MICROCOSMOS

Steven J. Hallam

The origin and diversity of life on Earth has inspired countless stories of creation, cataclysm, and rebirth since human beings first began to wonder. Our early ancestors likely directed much of their curiosity outwards, looking for traces of the origins of life in the visible world and in celestial objects. This began to change in the seventeenth century, when Anton van Leeuwenhoek gazed through the first microscope, revealing an enigmatic and previously invisible microbial world, the microcosmos. Since that time, we have learned how to read the evolutionary memory of Earth as it is encoded in this invisible world. Through careful study of conserved gene sequences and ancestral protein function, the emergence of cells and distributed metabolic networks, we can see how microbial life has both shaped and responded to changes in the Earth system over evolutionary time. Here, we may also find the inspiration and technical savvy needed to solve some of the most pressing environmental problems of our time.

At its most fundamental level, the replication and expression of

biological information is shared by all living things. Nucleic acids (DNA and RNA molecules) encode genetic blueprints for life, which are translated into proteins that build cellular structures and catalyze the biological reactions that drive cellular metabolism. This flow of biological information can be traced back to primordial chemical reactions in early Earth history and the emergence of self-replicating genes. Enclosure of these genes in lipid membranes gave rise to the first cells, which, in turn, became information-storage and -processing units in an evolving ecological network that continues to this day. By studying variations in the genetic information contained in organisms, we can reconstruct evolutionary relationships between extant (and extinct) life on Earth and understand how this life has both shaped and responded to changes in the planet's physical and chemical environment.

Take, for example, the genetic information encoded in ribosomes. These subcellular protein factories play a critical role in translating information from genes into the proteins that drive cellular metabolism. The biochemical activity encoded in the RNA structure of the ribosomal machinery can be traced back to a moment in deep evolutionary time when the first complex proteins were formed from simple building blocks (free amino acids) found in the early Earth environment. All cellular organisms contain ribosomes, and these structures have remained relatively constant over billions of years of evolution. However, over time, small changes in ribosomal RNA gene sequences (mutations) have randomly occurred and subsequently been inherited within organismal lineages. By comparing ribosomal RNA gene sequences from different organisms, we can place the organisms onto the proverbial Tree of Life, with the most closely

related species sharing the most similar pattern of accumulated mutations as a shared evolutionary memory.

Memory is also written into the structure and function of the proteins produced by ribosomes. Consider the protein Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which catalyzes a crucial step in photosynthesis, converting carbon dioxide and water into oxygen and sugar. RuBisCO is present in the vast majority of photosynthetic organisms – from cyanobacteria and moss to maple trees and water lilies – and it is directly responsible for most of the organic carbon naturally produced each year (about 100 billion tons globally). Despite its central role in Earth's carbon cycle, RuBisCO is an inefficient enzyme with a low binding affinity for carbon dioxide and an extremely slow reaction rate (it catalyzes about ten reactions per second, as compared to ten thousand to one hundred thousand reactions per second for other enzymes). The reasons for this conundrum become clearer when we consider the physical and chemical environment in which RuBisCO first evolved.

Approximately 3 billion years ago, a group of microorganisms (ancestors of the blue-green cyanobacteria) innovated a biochemical pathway that could separate water into oxygen and hydrogen, with the hydrogen used as a source of electrons needed for photosynthesis. Prior to the advent of these water-splitting reactions, Earth's atmosphere was rich in carbon dioxide and devoid of oxygen. Over time, free oxygen accumulated, altering the surrounding environment. The presence of oxygen posed a significant challenge for photosynthetic organisms because oxygen can react with RubisCO and decrease the enzyme's ability to bind carbon dioxide and produce sugars. As oxygen

continued to increase in the early atmosphere and as carbon dioxide levels dropped, the capacity for photosynthesis (and the presence of RubisCO) spread into new groups of organisms, including terrestrial plants. Yet, despite the large changes in atmospheric composition that occurred between the first appearance of photosynthetic bacteria (several billion years ago) and land plants (500 to 700 million years ago), the RubisCO proteins found in these two groups are highly similar, both in genetic sequence and functional properties. RubisCO thus represents a molecular fossil that has functional properties that reflect the rise of photosynthesis under environmental conditions radically different from those of the present day.

Moving beyond genes and gene products, we can also see traces of evolutionary memory in the structure and organization of cellular organisms. Single-cell microorganisms affiliated with archaea and bacteria (collectively known as prokaryotes) have a basic cell architecture: most metabolic activities are loosely organized within a central fluid (the cytoplasm) surrounded by the cell membrane. By contrast, eukaryotes – which include all plants, animals, and fungi – have far more complicated cell structures: the biochemical machinery driving metabolism is contained in subcellular compartments known as organelles. The evolutionary transition from prokaryotic to eukaryotic cells (believed to have occurred as much as 2.5 billion years ago) was a major turning point in the diversification of life on Earth. The processes that led to this evolutionary transition have had an enduring impact on the emergence of multicellular life forms and the structure of ecological networks.

Symbiosis, defined as a long-term biological interaction between

two different organisms, has been well documented across many organismal lineages. One of the more compelling and beautiful examples is that of the photosynthetic algae (zooxanthellae) that live inside the animal tissues of coral. These symbiotic algae give the coral reefs their striking colours and, more importantly, the ability to derive energy from photosynthesis. In the late 1960s, Lynn Margulis posited a theory for the origin of eukaryotic cells based on a process of endosymbiosis. She hypothesized that the structures inside the cells we now identify as organelles, including chloroplasts and mitochondria (sites of photosynthesis and respiration, respectively), arose from the incorporation of one cell type into another, resulting in a stable intracellular symbiosis. Although her ideas were initially rejected by the scientific community, supporting evidence accumulated over time. For example, eukaryotic organelles such as chloroplasts and mitochondria have their own genomes (separate from the nuclear genomes of the cells in which they are found), and they also have membranes, ribosomes, and modes of replication that are similar to the ancient bacteria from which they are believed to be descended. We now recognize that symbiosis is a fundamental organizing principle that records evolutionary memories at different levels of cellular complexity.

Memories encoded in the microcosmos represent 3.5 billion years of evolution, during which time microorganisms have developed metabolic pathways to harness energy and materials from the world around them. This process has fundamentally transformed the surface chemistry of Earth, and it has generated a deep reservoir of genomic diversity as microbial lineages diversified and came to occupy every conceivable metabolic niche. Today, there are an estimated nonillion

(or 1,000,000,000,000,000,000,000,000,000,000) prokaryotic microorganisms on Earth. Their abundance eclipses the number of stars in the known universe, the number of neurons in our brains, and all of our synapses combined. Collectively, prokaryotic genomes define a metabolic network with the potential to encode over 15 decillion (or 15,000,000,000,000,000,000,000,000,000,000,000) genes. Although many of these genes may encode redundant information (that is, proteins with the same function), their widespread distribution within microbial lineages guards against the loss of encoded functions and provides a molecular signature of biological diversity on Earth.

Over the past several decades, unprecedented technological innovation has enabled us to read the collective evolutionary memory of the microcosmos within single cells and whole microbial communities. Scientists and engineers are beginning to see the practical benefits of harnessing these memories in the development of next-generation biotechnologies. Through such efforts, we have come to understand how microbial communities function as groups of interacting cells that have evolved modes of metabolic cooperation. This cooperation is based on modular design principles in which metabolic processes are distributed among multiple community members. In this way, a complex metabolic process (such as the conversion of organic wastes into carbon dioxide and simple nutrients) is accomplished by a consortium of cells, each capable of a particular step in the overall biochemical pathway. This division of metabolic labour among and between microorganisms interacting in complex communities is not unlike the specialization of human actors in a complex economic system, albeit a more efficient and sustainable one.

By harnessing the metabolic problem-solving power of microbial communities, we can learn to optimize our use of natural resources and the production of energy and materials. Engineered microbial communities can perform complex tasks more effectively than can single cells, and they can be more resilient in the face of environmental perturbation. For example, plant biomass provides a renewable resource for energy and materials production and has the potential to replace our reliance on petroleum products. However, the lignin and sugar polymers that make up plant biomass must first be broken down into simpler building blocks before they can be used efficiently in a biorefining context. In nature, the capacity to decompose plant biomass is distributed among multiple interacting microorganisms. Studying the metabolic networks inherent in these communities provides practical wisdom in our search for efficient biorefining processes built on the same design principles that define microbial community structure and function in the world around us.

As we enter the Anthropocene, a new epoch of human experience on Earth, the conditions for life on the planet are rapidly changing in ways that were previously unimaginable and with impacts that are difficult to predict. Shared evolutionary memories encoded and stored in the microcosmos can be traced through the information content of genes, proteins, and cellular architecture across the Tree of Life. This diversity is also expressed in the collective metabolic potential of microbial communities in natural and engineered ecosystems. By reading the genomic sequences encoded in microbial communities, we can better understand the tendency of interacting cells to evolve modes of cooperation. Memory in the microcosmos can help us reimagine

innovation and redirect our own evolutionary trajectory through the development of bioprocesses that are more efficient and sustainable than any mode of energy or materials production in existence today.