



# NASA ASTROBIOLOGY INSTITUTE ANNUAL REPORT YEAR



[July 2001 – June 2002]

Project Report: Delivery of Organic Materials to Planets

## **Marine Biological Laboratory Executive Summary Principal Investigator: Mitchell Sogin**

*Organisms much like some of today's bacteria lived 3.5 billion years ago and were Earth's only inhabitants for as long as 2 billion years. Bacteria are remarkably adaptable: they exist in the bodies of all living organisms and on all parts of Earth—on land, in ocean depths, in arctic ice, in hot springs, and even in the stratosphere. They are largely responsible for the decay and decomposition of organic matter, soil formation, and the life-essential cycling of such chemicals as carbon, oxygen, nitrogen, and sulfur. A better understanding of how Earth's ubiquitous and simple early life forms evolved and developed into ever more complex organisms will help scientists identify and evaluate the habitability potential of extraterrestrial environments.*

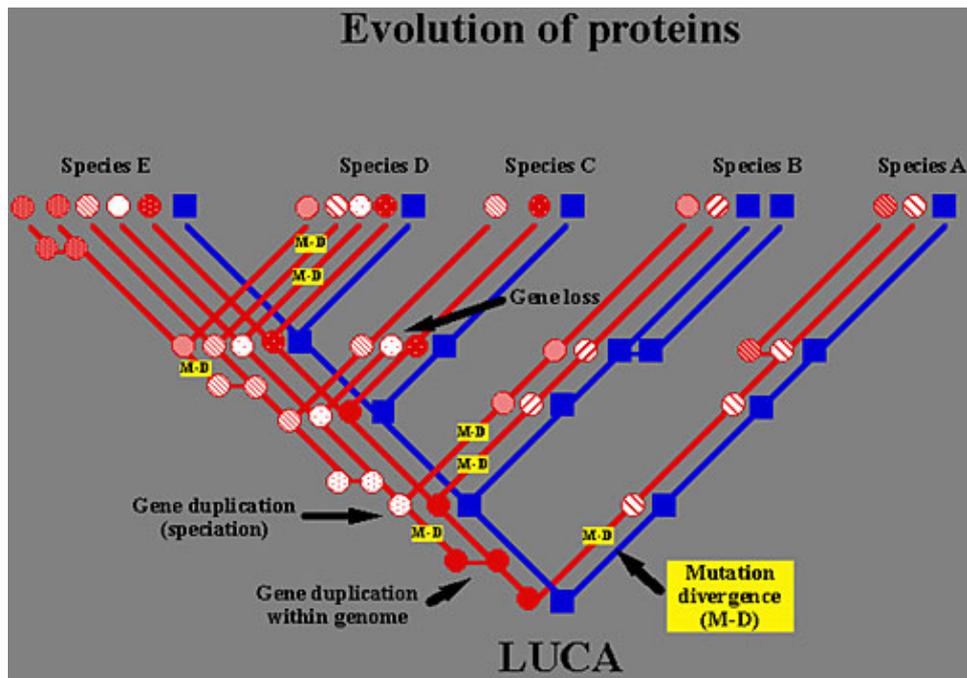
The astrobiology team at the Marine Biological Laboratory (MBL) studies the genomic basis of early life, its evolution into complex forms, and the dispersion of extant microorganisms on Earth and beyond. We harness the power of bioinformatics and molecular biology (1) to trace the gene duplication patterns and endosymbiotic processes that led to more complex genomes, (2) to identify the oldest eukaryotic lineages and the immediate protist ancestors of multi-cellular species, and (3) to explore the habitat range of prokaryotes and eukaryotes in extreme environments that might resemble those of early Earth or early Mars. Understanding the evolution and distribution of diverse single-cell organisms at the genome, organism, and ecological level will inform us about where and what to look for as we explore habitability and the distribution of life beyond our biosphere.

Microorganisms were the first descendants of Earth's pre-biotic world and they were the only form of life for at least 80 percent of our evolutionary history. Multi-cellular organisms, including plants and animals, are merely evolutionary after-thoughts that emerged from microbial ancestors 500 million to 1,000 million years ago. By the standards of multi-cellular plants and animals, single-cell organisms express relatively simple architectures, yet they transformed the atmosphere, waters, and upper sediments of Earth. Prokaryotes and single-cell eukaryotes continue to dominate our biosphere when measured in terms of genetic diversity and total biomass. Microorganisms are directly or indirectly responsible for all biogeochemical cycles on Earth. Cyanobacterial photosynthesis in the oceans produces more oxygen than do all higher plants. Microbial carbon re-mineralization, with and without oxygen, maintains the carbon cycle. Microorganisms control global utilization of nitrogen through nitrogen fixation, nitrification, and nitrate

reduction. All multi-cellular organisms are completely dependent on diverse microbial species for their continued survival. If life exists beyond Earth, compelling arguments hypothesize that such a world had its origins and continued manifestation in a microbial world. Any hope of understanding our evolutionary history or the kinds of life that might be encountered beyond Earth, depends on delineating the phylogenetic history, habitat-range, ecology, and diversity of microorganisms that shape our biosphere.

## Genomic Investigations

The origin of life was coincident with the accretion of sufficient genetic information in a minimal genome to define self-replicating organisms with evolutionary potential (Last Universal Common Ancestor, LUCA). We seek a description of gene families contained within early genomes and the diversification mechanisms that led to greater genome complexity with enhanced metabolic capability. Our bioinformatics group focuses on complex gene families from a single genome. By tracing the sequence and functional evolution of members of gene families, the minimal number of primordial genes that were ancestral to a spectrum of related activities has been identified (Liang, P. & Riley, M. (2001). A Comparative Genomics Approach for Studying Ancestral Proteins and Evolution. *Advances in Applied Microbiology*, 50: 39–72.).



**Figure 1.** Evolution of genes and proteins. Generation of families of genes/proteins from primordial ancestor genes through gene duplication and mutation. The diagram shows two genes in the Last Common Universal Ancestor (LUCA). The left-most ancestral gene undergoes duplication concomitant with generation of species, also duplication within a genome (horizontal lines connect the duplicates), and mutation which generates diversity (shown as changes in gene patterns). The left-most ancestral gene generates families of related genes, and these differ in kind and number from

species to species. The right-most gene undergoes less duplication and mutation.

For example, in *Escherichia coli* and other bacteria, Class III transaminases have three major regions of conservation: one binds the cofactor, pyridoxal phosphate, a second binds an amino acid donor, and a third is the core region. The close relationships of sequence and structure of transaminases reveal that they share a common transaminase ancestor. They differ in substrate specificity rather than in chemistry. Other protein families can differ not only in specificity of small molecule bound but also in the actual reaction catalyzed. The crotonase family and the thiamine– diphosphate–dependent decarboxylase family each may represent a single early ancestral protein. Each gave rise to a family of enzymes that carry out many kinds of reactions having only basic chemistry in common. Nahum, L.A., Goswami, S.K., Riley, M. (In preparation, 2002). A relatively few ancestral proteins could have generated the wealth of enzymes on Earth that catalyze many different kinds of reactions. It would appear that conserved features of super-families are signatures for very early molecules in the origin of life on Earth. Some of these signatures are particularly common in enzymes throughout the Tree of Life on Earth, such as ATP-binding sites and NAD-binding sites (Kerr, A., Riley, M. (Submitted, 2002). Superfamilies of proteins in the context of functional annotation: A study of NAD(P)–dependent dehydrogenases in *E. coli* K–12. *Omics*). One way to search for evidence of life beyond Earth would be to seek such signatures in proteins or fragments of proteins.

In contrast to gene duplication and metabolic diversification, genome streamlining and specialization occurs in response to the formation of obligate endosymbioses. Across phylogenetically independent lineages, the transition to such a lifestyle can be coupled with "reductive genome evolution." This process includes severe gene loss, changes in genome architecture, shifts in compositional bias, and elevated rates of evolution. For example, the loss of DNA repair genes may shift mutational biases and elevate evolutionary rates. We have shown that AT mutational bias largely drives amino acid and codon use in the genome of *Buchnera aphidicola* (643 kilobases or kb), the bacterial mutualist of aphids (Palacios, C. & Wernegreen, J.J. (In Press, 2002). A Strong effect of AT mutational bias on amino acid usage in *Buchnera* is mitigated at high expression genes. *Molecular Biology and Evolution*). A comparative framework for the better understanding of how the fundamental processes of mutation, selection, and genetic drift are altered in the context of endosymbiosis is being developed.

**Phylogenetic Studies** Analyses of ribosomal RNAs (rRNAs) reveal that three primary lineages trace their origins back to LUCA or to a population of primordial cells: the the Archaea, the Bacteria, and the Eukarya. The root of the Universal Tree of Life may lie within the Bacteria, while the Eukarya and Archaea shared a more recent common ancestor. The identity of the earliest diverging eukaryotes and lineages that were ancestral to animals and fungi remains an unsolved mystery. Unlike plants and animals, the geological record is silent for most microbial groups and interpretations of the most ancient, putative, prokaryotic fossils are under renewed scrutiny. Molecular phylogeny has to potentially fill this void in our understanding. The MBL Astrobiology

team explores both the phylogenetic history of eukaryotes and the range of conditions that are compatible with a eukaryotic lifestyle.

One striking feature of early rRNA "trees" is the basal location of taxa without mitochondria – such as the microsporidia, the trichomonads, and the diplomonad *Giardia*. But all of these are parasitic species and for at least one lineage, the microsporidia, the basal position in molecular trees may be artifacts associated with the limitations of tree inference technologies. We have turned our attention to studies of eukaryotes that on morphological grounds appear to be related to diplomonads. These so-called excavates include the free-living taxon *Carpodomonas*. Our analyses of small subunit ribosomal RNA genes and tubulins scatter the various excavate taxa across the diversity of eukaryotes. However, all phylogenies place the excavate taxon *Carpodomonas* as the closest relative of diplomonads and retortamonads (Simpson, A. G. B., Roger, A. J., Silberman, J. D. Leipe, D.D., Edgcomb, V. P., Jermini, L.S., Patterson, D. J., & Sogin, M. L. (In Press, 2002). Evolutionary history of 'early-diverging' eukaryotes: The excavate taxon *Carpodomonas* is a close relative of *Giardia*. *Molecular Biology and Evolution*). We conclude that free-living and parasitic taxa can represent distinct basal branches in the eukaryotic tree. The absence of mitochondria and peroxisomes in these organisms shows that eukaryotes could have evolved in the absence of elevated levels of oxygen, although we note that *Carpodomonas* has double membrane-bounded structures physically resembling hydrogenosomes. In some protists (e.g., anaerobic ciliates) hydrogenosomes are clearly degenerate mitochondria. However, the ancestry of these structures in basal eukaryotic lineages is still under debate.

Pelobionts might also represent a deep-branching eukaryotic lineage. They inhabit micro-oxic and anoxic environments and they lack stacked dictyosomes, mitochondria, and outer dynein arms in their flagellar apparatus. In contrast to molecular analyses of DNA-dependent RNA polymerase II that show the putative pelobiont *Mastigamoeba invertens* diverged early; our analyses of SSU (small subunit) ribosomal RNAs from several new pelobiont taxa position them near the eukaryotic "crown groups," but they are not specifically related to *M. invertens*. Based on re-examination of the *M. invertens* structure and its failure to converge with other pelobionts in molecular analyses, we conclude that this organism was not correctly identified. Furthermore we determined that Pelobionts are specifically related to Entamoebidae and that they do not represent a basal eukaryotic lineage (Edgcomb, V. P., Simpson, A.G.B., Amaral Zettler, L., Nerad, T.A., Patterson, D.J., & Sogin, M.L. (2002). Pelobionts are degenerate protists: Insights from molecules and morphology. *Molecular Biology and Evolution*, 19: 978–982).

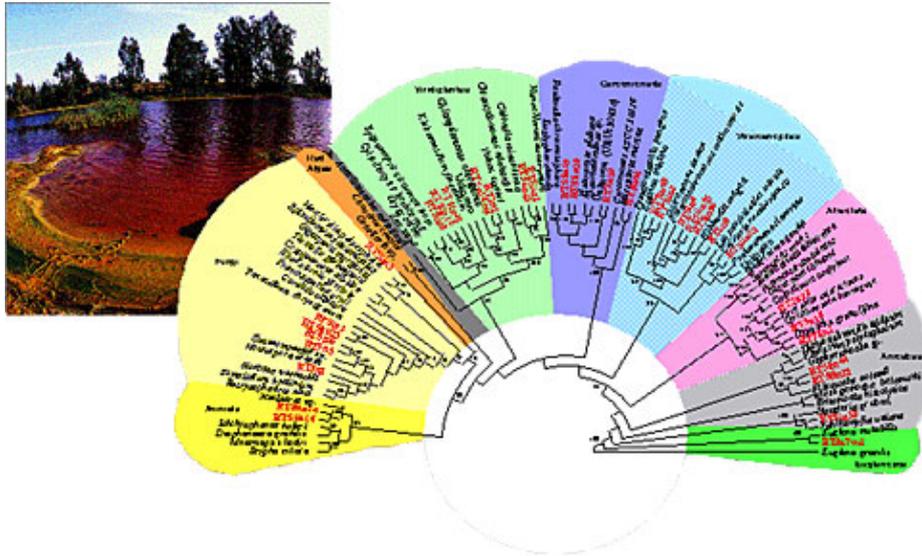
### **Studies of Extreme Environments**

Future clues about the evolutionary origins of eukaryotes will most likely come from studies of protists from anoxic, extreme environments. Of particular interest are warm anoxic sediments associated with hydrothermal vents and iron-rich environments that might be analogues for early life on Mars. Using sequence comparisons of PCR-amplified rRNAs, we have characterized both prokaryotic and eukaryotic diversity in hydrothermal vent environments of the

Guaymas Basin in the Gulf of California. These anoxic sediments and the overlying seawater harbor a mixture of genetically diverse protists (Edgcomb, V.P., Kysela, D.T., Teske, A., Gomez, A.D.V. & Sogin, M. L. (2002). Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the National Academy of Sciences*, 99(11): 7663–7668) whereas the bacterial and archaeal populations represent only a fraction of known prokaryotic phylogenetic diversity (Teske, A., Hinrichs, K.-U., Edgcomb, V., de Vera Gomez, A., Kysela, D., Sylva, S.P., Sogin, M.L., & Jannasch, H. W. (2002). Microbial Diversity of Hydrothermal Sediments in the Guaymas Basin: Evidence for Anaerobic Methanotrophic Communities. *Applied and Environmental Microbiology*, 68: 1994–2007). Many sequence-isolates represent novel protists, including early branching eukaryotic lineages or extended diversity within described taxa.

Two mechanisms, with overlapping consequences, account for the eukaryotic community structure of this environment. The adaptation to warm anoxic environments is superimposed against a background of widely distributed aerophilic protists, some of which may migrate into and survive in the sediment whereas others, for example, phototrophs, are simply deposited by sedimentary processes. The archaeal community in the Guaymas basin hydrothermal vent sediments is dominated by anaerobic methanotrophic archaea (ANME-1 and ANME-2 lineages). These anaerobic methanotrophs, in combination with sulfate-reducing syntrophs and methanogenic archaea, constitute an anaerobic microbial ecosystem that produces and oxidizes methane, and thus drives an oxygen-independent carbon cycle. Several phylogenetic lineages of novel, deeply branching key genes of sulfate reduction occur in these sediments, indicating diverse and possibly ancestral sulfate-reducing bacteria in this methane-oxidizing model ecosystem. A follow-up study is in progress (see below).

The MBL Astrobiology team has also joined forces with Spain's CAB to explore eukaryotic diversity in the Rio Tinto, a river in South Western Spain. This is an acidic (pH ~1.7–2.2), iron-rich (Fe concentrations of 20 mg/ml) drainage system. Our small, subunit, rDNA-based studies reveal an unexpectedly high, eukaryotic phylogenetic diversity, including several novel taxa, as well as cosmopolitan organisms never before reported from acidic/high-metal extremes. We identified many additional nonphotosynthetic lineages such as ciliates, amoebae, stramenopiles, and fungi that had escaped detection by traditional methods, which contrasts with the relatively low prokaryotic diversity. This remarkable study demonstrates that protists can thrive and dominate extremely acidic, heavy, metal-laden environments. Since some of these acidophiles are closely related to cultured neutrophiles, eukaryotes must have the ability to adapt from neutral to acidic environments over relatively short geological time-scales. Finally, eukaryotic extremophiles are more widely distributed and phylogenetically diverse than previously thought.



**Figure 2.** The Rio Tinto at Palma del Condado. The Rio Tinto gets its red color and name (“Tinto” means red in Spanish), from the high levels of iron dissolved in its acidic waters. A minimum evolution phylogeny for small subunit rRNAs using a likelihood model. Bold letters indicate environmental clones. “RT” indicates the sequence is from the Rio Tinto and “cul” indicates cultured species. Underlined taxa represent genera that have been identified in the river based on microscopic observation. Sampling sites were as follows: RT1, La Palma; RT3, Berrocal Upper; RT5i, the Origin, black filamentous biofilm; RT5ii, the Origin, green filamentous biofilm; RT7i, Anabel's Garden green biofilm; RT7ii, Anabel's Garden yellow biofilm. Bootstrap support values are shown and the scale-bar represents the number of substitutions per site. GenBank accession numbers AY082969–AY083001.

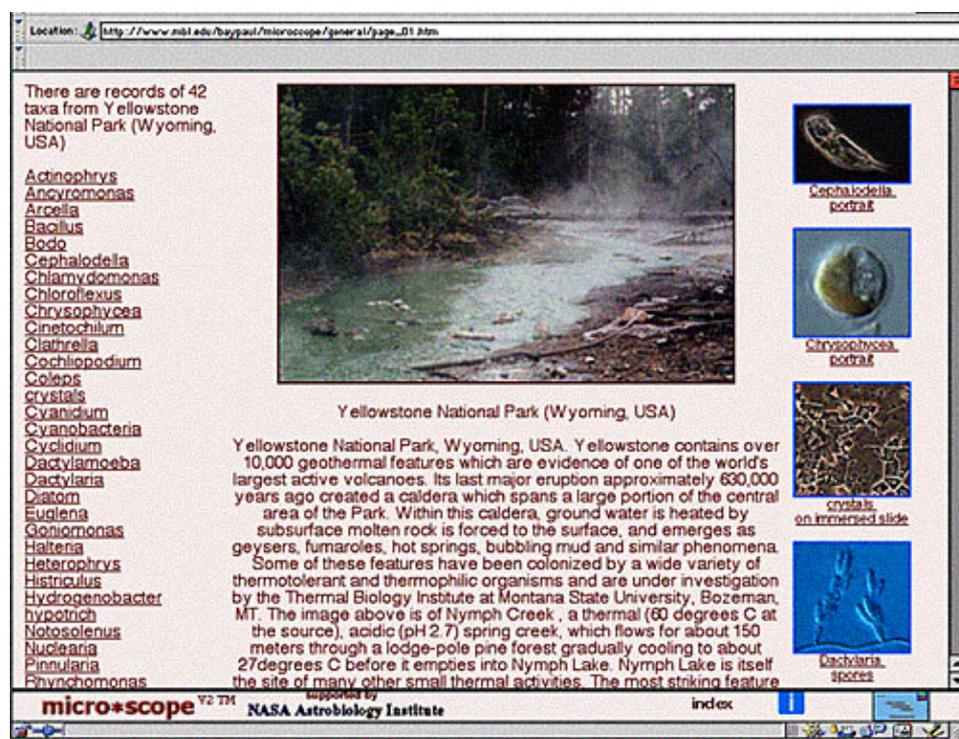
These studies identify organisms that live under acidic/high-metal extremes, but tell us little about how these organisms have adapted to such environments. It has been demonstrated that a 40,000-fold transmembrane  $H^+$  gradient in these protists is maintained by active mechanisms. The next step will be to isolate and characterize ion-transporting ATPases in a cultured acidophilic chlamydomonad and euglenid from the Rio Tinto, and to compare these genes with those from neutrophilic counterparts. We have partially sequenced members of the heavy-metal transporting ATPase family. We predict that special properties of these ion transporters allow protists to survive in the Rio Tinto.

### Ecogenomics

Despite the central role of microorganisms of untold diversity in shaping planetary environments, little is known about how they orchestrate and drive key biogeochemical cycles. The MBL Astrobiology team participates in three cross-institute projects organized through the Ecogenomics Focus group. The first is a collaboration with the University of Washington Astrobiology team in which the occurrence of dissimilatory sulfite reductase (DSR) genes in Guaymas and ultimately in the Guerro Negro is being explored. The second is a genomics-based study with Ames Research Center in which DNA

microarrays will be used to study gene expression of cyanobacterial isolates from the hypersaline mats of the Guerro Negro. The large-scale sequencing phase was recently completed and we are constructing DNA micorarrays. And third, with funding through the CAB, a high-throughput, molecular technique for studying microbial population structures is being developed. The technology is patterned after SAGE (serial analysis of gene expression) and it has the potential to accelerate studies of microbial population structures by at least an order of magnitude. We have tested the system using samples from Guaymas and a manuscript describing the technique is in preparation.

The evolution of genomes, microorganisms, and entire ecosystems unites the scientific investigations supported by the MBL Astrobiology program. We have constructed a unique WWW site micro\*scope (<http://www.mbl.edu/microscope>). It is neither a database nor a single site. It is an



**Figure 3.** Yellowstone National Park: an example of extreme environment represented in micro\*scope.

image-rich, research and teaching resource that re-aligns morphological, physiological, evolutionary, and ecological data for diverse microorganisms. A compilation of all synonyms, colloquial, or alternative names is used to find, collate, and index all taxon-related data through the internet. Micro\*scope provides identification guides (Lucid guides) for determining the identity of microbial isolates, interrogates the internet to gather information, and presents the assembled data from a biocentric perspective. Micro\*scope organizes information about microbial diversity according to multiple taxonomic hierarchies, adaptive kinds of organisms, for example, algae, flagellates, ciliates, fungi, and prokaryotes, and types of environments, for example,

marine, fresh water, sediments, and extreme. Although originally designed for outreach and education, micro\*scope also serves the research community. In the future, micro\*scope will include data from molecular and morphological databases in a phylogenetic context. The ultimate objective of micro\*scope is to provide peer-to-peer communications between internet teaching and research tools including, but not limited to, the Ribosomal Data Base (RDP) and the NASA Science Organizer.