Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 29, Issue Number 2, (2011)

Vocabulary of Definitions of Life Suggests a Definition

http://www.jbsdonline.com

Edward N. Trifonov[1,2]

[1]Genome Diversity Center, Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel
[2]Department of Functional Genomics and Proteomics, Faculty of Science, Masaryk University, Kotlarska 2, CZ-61137 Brno, Czech Republic

Abstract

Analysis of the vocabulary of 123 tabulated definitions of life reveals nine groups of defining terms (*definientia*) of which the groups *(self-)reproduction* and *evolution (variation)* appear as the minimal set for a concise and inclusive definition: Life is self-reproduction with variations.

Key words: Consensus; *Definientia*; Evolution; Origin of life; Self-reproduction; Variations; Vocabulary.

Over 100 of definitions of life exist today (1, 2) – learned opinions each one of which is, or has been in the past, defended not without a reason though generally met with skepticism. The skepticism is multiplied by the above number, leaving almost no chance for new formulations which, however, continue to appear. An excellent overview of the current status of the problem is given by a special issue of the journal "Origins of Life and Evolution of Biospheres" (3). Sixteen papers of this issue, expert opinions, display a variety of philosophical and historical aspects of defining life, and inevitable limitations of about every approach and view point.

The definitions are more than often in conflict with one another. Undeniably, however, most of them do have a point, one or another or several, and common sense suggests that, probably, one could arrive to a consensus, if only the authors, some two centuries apart from one another, could be brought together. One thing, however, can be done – sort of voting in absentia – asking which terms in the definitions are the most frequent and, thus, perhaps, reflecting the most important points shared by many. Such analysis is offered below, revealing those most frequent terms that may be used for tentative formulation of the consensus.

In the Table I the vocabulary of words used in 60-definition set of Barbieri (1) and 90-definition collection of Popa (2) is presented. The non-redundant total size of two collections is 123 definitions. All words of 3 or more letters are taken for the survey, excluding connective ones ("the", "and", "that", *etc*.). The words that appear more than 4 times in the collections are presented in the Table I (the full list is available by request). The "life", as *definiendum*, is at the top of the list. Inspection of the list reveals that amongst frequent words the ones closely related to, *e.g*., "life" group (such as "living", "alive") appear as well. This suggests combination of various words in groups by their common meaning. The Table II displays several such groups, topmost by their scores. Words of each group are present in at least 30% of the definitions analyzed. The groups of smaller size (not shown) contain, essentially, only words with the same root (*e.g*., definition, defined, defining, *etc*.).

Phone: +972 4 828 8096
Fax: +972 4 824 6554
E-mail: trifonov@research.haifa.ac.il

Thus, the consensus of the life definition patched from these nine *definientia* would be: Life is [System, Matter, Chemical (Metabolism), Complexity (Information), (Self-)Reproduction, Evolution (Change), Environment, Energy, Ability,…] where the square brackets correspond to some compact expression containing the words listed within. For example, one possibility is:

Life is *metabolizing material informational system* with *ability* of *self-reproduction* with *changes (evolution)*, which requires *energy* and suitable *environment.* [1]

Since the analysis described, inevitably, is not free from some arbitrariness in assignment of the words to this or another group, it would be desirable to compare it to a similar study conducted by an independent laboratory. The work of Kompanichenko (4) offers such independent analysis. In that work the definitions of life are taken from the collection of papers and definitions (5) that preceded the publication of Popa (2). Kompanichenko considered definitions given by 63 authors from the book of Palyi *et al.* (5), and extracted from the definitions 19 major "unique fundamental properties of biological systems". Instead of words the notions of the properties have been used, often expressed by several words, differently by different authors. The largest category, according to Kompanichenko, is "Capable of evolution including the increase of complexity, hierarchy and the display of self-perfecting logic" (close to "Evolution" in our list) – 32 authors. The fundamental property "Capability for self-reproduction" ("Reproduction" in our list) has been mentioned by 27 authors. Another fundamental property, "Capability for self-replication" (11 authors) also would go to "Reproduction" group, 38 authors together. Note that according to the "notion count" of Kompanichenko our categories "Evolution" and "Reproduction" appear at the top of the list, while by "word count" (this work) they rank more modestly. Then follow "Performance and control of metabolism, including autocatalysis, cyclic chemical processes, feedback loops, and active transport" ("Chemical" of our list) – 25 authors, "Ability to extract (free) energy and matter from the environment" ("Energy", "Matter", "Environment", and "Ability" in our list) – 16 authors, and "Capacity to accumulate, re-organize… and transmit genetic information" ("Complexity"/"Information") in our list) – 13 authors. Remaining 13 properties considered by Kompanichenko are mentioned by smaller number of authors (1 to 7) and are not reflected in our list of major terms. This approach, obviously, is not free from subjective assignments as well. It is remarkable, however, that eight of nine categories of our list of major terms are also at the top of the list of Kompanichenko.

Table I

List of most frequent words in the definitions of life.

Life	123	Organic	11	Internal	7	Capacity	5
Living	47	Alive	10	Replication	7	Different	5
System	43	Evolution	10	Being	6	Force	5
Matter	25	Materials	10	Change	6	Form	5
Systems	22	Reproduction	10	Characteristics	6	Functional	5
Environment	20	Existence	9	Entity	6	Highly	5
Energy	18	Defined	8	External	6	More	5
Chemical	17	Growth	8	Means	6	Mutation	5
Process	15	Information	8	Molecules	6	Necessary	5
Metabolism	14	Open	8	One	6	Network	5
Organism	14	Processes	8	Order	6	Objects	5
Organization	14	Properties	8	Organisms	6	Only	5
Complexity	13	Property	8	State	6	Organized	5
Ability	12	Reproduce	8	Things	6	Reactions	5
Itself	12	Through	8	Time	6	Self-reproduction	5
Able	11	Complex	7	Way	6	Some	5
Capable	11	Evolve	7	Based	5	Three	5
Definition	11	Genetic	7	Biological	5		

Table II

Groups of words with similar meaning.

LIFE	123	**COMPLEXITY**	13
living	47	information	8
alive	10	complex	7
being	6	other related words	46
biological	5	Sum	**74**
other related words	8	**REPRODUCTION**	10
Sum	**199**	reproduce	8
SYSTEM	43	replication	7
systems	22	self-reproduction	5
organization	14	other related words	33
organism	14	Sum	**63**
order	6	**EVOLUTION**	10
organisms	6	evolve	7
network	5	change	6
organized	5	mutation	5
other related words	40	other related words	20
Sum	**155**	Sum	**48**
MATTER	25	**ENVIRONMENT**	20
organic	11	external	6
materials	10	other related words	15
molecules	6	Sum	**41**
other related words	36	**ENERGY**	18
Sum	**88**	force	5
CHEMICAL	17	other related words	17
process	15	Sum	**40**
metabolism	14	**ABILITY**	12
processes	8	able	11
reactions	5	capable	11
other related words	26	capacity	5
Sum	**85**	other related words	1
		Sum	**40**

Although the extract [1] above may already appear as a reasonable definition, one, of course, would like to have it more concise, and desirably, containing components that are both necessary and sufficient, either alone or in combination. The possible shorter definitions would be, of course, subject of a thorough evaluation, inviting new discussions of the definition problem, but on a new basis of limited number of relevant terms. Below one rather plausible reduction is suggested. First, some of these consensus terms implicitly involve others. For example, existence of *metabolism* implies both *energy* and *material* supply which also represent *environment. Self-reproduction (replication)* appears to be the most inclusive term of the nine groups above, as it implies *metabolism* and *system* as well. That is, if the self-reproduction is going on, it can proceed only on condition that metabolism, system, energy and material supply are also in place. The *complexity (information)* can be considered also as product of *self-reproduction* with *changes (evolution),* on the evolutionary route from simple to complex. We are, thus, left with two independent notions: *self-reproduction* and *changes (evolution)*. None of these two implies another one. They, actually, exclude one another, as self-reproduction is exact copying, no changes, while changes can not relate to exact copying. These two notions can be combined in a third one: an *almost exact self-reproduction* or *self-reproduction with variations,* suggesting, thus, a tentative minimalistic definition.

Evolution (and natural selection) means changing inheritance, that is, causing variations in self-reproduction. As it has been said by Darwin (6): **"… if *variations* useful to any organic being ever do occur, assuredly individuals thus characterized will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance, these will tend to *produce offspring***

***similarly characterized*"** (6, Chapter 4, italics by ENT). The definition of life based on only two terms extracted from the vocabulary of definitions, and consistent with Darwin's views would be, thus:

"Life is *self-reproduction* with *variations*". [2]

Here the tandem *self-reproduction with variations* should be considered as one indivisible term of very clear Darwinian meaning. Of 123 different definitions only 18 contain this pair of *definientia,* in combination with other defining terms of the Table II. The most succinct among them is one by Oparin (7):

"Any system capable of *replication* and *mutation* is alive".

The vocabulary approach implemented in this work is conceptually close to the "Principal Component Analysis" (8, 9) which is applicable to large ensembles of quantitative data. The data are reduced to several independent ("orthogonal") components, and the major principal component is extracted that covers most of the data. Similarly, our vocabulary of definitions is reduced to several groups of words with different meaning. The same procedure of extraction of a single "principal component" has been used in derivation of consensus temporal order of engagement of amino acids in early evolution (10, 11). The order has been established as a consensus of a large number of rather different chronologies suggested by various authors. As this order correlates well with thermostability of respective codon-anticodon pairs, the reconstruction of the Evolutionary Chart of Codons became possible. From the properties of the Chart several important features of the earliest stages of molecular evolution have been predicted, and confirmed by sequence analyses (12, 13). The earliest steps of the evolution of the codons also suggested two major stages in the origin of life – self-reproduction (exact replication of the ideal RNA duplex in the above theory, one strand of which is repeating triplet GCC_n, while another strand is complementary GGC_n), and variations (appearance of point-mutated versions of GCC and GGC in the subsequent replications). That ended logically in the definition of life – *self-reproduction with variations* (14), identical to above [2]. In its earlier version – *almost precise replication* – it appeared in the collection of definitions of life gathered by Barbieri (1). (That formula has been excluded from the analysis above and did not enter the vocabulary).

According to this model-based definition, any experimental work involving the GCC_n* GGC_n replicator would border the life-nonlife transition. The definition of life, thus, is naturally required for the exploration, at least as a practical guide in the research (15). Thus, it is not purely philosophical and historical matter anymore. This was one of the motivations for the linguistic analysis described in the paper. The derived definition [2] is minimalistic both by *definientia* involved and, independently, by structurally minimal size of the presumed earliest replicator to which the definition fully applies (14).

One unforeseen property of the minimalistic definition is its generality. It can be considered as applicable not just to "earthly" life but to any forms of life imagination may offer, like extraterrestrial life, alternative chemistry forms, computer models, and abstract forms. It suggests a unique common basis for the variety of lives: all is life that copies itself and changes.

One important question to address: is the minimalistic definition both necessary and sufficient? Even most primitive forms of observable life are still too complex, to claim that they can be reduced to the above simple formula. The applicability of the definition can only be tested on much simpler artificial life-like models which one day will, hopefully, be designed and brought to life, by providing artificially

produced necessary ingredients. This day is not far away. A self-catalytic generation ("cross-replication") of plus-strand and minus-strand ribozymes from their constituent 14-mer and 52-mer RNA chains has been recently accomplished (16). The system also allows to introduce and to monitor evolutionary changes in the ribozymes. It does not include, though, the elementary template polymerization steps one would expect the simplest self-reproducing system to have. The experimental chemical template polymerization has been intensively studied during last two decades (for most recent review see (17)). An efficient complementary primer extension on C_{15} template has been achieved in "protocells" by using chemical RNA analog (18). No evolutionary changes have been observed so far in the system. Yet simpler setup, with oligo-riboA in aqueous solution has been developed, in which the chain of RNA could elongate indefinitely, apparently, due to formation of "complementary" contacts between polyA chains (19). In this work similar extension of oligoG ligated to oligoC has been observed as well, with incorporation of complementary Gs. "Creation" of a bacterial cell with chemically synthesized genome (20) has to be mentioned as another case of a system at the border life-non-life. This is, actually, a very large scale bacterial transformation where the transforming DNA has been, indeed, chemically synthesized according to natural design – previously fully sequenced genome, with some changes. The synthesized genome did replicate many rounds. However, it should be considered as very much assisted replication as it was provided with an initial natural cytoplasm. No evolutionary changes in the design have been monitored. Nearly fully artificial life with the properties described by the above minimalistic definition has been created many years earlier, by Sol Spiegelman – his famous monster of replicating degenerate products of mutating Q-beta RNA. The monster RNA versions have practically lost any sequence similarity to their viral ancestor, in the process of ingenious evolutionary game (21). This has been an assisted replication as well, since the experiments have been performed in presence of natural replicase. However, this protocol is, probably, the most promising starting model for eventual design of simplest truly artificial life, since for modern peptide chemistry the synthesis of an active analog of the replicase, perhaps, is as plausible as the chemically synthesized genome.

Acknowledgements

Author is grateful to anonymous reviewers for thoughtful comments, criticism and suggestions towards improvement of the paper.

References

1. M. Barbieri. *The Organic Codes. An introduction to Semantic Biology.* Cambridge University Press, Cambridge (2003).
2. R. Popa. In *Between Necessity and Probability: Searching for the Definition and Origin of Life.* Series: Adv Astrobiol Biogeophys, Springer, NY, pp. 197-205 (2004).
3. J. Gayon, C. Malaterre, M. Morange, F. Raulin-Cerceau, and S. Tirard (guest Eds.). Special Issue: Definitions of life. *Origins Life Evol Biospheres 40*, 119-244 (2010).
4. V. N. Kompanichenko. *Int J Astrobiol 7*, 27-46 (2008).
5. G. Palyi, C. Zucci, and L. Caglioti (Eds.). *Fundamentals of Life*. Elsevier SAS, Paris (2002).
6. C. Darwin. *Origin of species*, John Murray, London (1859).
7. A. I. Oparin. as referred to in (2) (1961).
8. K. Pearson. *Philosophical Magazine 2*, 559-572 (1901).
9. I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics, 2nd ed., Springer, NY (2002).
10. E. N. Trifonov. *Gene 261*, 139-151 (2000).
11. E. N. Trifonov. *J Biomol Struct Dyn 22*, 1-11 (2004).
12. E. N. Trifonov. *Isr J Ecol Evol 52*, 375-387 (2006).
13. E. N. Trifonov. *J Cosmology 10*, 3374-3380 (2010).
14. E. N. Trifonov. *Res Microbiol 160*, 481-486 (2009).
15. S. Pino, E. N. Trifonov, and E. Di Mauro. *Genomics, Proteomics and Bioinformatics 9*, 7-14 (2011).
16. T. A. Lincoln and G. F. Joyce. *Science 323*, 1229-1232 (2009).

17. J. S. Schrum, T. F. Zhu, and J. W. Szostak. *Cold Spring Harb Perspect Biol 2*, a002212 (2010).
18. S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco, and J. W. Szostak. *Nature 454*, 122-125 (2008).
19. G. Costanzo, S. Pino, F. Ciciriello, and E. Di Mauro. *J Biol Chem 284*, 33206-33216 (2009).
20. D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison III, H. O. Smith, and J. C. Venter. *Science 329*, 52-56 (2010).
21. D. R. Mills, F. R. Kramer, C. Dobkin, T. Nishihara, and S. Spiegelman. *Proc Natl Acad Sci USA 72*, 4252-4256 (1975).

Date Received: March 17, 2011

Communicated by the Editor Ramaswamy H. Sarma

Appendix

Full list of the words making groups of the Table II.

LIFE	123
living	47
alive	10
being	6
biological	5
animate	4
beings	3
animated	1
Sum	**199**
SYSTEM	43
systems	22
organization	14
organism	14
order	6
organisms	6
network	5
organized	5
assembly	3
building	3
components	3
composed	3
ensemble	3
aggregates	2
automata	2
consists	2
ordered	2
arrangement	1
automaton	1
automatons	1
biosystem	1
built	1
consisting	1
contain	1
containing	1
ensembles	1
multilevel	1
networks	1
orderly	1
organizational	1
organize	1
rearrangement	1
self-organization	1
self-organized	1
Sum	**155**
MATTER	25
organic	11
materials	10
molecules	6
material	4
compounds	3
nucleic	3
polymers	3
proteins	2
acids	2
fluid	2
substance	2
acid	1
aqueous	1
bioelements	1
carbon	1
carbon-based	1
molecule	1
monomers	1
oligosaccharides	1
polymer	1
polypeptides	1
polysaccharides	1
protein	1
proteinaceous	1
substances	1
water	1
Sum	**88**
CHEMICAL	17
process	15
metabolism	14
processes	8
reactions	5
molecular	4
production	3
metabolic	2
metabolize	2
chemistry	2
produces	2
decay	1
degradable	1
degradation	1
exchange	1
precursors	1
processing	1
produced	1
regeneration	1
reparation	1
synthesis	1
synthesize	1
Sum	**85**
COMPLEXITY	13
information	8
complex	7
code	4
entropy	3
knowledge	3
patterns	3
communicate	2
molecular-informational	2
pattern	2
program	2
reading	2
algorithmic	1
complicated	1
computational	1
digital	1
feedback	1
feedbacks	1
feedback-loops	1
informational	1
informationally	1
informationally-controlled	1
information-storage	1
instructional	1
instructions	1
low-entropy	1
maximally-complex	1
message	1
program-controlled	1
programs	1
self-correction	1
self-instruction	1
self-reading	1

(*Continued*)

Appendix (Continued)

sequence	1	mutability	1
signal	1	modifies	1
Sum	**74**	mutandis	1
REPRODUCTION	10	mutations	1
reproduce	8	mutatis	1
replication	7	variant	1
self-reproduction	5	Sum	**48**
self-replication	3	**ENVIRONMENT**	20
autopoiesis	2	external	6
autopoietic	2	conditions	3
multiplication	2	surroundings	3
proliferation	2	biosphere	1
replicate	2	condition	1
self-replicating	2	conditionally	1
self-reproduce	2	conditioned	1
self-reproducing	2	environmental	1
copies	1	environments	1
copying	1	medium	1
perpetuate	1	microenvironment	1
perpetuated	1	supply	1
proliferating	1	Sum	**41**
recreate	1	**ENERGY**	18
reproduces	1	force	5
reproducibility	1	thermodynamic	3
reproducing	1	engine	2
self-duplication	1	forces	2
self-generating	1	power	2
self-generation	1	powers	2
self-perpetuating	1	energetically	1
self-producing	1	energy-dependent	1
Sum	**63**	energies	1
EVOLUTION	10	engines	1
evolve	7	thermodynamical	1
change	6	thermodynamics	1
mutation	5	Sum	**40**
changes	4	**ABILITY**	12
evolutionary	2	able	11
mutate	2	capable	11
variation	2	capacity	5
errors	2	capacities	1
evolved	1	Sum	**40**
evolves	1		