



The discovery of archaea: from observed anomaly to consequential restructuring of the phylogenetic tree

Michael Fry¹

Received: 7 March 2023 / Accepted: 14 February 2024
© The Author(s) 2024

Abstract Observational and experimental discoveries of new factual entities such as objects, systems, or processes, are major contributors to some advances in the life sciences. Yet, whereas discovery of theories was extensively deliberated by philosophers of science, very little philosophical attention was paid to the discovery of factual entities. This paper examines historical and philosophical aspects of the experimental discovery by Carl Woese of archaea, prokaryotes that comprise one of the three principal domains of the phylogenetic tree. Borrowing Kuhn's terminology, this discovery of a major biological entity was made during a 'normal science' project of building molecular taxonomy for prokaryotes. Unexpectedly, however, an observed anomaly instigated the discovery of archaea. Substantiation of the existence of the new archaeal entity and consequent reconstruction of the phylogenetic tree prompted replacement of a long-held model of a prokarya and eukarya bipartite tree of life by a new model of a tripartite tree comprising of bacteria, archaea, and eukarya. This paper explores the history and philosophical implications of the progression of Woese's project from normal science to anomaly-instigated model-changing discovery. It is also shown that the consequential discoveries of RNA splicing and of ribozymes were similarly prompted by unexpected irregularities during normal science activities. It is thus submitted that some discoveries of factual biological entities are triggered by unforeseen observational or experimental anomalies.

Keywords Scientific discovery · Microbial phylogenetics · Ribosomal RNA-based phylogenetics · Archaea · Carl Woese

✉ Michael Fry
mickey@technion.ac.il

¹ Department of Biochemistry, Rappaport Faculty of Medicine, Technion–Israel Institute of Technology, Efron St., Bat Galim, POB 9649, Haifa 31096, Israel

1 Scientific discovery: philosophical issues

Scientific discovery of new theories or factual entities, (objects, systems, processes, etc.) is a component of some, but not all, multilayered endeavors of gaining new scientific knowledge about the natural world.¹ Opinions diverge on the boundaries of the term ‘discovery’.² Whereas some thinkers saw discovery as the instance of new scientific insight, (“eureka moment”) others stipulated that a discovery of new scientific theory or entity becomes recognized as such only after withstanding testing and substantiation.³

Whereas discovery of new theories and hypotheses was the subject of extensive philosophical discourse,⁴ philosophers paid much less attention to the philosophy of discovery of factual entities. Three factors contributed to this asymmetry. First, because philosophy of science of the nineteenth and twentieth centuries took theory-governed physics as its dominant model of scientific research, (Popper, 1959, 2002; Carnap, 1966; Kuhn, 1970b; Chalmers, 2013) discovery of theories was thought to be the primary impetus for scientific inquiry. Second, under this theory-first viewpoint versions of the theory-driven hypothetico-deductive method became heuristic norm of scientific research (Popper, 1959, 2002; Platt, 1964; Godfrey-Smith, 2003). Third, because of their vastly heterogeneous histories and intricate technical specificities, experimental discoveries of factual entities resisted systematic philosophical analyses.

This paper examines a prominent case of experimental discovery of an unforeseen major biological factual entity: archaea, a third super-kingdom (domain) of the evolutionary tree of life, (Sect. 2). This and analogous discoveries in biology, (Sect. 3) are put into philosophical context in Sects. 1 and 3.

¹ Scientific discovery is an important, but not an only or necessarily obligatory component of the multifarious processes of acquisition of new knowledge about the natural world. Addressing the multitude of philosophical questions and ideas on the scientific pursuit of knowledge about nature is beyond the scope of this paper and the relevant philosophical literature is much too vast to be adequately addressed here. This paper is limited to historical and philosophical thinking on discovery at large and particularly on discovery of factual biological entities. Cited publications on these topics were selected from a more expansive body of work.

² This paper uses lexical definitions of the term ‘discovery’: “[an] act of finding or learning something for the first time” (Britannica Dictionary) and “[a] process of learning something the was not known before or finding (...) something that was not known about before” (McMillan Dictionary).

³ Disavowing the narrow ‘eureka moment’ view of discovery, Carl Kordig wrote: “Real discoveries are more than initially plausible [hypothesis]. They are well established and justified. Only what is justified is a discovery” [(Kordig, 1978) p. 112]. According to Norwood Hanson, a discovery is certified as such only after it is crystallized into ‘finished research report’ [(Lund, 2010) p. 26]. Extrapolating from the discovery of bacteriophages, William Summers generalized on the process that follows discovery of a “fact”: “...during (that) integration process, the fact seems to expand into a stable discovery by accretion of other facts, contexts, theoretical explanations, and general understanding. Thus, discovery seems to be a process rather than event” (Summers, 2021).

⁴ It may be argued that since theories, unlike physical objects or systems, do not exist in the real world, their birth is a process of (mental) ‘generation’ or ‘creation’ rather than ‘discovery’. This view notwithstanding, philosophers of science routinely use the term ‘discovery’ to describe birth of theories.

1.1 Discovery as part of an overall context of scientific progress

Discovery of new theories or factual entities is an important contributing element in many cases of scientific advancement (Schickore, 2022). Progress in science poses difficult and unsettled philosophical issues, involving profound questions such as what progress is, how is it made, and what distinguishes progress from other human intellectual undertakings. These unresolved questions notwithstanding, beginning with Francis Bacon, philosophers of the modern era have framed different general models of scientific progress. Some of the more enterprising models were offered in the last 70 years, (i.e., Popper, 1959/2002, 1965; Lakatos, 1968, 1970; Feyerabend, 1970, 1993; Kuhn, 1962b; Chang, 2007, 2017). Selected cases in the histories of physics, chemistry, and biology appeared to conform to one or another model. However, because of the high divergence of the theoretical and empirical paths to epistemological justification, no single model could serve as a universal template for all cases of scientific advancement.

Kuhn's model of the progression of science, (Kuhn, 1962b, 1970a, 1970b), distinguished between 'normal' and 'revolutionary' phases of scientific research. The normal science phase was perceived as 'puzzle solving' activity that takes place under an umbrella of a consensually accepted theoretical and methodological framework (a 'paradigm' which is distinct from a model). However, according to Kuhn, anomalies that accumulate during normal scientific activity ultimately reach a level that contravenes the accepted paradigm. At such a breakpoint, reached through accumulation of many rather than a single anomaly, a new paradigm must be framed that can successfully accommodate and explain both the old and new ('anomalous') information. According to Kuhn, such paradigm shift entails incommensurability, i.e., stark epistemological incompatibility between the old and the new paradigms.

Kuhn drew support for his model from selected cases from the histories of physics and astronomy. However, in time the Kuhnian original model underwent adjustments and modifications, (including Kuhn's own) and was subjected to various criticisms [i.e., (Bird, 2002, 2005; Chike, 2021; Levit & Hossfeld, 2022; Sankey, 1993; Shan, 2020; Weinert, 2014)]. Also, the historical patterns of some landmark discoveries in the life sciences ill-fitted this framework (Fry, 2016b). Without delving into the question whether the discoveries of archaea, RNA splicing, and ribozymes fully or even partially conform to the Kuhnian framework, this paper borrows from his theory the concepts of 'normal science activity' and 'anomaly' as useful constructs for the historiographies of these three discoveries.

It is difficult, and likely impossible, to craft a philosophical model of scientific progress that can serve as a universal template for the highly variable histories of discoveries of factual biological entities. Yet, comparative examination of the histories of specific cases may reveal shared patterns of selected discoveries. This paper presents detailed history of the discovery of archaea, (Sect. 2) and outlines the histories of the initial stages of the discoveries of RNA splicing and split genes and of ribozymes, (Sect. 3). It is shown that these three landmark discoveries were similarly instigated by detection of unforeseen anomalies during normal science activity.

1.2 Philosophical notions on discovery of new hypotheses and theories

Contrasting the seventeenth century Baconian and Newtonian ideas that theories are born out of observation and experiment, the nineteenth century thinkers John Herschel and William Whewell placed generation of new scientific hypotheses ahead of experiment [(Herschel, 1831); (Whewell, 1847, 1858); for historical perspective see (Schickore, 2006)]. Whewell also contended that the instance of discovery of new scientific hypothesis is an inimitable occurrence that cannot be formulated into algorithm ('maxim') for the generation of unrelated other discoveries [(Whewell, 1858), aphorism III, section IV, p. 44]:

Scientific discovery must ever depend upon some happy thought, of which we cannot trace the origin; some fortunate cast of intellect, rising above all rules. No maxim can be given which inevitably lead to discovery.

Whewell also argued that beyond the initial 'happy thought', a discovery must include elements of articulation, development, testing, and corroboration.

In their venture to delineate the logical boundaries of science, early and mid-twentieth century philosophers, notably those of the logical positivism school, excluded the generation of new scientific theories from the realm of rational reasoning. Rather, they considered the discovery of a novel theory to be a 'mental jump' [(Wisdom, 1952), p. 49], 'free mental creation', [(Einstein, 1954), p. 291] 'an irrational element', or 'creative intuition', [(Popper, 1959/2002) p. 8]. Novel theories were thus seen as products of creative invention and not as outcomes of observed facts [(Hempel, 1966), p. 15].

Framed by the logical empiricist Hans Reichenbach, a two contexts model of discovery and corroboration of theories had significant impact on modern thinking on the discovery of scientific theories. This model divided the process of theory formation into an initial *context of discovery* of theory creation, and a subsequent *context of justification* in which the theory is evaluated, tested, and epistemically substantiated [(Reichenbach, 1938) pp. 6–7]. Because Reichenbach considered the context of discovery to be an intuitive a-rational instance, he relinquished philosophical analysis of this phase, consigning its study to psychologists, historians, and sociologists. Also, considered unamenable to logical analysis, the context of discovery could not produce a general algorithm (a so-called "discovery machine") for the generation of other discoveries. Reichenbach claimed that unlike the irrationality of the context of discovery, the context of justification was governed by domain-neutral rules of logic and was thus open to normative formulation [(Reichenbach, 1951), p. 231]:

The act of discovery escapes logical analysis; there are no logical rules in terms of which a "discovery machine" could be constructed that would take over the creative function of the genius. But it is not the logician's task to account for scientific discoveries; all he can do is to analyze the relation between given facts and a theory presented to him with the claim that it explains these facts. In other words, logic is concerned only with the context of justification.

The two-contexts model was subjected to many criticisms and modifications. Some philosophers disputed its sharp division between discovery and justification, arguing that the two contexts were intertwined [(Feyerabend, 1970) pp. 70–71; (Feyerabend, 1993) Chapter 4; (Mowry, 1985; Nickles, 1980; Schaffner, 1974)].⁵ Other thinkers maintained that accurate description of the process requires addition of one (Curd, 1980; Kordig, 1978; Laudan, 1980; Nickles, 1980; Schaffner, 1980) or even two (Goldman, 1983) intermediate contexts.

Some philosophers challenged the idea of a completely a-rational context of discovery, suggesting that it is guided by variants of abductive logic (Aliseda, 2006; Gabbay & Woods, 2005; Hanson, 1958a, 1958b, 1961; Magnani, 2001, 2009; Paavola, 2004, 2006). Other thinkers, however, rejected abduction as logical basis for discovery (Achinstein, 1970, 1987; Frankfurt, 1958; Harman, 1965, 1968; Kapitan, 1992; Nickles, 1980). Computational and artificial intelligence methods were offered more recently as potential generators of logical rules of discovery and of discovery-producing algorithms (Addis et al., 2016; Džeroski et al., 2007; Langley, 2000; Sozou et al., 2017).

Conversely, some philosophers argued that the context of justification is not entirely logical and that it involves unreasoned steps of raising intermediary auxiliary hypotheses and assessing their testability (Nickles, 1980, 1985; Putnam, 1991).

Despite the various criticisms and modifications, Reichenbach's original two contexts model and variants thereof are still central to much of the philosophical thought on discovery of scientific hypotheses and theories (Hoyningen-Huene, 1987; Schickore & Steinle, 2006).

1.3 Philosophical ideas on discovery of theories in biology

Taking cue from philosophical thinking on discoveries in physics, the relatively few philosophical studies of discovery in biology have dealt with genesis and justification of new biological theories and not of factual entities.

1.3.1 *The repressor theory of negative regulation of gene expression*

Based on results of their experiment on the inducible expression of the enzyme β galactosidase in *E. coli*, Arthur Pardee, François Jacob, and Jacques Monod framed a theory of repressor-controlled negative regulation of gene expression (Pardee et al., 1959).⁶ Kenneth Schaffner contended that the construction of this theory was neither consistent with the standard hypothetico-deductive model nor with the scheme of sharply differentiated contexts of discovery and of justification (Schaffner, 1974). He argued first that this theory was deduced from experimental results and was not

⁵ First introduced by Charles Peirce as a new type of logical reasoning distinct from inductive or deductive logic, abductive logic infers from extant data and especially from unexpected irregular phenomena (Fann, 1970; Kraus, 2003).

⁶ This so-called PaJaMa experiment, named after the last names of its three authors: Pardee, Jacob, and Monod, was described in detail in the original paper (Pardee et al., 1959) and in many subsequent articles [i.e., (Schaffner, 1974; Pardee, 2002, 2003); (Fry, 2016a) pp. 371–381]

the product of irrational ‘creative intuition’ as postulated in the hypothetico-deductive model. Second, the contexts of discovery and justification⁷ similarly entailed empirical and extra-empirical factors and inferences, making both a single continuum guided by unitary logic (Schaffner, 1974). Reexamining the historical evidence, Marcel Weber contested the thesis of a single discovery-justification continuum. In his reading the logic behind the genesis of the repressor theory differed from the reasoning of its justification. He also argued that the repressor model was generated by analogy to prior cases of enzyme repression and not by deduction from experimental results [(Weber, 2005) pp. 55–63].

1.3.2 Proposed models of logic-driven discoveries of biological theories

Lindley Darden delineated several potential reason-based procedures for the discovery of explanatory theories for unresolved problems in genetics and molecular biology. Under one model, inexplicable observation or data are solved by application of solutions to settled analogous past problems (Darden, 1980, 1982). It was later conjectured that inter-field connections may be better sources for new theories than solved past problems (Darden, 1980, 2006; Darden & Craver, 2002). Another reason-based model employed hindsight from historical cases to mold allegedly sufficient and general non-algorithmic strategy for the discovery of biological mechanisms⁸ (Darden, 1991, 2002, 2006, 2009). Taking the change of genetic theory from Mendel to Morgan as a test case, Darden contended that although her account did not correspond to the reasoning that geneticists of the period actually employed, it still could have generated similar historical change (Darden, 1991). Weber criticized, however, both the failure of the proposed procedure to faithfully reconstruct historical scientific developments and its claims of sufficiency and generality [(Weber, 2005), pp. 63–71].

1.4 Kuhn’s philosophical thinking on observational and experimental discovery of factual entities

In a relatively less noticed article that coincided with the first edition of his *Structure of Scientific Revolutions*, (Kuhn, 1962b) Kuhn considered the largely neglected philosophical issue of theory-free empirical discoveries of factual entities (Kuhn, 1962a). Dissecting historical cases of unpredicted observational or experimental discoveries, Kuhn outlined their general structure. Although all the studied cases were taken from the histories of physics and chemistry, his conclusions are highly relevant to discoveries in the life sciences.

⁷ Schaffner limited the context of justification in this case to circumstantial supporting evidence derived from the PaJaMa experiment itself or from inferences thereof. Under a more comprehensive view, justification (corroboration) of the repressor theory was accomplished only after the repressor was physically isolated and its specific binding to both the target regulatory DNA segment and to the inducer molecule were experimentally demonstrated (Gilbert & Müller-Hill, 1966, 1967).

⁸ Such strategy which ideally generates a plausible theory, differs from algorithm that produces guaranteed correct theory.

Kuhn first made a distinction between two types of discovery. One kind were discoveries of objects already predicted by theory,⁹ whereas discoveries of the other type, (which is common in biology) were not predicted by prior theory.¹⁰ Because they are entirely unanticipated, discoveries of the second category frequently catch the scientific community by surprise (Kuhn, 1962a). Generalizing from examined historical cases, Kuhn identified three shared features of discoveries of objects that were not predicted by theory. (a) Such discoveries begin with observational or experimental findings of anomalies. Although other scientists may also encounter such irregularities, only individuals with required aptitude and gift fully notice the anomaly and pursue its significance.¹¹ (b) In a second extended phase of additional observations and experimentation, the investigator strains to turn the anomaly into an established part of nature. (c) In a third and final stage, the discovery and its broader significance are adjusted, adapted, and assimilated by the professional community. Kuhn contended that by accepting the full implications of the discovery, scientists gain a new look at what was previously known (Kuhn, 1962a).

2 The discovery of archaea: puzzle-solving normal science unpredictably heralded restructuring of the phylogenetic tree

This paper contemplates the history and philosophical implications of the discovery by Carl Woese in the 1970s of archaea, a third domain of the phylogenetic tree. This discovery led in time to reconstruction of the tree from a two-branched one, comprised of prokarya and eukarya, to a tripartite tree of bacteria, archaea, and eukarya. The principal focus of this contribution is the history and philosophical connotations of the progression of Woese's project from a problem-solving 'normal science' investigation to anomaly-instigated discovery.

Carl Woese, (1928–2012) studied physics and mathematics at Amherst College and completed in 1953 a PhD research thesis in biophysics at Yale University. After sojourns at the University of Rochester, Yale, and General Electric Research Laboratory in Schenectady NY, he joined in 1960 the University of Illinois at Urbana-Champaign which remained his academic home until the end of his career. There he served as professor at the Institute of Genomic Biology which was posthumously renamed in his honor in 2015 'The Carl R. Woese Institute of Genomic Biology'.¹²

⁹ Representative cases of this category were the discoveries of radio waves, the neutrino, and elements predicted to fill empty spots in the periodic table (Kuhn, 1962a).

¹⁰ Representative cases of this category were the discoveries of oxygen, the planet Uranus, and X-rays (Kuhn, 1962a).

¹¹ In an 1854 lecture at the University of Lille, Louis Pasteur famously underscored the import of the skill and talent of the scientist in exploiting an opportune observation: "In the fields of observation, chance favours only the prepared mind".

¹² Woese's scientific achievements were amply recognized in his lifetime: MacArthur Fellowship (1984); elected member of the National Academy of Sciences USA (1988) and of the Royal Society (2006); the Dutch Royal Academy of Science's Leeuwenhoek Medal (1992); the US National Medal of Science (2000); and the Crafoord Prize in Biosciences (2003). For biographical details and appreciation

The original aim of Woese's work at the University of Illinois was to establish evolution-based taxonomy of prokaryotes by comparing their ribosomal RNA sequences. Although this project was marked by evolutionary approach to construction of classification and by application of innovative methodology, it was nevertheless framed by theory and practice of 'normal' molecular biology and microbiology of the time. However, several years into the project an unexpected anomalous finding heralded the discovery of archaea, a hitherto unknown principal domain of the evolutionary tree. Substantiation and gradual acceptance of this discovery ended in the replacement of a previous scheme of a two-domain, (prokarya-eukarya) phylogenetic tree by a new model of tripartite, (bacteria-archaea-eukarya) tree. Kuhn's terminology of normal (puzzle solving) science and paradigm-changing scientific revolution, [(Kuhn, 1970a); (Kuhn, 1970b) pp. 23–42] suits the progression from normal science of constructing molecular taxonomy for prokaryotes to the consequential discovery of archaea. This paper contends that this and other discoveries in the life sciences conform with Kuhn's idea of anomalies as instigators of discoveries that are unpredicted by prior theories [(Kuhn, 1962a) and 1.3].

2.1 Woese's initial aim: construction of bacterial phylogenetic tree

On a backdrop of disconcerting failure to establish evolutionarily meaningful bacterial taxonomy, (2.1.2) Woese launched in the late 1960s a research project to construct molecular-based prokaryotic phylogenetic trees (Albers et al., 2013; Gold, 2014; Goldenfeld, 2014). To put this endeavor in context, the next section briefly summarizes the history of taxonomy of all forms of life and particularly of prokaryotes.

2.1.1 *Changing schemes of the tree of life*

The earliest taxonomic system apportioned each living thing to one of two kingdoms—Plantae (plants) or Animalia (animals) (Whittaker, 1969). This classification was capsized in the seventeenth century when Antonie van Leeuwenhoek used his simple microscope to auspiciously discover a hitherto hidden vast world of miniscule unicellular organisms that he named animalcules. Merging this new knowledge with a large body of embryological, palaeontological, and systemic data, Ernst Haeckel constructed in the nineteenth century a new phylogenetic tree of life (Dayrat, 2003). The root of this tree represented a presumed common primordial ancestor of all living things and its trunk branched into three super-kingdoms (domains): Protista (unicellular organisms that do not form tissues), Plantae, and Animalia. Because both bacteria and blue-green algae lacked cell nucleus, Haeckel merged them into a single group, the Monera which ranked below the Protista (Haeckel, 1866). Herbert Copeland later argued that the differences between Monera and Protista warranted their separation into two distinct super-kingdoms and his proposed

Footnote 12 (Continued)

of Woese's scientific contributions see (Sapp, 2008; Nair, 2012; Sapp & Fox, 2013; Goldenfeld, 2014; Gold, 2014; Koonin, 2014; Moore, 2014; Luehrsens, 2014; Quammen, 2018).

version of the phylogenetic tree had thus four super-kingdom branches: Monera, Protista, Plantae, and Animalia (Copeland, 1938, 1947). Robert Whittaker classified Fungi as another independent super-kingdom such that his evolutionary tree comprised of five super-kingdom: Monera, Protista, Fungi, Plantae, and Animalia (Whittaker, 1959, 1969). However, the version of the phylogenetic tree that became dominant by the second half of the twentieth century was a parsimonious scheme of mere two super-kingdom branches, Prokarya and Eukarya, terms that were originally introduced by the French biologist Édouard Chatton, (Chatton, 1938). Under his classification, prokaryotes comprised of all unicells that were devoid of nucleus. However, in his terminology, eukaryotes were only monocellular nucleus bearing protists, (protozoa).¹³ Roger Stanier later expanded the eukaryotic super-kingdom to also include multicellular plants and animals (metaphyta and metazoa). Thus, whereas under Stanier's taxonomy the prokaryotic super-kingdom of anucleate bacteria remained unchanged, his eukaryotic domain encompassed every nucleus-containing mono- or multicellular organism (Stanier, 1961; Stanier & van Niel, 1962; Stanier et al., 1963), for historical reviews see (Katscher, 2004; Sapp, 2005). This dichotomous scheme of the tree of life was consensually adopted by biologists at the second half of the twentieth century (Corliss, 1989).

2.1.2 Bacterial taxonomy appeared to be indeterminable

Following the discovery of monocellular microscopic 'animalcules' in the seventeenth century, it was debated for almost two centuries whether they were animals or plants. The father of modern taxonomy, Carolus Linnaeus placed them in the 1767 edition of his *Systema Naturae* under the group 'Verms' (worms), class 'Chaos' and species 'Chaos infusorium'.¹⁴ Various classification systems that were introduced in the subsequent two centuries were first based on distinguishing morphologies of the microscopic organisms and, starting in the mid- to late nineteenth century, on some of their characteristic metabolic properties and distinctive biochemical constituents of different bacteria."

The Danish naturalist Otto Friedrich Müller arranged the animalcules, (named by then 'infusoria') into genera and species (Müller, 1773). Believing that the infusoria comprised of diverse fixed animal species,¹⁵ the German zoologist Christian Gottfried Ehrenberg classified them by shape into 22 families (Ehrenberg, 1838). Others who also adhered to the idea that infusoria comprised of fixed animal species later proposed simpler classification systems (Dujardin, 1841; Perty, 1852). Botanists of the nineteenth century rejected the perception of bacteria as animals arguing

¹³ For a history of the evolution of Chatton's classification and terminology see (Katscher, 2004).

¹⁴ This naming reflects the 18th century perception of a chaotic universe of microscopic organisms.

¹⁵ The belief that the bacterial world comprised of multitude of fixed species, (monomorphism) was thrown into turmoil in the second half of the 19th century. An alternative view (polymorphism) claimed that there is only a single or very few bacterial species and that their different shapes and even transformation into yeast and fungi are consequences of changing conditions. Robert Koch ultimately settled the dispute between monomorphism and polymorphism on the side of the former. He showed in classical experiments that isolated single types of bacteria retained their morphologies and properties over many generations [for a history of the debate and Koch's experiments see (Fry, 2020)].

instead that they were microscopic plants akin to fungi and algae. The German botanist Ferdinand Cohn classified bacteria by shape into four main groups: Sphaerobacteria (spherical cells, later named cocci), Bacterium (rod-shaped), Desmobacteria (filament-like), and Spirobacteria (screw-shaped, later termed spirilla) (Cohn, 1872, 1875, 1876). Other investigators proposed alternative bacterial classification systems that were also based on cell morphology and metabolism. However, failure to reach agreed biological definition of bacteria¹⁶ and to establish evolutionarily meaningful bacterial taxonomy, (Breed et al., 1957; Buchanan, 1925; Migula, 1897, 1907; Niel, 1955; Stanier & Niel, 1941) left leading mid-twentieth century microbiologists with a sense that construction of bacterial phylogenetic tree was beyond reach.¹⁷

Confronting this impasse, Woese undertook to establish genealogical tree of bacteria based on evolutionary changes in the nucleotide sequence of their ribosomal RNA (rRNA). Because bacteria were evolutionarily much more ancient than eukaryotes,¹⁸ the project was also motivated by hope of identifying a common root of the bacterial lineage. This could in turn shed light on the nature of the last universal common ancestor (LUCA) or last universal ancestor (LUA) which Woese named the 'progenote' (Woese, 1970; Woese & Fox, 1977a).

2.1.3 Origins of Woese's project

At an early phase of his work as independent researcher, Woese attempted to tackle the problem of the evolution of the genetic code. Briefly, he asked how recognition had developed during the earliest stages of life on earth between specific amino acids and their corresponding transfer RNA (tRNA) carriers (Woese, 1967). Although this endeavor proved to be unproductive, its underlining idea of evolution at the molecular level led Woese to the notion that the evolution of bacteria could be clocked by changes in component(s) of their universal and evolutionarily conserved protein biosynthesis machinery. This general idea was the basis for his ensuing successful effort to construct evolution-based taxonomy of bacteria that were then thought to belong to a single prokaryal domain. The project had at its onset both a formerly framed theoretical basis, (Zuckerkindl & Pauling, 1965a, 1965b) and an effective RNA sequencing methodology previously developed by Fred Sanger (Sanger et al., 1965). Specifically, relative evolutionary distances between different bacterial species and phyla were derived from quantified changes in the nucleotide sequences of their rRNA molecules. Graphically positioned according to their relative evolutionary distances, bacterial taxa appeared as branches of a constructed phylogenetic tree.

¹⁶ "Any good biologist finds it intellectually distressing to devote his life to the study of a group that cannot be readily and satisfactorily defined in biological terms; and the abiding intellectual scandal of bacteriology has been a clear concept of a bacterium." [(Stanier & Niel, 1962) p. 2].

¹⁷ "An eminent contemporary bacteriologist, van Niel, who is noted for his taxonomic studies on several groups of bacteria, has expressed the opinion that it is a waste of time to attempt a natural system of classification for bacteria, and that bacteriologists should concentrate instead on the more humble practical task of devising determinative keys." [(Stanier et al., 1957) p. 296].

¹⁸ Beginning in the early 1950s, paleobiologists identified Precambrian fossils of >3 billion years old bacteria [evidence that was already available at the onset of Woese's work was summarized by (Schopf, 1975)].

2.1.4 *The theoretical basis to Woese's bacterial phylogenetic project*

The French biologist Emile Zuckerkandl working in the first half of the 1960s in Linus Pauling's Caltech laboratory, catalogued substituted amino acids in homologous globin chains from different species.¹⁹ Positing that any particular protein evolves over time at a fairly constant rate, Zuckerkandl and Pauling hypothesized that the number of substituted amino acid in homologous proteins from two different species was proportional to the evolutionary time that elapsed since their divergence from a last common ancestor [(Zuckerkandl & Pauling, 1965a, 1965b), for history of their work and hypothesis see (Morgan, 1998)].²⁰ Under this so-called 'molecular clock hypothesis', comparative sequence data were used to build phylogenetic trees whose taxon-representing branches were placed at evolutionary temporal distances from one another. Woese applied a modified form of this hypothesis as the theoretical basis for his construction of bacterial genealogical trees. Reasoning that proteins are not universally distributed and that they do not necessarily preserve constant function throughout evolution, Woese compared instead changes in highly conserved nucleotide sequences of 16S rRNA molecules of different bacterial species. This RNA was considered a more reliable evolutionary chronometer because: (a) a clocklike behavior was guaranteed by the nearly random nature of changes in its sequence. (b) Sequence changes in paired species were proportional to their evolutionary distances. (c) The size of rRNA was large enough to yield quantitatively dependable information.

2.1.5 *Sanger's RNA sequencing technique provided Woese with a vital experimental tool*

To determine nucleotide sequences of bacterial rRNA, Woese and associates adopted a technique for the sequencing of short RNA fragments, (Sanger et al., 1965). Employing this method, they compared partial sequences of rRNA molecules from different species of bacteria. Extents of rRNA sequence variance between paired bacterial species were used to calculate their relative evolutionary distances in constructed phylogenetic trees (2.1.7).

Figures 1A and 1B (I) illustrate main elements of Woese's methodology. Briefly, to label their RNA, bacteria were fed a radioactive ³²P isotope of phosphorous. Next, a desired ³²P-labeled RNA species, (i.e., 5S rRNA, 16S rRNA etc.,) was isolated

¹⁹ Globins are proteins that form hemoglobins together with a non-proteinaceous heme molecule. Starting by comparison of patterns of tryptic oligopeptides of homologous globins from different species, (Zuckerkandl et al., 1960) Zuckerkandl later progressed to finer comparisons of their amino acid sequences (Zuckerkandl & Schroeder, 1961; Zuckerkandl & Pauling, 1962; Pauling & Zuckerkandl, 1963).

²⁰ About seven years before Zuckerkandl and Pauling introduced their hypothesis, Francis Crick raised an abstract idea that had no experimental support, of applying molecular evolutionary clock to build taxonomies: "Biologists should realise that before long we shall have a subject which might be called 'protein taxonomy' - the study of the amino acid sequences of the proteins of an organism and the comparison of them between species. It can be argued that these sequences are the most delicate expression possible of the phenotype of an organism and that vast amounts of evolutionary information may be hidden away within them." [(Crick, 1958) p. 142].

and then digested by T1 ribonuclease (RNase) that specifically nicked it terminally to each of its randomly situated guanine (G) residues (Fig. 1A). Oligomeric products of the enzymatic digestion were placed on paper and separated according to their lengths, base composition, and sequences by electrophoresis in two dimensions. Viewed by autoradiography, differently migrating spots of radioactive RNA oligomers formed typical ‘fingerprint’ patterns such as shown in Fig. 1B(I). In a final step, each spot was cut out, partially digested by ribonucleases other than T1 and its nucleotide sequence was determined by one dimensional electrophoresis of the digestion products (Fig. 1A).

2.1.6 Woese first presented his core ideas in a 1969 letter to Crick

Because of their shared interest in the evolution of the genetic code, Crick and Woese corresponded on occasion in the 1960s and 1970s.²¹ Pertinent to this study is a 1969 letter²² in which Woese presented his thoughts on mapping evolutionary distances between bacterial species based on sequence changes in their genes:

If we ever to unravel the course of events leading to the evolution (i.e., simplest) cells, I feel it will be necessary to extend our knowledge of evolution backward in time by billion years or so – i.e., backward into the period of actual “Cellular Evolution”. There is a possibility, though not a certainty, that this can be done by using the cell’s “internal fossil record” – i.e., the primary structure of various genes.

Woese further speculated that evolutionary change is best reflected by changes in sequences of RNA component(s) of the translation machinery:

The obvious choice of molecules here lies in the components of the translation apparatus. What more ancient lineages are there? A priori it seems impossible to evolve any structural gene without the capacity to translate the gene - making the evolution of some rudimentary translation machine necessarily a very early happening. Hopefully that machine was a direct lineal ancestor (both functionally and structurally) of the present one. Also, I feel (and you may too) that the RNA components of the machine hold more promise than (most of the) protein components.

While writing this letter Woese had already decided to use Sanger’s method for the sequencing of bacterial RNA.²³ However, being aware of his lack of the required technical expertise, he asked for Crick’s help in recruiting: “...some energetic young

²¹ Facsimiles of the Crick-Woese letters (1962-1973) are in the Wellcome Collection Web site: <https://wellcomecollection.org/works/ej5p7b5z/items?canvas=15>

²² Facsimile of this June 24, 1969, letter is in the Wellcome Collection: <https://wellcomecollection.org/works/ej5p7b5z/items?canvas=40>

²³ The molecular biologist Sol Spiegelman, Woese’s colleague in the Department of Microbiology, University of Illinois at Urbana-Champaign was recruited in 1969 by Columbia University. As Woese mentioned in his letter to Crick, upon departing for New York Spiegelman gave Woese his sequencing equipment. Woese’s acceptance of this apparatus suggests that by that time he had already made up his mind to apply Sanger’s methodology for his bacterial taxonomy project.

product of Fred Sanger's lab,²⁴ whose scientific capacities complement mine". Eventually, however, no such a person was required as his group became proficient in the use and even the improvement of Sanger's short RNA sequencing technique.

2.1.7 Variance in rRNA sequences was used as molecular clock to build bacterial phylogenetic trees

Starting in the early 1970s, Woese and associates embarked on yearslong experimental efforts to build genealogical trees for bacteria. Phyla and species were respectively represented as branches and sub-branches of the tree. The branches and sub-branches were placed according to their relative evolutionary distances that were proportional to the degrees of disparity between their rRNA sequences. Since ribosomes are components of the universal and indispensable translation apparatus, their RNA constituents were thought to have changed during evolution at more restrictive and slower rate than any other gene/protein.²⁵ It was thus reasoned that extents of change in nucleotide sequences of rRNA of different bacteria best reflect elapsed evolutionary time. Additionally, it was theoretically possible that monitoring rRNA sequences of diverse species could uncover evolutionarily early molecules ('molecular fossils') that may potentially unveil attributes of ancient versions of the translation machinery.

Because of the relative ease of the isolation, characterization, and sequencing of shorter RNA molecules, Woese initially compared in different bacterial species sequences of the ~120 nucleotide-long molecules of 5S ribosomal RNA²⁶ (Sogin et al., 1972). It was soon realized, however, that these molecules were too short to yield large enough numbers of mutational changes. He therefore switched to comparisons of sequences in different bacteria of the ~1500 nucleotides-long 16S RNA component of the smaller ribosomal subunit.²⁷ Because this RNA species exists in every organism and its function is conserved, it was deemed ideal evolutionary 'chronometer' that reflects true line of descent. This premise could potentially be wrong if horizontal (interspecies) gene transfer was a major contributor to genetic variation. However, evidence indicated that horizontal transfer of genes had

²⁴ Crick and Sanger were colleagues in the Cambridge MRC Unit.

²⁵ This premise was supported by prior results of RNA-DNA hybridization experiments that indicated that despite their conservation, (Yankofsky & Spiegelman, 1963) rRNA molecules from different organisms did diverge as reflected by the lower stability of heterologous (inter-species) ribosomal RNA-DNA hybrids relative to homologous (intra-species) hybrids (Moore & McCarthy, 1967; Bendich & McCarthy, 1970; Pace & Campbell, 1971).

²⁶ Discovered in 1964, (Rosset et al., 1964) 5S RNA, ('S' stands for Svedberg unit - a measure of the rate of sedimentation of a macromolecule under high centrifugal acceleration), was later found to be a part of the large subunit of every bacterial, archaeal, and eukaryotic ribosome. Full sequences of 5S RNA molecules from two bacterial species were determined prior to Woese's work (Brownlee et al., 1967; DuBuy & Weissman, 1971).

²⁷ Bacterial ribosomes (70S) comprise of two nucleoprotein subunits of 50S and 30S. The RNA components of these subunits are 23S + 5S for the 50S particle and 16S for the 30S subunit. Larger ribosomes (80S) of eukaryotic cells comprise of 60S and 40S nucleoprotein subunits whose respective RNA components are 28S + 5.8S + 5S and 18S.

having its individual nutritional requirements. It was thus technically nontrivial to establish conditions for cell growth and division at rates that allowed incorporation of high enough levels of ^{32}P into their rRNA. Once grown and properly labeled, cells were disintegrated and their radioactive 16S rRNA molecules were isolated. Following enzymatic digestion of the rRNA, product oligonucleotides were separated by 2D electrophoresis, viewed by autoradiography, isolated, and their nucleotide sequences were determined (Figs. 1A, B (I)). Typical comparison of a set of 16S rRNA oligomeric sequences of eight different species is shown in Figs. 1B (II-IV). First, frequencies of occurrence of identical fragment sequences were tabulated, (Figs. 1B (II)). Next, binary association coefficients, S_{ab} , were calculated and charted (Fig. 1B (III)).²⁹ Finally, rRNA sequences of the different species were graphically arranged by their S_{ab} values in a dendrogram of relative evolutionary distances between species (Fig. 1B (IV)).

For more than a decade, Woese and his coworkers tenaciously applied the described multi-step procedure for numerous bacterial species.³⁰ Years later Woese described his tedious routine (Woese, 2007):

My job was to determine the complete sequence of every oligonucleotide of significant length (five or more nucleotides) in primary pattern,³¹ which required the aforementioned “secondary” patterns. These in turn were created by removing little snippets of paper in the appropriate places in the original electropherogram and further digesting the oligonucleotide(s) therein (*in situ*) with one or a few ribonucleases of different cutting specificities than that of T1 RNase [...] From the one or several “secondary” taken from a primary spot, the exact sequence of the oligonucleotide(s) in the corresponding primary spot could (almost always) be deduced [...]

“Reading” a Sanger pattern was painstaking work, requiring a good fraction of the day to work up a single primary,” something I at the time had been doing for several days a week off and on for a long time. It was routine work, boring, but demanding full concentration. (There were days when I would walk home from work saying to myself: “Woese, you have destroyed your mind again today”).

The project was modestly opened in 1972 by comparison of catalogues of sequences of oligomeric products of T1 RNase digestion of 16S rRNA from *Escherichia coli* and *Bacillus megaterium* (Pechman & Woese, 1972). Soon thereafter Woese and

²⁹ Degree of genetic relatedness of two compared species was determined under the assumption that sequence identity between oligomers of 6 nucleotides or more is highly indicative of their common ancestry. Extents of genetic closeness of paired bacterial species were expressed by their association coefficient, S_{ab} , defined as the fraction of the total number of nucleotides in any pair of catalogues A and B of oligomers of 6 or more nucleotides that are found in sequences common to the two catalogues [(Woese, 1987) p. 228].

³⁰ Three to four weeks were needed to complete a catalogue of 16S rRNA oligomer sequences from one bacterial species (Sapp & Fox, 2013).

³¹ Primary pattern is that of the 2D electrophoretically separated oligonucleotides products of T1 RNase digestion of 16S rRNA. “Secondary” is the 1D electrophoretic resolution of products of digestion of the primary oligomers by ribonucleases other than T1 (see Fig. 1A).

associates modified and improved some aspects of the RNA sequencing method, (Uchida et al., 1974; Woese et al., 1976) and the scope of analyzed bacterial species was gradually expanded. Sequences of 16S rRNA were determined for diverse bacterial families that included, among others, Enterobacteriaceae, (Woese et al., 1974; Zablen et al., 1975) Cyanophyta, (Blue-Green algae), (Doolittle et al., 1975) photosynthetic bacteria, (Zablen & Woese, 1975) mesophilic and thermophilic Bacillaceae, (Fox et al., 1977b; Woese et al., 1976) and Mycoplasma (Woese et al., 1980b).³² By 1975 the Woese group had sequenced 16S rRNA oligonucleotides from about 30 bacterial species and by the end of that decade the number grew to 100 species (Woese, 2013).³³ Calculated association coefficients placed different bacterial species at relative evolutionary distances from one another, allowing construction of genealogical trees.³⁴ 16S rRNA-based bacterial systematics divide the eubacterial world into ten divisions, (phyla) each with its own subdivisions (Woese, 1987; Woese et al., 1985).

2.1.8 *What did the phylogeny project achieve and what remained unsolved*

At the time that Woese launched his phylogeny project in the late 1960s microbiologists had abandoned hope of establishing methodical taxonomy of bacteria. Instead, practitioners of medical, agricultural, and industrial bacteriology unsystematically grouped bacteria by their morphology and metabolism (2.1.2). Woese held, however, that rational bacterial classification system must be informed by evolution. Thus, taxa and individual species should be arranged in a phylogenetic tree according to their relative evolutionary distances. He further reasoned that such distances are best clocked by monitoring inter-species changes in the 16S rRNA elements of the universal and conserved translation machinery (2.1.7). It soon became evident that 16S rRNA-based taxonomies conflicted with phenotype-based phylogenetics. Thus, 16S RNA systematics identified some phenotypically similar bacteria as members of different phyla. Conversely, 16S rRNA genealogy categorized some phenotypically dissimilar species under the same divisions and subdivisions. Later replacement of the Sanger/Woese RNA sequencing method by newer techniques of sequencing full rRNA genes, (Brosius et al., 1978; Carbon et al., 1978), greatly accelerated the accumulation of 16S rRNA sequences and allowed growth of more detailed phylogenetic trees. Despite early resistance to the 16S rRNA-based phylogeny, it was gradually accepted as the standard method for estimating evolutionary distances and as the most reliable basis for building phylogenetic trees. This acceptance was highlighted in 2001 when Bergey's Manual of Systematic Bacteriology, the benchmark of bacterial classification, changed its phenotype-based systematics to 16S rRNA-based one

³² Following the introduction of the Sanger DNA sequencing technique, (Sanger et al., 1977) the number of sequenced 16S rRNA genes of different species and their deduced rRNA transcripts, grew at accelerating pace such that by 1994 sequences of 16S rRNAs were determined for more than 1500 prokaryotes (Olsen et al., 1994).

³³ For lists of the analyzed bacterial families and their places in phylogenetic trees see (Fox et al., 1980; Woese, 1987; Olsen et al., 1994).

³⁴ These trees represented relative evolutionary distances between species and phyla with no determinate time scale.

(Garrity et al., 2001).³⁵ In a broader context, Woese's powerful molecular approach for tracing the evolution of cells and their organelles invigorated the field of evolutionary biology at large.

A second objective of Woese's molecular phylogeny project remained, however, unfulfilled. It was anticipated that tracing of the root of the evolutionary tree would expose the origin of all living cells, an entity that Woese and Fox named the progenote (Woese & Fox, 1977a).³⁶ Already in his 1969 letter to Crick¹⁸ Woese insinuated that tracking bacterial evolution back to its beginning may uncover a primordial living form at the cusp of cellular evolution:

If we are ever to unravel the course of events leading to the evolution of the procaryotic (i.e., simplest) cells, I feel it will be necessary to extend our knowledge of evolution backward in time by a billion years or so -- i.e., backward into the period of actual "Cellular Evolution.

In subsequent years Woese, (Woese, 1998b, 2000, 2002) and many other researchers conjectured extensively on the possible nature of the progenote. However, to this day no common root has been identified for the bacterial, archaeal, and eukaryotic phylogenetic trees, and the nature of the progenote remains elusive.³⁷

2.2 The discovery of archaea: an unforeseen model-changing outcome of the bacterial phylogeny project

Archaea, a principal third branch of the phylogenetic tree, were unexpectedly discovered in 1976/77 during Woese's bacterial molecular taxonomy project. This section describes this discovery, the controversy that it raised, and early phases of its substantiation and acceptance up to the early 1990s.

2.2.1 *The start: experimenting with methanogenic prokaryotes*

Construction of an all-inclusive phylogenetic tree necessitated a broad base of 16S rRNA sequences from diverse bacterial species. Provided with different strains of bacteria by various bacteriologists, the Woese laboratory amassed by 1976 sequences of 16S rRNA from about 60 bacterial species and from few eukaryotes.³⁸

³⁵ Capacious whole genome data that were gathered in the recent two decades bred alternative bacterial classification systems. Yet, no consensus had been reached yet on taxonomic and nomenclatural standards for bacteria (Hugenholtz et al., 2021).

³⁶ As originally proposed, (Woese & Fox 1977a) the hypothetical progenote ancestor of all forms of life existed before genotype and phenotype became firmly coupled [for the presumed characteristics of the progenote and its relationship to the putative Last Universal Common Ancestor (LUCA) see (Gogarten, 2019)].

³⁷ Very recent experimental results pointed to members of the Asgard archaeal phylum as likely ancestors of eukaryotes, (2.2.6 below). However, these ancient cells are positioned at the archaea/eukaryotes branching point, downstream from the hypothetical progenote ancestor.

³⁸ Because of their larger size and high content of modified nucleotides, 18S rRNA chains of the smaller eukaryotic ribosomal subunit were harder to sequence than their analogous prokaryotic 16S rRNA chains.

One bacteriologist with whom Woese had conferred was Ralph Wolfe, his departmental colleague and expert on methanogenic bacteria. These morphologically heterogeneous and strictly anaerobic prokaryotes metabolically produce methane by reducing carbon dioxide in the presence of hydrogen. Although Woese was keen to add 16S rRNA of methanogens to his catalogue, their standard culturing conditions precluded presence of high enough levels of the ^{32}P isotope. George Fox, then a postdoctoral associate in the Woese laboratory, discussed the problem with William (Bill) Balch, a doctoral student in the Wolfe laboratory.³⁹ Ultimately, Balch devised a method to grow methanogens anaerobically in pressurized atmosphere of carbon dioxide and hydrogen (Balch & Wolfe, 1976). Importantly, these conditions allowed addition of the necessary levels of ^{32}P without exposure to oxygen or contamination (Fox, 2015; Sapp & Fox, 2013; Wolfe, 2014).

2.2.2 *An anomaly set the stage for a breakthrough discovery: sequences of methanogenic 16S rRNA oligomers differed from those of known prokaryotes*

A moment of revelation occurred in June 1976 when Woese had first looked at a primary fingerprint pattern of oligonucleotides that Linda Magrum and George Fox derived from 16S rRNA of a methanogenic strain. Absent from this pattern were two spots of modified oligonucleotides that hallmarked every hitherto analyzed prokaryotic 16S rRNA. Alerted by this anomalous pattern, Woese went on to determine nucleotide sequences of the resolved methanogenic rRNA oligomers. Strikingly, sequences of some oligomers were unlike those of known prokaryotes, and several were typical to eukaryotes. Here is his retrospective description of the anomaly and his sense of a significant discovery [(Woese, 2007), bold in the origin]:

The more oligos I sequenced, the less prokaryotic it felt, as signature oligo failed to turn up. However, a number of them were still there, as, surprisingly, were some oligos from the eukaryotic signature. [...] I rushed to share my out-of-biology experience with George, a skeptical George Fox to be sure.⁴⁰ [...] whatever skepticism he initially evinced quickly dissipated. Yes, he agreed, there probably was something else out there: it wasn't just prokaryotes and eukaryotes all the way down. That was a heady thought, novel enough that we sensed trouble in trying to convince other biologists of that idea. Little did we know how **much** trouble there would be.

Ralph Wolfe also described in his own words the surprise and wonder that the unanticipated results evoked (Wolfe, 2014):

³⁹ Fox and Balch have known one another since 1973 when they attended together a course in microbial ecology at the Woods Hole Marine Laboratory.

⁴⁰ Fox's skepticism was due to the trivial possibility that because of mix-up, the source of the examined 16S rRNA was not the methanogenic microorganism but some other organism(s) (eukaryotic? eukaryotic/prokaryotic mixture?). This alternative explanation was excluded after rRNA chains from the original methanogen and from a different methanogenic species yielded again exceptional non-prokaryotic and non-eukaryotic fingerprint patterns and nucleotide sequences.

...his [Woese's] response was "Wolfe, these methanogens are not bacteria." "Of course they are Carl; they look like bacteria." "They are not related to any bacteria I've seen."

Because Woese had spent 10 years developing the method and analyzing the 2-D chromatographic patterns of T₁ endonuclease digestion patterns of ³²P-labeled 16S rRNA from 60 different bacteria, he was easily able to discern that the methanogens were different!

Although evidence was rather slim at that early stage, Woese nonetheless was quick to proclaim that a new form of life has been discovered.⁴¹ The ciliatologist David Nanney and Linda Magrum of the Woese laboratory independently suggested to name these microorganisms 'archaebacteria'. This early name was later changed to 'archaea', a term that this paper henceforth uses. To underscore the distinction between archaebacteria (archaea) and non-archaeal prokaryotes, the latter were christened eubacteria.

2.2.3 *The discovery of archaebacteria was made public*

Woese and Fox presented their first molecular evidence for the distinctiveness of archaea from eubacteria and eukaryotes in a short paper in the November 1, 1977 issue of the *Proceedings of the National Academy of Science USA* (Woese & Fox, 1977b). The evidential centerpiece of this report was a table of measured values of association coefficients (S_{ab}) of paired 18S rRNA sequences from three eukaryotic species; 16S rRNA of five eubacteria; chloroplasts rRNA derived from *Lemna* aquatic plant;⁴² and 16S rRNA of four methanogenic archaeal species. These data clearly indicated that degrees of sequence similarity were significantly higher for intra-group paired sequences (i.e., eukaryotic with eukaryotic species, etc.) than for inter-group pairs (i.e., eukaryotic with eubacterial species, etc.). Most importantly, whereas 16S rRNA sequences of the four archaeal species had high degree of similarity, their resemblance to eubacterial or eukaryotic rRNA sequences was much lower. Based on these findings, Woese and Fox boldly proposed that the then prevailing prokaryotes/eukaryotes bipartite tree of life should be replaced by a tripartite tree that branched into 'three aboriginal lines of descent':⁴³ (a) Eubacteria that comprised of all 'typical' bacteria; (b) Archaebacteria (which at that early stage

⁴¹ This assertion was hinted by earlier unpublished finding in the Woese laboratory that the sequence of methanogen 5S RNA was more similar to that of eukaryotes than to prokaryotes (Sapp & Fox, 2013)

⁴² As noted, (Woese & Fox, 1977b) the similarity between chloroplast and bacterial rRNA sequences provided supporting molecular evidence to the hypothesis that the origin of chloroplasts were protobacteria which after engulfment by proto-eukaryotic cells, became stable endosymbionts.

⁴³ First termed 'Primary Kingdoms' or 'Ur kingdoms', (Woese & Fox, 1977b) these three lines of descent were later renamed 'Domains' (Woese et al., 1990).

comprised of only methanogenic bacteria)⁴⁴ and; (c) Eukaryotes (Eukarya) (Woese & Fox, 1977b).⁴⁵

2.2.4 *The claim that archaea constitute a new phylogenetic domain met with skepticism and hostility*

Concurrently with the publication of the formal technical report, (Woese & Fox, 1977b) the *New York Times* published a front page news story under the heading: “*Scientists Discover a Form of Life that Predates Higher Organisms*”⁴⁶ (Fig. 2). This audaciously headlined article prompted many microbiologists, including the 1969 Nobel Prize laureate Salvador Luria, to call and admonish Ralph Wolfe (Wolfe, 2014):

The immediate response of the scientific community to the press release was negative with disbelief and much hostility, especially among microbiologists. Scientists were suspicious of scientific publication in newspapers, and only a very few were familiar with the use of 16S rRNA oligonucleotides to define relationships among organisms. Among the phone calls that I received the morning of November 3, the one by S. E. Luria was the most civil and free of four-letter words. Luria was a Professor of Microbiology, when I joined the Department at Illinois in 1953 and had later moved to MIT.

Luria: “Ralph, you must dissociate yourself from this nonsense, or you’re going to ruin your career!”

“But, Lu, the data are solid and support the conclusions: they are in the current issue of PNAS.”

Luria: “Oh yes, my issue just arrived.”

“If you would like to discuss the paper after you have had a chance to look at it, give me a ring.”

He did not call again. I wanted to crawl under something and hide.

Secondary sources suggested that many microbiologists, who have also mostly relied on the *Times* story and not on the *Proceedings* paper, shared Luria’s negative impressions of the discovery (Morell, 1997). In general, Woese was regarded by mainstream researchers of the time as a marginal individual doing dubious science.⁴⁷

⁴⁴ Eubacteria and archaeobacteria were both defined as prokaryotes based on their unicellularity, absence of membrane-enveloped nucleus, and lack of intracellular organelles.

⁴⁵ Prior to the negative reception of the published paper, (2.2.4 below) it encountered resistance already upon submission to the *Proceedings*. One reviewer, criticizing the methodology of the work and rebuffing the claimed three-domain phylogenetic tree, recommended that the paper be rejected (Pace et al., 2012).

⁴⁶ Lyons, Richard D. *New York Times*, Vol. CXXVII, No. 43748, November 3, 1977; <https://timesmachine.nytimes.com/timesmachine/1977/11/03/issue.html>. The *Times* was informed about the discovery by NASA, which funded in part Woese’s research work.

⁴⁷ “Woese’s solitary years at his light table [used to visualize autoradiograms] had left him with a reputation as an odd person, “a crank, who was using a crazy technique to answer an impossible question” as one researcher put it. “His tiny snippets of rRNAs were considered too fragmentary to be reliable

Soon after the initial description of the four methanogenic archaeal species, (Woese & Fox, 1977b) the Woese and Wolfe laboratories collaborated to construct a phylogenetic tree from 16S rRNA sequences of 10 methanogenic archaeal species and 3 eubacterial reference species (Fox et al., 1977a). Despite being phenotypically heterogenous, all ten archaea proved to be members of sub-branching single line of descent, distinctly different from the eubacterial one (Fig. 3A). Regardless of these and subsequent corroborative results, for the next several years microbiologists and evolutionary biologists continued to contest Woese's three-domain phylogenetic tree [for description of Woese's struggles with largely behind-the-back criticisms see (Morell, 1997)]. Yet, the weight of progressively accumulating evidence gradually persuaded most interested scientists to accept archaea as an independent line of descent separate from the eubacterial and eukaryotic phylogenetic domains (2.2.5). Even so, some prominent evolutionists persisted in questioning the validity of rRNA-based classification and held on to the model of bipartite (prokaryotic and eukaryotic) evolutionary tree (2.2.6).

2.2.5 *Amassed evidence buttressed the perception of archaea as a distinct phylogenetic domain*

After his and Fox's first report of archaea, (Woese & Fox, 1977b) Woese endeavored to experimentally establish this group as a domain separate from Eubacteria and Eukarya⁴⁸. Indeed, he and others built up a body of experimental data that significantly strengthened this premise. The notion of archaea as distinctive evolutionary domain was independently advanced by discoveries of German and American laboratories⁴⁹ of their idiosyncratic biochemical features. Thus, even before they have been defined as archaea, cell walls of some methanogenic, halophilic, and thermoacidophilic prokaryotes,⁵⁰ were found to have exclusive components, (Kandler & Hippe, 1977; Kandler & König, 1978). Also, these microorganisms had distinguishing ether-linked membrane lipids, (Kates et al., 1966; Langworthy et al., 1972; Tornabene & Langworthy, 1979) eukaryotic-like components of RNA polymerase, (Zillig et al., 1978, 1979, 1980) and non-bacterial translation elongation factor (Kessel & Klink, 1980, 1982). In parallel, the Woese team showed that the 16S

Footnote 47 (Continued)

indicators of evolutionary relationships" said [Norman] Pace. The Yale University molecular biologist Alan Weiner recalled that many leading biologists thought Woese was "crazy," and that his RNA tools couldn't possibly answer the question he was asking." (Morell, 1997).

⁴⁸ Parallel to this main objective, Woese also continued his original bacterial taxonomy project, (Woese et al., 1976; Fox et al., 1977b; Fox et al., 1980; Woese et al., 1980b; Woese et al., 1985; Woese, 1987) and studied secondary structure of 16S rRNA (Woese et al., 1983; Gutell et al., 1985)

⁴⁹ Prominent among those were the University of Munich laboratory of Otto Kandler, (archaeal cell wall chemistry) Wolfram Zillig's laboratory at the Munich Max-Planck Institute, (archaeal DNA dependent RNA polymerase) and the South Dakota University laboratory of Thomas Tornabene, (archaeal membrane lipids).

⁵⁰ Found in nature in salt-saturated brine, halophilic archaea, (haloarchaea), require high concentration of NaCl (>2 M) for their survival and growth. Thermoacidophile archaea grow in acidic environment (pH 2-4) at high temperatures (80 to >100 °C).

rRNA sequence of the halophile *Halobacterium halobium* marked it as a member of the archaeal group (Woese et al., 1978). Integrating evidence on the emblematic sequences of archaeal 5S and 16S rRNA and on the characteristic chemistry of their cell walls and lipids, Woese and associates proposed that in addition to methanogens the archaeal domain also included halophiles and thermoacidophiles (Woese et al., 1978).⁵¹ Subsequently gathered experimental data amply corroborated this portending proposal. One early example of such supporting evidence is shown in Fig. 3B. Here the distinction between a eubacterial species, (*E. coli*) and halophilic archaeal species, (*H. volcanii*) was demonstrated by the different folding of their 16S rRNA into secondary structures (Gutell et al., 1985).

2.2.6 Relationship of archaea to the eubacterial and eukaryotic domains

Early studies indicated that despite their prokaryotic phenotypes, archaea were more similar in some respects to eukaryotes than to eubacteria. Archaeal eukaryotic-like genes included, among others, 5S rRNA, (Hori & Osawa, 1979) subunit structure of RNA polymerase, (Berghofer et al., 1988; Huet et al., 1983) translation elongation factors, (Iwabe et al., 1989; Lechner & Böck, 1987; Lechner et al., 1988, 1989) subunits of ATPase (Iwabe et al., 1989) and some ribosomal proteins (Matheson et al., 1980). Indeed, the first completely sequenced genome of an archaeon, *Methanococcus jannaschii*, revealed that genes involved in transcription, translation, and DNA replication were comparable to their eukaryotic paralogues whereas genes related to metabolism, energy generation, and cell division, resembled those of bacteria. Importantly, however, analysis showed that the evolutionary lineage of this archaeon was different from those of eubacteria or eukarya (Bult et al., 1996). Computer aided genomic analysis of two different methanogenic archaeal species identified genes that were more similar to bacterial than to eukaryotic paralogues, others that were closer to eukaryotic than to bacterial genes, whereas a portion of the genome was exclusively archaeal (Koonin et al., 1997; Smith et al., 1997).⁵² This last group of archaea-specific protein-encoding genes constituted an archaeal genomic signature (Graham et al., 2000).

As noted, Woese's discovery of archaea and his assertion that archaea constituted an evolutionary domain independent from the eubacterial and eukaryotic domains was initially met with skepticism and resistance. However, molecular and biochemical data gathered in the late 1980s convinced most scientists that archaea did indeed constitute a third evolutionary domain distinct from both the eubacterial

⁵¹ Early identified archaea were all extremophiles that grew in harsh habitats such as high-salt brines, hot geothermal environments, or strictly anoxic milieu in which even a trace of oxygen is lethal. Because such environments were conceived to mirror conditions at the beginning of life on earth, archaea were seen as the likely earliest living forms that preceded bacteria in evolution. However, later discoveries of mesophilic archaea that grow under temperate conditions, (DeLong, 1992; Fuhrman et al., 1992; DeLong, 1998) suggested that despite their referential name, archaea were not older than eubacteria (Eme & Doolittle, 2015).

⁵² Although these early studies identified in the examined archaeal genome higher proportions of bacteria-like than eukaryote-like genes, later discovered archaeal phyla were found to be more closely related to eukarya (see 2.2.6).

and eukaryotic domains. Ultimately, a ‘three-legged stool’ model of a tripartite phylogenetic tree was adopted (Fig. 3C).⁵³ For a time, it remained unknown whether the three domains were equally separated from one another or alternatively, two of the domains were closer to one another than to the third one. To answer this question, two teams applied paralogous rooting phylogenetic bioinformatics technique, (Schwartz & Dayhoff, 1978) showing that archaea and eukarya were evolutionary closer to one another than to eubacteria, (Gogarten et al., 1989; Iwabe et al., 1989) (Fig. 3C).

2.2.7 Beyond the initial discovery—current state of archaeal research

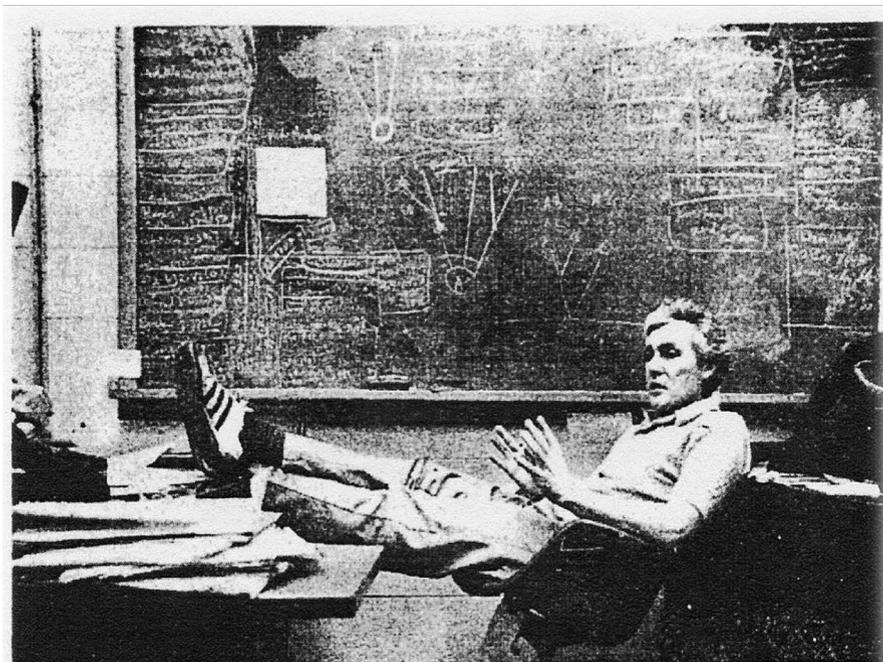
As more than 99% of all microorganisms cannot be cultivated by standard methods, (Amann et al., 1995) studies of archaea were initially hindered by inability to grow in culture most of these microorganisms. This obstacle was removed with the introduction of cloning and PCR amplification techniques that allowed isolation and sequencing of 16S rRNA genes from DNA of unculturable archaea directly collected from the environment (Amann et al., 1995; Pace et al., 1985, 1986). Detecting in this way large numbers of new types of archaea expanded the archaeal domain from just two phyla in the early 1990s, to more than 20 phyla today (Geesink & Ettema, 2022). Of the many new inroads that these advances opened, perhaps the most noteworthy was the identification of the archaeal Asgard phylum that comprises of some members that appear to be closest to the prokarya-eukarya boundary.⁵⁴

3 Normal science activities engendered other unanticipated model-changing discoveries

The discovery of archaea emerged unexpectedly during Woese’s methodological effort to construct a molecular-based evolutionary tree for bacteria. This taxonomic endeavor was framed by a model of a single prokaryotic domain that comprised of all unicells with no nucleus. This domain allegedly bifurcated in time into two-part evolutionary tree of prokaryotic and eukaryotic domains. Thus, because it operated within the consensually accepted theoretical framework of the day, the bacterial classification enterprise can be characterized as puzzle-solving normal science activity [(Kuhn, 1970a); (Kuhn, 1970b) pp. 23–42]. However, an anomaly noticed during this ‘normal science’ project, engendered the unforeseen landmark discovery

⁵³ Yet, prominent traditional evolutionists such as Ernst Mayr, (Mayr, 1990, 1998) and Lynn Margulis, (Margulis, 1996) insisted that bacteria and archaea belonged to a single prokaryote domain. Woese countered that Mayr’s taxonomy was focused on the relatively young eukaryotic branch of the phylogenetic tree. By contrast, bacteria and archaea had much longer evolutionary histories that engendered their significantly higher diversity (Woese, 1998a). Mounting evidence eventually corroborated the division between eubacteria and archaea and the three-domain model of the phylogenetic tree prevailed.

⁵⁴ Discussion of this remarkable finding is beyond the scope of the present paper. Some recent developments in the rapidly moving study of Asgard archaea can be found in (Zaremba-Niedzwiedzka et al., 2017; Cunha et al., 2018; Cunha et al., 2022; Hatano et al., 2022; Rodrigues-Oliveira et al., 2023; Wu et al., 2022).



The New York Times/Phil Greer

Dr. Carl R. Woese, leader of research team, in his office at the University of Illinois. Photo at right shows the newly discovered microorganism: top, a chain of two organisms, each one-thousandth-of-a-millimeter long; center, a cross section of the chain; bottom, an organism dividing into four cells.

Scientists Discover a Form of Life That Predates Higher Organisms

By RICHARD D. LYONS

Special to The New York Times

URBANA, Ill., Nov. 2—Scientists study—[the group investigating the evolution of

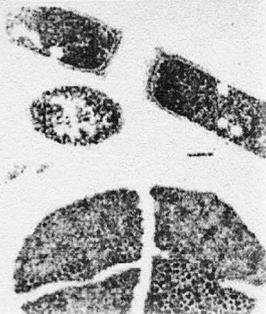


Fig. 2 Headline of the November 3, 1977, *New York Times* front page news story on the discovery of archaea

of archaea. That archaea embodied a distinct domain of the phylogenetic tree was corroborated within a relatively short time. Comparative studies of molecular and biochemical features of archaea relative to those of eubacteria and eukarya, led to realization that the earliest branching point of the evolutionary tree was at the split between the archaeal and eubacterial domains. It was further recognized that this separation was followed by splitting of the eukaryotic from the archaeal domain. Thus, the traditional old view of a bipartite tree in which eukarya branched out from their prokaryotic forebears, had to be replaced by tripartite tree that first branched into eubacterial and archaeal domains and then had the eukarya split from their archaea antecedents. The extension of the tree to three instead of two domains and

the new understanding of the lineages and temporal relationships between these domains forced a change from a model of a bipartite tree to a new model of a tripartite tree.

The described pattern of progression from normal science activity to emerging anomaly and then to a model-changing discovery, is not exclusive to the case of archaea. Two historical cases that are briefly described below illustrate that a similar pattern characterized the discovery of other major biological entities.

3.1 First case: anomaly led to the discovery of split genes and RNA splicing

Systematic experimental studies in the 1960s established that molecules of phage and bacterial messenger RNA (mRNA) were colinear with their encoding genes (Sarabhai et al., 1964; Yanofsky et al., 1964, 1967). Eukaryotic cells and their viruses posed, however, a mystery since their pre-mRNA nuclear transcripts, (so-called heterogenous nuclear RNA; hnRNA) were larger by up to tenfold than the cytoplasmic mRNA molecules (Hiatt, 1962; Scherrer & Darnell, 1962). Experimental studies of the molecular structure of hnRNA and of cytoplasmic mRNA demonstrated their respective precursor-product relationships but did not answer the riddle of their different sizes [reviewed in (Fry, 2016a) pp.495–505]. The two independent respective research groups of Phillip Sharp and Richard Roberts,⁵⁵ concomitantly solved this enigma by showing that genes of human adenovirus, (and later of eukaryotic cells) included transcribed coding segments (‘exons’) and non-coding intervening sequences (‘introns’). Following transcription of a complete gene, the hnRNA transcripts were reduced in length by removal of their introns whereas the exons were rejoined to form shorter translatable cytoplasmic mRNA molecules, [(Berget et al., 1977; Chow et al., 1977); reviewed in (Fry, 2016a) pp. 505–517]. Notably, the original aims of the Sharp and the Roberts teams were not to solve the puzzle of the different sizes of the precursor hnRNA and its mRNA product. Taking the Sharp case as an example, his original ‘normal science’ project, which was based on accepted theory and methodologies of the time, was aimed at determining which adenovirus genes were transcribed into their mRNA molecules at different stages of the virus lytic cycle. Without going into technical details, suffice it to say that Sharp and his associates used electron microscopy to view hybrids of a specific viral gene with its mRNA transcript. An anomaly was observed, however, when the electron micrographs revealed looping out of single strands of the DNA from their double-stranded hybrid regions with the mRNA (Berget et al., 1977). Arnold Berk, who was at the time a postdoctoral fellow in the lab, later described Sharp’s reaction to this unexpected ‘anomaly’ (Berk, 2016):

Sue Berget, and Claire Moore went down to the EM in the MIT Cancer Center a couple of floors below the Sharp laboratory. I was working on my own projects that day and anxiously awaited news of the results. After a couple of hours or so, Phil came back into the laboratory looking somewhat stunned, a

⁵⁵ Sharp and his coworkers Susan Berget and Claire Moore, were at the MIT. Roberts and his associates Louise Chow, Richard Gelinas, and Tom Broker were working at the Cold Spring Harbor Laboratory.

very unusual expression for Phil. “Did you see a loop?,” I asked him anxiously. By this time I had been in the Sharp laboratory for about 9 mo, and I had never heard Phil use anything but the mildest forms of profanity. Nonetheless, Phil excitedly responded: *There are three [profanity] loops!*

Substantiation of this ‘eureka moment’-style observation, together with the independent parallel discoveries by the Roberts team, ended up in dismissal of the concept of universal colinearity of mRNA molecules with their encoding genes and in its replacement by a new model under which colinearity marked only bacterial mRNA whereas mRNA of eukarya was non-colinear.

Although different in technical details, Roberts’ project was also not originally aimed at solving the hnRNA–mRNA relationships puzzle. There too, an unexpected observed ‘anomaly’ led to an independent discovery of split genes and RNA splicing (Chow et al., 1977).

3.2 Second case: anomaly leading to the discovery of catalytic RNA

Until the very early 1980, it was unanimously believed that only proteins could act as biological catalysts (enzymes). However, unanticipated findings that emerged during ‘normal science’ experimental projects led to a recognition that some RNA species were also capable of conducting biological catalysis. The discovery of RNA enzymes (‘ribozymes’) brought about a far-reaching conceptual change from protein-exclusive biological catalysis to catalysis by both proteins and RNA.

Catalytic RNA molecules were independently and concomitantly discovered in the Yale University laboratory of Sidney Altman and in the laboratory of Thomas Cech at the University of Colorado. The original objective of Cech’s ‘normal science’ project was to identify proteins that regulate the transcription of the nucleolar ribosomal RNA coding genes (rDNA) in the ciliated protozoan *Tetrahymena*. These genes were known beforehand to include a spliced intervening sequence. When Art Zaug and Cech followed transcription of the rDNA in crude extracts of *Tetrahymena* nuclei they discerned both mature rRNA transcripts and an excised fragment of the rDNA intervening sequence (Zaug & Cech, 1980). To isolate from the nuclear extract presumed protein factor(s) that excise the intervening sequence, Cech and associates used as substrate purified unprocessed rRNA. As anticipated, incubation of this rRNA with the extract resulted in appearance of the excised intervening sequence. However, an entirely unexpected result was that the same excised fragment appeared in control samples that contained purified rRNA but no extract. Initially thinking that this anomaly must have been a mistake, Cech told Zaug: “*Well Art, this looks very encouraging, except you must have made a mistake making up the control sample*” (Cech, 1990). Repeated experiments, however, yielded the same result, hinting that the primary rRNA transcript may act on itself to excise the intervening sequence and rejoin its two flanking segments. Much work had been subsequently invested in substantiation of this surprising observation and in the elucidation of the detailed mechanism of the self-catalyzed splicing of the rRNA [reviewed in (Cech, 1988, 1990)]. Thus, in this case too, a project that began as

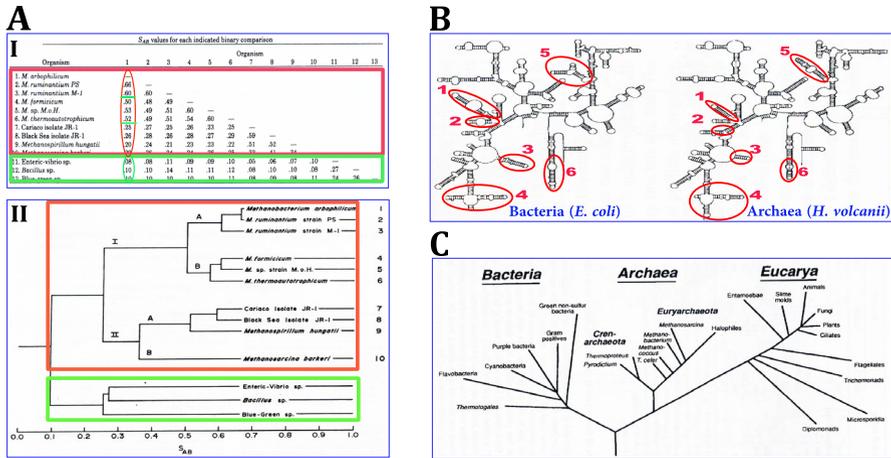


Fig. 3 16S rRNA sequences support a model of archaea as one of three branching domains of the phylogenetic tree. **A** Construction of phylogenetic tree of archaea and eubacteria. **A (I)** Ten methanogenic archaeal species were separated by the association coefficients (S_{ab}) of their paired 16S rRNA sequences²⁶ into three internally related groups of close, intermediate, and low similarity to reference sequence of *M. arborophilicum*. Yet, the 16S rRNA sequences of all three methanogenic subgroups were more similar to one another than to 16S rRNA of three eubacterial species. **A (II)** Arrangement of the eubacterial and archaeal 16S rRNA sequences by their association coefficient-derived evolutionary distances yielded a phylogenetic tree with two separate branches of archaeal and eubacterial domains [adapted from (Fox et al., 1977a)]. **B** 16S rRNA chains of a bacterium (*E. coli*) and halophilic archaeon (*H. volcanii*) had different patterns of secondary structure regions modeled after (Woese et al., 1983). Locales of divergent secondary structure are marked by numbered red ellipses [adapted from (Gutell et al., 1985)]. **C** Early version of the three-domain model of the phylogenetic tree (Woese, 1993). The archaeal and eukaryotic domains were closer to one another than to the eubacterial domain (2.2.7). As additional archaeal phyla were discovered, the archaeal domain in more recent versions of the tree has greater number of sub-branches (Eme et al., 2017; Hug et al., 2016)

‘normal science’ activity yielded unanticipated anomaly which inaugurated verified discovery that replaced the model of proteins as the only biological catalyzers by new understanding the biological catalysis is carried out by some molecular species of both protein and RNA.

That RNA can act as catalyst was discovered in a concomitant independent study of a different system. This work of Altman and associates also began as a ‘normal science’ project that aimed at the elucidation of the mechanism of cleavage of pre-tRNA molecules by the ribonucleoprotein enzyme Ribonuclease P (RNase P). However, unexpected observations led to the discovery that the RNA subunit of RNase P acted alone to catalyze cleavage of the pre-tRNA whereas the protein subunit, which was devoid of catalytic activity, only accelerated the RNA-catalyzed reaction [(Guerrier-Takada et al., 1983); reviewed in (Altman, 1990, 2000)].

4 Anomaly triggered discovery is but one of several paths to the discovery of factual biological entities

As was shown above, the discoveries of the factual biological entities—archaea, RNA splicing, and catalytic RNA conformed to the Kuhnian model discoveries triggered by observational or experimental anomalies (Kuhn, 1962a, 1962b). However, a closer look reveals that this model is by no means a fit-all universal path to discoveries in the life sciences and that alternative approaches also led to discoveries of new unpredicted factual biological entities. This concluding section offers a passing glance at historically proven other effective strategies for discovery of factual entities in biology.

4.1 Discovery by deployment of specific instruments or techniques

Use of novel devices or experimental methodologies effectively opened ways to discoveries of new biological entities. For instance, deployment of microscopes of specific resolving power propelled discoveries of certain cell types, of subcellular structures, and of viruses, (Bonifacino, 2020; Hayat, 1987). In other cases use of the ultracentrifuge prompted the discoveries of mitochondria and microsomes (Claude, 1944) and of lysosomes (De Duve et al., 1953, 1955).

4.2 Discovery by interrogation of large datasets

Another non-Kuhnian path to discovery employs computer-aided searches of large ‘Omics’ databases. In general, computer-aided queries of large databases endeavor to discover hitherto hidden entities or regularities (Leonelli, 2013; Philippi & Köhler, 2006). Just few of the myriad examples of efficacious deployment of this approach are the interrogation of the human genome for the discovery of multiple disease genes (Antonarakis, 2021) or of drugs that are targeted at specific sub-populations or persons (Bachtiar et al., 2019). Likewise, probing of the proteome database discovered diagnostically important cancer biomarkers, (Kang, 2021) and new targets for anti-cancer drugs (Kurimchak et al., 2020).

4.3 Theory-free experimental discoveries

Many major discoveries were made by stepwise methodological investigations with no guiding theory and with only an open biological question [for recent analysis see (Fry, 2022)]. This general approach had been successfully deployed for the discoveries of the protein biosynthesis (‘translation’) apparatus, (Rheinberger, 2006) the cyclic AMP signaling molecule, (Sutherland, 1970, 1992) and the non-lysosomal ubiquitin–proteasome system of intracellular protein breakdown (Hershko, 2005).

4.4 Potential future approach to discovery: employment of artificial intelligence

As discussed in Sect. 1, because scientific discoveries were traditionally viewed as reason-defying unique occurrences, discovery-generating algorithms were thought to be unattainable. However, very recent developments in computer sciences introduced a prospect of using generative artificial intelligence to create scientific hypotheses and accelerate and even produce new discoveries. Although substantial hurdles, such as poor datasets quality and stewardship, stand on the way of making this possibility a reality, deployment AI methods to generate new scientific discoveries is currently holding enticing potential (Wang et al., 2023).

Acknowledgements The author thanks Dr. Iris Fry for her critical reading of of the manuscript and the two anonymous reviewers for their insightful comments and suggestions.

Funding Open access funding provided by Technion - Israel Institute of Technology.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflicts of interest None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Achinstein, P. (1970). Inference to scientific laws. In R. H. Stuewer (Ed.), *Historical and philosophical perspectives of science* (pp. 87–111). University of Minnesota Press.
- Achinstein, P. (1987). Scientific discovery and Maxwell's kinetic theory. *Philosophy of Science*, 54(3), 409–434.
- Addis, M., Sozou, P., Lane, P., & Gobet, F. (2016). Computational scientific discovery and cognitive science theories. In V. C. Mueller (Ed.), *Computing and philosophy* (pp. 83–97). Springer.
- Albers, S. V., Forterre, P., Prangishvili, D., & Schleper, C. (2013). The legacy of Carl Woese and Wolfram Zillig: From phylogeny to landmark discoveries. *Nature Reviews Microbiology*, 11(10), 713–719.
- Aliseda, A. (2006). *Abductive reasoning*. Springer.
- Altman, S. (1990). Enzymatic cleavage of RNA by RNA (Nobel Lecture). *Angewandte Chemie International Edition in English*, 29(7), 749–758.
- Altman, S. (2000). The road to RNase P. *Nature Structural Biology*, 7(10), 827–828.
- Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, 59(1), 143–169.

- Antonarakis, S. E. (2021). History of the methodology of disease gene identification. *American Journal of Medical Genetics Part A*, 185(11), 3266–3275.
- Bachtiar, M., Ooi, B. N. S., Wang, J., Jin, Y., Tan, T. W., Chong, S. S., & Lee, C. G. L. (2019). Towards precision medicine: Interrogating the human genome to identify drug pathways associated with potentially functional, population-differentiated polymorphisms. *The Pharmacogenomics Journal*, 19(6), 516–527.
- Balch, W. E., & Wolfe, R. S. (1976). New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressureized atmosphere. *Applied Environmental Microbiology*, 32(6), 781–791.
- Bendich, A. J., & McCarthy, B. J. (1970). Ribosomal RNA homologies among distantly related organisms. *Proceedings of the National Academy of Sciences USA*, 65(2), 349–356.
- Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences USA*, 74(8), 3171–3175.
- Berghofer, B., Krockel, L., Kortner, C., Truss, M., Schallenberg, J., & Klein, A. (1988). Relatedness of archaeobacterial RNA polymerase core subunits to their eubacterial and eukaryotic equivalents. *Nucleic Acids Research*, 16(16), 8113–8128.
- Berk, A. J. (2016). Discovery of RNA splicing and genes in pieces. *Proceedings of the National Academy of Sciences USA*, 113(4), 801–805.
- Bird, A. (2002). Kuhn's wrong turning. *Studies in History and Philosophy of Science*, 33(3), 443–463.
- Bird, A. (2005). Naturalizing Kuhn. *Proceedings of the Aristotelian Society*, 105(1), 99–117.
- Bonifacino, J. S. E. (2020). *Current protocols in cell biology: Microscopy*. Wiley.
- Breed, R. S., Murray, E. G. D., & Smith, N. R. (Eds.). (1957). *Bergey's manual of determinative bacteriology* (7th ed.). Williams and Wilkins.
- Brosius, J., Palmer, M. L., Kennedy, P. J., & Noller, H. F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, 75(10), 4801–4805.
- Brownlee, G. G., Sanger, F., & Barrell, B. G. (1967). Nucleotide sequence of 5S-ribosomal RNA from *Escherichia coli*. *Nature*, 215(5102), 735–736.
- Buchanan, R. E. (1925). *General systematic bacteriology*. Williams and Wilkins.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., et al. (1996). Complete genome sequence of the methanogenic archaeon. *Methanococcus Jannaschii Science*, 273(5278), 1058–1073.
- Carbon, P., Ehresmann, C., Ehresmann, B., & Ebel, J. P. (1978). The sequence of *Escherichia coli* ribosomal 16 S RNA determined by new rapid gel methods. *FEBS Letters*, 94(1), 152–156.
- Carnap, R. (1966). *Philosophical foundations of physics: An introduction to the philosophy of science*. Basic Books.
- Cech, T. R. (1988). Conserved sequences and structures of group I introns: Building an active site for RNA catalysis—a review. *Gene*, 73(2), 259–271.
- Cech, T. R. (1990). Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from *Tetrahymena*. *Bioscience Reports*, 10(3), 239–261.
- Chalmers, A. F. (2013). *What is this thing called science?* University of Queensland Press.
- Chang, H. (2007). Scientific progress: Beyond foundationalism and coherentism. *Royal Institute of Philosophy Supplements*, 61, 1–20.
- Chang, H. (2017). Operational coherence as the source of truth. *Proceedings of the Aristotelian Society*, 117(2), 103–122.
- Chatton, É. P. L. (1938). *Titres et travaux scientifiques (1906–1937)*. E. Sottano.
- Chike, A. B. (2021). Karl popper's critique of Thomas Kuhn's concept of normal science: An evaluation. *African Journal of Social Sciences and Humanities Research*, 4(3), 105–115.
- Chow, L. T., Gelinias, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1), 1–8.
- Claude, A. (1944). The constitution of mitochondria and microsomes, and the distribution of nucleic acid in the cytoplasm of a leukemic cell. *Journal of Experimental Medicine*, 80(1), 19–29.
- Cohn, F. (1872). Untersuchungen über Bacterien. *Beiträge zur Biologie der Pflanzen*, 1(2), 127–224.
- Cohn, F. (1875). Untersuchungen über Bacterien II. *Beiträge zur Biologie der Pflanzen*, 1(3), 141–207.
- Cohn, F. (1876). Untersuchungen über Bacterien IV. *Beiträge zur Biologie der Pflanzen*, 2(2), 249–276.
- Copeland, H. F. (1938). The kingdoms of organisms. *The Quarterly Review of Biology*, 13(4), 383–420.
- Copeland, H. F. (1947). Progress report on basic classification. *American Naturalist*, 81(800), 340–361.

- Corliss, J. O. (1989). Protistan diversity and origins of multicellular/multitissued organisms. *Bollettino di Zoologia*, 56(3), 227–234.
- Crick, F. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, 138–163.
- Curd, M. V. (1980). The logic of discovery: An analysis of three approaches. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 201–219). Springer.
- Da Cunha, V., Gaia, M., & Forterre, P. (2022). The expanding Asgard archaea and their elusive relationships with Eukarya. *Mlife*, 1(1), 3–12.
- Da Cunha, V., Gaia, M., Nasir, A., & Forterre, P. (2018). Asgard archaea do not close the debate about the universal tree of life topology. *PLoS Genetics*, 14(3), e1007215.
- Darden, L. (1982). Artificial intelligence and philosophy of science: Reasoning by analogy in theory construction. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 147–165.
- Darden, L. (1980). Theory construction in genetics. In T. Nickles (Ed.), *Scientific discovery: Case studies, boston studies in the philosophy of science* (Vol. 60, pp. 151–170). D. Reidel.
- Darden, L. (1991). *Theory change in science: Strategies from mendelian genetics*. Oxford University Press.
- Darden, L. (2002). Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining. *Philosophy of Science*, 69(S3), S354–S365.
- Darden, L. (2006). *Reasoning in biological discoveries: Essays on mechanisms, interfield relations, and anomaly resolution* (Cambridge Studies in Philosophy and Biology). Cambridge University Press.
- Darden, L. (2009). Discovering mechanisms in molecular biology. In J. Meheus & T. Nickles (Eds.), *Models of discovery and creativity* (pp. 43–55). Springer.
- Darden, L., & Craver, C. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(1), 1–28.
- Dayrat, B. (2003). The roots of phylogeny: How did Haeckel build his trees? *Systematic Biology*, 52(4), 515–527.
- De Duve, C., Gianetto, R., Appelmans, F., & Wattiaux, R. (1953). Enzymic content of the mitochondrial fraction. *Nature*, 172(4390), 1143–1144.
- De Duve, C., Pressman, B. C., Gianetto, R., Wattiaux, R., & Appelmans, F. (1955). Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. *The Biochemical Journal*, 60(4), 604–617.
- DeLong, E. F. (1992). Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences USA*, 89(12), 5685–5689.
- DeLong, E. F. (1998). Everything in moderation: Archaea as ‘non-extremophiles.’ *Current Opinion in Genetics and Development*, 8(6), 649–654.
- Dickerson, R. E. (1980). Evolution and gene transfer in purple photosynthetic bacteria. *Nature*, 283(5743), 210–212.
- Doolittle, W. F., Woese, C. R., Sogin, M. L., Bonen, L., & Stahl, D. (1975). Sequence studies on 16S ribosomal RNA from a blue-green alga. *Journal of Molecular Evolution*, 4(4), 307–315.
- DuBuy, B., & Weissman, S. M. (1971). Nucleotide sequence of *Pseudomonas fluorescens* 5 S ribonucleic acid. *Journal of Biological Chemistry*, 246(3), 747–761.
- Dujardin, F. (1841). *Histoire naturelle des zoophytes. Infusoires, comprenant la physiologie et la classification de ces animaux, et la manière de les étudier à l'aide du microscope*. Librairie Encyclopédique de Roret.
- Džeroski, S., Langley, P., & Todorovski, L. (2007). Computational discovery of scientific knowledge. In S. Džeroski & L. Todorovski (Eds.), *Computational discovery of scientific knowledge: Introduction, techniques, and applications in environmental and life sciences* (pp. 1–14). Springer.
- Ehrebner, C. G. (1838). *Die Infusionsthierehen als Vollkommene Organismen. Ein Blick in das Tiefere organische Leben der Natur*. von Leopold Voss.
- Einstein, A. (1954). *Ideas and opinions* (S. Bargmann, Trans.). Crown.
- Eme, L., & Doolittle, W. F. (2015). Archaea. *Current Biology*, 25(19), R851–R855.
- Eme, L., Spang, A., Lombard, J., Stairs, C. W., & Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nature Reviews Microbiology*, 15(12), 711–723.
- Fann, K. T. (1970). *Peirce's theory of abduction*. Martinus Nijhoff.
- Feyerabend, P. (1970). Against method: Outline of an anarchistic theory of knowledge. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science* (Vol. 4, pp. 17–130). Minneapolis University of Minnesota Press.

- Feyerabend, P. (1993). *Against method* (3rd ed.). Verso.
- Fox, G. E. (2015). Archaea. In J. Trefil (Ed.), *Discoveries in modern science: Exploration, inventions* (1st ed., pp. 35–37). Macmillan.
- Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S., & Woese, C. R. (1977a). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proceedings of the National Academy of Sciences USA*, 74(10), 4537–4541.
- Fox, G. E., Pechman, K. R., & Woese, C. R. (1977b). Comparative cataloging of 16s ribosomal ribonucleic acid: Molecular approach to prokaryotic systematics. *International Journal of Systematic and Evolutionary Microbiology*, 27(1), 44–57.
- Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., et al. (1980). The phylogeny of prokaryotes. *Science*, 209(4455), 457–463.
- Frankfurt, H. G. (1958). Peirce's notion of abduction. *Journal of Philosophy*, 55(14), 588–597.
- Fry, M. (2016a). *Landmark experiments in molecular biology*. Elsevier.
- Fry, M. (2016b). Dissolution of hypotheses in biochemistry: Three case studies. *History and Philosophy of the Life Sciences*, 38(4), 17.
- Fry, M. (2020). Ontologically simple theories do not indicate the true nature of complex biological systems: Three test cases. *History and Philosophy of the Life Sciences*, 42(2), 17.
- Fry, M. (2022). Question-driven stepwise experimental discoveries in biochemistry: Two case studies. *History and Philosophy of the Life Sciences*, 44(2), 12.
- Fuhrman, J. A., McCallum, K., & Davis, A. A. (1992). Novel major archaeobacterial group from marine plankton. *Nature*, 356(6365), 148–149.
- Gabbay, D., & Woods, J. (2005). *A practical logic of cognitive systems. The reach of abduction: Insight and trial* (Vol. 2). Elsevier.
- Garrity, G. M., Boone, D. R., & Castenholz, R. W. (2001). *Bergey's manual of systematic bacteriology*. Springer.
- Geesink, P., & Ettema, T. J. G. (2022). The human archaeome in focus. *Nature Microbiology*, 7(1), 10–11.
- Gilbert, W., & Müller-Hill, B. (1966). Isolation of the lac repressor. *Proceedings of the National Academy of Sciences USA*, 56(6), 1891–1898.
- Gilbert, W., & Müller-Hill, B. (1967). The lac operator is DNA. *Proceedings of the National Academy of Science USA*, 58(6), 2415–2421.
- Godfrey-Smith, P. (2003). *Theory and reality*. University of Chicago Press.
- Gogarten, J. P. (2019). The progenote, last universal common ancestor, and the root of the cellular tree of life. In V. M. Kolb (Ed.), *Handbook of astrobiology* (pp. 521–525). CRC Press.
- Gogarten, J. P., Kibak, H., Dittich, P., Taiz, L., Bowman, E. J., Bowman, B. J., et al. (1989). Evolution of the vacuolar H⁺-ATPase: Implications for the origin of eukaryotes. *Proceedings of the National Academy of Sciences USA*, 86(17), 6661–6665.
- Gold, L. (2014). Carl Woese in Schenectady. *RNA Biology*, 11(3), 236–238.
- Goldenfeld, N. (2014). Looking in the right direction: Carl Woese and evolutionary biology. *RNA Biology*, 11(3), 248–253.
- Goldman, A. I. (1983). Epistemology and the theory of problem solving. *Synthese*, 55(1), 21–48.
- Graham, D. E., Overbeek, R., Olsen, G. J., & Woese, C. R. (2000). An archaeal genomic signature. *Proceedings of the National Academy of Sciences USA*, 97(7), 3304–3308.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 Pt 2), 849–857.
- Gutell, R. R., Weiser, B., Woese, C. R., & Noller, H. F. (1985). Comparative anatomy of 16-S-like ribosomal RNA. In W. E. Cohn & K. Moldave (Eds.), *Progress in nucleic acid research and molecular biology* (Vol. 32, pp. 155–216). Academic Press.
- Haeckel, E. (1866). *Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch Begründet durch die von Charles Darwin reformirte Descendenz-Theorie*. Reimer.
- Hanson, N. R. (1958a). The logic of discovery. *Journal of Philosophy*, 55(25), 1073–1089.
- Hanson, N. R. (1958b). *Patterns of discovery: An inquiry into the conceptual foundations of science* (1st ed.). Cambridge University Press.
- Hanson, N. R. (1961). Is there a logic of scientific discovery? In H. Feigl & G. Maxwell (Eds.), *Current issues in the philosophy of science* (pp. 20–35). Holt Rinehart and Winston.
- Harman, G. H. (1965). The inference to the best explanation. *The Philosophical Review*, 74(1), 88–95.

- Harman, G. H. (1968). Enumerative induction as inference to the best explanation. *The Journal of Philosophy*, 65(18), 529–533.
- Hatano, T., Palani, S., Papatziadou, D., Salzer, R., Souza, D. P., Tamarit, D., et al. (2022). Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nature Communications*, 13(1), 3398.
- Hayat, M. A. (Ed.). (1987). *Correlative microscopy in biology: Instrumentation and methods*. Elsevier.
- Hempel, C. G. (1966). *Philosophy of natural science*. Prentice-Hall.
- Herschel, J. F. W. (1831). *Preliminary discourse on the study of natural philosophy*. Longman, Rees, Orme, Brown, and Green.
- Hershko, A. (2005). The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. *Cell Death & Differentiation*, 12(9), 1191–1197.
- Hiatt, H. H. (1962). A rapidly labeled RNA in rat liver nuclei. *Journal of Molecular Biology*, 5(2), 217–229.
- Hori, H., & Osawa, S. (1979). Evolutionary change in 5S RNA secondary structure and a phylogenetic tree of 54 5S RNA species. *Proceedings of the National Academy of Sciences USA*, 76(1), 381–385.
- Hoyningen-Huene, P. (1987). Context of discovery and context of justification. *Studies in History and Philosophy of Science Part A*, 18(4), 501–515.
- Huet, J., Schnabel, R., Sentenac, A., & Zillig, W. (1983). Archaeobacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *EMBO Journal*, 2(8), 1291–1294.
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nature Microbiology*, 1(5), 16048.
- Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D., & Soo, R. M. (2021). Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME Journal*, 15(7), 1879–1892.
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., & Miyata, T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences USA*, 86(23), 9355–9359.
- Kandler, O., & Hippe, H. (1977). Lack of peptidoglycan in the cell walls of *Methanosarcina barkeri*. *Archives of Microbiology*, 113(1), 57–60.
- Kandler, O., & König, H. (1978). Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria. *Archives of Microbiology*, 118(2), 141–152.
- Kang, U. B. (2021). Proteomic interrogation in cancer biomarker. In D. Y. Noh, W. Han, & M. Toi (Eds.), *Advances in experimental medicine and biology*. Springer.
- Kapitan, T. (1992). Peirce and the autonomy of reasoning. *Erkenntnis*, 37, 1–26.
- Kates, M., Palameta, B., Joo, C. N., Kushner, D. J., & Gibbons, N. E. (1966). Aliphatic diether analogs of glyceride-derived lipids. IV. The occurrence of di-O-dihydrophytylglycerol ether containing lipids in extremely halophilic bacteria. *Biochemistry*, 5(12), 4092–4099.
- Katscher, F. (2004). The history of the terms prokaryotes and eukaryotes. *Protist*, 155(2), 257–263.
- Kessel, M., & Klink, F. (1980). Archaeobacterial elongation factor is ADP-ribosylated by diphtheria toxin. *Nature*, 287(5779), 250–251.
- Kessel, M., & Klink, F. (1982). Identification and comparison of eighteen Archaeobacteria by means of the Diphtheria toxin reaction. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene. C: Allgemeine, Angewandte und Ökologische Mikrobiologie*, 3(1), 140–148.
- Koonin, E. V. (2014). Carl Woese's vision of cellular evolution and the domains of life. *RNA Biology*, 11(3), 197–204.
- Koonin, E. V., Mushegian, A. R., Galperin, M. Y., & Walker, D. R. (1997). Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Molecular Microbiology*, 25(4), 619–637.
- Kordig, C. R. (1978). Discovery and justification. *Philosophy of Science*, 45(1), 110–117.
- Kraus, M. (2003). Charles S. Peirce's theory of abduction and the Aristotelian enthymeme from signs. In F. H. Van Eemeren, J. A. Blair, C. A. Willard, & A. F. Snoeck Henkemans (Eds.), *Anyone who has a view: Theoretical contributions to the study of argumentation* (pp. 237–254). Springer.
- Kuhn, T. S. (1962a). Historical structure of scientific discovery. *Science*, 136(3518), 760–764.
- Kuhn, T. S. (1962b). *The structure of scientific revolutions* (1st ed.). University of Chicago Press.
- Kuhn, T. S. (1970a). Logic of discovery or psychology of research? In I. Lakatos & A. Musgrave (Eds.), *Criticism and growth of knowledge* (pp. 4–10). Cambridge University Press.
- Kuhn, T. S. (1970b). *The structure of scientific revolutions* (2nd ed.). The University of Chicago Press.

- Kurimchak, A. M., Herrera-Montávez, C., Brown, J. R., Johnson, K. J., Sodi, V., Srivastava, N., et al. (2020). Functional proteomics interrogation of the kinome identifies MRCKA as a therapeutic target in high-grade serous ovarian carcinoma. *Science Signaling*, *13*(619), eaax8238.
- Lakatos, I. (1968). Criticism and methodology of scientific research programmes. *Proceedings of the Aristotelian Society*, *69*(1), 149–186.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (Vol. 4, pp. 91–195). Cambridge University Press.
- Langley, P. (2000). The computational support of scientific discovery. *International Journal of Human-Computer Studies*, *53*(3), 393–410.
- Langworthy, T. A., Smith, P. F., & Mayberry, W. R. (1972). Lipids of thermoplasma acidophilum. *Journal of Bacteriology*, *112*(3), 1193–1200.
- Laudan, L. (1980). Why was the logic of discovery abandoned? In T. Nickles (Ed.), *Scientific discovery, logic, and rationality, boston studies in the philosophy of science* (Vol. 56, pp. 173–183). Reidel.
- Lechner, K., & Böck, A. (1987). Cloning and nucleotide sequence of the gene for an archaeobacterial protein synthesis elongation factor Tu. *Molecular and General Genetics*, *208*, 523–528.
- Lechner, K., Heller, G., & Böck, A. (1988). Gene for the diphtheria toxin-susceptible elongation factor 2 from *Methanococcus vannielii*. *Nucleic Acids Research*, *16*(16), 7817–7826.
- Lechner, K., Heller, G., & Böck, A. (1989). Organization and nucleotide sequence of a transcriptional unit of *Methanococcus vannielii* comprising genes for protein synthesis elongation factors and ribosomal proteins. *Journal of Molecular Evolution*, *29*(1), 20–27.
- Leonelli, S. (2013). Data-intensive research. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of systems biology* (pp. 545–548). Springer.
- Levit, G., & Hossfeld, U. (2022). Ernst Mayr's critique of Thomas Kuhn. *Epistemology & Philosophy of Science*, *59*(4), 163–180.
- Luehrsen, K. R. (2014). Remembering Carl Woese. *RNA Biology*, *11*(3), 217–219.
- Lund, M. (2010). *N. R. Hanson: Observation, discover, and scientific change*. Humanity Books.
- Magnani, L. (2001). *Abduction, reason, and science: Processes of discovery and explanation*. Kluwer.
- Magnani, L. (2009). Creative abduction and hypothesis withdrawal. In J. Meheus & T. Nickles (Eds.), *Models of discovery and creativity* (pp. 95–126). Springer.
- Margulis, L. (1996). Archaeal-eubacterial mergers in the origin of Eukarya: Phylogenetic classification of life. *Proceedings of the National Academy of Sciences USA*, *93*(3), 1071–1076.
- Matheson, A. T., Yaguchi, M., Balch, W. E., & Wolfe, R. S. (1980). Sequence homologies in the N-terminal region of the ribosomal 'A' proteins from methanobacterium thermoautotrophicum and halobacterium cutirubrum. *Biochimica et Biophysica Acta*, *626*(1), 162–169.
- Mayr, E. (1990). A natural system of organisms. *Nature*, *348*(6301), 491–491.
- Mayr, E. (1998). Two empires or three? *Proceedings of the National Academy of Sciences USA*, *95*(17), 9720–9723.
- Migula, W. (1897). *System der Bakterien. Handbuch der Morphologie Entwicklungsgeschichte und Systematik der Bakterien*. Gustav Fischer.
- Migula, W. (1907). Allgemeine morphologie. In F. Lafar (Ed.), *Handbuch der technischen Mykologie (Vol. 1, pp. 29–149)*. G. Fischer.
- Moore, P. B. (2014). Carl Woese: A structural biologist's perspective. *RNA Biology*, *11*(3), 172–174.
- Moore, R. L., & McCarthy, B. J. (1967). Comparative study of ribosomal ribonucleic acid cistrons in enterobacteria and myxobacteria. *Journal of Bacteriology*, *94*(4), 1066–1074.
- Morell, V. (1997). Microbiology's scarred revolutionary. *Science*, *276*(5313), 699–702.
- Morgan, G. J. (1998). Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. *Journal of the History of Biology*, *31*(2), 155–178.
- Mowry, B. (1985). From Galen's theory to William Harvey's theory: A case study in the rationality of scientific theory change. *Studies in History and Philosophy of Science Part A*, *16*(1), 49–82.
- Müller, O. F. (1773; 1774). *Vermium terrestrium et fluviatilium, seu animalium infusorium, helminthicorum, et testaceorum, non marinorum, succincta historia* (Vol. 1 and 2). Heineck et Faber.
- Müller, O. F. (1786). *Animalcula infusoria fluviatilia et marina quae detexit, systematice, descripsit et ad vivum delineare curavit*: Möller
- Nair, P. (2012). Woese and Fox: Life, rearranged. *Proceedings of the National Academy of Sciences USA*, *109*(4), 1019–1021.
- Nickles, T. (1980). Introductory essay: Scientific discovery and the future of philosophy of science. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 1–59). Springer.

- Nickles, T. (1985). Beyond divorce: Current status of the discovery debate. *Philosophy of Science*, 52(2), 177–206.
- Olsen, G. J., Woese, C. R., & Overbeek, R. (1994). The winds of (evolutionary) change: Breathing new life into microbiology. *Journal of Bacteriology*, 176(1), 1–6.
- Paavola, S. (2004). Abduction as a logic and methodology of discovery: The importance of strategies. *Foundations of Science*, 9(3), 267–283.
- Paavola, S. (2006). *On the origin of ideas: An abductivist approach to discovery* (Philosophical studies from the University of Helsinki). Department of Philosophy, University of Helsinki.
- Pace, B., & Campbell, L. L. (1971). Homology of ribosomal ribonucleic acid diverse bacterial species with *Escherichia coli* and *Bacillus stearothermophilus*. *Journal of Bacteriology*, 107(2), 543–547.
- Pace, N. R., Sapp, J., & Goldenfeld, N. (2012). Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. *Proceedings of the National Academy of Sciences USA*, 109(4), 1011–1018.
- Pace, N. R., Stahl, D. A., Lane, D. J., & Olsen, G. J. (1985). Analyzing natural microbial populations by rRNA sequences. *ASM News*, 51(1), 4–12.
- Pace, N. R., Stahl, D. A., Lane, D. J., & Olsen, G. J. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. In K. C. Marshall (Ed.), *Advances in microbial ecology* (pp. 1–55). Springer.
- Pardee, A. B. (2002). PaJaMas in Paris. *Trends in Genetics*, 18(11), 585–587.
- Pardee, A. B. (2003). The PaJaMa experiment. In A. Ullmann (Ed.), *Origins of molecular biology: A tribute to jacques monod* (pp. 133–142). ASM Press.
- Pardee, A. B., Jacob, F., & Monod, J. (1959). The genetic control and cytoplasmic expression of “Inducibility” in the synthesis of β -galactosidase by *E. coli*. *Journal of Molecular Biology*, 1(2), 165–178.
- Pauling, L., & Zuckerkandl, E. (1963). Chemical paleogenetics. Molecular “restoration studies” of extinct forms of life. *Acta Chemica Scandinavica*, 17, 9–16.
- Pechman, K. J., & Woese, C. R. (1972). Characterization of the primary structural homology between the 16s ribosomal RNAs of *Escherichia coli* and *Bacillus megaterium* by oligomer cataloging. *Journal of Molecular Evolution*, 1(3), 230–240.
- Perty, M. (1852). *Zur Kenntniss kleinster Lebensformen nach Bau, Funktionen, Systematik, mit Specialverzeichniss der in der Schweiz beobachteten*. Jent & Reinert.
- Philippi, S., & Köhler, J. (2006). Addressing the problems with life-science databases for traditional uses and systems biology. *Nature Reviews Genetics*, 7(6), 482–488.
- Platt, J. R. (1964). Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, 146(3642), 347–353.
- Popper, K. (1959/2002). *The logic of scientific discovery*. Routledge Classics.
- Putnam, H. (1991). The “corroboration” of theories. In R. Boyd, P. Gasper, & J. D. Trout (Eds.), *The philosophy of science* (pp. 121–137). MIT Press.
- Quammen, D. (2018). *The tangled tree: A radical new history of life*. Simon & Schuster.
- Reichenbach, H. (1938). *Experience and prediction*. University of Chicago Press.
- Reichenbach, H. (1951). *The rise of scientific philosophy*. University of California Press.
- Rheinberger, H. J. (2006). A history of protein biosynthesis and ribosome research. In K. H. Nierhaus & D. Wilson (Eds.), *Protein synthesis and ribosome structure: Translating the genome* (pp. 1–55). Wiley.
- Rodrigues-Oliveira, T., Wollweber, F., Ponce-Toledo, R. I., Xu, J., Rittmann, S. K. M. R., Klingl, A., et al. (2023). Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature*, 613(7943), 332–339.
- Rosset, R., Monier, R., & Julien, J. (1964). Les ribosomes d’*Escherichia coli*. I. Mise en évidence d’un RNA ribosomique de faible poids moléculaire. *Bulletin de la Société Chimique (Paris)*, 46, 87–109.
- Sanger, F., Brownlee, G. G., & Barrell, B. G. (1965). A two-dimensional fractionation procedure for radioactive nucleotides. *Journal of Molecular Biology*, 13(2), 373–398.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA*, 74(12), 5463–5467.
- Sankey, H. (1993). Kuhn’s changing concept of incommensurability. *British Journal of the Philosophy of Science*, 44(4), 759–774.
- Sapp, J. (2005). The prokaryote-eukaryote dichotomy: Meanings and mythology. *Microbiology and Molecular Biology Reviews*, 69(2), 292–305.
- Sapp, J. (2008). The iconoclastic research program of Carl Woese. In O. Harman & M. R. Dietrich (Eds.), *Rebels, mavericks, and heretics in biology* (pp. 302–320). Yale University Press.

- Sapp, J., & Fox, G. E. (2013). The singular quest for a universal tree of life. *Microbiology and Molecular Biology Reviews*, 77(4), 541–550.
- Sarabhai, A. S., Stretton, A. O., Brenner, S., & Bolle, A. (1964). Co-linearity of the gene with the polypeptide chain. *Nature*, 201(4914), 13–17.
- Schaffner, K. F. (1974). Logic of discovery and justification in regulatory genetics. *Studies in History and Philosophy of Science Part A*, 4(4), 349–385.
- Schaffner, K. F. (1980). Discovery in the biomedical sciences: Logic or irrational intuition? In T. Nickles (Ed.), *Scientific discovery: Case studies Boston studies in the philosophy of science* (Vol. 60, pp. 171–205). D. Reidel.
- Scherrer, K., & Darnell, J. E. (1962). Sedimentation characteristics of rapidly labelled RNA from HeLa cells. *Biochemical and Biophysical Research Communications*, 7(6), 486–490.
- Schickore, J. (2006). A forerunner? — perhaps, but not to the context distinction. William Whewell's germano-cantabrigian history of the fundamental ideas. In J. Schickore, & F. Steinle (Eds.), *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction* (Vol. 14, pp. 57–77), (Archimedes: New Studies in the History and Philosophy of Science and Technology). Springer.
- Schickore, J. (2022). Scientific discovery. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Stanford Metaphysics Research Lab.
- Schickore, J., & Steinle, F. (Eds.). (2006). *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction* (Vol. 14) (Archimedes: New Studies in the History and Philosophy of Science and Technology). Springer.
- Schopf, J. W. (1975). Precambrian paleobiology: Problems and perspectives. *Annual Review of Earth and Planetary Sciences*, 3, 213–249.
- Schwartz, R. M., & Dayhoff, M. O. (1978). Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science*, 199(4327), 395–403.
- Shan, Y. (2020). Kuhn's "wrong turning" and legacy today. *Synthese*, 197(1), 381–406.
- Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., et al. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: Functional analysis and comparative genomics. *Journal of Bacteriology*, 179(22), 7135–7155.
- Sogin, S. J., Sogin, M. L., & Woese, C. R. (1972). Phylogenetic measurement in prokaryotes by primary structural characterization. *Journal of Molecular Evolution*, 1(2), 173–184.
- Sozou, P. D., Lane, P. C. R., Addis, M., & Gobet, F. (2017). Computational scientific discovery. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 719–734). Springer.
- Stanier, R. Y. (1961). La place des bactéries dans le monde vivant. *Annales de l'Institut Pasteur (Paris)*, 101, 297–312.
- Stanier, R. Y., Doudoriff, M., & Adelberg, E. A. (1957). *The microbial world* (1st ed.). Prentice Hall.
- Stanier, R. Y., Doudoroff, M., & Adelberg, E. A. (1963). *The microbial world* (2nd ed.). Prentice-Hall.
- Stanier, R. Y., & Van Niel, C. B. (1941). The main outlines of bacterial classification. *Journal of Bacteriology*, 42(4), 437–466.
- Stanier, R. Y., & van Niel, C. B. (1962). The concept of a bacterium. *Archiv für Mikrobiologie*, 42(1), 17–35.
- Summers, W. C. (2021). The discovery of bacteriophages and the historical context. In D. R. Harper, S. T. Abedon, B. H. Burrowes, & M. L. McConville (Eds.), *Bacteriophages: Biology, technology, therapy* (pp. 387–400). Springer.
- Sutherland, E. W. (1970). On the biological role of cyclic AMP. *Journal of the American Medical Association—JAMA*, 214(7), 1281–1288.
- Sutherland, E. W. (1992). Studies on the mechanism of hormone action (1971 Nobel Lecture). In J. Lindsten (Ed.), *Nobel lectures in physiology or medicine 1971–1980* (pp. 2–22). World Scientific.
- Tornabene, T. G., & Langworthy, T. A. (1979). Diphytanyl and dibiphytanyl glycerol ether lipids of *Methanogenic archaeobacteria*. *Science*, 203(4375), 51–53.
- Uchida, T., Bonen, L., Schaup, H. W., Lewis, B. J., Zablen, L., & Woese, C. (1974). The use of ribonuclease U2 in RNA sequence determination. Some corrections in the catalog of oligomers produced by ribonuclease T1 digestion of *Escherichia coli* 16S ribosomal RNA. *Journal of Molecular Evolution*, 3(1), 63–77.
- Van Niel, C. B. (1955). Classification and taxonomy of the bacteria and bluegreen algae. In E. B. Babcock, J. W. Durham, & G. S. Myers (Eds.), *A century of progress in the natural sciences, 1853–1953* (pp. 89–114). California Academy of Sciences.

- Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972), 47–60.
- Weber, M. (2005). *Philosophy of experimental biology*. Cambridge University Press.
- Weinert, F. (2014). Lines of descent: Kuhn and beyond. *Foundations of Science*, 19(4), 331–352.
- Whewell, W. (1847). *The philosophy of inductive sciences, founded on their history* (Vol. 2). John W. Parker.
- Whewell, W. (1858). *Novum Organon renovatum*. John W. Parker & Son.
- Whittaker, R. H. (1959). On the broad classification of organisms. *The Quarterly Review of Biology*, 34(3), 210–226.
- Whittaker, R. H. (1969). New concepts of kingdoms of organisms. *Science*, 163(3863), 150–160.
- Wisdom, J. O. (1952). *Foundations of inference in natural science* (Vol. 11). Methuen.
- Woese, C. R. (1967). *The genetic code : The molecular basis for genetic expression (Modern Perspectives in Biology)*. Harper & Row.
- Woese, C. R. (1970). Organization and control in prokaryotic and eukaryotic cells. In H. P. Charles & B. C. J. G. Knight (Eds.), *20th symposium, society for general microbiology* (pp. 39–54). Cambridge University Press.
- Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews*, 51(2), 221–271.
- Woese, C. R. (1993). The archaea: Their history and significance. In M. Kates, D. J. Kushner, & A. T. Matheson (Eds.), *New comprehensive biochemistry* (Vol. 26, pp. vii–xxix). Elsevier.
- Woese, C. R. (1998a). Default taxonomy: Ernst Mayr's view of the microbial world. *Proceedings of the National Academy of Sciences USA*, 95(19), 11043–11046.
- Woese, C. R. (1998b). The universal ancestor. *Proceedings of the National Academy of Sciences USA*, 95(12), 6854–6859.
- Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences USA*, 97(15), 8392–8396.
- Woese, C. R. (2002). On the evolution of cells. *Proceedings of the National Academy of Sciences USA*, 99(13), 8742–8747.
- Woese, C. R. (2007). The birth of the archaea: A personal retrospective. In R. A. Garret & H.-P. Klenk (Eds.), *Archaea: Evolution, physiology, and molecular biology*. Blackwell.
- Woese, C. R. (2013). How we do, don't, and should look at bacteria and bacteriology. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The prokaryotes: Prokaryotic biology and symbiotic associations* (pp. 3–20). Springer.
- Woese, C. R., & Fox, G. E. (1977a). The concept of cellular evolution. *Journal of Molecular Evolution*, 10(1), 1–6.
- Woese, C. R., & Fox, G. E. (1977b). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences USA*, 74(11), 5088–5090.
- Woese, C. R., Gibson, J., & Fox, G. E. (1980a). Do genealogical patterns in purple photosynthetic bacteria reflect interspecific gene transfer? *Nature*, 283(5743), 212–214.
- Woese, C. R., Gutell, R., Gupta, R., & Noller, H. F. (1983). Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiology Reviews*, 47(4), 621–669.
- Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences USA*, 87(12), 4576–4579.
- Woese, C. R., Magrum, L. J., & Fox, G. E. (1978). Archaeobacteria. *Journal of Molecular Evolution*, 11(3), 245–251.
- Woese, C. R., Maniloff, J., & Zablén, L. B. (1980b). Phylogenetic analysis of the mycoplasmas. *Proceedings of the National Academy of Sciences USA*, 77(1), 494–498.
- Woese, C. R., Sogin, M., Stahl, D., Lewis, B. J., & Bonen, L. (1976). A comparison of the 16S ribosomal RNAs from mesophilic and thermophilic bacilli: Some modifications in the sanger method for RNA sequencing. *Journal of Molecular Evolution*, 7(3), 197–213.
- Woese, C. R., Sogin, M. L., & Sutton, L. A. (1974). Procaryote phylogeny. I. Concerning the relatedness of *Aerobacter aerogenes* to *Escherichia coli*. *Journal of Molecular Evolution*, 3(4), 293–299.
- Woese, C. R., Stackebrandt, E., Macke, T. J., & Fox, G. E. (1985). A phylogenetic definition of the major eubacterial taxa. *Systematic and Applied Microbiology*, 6, 143–151.
- Wolfe, R. (2014). The archaea: A personal overview of the formative years. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The prokaryotes: Other major lineages of bacteria and the archaea* (pp. 3–9). Springer.

- Wu, F., Speth, D. R., Philofof, A., Cremiere, A., Narayanan, A., Barco, R. A., et al. (2022). Unique mobile elements and scalable gene flow at the prokaryote-eukaryote boundary revealed by circularized Asgard archaea genomes. *Nature Microbiology*, 7(2), 200–212.
- Yankofsky, S. A., & Spiegelman, S. (1963). Distinct cistrons for the two ribosomal RNA components. *Proceedings of the National Academy of Sciences USA*, 49(4), 538–544.
- Yanofsky, C., Carlton, B. C., Guest, J. R., Helinski, D. R., & Henning, U. (1964). On the colinearity of gene structure and protein structure. *Proceedings of the National Academy of Sciences USA*, 51(2), 266–272.
- Yanofsky, C., Drapeau, G. R., Guest, J. R., & Carlton, B. C. (1967). The complete amino acid sequence of the tryptophan synthase A protein (alpha subunit) and its colinear relationship with the genetic map of the A gene. *Proceedings of the National Academy of Sciences USA*, 57(2), 296–298.
- Zablen, L., Bonen, L., Meyer, R., & Woese, C. R. (1975). The phylogenetic status of *Pasteurella pestis*. *Journal of Molecular Evolution*, 4(4), 347–358.
- Zablen, L., & Woese, C. R. (1975). Procaryote phylogeny IV: Concerning the phylogenetic status of a photosynthetic bacterium. *Journal of Molecular Evolution*, 5(1), 25–34.
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Backstrom, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, 541(7637), 353–358.
- Zaug, A. J., & Cech, T. R. (1980). In vitro splicing of the ribosomal RNA precursor in nuclei of *Tetrahymena*. *Cell*, 19(2), 331–338.
- Zillig, W., Stetter, K. O., & Janeković, D. (1979). DNA-Dependent RNA polymerase from the Archaeobacterium *sulfolobus acidocaldarius*. *European Journal of Biochemistry*, 96(3), 597–604.
- Zillig, W., Stetter, K. O., & Tobien, M. (1978). DNA-dependent RNA polymerase from *Halobacterium halobium*. *European Journal of Biochemistry*, 91(1), 193–199.
- Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H., & Scholz, I. (1980). The *Sulfolobus*–“*Caldariella*” group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Archives of Microbiology*, 125(3), 259–269.
- Zuckerkindl, E., Jones, R. T., & Pauling, L. (1960). A comparison of animal hemoglobins by tryptic peptide pattern analysis. *Proceedings of the National Academy of Science USA*, 46(10), 1349–1360.
- Zuckerkindl, E., & Pauling, L. (1962). Molecular disease, evolution, and genic heterogeneity. In M. Kasha & B. Pullman (Eds.), *Horizons in biochemistry* (pp. 189–225). Academic Press.
- Zuckerkindl, E., & Pauling, L. (1965a). Evolutionary divergence and convergence in proteins. In V. Bryson & H. Vogel (Eds.), *Evolving genes and proteins* (pp. 97–166). Academic Press.
- Zuckerkindl, E., & Pauling, L. (1965b). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, 8(2), 357–366.
- Zuckerkindl, E., & Schroeder, W. A. (1961). Amino-acid composition of the polypeptide chains of gorilla haemoglobin. *Nature*, 192(4806), 984–985.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.