Bioinformatics: The Information in Life

Don Johnson

Ph.D. Chemistry: Michigan State Univ.

Ph.D. Computer & Information Sciences: U of Minn

Topics for This Presentation

The nature data versus 3 types of bio-information

The roles of chance and probability

Information and its processing systems in every cell

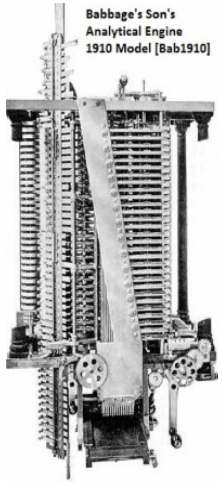
Information theory ramifications

Information and evolution

Unanswered problems

Life as Computer System?

Mechanical computer designed 1837



"The machine code of the genes is uncannily computer-like. Apart from differences in jargon, the pages of a molecular biology journal might be interchanged with those of a computer engineering journal." Dawkins River Out of Eden: A Darwinian View of Life, p17

"The information content of a simple cell has been established as around 10^{12} bits, comparable to about a hundred million pages of the Encyclopaedia Britannica." Carl Sagan, "Life," Encyclopaedia Britannica: 22, 1997, p964-981

Data vs 3 Kinds of Information

Data may or may not have meaning

Binary is the smallest base to hold data in a bit

A binary digit (bit) can represent any 2 possibilities

married/single resident/nonresident male/female

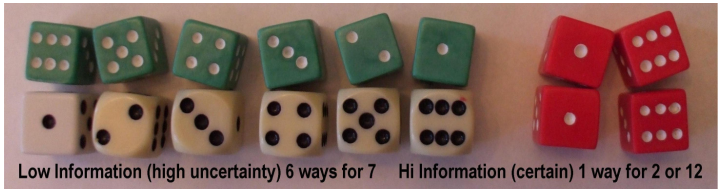
If 110 is married nonresident female,

001 is single resident male (arbitrary)

Information: contingency ruling out other possibilities

- **Functional: useful/purposeful/meaningful**
- **Prescriptive: instructional/algorithmic choices**
- **Shannon: reduction of possibilities or uncertainty**

Shannon Information (information theory)



Purely probability-based – functionality not required

Redundant patterns provide no additional information

"junkjunkjunkjunk": only1st is information

Shannon info defines limits on info storage or transmission

e.g. – Zip compresses file retaining Shannon info

Random data (0 functional info) has maximum Shannon info

Cannot be compressed using a more concise alphabet

Examples of Coded Information & Data

- Random coin tosses/1010100 = ASCII 'T' (head=1)
- Random die throws/Minneapolis area code

Chance can't produce functional coded information

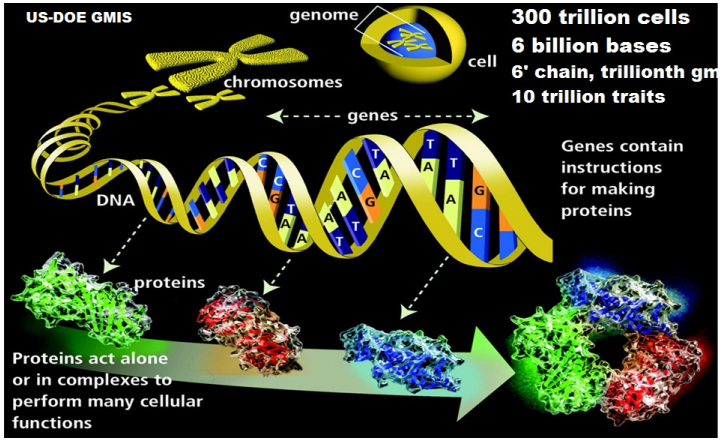
Pattern match probability: 1/128 (coins), 1/216 (dice)

Information determined by protocol rules, not law

Sender and receiver must agree on arbitrary rules

1100101 means 'e' using ASCII (e.g. for printer)

Simplified Genetic Code for Protein Construction



Simplified View of Life's Incredible Complexity

25,000 genes (many overlapping to produce >100,000 proteins)

"A single gene can potentially code for tens of thousands of different proteins... It's the way in which genes are switched on and off, though, that has turned out to be really mind-boggling, with layer after layer of complexity emerging" Le Page, "Genome at 10," New Scientist, 6/16/10.

Digital (base 4) self-correcting encoded information

Group of 3 1-of-4 bases (ACGT) : 4^3 (= 64) possible codons

20 amino acids for proteins redundantly codon-specified

Information in 1 teaspoon of DNA: all people + all books

Even "simplest" organism's DNA has >150,000 nucleotides

DNA, proteins, etc. must be fully-formed/functional

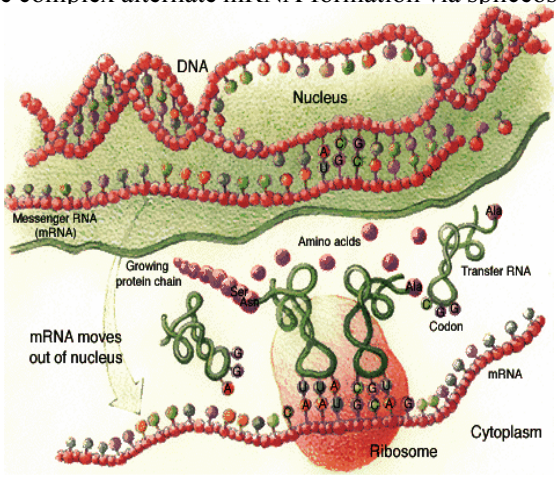
>2000 enzyme proteins enable reactions

Slowest non-enzymatic reaction would take a trillion yrs

"Human DNA is like a computer program but far, far more advanced than any software we've ever created." Bill Gates, The Road Ahead, p.228.

Simplified DNA Transcription/Translation Process

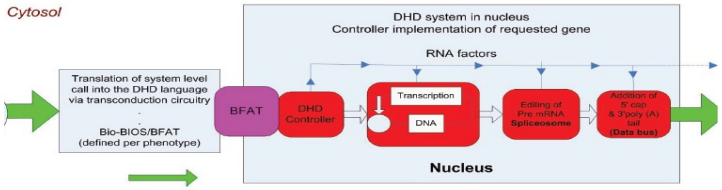
(more complex alternate mRNA formation via spliceosomes)



Information Systems in Life

- **Genetic system is a preexisting operating system**
- **Specific genetic program (genome) is an application**
- **Native language has codon-based encryption system**
- **Codes are read by enzyme "computers" with own OS**
- **Enzyme's output is to another OS in a ribosome**
- **Codes are decrypted and output to tRNA computers**
- **Codon-specified amino acid is transported to protein construction site**
- **In each cell, there are multiple OSs, multiple programming languages, encoding/decoding hardware and software, specialized communications systems, error detection/correction mechanisms, specialized input/output channels for organelle control and feedback, and variety of specialized "devices" to accomplish the tasks of life.**

A comparative approach for the investigation of biological information processing, D'Onofrio & An, Theoretical Biology and Medical Modeling, 1/21/10



Disk/DNA properties & functional equivalences are compared

Chromosome/partition, file/gene, fragmentation/epigenome

"The cell is viewed as a complete computational machine in terms that are akin to a multi-core computer cluster, where there is a centralized memory and instruction set, yet computational tasks are distributed among distinct processing elements… as a system with centralized memory with multi-access capability leading to distinct computing units."

Shannon Channel Capacity

DNA/RNA/protein system uses info theory equations:

Discrete because all alphabet symbols are defined

Memoryless because there's no history dependence

Unconstrained because any symbol may follow

Crick's "Central dogma" predicted by info theory

Information transfer from protein to RNA impossible

(20 to 64 symbols would exceed channel capacity)

Life's initial alphabet was at least that of codon since any predecessor MUST be at least that of successor

Algorithmic Prescriptive Information (PI) in Life

Biosemiotics: cybernetic sign-systems in life

Chance & law can't explain decision nodes (choice)

PI is intrinsically formal, but implemented physically

Abel, "The Biosemiosis of Prescriptive Information," Semiotica:174-1, 2009, p1-19

"No rational scientific basis exists for blindly believing in a relentless uphill push by mere physicality toward formal algorithmic optimization" Abel & Trevors, "Self-Organization vs Self-Ordering events in Life-Origin Models," Physics of Life Rev:3, 2006, p211-228.

Self-ordering is not self-organization (purposeful)

"The Origin-of-Life Prize® ... will be awarded for proposing a highly plausible natural-process mechanism for the spontaneous rise of genetic instructions in nature sufficient to give rise to life."

Neo-Darwinian Biology

Richard Dawkins: "Each nucleus ... contains a digitally coded database larger, in information content, than all thirty volumes of the Encyclopedia Britannica" "Each successive change in the gradual evolutionary process was simple enough, relative to its predecessor, to have arisen by chance… Even if the evidence did not favour it [evolution by natural selection], it would still be the best theory available!" "Mutation is not an increase in true information content, rather the reverse." ***Climbing Mount Improbable, Blind Watchmaker, Information Challenge***

"The failure to observe even one mutation that adds information is more than just a failure to support the theory. It is evidence against the ... neo-Darwinian theory." ***(Spector, Not By Chance, p160)***

Information Increase Moving up Tree

- **The simplest life has only 267,000 information bits**
- **Human DNA has over 6 billion information bits**
- **Based on functional information, simplest life is $10^{300,000,000}$ more probable than man**
- **No mechanism to produce ANY net info increase**

  New functionality offset by functionality loss, e.g.--

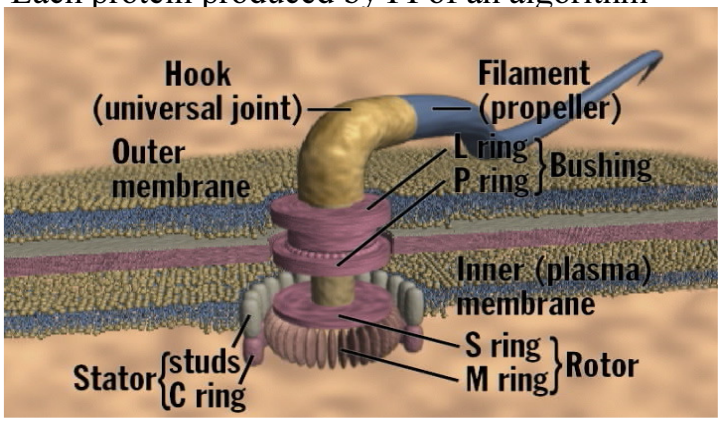
  Single mutation causes sickle cell anemia

  Frame-shift may enable nylon-digesting bacteria
- **"We must concede there are presently no detailed Darwinian accounts of the evolution of any biochemical or cellular system, only a variety of wishful speculations"** Harold, The Way of the Cell, 2001, p205.

Bacterial Flagellum: Irreducibly Complex

48+ proteins (>30 unique): $<10^{-5250}$ probability

Each protein produced by PI of an algorithm



Evolution via Natural Genetic Engineering

"Molecular cell biology has revealed a dense structure of information-processing networks …The natural genetic engineering functions that mediate genome restructuring are activated by multiple stimuli…One of the traditional objections to Darwinian gradualism has been that it is too slow and indeterminate a process to account for natural adaptations, even allowing for long periods of random mutation and selection … natural genetic engineering … employs a combinatorial search process based upon DNA modules that already possess functionality … Such a cognitive component is absent from conventional evolutionary theory because 19th and 20th century evolutionists were not sufficiently knowledgeable about cellular response and control networks." James A Shapiro, "Mobile DNA and evolution in the 21st century," Mobile DNA 1/25/10

Darwinism Revisited

"**The complexity of biology has seemed to grow by orders of magnitude… Biology's new glimpse at a universe of non-coding DNA — what used to be called 'junk' DNA — has been fascinating and befuddling… the signaling information in cells is organized through networks of information rather than simple discrete pathways. It's infinitely more complex.**" Erika Hayden, "Life is Complicated," Nature, 4/10, p664-667

"**Much of the vast neo-Darwinian literature is distressingly uncritical... Natural selection has shown insidious imperialistic tendencies."** Fodor & Piattelli-Palmarini, "Survival of the fittest theory: Darwinism's limits," New Scientist, 2/3/10

"Natural selection is not a mechanism, it's the process by which the results of evolution are sorted." Bruce Runnegar, p188 of The Altenberg 16: An Exposé of the Evolution Industry, 2010 (Mazur)

Summary

Data is functional information only if it can be used

Life is incredibly complex and information rich

Is information science incorrect?

- **Can chance produce complex functional information?**
- **Can multiple mutational information losses cause gain?**
- **Can chance produce codes or formal protocols?**
- **Was life's first code simpler than the current codon code?**
- **Can chance write prescriptive algorithms (programs/OSs)?**
- **Can chance create genetic engineering capability?**

Scenarios proposed inadequately address information

- **Assertions for origins of life & species need verification**
- **Other avenues may provide more fruitful paths**

Science speculation is inappropriate for non-scientists

Computer Simulations & Artificial Life

Dawkins *(Scientific American, 6/88)* randomly changed:

"WDLTMNLT DTJBKWIRZREZLMQCO P"

to produce on the 43rd try:

"METHINKS IT IS LIKE A WEASEL"

He knew the goal in advance and stopped mutation if correct proving that programmers can solve problems using computers.

"Everywhere on the apparatus and in the 'genetic algorithms' appear the scientist's fingerprints: the 'fitness functions' and 'target sequences.' These algorithms prove what they aim to refute: the need for intelligence and teleology [targets] in any creative process." George Gilder, "Evolution and Me," National Review, 7/17/06

"Neglect of key factors or unrealistic parameter settings permit conclusions to be claimed which merely reflect what the decision maker intended a priori." Royal Truman, "Evaluation of Neo-Darwinian Theory Using the Avida Platform," PCID 3.1.1, 11/04.

POL highlights the informational aspects of life that are usually overlooked or ignored in chemical and biological evolutionary. Each cell of an organism has millions of interacting computers reading and processing digital information using algorithmic digital programs and digital codes to communicate information. Most scientists have been attempting to use physical science to explain life's information domain, a practice which has no scientific justification. For more info see scienceintegrity.net PoL link