Response to Comment on "The 1.2-Megabase Genome Sequence of Mimivirus"

Although alternative viewpoints are now increasingly voiced (*1*–*4*), the tradition is to deny viruses the status of bona fide living organisms and to a priori doubt the capacity of phylogenetic analyses to investigate their deepest origin. From this viewpoint, viruses are merely seen as rapidly evolving "bags of genes," more or less indiscriminately acquired from their hosts, and thus cannot be attributed a genealogy of their own. The comment by Moreira and López-García (*5*) defends this traditional view by arguing that Mimivirus tyrosyl-tRNA synthetase (TyrRS) was acquired from its amoebal host *Acanthamoeba polyphaga* and generalizes this conclusion to the whole viral genome.

The problems arising from the propensity of aminoacyl-tRNA synthetase (aaRS) to exhibit horizontal gene transfer (HGT) (*5*) were addressed in our study (*1*). Removing aaRS sequences from the phylogenetic analysis had no influence on the position of Mimivirus in the tree of life. Mimivirus TyrRS does exhibit closer similarity to the *Entamoeba histolytica* homolog (54% identical residues) than to other eukaryotic homologs (*Oryza sativa*, 45% identity), making it a likely candidate for HGT. However, the tree built with TyrRS sequences [figure 1 in (*5*)] is inconsistent with the accepted species classification of amoeboid eukaryotes (compare, for example, the relative positions of plasmodium, dictyostelium, and cryptosporidium) (*6*). Thus, the direction of this putative gene transfer (virus to amoeba or the converse) remains uncertain. Mimivirus TyrRS has now been characterized and found functional. It is tyrosine specific (*7*) and optimally active on eukaryotic-type $tRNA^{Tyr}$. The *TyrRS* gene is not a piece of junk DNA.

To further examine whether, as a whole, the Mimivirus genome exhibits similarity with amoeba, we took advantage of the recently released *E. histolytica* HM-1:IMSS complete genomic sequence (*8*). The corresponding 9722 predicted protein sequences were downloaded through the National Center for Biotechnology Information Entrez system. We determined 87 reciprocal best matching open reading frame (ORF) pairs between Mimivirus and *E. histolytica* protein sequence data sets.

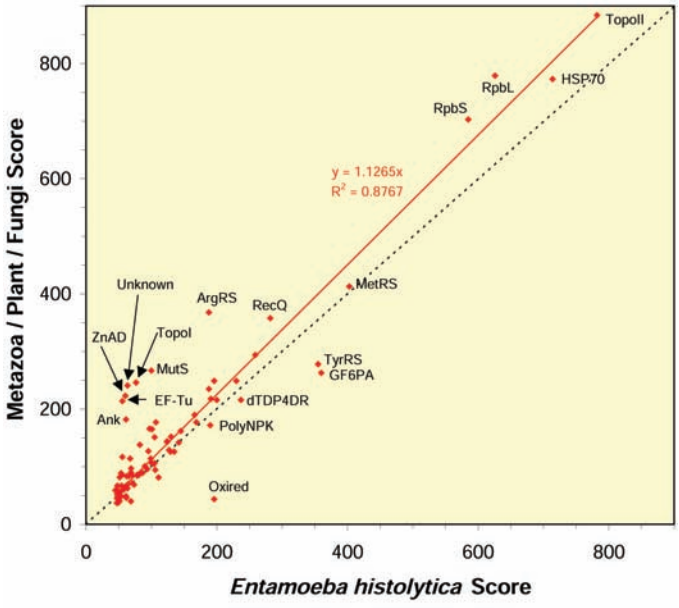

Fig. 1. Distribution of BLAST scores for 87 Mimivirus ORFs that exhibit reciprocal best matching ORFs in the protein sequence data of *E. histolytica* HM-1:IMSS. The horizontal and vertical axes represent best BLASTP scores against *E. histolytica* and a nonredundant sequence data set for metazoa, plants, and fungi, respectively. A regression line is highlighted in red. The black dotted line corresponds to equal scores along both axes. Ank, ankyrin-containing protein L863; ArgRS, arginyl-tRNA synthetase; dTDP4DR, dTDP-4-dehydrorhamnose reductase; EF-Tu, guanosine triphosphate–binding elongation factor eF-Tu; GF6PA, glucosamine-fructose-6-phosphate aminotransferase; HSP70, 70-kD heat-shock protein; MetRS, methionyl-tRNA synthetase; MutS, DNA mismatch repair adenosine triphosphatase MutS; Oxired, putative oxidoreductase R665; PolyNPK, polynucleotide phosphatase/kinase; RpbL, RNA polymerase II largest subunit; RpbS, RNA polymerase II second largest subunit; RecQ, helicase RecQ; TopoI, bacterial type topoisomerase I; TopoII, topoisomerase II; TyrRS, tyrosyl-tRNA synthetase; ZnAD, Zn-dependent alcohol dehydrogenase; Unknown, R521.

The 87 Mimivirus ORFs were then searched against a database containing all the currently available sequences from metazoa, plants, and fungi. The distribution of the similarity scores is shown in Fig. 1. This graph clearly demonstrates that only a handful of Mimivirus genes (four besides TyrRS) are more similar to their *E. histolytica* homologs than to homologs from other eukaryotic kingdoms. Thus, if one considers that *E. histolytica* is a good representative of all amoeba, our results clearly indicate that putatively host-acquired genes constitute a negligible part of the Mimivirus genome. Furthermore, comparing the Mimivirus-predicted proteome to 13,624 expressed sequence tags (*9*) generated from *Acanthamoeba castellani* (closer to *A. polyphaga*) has not yet provided any strong evidence for additional HGT (*10*).

To settle this matter further, we then estimated the total fraction of Mimivirus genes that could have been horizontally acquired from any source. We applied the method proposed by Nakamura *et al*. (*11*) to the 363 Mimivirus ORFs that exhibit recognizable homologs in other organisms and found that 8.3% of them (30 ORFs) likely originated from recent HGT. For comparison, Nakamura *et al*. (*11*) reported that an average of 15% of the ORFs in prokaryotic genomes are associated with recent HGT, ranging from 0.5% for *Buchnera* to 25% for *Methanosarcina acetivorans*.

More generally, a common characteristic of viral genomes, when compared with cellular organisms, is their much larger fraction of ORFs exhibiting no similarity in protein databases (60% for Mimivirus, 30% for *E. histolytica*). This is contrary to what is expected from genomes mostly built through HGT. Proponents of the "bag of genes" viewpoint would argue that all traces of similarity have been erased as a result of accelerated evolution in viruses. Yet, many viral genes (including those for DNA and RNA polymerases, various nucleases, and capsid proteins) have maintained easily recognizable sequence similarities. Overall, the hypothesis that viral genomes are mostly built from randomly stolen genes is not the most parsimonious.

Our alternative view is that ancestral DNA viruses may have originated prior to the Darwinian threshold (*12*), eventually participating in the mixing of bacterial and archaeal genes that led to the emergence of the eukaryotic cell. In that context, we do not expect viral genes to exhibit perfectly identical phylogenies. Nevertheless, they should all appear of ancient origin and be globally equidistant from the extant major eukaryotic kingdoms, as we observed (*1*). HGT probably happened but was not a major force in shaping the complexity of the

Mimivirus genome. The currently available data all suggest deep evolutionary origins for Mimivirus and other large DNA viruses.

Hiroyuki Ogata and Chantal Abergel
Structural and Genetic Information Laboratory
CNRS-UPR2589, IBSM, 13402 Marseille, France

Didier Raoult
Unité des Ricketssies, CNRS-UMR6020
University of Mediterranée
13385 Marseille, France

Jean-Michel Claverie
Structural and Genetic Information Laboratory
CNRS-UPR2589, IBSM, 13402 Marseille, France
E-mail: Jean-Michel.Claverie@igs.cnrs-mrs.fr

References and Notes

1. D. Raoult *et al.*, *Science* **306**, 1344 (2004).
2. D. P. Mindell, L. P. Villarreal, *Science* **302**, 1677 (2003).
3. P. Forterre, *Res. Microbiol.* **154**, 223 (2003).
4. L. M. Iyer, L. Aravind, E. V. Koonin, *J. Virol.* **75**, 11720 (2001).
5. D. Moreira, P. López-García, *Science* **308**, 1114 (2005); www.sciencemag.org/cgi/content/full/308/5725/1114a.
6. S. I. Nikolaev *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8066 (2004).
7. C. Abergel *et al.*, *Acta Cryst.* **F61**, 212 (2005).
8. B. Loftus *et al.*, *Nature* **433**, 865 (2005).
9. Protist EST Program; http://megasun.bch.umontreal.ca/pepdb/pep_main.html.
10. R. Watkins, personal communication.
11. Y. Nakamura *et al.*, *Nat. Genet.* **36**, 760 (2004).
12. C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8742 (2002).

8 March 2005; accepted 26 April 2005
10.1126/science.1111195