

---

*This copy is for your personal, non-commercial use only.*

---

**If you wish to distribute this article to others**, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

**Permission to republish or repurpose articles or portions of articles** can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of November 6, 2014 ):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/346/6210/698.full.html>

This article **cites 15 articles**, 8 of which can be accessed free:

<http://www.sciencemag.org/content/346/6210/698.full.html#ref-list-1>

## EVOLUTION

# Searching for new branches on the tree of life

Is there undiscovered life that differs fundamentally from that in the three known domains?

By **Tanja Woyke** and **Edward M. Rubin**

Ever since Woese's seminal work nearly 40 years ago (1), life has been divided into three domains: Archaea, Bacteria, and Eukaryota. But could there be life that does not fit into any of these domains? Novel techniques for exploring microbes that cannot readily be grown in the laboratory offer hope that scientists can discover such life, if it exists (see the first figure). These methods include metagenomics, which involves the sequencing of DNA extracted from environmental samples, and single-cell genomics, where individual microbial cells are isolated from the environment and their genomes amplified and sequenced.

On the basis of these and other approaches, we propose that microbial life be operationally divided into three categories: explored, unexplored, and undiscovered. It is among the latter that potential signs of additional branches on the tree of life beyond the three known domains may be found.

**EXPLORED, UNEXPLORED, UNDISCOVERED.** The explored category encompasses microorganisms that can be cultivated in the laboratory. The unexplored category includes uncultivated organisms present in

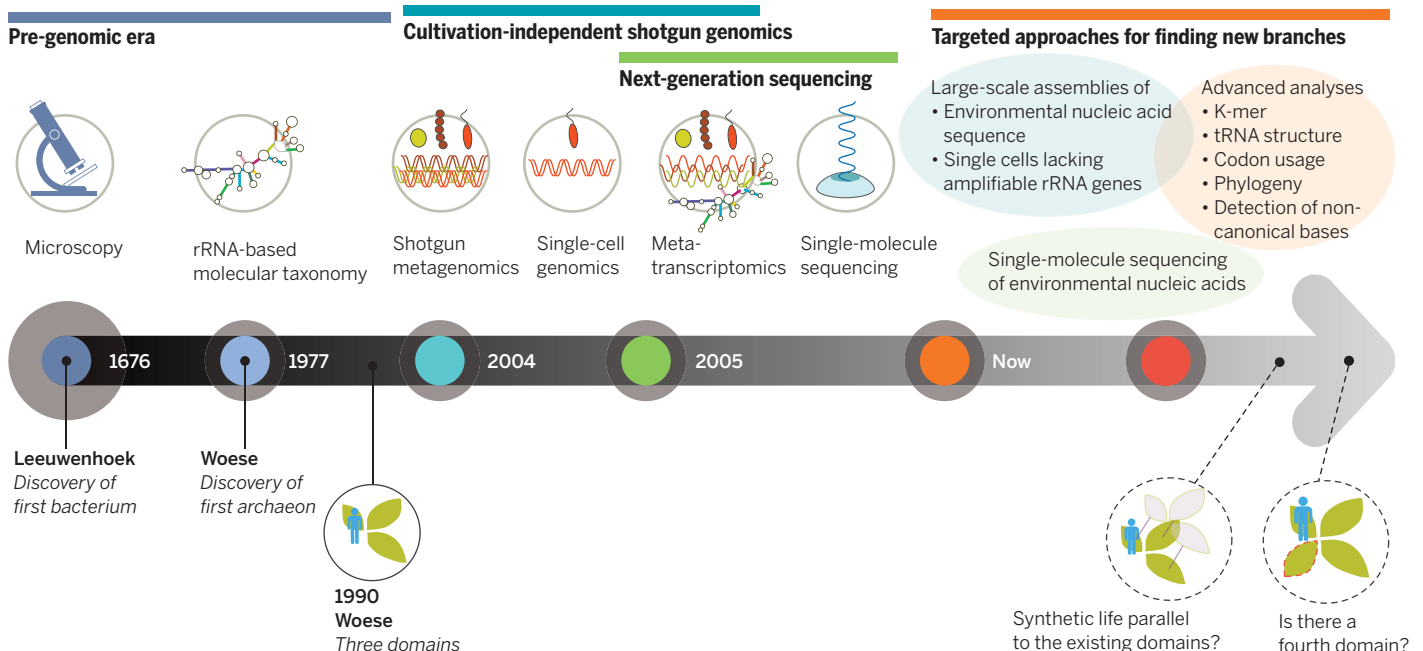
environmental samples, whose existence is known only through their molecular signatures and occasionally from partial genome assemblies obtained through metagenomics (2) and single-cell genomics studies (3). The sequence of the 16S ribosomal RNA (rRNA) gene, which can be amplified from environmental samples with a set of "universal" primers, has been extensively used as a molecular signature to assess the microbial diversity in a given sample, and to build phylogenetic trees. Taxa from the unexplored category dwarf the explored in both numbers and diversity (4). Beyond these organisms for which we have ribosomal barcodes or other molecular signatures resides the as yet undiscovered life—the putative organisms that have thus far eluded our detection.

Undiscovered life, if it exists, is either absent at the locations of existing environmental surveys or is missed by current approaches. There are reasons to believe that current approaches may indeed miss taxa, particularly if they are very different from those that have so far been characterized. The "universal" primers used to detect 16S rRNA genes from bacteria and archaea in environmental samples can miss major lineages because of primer mismatches (5). Similarly, the selection of specific single cells

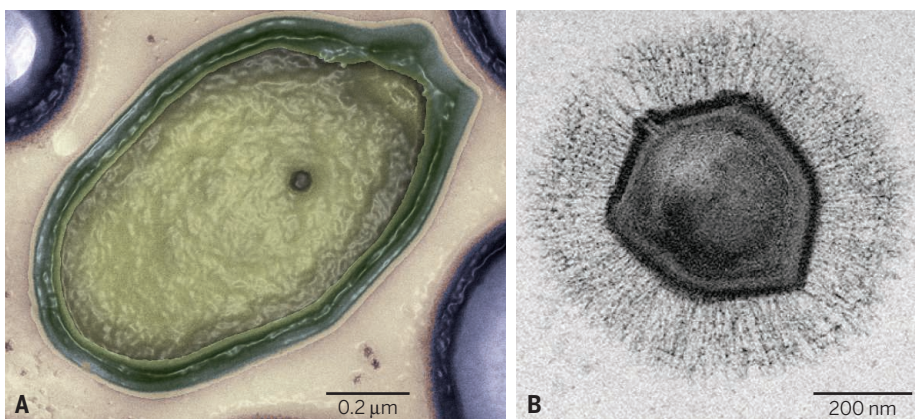
from environmental samples for genome sequencing has been based on rRNA gene identity, thus also relying on these universal primers. Organisms whose 16S rRNA genes are not recognized by the primers would not be detected using this approach. Past explorations of available metagenomic data sets have focused on the discovery of matches to the known genes and genomes—an analysis that is naturally biased against uncovering completely novel life. Finally, although we may soon have petabytes of metagenomic sequence data, samples have been collected from only a minute fraction of Earth's countless different environments.

Recognizing these limitations, it is reasonable to speculate that undiscovered and highly divergent branches of life may exist, possibly represented by domains whose marker genes differ extensively from those of the bacterial and archaeal branches on the tree of life. Refined strategies, involving both the application of new approaches and access to previously unexplored habitats, will be required for their discovery.

**THE SEARCH FOR UNDISCOVERED LIFE.** Approaches to further explore the diversity of microbial life will need to include the expansion and optimization of methods used



**The search for new major branches on the tree of life.** Cultivation-independent methods, novel sequencing technologies, and analytical approaches can be directed toward the detection of life outside currently established domains.



**Parasitic remnants of a fourth domain?** Electron microscopy images of (A) pandoravirus dulcis particle and (B) acanthamoeba polyphaga mimivirus particle. These viruses may be remnants of a fourth domain. Cellular entities from this domain either went extinct or remain to be found.

to capture the genomes of unexplored (uncultivated) organisms (see the first figure). Single-cell sequencing with microfluidic and cell-sorting approaches, focused specifically on cells lacking amplifiable rRNA genes, is a high-throughput strategy to search for novel organisms. Massive-scale metagenomic sequencing of environmental DNA and RNA samples should, in principle, generate sequence data from any entity for which nucleic acids can be extracted. The analysis of such sequence data with novel computational methods focused specifically on discovering outliers to previously defined life is another powerful means to explore the unknown. Assembled contiguous stretches of sequence data generated from environmental samples can be mined for unusual features in nucleotide composition, transfer RNA structures, and codon usage, as well as phylogenetic placement of rRNA and other marker genes (6). This would facilitate the detection of biological outliers and genomic fragments with deep phylogenies.

Finally, the application of single-molecule sequencing technologies that can recognize modified or nonstandard bases is another approach for the detection of life that may differ from the life we know. Sequencing to date has mostly been limited to the detection of the canonical four bases. Emerging new techniques, such as single-molecule real-time DNA sequencing (7) and nanopore-based single-molecule sequencing (8), have the potential to allow the recognition and characterization of environmental organisms with base modifications and compositions distinct from the four standard bases and their currently described modifications.

Beyond a new set of technologies and methodologies, the choice of suitable envi-

ronmental niches will be critical in cataloging the diversity of life. Although discoveries may be made by mining existing data sets from explored environments, future searches for deeply branching clades should include inhospitable and isolated environments. These sites would be expected to be preferred niches for early life, potentially sheltered from more modern microbial competitors. This would include hypoxic subsurface sites with environmental conditions predating the Great Oxidation Event that occurred about 2.3 billion years ago. Support for the idea that isolated hypoxic environments may be preferred niches for early life comes from observations that anaerobic niches deep within Earth's crust tend to harbor acetogens and methanogens; these species represent ancient branches of autotrophs in the bacterial and archaeal domains (9).

Using some of the aforementioned strategies, scientists are already speculating on the possible discovery of a “fourth domain.” Phylogenetic analysis of marker genes from metagenomic data revealed the presence of deep, novel branches thought to be occupied by either a cellular entity or novel viruses (6). Extremely large and unusual DNA viruses such as mimivirus and pandoravirus (see the second figure) (10) have also been shown to contain marker genes with deep phylogenetic roots between the archaeal and eukaryal domains (11). One interpretation of these findings is that there was an ancestral, cellular lineage that gave rise to these viruses but went extinct as a cellular entity, and that its genome now only is present as a so-called parasitic fourth domain (11). It is possible, however, that this cellular precursor has simply not yet been detected and still exists awaiting discovery. Even more speculative is the idea that an RNA world still exists in a niche of favorable conditions. A recent survey on the distribution of viruses among the domains of life found RNA viruses to be solely

associated with bacteria and eukaryotes; the authors hypothesize that these viruses are remnants of an RNA world (12). Under the appropriate conditions, cellular entities with RNA genomes may still persist.

**A BRAVE NEW WORLD.** While the search for new life is focused on organisms that exist in nature, a parallel effort is under way to create fundamentally new organisms in the laboratory. Organisms with an expanded genetic alphabet of six nucleotides (13) and with noncanonical codon usage (14) have already been built, representing human-designed variations of existing branches. This quest of synthetic biologists to build radically novel organisms also offers possible models for unusual varieties of life that may be sought in nature. A recent computational analysis of a global metagenomic data set revealed an unexpected abundance of environmental organisms with diverse noncanonical codon usage in nature (15)—a feature that has been targeted for creation in the laboratory through genome engineering (14).

An advanced toolkit of powerful genomic technologies is now poised to generate and mine increasingly large data sets for hints of life that differs strikingly from the life cataloged thus far. The discovery of new building blocks and organisms from a new domain would likely have major implications for biotechnology, agriculture, human health, and synthetic biology efforts. It might also elucidate the early evolution of the domains and their divergence from the last universal common ancestor. Irrespective of the depth at which newly discovered branches may be anchored in the tree of life, the quest to find them will likely reveal unexpected and valuable insights about the fruits of more than 3 billion years of biological tinkering. ■

#### REFERENCES AND NOTES

1. C. R. Woese, G. E. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088 (1977).
2. K. C. Wrighton *et al.*, *Science* **337**, 1661 (2012).
3. C. Rinke *et al.*, *Nature* **499**, 431 (2013).
4. N. R. Pace, *Science* **276**, 734 (1997).
5. B. J. Baker *et al.*, *Science* **314**, 1933 (2006).
6. D. Wu *et al.*, *PLOS ONE* **6**, e18011 (2011).
7. J. Eid *et al.*, *Science* **323**, 133 (2009).
8. G. F. Schneider, C. Dekker, *Nat. Biotechnol.* **30**, 326 (2012).
9. W. F. Martin, *FEBS Lett.* **586**, 485 (2012).
10. N. Philippe *et al.*, *Science* **341**, 281 (2013).
11. J. M. Claverie, *Commun. Integr. Biol.* **6**, e25685 (2013).
12. A. Nasir, P. Forterre *et al.*, *Front. Microbiol.* **5**, 194 (2014).
13. D. A. Malyshev *et al.*, *Nature* **509**, 385 (2014).
14. M. J. Lajoie *et al.*, *Science* **342**, 357 (2013).
15. N. N. Ivanova *et al.*, *Science* **344**, 909 (2014).

#### ACKNOWLEDGMENTS

Supported by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, under contract no. DE-AC02-05CH11231. We thank N. Kyrpides, N. Ivanova, and A. Visel for critical feedback.

10.1126/science.1258871

DOE Joint Genome Institute, Walnut Creek, CA 94598, USA, and Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. E-mail: emrubin@lbl.gov