

Biology of Digital Organisms: How Language and Tools Construct Reality

Orly (Kramash) Stettiner¹

¹Program in Science, Technology & Society (STS), Bar-Ilan University, Israel
orlyst@netvision.net.il

Abstract

When Christopher Langton first coined the term "artificial life" and organized the first conference of the nascent field in 1989, he envisioned that "We would like to build models that are so life-like that they cease to become models of life and become examples of life themselves." (Langton 1989). When Thomas Ray referred to his Tierra creatures four years later, he said: "These are not models of life, but independent instances of life" (Ray 1993).

Katherine Hayles, the American postmodern literary critic was startled by this vision and wondered how it was possible, in the late twentieth century, to "believe, or at least claim to believe, that computer codes are alive? And not only alive, but natural?" (Hayles 1996). The American philosopher of science Evelyn Fox Keller supported Hayles's view and generalized it into the linguistic domain (Keller 2002).

In this paper we briefly describe Hayles and Fox-Keller's claims, which will follow by an extended examination of how the usage of language, visualization and analysis tools have continued to construct and shape the field of ALife in the decade since their articles were published. Through this inspection, we suggest that the extensive usage of biological terminology and tools may give researchers a false impression regarding the validity and scientific significance of the experiments involving artificial simulated "organisms".

Introduction

Hayles's Narratives

Katherine Hayles concentrated on stories told about, and through, the evolving computer programs, where she identifies two levels of narratives, multilayered systems of metaphors: The textual-visual level and the strategic level.

The first level refers to the textual and visual representations of the environments, in which the artificial creatures develop. "In these representations", she said, "authorial intention, anthropomorphic interpretation, and the program's operations are so interwoven that it is impossible to separate them". Biomorphic naming such as "birth\death", "mother\daughter cell", "ancestor", "parasite", and their redefined interpretations reveal an intention of enabling a dynamic emergence of evolutionary processes within the computerized environment. The visual depiction of the code, in the form of sequenced images, imposes a feeling of real existence of living creatures within the computer. Ignoring the fact that the code is actually the organism and vice versa, the "creatures" gain a phenotypic

expression, a "body"- both visually (through specific shape, size and color) and verbally.

The Strategic level includes arguments and "political" strategies, used to position ALife as a research field within Theoretical Biology. New possible forms of "life-as-it-could-be" emerge spontaneously and evolve within the computer. Such attempts to synthesize life-like behaviors from simple rules and building blocks are claimed to complement the traditional analytic biological sciences that deal with "life-as-we-know-it". The essence of life, narrowed into complex logical forms, is claimed to be independent of medium. Thus, the ALife programs, considered alive themselves, are worth of studying as alternative evolutionary silicon-based pathways, which becomes a model for understanding the natural processes.

These narratives, as Hayles claimed, translate the operations of computer codes into dramatic biological analogues of a Darwinian struggle for survival and reproduction, the rise and fall of races, invented strategies for effective evolution, cooperation and competition.

Fox Keller's Lexicon

The American philosopher of science Evelyn Fox Keller supported Hayles's view, emphasizing the linguistic domain. Referring to the extensive biological lexicon, which ALife researchers developed for interpreting their models, she wrote, "it adds substantively to the sense of proximity to the real-life examples for which they aim" (Keller 2002, p. 277). Much like Hayles' note that "the organism is the code and the code is the organism", Fox Keller emphasized a persistent ambiguity and even identity of the words "genome", "program" and "body" of digital organisms, as a central agenda.

Fox Keller emphasized the increasingly narrowing and illusory gap between computers and organisms, as reflected by the terms "computational biology" and "biological computation", wondering if this convergence (both material and conceptual) can lead to an indistinguishable gap between the living and non-living. In her analysis, Fox Keller makes two major statements, regarding the achievements, in her view, of ALife simulations at the time of writing (2002).

(1) "The failure (at least to date) to generate the kinds of complex mechanisms observed in biological evolution weakens the claim of such models to enhance our understanding of life-as-we-know-it" (p. 281).

(2) "The models of A-Lifers have thus far failed to engage much interest among their biological colleagues" (p. 283).

At the end of this paper we shall examine the current validity of these two statements.

Over a decade has passed since the prognosis of Hayles and Fox-Keller. The discipline of Artificial Life and digital evolving organisms has become a mainstream, almost fully accepted and established field of research, at least as a method of studying evolution, with its own conferences and publications.

An investigation into some of the leading research reports in this field shows that the narrow trail, which Hayles and Fox Keller identified a decade ago, was followed by others and broadened into an actual highway. To demonstrate this trend we selected to focus on three levels, which we identify as central to the process: The linguistic level, the methodologies and analysis tools level and the human factor.

The Linguistic Level

At the linguistic level, we join and reinforce Hayles and Fox Keller, through the detection of an extensive and seemingly deliberate massive usage of re-defined biological concepts and anthropomorphisms of digital organisms, creating a vocabulary that becomes the mainstream convention.

The famous Swiss linguist Ferdinand de Saussure (1857-1913), the chief forerunner of structural linguistics, defined language as a collective product of social interactions. A meaning of a word is determined through differences, relative to other words or concepts. These differences structure our perception, and thus- language constructs our perceived reality. Therefore, naming digital objects after salient biological ones (like genotype or mutation) automatically categorizes them in the biological sphere.

During the 1930's, it was the American linguistic Benjamin-Lee Whorf, who advocated the idea that the structure of a language affects the way in which its speakers perceive and conceptualize their world and even their cognitive processes.

In the 1970's, philosophers and other scholars recognized the importance of language as a structuring agent, in what is known as 'the linguistic turn'. This turn began with the post-structuralism movement that followed De-Saussure and included influential theorists, such as Michel Foucault and Jacques Derrida. Language turned from being a tool for communicating messages into the message itself. Through daily usage of language we constitute our reality in an ongoing process of construction, modification and redefinition.

Genetic Algorithms became popular within computer science, with the early studies of Cellular Automata by John Holland (1975). Genetic Algorithms- by their very name and biological origin- naturally used terms adopted from Genetics: A population of randomly generated strings (called chromosomes) constitutes the genotype (or the genome), which encode candidate solutions (called individuals or phenotypes) to an optimization problem. In each generation, the fitness of every individual is evaluated, which serves as a basis for stochastic selection and modification (through recombination and possible random mutations) to form a new evolved population.

During the last decade, this preliminary basic lexicon was warmly adopted and widely extended to describe the digital

organisms, which live and evolve within machines, such as Avida.

The "Avidians", as the citizens of Avida are affectionately called, "can send messages to each other, produce and consume resources, and sense and change their environment's properties". They might be required to "communicate with neighboring organisms" (McKinley, Cheng et al. 2008). They "consume resources and generate by-products that can themselves serve as resources for other individuals" (Yedid, Ofria et al. 2008).

These descriptions and others attribute to these creatures traits and capabilities usually attributed to living creatures, specifically the ability to choose and to make knowledgeable independent decisions. This, in addition to the recurring statement that "Avida does not simulate evolution- it is an instance of evolution" (Pennock 2007), seems like a persistent attempt to designate the digital organisms a status of being an "instance of life" or even "an instance of an intelligent life".

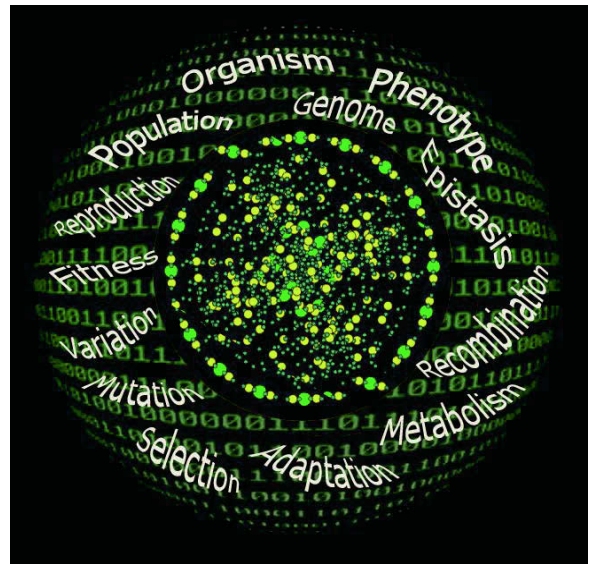


Figure 1: Schematic demonstration of language usage by digital organisms' researchers.

The basic vocabulary created for the original Genetic Algorithms has since expanded vastly to include every aspect of the biological research. It's not just organisms and population, genotype and phenotype, crossover and mutations, birth and death. It now also involves the Metabolism and Metabolic Pathways, Adaptation, Genetic drift and Fixation, Implicit and Explicit mutations, Sexual and a-sexual reproduction, single and multiple niche worlds, and even synergistic and antagonistic Epistasis.

Looking back over three decades ago, one might recall a brilliant critic paper written by an AI researcher, Drew McDermott, who attempted to ridicule some of the common Artificial Intelligence conceptual and linguistics trends, which he calls "mistakes" (McDermott 1976). One such trend was the use of wishful mnemonics: identifiers of programs or data structures named after grand concepts such as "Understand", "Resolve", "Think" or "Associate", describing their desired purpose but not their actual functionality. Referring to the "Is

a..." relation, which is commonly used by AI programmers, McDermott writes:

"Concepts borrowed from human language must shake off a lot of surface-structure dust before they become clear. "Is" is a complicated word, syntactically obscure. We use it with great facility, but we don't understand it well enough to appeal to it for clarification of anything. If we want to call attention to the "property inheritance" use, why not just say *INHERITS-INDICATORS*? Then, if we wish, we can prove from a completed system that this captures a large part of what "is a" means.... People reason circularly about concepts like "is a". Even if originally they were fully aware they were just naming *INHERITS-INDICATORS* with a short, friendly mnemonic, they later use the mnemonic to conclude things about "is a"." (McDermott 1976).

A similar adoption of concepts and naming occurs within ALife, which partly evolved as a separate field in AI, only here the concepts are not borrowed from natural language but rather from Biology. Having recognized the importance of language as a tool for reality construction, we find this usage of biological vocabulary a result of a deliberate affiliation and self-identification on the part of ALife researchers with the discipline of biology, rather than that of computer science. The selected language is becoming a message, it constitutes a modified reality, in which all organisms are alike, the living and the digital, all go through the same evolutionary processes, all become alive, reproduce and finally die, all struggle for survival, where the fittest has the best chances, all compete and sometimes collaborate, they are all actually the same.

Methodologies, Assessment and Analysis tools

A most fascinating aspect we identified in the process of adopting ALife into mainstream biology is the extensive usage of methodologies, assessment and analysis tools, widely adopted by A-Lifers from molecular biology, evolutionary biology and bioinformatics. We hereby demonstrate five of these tools:

(1) **Fitness Landscapes or Adaptive Landscapes** are concepts created in evolutionary biology, first introduced by Sewall Wright in 1932, to visualize the relationship between genotypes (or phenotypes) and their reproductive success, which is referred to as the "fitness" and visualized by the height at each point of the landscape.

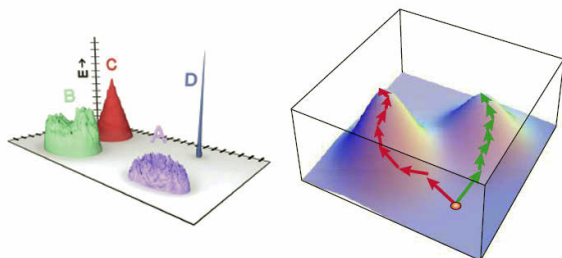


Figure 2: Fitness landscapes of living organisms, right (Elena and Lenski 2003) and of digital organisms, left (Hazen, Griffin et al. 2007)

The concept has gained importance in evolutionary optimization problems, followed by its adoption for describing the fitness function of digital organisms.

In Fig. 2, hypothetical fitness spaces are sketched (on the right) to describe the dynamics of evolutionary adaptation of bacteria and viruses (Elena and Lenski, 2003), while on the left, the fitness function space is shown, describing four classes of possible sequence solutions for digital populations evolving on Avida (Hazen et al., 2007).

(2) **Phylogenetics** is the study of evolutionary relations among biological species or populations, which is usually discovered through molecular sequencing and morphological data. Phylogenetic depth refers to the cumulative number of generations or lineages by which organisms differ from their common ancestor. A usage of this tool was presented, for example, in *Evolutionary Biology* magazine for the analysis of a large protein database, and in *"Systematic Biology"* journal to show the speciation of two groups of beetles

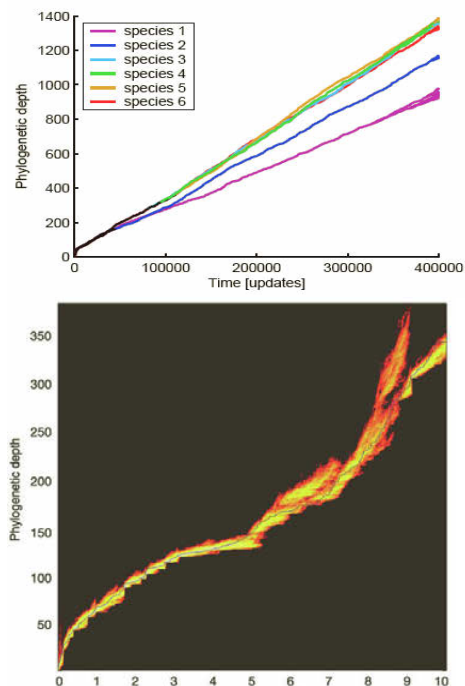


Figure 3: Phylogenetics analysis of digital organisms as presented by (Lenski et al, 2003) on the right and (Chow et al. 2004) on the left.

(Aphanarthrum and Coleobothrus).

Researchers of digital organisms adopted the concept and tool to present phylogenetic depth analysis, based on the number of generations in which an organism's genotype differs from its parents, where the colors indicate the relative abundance of genotypes at a specific depth (Lenski et al, 2003), (Chow et al. 2004).

(3) **Gene Expression Profiling Analysis** – also known as Functional Genomics Array, is a strategy developed by molecular biologists to describe genes functions and interactions. The profiling can present the measurement of the activity (expression) of thousands of genes at once, to create a global picture of cellular functionality under specific conditions. Each column usually represents a specific

experimental condition, whereas each row stands for a particular gene. A color-coded scale is used, where red generally represents expression greater than a certain reference, green is less than that reference, and gray or white is missing or excluded data (for example, Fig. 4A from (Glaser 2011), Fig. 4B from (Shoemaker et al. 2001).

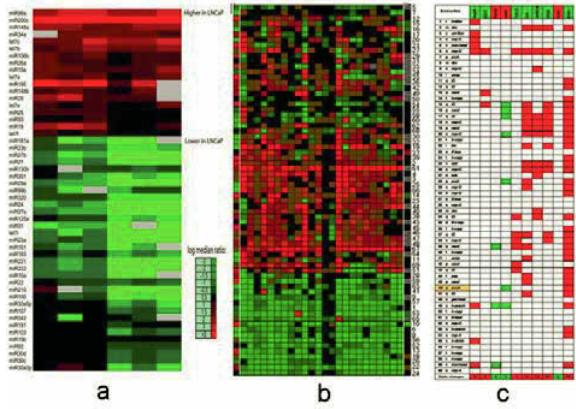


Figure 4: Gene expression profiling as used for living organisms: (a) by (Glaser 2011), (b) by (Shoemaker et al. 2001) and for digital organisms (c) by (Adami 2006).

The strategy was adopted by ALife researchers (e.g. (Adami 2006), as shown in Fig. 4C) to show the effect of knocking out individual computational instructions on the functions (which are actually the Genes) of a digital organism. The color codes look familiar: white indicates an unaffected function, red signifies a turned off (not expressed) function, whereas green signals functions that are turned on (the gene is expressed).

The overall resemblance between the original and adopted tool is remarkable and seems to be deliberate.

(4) **Sequence Alignment** is widely used in bioinformatics to arrange sequences of DNA or RNA (composed of nucleotides) or proteins (composed of amino acids), to identify regions of similarity, reflecting conserved regions, a consequence of functional, structural or evolutionary relationships between the sequences (e.g. (Brachner et al., 2012) (Karpinets et al., 2010), Fig.5A, 5B).

Researchers of digital organisms adopted this technique (Fig.5C, Adami 2000), to visually present the genome sequences of an entire Avida population, existing at a specific generation. The sequencing results demonstrate- at each genome site- the level of entropy, meaning- how variable or conserved this site is. Red sites are highly variable, whereas blue sites are conserved (having low entropy), as can be expected from the common practice in genomic sequencing of living organisms.

(5) **Gel Electrophoresis** is a well known procedure used in molecular biology to separate and sort a mixed population of DNA and RNA fragments by length, or proteins by electric charge, when they are made to move through a gel, usually made of PolyAcrylamide. DNA may be visualized using Ethidium Bromide which, when intercalated into DNA, fluoresce under ultraviolet light.

Recently, an attempt to sort and recognize digital viruses (referred to as "malware" or "cyber organisms") were made, using the same adopted procedure (Jaenisch 2010). The digital viruses, considered to be a collection of polypeptides, forming information-bearing protein structures, were analyzed using a mathematical analog to the 2-Dimensional PolyAcrylamide gel electrophoresis process, where again- the colors and general appearance seem familiar (Fig. 6).

These examples of methodologies and analysis tools, which originated in mainstream fields (such as molecular or

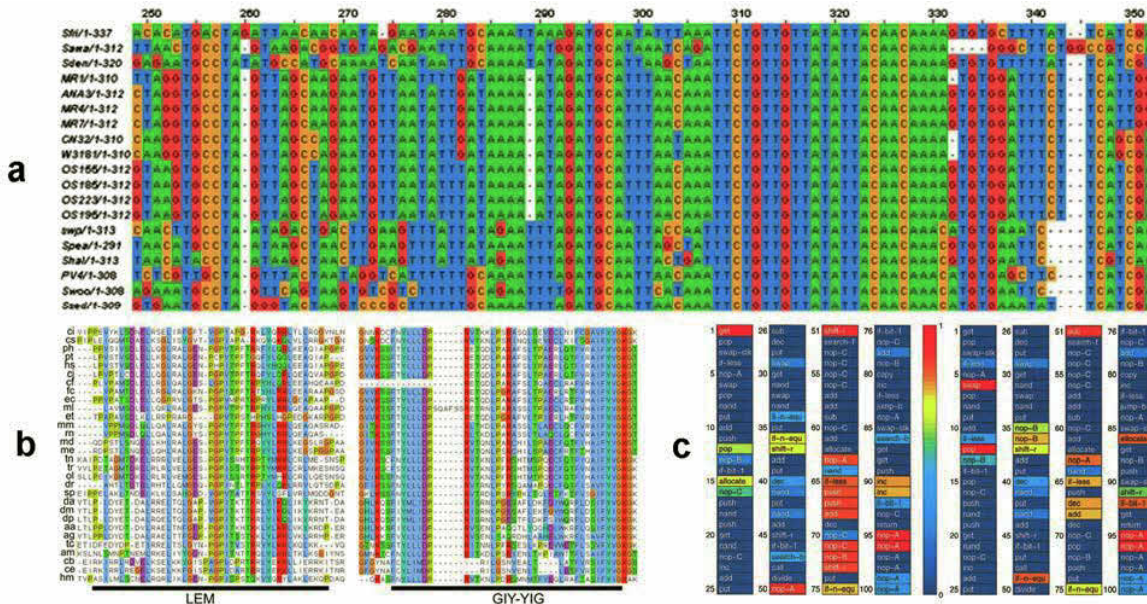


Figure 5: Examples of sequence alignments used for living organisms ((a) From (Karpinets et al., 2010), (b) From (Brachner et al., 2012)), and for digital organisms, (c) from (Adami 2000).

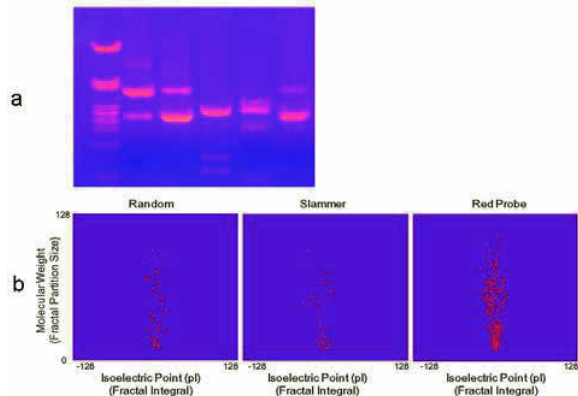


Figure 6: Gel electrophoresis examples as used for (a) proteins, from Wikimedia Commons http://commons.wikimedia.org/wiki/File:Gel_electrophoresis_2.jpg and (b) digital viruses (Jaenisch 2010).

evolutionary biology), demonstrate a remarkable visual and conceptual resemblance to the originals. A layman reader of an academic paper or, more importantly, a wet biologist or theoretical evolutionist, seeing the results of a gene expression profile or an electrophoresis slide, might only see the expected familiar visual structure and colors, and will not necessarily pause to make the distinction between the analyzed objects. This obviously results in the intensification of the illusion of similarity and even identity between the living and the artificial species.

Human factor, publications and citations

Finally, we identify the human factor as having a major influential effect on the increased acceptance of ALife into mainstream biology. The discipline today is composed of a mixture of computer scientists, engineers and leading biologists. Names like Richard Lenski, a distinguished professor of Microbial Ecology, who is well known for his long-term *E. coli* evolution experiments, certainly provide academic credibility, when conducting evolution experiments on other kinds of organisms, either in an actual petri-dish or in a so called "virtual-petri-dish". Christopher Adami, a professor of Microbiology and Molecular Genetics, can easily share the methodologies and assessment tools, used to analyze the genetic traits of Avidians with other Molecular Genetics researchers. As a result, the digital organisms' research has become a mainstream instrument that carries the knowledge and methodologies of related biological fields into the computerized artificial domain. The analogy between *E. coli* and digital organism is easily made, based on their visually comparable circular genome (Fig. 7).

Consequently, we identified an increased reliance of digital evolution studies on core-biology publications and lab reports, and vice versa. Relationships between digital organisms are compared to those reported for long-term bacterial experiments (e.g. (Yedid, Ofria et al. 2008)). ALife papers regularly quote dozens of biological articles, written by Zoologists, evolutionary biologists and molecular biologists,

which- to the reader- gives a definite impression of being part of a much larger corpus of scientific biological research. Side by side, an increasing number of so-called "pure" biologists conduct experiments on digital organisms, quoting such digital findings and support their real-life conclusions by these results. Digital and Living organisms' based experiments are quoted side-by-side, as reliable sources of theories that explain, for example, the evolution of complex traits. More and more biologists seem to believe that digital organisms' evolutionary mechanism can produce new scientific knowledge and explanation. Such papers are widely found in publications such as: *Journal of Evolutionary Biology*, *Genome Research*, *Journal of Molecular Evolution*, *Cellular and Molecular Life Science*, and others.

One can read, for example, a quote from a group of Developmental Biology Researchers saying:

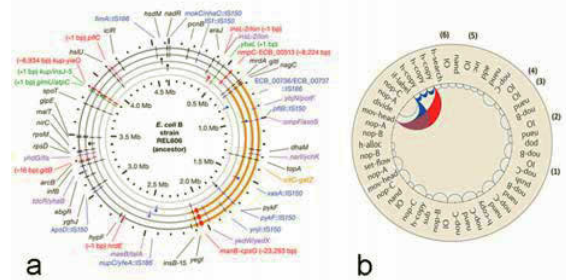


Figure 7: A Scheme of (a) *E. coli* genome (Barrick, et al. 2009) and (b) of and Avidian genome (Adami 2006)

"Although the digital creatures in these experiments embody an abstract view of life... it is highly informative to observe, in a laboratory microcosm, the ability of ever-more fit organisms to emerge while less fit variants disappear from the population..." (McAdams, Srinivasan et al. 2004)

Or a group of biologists conducting Zoological research that writes:

"In so far as the evolution of complex biological structure, under the influence of natural selection, can be expected to follow the same logical rules as these computer simulations [of (Lenski, Ofria et al. 2003). O.S], then the correlated progression model is corroborated." (Kemp 2006).

These developments can easily fix the notion of resemblance and equal academic credibility between the digital computer simulations and in-vivo laboratory experiments of evolving organisms.

Historically, such cross-discipline personal influence can be detected. A well-known example is Erwin Schrödinger's 1944 monograph "*What is Life? The Physical Aspect of the Living Cell*", a book considered one of the most influential scientific books in the twentieth century. Schrödinger, as a Nobel Laureate Physicist had its greatest influence on physicists, who were inspired to emigrate to the field of biology. But, according to the biographer Walter Moore, the book encouraged biologists to think more rigorously, in terms of mathematically formulated and physically testable models, bringing physics to the attention of biologists as well as biology to the attention of physicists (Ceccarelli 2001, p. 63).

Summary and conclusions

Katherine Hayles said, in her previously mentioned work, that "ALife researchers joke, that ALife is a solution in search for a problem". This is no more.

In 2003, Chris Adami was quoted to have said, when he referred to Avida:

"I wanted this digital life system to be an experimental system just like, let's say, Rich Lenski and *E. coli* bacteria...." (O'Neill, B., 2003).

It seems today that this vision has been fully realized. Having revisited Katherine Hayles and Evelyn Fox-Keller with an updated inspection, we feel strongly that their initial feeling was correct.

In contrast to Fox-Keller's belief, that artificial models will not enhance our understanding of life-as-we-know-it, and her statement that these models "failed to engage much interest among their biological colleagues", we witness biology experts today that quote ALife experiments and treat them as authorized and reliable sources of information on questions that used to be purely in-vivo or in-vitro issues. At the same time, ALifers refer to actual biological dilemma and theories, which they suggest to contribute to, with their powerful and effective computerized tools and through the vivacious, energetic, struggling and reproductive citizens of Avida.

A clear trend of building narratives around digital organisms was identified, which is primarily based on a wide and thorough linguistic biology-based lexicon. Analysis tools and methodologies have been adopted from molecular and evolutionary biology and seamlessly converted into the digital domain, to produce an illusionary feeling that one actually reads and talks about similar entities, with comparable credibility and analogous results. Significant support has been added to this conceptual construction during recent years by leading researchers, who conduct parallel experiments on both living and digital organisms, producing publications that enhance the validity and scientific significance of experiments involving artificial simulated "organisms" and consequently render the narrow gap unnoticeable.

References

- Adami, C., Ofria, C., & Collier, C. T. (2000). Evolution of biological complexity. *PNAS (Proc. National Academy of Science)*, 97(9), 4463-4468.
- Adami, C. (2006). Digital genetics: unravelling the genetic basis of evolution. *Nature Reviews Genetics*, 7, 109-118.
- Barrick, J. E., et al. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461(7268), 1243-1247.
- Brachner, A. et al. (2012). The endonuclease Ankle1 requires its LEM and GIY-YIG motifs for DNA cleavage in vivo. *J. Cell Science*, 125:1048-1057. JCS098392 Supplementary Material.
- Ceccarelli, L. (2001). *Shaping Science with Rhetoric: The Cases of Dobzhanski, Schrodinger and Wilson*. Chicago and London, The University of Chicago Press.
- Glaser, V. (2011). Sequencing Alters Medical Genetics. *Genetic Engineering & Biotechnology News*, 31(21):34-36.
- Harris, R. 1988. *Language, Saussure and Wittgenstein: How to play games with words*. Routledge.
- Hayles, K. (1996). Narratives of Artificial Life. *Futurenatural: Nature, Science, Culture*. M. M. George Robertson, Lisa Tickner, Jon Bird, Barry Curtis and Tim Putnam (editors). London, Routledge: 147-164.
- Hazen, R. M., P. L. Griffin, et al. (2007). Functional information and the emergence of biocomplexity. *Proc. National Academy of Science*, 104(Suppl. 1): 8574-8581.
- Jaenisch, H. (2010). Digital Virus Recognition. *SPIE Newsroom*. DOI 10.1117/2.1201005.002933.
- Karpinets, T.V. et al. (2010). Shewanella knowledgebase: integration of the experimental data and computational predictions suggests a biological role for transcription of intergenic regions. *Database*: 2010, Article ID baq012, doi:10.1093/database/baq012.
- Keller, E. F. (2002). *Making Sense of Life: Explaining Biological Development with Models, Metaphors and Machines*. Cambridge, Massachusetts, Harvard University Press.
- Kemp, T. S. (2006). The origin of mammalian endothermy: a paradigm for the evolution of complex biological structure. *Zoological Journal of the Linnean Society*, 147: 473-488.
- Langton, C. G. (1989). Artificial Life- the proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems. *Artificial Life*, Redwood City, CA, Addison-Wesley
- Lenski, E. R., C. Ofria, et al. (2003). The evolutionary origin of complex features. *Nature*, 423: 139-144.
- McAdams, H. H., B. Srinivasan, et al. (2004). The evolution of genetic regulatory systems in bacteria. *Nat Rev Genet*, 5(3): 169-178.
- McDermott, D. (1976). "Artificial Intelligence Meets Natural Stupidity." SIGART Newsletter 57: 4-9.
- McKinley, P., B. H. C. Cheng, et al. (2008). Harnessing Digital Evolution. *IEEE Computer*, 41(1): 54-63.
- O'Neill, B. (2003). Digital Evolution. *PloS Biology*, 1(1): 11-14.
- Shoemaker, D. et al. (2011). Experimental annotation of the human genome using microarray technology. *Nature*, 409: 922-927.