How evolution designs living matter

Ard Louis
Propositiones ad acuendos iuvenes
“Problems to sharpen the young”

• Problem 13:

“A king ordered his servant to collect an army from 30 manors, in such a way that from each manor he would take the same number of men he had collected up to then. The servant went to the first manor alone; to the second one he went with one other; to the next he took three with him, How many were collected from the 30 manors?”

\[2^{30} - 1 = 1,073,741,823\] (1 billion) soldiers,

Alcuin of York
735-804
A barley-corn: to a single barley-corn I increased, 2 barley-corns in the 1st day; 4 barley-corns in the 2nd day; 8 barley-corns in the 3rd day; .. ..

30 2 ‘thousand’ 7 ‘hundred’ 37 talents 1/2 mina 2 1/3 shekels 4 barley-corns in the 30th day.

Mari 08613 tablet
(Old Babylonian. 1900-1600 BC)

Jöran Friberg (2005), *Unexpected Links Between Egyptian and Babylonian Mathematics*
$2^{64} - 1 = 18,446,744,073,709,551,615$

Shah-nama (Persian: شاهنامه Šāhnāme, "The Book of Kings")
by Ferdowsi (فردوسي) 940 – 1020
History of life on earth

- Late heavy bombardment
- Origin of life?
- Earth forms from accretion disk

**The Evolution of Life**

- **Precambrian**
  - 700: Simple multicellular organisms evolve
  - 2,100: Oldest eukaryotic fossils
  - 2,500: Oxygen begins to accumulate in atmosphere
  - 3,500: Oldest prokaryotic fossils

- **Paleozoic**
  - 540: Plants colonize land
  - 360: First insects
  - 370: Amphibians appear
  - 340: Reptiles appear

- **Mesozoic**
  - 245: First dinosaurs and mammals
  - 230: First flowering plants

- **Cenozoic**
  - 141: Birds evolve from reptiles
  - 66.4: Mass extinction

- Present time
  - 0.0: Advent of modern humans
1859: Variation and Natural Selection
Arrival of the fittest?

where does variation come from?

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

*Species and Varieties. Their Origin by Mutation.* Chicago: Open Court (1904)

Natural selection as a sieve?  

Hugo de Vries 1848-1935
Modern Synthesis

R.A. Fisher
1890 – 1962

JBS Haldane
1892 – 1964

Sewall Wright
1889 – 1988
How big is evolutionary search space?
Estimates of the total number of genes in the cells of higher organisms range from 1000 up ... With 10 alleomorphs in each of 1000 loci, the number of possible combinations is $10^{1000}$ which is a very large number. It has been estimated that the total number of electrons and protons in the whole visible universe is much less than $10^{100}$

hyperastronomical

“The population is thus confined to an infinitesimal portion of the
field of possible gene combinations”

1. Nature can only explore an unimaginably small fraction of all theoretically possible genomes.

“The chance that a random combination is as adaptive as those characteristic of
the species may be as low as $10^{-100}$ and still leave room for $10^{800}$ separate peaks
[adaptive gene combinations], each surrounded by $10^{100}$ more or less similar
combinations. “

2. The current instantiation of genetic possibilities (life as we know it) is largely contingent,
since it could just have well occupied a different part of genotype space.
“...under biparental reproduction a very low rate of mutation balanced by moderate selection is enough to maintain a practically infinite field of possible gene combinations within the species“

3. The variation for natural selection to act on is abundant.

4. Variation does not introduce a significant bias in evolutionary trajectories
Where does evolutionary novelty come from?

MS: selection (not variation) is primary causal force

*Evolution is not primarily a genetic event. Mutation merely supplies the gene pool with genetic variation; it is selection that induces evolutionary change.*

Evolution and hyper-astronomical numbers

Proteins: linear chains made from an alphabet of 20 amino acids

Hoyle Paradox
all combinations pf length 100 proteins would weigh more than the visible universe.

Fred Hoyle 1915-2001
what does search space look like?
Hammerhead ribozyme keeps emerging from SELEX in-vitro evolution

RNA: alphabet of 4 nucleotides, so $L=55$ means $4^{55} \sim 10^{33}$

Convergent Evolution?

North America:
Placental Sabre-toothed cat

South America”
Marsupial Sabre-toothed cat
Convergent Evolution?

compound eye

camera eye
Convergent Evolution?

- Enormous number of examples ... from proteins to vision up to societies to intelligence.
- [http://www.mapoflife.org](http://www.mapoflife.org)
- Why all this convergence? variation or selection or something else?

The principal aim of this book has been to show that the constraints of evolution and the ubiquity of convergence make the emergence of something like ourselves a near-inevitability. Simon Conway Morris, “Life’s Solution”, (CUP 2003) pp328
Biological self-assembly

- Can we understand?
  - Self-assembly
  - Evolution
- Can we emulate (nanotechnology?)

Movie from: Keiichi Namba, Osaka ERATO project
Virus self-assembly

viruses assemble from identical capsomeric units
“computer virus” self-assembly

Computer viruses?

Biological self-assembly
Science is fun!
Protein folding (self-assembly)

Levinthal Paradox (1968):
150 amino acids
~10 angles between them
~$10^{150}$ different states.
How does protein find its folded native structure?

The search is not random!
Energy landscape for a virus capsid

The search is not random!
Evolution and hyper-astronomical numbers

Proteins: linear chains made from an alphabet of 20 amino acids

**Hoyle Paradox**

100 residue protein; 20 amino acids => combinations would weigh more than all the atoms in the visible universe.

Fred Hoyle 1915-2001
how are genotypes distributed over phenotypes?

$$N_G >> N_P$$

Neutral theory of evolution:
Kimura 1968
King and Jukes 1969

Lots of mutations are neutral …
e.g. proteins can have the same function with > 50% sequence dissimilarity

Motoo Kimura
1924-1994
how are genotypes distributed over phenotypes?

$N_G \gg N_P$

a) Approx. equal number of genotypes per phenotype

b) Highly unequal number of genotypes per phenotype: this is bias
Model GP map: RNA secondary structures

\[ N_G = 4^L \]

\[ N_P = 0.02 \times 1.93^L \]

\[ N_G >> N_P \]

Genotype: GCGGAUAUCUGCAAU

Phenotype: (((((.....)))....)

Sequence Structure
L=15: $4^{15} \sim 1 \times 10^9$ sequences -- 431 phenotypes -- but 26/431 take up 50% of G-space.
Convergent Evolution?

Hammerhead ribozyme keeps emerging from SELEX in-vitro evolution

RNA: alphabet of 4 nucleotides, so $L=55$ means $4^{55} \approx 10^{33}$

larger RNA?

RNA GSS–rank plot, L=55

GSS

10^0

10^5

10^10

10^15

10^20

10^25

Rank (M~10^{14})

x 10^{13}
RNA topology dominated by variation

hammerhead ribozyme
protein quaternary structure

Proteins self-assemble into quaternary structure

• Can we understand?
  – Assembly
  – Evolution?
• Can we emulate?

Movie from: Keiichi Namba, Osaka ERATO project
Neutral Space Topology

1) $N_G \gg N_P$ (number of genotypes $\gg$ number of phenotypes)

2) Genotype set (GS) size per phenotype is highly skewed

Why the skew??
Algorithmic information theory

Kolmogorov complexity $K(s)$ is roughly speaking the length of the shortest computer program in some fixed language $L$ that produces $s$ as an output, minimized over all languages $L$.

A.N. Kolmogorov  1903-1987  
G.J. Chaitin  1947--  
R. Solomonoff  1926-2009
Universal probability and the algorithmic nature of the world

\[ P(x) = \sum_l 2^{-l} \approx 2^{-k} \]

P(x) = probability that a random sequence, fed into a universal Turning machine, produces the output x.

The sum is over all halting programs of length l that produce x.

intuitively: simpler outputs are found more often.

Coding Theorem by Solomonoff and Levin.

deep connections to Occam’s razor and Bayes priors.

The Oxford crew ...
Biological networks: interacting many-body systems

Why are there so few genes?

Complexity comes from the interactions (systems biology)

Model as networks of differential equations?

Random Matrix Theory

What are the collective modes?

How do they evolve?

Transcriptional network for yeast: *Saccharomyces cerevisiae*