Are there recipes how to handle complexity?

Biological evolution creates complex entities and knows how to master them

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

Complexity Primer

London, Law Society, 08.05.2008
Web-Page for further information:

http://www.tbi.univie.ac.at/~pks
Catastrophic weather phenomena – storm, lightning, tornado and hurricane
The Mayas of Chichen Itza
Pyramid, Chaac, and cenote sagrada
Raleigh-Bénard convection and hurricane formation
\[
\frac{\partial u}{\partial t} = D_u \nabla^2 u + f(u, v) \\
\frac{\partial v}{\partial t} = D_v \nabla^2 v + g(u, v)
\]

\[u = u(x, y, z, t) \text{ and } v = v(x, y, z, t)\]

Change in local concentration =

= diffusion + chemical reaction

Alan M. Turing, 1912-1954

Nonequilibrium patterns from chemical self-organization:

Liesegang rings in precipitation from oversaturated solutions, periodic patterns in the Belousov-Zhabotinskii reaction, and stationary Turing patterns.

Turing pattern:
Boissonade, De Kepper 1990
Color patterns on animal skins
Different forms of mimicry observed in nature
Bates' mimicry

milk snake

false coral snake

coral snake

Emsley's or Mertens' mimicry

Different forms of mimicry observed in nature
Skin patterns in an inbred strain of cats

Parents and daughter
Genotype, Genome

GCGGATTTAGCTAGTTGGAGAGCGCCAGACTGAAGATCTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

Unfolding of the genotype

biochemistry  molecular biology
structural biology  molecular evolution
molecular genetics  systems biology
bioinformatics

development

genetics  epigenetics  environment

cell biology  developmental biology
neurobiology  botany
zoology  anthropology
ecology

Phenotype
Duplication of genetic information

Deoxyribonucleic acid – DNA
The carrier of digitally encoded information
A sketch of cellular information processing
A sketch of a genetic and metabolic network
The reaction network of cellular metabolism published by Boehringer-Ingelheim.
The citric acid or Krebs cycle (enlarged from previous slide).
Three necessary conditions for Darwinian evolution are:

1. **Multiplication**,  
2. **Variation**, and  
3. **Selection**.

**Multiplication** is a basic property of all cells in germ lines.

**Variation** through mutation and recombination operates on the **genotype** whereas the **phenotype** is the target of **selection**. **Variations**, mutations or recombination events, occur **uncorrelated** with their effects on the **selection** process.

**Selection** is a consequence of finite population sizes.

All conditions can be fulfilled not only by cellular organisms but also by nucleic acid molecules in suitable cell-free experimental assays.
Variation of genotypes through mutation and recombination
Variation of genotypes through mutation

- Parent sequence
- Point mutation
- Insertion
- Deletion
Chemical kinetics of molecular evolution

\[
\frac{dx_i}{dt} = \sum_{j=1}^{N} Q_{ij} f_j(x) x_j - x_i \Phi(x), \quad i = 1, 2, \ldots, N
\]

\[\Phi(x) = \sum_{j=1}^{N} f_j(x) x_j, \quad \sum_{j=1}^{n} x_j = 1, \quad \text{and} \quad N = 4^n\]

Chemical kinetics of molecular evolution

Formation of a quasispecies in sequence space
Formation of a quasispecies in sequence space
Formation of a quasispecies in sequence space
Formation of a quasispecies in sequence space
Uniform distribution in sequence space

Mutant cloud

Uniform distribution in sequence space
Quasispecies

Driving virus populations through threshold

The error threshold in replication
Antiviral strategy on the horizon

Error catastrophe laid its conceptual origin in the middle of the 1970s century, when the consequences of mutations on oncogenes involved in progeny synthesis, as a theory of stage in these circumstances, became biologically perceived differently from today. Infections diseases were regarded as a fleeting innate which would be damaged throughout the use of antibiotics and antiviral agents. Microbial variation, although known in some cases, was not thought to be a significant problem for the disease control. Variations in differentiated organisms were seen as resulting essentially from exchanges of genetic material associated with sexual reproduction. The problem was to predict the mechanics of microorganisms, expressions of genetic information, and metabolism. Few pointed that genetic change was occurring at rates in all organisms, and still fewer recognized Darwinian principles as essential to the biology of pathogenic viruses and cells. Populations genetics rarely used bacteria or viruses as experimental systems to define concepts in biological evolution. The ever-growing genetic polymorphism among individuals of the same biological species came as a surprise when the first results on comparative electrophoretic mobility of enzymes were obtained. The advent of the virus DNA, RNA, and rapid nucleic acid sequencing technique, molecular analyses of enzymes reinforced the conclusion of extreme inter-individual genetic variation within the same species. Now, due largely to sparseness progress in comparative genomics, we see cellular DNAs, both prokaryotic and eukaryotic, as highly dynamic. Many cellular processes, including such essential functions, such as transferring events as genome replication, transcription, and translation, are increasingly perceived as inherently inaccurate. Unreliable, and in contrast to bacteria, are among the more advanced examples of exploitation of replication inaccuracy for survival.

Error catastrophe, the loss of meaningful genetic information through random genetic variation, was formulated in quantitative terms as a consequence of quasimutants theory, which was first developed to explain self-organization and adaptability of primitive replicates in early stages of life. Recently, a conceptual extension of error catastrophes that could be defined as "induced genetic destruction" has emerged in a possible natural strategy. This is the topic of the current special issue of *Virus Research*.

Foreword: authors of this issue of the special issue share the notion that the control of viral disease is short-term adaptability of viral pathogens. Adaptability of viruses follows the same Darwinian principle that have shaped biological evolution over eons; that is, repeated rounds of reproduction with genetic variation, competition, and selection, often perpetuated by random events such as climatic fluctuations in population size. However, with variations the consequences of the operation of these same Darwinian principles are felt within very short times. Short-term evolution (within hours and days) can be observed with some cellular pathogens, with subsets of normal cells, and cancer cells. The nature of RNA virus pathogenesis is the potential antiviral strategy, and forcing the virus to cross the critical error threshold for maintenance of genetic information is one of them.

The contributions to this volume have been chosen to reflect current lines of evidence (manuscripts and experimental results) in which natural designs based on genetic determinism varied upon viruses are being constructed. Theoretical studies have explored the copying error conditions that must be fulfilled by any information-bearing replication system for the essential genetic information to be transmitted. Closest related to the theoretical developments have been numerous experimental studies on quasimutants dynamics and their multiple biological manifestations. The latter can be summarized by saying that RNA viruses, by virtue of accuracy at mutation spectra rather than defined genetic entities, remarkably expose their potential to overcome selective pressures correlated to their replication hindrance, the high error rate mutations in clinical practice and the design of vaccines for a number of canaries RNA virus-associated diseases, are currently provided by a degree of uncertainty. Another line of research is the understanding of copying fidelity by viral replication, named at understanding the molecular basis of quasimutants activities. Error catastrophe as a potential new antiviral strategy remains an important issue by the observation that viruses (a licensed antiviral nucleic acid analogue) may be inserted, in recent systems, its natural activity through enhanced transcription.

The idea to prepare this special issue came as a kind invitation of Ulrich Desselberger, former Editor of *Virus Research*, and then asked enthusiastically by Luis Esteban, recently appointed as Editor of *Virus Research*. I take this opportunity to thank Ulrich, Luis and the Editor-in-Chief of *Virus Research*, Bruno Mele, for their continued interest and support to the research on viral evolution over the years.

My thanks go also to the 10 authors who despite their busy schedules have taken time to produce excellent manuscripts, as Elsevier staff for their prompt responses to my requests, and last but not least, to Mrs. Lourdes Herrero from Centro de Biología Molecular “Severo Ochoa” for her patient dealing with the correspondence with authors and the final organization of the issue.

Esteban Esteban
Unidades Asociadas Madrid
Centro de Biología Molecular “Severo Ochoa”
Consejo Superior de Investigaciones Científicas
Casillón and Coalición
Madrid. Spain.
Tel: +34 91 407 8455/9; fax: +34 91 407 4700
E-mail address: econz@cnmc.es
Available online 6 December 2004
Motoo Kimura's Populationsgenetik der neutralen Evolution.

Evolutionary rate at the molecular level. 

*The Neutral Theory of Molecular Evolution.*
Neutral network

Sequence space

Structure space
N = 7

Neutral networks with increasing $\lambda$
$N = 24$

Neutral networks with increasing $\lambda$
$N = 68$

Neutral networks with increasing $\lambda$

Neutral network

$\lambda = 0.20, \ s = 229$
A sketch of optimization on neural networks
An example of selection of molecules with predefined properties in laboratory experiments
Secondary structure of the tobramycin binding RNA aptamer with $K_D = 9$ nM

Application of molecular evolution to problems in biotechnology
Results from molecular evolution in laboratory experiments:

• Evolutionary optimization does not require cells and occurs in molecular systems too.

• *In vitro* evolution allows for production of molecules for predefined purposes and gave rise to a branch of biotechnology.

• Direct evidence that neutrality is a major factor for the success of evolution.

• Novel antiviral strategies were developed from known molecular mechanisms of virus evolution.
The bacterial cell as an example for the simplest form of autonomous life

Escherichia coli genome:
4 million nucleotides
4460 genes

The structure of the bacterium *Escherichia coli*
**E. coli:**  
Genome length $4 \times 10^6$ nucleotides  
Number of cell types 1  
Number of genes 4 460

Four books, 300 pages each

**Man:**  
Genome length $3 \times 10^9$ nucleotides  
Number of cell types 200  
Number of genes $\approx 30 000$

A library of 3000 volumes,  
300 pages each

Complexity in biology
Evolution does not design with the eyes of an engineer, evolution works like a tinkerer.

Francois Jacob, Pantheon Books, New York 1982
A model for the genome duplication in yeast 100 million years ago

WHAT IS A GENE?

The idea of genes as beads on a DNA string is fast fading. Protein-coding sequences have no clear beginning or end and a DNA string is a key part of the information package, reports Helen Pearson.

Gene is not a typical four-letter word. It is not offensive. It is never dropped out of TV shows. And where the meaning of most four-letter words is too clear, that of gene is not. The more expert scientists become in molecular genetics, the less easy it is to be sure about what, if anything, a gene actually is.

Rick Young, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, says that when he first started teaching as a young professor two decades ago, it took him about two hours to teach fresh-faced undergraduates what a gene was and the nuts and bolts of how it worked. Today, he and his colleagues need not more than fifteen minutes to convey the concept of the gene, and that's not because the students are any less bright. "It takes a whole semester to teach this stuff to hardened graduates," Young says. "It used to be you could give a one-off definition and now it's much more complicated."

In classical genetics, a gene was an abstract concept — a unit of inheritance that ferried a characteristic from parent to child. As biochemistry came into its own, those characteristics were associated with enzymes or proteins, one for each gene. And with the advent of molecular biology, genes became real, physical things — sequences of DNA which when converted into strands of so-called messenger RNA could be used as the basis for building the associated protein piece by piece. The great coiled DNA molecules of the chromosomes were seen as long strings on which gene sequences sat like beads on a string. This picture is still the working model for many scientists. But those at the forefront of genetic research see it increasingly old fashioned — a crude approximation that, at best, hides fascinating new complexities and, at worst, blinds its users to useful new paths of enquiry.

Information, it seems, is parcelled out along chromosomes in a much more complex way than was originally supposed. RNA molecules are not just passive conduits through which the gene's message flows into the world but active regulators of cellular processes. In some cases, RNA may even pass information across generations — normally the sole preserve of DNA.

An eye-opening study last year raised the possibility that plants sometimes rewrite their DNA on the basis of RNA messages inherited from generations past. A study on page 469 in this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals. If this type of phenomenon is indeed widespread, it "would have huge implications" says evolutionary geneticist Laurence Hurst at the University of Bath, UK. "All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Reis at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the International Encyclopedia of DNA Elements (ENCODE) project. The pilot phase of ENCODE involves scrutinizing roughly 1% of the human genome in unprecedented detail; the aim is to find all the sequences that serve a useful purpose and explain what that purpose is. "When we started the ENCODE project I had a different view of what a gene was," says contributing researcher Federica Giorgi at the Center for Genomic Regulation in Barcelona. "The degree of complexity we've seen was not anticipated."

Under fire

The first of the complexities to challenge molecular biology's paradigm of a single DNA sequence encoding a single protein was alternative splicing, discovered in viruses in 1977 (see 'A dead track' on p.468). Most of the DNA sequences described in humans have modular arrangement in which exons, which carry the instructions for making proteins, are interspersed with non-coding introns. In alternative splicing, the cell snips out introns and sews together the exons in various different orders, creating messages that can code for different proteins. Over the years geneticists have also documented overlapping genes, genes within genes and countless other weird arrangements (see 'Modelling over genes, over DNA' on p.469).

Alternative splicing, however, did not in itself require a drastic reappraisal of the notion of a gene: it just showed that some DNA sequences could describe more than one protein. Today's assault on the gene concept is more far-reaching, fuelled largely by studies that show the previously unimaginable scope of RNA.

The one gene, one protein idea is coming under particular attack from researchers who are comprehensively extracting and analyzing the RNA messages, or transcripts, manufactured by genes, including the human and mouse genome. Researchers led by Thomas Gingeras at the company Affymetrix in Santa Clara, California, for example, recently studied all the transcripts from eleven chromosomes across eight human cell lines and worked out precisely where on the chromosomes each of the transcripts came from. The picture these studies paint is one of mind-boggling complexity. Instead of discrete genes faithfully mass-producing identical RNA transcripts, a teaspooning mass of transcription converts many segments of the genome into multiple RNA ribbons of differing length. These ribbons can be generated from both strands of DNA, rather than from sense as was conventionally thought. Some of these transcripts come from regions of DNA previously identified as holding protein-coding genes. But many do not. "It's somewhat revolutionary," says Gingeras colleague Philip Karpman. "We've come to the realization that the genome is full of overlapping transcripts."

Other studies, one by Craig Venter and one by geneticist Rotem Sorek, both at the Washington University, Israel, and his colleagues, have hinted at the reasons behind the mass of transcription. The two teams investigated occasional reports that transcripts can start at a DNA sequence associated with one protein and run straight through into the gene for a completely different protein, producing a chimeric transcript. By delving into databases of human RNA transcripts, Gaige's team estimate that 4-5% of the RNA in regions conventionally recognized as genes is transcribed in this way. Producing fused transcripts could be one way for a cell to generate a greater variety of proteins from a limited number of exons. This idea is also one of the central points to emerge from the ENCODE project when its results are published later this year.

Karpman and others say that they have documented many examples of transcripts in which protein-coding exons from one part of the genome correlate with exons from another.

The difficulty to define the notion of "gene".

Helen Pearson, Nature 441: 399-401, 2006
ENCODE stands for **ENCyclopedia Of DNA Elements**.

**ENCODE Project Consortium.**
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.  
Fast and frugal heuristics use simple rules for

• guiding search for information,
• stopping search, and
• decision making.

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks