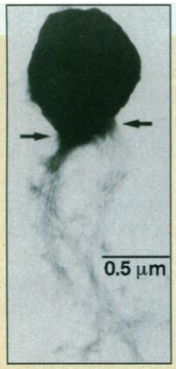
LETTERS

Relations



W. JACK JONES

The significance of a "complete genome sequencing" of an archaeon for our understanding of metabolism and taxonomy is discussed (right, *Methanococcus jannaschii*). The effort to sequence "the human genome" raises questions about cost, privacy, and "social stigmas." Although debating creationists is likened to "a boxing match," readers are encouraged to engage in "open, fair, honest, and well-informed disputation." "A robust theory" like evolution is said to have "nothing to fear from contradictory data."

Methanococcus Genome

The 23 August article by C. J. Bult *et al.* (p. 1058) about the sequencing of the *Methanococcus jannaschii* genome implies that the archaea tree theory has been confirmed by "complete genome sequencing and analysis," and the accompanying Research News article (V. Morell, p. 1043) contains a photo caption which prominently announces that "[t]he new archaeon sequence vindicates Carl Woese's theory that life is divided into three domains." However, a basic tenet of molecular evolution holds that a minimum of four groups is needed to test alternative unrooted trees (*1*). Only three types of genomes were available to Bult *et al.* (eubacteria, methanogens, and eukaryotes). Therefore, any statement that this complete genome analysis vindicates Woese's theory is unsupported.

The archaeal theory and the eocyte theory are based on two competing hypotheses relating the following four relevant taxa (*2*): eubacteria, methanogens (or better yet, halobacteria), eocytes (crenarchaeota), and eukaryotes. The archaeal theory proposes that the eukaryotes share a most recent, common ancestor with all archaea [halobacteria, methanogens, and crenarchaeotes (eocytes)]. The eocyte theory proposes that eukaryotes share a most recent, common ancestor only with the eocytes (crenarchaeotes). While the first analyses of tree topology in which 18S recombinant RNA sequences were used generally favored the archaeal tree, analyses of the EF-1α gene since 1992 have predominantly favored the eocyte tree (*3*).

Discovering the origin of the eukaryotes is one of the central questions of molecular evolution, and we now have the chance to resolve it. Both the archaeal theory and the eocyte theory are based on testable mutually exclusive tree topologies. If one theory is right, the other must be wrong. In the end, when the necessary genomes are available, and when methods are developed to properly test these theories at the genome level, we will know the answer.

James A. Lake
Maria Rivera
Molecular Biology Institute, University of California, Los Angeles, CA 90024–1570, USA
E-mail: lake@mbi.ucla.edu

References and Notes

1. D. M. Hillis and C. Moritz, *Molecular Systematics* (Sinauer, Sunderland, MA, 1990).
2. M. C. Rivera and J. A. Lake, *Science* **257**, 74 (1992).
3. Favor eocyte tree: S. L. Baldauf, J. D. Palmer, W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7749 (1996); T. Hashimoto and M. Hasegawa, *Adv. Biophys.* **32**, 73 (1996); J. A. Lake, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1457 (1994); B. Runnegar, *Early Life on Earth*, S. Bengston, Ed. (Cambridge Univ. Press, Cambridge, UK, 1993); M. Hasegawa, T. Hashimoto, J. Adachi, in *Origin and Evolution of Prokaryotic and Eukaryotic Cells*, H. Hartman and K. Matsuno, Eds. (World Scientific, Singapore, 1993), pp. 107–130; B. Cousineau, C. Cerpa, J. Lefebvre, R. Cedergren, *Gene* **120**, 33 (1992). Favors archaeal tree: R. Creti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3255 (1994).

Bult *et al.* state (p. 1067) that "[t]he ability to fix nitrogen has been demonstrated in a number of methanogens . . . and all the genes necessary for this pathway have been identified in M. *jannaschii* (Table 1)." However, when one searches their table 1 for nitrogen fixation genes, under the heading "Nitrogen metabolism," one finds only three possible *nif* genes: "nifB prot[ein]" and two labeled "nitrogenase RDase [reductase]" and "nitrogenase RDase rel[ated] prot[ein]." This set is far from sufficient to permit the organism to fix nitrogen. We then searched the whole sequence for possible orfs with a similarity to NifD and NifK, the two components of nitrogenase itself. They do not seem to be present, nor are the proteins required for molybdenum transport, molybdenum cofactor synthesis, or cofactor insertion into nitrogenase.

Robert Haselkorn
William J. Buikema
Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637, USA

Now that Woese's division of life on Earth into three categories (prokaryotes, eukaryotes, and archaebacteria, or their rechristenings as bacteria, eukarya, and archaea) is gaining acceptance, it is time to agree on an official name for this highest taxonomic level. Rather than using designations such as "branches," "domains," "divisions," and the like, I propose the term "empire." Thus, the animal kingdom, the plant kingdom, and the fungal kingdom, would belong to the eukaryote empire; cyanobacteria and purple bacteria would be kingdoms in the prokaryote (or bacterial) empire; and euryarchaeota and crenarchaeota would be among the kingdoms in the archaean empire. For international use, empire/kingdom/phylum could become imperium/regnum/phylum, reverting to the Latin, of which Karl von Linné ("Linnaeus") surely would have approved.

Solomon W. Golomb
Communication Sciences Institute, University of Southern California, Los Angeles, CA 90089–2565, USA

Response: Lake and Rivera appear to confuse two issues, the uniqueness of the archaea and their relationships to the other major groups of organisms. The original definition of the archaea as a group of organisms with a distinct set of shared sequences and common physiological characteristics remains unshaken (and unaddressed) by their arguments. What this single genome, or any single archaeal genome, obviously cannot do is address fully the matter of the relationships of the archaea to other major groups. As noted by Lake and Rivera, that requires corresponding sequence data which allows one to span the full phylogenetic breadth of the archaea, that is, a significant amount of sequence data representing the crenarchaeota. But phylogenetic analyses of genome data must be taken in toto, rather than being focused on only a single gene, as Lake and Rivera have done with EF-1α. In our initial description of the *M. jannaschii* genome, we did not attempt to include phylogenetic analyses of all the genes, as such comprehensive detailed analyses would have substantially delayed the publication of the genome. We look forward to having additional data from members of the crenarchaeota, which should be available when the genomes of *Pyrobaculum aerophilum* and *Sulfolobus solfataricus* are sequenced.

The rooted tree of Lake and Rivera would, if correct, have an interesting consequence. This tree suggests that ancient ancestors of human beings would have possessed most of the characteristics shared by *Methanococcus* and crenarchaeotes. Hence, just as our ancestors of roughly 450 million years ago are called fish (even though we are sufficiently modified that it is not useful to think of us as fish), our ancestors of some billions of years ago would be called archaea.

Complete genome sequencing is a foundation for comprehensive characterization and interpretation of the biology of an organism. The approach that we have chosen for characterizing the *M. jannaschii*, *Haemophilus influenza*, and *Mycoplasma genitalium* genomes includes distinct steps. First, after obtaining the complete genome sequence, we try to establish the identification of the database sequences most similar to those in the genome; results for *H. influenza* and *M. genitalium* were presented, respectively, in two previous papers (R. Fleischmann *et al.*, Research Article, 28 July 1995, p. 496; C. Fraser *et al.*, Research Article, 20 Oct. 1995, p. 403) and are electronically available through The Institute for Genomic Research (TIGR) World Wide Web server

(http://www.tigr.org). This step in the analysis, while relatively stable, evolves with time as more closely related sequences appear in the databases. The assignment of function to genes is less direct (and often less certain). It requires synthesis of much data, including the spectrum of functions represented by related sequences (which in turn relies on the availability and accuracy of annotated functions for these related sequences), information about the presence (or absence) of genes for other functions, and information about the organism itself. The functional interpretation of genomic sequences therefore improves with time and with the input and suggestions of diverse researchers.

With regard to nitrogen fixation and *M. jannaschii*, Haselkorn and Buikema are probably correct. However, the *M. jannaschii* genome contains a large percentage of genes new to biology that are of unknown function. Bioinformatics and sequence comparisons can lead to the generation of many hypothesis, including our own, that must be tested and verified experimentally. We applaud the interest taken by our colleagues and encourage further constructive comment. A large number of other useful contributions have been made through the TIGR Internet site. Our goal is to provide an environment in which this information can be collected in a coherent manner, updated, and made available to the world.

Gary J. Olsen
Carl R. Woese
Department of Microbiology, University of Illinois, Champaign-Urbana, Urbana, IL 61801, USA
Owen White
J. Craig Venter
The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

■

Whose Genes Are They and How Can We Identify Them?

The new policy of the National Center for Human Genome Research (NCHGR) on informed consent for DNA sources for the Human Genome Project (E. Marshall, News & Comment, 27 Sept., p. 1788) may protect the identity of donors at a high price for the image of human genetic research. The need for detailed informed consent for DNA sources cannot be questioned. The issue is how the research effort manages the identity of the new sources and the justification for anonymity. Anonymity should not be required for donor protection if the NCHGR collaborates with consenting DNA donors who are at low risk of adverse psychosocial effects [for example, those of a mature age (say, 75) with no children or who are retired and on Medicare].

The problem with strict anonymity is the message it broadcasts about the nature of genetic information. The Human Genome Project will be an important landmark in the history of science and medicine. There is a public fascination with this effort that will only increase as "the sequence" is completed. Yet secrecy surrounding the often-asked question about the identity of the source will raise troubling questions. Why are the donors being hidden? What kind of threat does genetic analysis pose? Is this information about which we should be afraid or ashamed? Why are we paying billions for this information? Ironically, the elaborate mechanisms developed to protect the identity of the DNA sources through the new policy may foster the very social stigmas that the NCHGR seeks to avoid. While great care must be taken in the conduct of clinical genetic testing (*1*), overstating the risks will hinder the beneficial applications that justify the project and augment the psychosocial risks. Also, the NCHGR would make a strong