

Origins of Life Research: A Bibliometric Approach

Arsev Umur Aydinoglu & Zehra Taşkın

Abstract

This study explores the collaborative nature and interdisciplinarity of the origin(s) of life (OoL) research community. Although OoL research is one of the oldest topics in philosophy, religion, and science; to date there has been no review of the field utilizing bibliometric measures. A dataset of 5647 publications that are tagged as OoL, astrobiology, exobiology, and prebiotic chemistry is analyzed. The most prolific authors (Raulin, Ehrenfreund, McKay, Cleaves, Cockell, Lazcano, etc.), most cited scholars and their articles (Miller 1953, Gilbert 1986, Chyba & Sagan 1992, Wächtershäuser 1988, etc.), and popular journals (*Origins of Life and Evolution of Biospheres* and *Astrobiology*) for OoL research are identified. Moreover, interdisciplinary research conducted through research networks, institutions (NASA, Caltech, University of Arizona, University of Washington, CNRS, etc.), and keywords & concepts (astrobiology, life, Mars, amino acid, prebiotic chemistry, evolution, RNA) are explored.

Keywords: Origins of life, astrobiology, exobiology, prebiotic chemistry, bibliometrics

Introduction & Background

Origin of life research has been one of the oldest topics for philosophy, religion, and science; however, recently, the scientific community thinks that it would be better to call it “origins of life” (OoL) as we do not know how the transition from an abiotic Earth to a biotic Earth occurred and also whether there is more than one developmental path for life to emerge or whether some early forms of life have already gone extinct leaving no record (Scharf et al., 2015; Davies, 2007; Hartman, 1998). “Origins of life” may describe the process better and in this study we will use this term.

Although much has been accomplished in OoL research, the knowledge generated has often been hidden in isolated domains and progress has often been fragmented. The interdisciplinary nature of the phenomenon requires larger and better integrated approach by researchers recruited from different domains yet working collaboratively. Research networks such as the NASA Astrobiology Institute’s (NAI) teams, the Simons Collaboration on the Origins of Life (SCOL), and the ELSI Origins Network (EON) aim to build a cohesive and lasting global network of researchers working on the OoL. Nevertheless, a survey of the state of the art of OoL research will be useful for the coordination and success of such initiatives.

Since the early twentieth century, there have been scientific studies investigating how life originated. In the 1920s Oparin (1952) theorized about early Earth conditions and the OoL but did not conduct experiments. Haldane introduced a similar modern concept for abiogenesis and their contribution is known as the Oparin-Haldane Hypothesis (Lazcano, 2010; Mesler & Cleaves, 2015). Schrödinger (1944) introduced the idea of an “aperiodic crystal” to explain genetic information in his famous book entitled “*What is Life?*” that inspired a generation of scientists. Since the famous Urey-Miller experiment, the primordial soup concept has become a cornerstone in the field and the number of studies has increased. Between the 1950s and 1990, there were important studies by scholars such as Miller, Sagan, Lederberg, Calvin, Horowitz, Fox, Woese, Margulis, and Lovelock (Dick & Strick, 2005; Mesler & Cleaves II, 2015); however, the field really grew exponentially during the 1990s –after the controversial Science article on the Allan Hills 84001 meteorite (McKay et al., 1996) and the establishment of the NAI (Blumberg, 2003) which provided a continuous funding and organizational structure to researchers in the U.S. In the last decade, we find many synthesis works, books and textbooks that cover a diverse body of literature, delve into the history of OoL and OoL research, and recommendations for research directions (Luisi 2006; Des Marais et al. 2008; Plaxco and Gross 2011; Impey, Lunine & Funes 2012; Mesler & Cleaves 2015; Cockell 2015; Scharf et al. 2015; Smith & Morowitz 2016; Domagal-Goldman et al. 2016; Horneck et al. 2016); however, few of them utilize bibliometric tools to understand how the OoL field has progressed.

Of the publications that utilize bibliometric tools, the focus is given to astrobiology and interdisciplinary collaboration. For instance, Brazelton and Sullivan (2009) investigated the emergence of astrobiology as a new interdisciplinary domain by comparing it to the emergence of geology in the 1800s. Gazan's group (Gowanlock & Gazan 2013; Miller, Gazan & Still 2014) utilized ~700 publications published by NAI-funded researchers to understand collaboration and identify connections between research domains. Finally, Taskin & Aydinoglu's (2015) study investigated publication practices of scholars, interdisciplinary collaboration, and international collaboration through publications between 2008 and 2012 that were funded by the NAI.

Although, these views and interpretations of OoL research are quite valuable, it has to be admitted that the notions we have about the field are not based on scholarly output and bibliometrics. Furthermore quantitative measures of the OoL domain are lacking. This study addresses that gap through creating an inventory of 5,647 OoL related publications, identifying the disciplines & domains involved in OoL research, journals that publish OoL research, researchers in the field and their affiliations, seminal articles in the field, etc. OoL, like astrobiology, is and needs to be highly multidisciplinary field. Here multidisciplinary research is used in the sense that different fields speak as separate voices, it is an ad hoc mix, they retain their disciplinary identity whereas interdisciplinary research is integrative and a more holistic understanding of the core problem is formed together with a new community and jargon, and coordinated and collaborative efforts and even an organizational framework is observed (Wagner et al. 2011; Porter et al. 2006; Stokols et al. 2003; Rosenfield 1992). Thus, for OoL to be a sustainable field, it has to be interdisciplinary. Quantifying and measuring the scholarly outputs of OoL researchers and whether and how they are connected to each other is critical for a better understanding of the field.

Methods

Data Generation

A panel that consisted of the ELSI Origins Network (EON¹) management team² and the co-authors of this study identified keywords that were relevant to OoL research. These keywords were “origin of life”, “origins of life”, “astrobiology”, “exobiology”, and “prebiotic chemistry”. Using these keywords we conducted a full-text article search in the *Web of Science™ Core Collection* database to create our dataset. Once the dataset was created, the EON team checked the dataset to ensure its quality and relevance.

As of June 14, 2016 our search retrieved 5,647 publications indexed in the *Web of Science™ Core Collection*. Of the 5,647 publications, 57.4% are tagged as articles, 12% as proceedings, 8.8% as reviews, and 7.5% as both articles and proceedings. The number of publications (dark red line) and the citations they received (blue line) per year can be found in Figure 1.

¹ EON is a world-wide interdisciplinary network centered at Earth-Life Science Institute, Tokyo Institute of Technology to conduct research on the origins of life and establish a collaborative research community. The initial funding for EON came from the John Templeton Foundation. (<http://eon.elsi.jp/about-us/>)

² At the time of the study the EON management team consisted of Piet Hut, Henderson Cleaves, Nathaniel Virgo, and Jun Kimura, all of whom are involved in OoL research.

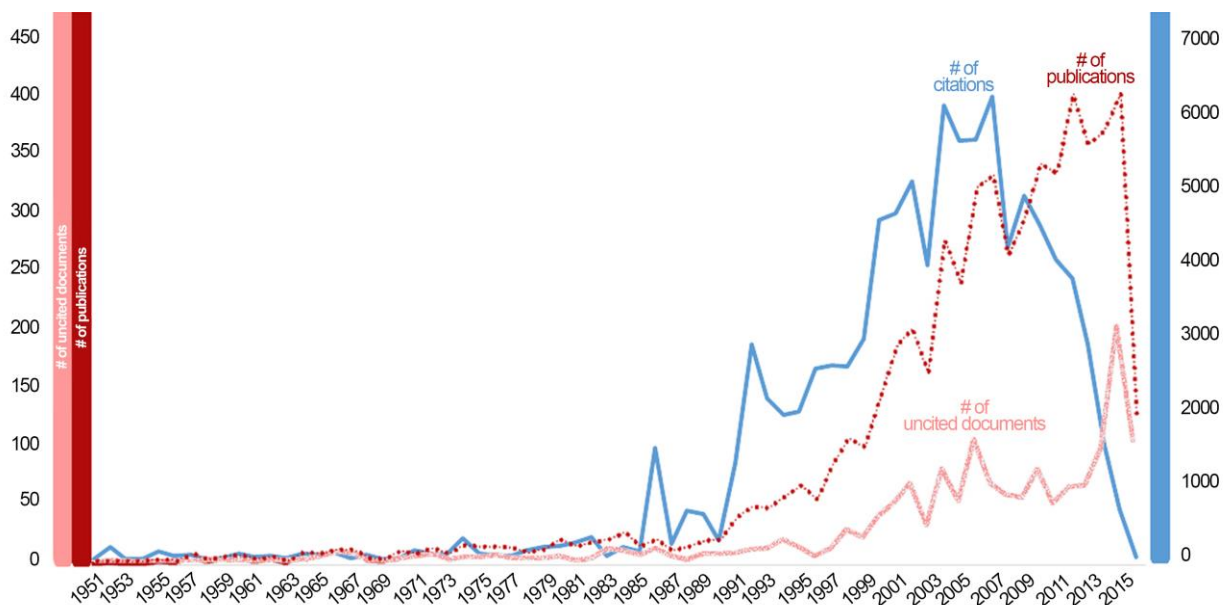


Fig 1. Number of publications, Number of citations, and Number of uncited papers per year

As can be seen, there were a few publications in OoL every year since 1951. A dramatic increase, however, happened in the 1990s. This jump can be explained by the interest and funding generated by the famous and controversial McKay et al. (1996) study regarding the possibility of evidence for life in the AL84001 Martian meteorite and the establishment of the NASA Astrobiology Institute, which has become a major funder of research in this domain for almost two decades. A rough calculation of NAI Cooperative Agreement Notice Grants gives us around \$350 million in seven cycles since its inception (NAI, 2015; NSPIRES, 2015). Citations have also tripled in the last 20 years which shows that the field is still developing. The reason for the drop for the last three years is that many recent publications have not yet received citations and there is a delay in citation measures. It is expected that these recent publications will receive citations in upcoming years. The number of publications that have yet to receive citations is also shown in the figure. The peak around 2013 indicates that there is an approximately 2 to 5 year delay in citations for many publications.

Visualization and Mapping

Visualization of connections between authors, institutions and countries has been accepted as one of the most important part of bibliometric studies (Chen, 2006, p. 1). In this context, VosViewer (<http://www.vosviewer.com>) and CiteSpace (<http://cluster.cis.drexel.edu/~cchen/citespace>) visualization tools, which are freely available software for bibliometric studies, are used to create visualizations and to calculate betweenness centrality scores (explained in Table 1) in this study. Detailed information about creating maps by using these software is available on the manuals of the software (Chen, 2016; van Eck and Waltman, 2016).

Analysis

A number of bibliometric analyses were conducted to investigate the professional OoL literature. Although these methods are common among scientometrics and library and information science scholars; for the readers of *Origins of Life and Evolution of Biospheres*, we provide short definitions (Table 1).

Table 1. Types of bibliometric analysis used in the study

Analysis	Definition
Co-authorship	Identifies active & productive authors and their collaborators. A network analysis provides information regarding who collaborates with whom and the names of prominent scholars in the field.
Journal analysis	Identifies the venues in which OoL research to be published. As OoL research is interdisciplinary, there may be more journals than a single researcher can follow. Bringing all these journals to the attention of the greater OoL community would be beneficial for general scholarship and research.
Citing	Identifies the citations OoL researchers use to see which research they cite and from which journals & fields. This helps to see who, and which domains intellectually, feed the community.
Keyword analysis	This is about analyzing the words and phrases in the titles, keywords or abstracts of the articles. We can both see which keywords are more of interest to the community and identify connections among them (whether they are used together frequently). This helps identify hidden connections among concepts, especially in large interdisciplinary domains such as the OoL.
Institutional collaboration	This analysis identifies the affiliations of authors. Through this analysis, it is possible to identify institutions that support OoL research and collaborative connections between institutions.
Geo/location collaboration	Through this analysis, national and international collaboration networks can be identified.
Betweenness centrality score	A measure of a node's centrality in a network, equal to the number of shortest paths from all vertices to all others that pass through that node. It is used whenever network analysis is applied and network density is measured. Measuring the density of a network gives us a ready index of the degree of dyadic connection in a population (Hanneman & Riddle, 2005).
Co-occurrence	This measures documents which are cited together in an article.

Our dataset consists of only publications that are indexed by the *Web of Science™ Core Collection* database. Web of Science indexes around 12,000 journals; however, there are other indexes with wider coverage such as Scopus with around 20,000 journals. We also did not cover any books. That being said, the *Web of Science™ Core Collection* database is used by many bibliometric studies and together with Elsevier's Scopus has become an industry standard for bibliometrics and scientometrics.

Findings

Productivity and Connectivity

Table 2 provides the names for the most productive scholars in the wider OoL research worldwide as measured by these metrics. Most of the names are from the NAI network which can be found on the NAI Directory website – nai.nasa.gov/directory (Aydinoglu et al., 2016; NAI, 2017). Scholars who are in the NAI network are marked by † in Table 2 below. There are 19 names which is not surprising given that the NAI has been funding astrobiology research for almost two decades (NAI, 2016). The second column provides the names of scholars who are most connected, who serve as a node among the different research domains in OoL. The * near the names shows that these names exist on both lists (productivity & connectivity). There are 11 names that exist on the both list which indicates that productive authors also play an important role in bringing together the diverse research networks of OoL research together.

Table 2. Most productive scholars in wider OoL research worldwide

	Name	Frequency		Name	Centrality
1	Raulin F†	60	1	Brucato JR	0.31
2	Ehrenfreund P*†	48	2	Saladino R	0.29
2	Cockell CS*†	48	3	Westall F*†	0.26
4	Brack A*†	45	4	Ehrenfreund P*†	0.24
4	Horneck G	45	5	Cleaves HJ*†	0.14
6	McKay CP*†	44	6	Glavin DP†	0.13
7	Edwards HGM*	41	7	Lazcano A*†	0.12
7	Luisi PL	41	7	Di Mauro E*	0.12
9	Lazcano A*†	36	7	Pascal R	0.12
10	Hoover RB†	34	10	Coll P*†	0.11
10	Ferris JP†	34	11	Cockell CS*†	0.1
12	Westall F*†	33	11	Selsis F	0.1
13	Kobayashi K	32	13	Rettberg P†	0.09
14	Coll P*†	31	13	Rozanov AY	0.09
14	Lammer H	31	15	McKay CP*†	0.08
16	Di Mauro E*	30	16	Brack A*†	0.07
16	Russell MJ†	30	16	Dworkin JP†	0.07
18	Bada JL†	29	16	Allamandola LJ†	0.07
18	Hazen RM†	29	16	Reitz G	0.07
20	Maurel MC*	28	20	Edwards HGM*	0.06
21	Cleaves HJ*†	27	20	Maurel MC*	0.06
21	Szopa C	27			
21	Schwartz AW	27			
21	Szostak JW†	27			

Since OoL research has been around for more than half a century, a temporal analysis of the co-authorship networks is possible, which is provided in Fig.2. We see two things when we look at early OoL research: the scholars that are relevant and influential and the scholars that become less relevant over time. The latter, which are represented in blue and green in figure 2, have not been cited significantly for some time. This means that current research trends are not interested in them (their theories and results may have been discarded, their methodologies may have been surpassed, maybe there are better and more precise instruments now, etc.). We thus focus our attention on the research that is evidently deemed more relevant as measured by citation in this database. We divide this group into two based on the date of publication. The first group consists of Miller, Wächtershäuser, Chyba, Sagan, Eigen, Orgel, Shapiro, Oparin, Huber, Gilbert, Oró, Ponnampereuma, and others who can be considered the founding scholars of OoL, some of whom may not be alive and/or scholarly active today. They have blue or green at the center and red at the periphery in fig. 2. The second group is fragmented. Starting from the center (and going clockwise), there are a handful of nodes. Schopf (paleobiology); Bada (amino acids and other organic compounds geochemistry); and Sagan; Chyba and Mojzsis (early Earth & geomicrobiology). The last node is important as it is an originating node for a number of networks (Brownlee, Cronin; Bernstein, McKay, Elsila) which are also connected to another group of current research networks (such as Glavin, Pizzarello, Caro, Sephton, Martins, Callahan, Nuevo, etc.) towards the right of Fig. 2. Cockell (extremophiles) seems to not have many connections, yet he is an influential scholar as being a bridge between past and current research (green & red lines). A red line with some branches can be seen on the right hand-side of Fig. 2. Researchers here, including Squyres, Christensen, Ehlmann, Mustard, Grotzinger, Westall, Parnell and so on focus on solar system bodies. The node towards the bottom of Fig. 2 with Kasting, Kopparapu, Segura, Kaltenegger, Seager, Charbonneau, Borucki, Selsis, and others investigates planetary atmospheres, exoplanets, biosignatures, and habitability.

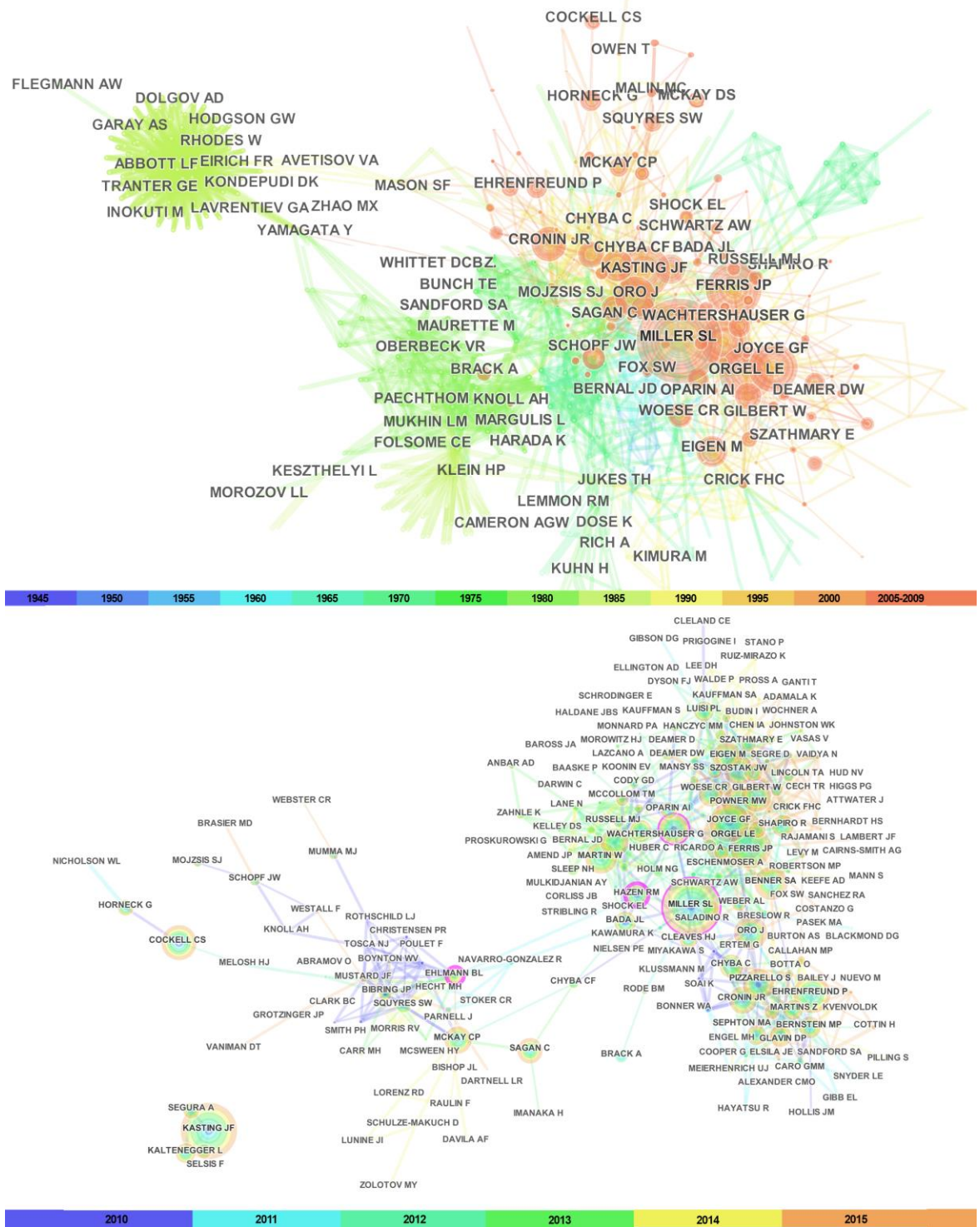


Fig 2. Temporal analysis of author co-citation network between 1945 and 2015
http://yunus.hacettepe.edu.tr/~ztaskin/author_co-citation.png

*How to read this graph: Each node represents a scholar and a line represents that these nodes are cited together. The thickness represents frequency. Colors show the date with five-year increments (see the color bar at the bottom of the figure for details).

Although almost all of the authors shown in fig. 2 conducts interdisciplinary research, this section of the visual has more life scientists, and their research focused on how chemistry becomes biology. In addition to the inactive scholars mentioned earlier (starting from bottom to top) Hazen (mineralogy); Russell (hydrothermal vents), Ferris (clay minerals & RNA); Cleaves (abiological organic synthesis); Kauffman (complex systems), Deamer (nucleic acid & self-assembly), Benner (synthetic biology & paleogenetics), Szostak (chemical to biological evolution), and Morowitz (thermodynamics) are to name a few in this part of the figure.

Journals

In order to review journals in a field, there are some bibliometric measures such as Bradford’s law (diminishing returns of reference searches), Zipf’s law (a power distribution law used in many fields), and the 80/20 rule (Garfield, 1980). Originally developed for economics, the 80/20 rule found an application in bibliometrics, arguing that for an established field 80% of publications in that field are published in 20% of the journals (Egghe, 1986). There are 1139 journals in our dataset and 80% of the publications are in 260 (22%) of these journals. OoL research thus fits into the 80/20 rule.

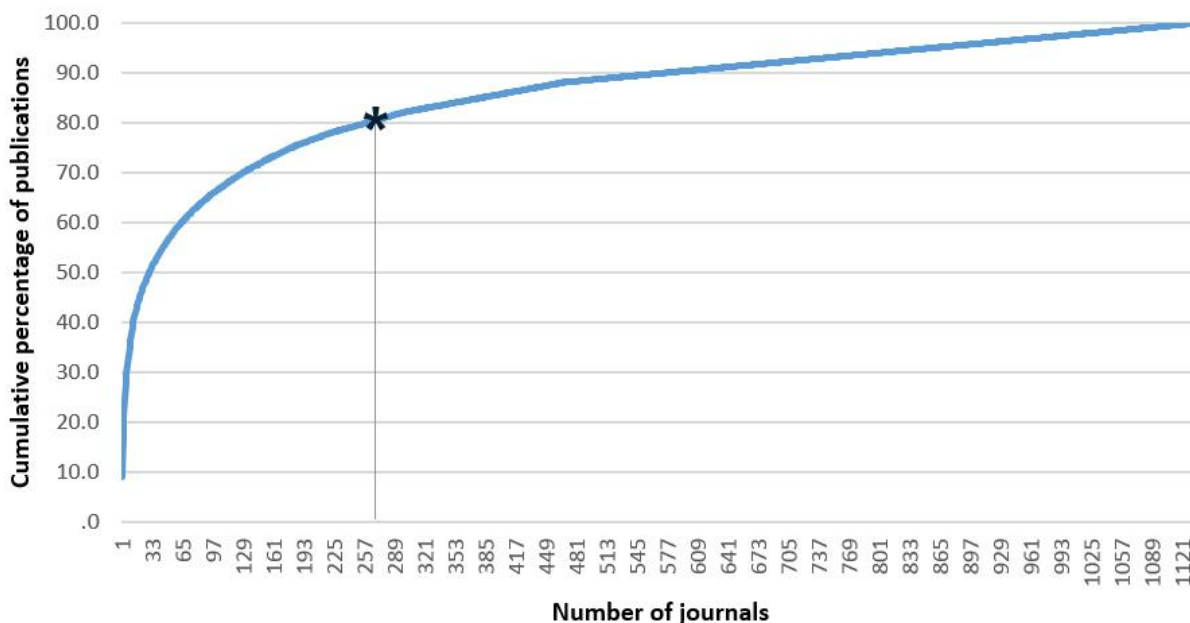


Fig. 3 – Number of journals and cumulative percentage of publications

As expected from a multidisciplinary field, these journals belong to variety of subject areas from Biology to Astronomy & Astrophysics to Geosciences to Multidisciplinary and even to Philosophy. Below Table 3 provides the top-12 most common journals for OoL research. *Origins of Life and Evolution of the Biospheres* and *Astrobiology* are the most preferred venues to publish OoL research. Although, both journals do not rank very high (please see ranking scores in Table 3 below, next to the journal title in parentheses; for instance in Biology, *Origins of Life and Evolution of Biospheres* ranks 55 out of 86 journals); due to their interdisciplinary focus, they reach a wider audience which suits very well to both astrobiology and OoL domains. *Nature*, *Science*, and *PNAS*, have the highest impact factors, are also in the top-10 list of multidisciplinary journals that publish OoL relevant research. *Astrobiology* and the *International Journal of Astrobiology* started in early 2000s, and provided scholars a new venue to publish relevant research.

Table 3. Top-12 most common journals for OoL publications in this dataset. In parentheses, ranking of a journal in a discipline can be found together with the number of journals in that discipline. The complete list of journals is given in the appendix.

Title	# of pubs	% of pubs
Origins of Life and Evolution of the Biosphere (55/86 Biology)	507	8.9%
Astrobiology (25/61 Astronomy & Astrophysics; 21/86 Biology; 41/184 Geosciences, Multidisciplinary)	451	7.9%
Astrophysical Journal (8/61 Astronomy & Astrophysics)	220	3.8%
Icarus (19/61 Astronomy & Astrophysics)	214	3.7%
International Journal of Astrobiology (42/61 Astronomy & Astrophysics; Biology 50/86; 134/170 Geosciences Multidisciplinary)	169	2.9%
Nature (1 of 63 Multidisciplinary)	111	1.9%
PNAS (4 of 63 Multidisciplinary)	107	1.8%
Planetary and Space Science (32/61 Astronomy and Astrophysics)	107	1.8%
Journal of Molecular Evolution (213/289 Biochemistry & Molecular Biology; 34/45 Evolutionary Biology, 120/165 Genetics & Heredity)	88	1.5%
Astronomy & Astrophysics (12/61 Astronomy and Astrophysics)	85	1.5%
Journal of Theoretical Biology (25/86 Biology; 14/56 Mathematical & Computational Biology)	78	1.3%
Science (2 of 63 Multidisciplinary)	75	1.3%

The top-12 journals that the OoL researchers identified in this dataset cite most are not that different but the order is: Nature (3407), Science (3385), PNAS (2226), Origins of Life and Evolution of Biospheres (2068); Icarus (1640); Astrobiology (1457); Journal of American Chemical Society (1246); Journal of Molecular Evolution (1192); the Astrophysical Journal (1114); Geochimica et Cosmochimica Acta (1042); Planetary and Space Science (935); and Astronomy & Astrophysics (874). (The complete list can be found in the appendix). The list is important as these are the journals that provide the inspiration for and validation of OoL research.

Keyword Analysis

In the keyword analysis only keywords & concepts mentioned together more than 10 times are analyzed. This provided 500 terms in seven clusters. Figure 4 only includes 60% of this data for clarity. (The static image shows less, for details see <https://goo.gl/VMZDr1>).

The seven clusters for OoL research are as follows: The first cluster (purple) seems to be related to exoplanet research. The relevant keywords here are “astrobiology”, “Earth”, “exoplanet”, “terrestrial planet”, “planetary systems”, “main sequence stars”, “planet”, “star”, “habitable zone”. The second cluster (red) is mostly about Mars but not limited to it as “Europa” is a keyword here too. The keywords in this cluster are “Mars”, “exobiology”, “water”, “surface”, “environment”, “methane”, “radiation”, “meridian-planum”, “bacteria”. The third cluster (blue) is related to astrochemistry. “astrochemistry”, “chemistry”, “spectroscopy”, “atmosphere”, “meteorites”, “molecules”, “organic molecules”, “Murchison meteorite”, “comets”, “Titan”, “dust”, “solar system” are some of the prominent keywords in this cluster. Prebiotic chemistry (green) is the fourth cluster. “amino acid”, “prebiotic chemistry”, “prebiotic synthesis”, “chemical evolution”, “montmorillonite”, “aqueous-solution”, “primitive earth”, “hydrogen-cyanide”, “early Earth”, “glycine” are principle keywords here. The fifth cluster (yellow) is associated with the origins of life. The keywords here (some overlap, and thus are not clearly seen in Figure 4, please refer to the website for details) are “origins of life”, “origin”, “life”, “evolution”, “RNA”, “RNA world”, “metabolism”,

“protein”, “*Escherichia-coli*”, “selection”, “replication”, “emergence”, “protocell”. The last two clusters are quite small compared to the first five, yet they are significant enough to appear on the network map. The sixth one (turquoise) has “chirality”, “homochirality”, “asymmetric autocatalysis”, “symmetry-breaking”, and “enantiomeric excess” and the seventh one (light blue) has “science”, “education”, and “communication”. The latter emphasizes the importance of education and outreach efforts in OoL studies.

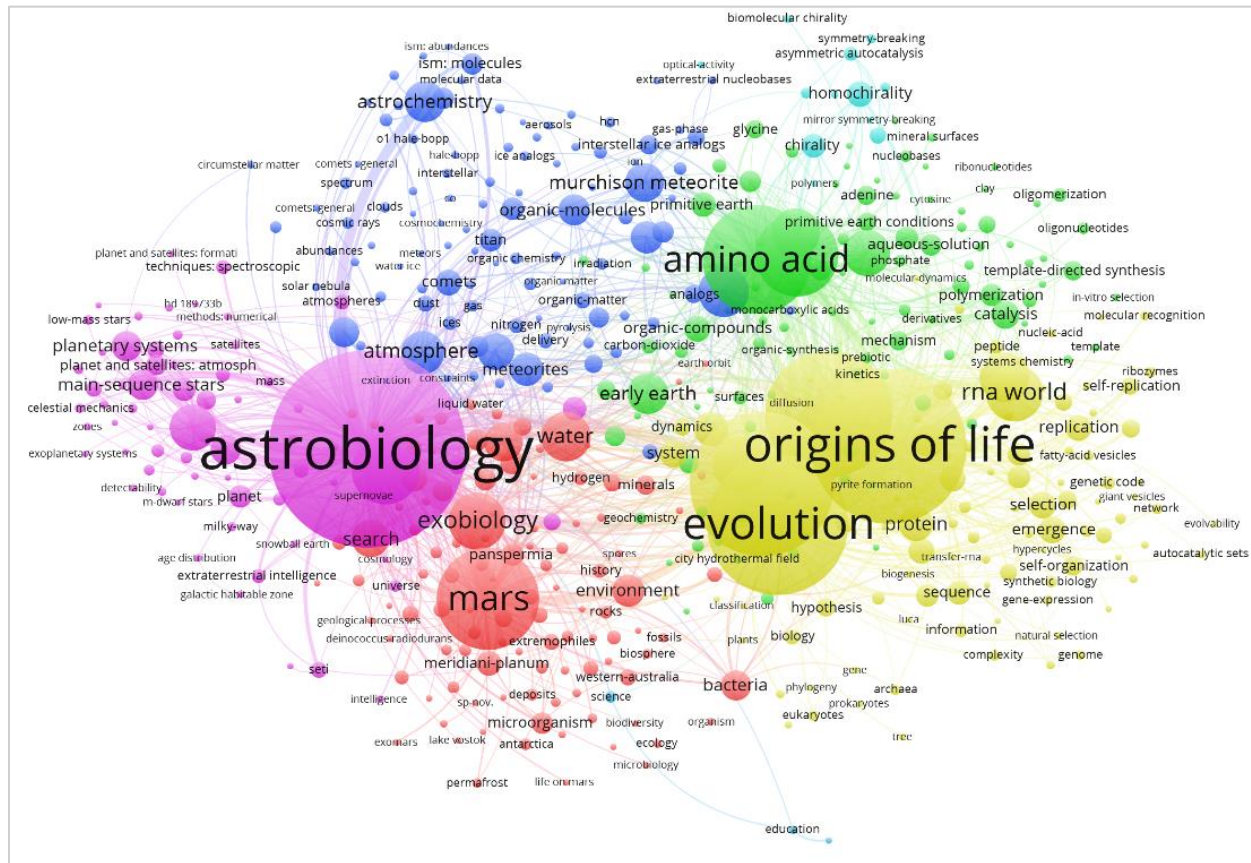


Figure 4. Keyword map of OoL papers (for full map on VosViewer: <https://goo.gl/VMZDr1>)

Co-cited Papers in the Field

The OoL papers identified here cite collectively 126,105 unique publications. Only the papers that have been co-cited more than 20 times are analyzed. A citation map only considers the number of citations whereas a co-citation map considers the citations that are cited together which helps describe a more relevant and comprehensive literature that maybe suitable for interdisciplinary fields such as OoL. According to the cited references, there are five clusters of literature (see Fig 7). Miller’s paper (*Miller S, 1953, A production of amino acids under possible primitive Earth conditions, Science*) is at the heart of OoL research and the only paper that is cited by every cluster and belongs to the purple cluster. That cluster also has the Chyba & Sagan paper on the sources of organic molecules on early Earth (*Chyba C & Sagan C, 1992, Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: An inventory for the origins of life, Nature*). The other prominent papers in the cluster are *Cronin JR & Pizzarello S, 1997, Enantiomeric excesses in meteoritic amino acids, Science*; *Oró, 1961, Comets and formation of biochemical compounds on primitive Earth, Nature*; and *Bernstein et al., 2002, Racemic amino acids from the ultraviolet photolysis of interstellar ice analogues, Nature*. The broad focus here is the production of amino acids under early Earth conditions or their delivery from meteorites. The red cluster, in terms of citations received, is led by Gilbert’s seminal article (*Gilbert, 1986, Origin of life: The RNA world*); followed by *Eigen M, 1971, Selforganization of matter and evolution of biological macromolecules,*

Naturwissenschaften; Szostak JW, Bartel DP, Luisi PL, 2001, *Synthesizing life, Nature*; and Orgel LE, 2004, *Prebiotic chemistry and the origin of the RNA world, Critical Reviews in Biochemistry and Molecular Biology*. These studies investigate RNA and prebiotic chemistry. The green cluster has Wächtershäuser G, 1988, *Before enzymes and templates: Theory of surface metabolism, Microbiological Reviews*; Wächtershäuser G, 1992, *Groundworks for an evolutionary biochemistry, Proceedings of the National Academy of Sciences*; Huber C, Wächtershäuser G, 1997, *Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions, Science*; Russell MJ & Hall AJ, 1997, *The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front, Journal of the Geological Society*. The chemistry of metabolic cycles, iron-sulfur compounds, and hydrothermal vents are the common themes in this cluster. The green, blue, and red clusters are tightly connected to each other, indicating shared relevant literatures, compared to the rest of the network. The purple cluster is mostly connected to the yellow & blue clusters. The blue cluster, detailing prebiotic chemistry in the origins of life, relatively receives less citations. Finally, the yellow cluster (microfossils and evidence of life on the early Earth) is the least connected literature. The publications here are connected to only the purple cluster through Miller and Chyba. The publications are more recent too. McKay 1996 *Science*, Benner 2000 *PNAS*, Kasting JF, 1993, *Earth's early atmosphere, Science*; Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutman AP, Friend CRL, 1996, *Evidence for life on Earth before 3,800 million years ago, Nature*; and Schopf JW, 1993, *Microfossils of the early Archean apax chert: New evidence of the antiquity of life, Science* are the prominent ones. (Please see the Appendix for the full titles of the most cited 40 publications in OoL)

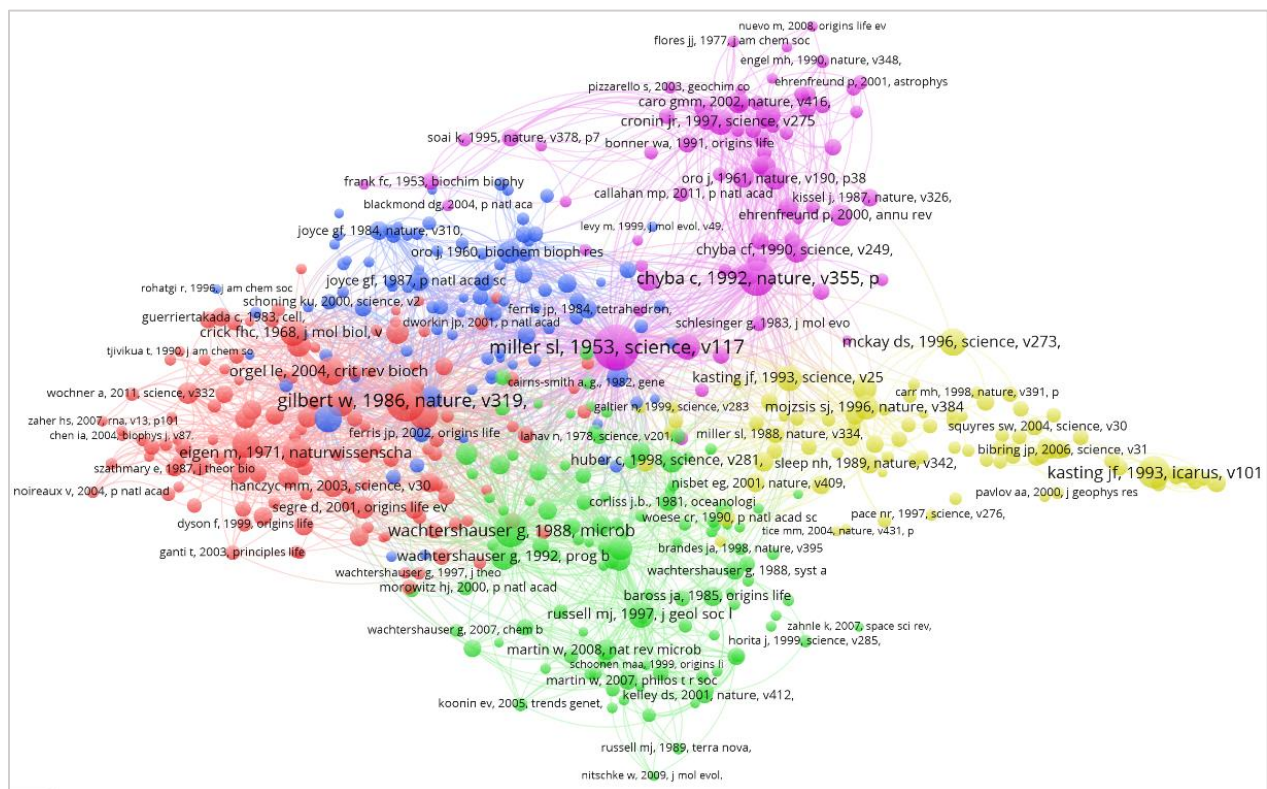


Figure 5. Mostly cited papers in the field (for full map on VosViewer: <https://goo.gl/6NoUGR>)

Institutional affiliations of the OoL scholars

Among the researchers who publish in the OoL field, NASA and Caltech (NASA staff at the Jet Propulsion Laboratory are employed by Caltech, not NASA) employ the most. NASA is a behemoth when it comes to

OoL research –being not only the most productive but also the most connected to other nodes in the network. However, that is not surprising as, through the Astrobiology Program, NASA is the main funder of astrobiology and exobiology research³. The University of Washington team and Caltech (with and without JPL staff) are important players in the NAI; they are scholarly productive and well-connected to the rest of the network. The University of Colorado team was an NAI team for 10 years, then not for seven years, and is now again a member, and an important hub for OoL research. Only The University of California San Diego is not among the U.S. institutions that have not been awarded a CAN (Cooperative Agreement Notice⁴) by the NAI. The University of Arizona and Harvard University was NAI teams previously (Harvard University is still active through the MIT team). Their research is quite interdisciplinary, bringing different expertise to the table. Outside the U.S., CNRS (the French National Center for Scientific Research), UCL (University College London), The University of Paris VI (Pierre and Marie Curie University), CSIC (Spanish National Research Council), University of Tokyo, and Imperial College are the most prominent institutions. The network map in Figure 6 that is generated through co-authors’ affiliations shows how these institutions collaborate with each other. There are four major clusters, with NASA at the center.

Table 4. Institutional affiliation of the co-authors of OoL publications

Institution	Freq	Institution	Cent
1 NASA*	356	1 NASA*	0.38
2 California Institute of Technology*	134	2 University of Arizona *	0.24
3 University of Arizona *	85	3 University of Washington*	0.16
4 Russian Academy of Sciences	83	4 Harvard University*	0.15
5 CNRS*	70	5 University of Leicester	0.13
6 University of Colorado *	60	6 California Institute of Technology*	0.12
7 SETI Institute	59	7 CNRS*	0.12
8 University of Washington*	58	8 Johns Hopkins University	0.12
9 National Autonomous University of Mexico	58	9 UCL*	0.11
10 Harvard University*	51	10 CSIC*	0.11
11 UCL*	46	11 University of California, Santa Cruz	0.11
12 Arizona State University	44	12 University of Paris VI *	0.10
13 University of California San Diego *	40	13 Tokyo Institute of Technology	0.10
14 University of Hawaii at Manoa	37	14 Paris Observatory	0.10
15 University of Paris VI *	36	15 University of California San Diego *	0.09
16 CSIC*	34	16 University of Tokyo *	0.09
17 University of California, Berkeley	34	17 National Research Council, Italy (CNR)	0.09
18 University of Tokyo *	32	18 United States Geological Survey	0.08
19 Imperial College London *	32	19 University of Colorado *	0.07
20 Open University	32	20 Imperial College London *	0.07
		21 University of Edinburgh	0.07
		22 Leiden University	0.07

³ It has to be mentioned that there are other programs that support the community. For instance, University of Washington received an Integrative Graduate Education and Research Training (IGERT) award and Rensselaer Polytechnic Institute a NASA Specialized Center of Research and Training (NSCORT) award. The University of California San Diego also received an NSCORT award.

⁴ CAN’s are five-year funding programs that support astrobiology research on a team scale.

Table 5 – Number of OoL publications and betweenness centrality scores based on country

Rank	Country	Frequency	Centrality	Rank	Country	Frequency	Centrality
1	USA	2123	0.28	11	China	116	0
2	France	429	0.3	12	Switzerland	84	0.04
3	England	411	0.24	13	Brazil	80	0
4	Germany	306	0.13	14	Austria	77	0.02
5	Italy	261	0.1	15	Mexico	76	0.01
6	Spain	210	0.05	16	Scotland	72	0.03
7	Japan	203	0.06	17	Israel	69	0
8	Russia	137	0.05	18	Australia	65	0
9	Netherlands	128	0.03	19	Belgium	59	0.06
10	Canada	127	0.04	19	India	59	0

Disciplines & Domains

This network is developed through the Web of Science Journal Category dataset. Web of Science tags each of the journals in their database with at least one subject category, sometimes more. For instance, *Gondwana Research* is tagged with *Geology* whereas *Astrobiology* has three tags *Astronomy & Astrophysics*, *Biology*, and *Geosciences*. An article is tagged with the Journal it is published in.

The network for the journal subject categories that publish OoL research is provided in Fig. 7 below. Almost every scholarly field can be related somehow to OoL research. *Astronomy and Astrophysics* and *Life Sciences & Biomedicine* are the two big pivot nodes, which are joints between different networks. The pivot nodes are either the common nodes shared by two networks or gateway nodes that are connected by internetwork links (Chen, 2006, p. 287). *Geology* is a big node but not a pivot. Other pivot nodes include *Chemistry*, *Biochemistry & Molecular Biology*, *Science & Technology*, *Computer Science*, *History & Philosophy of Science*, and *Philosophy*.

Moreover, the humanities (*Religion*, *Philosophy*, *History & Philosophy of Sciences*) and social sciences (*Social Sciences*, *Social Issues*, *Education & Educational Research*), are part of the disciplinary network, though they are not very recent (see Fig. 7).

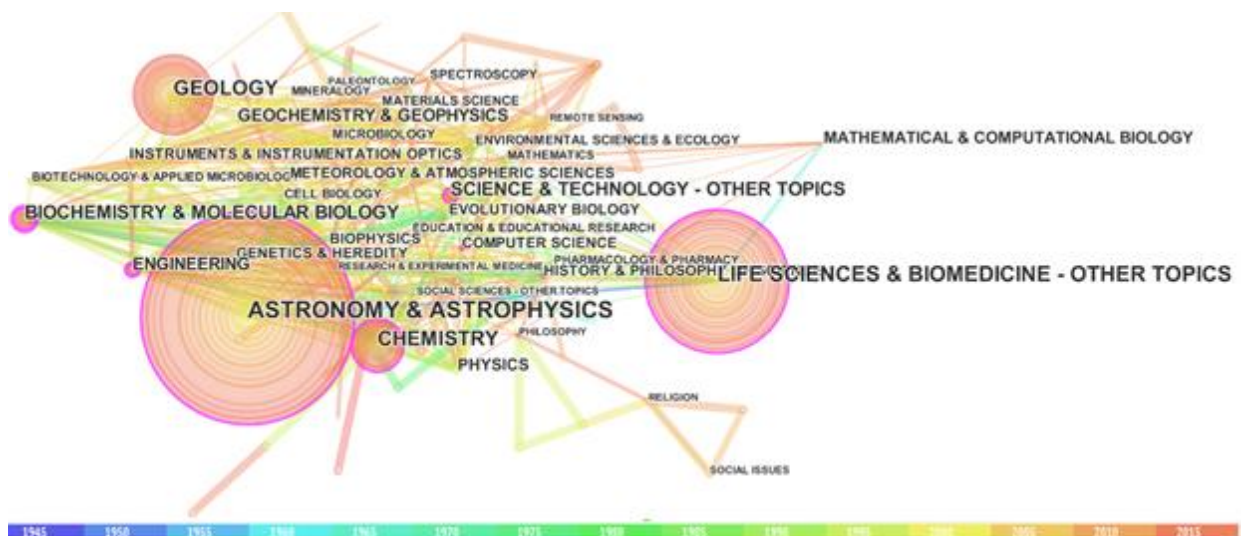


Figure 7. Network map of OoL relevant research publishing journals' subject categories

How to read this graph: Each node represents a scholar and a line represents that these nodes are cited together. The thickness represents frequency. Colors show the date with five-year increments, see the color bar at the bottom of the figure for details..

Conclusion

This study provides insights on OoL research utilizing bibliometric tools. The analysis focuses on journal articles and proceedings in OoL, astrobiology, exobiology, and prebiotic chemistry topics. We report a continuous growth in publication in the OoL field -especially after the mid-1990s (around the time of the controversial Allan Hills 84001 meteorite study and the establishment of the NASA Astrobiology Institute). Since then publications in OoL research have increased exponentially. We tracked the chronology of OoL research and the research topics through co-authorship networks. This study identifies the need for a network to connect different OoL research networks, do synthesis work, and make sense of disparate OoL research, and thus, create a well-connected & integrated OoL research community.

The field is true to its collaborative premise according to this data. Researchers from different departments and countries direct their efforts to address OoL questions. The U.S., France, England, and Germany are the leading national actors in the field in terms of publication in this database. NASA, the California Institute of Technology, the University of Washington, the CNRS, the University of Arizona, and Harvard University are among the most productive institutions in this regard. However, there are also individual nodes and researchers that are important to the network.

The paper also identifies research trends and topics through clustering of keywords and titles of publications through co-occurrence which provides a basis for examining collaboration patterns. Here, exoplanets, missions to Mars and Europa, astrochemistry, prebiotic chemistry, RNA world, and origin(s) of life are the biggest clusters and there are many connections between these clusters. Astrobiology is the dominant keyword (Figure 4). We suspect that NAI is the reason for that as they are a big funder, if not the biggest, providing an institutional body to the community while facilitating lots of activities from conferences to workshops to summer/winter schools. An institutional body that catalyzes interdisciplinary research, such as NASA, can easily promote the term “astrobiology”, hence the increased use of the term. The term “astrobiology” was introduced in the 1990s (Dick & Strick, 2005) and has already surpassed the other keywords that have been around for much longer such as exobiology, origin of life, and evolution. In addition, two relatively new astrobiology journals (*Astrobiology* and *International Journal of Astrobiology*) are promoting the term while providing publishing venue for researchers.

To add more on institutional support, for almost 15 years, NASA has been the only game in town in regards to supporting collaborative research. However, the recently established Earth-Life Science Institute’s (ELSI) impact may be demonstrated in the coming years as their per annum budget is close to the half of what the NAI has provided for (~\$350million/18 years). ELSI’s kick-off funding was \$53million for 5 years and received another \$50 for the second 5 years (WPI 2012; Kaufman 2017). Thus, if we accept that more funding results in more research output, more ELSI (and EON) affiliated researchers are expected to appear in network maps in the coming decade. EON may catalyze the transformation from multidisciplinary research to interdisciplinary research by creating and sustaining a community. The European Space Agency does not have a specific astrobiology program; the activities are under “Human Spaceflight & Robotic Exploration” and “Space Science” programs (Horneck et al., 2016; ESA, 2017). Data for Russia and China were not readily available for comparison.

As for looking at the NAI’s impact on the field, we need to compare another study (Taskin & Aydinoglu, 2014) that investigated NAI. Although that study covered only four years with a smaller dataset (between 2008 and 2012 & ~1200 publications), and to a degree overlaps with the OoL community, there are similarities and differences that need to be highlighted. Both astrobiology and OoL fields demonstrate interdisciplinary traits. A cadre of researchers from astrophysical sciences to biological sciences to

geophysical sciences are publishing in both fields. Both studies find that multidisciplinary journals are principally used to publish; even though, OoL relies on two main journals (*Origins of Life and the Evolution of Biospheres* and *Astrobiology*).

Although interdisciplinary fields the prominence of astrophysics/astronomy is obvious in both astrobiology and OoL. The keywords Mars, exoplanets, habitability, exoplanet, etc. suggests that an important part of their field is driven by space exploration. After the Kepler Space Telescope, exoplanet research has almost become a subdiscipline, and more studies are needed to quantify its impact. Still, geological sciences and life sciences are important part of OoL research.

As for the scholars in the field, the density of the co-authorship network of the NAI, although it had some prominent authors, never reached to a significant degree (probably due to investigating only four years of publication record) whereas for OoL research Miller, Chyba, Orgel, Russell, Wächtershäuser, Gilbert, Kasting, Eigen and others' studies constitute a recognizable core literature.

Another difference is the existence of publications in humanities and social sciences regarding the OoL. NAI teams consist of "hard" scientists mostly and focus on the science itself but not the philosophy or the history of the science. We do not see many synthesis works that covers the whole domain. OoL, on the other hand, is a broader area of study. Philosophers, social scientists, and theologians publish studies regarding the OoL. Of course we acknowledge that for a comprehensive comparison of astrobiology and OoL, the analysis of four years of NAI publications is not enough; all publications since the establishment of NAI, as it denotes the start of astrobiology, from 1997 need to be studied. In a future study, we would like to identify all publications resulting from the NASA Astrobiology Program (NAI and Exobiology Program) in the OoL dataset and measure the impact of NASA in the field.

Acknowledgement

This study was funded by an EON Seed Grant (ELSI Origins Network Seed Grant).

This project/publication was made possible through the support of a grant from the John Templeton Foundation provided through the Earth-Life Science Institute of the Tokyo Institute of Technology. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation or the Earth-Life Science Institute.

Bibliography

- Aydinoglu AU, Allard S, Mitchell C (2016) Measuring diversity in disciplinary collaboration in research teams. *Research Evaluation* 25(1): 18-36.
- Blumberg B (2003) The NASA Astrobiology Institute: Early history and organization. *Astrobiology* 3: 463-470.
- Brazelton WJ & Sullivan WT (2009) Understanding the nineteenth century origins of disciplines: Lessons for astrobiology today? *International Journal of Astrobiology* 8:257-266.
- Chen C (2006) *Information visualization: Beyond the horizon*. Springer Verlag, London.
- Chen C (2016) *How to use CiteSpace*. Lean Publishing, Canada.
- Cockell CS (2015) *Astrobiology: Understanding life in the universe*. Wiley-Blackwell, UK.
- Davies P (2007) Are aliens among us?. *Scientific American* 297(6): 62-69
- Des Marais D, Nuth JA, Allamandola LJ, Boss AP, Farmer JD, Hoehler TM, Jakosky BM, Meadows VS, Pohorille A, Runnegar B, & Spormann AM (2008) The NASA astrobiology roadmap. *Astrobiology* 8:715-30.

- Dick, SJ & Strick JE (2005) *The living Universe: NASA and the development of astrobiology*. Rutgers University Press: News Brunswick.
- Domagal-Goldman SD, Wright KE, Adamala K, Arina de la Rubia L, Bond J, Dartnell LR, Goldman AD, Lynch K, Naud ME, Paulino-Lima IG, Singer K, Walter-Antonio M, Abrevaya XC, Anderson R, Arney G, Atri D, Azúa-Bustos A, Bowman JS, Brazelton WJ, Brennecke GA, Carns R, Chopra A, Colangelo-Lillis J, Crockett CJ, DeMarines J, Frank EA, Frantz C, de la Fuente E, Galante D, Glass J, Gleeson D, Glein CR, Goldblatt C, Horak R, Horodyskyj L, Kaçar B, Kereszturi A, Knowles E, Mayeur P, McGlynn S, Miguel Y, Montgomery M, Neish C, Noack L, Rugheimer S, Stüeken EE., Tamez-Hidalgo P, Walker SI, and Wong T (2016) *The Astrobiology Primer v2. 0*. *Astrobiology* 16(8):561-653.
- van Eck NJ and Waltman L (2016) *VosViewer manual: Manual for VosViewer version 1.6.5*. CWTS, Leiden.
- Egghe, L (1986). On the 80/20 rule. *Scientometrics*, 10: 55-68
- European Space Agency (2017) *Funding*. Retrieved on May 1, 2017 from http://www.esa.int/About_Us/Welcome_to_ESA/Funding
- Garfield, E (1980) Bradford's law and related statistical patterns. *Essays of an information scientist*, 4(19): 476-483
- Gowanlock M. & Gazan R. (2013) Assessing researcher interdisciplinarity: A case study of the University of Hawaii NASA Astrobiology Institute. *Scientometrics* 94:133-161.
- Hanneman, RA & Riddle M (2005) *Introduction to social network methods*. University of California, Riverside, Riverside, CA.
- Hartman H (1998) Photosynthesis and the origin of life. *Origins of Life and Evolution of Biospheres* 28: 515-521
- Horneck G, Walter N, Westall F, Grenfell JL, Martin WF, Gomez F, Leuko S, Lee N, Onofri S, Tsiganis K, Saladino R, Pilat-Lohinger E, Palomba E, Harrison J, Rull F, Muller C, Strazzulla G, Brucato JR., Rettberg P, and Capria MT (2016) *AstRoMap European astrobiology roadmap*. *Astrobiology* 16, 201–243.
- Impey C, Lunine J, Funes J (2012) *Frontiers of astrobiology*. Cornell University Press: New York.
- Kaufman, M (2017) *Research center: A hub for origins of life studies*. *AstroBio Magazine*. Accessed on April 21, 2017, retrieved from <http://www.astrobio.net/alien-life/research-center-hub-origins-life-studies/>
- Lazcano A (2010) Historical development of origins research. *Cold Spring Harbor Perspectives in Biology* 2(11): doi:10.1101/cshperspect.a002089
- Luisi PL (2006) *The emergence of life: from chemical origins to synthetic biology*. Cambridge University Press, Cambridge.
- Mesler B & Cleaves II HJ (2015) *A brief history of creation: Science and the search of the origin of life*. W.W. Norton & Company, New York.
- Miller LJ, Gazan R & Still S (2014) Unsupervised document classification and visualization of unstructured text for the support of interdisciplinary collaboration. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW 2014)*, 15-19 February, Baltimore, MD, (pp: 1033-1042).
- NASA Astrobiology Institute (2015) *Funding opportunities*. Retrieved on April 1, 2017 from <https://nai.nasa.gov/funding>

- NASA Astrobiology Institute (2016) About NAI. Retrieved on April 2, 2017 from <https://nai.nasa.gov/about>
- NASA Astrobiology Institute (2017) NASA Astrobiology Directory. Retrieved on April 2, 2017 from <https://nai.nasa.gov/directory>
- NASA Solicitation and Proposal Integrated Review and Evaluation System [NSPIRES] (2017). Closed/Past Solicitations and Selections. Retrieved on April 1, 2017 from <https://nspires.nasaprs.com/external/solicitations/solicitations.do?method=closedPastInit&stack=push>
- National Research Council (2003) Life in the Universe: An Assessment of U.S. and International Programs in Astrobiology. Washington, DC: The National Academies Press.
- OECD (2016a), “Expenditure on R&D”, in OECD Factbook 2015-2016: Economic, Environmental and Social Statistics, OECD Publishing, Paris.
- OECD (2016b), “Researchers”, in OECD Factbook 2015-2016: Economic, Environmental and Social Statistics, OECD Publishing, Paris.
- Oparin, A. (1952) The origin of life. Dover: New York.
- Plaxco KW & Gross M (2011) Astrobiology: A brief introduction. Johns Hopkins University Press: Maryland
- Porter AL, Roessner JD, Cohen AS, and Perreault M (2006) Interdisciplinary research: Meaning, metrics and nurture. *Research Evaluation* 15(3):187-195
- Rosenfield PL (1992) The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science and Medicine* 35(11):1343-1357.
- Scharf C, Virgo N, Cleaves HJ, Aono M, Aubert-Kato N, Aydinoglu A, Barahona A, Barge LM, Benner SA, Biehl M, Brassler R, Butch CJ, Chandru K, Cronin L, Danielache S, Fischer J, Hernelund J, Hut P, Ikegami T, Kimura J, Kobayashi K, Mariscal C, McGlynn S, Menard B, Packard N, Pascal R, Pereto J, Rajamani S, Sinapayen L, Smith E, Switzer C, Takai K, Tian F, Ueno Y, Voytek M, Witkowski O, & Yabuta H (2015) A strategy for origins of life research. *Astrobiology*, 15(12): 1031-1042. doi:10.1089/ast.2015.1113
- Schrödinger, E. (1944) What is life? The physical aspect of the living cell. Cambridge University Press: UK.
- Smith E, Morowitz HJ (2016) The origin and nature of life on Earth: The emergence of the fourth geosphere. Cambridge University Press, UK.
- Stokols D, Fuqua J, Gress J, Harvey R, Phillips J, Baezconde-Garbanati L, Unger J, Palmer P, Clark MA, Colby SM, Morgan G, and Trochim W (2003) Evaluating transdisciplinary science 5(Suppl_1):S21:S39. doi: 10.1080/14622200310001625555
- Taskin Z & Aydinoglu AU (2015) Collaborative multidisciplinary astrobiology research: A bibliometric study of the NASA Astrobiology Institute. *Scientometrics* 103:1003-1022. doi: 10.1007/s11192-015-1576-8
- Wagner CS, Roessner JD, Bobb K, Klein JT, Boyack JW, Keyton J, Rafols I, Börner K (2011) Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* 165:14–26. doi:10.1016/j.joi.2010.06.004
- World Premiere International (WPI) (2012) Proposal. Accessed on April 30, 2017 retrieved from http://www.jsps.go.jp/english/e-toplevel/data/201312/1_b_tokyo_e.pdf