

Ribosomal RNA, the lens into life

DANIEL MCDONALD,^{1,2} ZHENJIANG XU,² EMBRIETTE R. HYDE,³ and ROB KNIGHT^{3,4}

¹Department of Computer Science, University of Colorado at Boulder, Boulder, Colorado 80309, USA

²BioFrontiers Institute, University of Colorado at Boulder, Boulder, Colorado 80303, USA

³Department of Pediatrics, University of California San Diego, La Jolla, California 92093, USA

⁴Department of Computer Science & Engineering, University of California San Diego, La Jolla, California 92093, USA

Our world is, to a first approximation, a microbial world—even our own lives are evolutionarily and molecularly linked to that which we cannot see. This “invisible” world contains nearly unfathomable molecular and genetic diversity. Throughout the 20th century, a metaphor of “war against disease” prevailed: Microbes were hunted under the guise of health and agriculture, the goal being the eradication of “pathogens” at a global scale. Microbes were considered at best a nuisance and at worst a threat: a target for technological solutions. Humanity invested heavily, with the broad support of the scientific community, in the advancement of weapons for the war against microbes. We sought out novel antibiotics and antimicrobials as our enemy mounted resistance.

However, despite the profile of microbial pathogens as sources of human suffering needing to be eradicated, the scientific mindset still had room to respect the unknown. The evidence of what we might miss were there from the beginning. Even Pasteur noted that many microbes were beneficial, especially in industry, and doubted that humans could live without their microbial guests. Beijerinck wrote in the early 1900s that “...in its primitive form life is like fire, like a flame borne by the living substance;—like a flame which appears in endless diversity and yet has specificity within it.” Now, by viewing the microbial world through the lens of RNA, we are beginning to see that this flame of diversity can be a powerful ally.

In the 20 years since it was founded, *RNA* has been a key partner in a dramatic expansion of our ability to see into the microbial world, including the subtle effects of the beneficial microbes that overwhelmingly outnumber pathogens. In particular, using ribosomal RNA not just as an object of study in itself but as a tool to investigate the structure of microbial communities has led to a fuller appreciation that, in nature, no organism lives in isolation. Advances in molecular techniques, improvements to sequencing technologies, expansion of databases, and development of scalable software

have all played key roles in characterizing microbial communities in environments ranging from the deep ocean to the air to our own bodies. In many ways, these new insights have come from seeing RNA as a tool, at higher and higher levels of organization.

The first of these conceptual leaps came from the work of Carl Woese and George Fox in the late 1970s, who realized that universally conserved components of the translation apparatus, first the 5S and then the longer small subunit ribosomal RNA gene, could be not just objects of study in the context of their role in translation, but could also be utilized as key markers for evolution that could be used to interpret large-scale patterns of the evolution of life. The reorganization of life’s diversity from a five-kingdom model (plants, animals, fungi, protists, and bacteria) to a three-domain model (bacteria, archaea, and eukaryotes), with the revelation that most diversity was found not in organisms visible to the naked eye but rather in the microbial world, was in many ways as fundamental to our understanding of our place in the universe as the Copernican shift from a geocentric to a heliocentric model. This idea, that rRNA could be used to read out an organism’s place in the tree of life, rapidly led to a proliferation of rRNA sequences in the database, as many investigators sought to place their favorite organisms on this universal tree of life.

The second conceptual leap was to realize that in addition to placing known organisms on the tree of life, rRNA used as a universal marker could also place unknown organisms, and thus act as a tool for identifying novel members of microbial communities. Norm Pace and members of his laboratory developed environmental PCR, which allowed direct investigation of the organisms in an environmental sample without resorting to culture-based techniques that can only grow a small fraction of the microbes in a given sample. Norm, together with David Lane, Mitch Sogin, Phil Hugenholtz, and many others rapidly expanded the tree of life to encompass many previously unsuspected lineages, the vast majority of which still have not been cultured successfully today. As

Corresponding author: robknight@ucsd.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050799.115>. Freely available online through the *RNA* Open Access option.

© 2015 McDonald et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

SSU rRNA sequences began to amass from a multitude of environments, identifying the kinds of communities in each of those environments, the need for comprehensive references arose. These resources, including the Ribosomal Database Project, Greengenes, and SILVA, centralized our growing knowledge of the various lineages of microbes that were out there, and especially in the case of Greengenes, highlighted the vast “dark” portions of the tree in which no organisms have been cultured. Of the phyla recognized within the kingdom Bacteria, approximately 30 are represented by organisms that have been successfully cultured, whereas 50–100 candidate phyla have been reported by different authors. As resources grew, taxonomies based on phylogeny highlighted the gross errors present in the published literature, often stemming from earlier classifications based on morphology or biochemistry, which evolve in a far less clock-like fashion than DNA sequences.

The third conceptual leap was to realize that as well as comparing organisms to one another in the context of a single community, the universal tree of life could be used to compare whole communities to each other, thus providing an overall picture of which factors drive microbial community composition. The initial idea for this came from a discussion after one of Norm’s lab meetings, where the problem of comparing five communities from one of Jeff Walker’s studies on endolithic communities (microbes that live in rocks) arose. Andy Martin had devised a test called the P test, which could use the tree to test whether two communities were identical. The problem was that all the communities were statistically significantly different, and where do you go from there? One of us (Rob Knight) realized that the goal wasn’t so much to find out whether the communities were different, because all microbial communities will differ in some respect, but rather to tell how far apart evolutionarily they were from one another. With this distance metric in hand, it would then be possible to perform ecological techniques such as ordination based on distances from the phylogenetic tree, and to find out which environmental factors were most important in determining which microbes live in which communities. Cathy Lozupone completed a very exciting PhD thesis in Rob’s lab developing this distance metric, called UniFrac (for “Unique Fraction”, i.e., the fraction of the evolutionary tree unique to one lineage or another) and set up an easy-to-use web site in conjunction with another graduate student, Micah Hamady. The Knight lab then proved UniFrac’s metric properties together with Manuel Lladser. Shortly after developing UniFrac, Cathy set out to ask a simple and straightforward question: Is there a primary environmental factor that differentiates microbial communities? Through an analysis that spanned over 200 papers (although she had to read over 400 to find the ones that had used general-purpose primers and that had deposited their data), Cathy saw that communities associated with saline environments were systematically different than those associated with non-saline, with brackish environments such as estuar-

ies falling between them (this factor, a long way down the list of factors examined, outweighed temperature, pH, etc.). Following this remarkable observation, Ruth Ley asked how communities associated with the mammalian gut fit in, and much to our surprise, the most important factor was whether the community was environmental or host-associated: This was twice as important as the saline/non-saline split in environmental samples. This observation suggests that the selective pressures put forth by the gastrointestinal tract are more important than those present in any other environment.

UniFrac came along at just the right time because following on the heels of the Human Genome Project, “next generation sequencing” instruments enabled a completely new scale of DNA sequencing. Coupled with environmental PCR, we and other labs began to generate tremendous amounts of SSU rRNA amplicon sequence data. For example, in late 2007, the core facility down the hall from the Knight lab was still charging \$8 per sequence; shifting to the 454 instrument, in the first deployment of our highly multiplexed barcoded sequencing protocol with Kirk Harris, Jeff Walker, Nick Gold, and Micah Hamady, we collected 500,000 sequences for \$12,000. If we had done this the old way, it would have cost \$4 million, and they wouldn’t be finished with the sequencing yet! Together with other pioneers in amplicon barcoding such as Rick Bushman and Mitch Sogin, and computational tools such as QIIME, developed in my lab by Greg Caporaso, Jesse Stombaugh, Justin Kuczynski, and a host of others, we enabled these protocols to be deployed in a huge range of environmental samples. These sequences came from every conceivable environment on the planet, and previously unobserved life was popping up everywhere.

Enter the poop. Through a back of the envelope calculation in the 70s, Dwayne Savage estimated the total number of microbial cells in and on the human body to be up to 100 trillion; 10-fold more than the number of human cells in a body. The vast majority of these microbial cells reside in the large intestine, and the structure of these communities can be observed (in approximation) by proxy through the end result of last night’s dinner. Changes in microbial communities have now been linked to a wide range of conditions, including inflammatory bowel disease, colon cancer, cardiovascular disease, and rheumatoid arthritis in humans, and, in mouse models, multiple sclerosis, depression, and even autism. However, simply observing a community does not provide a mechanistic understanding of the interplay between the microbiome and the host. To answer this question, we need hosts, such as mice, that completely lack microbes and which can be experimentally inoculated with known microbes under controlled conditions. Jeff Gordon’s group, through an interest in determining the causes of obesity in the mid-2000s, brought the use of gnotobiotic mice to scale, developing a framework for assessing causality in microbiome studies. Remarkably, phenotypes can be transferred not just from mouse to mouse but even from human to mouse by transferring the microbiome. For example, germ-

free mice, when inoculated with the communities from humans who have Kwashiorkor, a pervasive wasting disease in developing countries, will develop similar symptoms. Similarly, germ-free mice inoculated with bacteria from an obese human will gain significantly more adipose tissue than those inoculated with the microbiome of a healthy human. These phenotypic changes are coupled to changes in the microbial community, which can be read out from the RNA. These results cemented the idea that the microbiome should be thought of as a vital organ, and, much as we take care to exercise for cardiovascular health, it is imperative that we take care of our microbial friends through eliminating unnecessary antibiotic use, regularly consuming dietary fiber (which feeds the butyrate producers in the large intestine), and other measures.

The fourth, and most recent, conceptual leap is occurring now: Rather than using the microbial communities as objects of study in and of themselves, we can use the communities as a tool to read out environmental or medical conditions. For example, we can tell today if a person is lean or obese with 90% accuracy based on the microbial community in a person's stool: On the one hand, there are easier ways to tell if someone is fat, but on the other we can only do this with 58% accuracy using every human gene ever linked to obesity by genome-

wide association studies. Similar potential has been shown for reading out diabetes and cirrhosis from the human microbiome, and for reading out soil pH and nitrogen content, oil pollution in marine water and sediment, stress in populations of primates, post-mortem interval of a corpse, and a wide range of other conditions. An exciting new frontier is in the dynamics of the microbiome, where it may actually be how your microbiome changes, rather than a static snapshot, that can best be linked to your physiological state.

Taken together, what is remarkable is that in such a short timespan, ribosomal RNA has moved from being an object of study in its biochemical role, to a tool for placing organisms on a phylogenetic tree, to a tool for understanding who lives in a given microbial community, to a tool for relating communities to one another, and most recently to a tool for reading out the properties of an organism or environment for the communities it harbors. As sequencing technologies and software to interpret all those sequences continue to advance, it will be fascinating to see how these applications, and conceptual advances we have not yet even begun to anticipate, will cement the central role of RNA as a marker for the microbial world, an instrument to conserve precious microbial biodiversity, and an enabling technology to improve human and environmental health.



RNA

A PUBLICATION OF THE RNA SOCIETY

Ribosomal RNA, the lens into life

Daniel McDonald, Zhenjiang Xu, Embriette R. Hyde, et al.

RNA 2015 21: 692-694

-
- Open Access** Freely available online through the *RNA* Open Access option.
- Creative Commons License** This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.
- Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).
-



Rudi Micheletti uses LNA™
GapmeRs to silence cardiac lncRNAs
www.exiqon.com/gapmers

EXIQON

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>
