Biological complexity and Evolution

Marco V. Bayas December 11, 2001

Abstract.

Biological complexity is intimately related to the amount of information in a given system. The way in which information arises and consequently how complexity increases in biological systems is discussed. Both are consequences of the principle of natural selection introduced by Darwin.

Introduction.

Before start talking about biological complexity is convenient to give an appropriate definition of it. This definition should allow us to quantify biological complexity and consequently make it susceptible of systematic study. We could start pointing out that biological systems are more complex than those with the same material components but without life. Therefore, any definition of biological complexity will be intrinsically related to a definition of life. But there is not a completely satisfactory definition for life.

As we know now, all of the characteristics of a living being are encoded in its DNA. So, the study of biological complexity should start with the study of the complexity of the genetic information. Because the genetic information lies in the DNA as a linear sequence of symbols, the quantification of biological complexity takes us ultimately to the quantification of the information that can be put into a sequence of nucleotides.

In this article, we will analyze one way of quantify biological information and how its evolution can be studied.

Information Theory and Complexity

Considering the genetic content of organisms, we can say that certainly eukaryotes are more complex than prokaryotes and, human beings are more complex than amoeba. But, how more complex?. One can answer this question using the concepts of has been given by the theory of information.

In the context of the information theory, the information content is the average number of binary decisions (bits) that are necessary in order to identify unambiguously a particular sequence of symbols [3]. A close related quantity is the entropy defined in Shannon's information theory. The entropy (S) represents the expected number of bits

required to specify the state of a physical object given a distribution of probabilities; that is, it measures how much information can potentially be stored in it. In a genome, for a site *i* that can take four nucleotides with probabilities $P_C(i)$, $P_G(i)$ $P_A(i)$ and $P_T(i)$, the entropy is [1]

$$S_i = -\sum_{j}^{C,G,A,T} P_j(i)\log_2 P_j(i)$$

The complexity of genomes has been defined [1] as the amount of information they encode about the world in which they have evolved[1]. Consequently, for an organism with N base pairs the complexity has been defined as:

$$C = N - \sum_i S_i$$

This way of quantify the complexity allows the study of its evolution.

Evolution of Complexity

Natural selection is the mechanism of biological evolution. Evolution is the steady generation of information, information that is written down in the genes of living organisms. In the same way complexity, built upon simplicity, has accumulated throughout biological evolution constrained by natural selection.

Evolution, in the Darwinian sense requires a population of organisms susceptible of self-reproduction, mutagenicity, and metabolism. Simulation of evolution by natural selection can be performed in systems where these conditions are held. One of these systems involves "digital organisms" in the "Avida environment" [1],[4].

The Avida systems host populations of self-replicating computer programs in a complex and noisy environment, within a computer's memory. The evolution of these, "digital organisms" typically takes only a few seconds for populations of the order of 10^3 - 10^4 programs. This is, of course advantage over true biological systems. Simulations, using Avida systems have demonstrated that only mutations that reduce the entropy are kept while mutations that increase it are eliminated [1].

Discussion.

Natural selection is the mechanism that allows biological complexity to increase. It makes possible the appearance of more efficient organisms. The increase in efficiency for survival is a consequence of the increasing of the amount of information a given organism has about its environment.

[1] This is also known as "physical complexity" [1], [2]

The change in the content of information requires a change in the probability distribution of the nucleotides. There is no perturbation that can change the probability distribution in the system, as long as the external conditions are unaltered. Thus, information cannot arise in systems in thermodynamical equilibrium. Actually, living systems are open systems "working" in stationary conditions.

As mentioned by Adami et al [1], the Maxwell Demon mechanism is at work during all phases of evolution allowing a continuous decreasing of entropy and consequently an increase of complexity in the natural world. This is the behavior observed in the evolution of living organisms as well. We could say that the distinctive characteristic of life is the continuous increment of complexity.

References

1. Adami, C., Ofria, C. & Collier, T. (2000). *Evolution of Biological Complexity*. Proc. Nat. Acad. Sci. USA. 97: 4463-4468

2. Adami, C. & Cerf, N.J. (2000), *Physical complexity of symbolic sequences.* Physica D. 137: 62-69.

3. Eigen, M. (1992). *Steps toward life. A perspective on evolution.* Oxford university press.

4. Digital Life Laboratory. http://dilab.caltech.edu/avida