

HYPOTHESIS-BASED, COMMUNITY-SUPPORTED ORGANIZATION OF SCIENTIFIC INFORMATION IN ASTROBIOLOGY.

A. Pohorille¹ and R. Keller², ¹ Exobiology Branch, MS 239-4, NASA-Ames Research Center, Moffett Field, CA 94035, Andrew.Pohorille@nasa.gov, ² Intelligent Systems Division, MS 269-2, NASA-Ames Research Center, Moffett Field, CA 94035, Richard.Keller@nasa.gov.

Introduction: Astrobiology is a scientific discipline devoted to the study of the origins, evolution distribution and future of life in the universe. It often deals with objects and processes that are remote in time and space and, therefore, cannot be subjected to direct experimentation. Instead, their nature has to be inferred from indirect observational, experimental or theoretical evidence. This evidence is drawn from a number of different fields, such as molecular biology, ecology, planetary science, astronomy, chemistry, geology and information science. As a consequence, astrobiology is inherently interdisciplinary, strongly driven by hypotheses that cannot be unambiguously proven and, therefore, are often subjects of intense scientific debate. This poses a number of challenges for the scientific community. How to remain well informed about different lines of arguments and evidence regarding hypotheses in the field, which could be often very diverse, exceeding expertise of a single scientist? How to disseminate new research results such that their essence is understood and appreciated by a broad range of scientists with different backgrounds? How to transfer broad knowledge in the field in an unbiased way to the new generation of astrobiologists? These issues are essential for effective communication, collaboration and education, which in turn are required for long-term progress in the field.

Hypothesis Browser: We have addressed the challenges described in the Introduction through developing a novel, web-based, hypothesis-driven, community-supported system for organizing and disseminating scientific information. Our prototype system consists of three components: (1) a database of scientific information derived from scientific papers and indexed *via* the hypotheses it supports or contradicts, (2) arguments for and against each hypothesis organized and presented to the user using a web browser interface, and (3) an “ecology of hypotheses” created by connecting each hypothesis to other, relevant hypotheses. The underlying idea is that arguments supporting or contradicting scientific hypotheses are the most frequently sought scientific information and, therefore, presenting them in a structured environment is much more useful to researchers and students at every level than literature searches based on keywords or other similar criteria.

To represent information of interest we have developed an ontology that consists of five concepts: (i) hypothesis, (ii) argument category, (iii) argument, (iv) evidence and (v) research paper. These concepts and

the relationships between them, shown in Fig. 1, reflect the organization of information displayed in our Hypothesis Browser tool. Although designed independently, this ontology bears a relationship to the SWAN discourse ontology [1], a more complex ontology developed to support hypothesis tracking in the Alzheimer’s medical community.

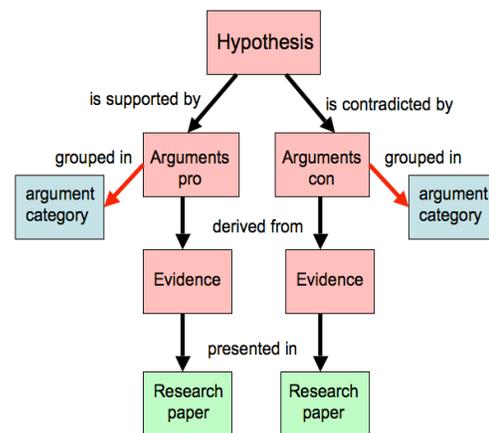


Fig. 1. Schematic representation of the hypothesis browser ontology

Upon entry to our Hypothesis Browser web site, a user is presented with a set of hypotheses under consideration by the scientific community. Browsing to view a specific hypothesis, the user sees a two-column display (Fig. 2) with the hypothesis on the left and a counter-hypothesis on the right. Underneath the hypothesis and counter-hypothesis, the system enumerates a list of arguments supporting and contradicting the hypothesis, as illustrated in Figs. 3. Conceptually similar arguments are grouped within an argument category. Each argument is derived from evidence presented in scientific papers. Users can enter comments on arguments and evidence, which stimulates scientific discourse. We have considered, but not implemented, various methods of notifying users when specific hypotheses or arguments of interest are commented upon by other users. These methods include email and RSS subscription, as well as simple notification upon login to the system.

Papers are indexed by their authors or other experts in the field, making the system a community effort. This indexing process is accomplished through a convenient web form. In addition to the full citation (Fig. 4), the author provides short statements describing

his/her evidence and argument (if applicable), and specifies supporting or contradicting links to hypotheses and related arguments. Currently, we have indexed approximately 100 papers in the origin of life area to seed and test the system.

System Status: With the initial design and development of the Hypothesis Browser complete, our next step is to begin limited user testing to validate the prototype system and gather suggestions for improvements. In response to these suggestions, we anticipate completing another round of design and development before releasing the system to an “early adopter” group of astrobiology scientists.

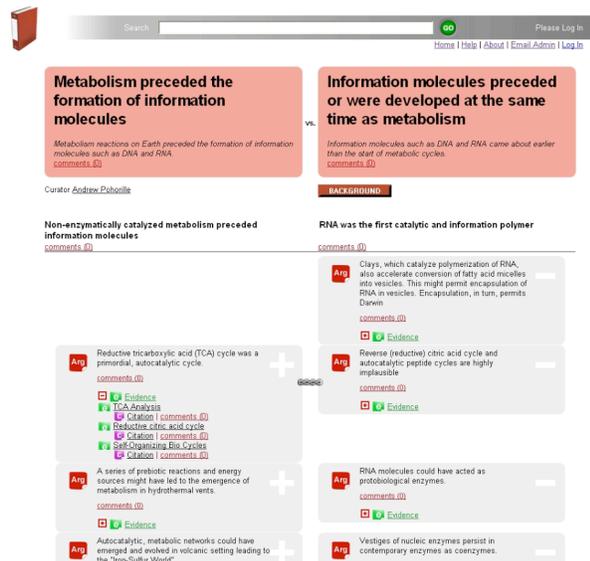


Fig. 2. Screen shot of the Hypothesis Browser.

Summary: The Hypothesis Browser described here is a digital resource that promotes disseminating scientific information in the field of astrobiology, facilitates scientific discourse and encourages interdisciplinary collaborations. In contrast to encyclopedic knowledge, the providers of information do not have to be objective. Even if a specific researcher supplies facts are biased for or against a given hypothesis, the overall body of evidence from the scientific community should provide a balanced viewpoint. In addition, the proposed structure is particularly suitable for educational purposes because it encourages students to think critically and learn to formulate and test hypotheses rather than accumulate isolated facts. Also, the system can be readily extended to other fields of science, medicine, engineering, as well as social and political arena.

Acknowledgements: This work was supported by a grant from Google.

References: [1] Ciccarese P. et al. (2008) *J. Bio-med. Inform.*, 41, 739-51.

Metabolism preceded the formation of information molecules

Metabolism reactions on Earth preceded the formation of information molecules such as DNA and RNA.

[comments \(0\)](#)

Curator [Andrew Pohorille](#)

Non-enzymatically catalyzed metabolism preceded information molecules

[comments \(0\)](#)

Arg Reductive tricarboxylic acid (TCA) cycle was a primordial, autocatalytic cycle. +

[comments \(0\)](#)

[Evidence](#)

[TCA Analysis](#)

[Citation | comments \(0\)](#)

[Reductive citric acid cycle](#)

[Citation | comments \(0\)](#)

[Self-Organizing Bio Cycles](#)

[Citation | comments \(0\)](#)

Arg A series of prebiotic reactions and energy sources might have led to the emergence of metabolism in hydrothermal vents. +

[comments \(0\)](#)

[Evidence](#)

Arg Autocatalytic, metabolic networks could have emerged and evolved in volcanic setting leading to the "Iron-Sulfur World". +

[comments \(0\)](#)

[Evidence](#)

Arg Energetically favorable, prebiotic reactions in mild aqueous conditions might have led to the emergence of metabolic and autocatalytic networks. +

[comments \(0\)](#)

[Evidence](#)

Fig. 3. Close-up of screen shot in Figure 2

CITATION

Universality in intermediary metabolism

Authors: Smith, E. and Morowitz, H. J.

Reference: *Proc. Natl. Acad. Sci. USA* (2004) 101, 13168-13173

Keywords: *reductive tricarboxylic acid cycle, autocatalytic cycle, primordial metabolism, universal metabolism*

We analyze the stoichiometry, energetics, and reaction concentration dependence of the reductive tricarboxylic acid (rTCA) cycle as a universal and possibly primordial metabolic core. The rTCA reaction sequence is a network-autocatalytic cycle along the relaxation pathway for redox couples in nonequilibrium reducing environments, which provides starting organic compounds for the synthesis of all major classes of biomolecules. The concentration dependence of its reactions suggests it as a precellular bulk process. We propose that rTCA is statistically favored among competing redox relaxation pathways under early-earth conditions and that this feature drove its emergence and also accounts for its evolutionary robustness and universality. The ability to enhance the rate of core reactions creates an energetic basis for selection of subsequent layers of biological complexity.

[EXPORT](#) [OK](#)

Fig. 4. Citation display