What is the minimal genome?

by David W. Ussery, Center for Biological Sequence Analysis, Dept. Systems Biology, The Technical University of Denmark, Kgs. Lyngby, Denmark.

COMMENTS ON **'Genetics Indicates Extra-terrestrial Origins for Life:**
The First Gene - Did Life Begin Following the Big Bang?', by Rhawn Joseph and N. Chandra Wickramasinghe, J. Cosmology 16: (2011).

An underlying assumptions in this paper is stated in the second sentence of the abstract: 'Life, as we know it, requires genetic information and no fewer than 382 genes.' This is based on studies with the parisitic bacterium Mycoplasma genitalium, with severely limited metabolism. This genome is famous in the bacterial genomics world as it was the second sequenced bacterial genome, and at the time thought to be the smallest bacterial genome. However, we now know of several other bacteria with reduced genomes smaller than M. genitalium. The 382 essential genes in M. genitalium are based on experiments where individual genes were deleted. However, just because a gene is essential for M. genitalium under ones set of growth conditions does not necessarily mean that this gene is necessary for all life, nor is it safe to infer that these genes would be found in other organisms. For example, the genome of Hodgkinia cicadicola encodes a mere 169 proteins, and contains another 19 structural RNA genes, for a total of 189 genes, a bit less than half of the 'minimal number' mentioned by Joseph and Wickramasinghe.

I think it's important to discuss the difference between a 'minimal genome' and an optimal, 'streamlined genome'. In the field of synthetic biology, there has been much talk of the 'smallest genome'; the smallest free-living bacteria contains more than a thousand genes, and perhaps the best 'minimal genome' candidate for synthetic biology is Prochlorococcus marinus, with more than 1800 genes (DeWall and Cheng, 2011). I had mentioned in a previous commentary for Journal of Cosmology (on synthetic biology) that Craig Venter wound up using a Mycoplasma genome that contained around a thousand genes, or twice the size of M. genitalium (Ussery 2010). The point here is that robustness is important in biology, and this is achieved in part by having redundant, overlapping parts. As James Shapiro says in his recent book on evolution, the emergence of novel functions often comes from the 'retention, duplication, and diversification of evolutionary inventions' (Shapiro, 2011). It seems very reasonable to me that there were many early genes which were not retained; we have no record of them today. And many genes in existence today are vastly divergent (different) from their ancestors - so different that we might not be able to detect two highly divergent genes coming from the same ancient gene duplication millions of years ago.

In my opinion, finding the 'minimal genome', especially in terms of a set of primordial set of genes early on in evolution, is a difficult task. Having looked at more than a thousand bacterial genomes, we find that there is not a single protein that is conserved amongst all. We published this last year (Lagesin, Ussery, Wassenaar 2010), and now with more genomes we even find that within some phyla it appears that very few (or no) genes are conserved. Finding zero proteins conserved sounds pretty heretical on the one hand, but on the other, if we look at the trend, based on past publications from various groups around the world, this is not surprising. The number has been steadily declining, from 256 universally conserved genes proposed by Mushegian and Koonin in 1996 (based on only two genomes), to a mere 31 conserved genes found by Ciccarelli and others ten years later (based on 191 genomes). Clearly as more genomes are added for the comparison, the number continuously gets smaller. Of course, there is undoubtedly a set of functions that are necessary for life, but the systems doing these functions do not themselves have to be conserved at the level of amino acid sequence. It is likely that the first 'genes' were in fact not even proteins at all, but rather small functional RNAs. We can still still in effect molecular fossils of some of these ancient complexes in structures such as the ribosome (which is a complicated mixture of proteins and rRNAs), telomerase, and other RNA/protein structures.

Thus, I think it might in some ways make sense that we don't see a large set of 382 conserved proteins across all organisms. First, I think that our assumption of being able to use full-length protein sequences to track history back very far just simply won't work, as too much divergence has happened over time. Second, from my perspective, based on sequenced bacterial genomes, we are essentially looking at a snapshot of life as it is today, after billions of years of evolution. It is kind of like examining the outer edge of a balloon for all modern life, and trying to extrapolate back to what the branches are underneath.

In summary, the underlying assumption - that life must have at least 382 genes - does not hold up to what we've already observed based on sequencing a few thousand bacterial genomes. This is a tiny sampling of the vast bacterial population in the world today. There is undoubtedly much we do not know. Having said that, in science it is often much easier to prove an idea is false, than to prove something is true. It seems clear to me that the conclusion argued in this paper that life must have been present before the formation of the earth 4.5 billion years ago is clearly built upon an assumption that is not correct.

References:

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, and Bork, P, 'Toward Automatic Reconstruction of a Highly Resolved Tree of Life', Science, 311:1283-1285, (2006).

DeWall MT and Cheng DW, 'The Minimal Genome - A metabolic and environmental comparison', Briefings in Functional Genomics, 10:312-315, (2011).

Lagesen K, Ussery DW, and Wassenaar TM, 'Genome update: the 1000th genome - a cautionary tale', Microbiology, 156:603-609, (2010).

Mushegian AR and Koonin EV, 'A minimal gene set for cellular life derived by comparison of complete bacterial genomes', Prod. Natl. Acad. Sciences, USA, 93:10268-10273, (1996).

Shapiro J: Evolution: A View from the 21st Century, Upper Saddle River, New Jersey: FT Press Science; 2011.

Ussery DW, 'One Small Step for Bacteria, or One Giant Leap for Mankind?', Journal of Cosmology, 8:3 (2010).
http://JournalofCosmology.com/ArtificialLife100.html