 31
 (2002) 64



- Dictionary of constituent parts, or network motifs, for transcriptional interactions in bacteria
- Shown here is the entire transcriptional regulation network of *E. coli*
- Nodes represent collections of genes, and the lines represent regulation of these genes
- The dictionary of network motifs

Advantage of Hierarchy

- Can evolutionary potential of a set of mutational events be quantified?
- Hierarchy and modular structure fundamental to evolution?
- What general statements can be made?
- Mathematically?

—

$$DX R[X] = hRi_X \frac{1}{4} R[X_0]$$

The Fossil Record

- Evolution does seem to speed up over time
- Fossil records indicate that save for mass extinctions, the speed and complexity of evolution increases over time
 - The first, single-cell life forms evolved 3.5-4 billion years ago, only 0.5-1 billion years after the formation of earth
 - It took another 2.5-3 billion years for multicellular organisms to appear
 - It took a final one billion years for all of the multicellular species to evolve into being

Why Does Life Evolve to Evolve?

- What is the underlying pressure for evolution to speed up over time, say by the emergence of modularity and hierarchy?
- Whatever the selective force for rapid evolution is, it must be consistent with causality
- It seems likely that a changing environment selects for adaptable evolutionary frameworks Earl and Deem, *PNAS* **101** (2004) 11531
- Competition different evolutionary frameworks leads to selection for the most efficient dynamics

Is Modularity Inevitable?

- Is modularity of structure a typical or special case?
- That is, what is the probability that a modular structure will occur in a general evolutionary system?
- By asking whether modularity is inevitable, and thus what is the probability that life will evolve via a hierarchy of mutational events, we may understand the structure that we observe today in biology without the need to resort to the anthropic or intelligent design argument
- By way of analogy to another area of physics, one of the questions contemplated by string theorists is the following: Given an enormous number of possible universes, is the universe that we inhabit reasonably likely?
- An analogy with thermodynamics illustrates the type of answer that we are seeking, albeit in a system that seems simpler than biology
 - In thermodynamics the observed value of energy or density or pressure for a large system is equal to the quantity's value in the most likely state of the system

The Argument for Modularity

- By being modular, a system may be more robust to perturbations and more evolvable
 - H. A. Simon, *Proc. Amer. Phil. Soc.* **106** (1962) 467
 - G. P. Wagner and L. Altenberg, *Evolution* **50** (1996) 967
 - M. E. Csete and J. C. Doyle, *Science* **295** (2002) 1664
 - H. Kitano, *Nature Rev. Gen.* **5** (2004) 826
 - P. Oikonomou and P. Cluzel, *Nature Phys.* **2** (2006) 532
- There is an implicit selective pressure for evolvability in a changing environment
 - D. J. Earl and M. W. Deem, *PNAS* **101** (2004) 11531
- Modularity increases evolvability in the presence of large genetic moves (HGT, recombination, super/co-infection)
 - J. A. Shapiro, *BioEssays* **27** (2005) 122; *Gene* **345** (2005) 91
 - N. Goldenfeld and C. Woese, *Nature* **445** (2007) 369
 - L. D. Bogarad and M. W. Deem, *PNAS* **96** (1999) 2591
- Thus, a changing environment should implicitly select for modularity
 - H. Lipson *et al.*, *Evolution* **56** (2002) 1549; A. Gardener and W. Zuidema, *Evolution* **57** (2003) 1448
 - E. A. Variano, J. H. McCoy and H. Lipson, *PRL* **92** (2004) 188701 (stability)
 - M. W. Deem, *Physics Today*, January 2007, 42-47

Genetic Moves Conjugate to Modular Protein Structure Enhance Evolvability

- Regulation (yeast 6000, human 21 000 genes)
- Timing of protein expression is basis for dog breed diversity (synteny)
- Alternative splicing
- VDJ recombination in immune system
- Exon shuffling (Walter Gilbert)
- Transposons and retrotransposons
- Horizontal transfer

N. Goldenfeld and C. Woese, *Nature* **445** (2007) 369

J. A. Shapiro, *BioEssays* **27** (2005) 122; *Gene* **345** (2005) 91

Modular Genetic Moves Efficient

- Experimentally

- W. P. C. Stemmer, *Nature* **370** (1994) 389: fucosidase -> galactosidase
- J. C. Moore, H-M. Lin, O. Kuchner, and F. H. Arnold, *J. Mol. Biol.* **272** (1997) 336: cP450 functionality

- Theoretically

- L. D. Bogarad and M. W. Deem, *PNAS* **96** (1999) 2591

Table 1. Results of Monte Carlo simulation of the evolution protocols

Evolution method	Starting energy	Evolved energy	Achieved binding constant
Amino acid substitution	-17.00	-23.18	1
DNA shuffling	-17.00	-23.83	100
Swapping	0	-24.52	1.47×10^4
Mixing	0	-24.88	1.81×10^5
Multipool swapping	0	-25.40*	8.80×10^6 *

Simplified Model of Evolution

- Population of individuals
- Each replicates at rate r_i
- Changes also by mutation at rate μ_{ij}
- Hierarchy of structure implies r_i partially linearly decomposable according to domains
- Hierarchy of mutation implies μ_{ij} connects i and j related by a change of a domain
- Evolution is efficient if mutational events are complementary to domain structure

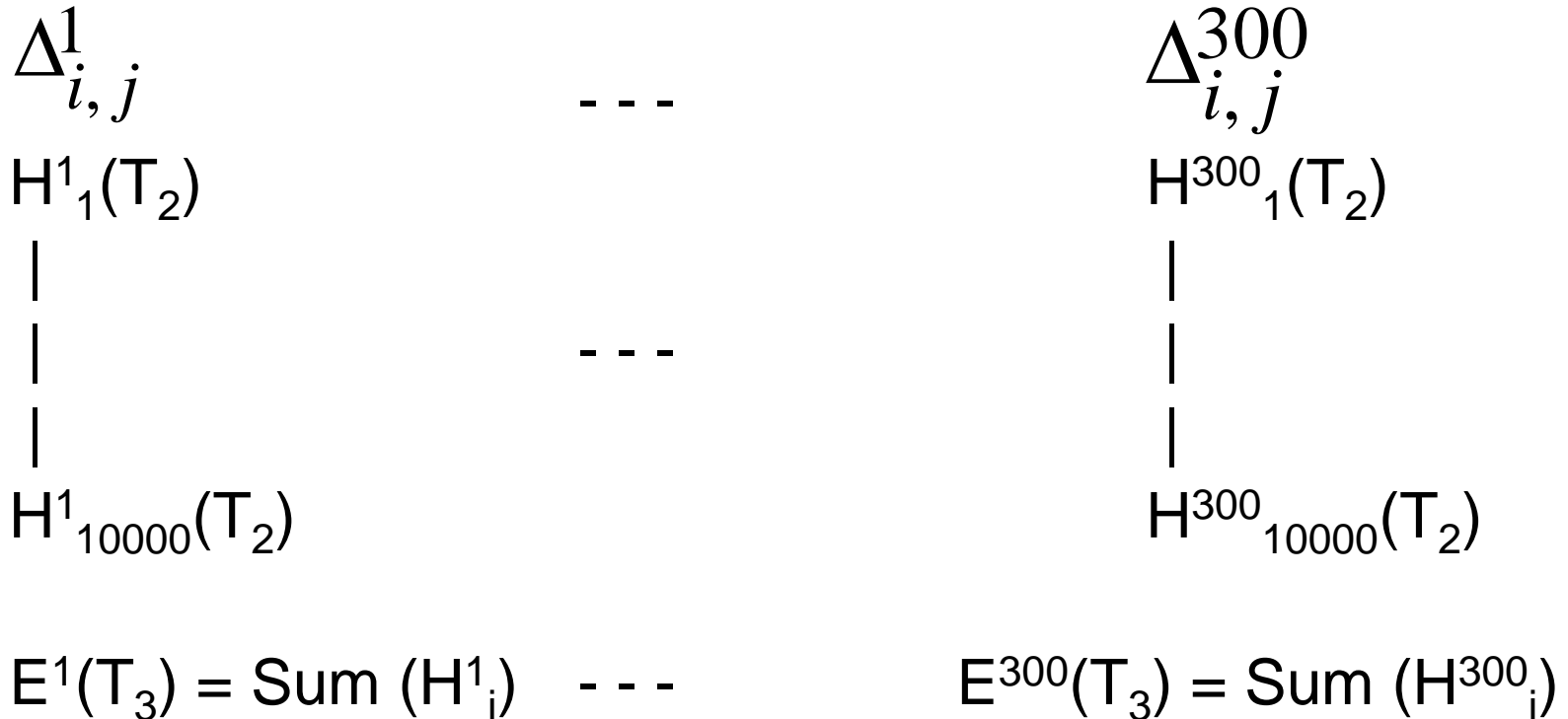
Spontaneous Emergence of Modularity in a Population of Evolving Individuals

Spin glass form of replication rate (fitness)

$$H = \frac{1}{2\sqrt{N_D}} \sum_{i \neq j} \sigma_{i,j}(s_i, s_j) \bullet \Delta_{i,j}$$

- $1 \leq i \leq N$, $N=120$ is the size of genome
- s_i is the sequence (amino acid, allele, etc) at position i
- $\sigma_{i,j}(s_i, s_j)$ is the interaction matrix, representing the environment
- $\Delta_{i,j}$ is the connection (adjacency) matrix of 0 or 1, reflecting the structure of the interactions; $\Delta_{i,i} = \mathbf{0}$ $\Delta_{i,i\pm 1} = \mathbf{1}$
- $N_D = \sum_{i>j+1} \Delta_{i,j} = 346$ is the fixed, total number of connections of each of the $D=300$ structures

Dynamics of Evolution



Top 5% Δ 's selected based on their E, and reproduced
 Δ 's undergo mutation

Dynamics of Evolution

- $D=300$ structures, $\Delta_{i,j}^\alpha$, each with 1000 associated sequences
- Environment, represented by $\sigma_{i,j}(s_i, s_j)$, changes with magnitude p and period T_2
- Three different time scales: $T_1=1$ (rapid sequence evolution), T_2 (moderate environmental change), and $T_3=10000 T_2$ (slow evolution of the structure of the connections)
- Dynamics
 - Sequence: point mutation and horizontal gene transfer
 - Environment: random change
 - Structure: point mutation construction/destruction of connections

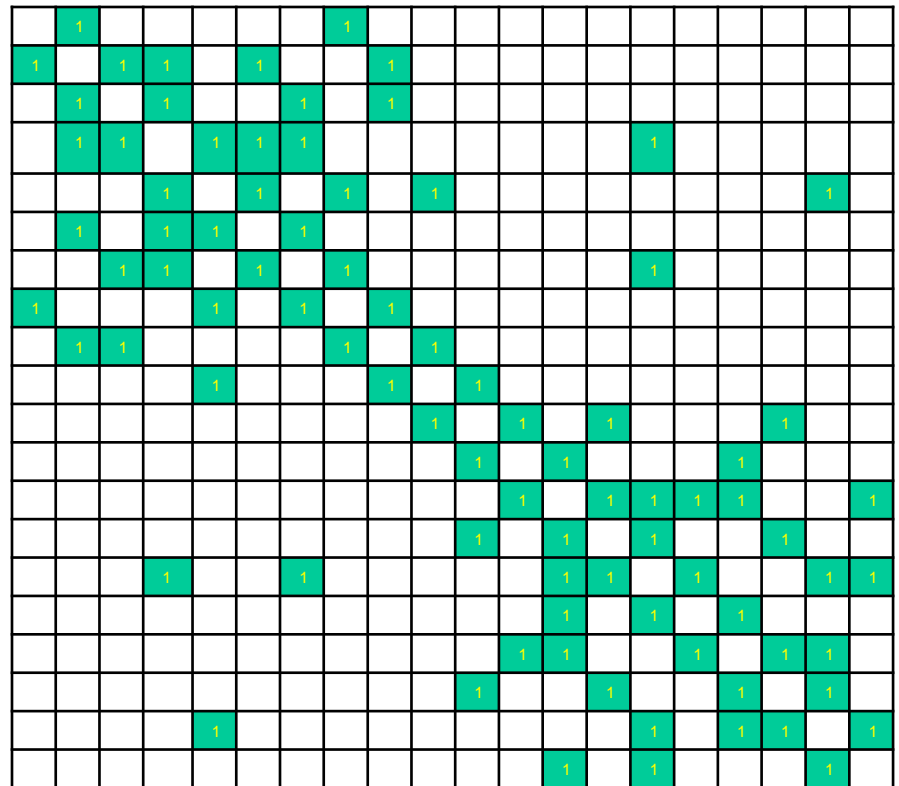
Definition of Modularity

Definition of Modularity

$$M^\alpha = \sum_{i>j+1,k} \Delta_{i+10k,j+10k}^\alpha, \quad M = \frac{1}{D} \sum_{\alpha} M^\alpha$$

$1 \leq i, j \leq 10$

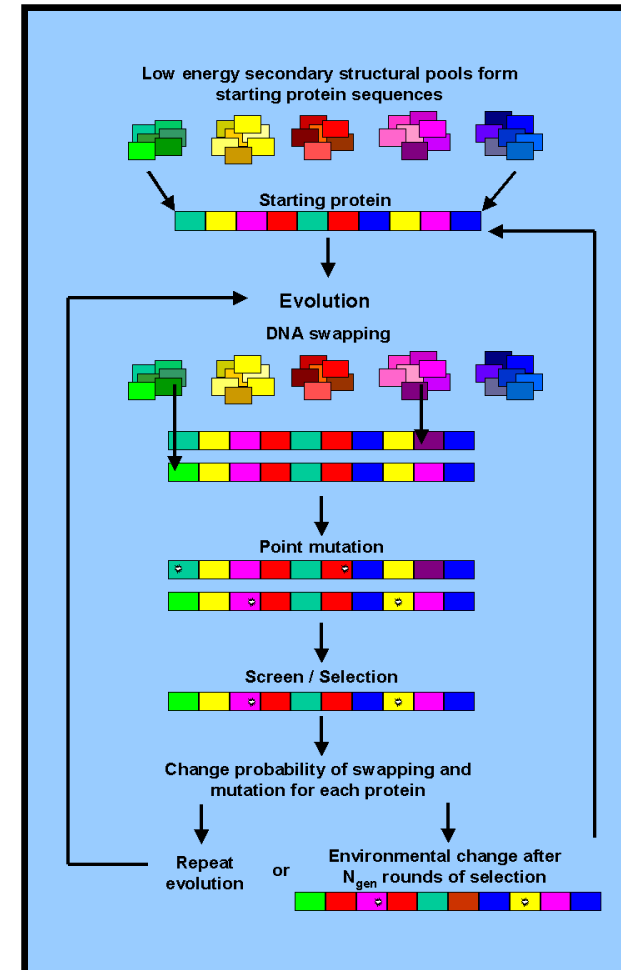
k denotes the k^{th} diagonal block in the $\Delta_{i,j}^\alpha$ connection matrix



Environmental Change and Selection

- Population of 1000 proteins
- After evolution we select the 50% most viable proteins and repeat
- System evolves for T_2 rounds of selection and we then impose an environmental change (frequency = $1/T_2$)
- Magnitude of environmental change characterized by parameter p (probability to change random matrix elements)
- Study for different frequency and magnitude of environmental change

Schematic



Spontaneous Emergence of Modularity in a Model of Evolving Individuals

Modularity: Details of the Computation

A small set of assumptions leads to modularity

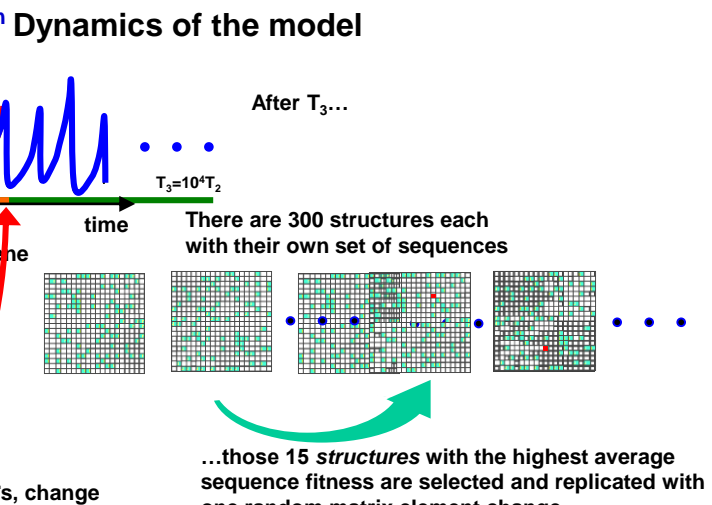
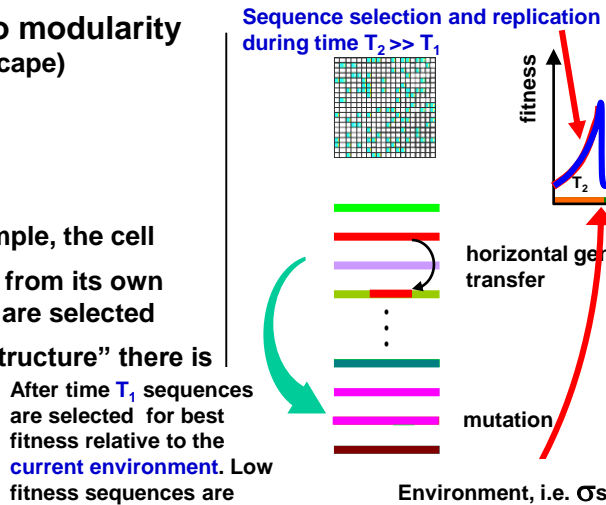
- Slow evolution (e.g. rugged fitness landscape)
- Changes in environment
- Horizontal gene transfer

Protein evolution

- The environment for a protein is, for example, the cell
- As the cell experiences varying demands from its own environment changes in protein function are selected
- Since “protein function” = “sequence + structure” there is selection pressure on both

The model

- The fitness $H^\alpha(s^{\alpha,k})$ of a given protein is a function of its sequence and structure
- A sequence is a string of amino acids
- The structure is encoded in a “connection matrix”
- The couplings $\sigma_{i,j}$ represent the effects of the environment
 - For a given environment specific protein functions are most fit
 - To optimize protein fitness natural selection drives protein sequence and structure to maximize the contributions of the $\sigma_{i,j}$

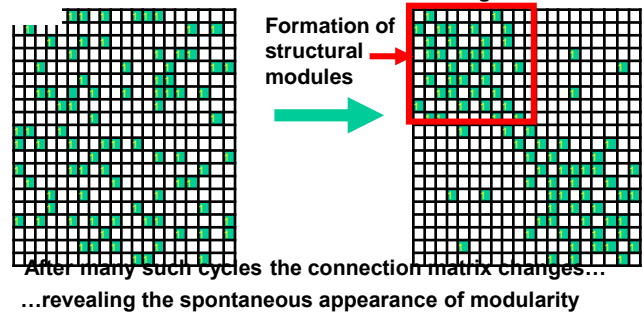
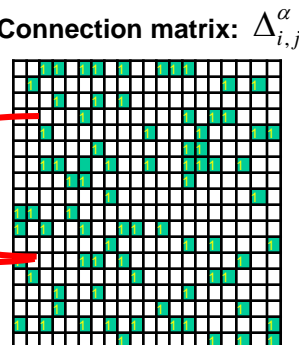


$$H^\alpha(s^{\alpha,k}) = -\frac{1}{2\sqrt{N_D}} \sum_{i \neq j} \sigma_{i,j} (s_i^{\alpha,k}, s_j^{\alpha,k}) \cdot \Delta_{i,j}^\alpha$$

- Amino acid sequences**
- gly-leu-ala-thr- ...
 - ala-his-cys-asp- ...
 - gly-his-thr-asp- ...
 - leu-cys-cys-leu- ...

Matrix entry of “1”: points in sequence are connected

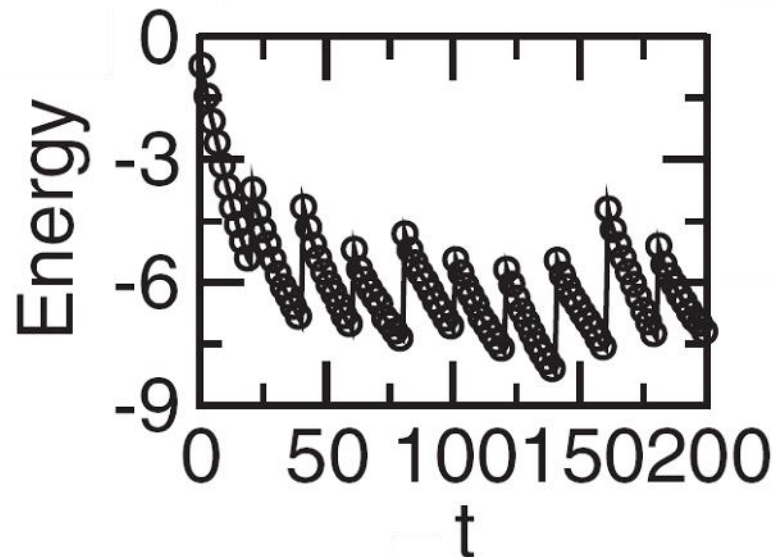
Matrix entry of “0”: points in sequence are not connected



Sun and Deem
 Phys. Rev. Lett. 2007;
 99:228107

Energy Dynamics for Given Δ^α

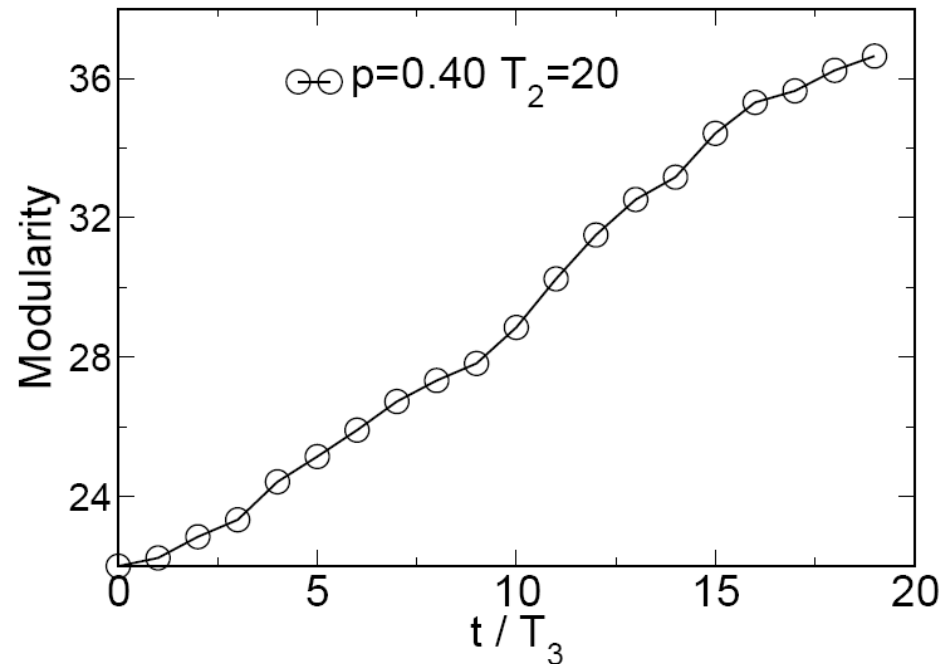
- Energy rises with environmental change and evolves within one environment
- e.g. $p = 0.4$, $T_2 = 20$



Spontaneous Symmetry Breaking and Emergence of Modularity

- Generate Δ^α randomly
 $M_0=22$, no net modularity
- Spontaneous emergence of modularity
- $\delta M = M - M_0 \neq 0$

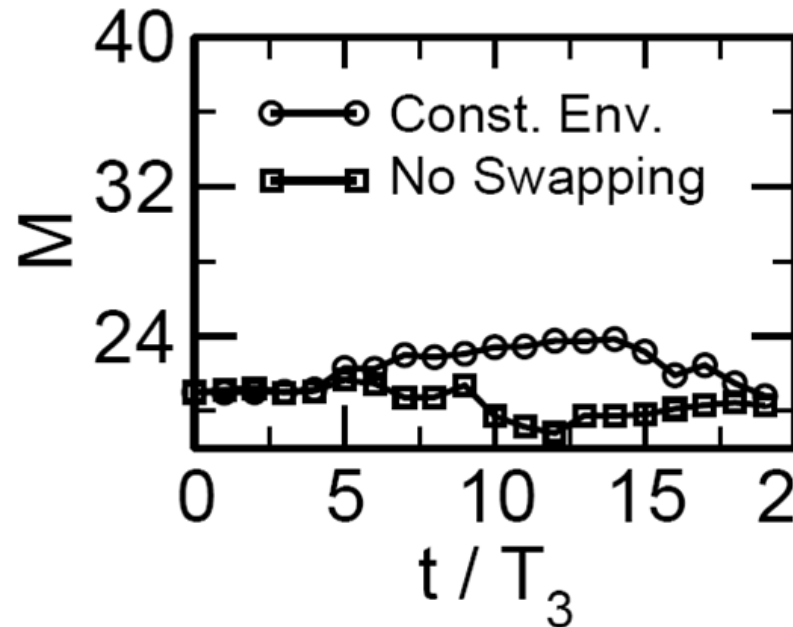
A symmetry breaking event (permutation symmetry)



J. Sun and M. W. Deem,
PRL **99** (2007) 228107

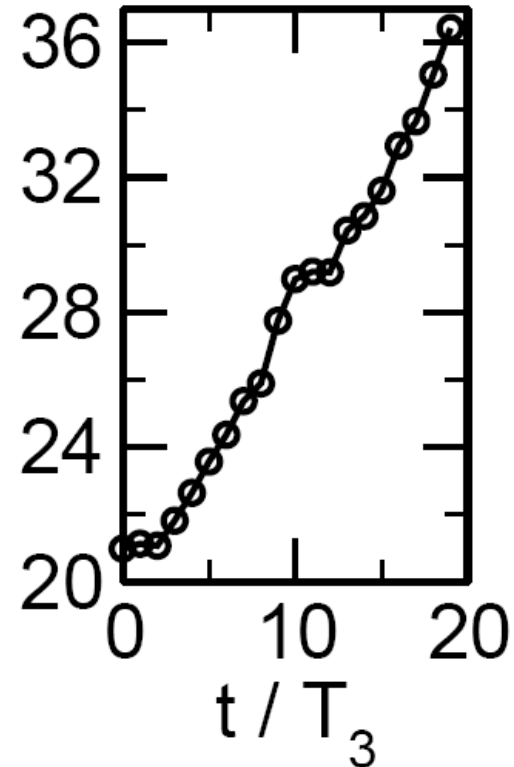
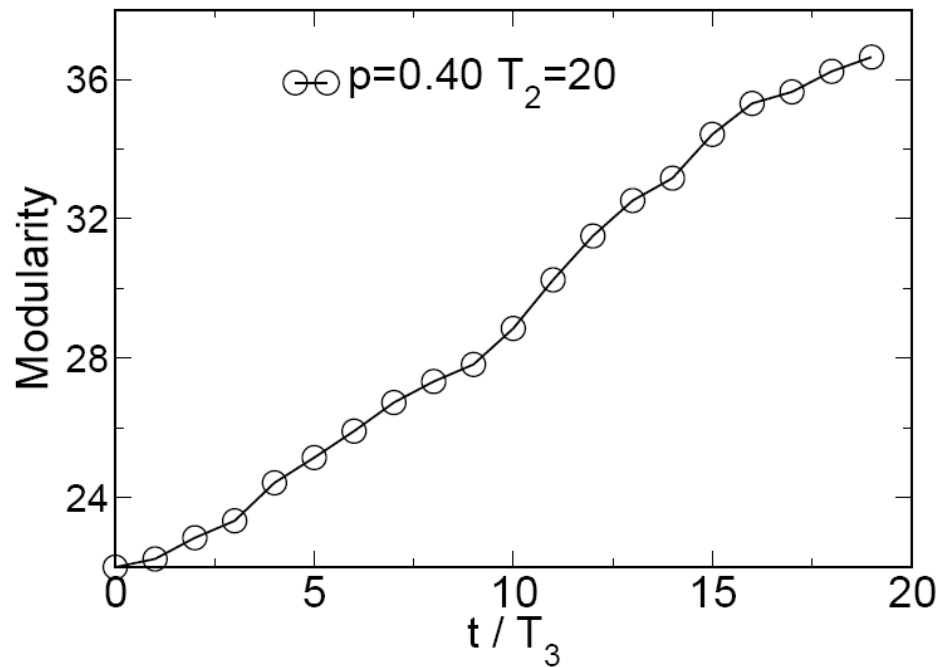
Control Experiment

- If $p=0$, modularity does not increase
- If no HGT, modularity does not increase



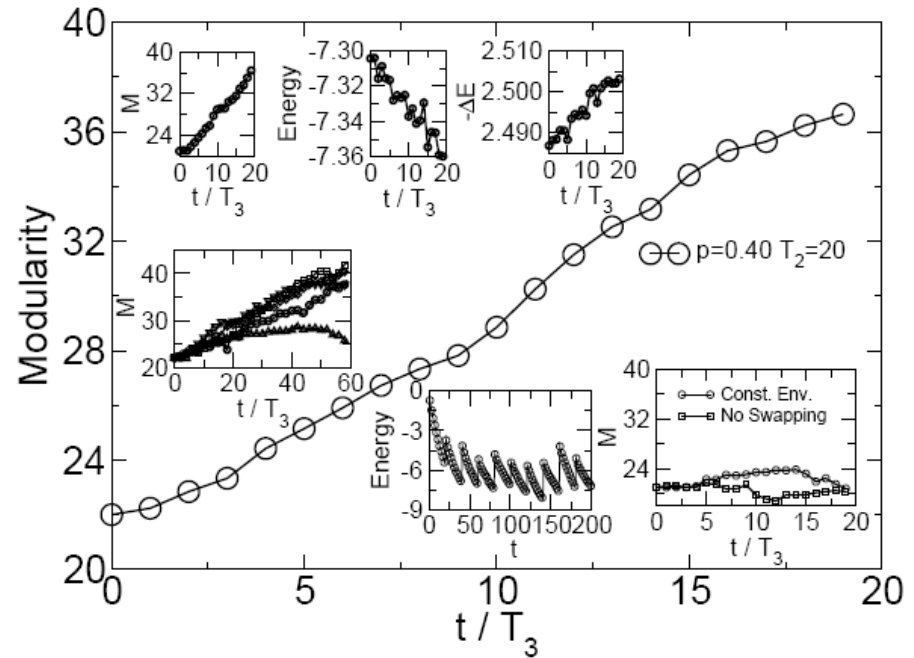
Scale-Free Initial Network

- Use a random, correlated adjacency matrix
- Barabassi method: $\gamma=3$
- Result identical to random network



Cumulative Fitness Increases

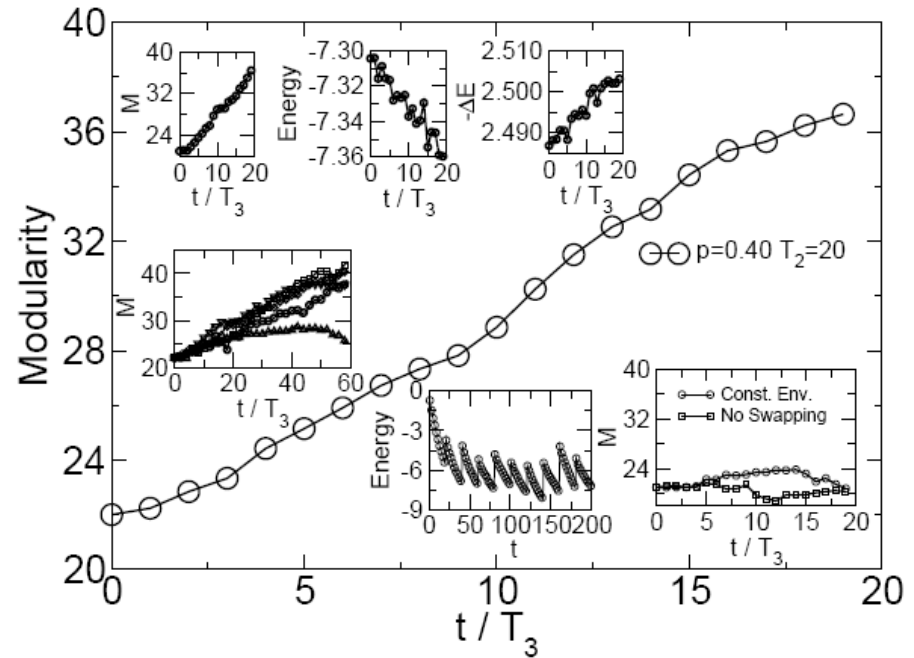
- Selection is for replication rate (fitness)
- Fitness increases over time due to emergence of modularity



J. Sun and M. W. Deem,
PRL **99** (2007) 228107

Increase of Evolvability

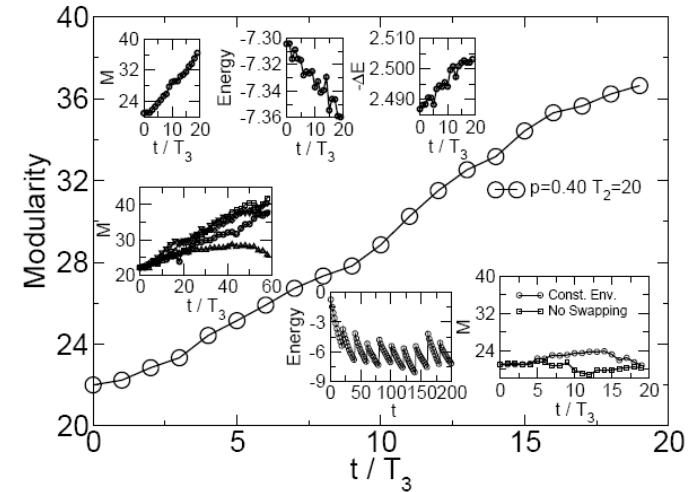
- Selection is for replication rate (fitness)
- Implicit selection for evolvability
- Evolvability characterized by response function:
 $-\Delta E / (10^4 T_2)$
- Evolvability increases



J. Sun and M. W. Deem,
PRL **99** (2007) 228107

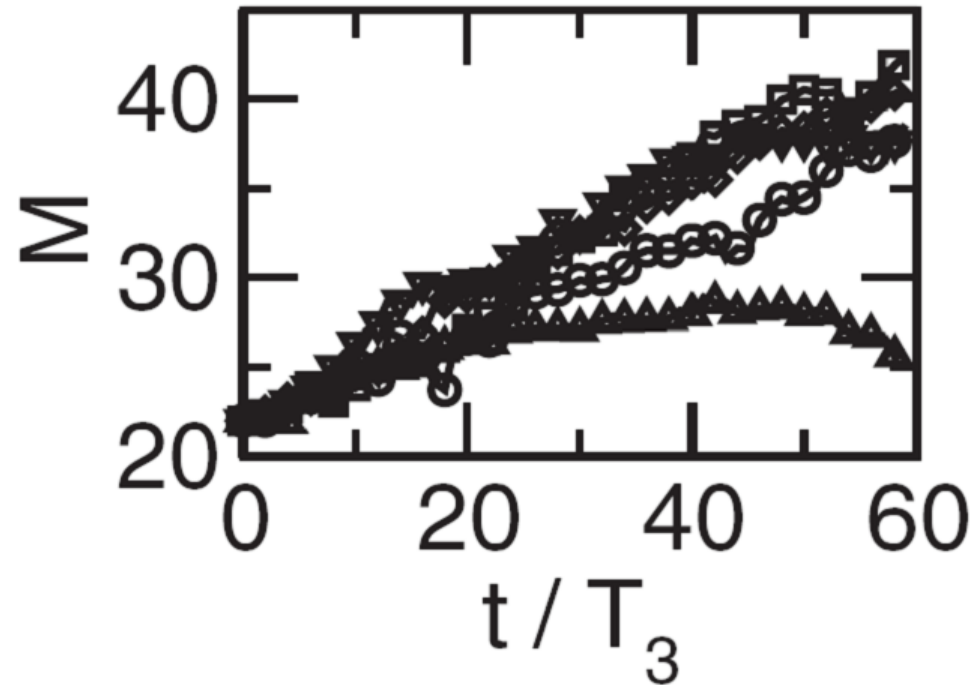
Fitness Change within One Environment

- Within one environment, sequence mutation and selection lead to increased fitness
- Environmental change then destroys some of these gains
- Gains are enhanced due to emergence of modularity over (long) time



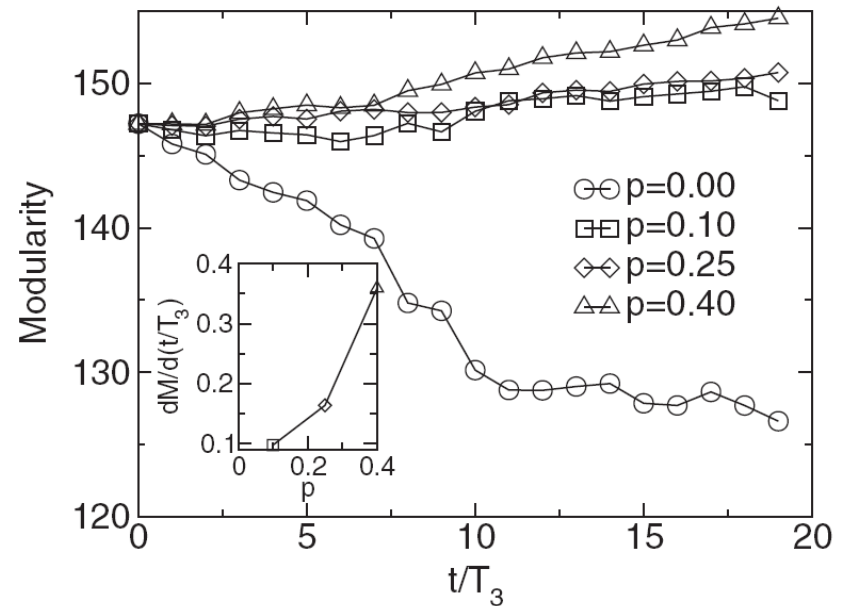
Incommensurate, Random Swapping Leads to Modularity

- If HGT occurs at a random position, with a random length, modularity also arises
- Average lengths of 10,20,40,5
- Modularity measured as before
- Fixed position and length HGT is biologically motivated: non-coding DNA >> exons



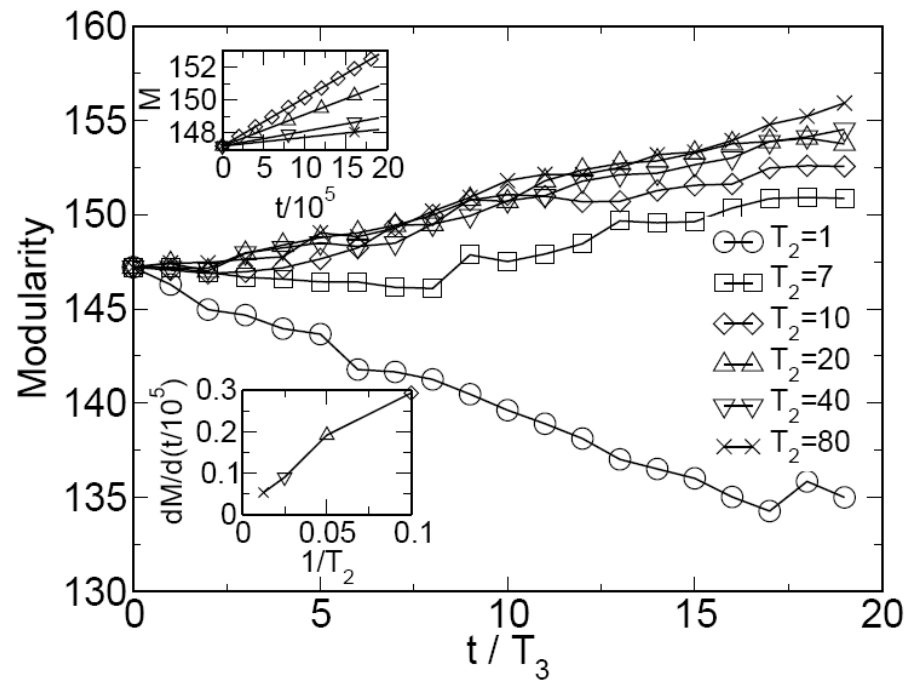
Modularity Is a Function of Magnitude of Environmental Change

- Modularity increases in a changing environment
- Insufficient environmental change leads to decay of modularity
- Velocity of modularity growth depends on magnitude



Modularity is a Function of Frequency of Environmental Change

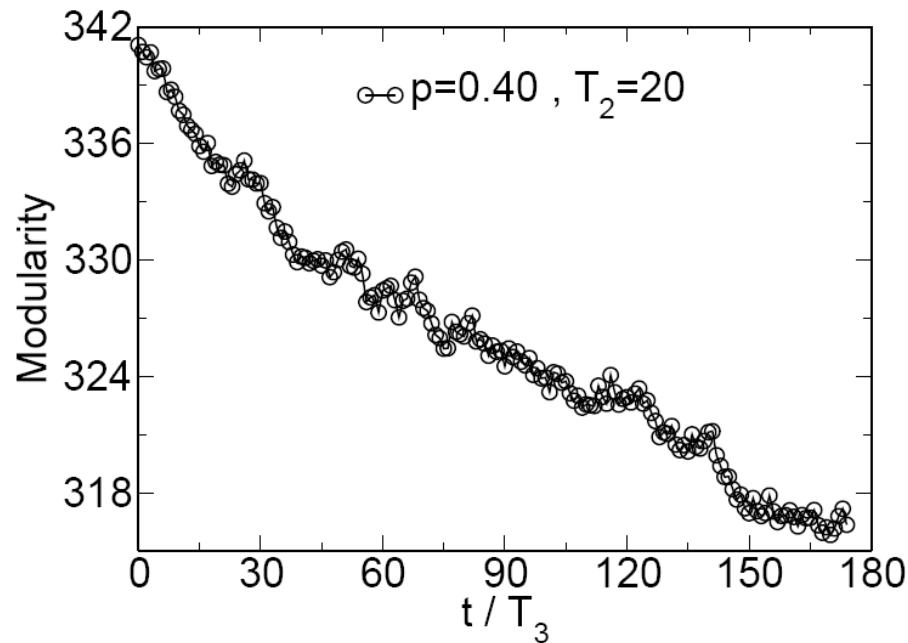
- The velocity of modularity growth depends on the frequency ($f=1/T_2$) of the environment change
- At high frequency ($1/T_2 > 1/5$), modularity decays
- Modularity growth linear in frequency at low frequency ($1/T_2 < 1/10$)



J. Sun and M. W. Deem,
PRL **99** (2007) 228107

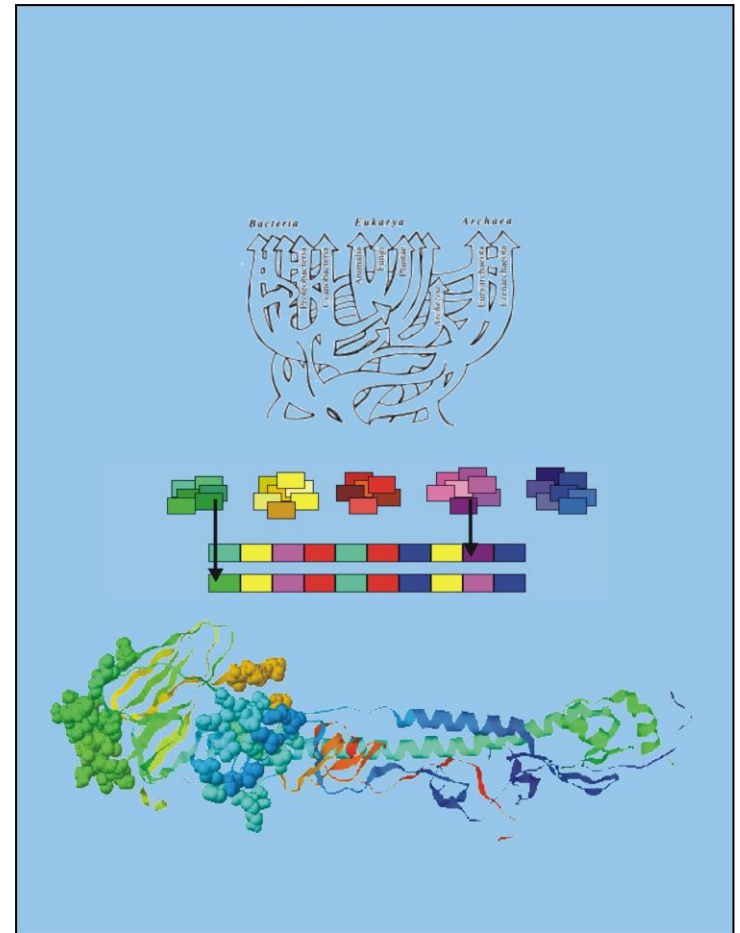
Steady-State Modularity

- Modularity evolves to a finite value:
 $22 \leq M \leq 346$
- There is a competition between the implicit selection for modularity due to the changing environment and the destruction of modularity due to random mutation



A Model Evolving System

- Model described at the individual level
- Or, it is a coarse-grained model of an arbitrary evolving system: amino acids, secondary structures, domains, proteins, multi-protein complexes, pathways, organelles, cells, organs, individuals, species, and so on

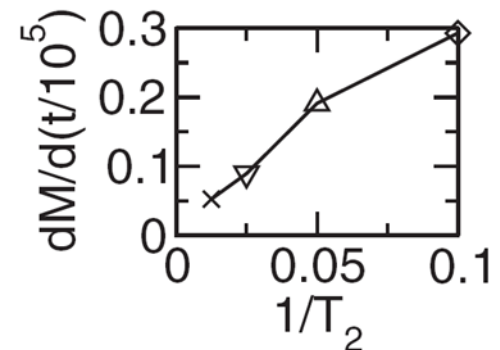


Conjecture

- Axioms
 - Fitness landscape has many local optima
(Evolution occurs slowly)
 - Environment is changing
 - Horizontal gene transfer exists
- Conjecture
 - Under these conditions, modularity will spontaneously develop

$$p_E \approx p_0 = \frac{M'}{R}$$

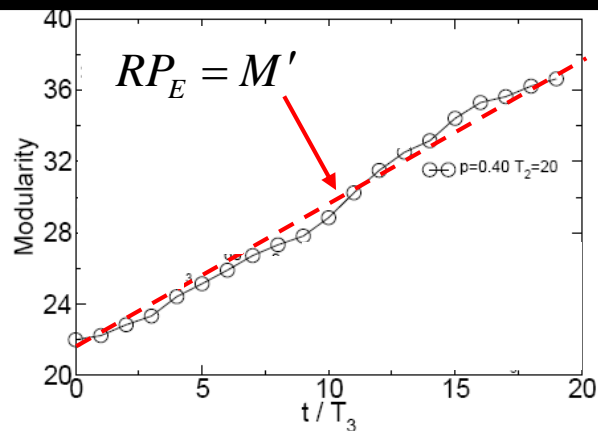
- e.g. $p_E = p/T_2$
 $R =$ density of local optima
 $M' = dM / dt$



Spontaneous Emergence of Modularity in a Model of Evolving Individuals

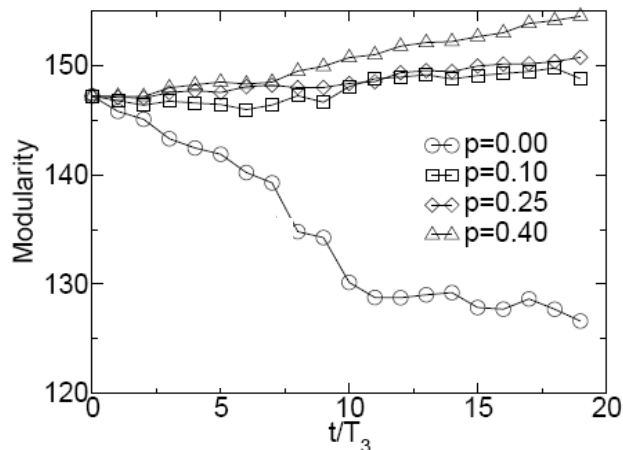
Modularity

$P_E = I - E - P_0$	E is the operator that per unit time produces the new environment
$M_I \rightarrow P_0$	M_I is the initial condition
$E = E(p, T_2)$	R is the density of fitness optima in sequence space
R	$M^\alpha = \sum_{k=0}^{11} \sum_{i=1, j=i+2}^{10} \Delta_{10k+i, 10k+j}^\alpha$ the projection of near-diagonal elements in an adjacency matrix
M^α	M' is the time derivative of the average of M^α over the population of structures
M'	



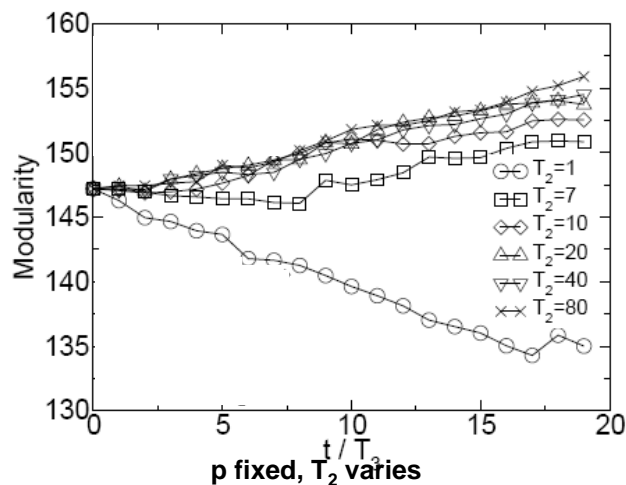
Spontaneous emergence of growing modularity, $M > M_0$, the random distribution of structural connections, as the system evolves. The slope is

$$M' = RP_E.$$



Notice how the rate at which modularity grows is positively correlated with increasing p

T_2 fixed, p varies



Notice how the rate at which modularity grows is positively correlated with increasing T_2

p fixed, T_2 varies

T_2 and T_3 are the sequence and structure selection times

Environmental variability and modularity of bacterial metabolic networks

Merav Parter, Nadav Kashtan and Uri Alon*

Address: Molecular Cell Biology Department, Weizmann Institute of Science, Rehovot 76100, Israel

Email: Merav Parter - merav.parter@weizmann.ac.il; Nadav Kashtan - nadav.kashtan@weizmann.ac.il; Uri Alon* - uri.alon@weizmann.ac.il

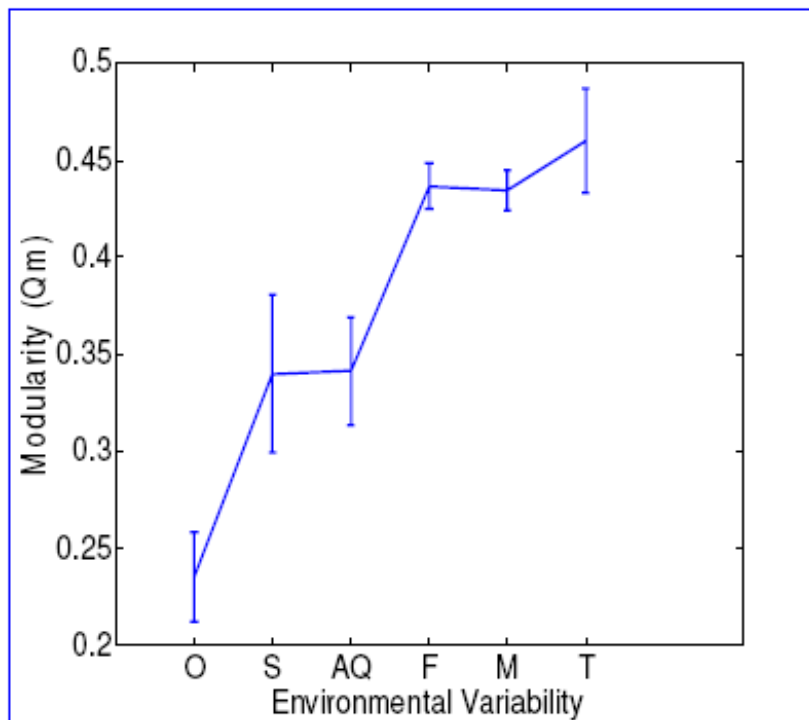
* Corresponding author

Published: 23 September 2007

BMC Evolutionary Biology 2007, **7**:169 doi:10.1186/1471-2148-7-169

Received: 17 May 2007

Accepted: 23 September 2007



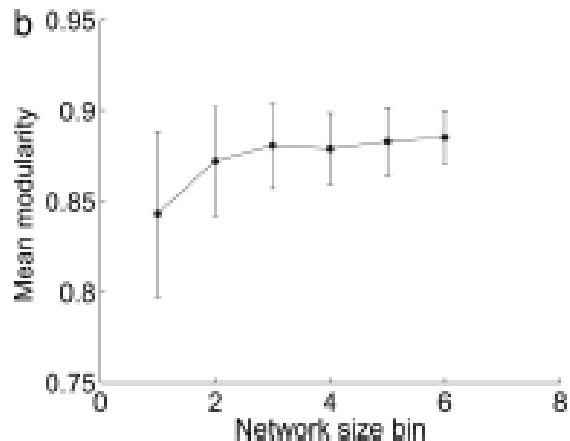
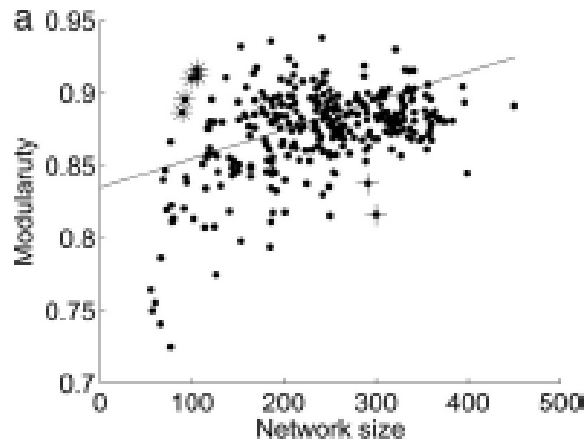
Relation between environmental variability and modularity. Normalized modularity measure (Q_m) of bacterial metabolic networks versus the environmental class of the organism. Environments are ordered according to their variability ranging from O (obligate), the least variable to T (terrestrial), the most variable. Mean and standard error of Q_m are presented for each environmental class.

The evolution of modularity in bacterial metabolic networks

Anat Kreimer^a, Elhanan Borenstein^{1*}, Uri Gophna⁵, and Eytan Ruppin¹¹

^aSchool of Mathematical Science, ¹Department of Molecular Microbiology and Biotechnology, Faculty of Life Sciences, and ¹¹School of Computer Science and School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; ⁵Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020; and ¹Santa Fe Institute, Santa Fe, NM 87501

6976–6981 | PNAS | May 13, 2008 | vol. 105 | no. 19



Three main determinants of modularity:

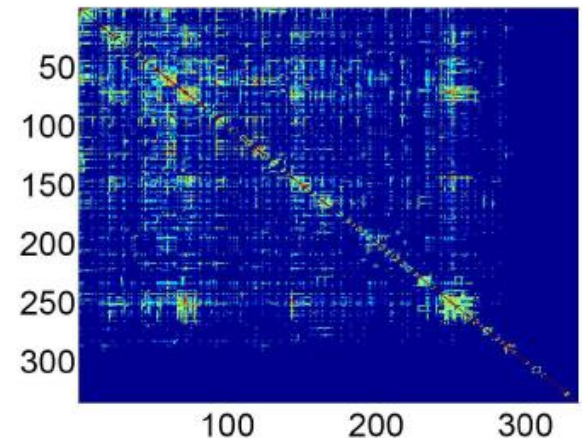
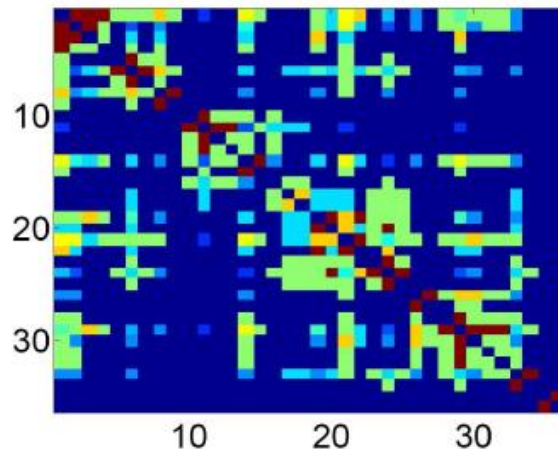
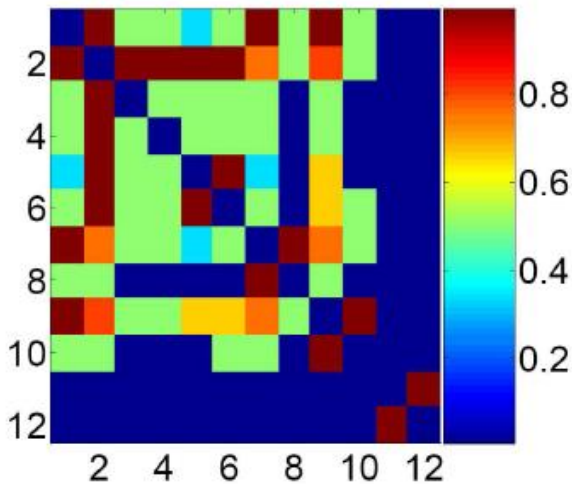
1. network size is an important topological determinant of network modularity.
2. several environmental factors influence network modularity; even among the pathogens, those that alternate between two distinct niches, such as insect and mammal, tend to have relatively high metabolic network modularity.
3. HGT is an important force that contributes significantly to metabolic modularity.

Modularity Growth in Protein-Protein Interaction Network

- Network of protein-protein interactions in *E. coli* and *S. cerevisiae*

$$TOM_{ij} = \frac{k_i a_{ik} a_{kj} + a_{ij}}{\min(k_i; k_j) + 1}$$

- Network reordered by hierarchical clustering algorithm

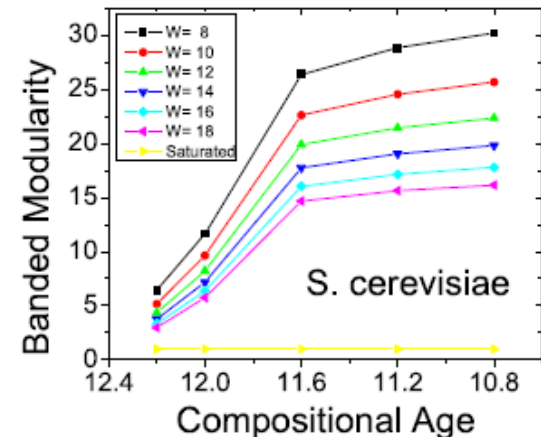
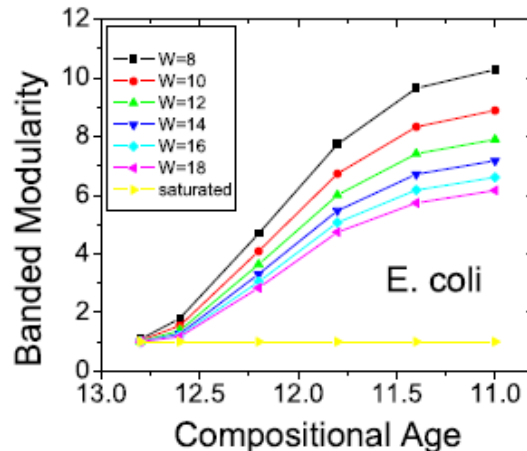
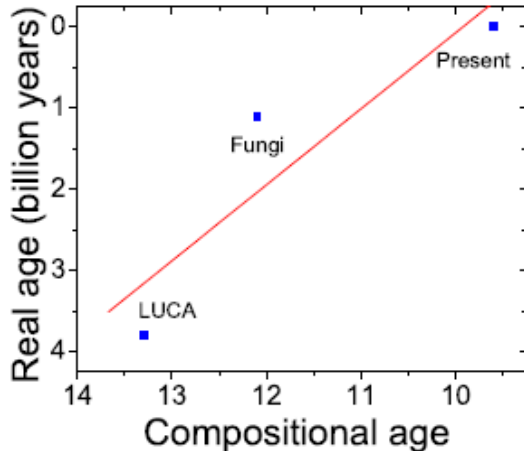


Growth of Modularity

- Measure interactions along diagonal

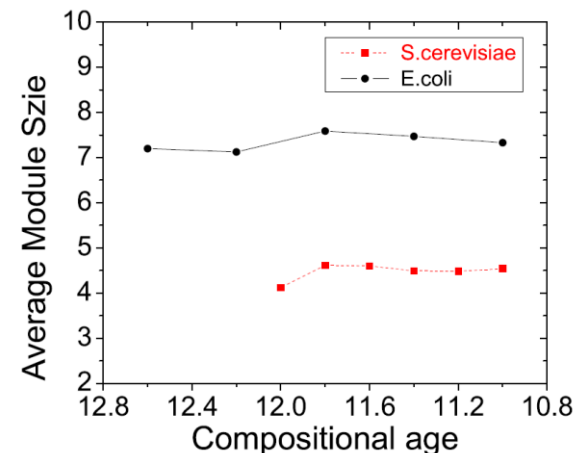
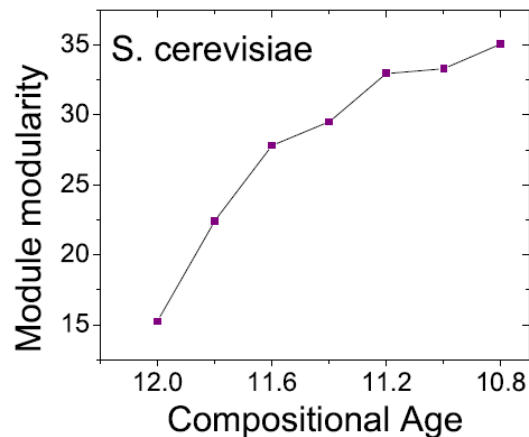
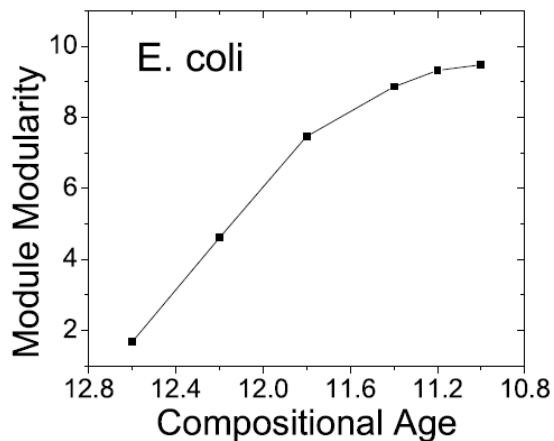
$$M = \frac{\sum_{0 \leq j < i < j < W} a_{ij}}{\sum_{i \in J} a_{ij}} \propto \frac{\sum_{0 < j < i < j < W} a_{ij}}{\sum_{i \in J} a_{ij}}$$

- Modularity grows with time!



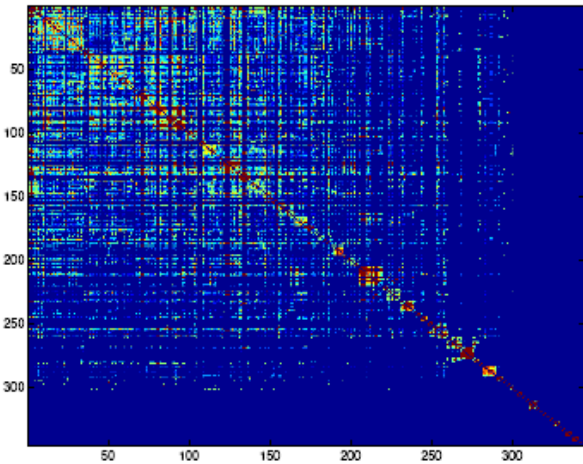
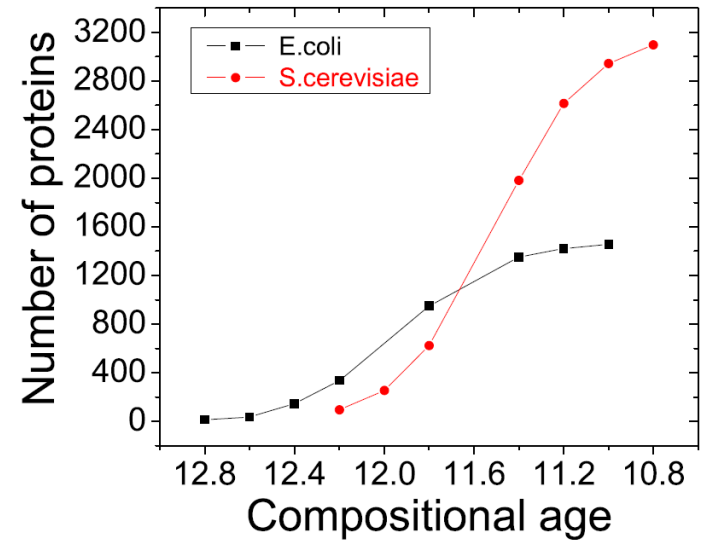
Modularity vs. Banded Localization

- Define modules along diagonal by when the interaction decays to 0.2 of maximal value (average module size ~ constant)
- Measure a_{ij} in these modules
- Modularity so-measured grows

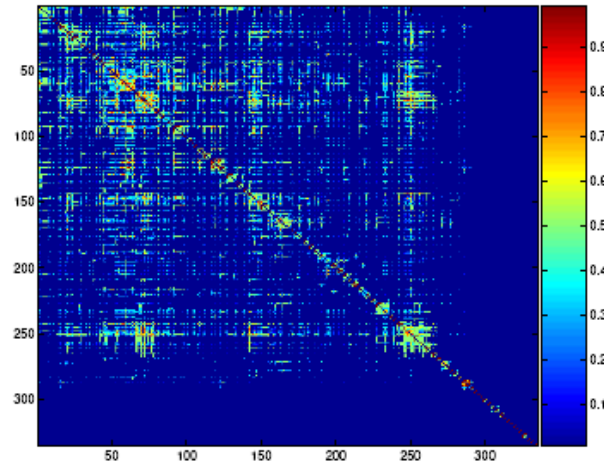


Network Properties

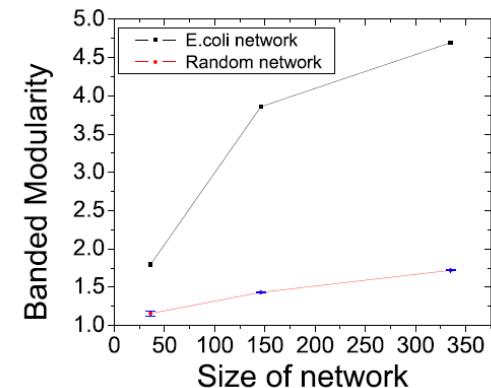
- There are more proteins at younger ages
- Networks formed by randomly chosen proteins do not display modularity growth



Random network

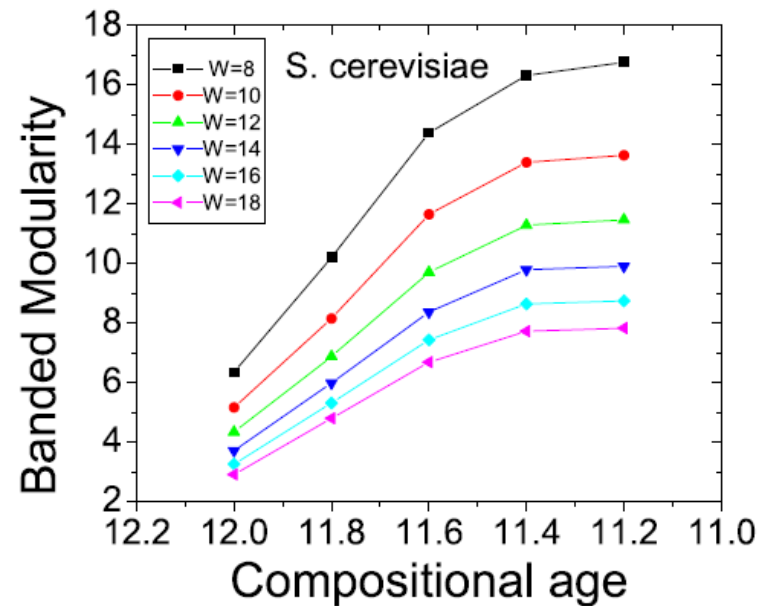
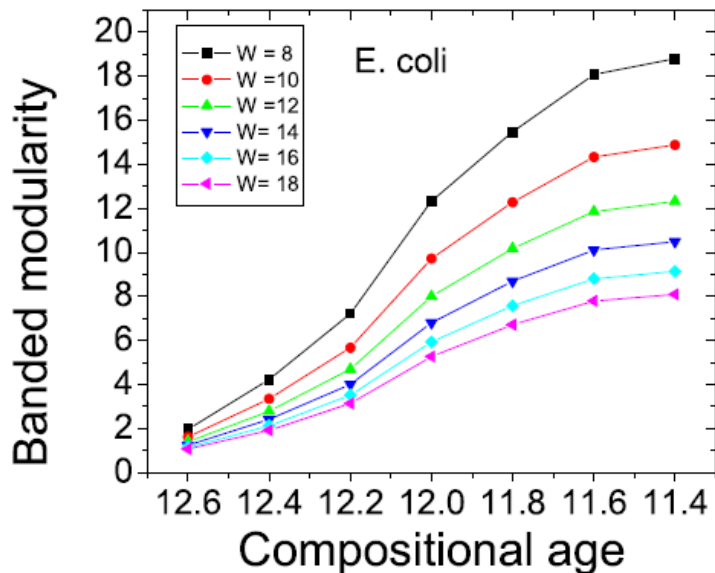


E. coli, 12.2 ca



Modularity in Domain-Domain Interaction Network

- Consider the domain-domain interaction network, rather than the protein-protein interaction network
- Modularity grows



Newman's Modularity

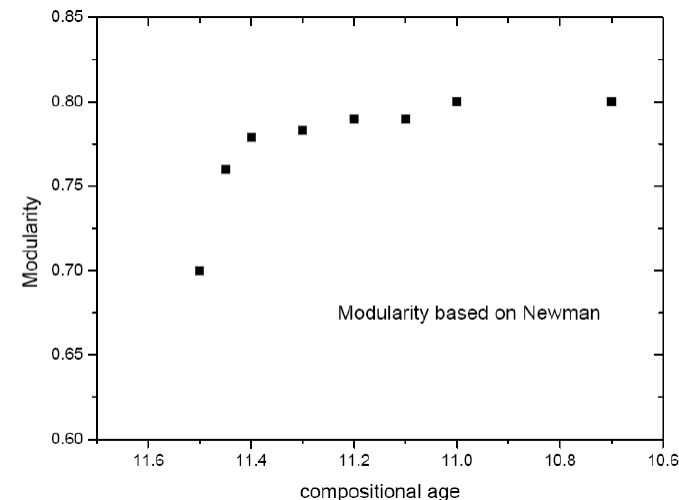
- Define

$$M = \max_P M(P) = \max_P \sum_{i=1}^m \frac{l_i}{L} \left(\frac{d_i}{2L} - \left(\frac{d_i}{2L} \right)^2 \right)$$

R. Girvan, M. E. J. Newman, *Phys. Rev. E* **69** (2004) 026113

M. Salestardo, R. Guimera, A. Moreira, and L. Nunesamaral, *PNAS* **104** (2007) 15224

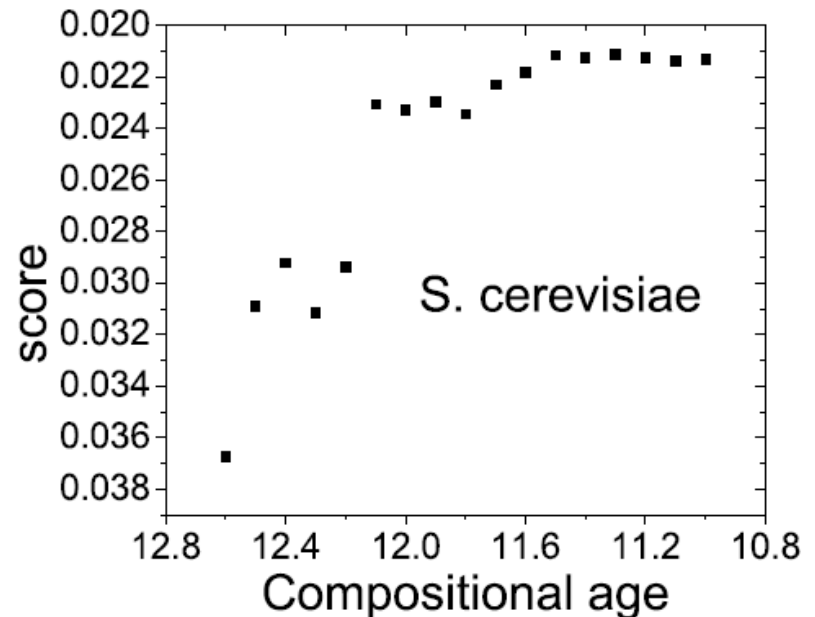
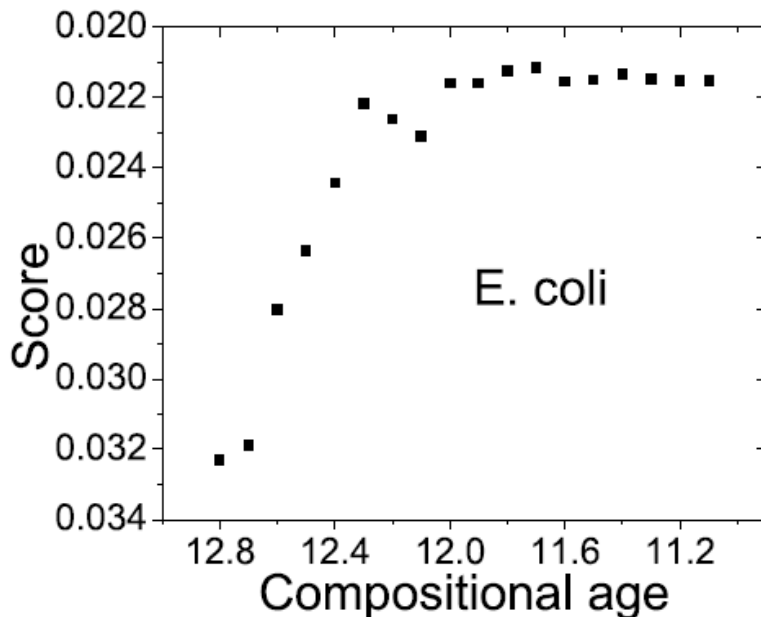
- Where L is the number of links in the network, l_i is the number of links in module i , d_i is the sum of degrees of the nodes in module i , and m is the number of modules in partition P
- Han's data: 1376 proteins



Domain Modularity: Another Definition

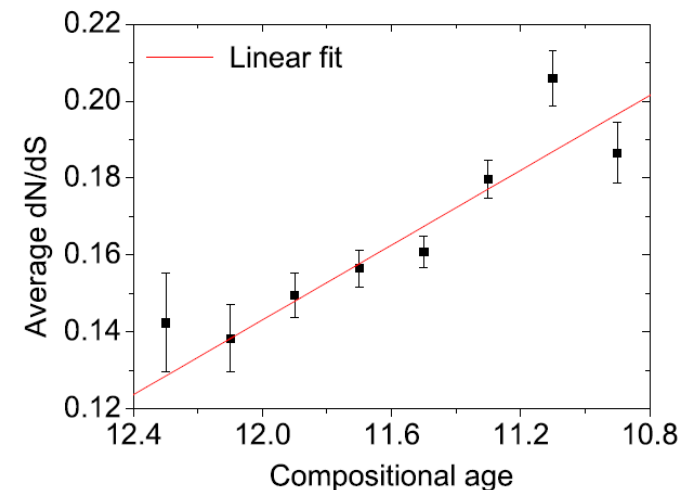
- Measure fraction of domains in protein A with which other proteins interact

$$M = \frac{1}{2N} \sum_{i=1}^N \left(\frac{I_i^A}{D_i^A L_B^{2=3}} + \frac{I_i^B}{D_i^B L_A^{2=3}} \right)$$



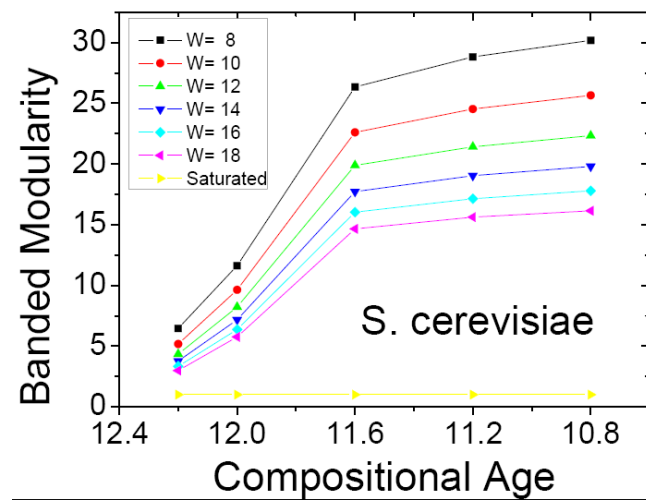
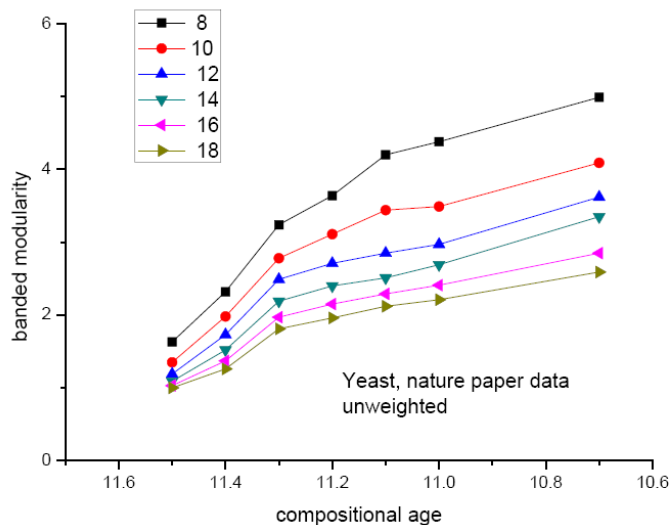
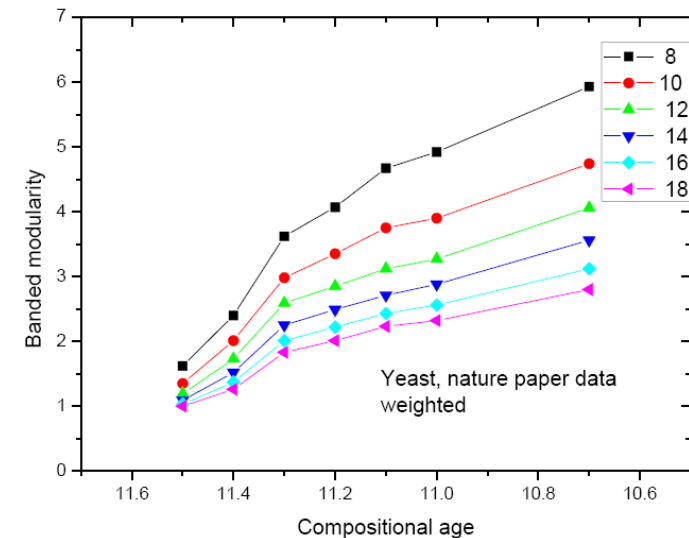
Rate of Evolution

- dN/dS is conventional measure of rate of evolution due to point mutation
- Yeast genes
- New genes are evolving more rapidly than old genes
- $R^2 = 0.81$

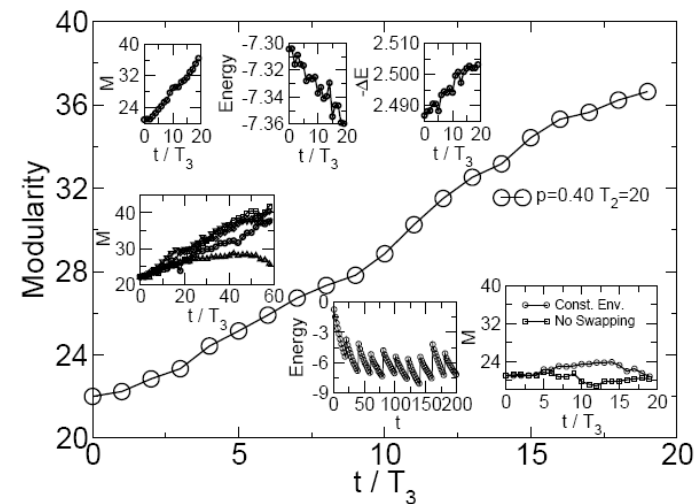


Higher Order Modularity: Hierarchy

- Identify modules, then construct adjacency matrix of modules, weighted or binary
- Note yeast appeared at about 12.1 ca



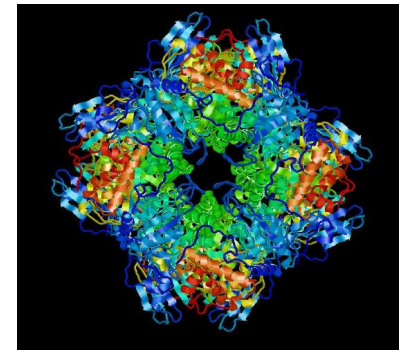
Experimental Tests of Spontaneous Emergence of Modularity



- Single, positive stranded RNA virus; in vitro and in vivo rhinovirus
 - Defective interfering (DI) particles
 - Do they bank diversity?
 - Is that transmitted to infectious particles (virus)?
 - Vary immune pressure; does DI/I ratio increase with variation?
- Supra-genome effect in prokaryotic model
 - Population level banking of diversity
 - Hosts: mice (*S. pneumoniae*) or chinchilla (*H. Influenza*)
 - Vary host immune system
 - Measure fraction of diversity not in an individual (degree of supra-genome effect)
- Gnotobiotic miniature pigs; VDJ recombination
 - Measure naïve B cell diversity
 - As a function of exposure to controlled environmental antigens, e.g. viruses and bacteria
 - Increased VDJ recombination \equiv modularity development
- Regulatory networks evolving under changing conditions (Tim Cooper, UH)

Conclusions

- Spontaneous emergence of modularity in a population of individuals in a changing environment
- The velocity of modularity increase depends on the magnitude (p) and frequency (f) of the environment change
- Selection in a changing environment generically leads to modularity in the presence of horizontal gene transfer
- A symmetry breaking event
- Beautiful hierarchical structures observed in nature may be a result of selection for evolvability
- Need not necessarily rely on intelligent design or the anthropic principle



Acknowledgements

- Members of the FunBio team for stimulating discussions (DARPA)