

Yet viruses cannot be included in the tree of life

Purificación López-García and David Moreira

We welcome the comments on our Review article (Ten reasons to exclude viruses from the tree of life. *Nature Rev. Microbiol.* 7, 306–311 (2009))¹ by Bayry and colleagues (Reasons to include viruses in the tree of life. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c1))²; Navas-Castillo (Six comments on the ten reasons for the demotion of viruses. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c2))³; Claverie and Ogata (Ten good reasons not to exclude viruses from the evolutionary picture. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c3))⁴; Ludmir and Enquist (Viral genomes are part of the phylogenetic tree of life. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c4))⁵; Koonin and colleagues (Compelling reasons why viruses are relevant for the origin of cells. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c5))⁶; and Raoult (There is no such thing as a tree of life (and of course viruses are out!) *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c6))⁷.

They offer us an opportunity to clarify our ideas further and to dissipate some confusion owing to the mixing of very different concepts and levels of interpretation regarding the actual possibility of including viruses in a tree of life (TOL). We realize that much of this confusion comes from the fact that many virologists and other biologists are not familiar with the theory and practice of molecular phylogeny. Their reaction is therefore more sentimental than rational, as if denying the possibility of including viruses in the TOL would imply dismissing their importance in evolution. First, we emphasize that we agree that viruses have been, and still are, important for the evolution of cellular organisms. They are vehicles of horizontal gene transfer (HGT) between cells, contribute to the acceleration of gene evolutionary rates and exert fundamental selection pressures on their host populations, thereby generating and regulating biodiversity (for example, through ‘kill-the-winner’ strategies⁸). However, their importance in biological evolution does not necessarily authorize their inclusion in the TOL. Let us make clear the essential points of our opinion article

in a more systematic way. We stated that viruses do not belong in the TOL based on two types of arguments: epistemological arguments, which refer to what ‘should not be done’, and methodological arguments, which refer to what ‘cannot be done’ and are more practically conclusive.

Our first argument relates to the definition of life and whether viruses are alive. It is epistemological in nature and deals with the way in which humans conceptualize their surrounding natural world. It is neither meta-physical, as Koonin *et al.*⁶ state, nor religious (even much less so) as unfoundedly claimed⁷. As biological scientists, we deal exclusively with empirical data from the material world, from which we extract information that allow hypotheses to be constructed and tested through the hypothetico–deductive methodology that is the common practice in our discipline. Any form of spirituality, including religion⁷ or intelligent design⁴, is out of the scope of our scientific activities. We disagree that defining life should be left to philosophers alone⁶, as it is biologists who actually study life. Defining life, as defining species, is a problematic issue owing to the difficulty of delimiting barriers to some kind of natural continuum. However, once a definition is established it should be applied logically. When some biologists say that viruses are alive, they are accepting some kind of definition. A logical syllogism would ensue: if viruses were alive they ‘should be’ placed in the TOL (epistemologically, an ideal construction in which all living entities would have a place) whereas if viruses were not alive they ‘should not’. The problem is that current definitions of life based on self-replication and evolution do not accommodate viruses because viruses cannot self-replicate. This is not because of the occurrence or absence of ribosomes, as is often thought^{5,9}, but because of the fact that viruses, unlike cells (including cellular parasites), are unable to transform energy and matter (that is, to actively generate order from disorder). In this sense, we agree with van Regensmortel¹⁰ that viruses, as genes or transposons, are biological elements but not living entities. By contrast, a self-replicating molecule, for example a self-replicating ribozyme, would be alive

[LINK TO ORIGINAL ARTICLE](#)
[LINK TO INITIAL CORRESPONDENCE C1](#)
[LINK TO INITIAL CORRESPONDENCE C2](#)
[LINK TO INITIAL CORRESPONDENCE C3](#)
[LINK TO INITIAL CORRESPONDENCE C4](#)
[LINK TO INITIAL CORRESPONDENCE C5](#)
[LINK TO INITIAL CORRESPONDENCE C6](#)

according to the above definition. However, viruses are not self-replicating ribozymes or DNAzymes and consequently cannot be qualified as living. Nevertheless, one can avoid defining life or use a more inclusive definition to accommodate viruses. Let us accept for the benefit of this discussion that anything that can be replicated (that is, does not necessarily self-replicate) and evolves is alive. Viruses and, by the same token, languages, popular legends, cooking recipes and human technology, would be alive. We contend that it would still be impossible to place viruses in the TOL because of purely methodological reasons (nine reasons out of the ten that we proposed¹).

Regarding the methodological arguments to exclude viruses from the TOL, we must differentiate between two types of hypotheses and we must also explain what a TOL is in this context: a molecular phylogenetic tree based on universally conserved (core) genes that are supposed to reflect organismal evolution. Whether this molecular phylogenetic tree truly reflects the most important lines of organismal evolution is a matter of discussion but, in practice, it is helpful and the basis of natural systematics (for an example, see the [Tree of Life Web Project](#)). Paradoxically, those that propose that a TOL does not exist because it is blurred by HGT use molecular phylogenies to show HGT¹¹ and recognize the existence of archaea, primarily defined as an independent domain of life by their segregation in a conserved gene-based TOL.

The first type of hypotheses, sustained by Koonin *et al.*¹² and others before them¹³, propose that viruses pre-date cells in biological evolution. Koonin *et al.*^{6,12} recognize that “viruses do not have universal genes” but claim that there is a core of ancient (precellular) virus hallmark genes that have persisted to date, dismissing convergence and HGT as explanations for their widespread distribution. We acknowledge that many viral genes are shared within viral families, but we still claim that dating them back to a common precellular origin is impossible because viruses are polyphyletic and because structural convergence and HGT are well-attested phenomena that can account for that distribution. Although simple structures are

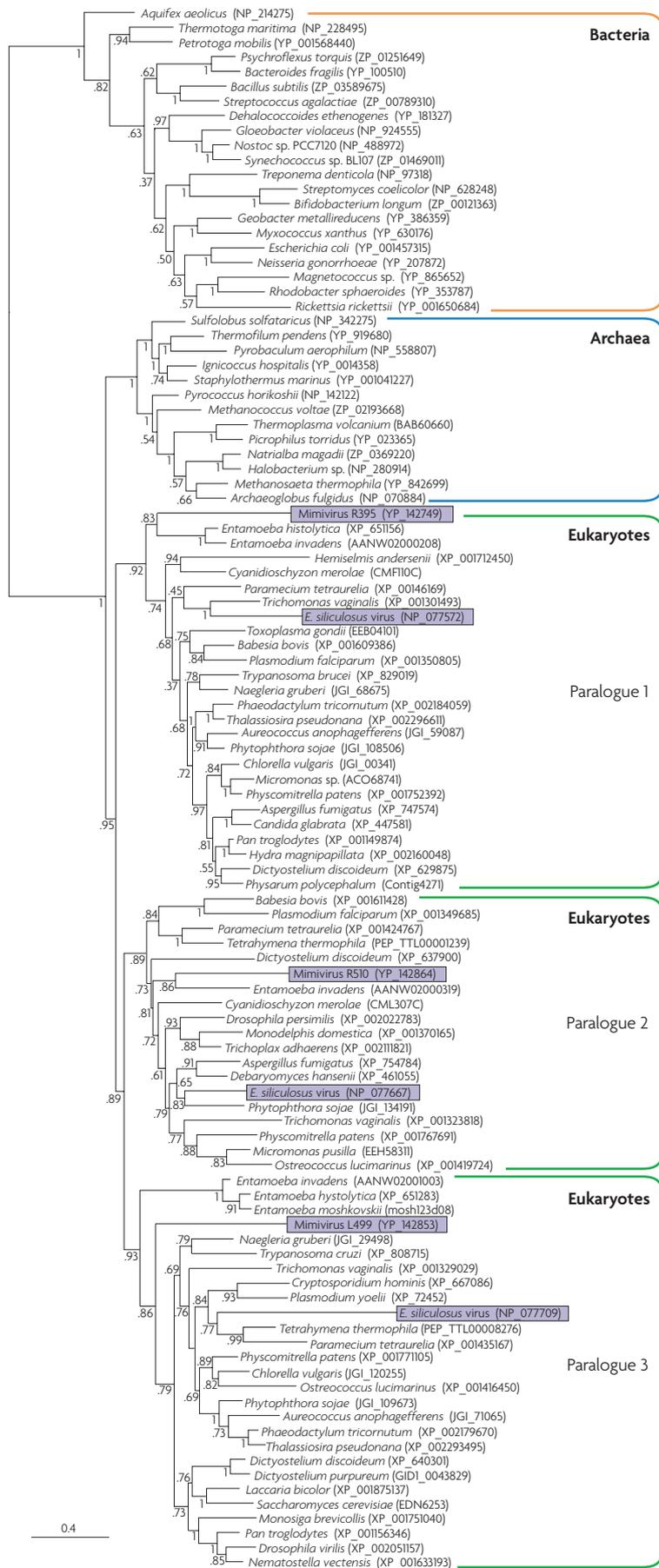


Figure 1 | Taxon-rich phylogenetic tree of clamp loader proteins. This 106-taxon tree was constructed on 174 conserved sites with the Bayesian approach implemented in PhyloBayes, using a mixture model (CAT) that was less sensitive to compositional bias and evolutionary rate heterogeneity between species¹⁸. Note that, in contrast to Claverie and Ogata's 20-taxon tree⁴, there is not a single eukaryotic group but three distinct paralogues and that, as a consequence of the richer taxonomic sampling, viral sequences emerge within (and not at the base of) the eukaryotic groups, suggesting that they were acquired by horizontal gene transfer from their eukaryotic hosts. Numbers at nodes are posterior probabilities. Sequence accession numbers are given in parentheses.

more prone to convergence, structures that are more complex can also be affected by it (for example, the presence of complex eyes in some unicellular dinoflagellates is an amazing case of convergence with metazoan eyes¹⁴). Koonin *et al.* maintain that a “virus-like stage of precellular evolution appears inevitable”. There is some ambiguity about what ‘virus-like’ implies (not all genetic elements are viruses; the distinction is important in this debate). If virus-like implies self-replicating elements, we might agree, but if it implies genetic elements that are dependent for replication on other systems (what viruses actually are), we disagree. It would seem more logical that life derives from those systems that were already able to replicate than from their molecular parasites (these might, however, foster the evolution of such self-replicating systems, just as viruses contribute to cell evolution today). Nonetheless, the problem is that this hypothesis cannot be tested. It cannot be proved or disproved by phylogenetic analysis with contemporary empirical data. Consequently, hypotheses proposing that viruses antedate cells, however unlikely they might be (it is difficult to imagine that parasites pre-date their hosts), cannot be falsified and remain valid. However, proponents of such models would agree with us that viruses cannot be placed in a phylogenetic TOL. This is methodologically impossible as they share no universally conserved genes with cells. If such genes ever existed, their sequences have evolved beyond homology recognition, losing all phylogenetic signal from putative pre-cellular times.

The second type of hypotheses on virus-cell relationships, including those of Raoult *et al.*¹⁵ and Claverie and Ogata⁴, claim that viruses (particularly large DNA viruses such as the Mimivirus) can indeed be placed in a TOL on the basis of phylogenetic analysis of their conserved cell-like genes (those that

are used to construct the cellular, however imperfect, TOL by molecular phylogeny). On the basis of these analyses, viruses would define a fourth domain of life^{4,15}. Ironically, Raoult, who was the first to propose that the “Mimivirus appears to define a new branch distinct from the three other domains” (REF. 15), partly motivating our article¹, now contradicts himself saying that there is no TOL⁷. Claverie and Ogata, on the contrary, maintain that position⁴ and present an additional tree of the clamp loader protein from Mimivirus (MIMI_R395) and from *Ectocarpus siliculosus* virus-1 (ESV-1) with their cellular homologues. The viruses appear at the base of eukaryotes, which is taken as “evidence of deep Mimivirus gene ancestry” (REF. 4). This kind of assertion can be tested. Using proper phylogenetic analyses, the vast majority of genes shared by Mimivirus and cells can be shown to have been acquired by recent (by contrast to pre-cellular or pre-domain diversification) HGT from their cellular hosts (and associated bacteria) and not the other way around¹⁶. The apparent basal positions of Mimivirus genes are easily explained both by long-branch attraction artefacts owing to the higher evolutionary rate of these genes in viruses and by poor taxon sampling. The case of the clamp loader is illustrative, as Claverie and Ogata’s tree includes a poor representation of taxa⁴. Homologues from the lineages to which Mimivirus and ESV-1 hosts belong (amoebae and stramenopiles) are excluded. A taxon-rich, detailed phylogenetic analysis shows a more complicated history for this gene, which has three paralogues in eukaryotes. Interestingly, Mimivirus also has three copies of the gene. However, instead

of being at the base of the three eukaryotic branches (which would be expected if the Mimivirus genes were ancestral), each Mimivirus gene appears nested within its respective paralogue group close to amoebae genes (FIG. 1). Similarly, ESV-1 appears nested within the eukaryotic groups (the slowest evolving ESV-1 gene — paralogue 2 — even branches with a stramenopile sequence), far from any of the Mimivirus copies. This demonstrates that the viral clamp loader genes were recently, and independently, acquired by HGT from eukaryotic hosts, and highlights the importance of adequate taxonomic sampling¹⁷. By showing their tree of the clamp loader, Claverie and Ogata provide a further example of what ‘cannot be done’ from a molecular phylogenetic point of view: concluding that large DNA viruses form an independent domain in the TOL from a poor phylogenetic analysis using genes that can be shown to have been acquired by HGT from cells. This and previous analyses argue for massive cell-to-virus HGT^{1,16}. Consequently, these hypotheses proposing that viruses define a fourth domain of life that can be placed in a TOL can be, and in our opinion have been, experimentally refuted.

In conclusion, even if viruses were considered to be alive and pre-date primitive cells, viruses could not be placed in a universal TOL by purely methodological reasons owing to the absence of shared genes and/or the loss of phylogenetic signal over billions of years of evolution. The claim that viruses can be placed in a TOL using cell-like genes is based on artefactual results and can be shown to be wrong.

Purificación López-García and David Moreira are at the Unité d’Ecologie, Systématique et Evolution, CNRS UMR 8,079, Université Paris-Sud 11, bâtiment 360, 91405, Orsay Cedex.
e-mail: puri.lopez@u-psud.fr, david.moreira@u-psud.fr

1. Moreira, D. & López-García, P. Ten reasons to exclude viruses from the tree of life. *Nature Rev. Microbiol.* **7**, 306–311 (2009).
2. Hegde, N. R., Maddur, M. S., Kaveri, S. V. & Bayry, J. Reasons to include viruses in the tree of life. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c1).
3. Navas-Castillo, J. Six comments on the ten reasons for the demotion of viruses. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c2).
4. Claverie, J. M. & Ogata, H. Ten good reasons not to exclude giruses from the evolutionary picture. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c3).
5. Ludmir, E. B. & Enquist, L. W. Viral genomes are part of the phylogenetic tree of life. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c4).
6. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. Compelling reasons why viruses are relevant for the origin of cells. *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c5).
7. Raoult, D. There is no such thing as a tree of life (and of course viruses are out!). *Nature Rev. Microbiol.* 29 June 2009 (doi: 10.1038/nrmicro2108-c6).
8. Hoffmann, K. H. *et al.* Power law rank-abundance models for marine phage communities. *FEMS Microbiol. Lett.* **273**, 224–228 (2007).
9. Raoult, D. & Forterre, P. Redefining viruses: lessons from Mimivirus. *Nature Rev. Microbiol.* **6**, 315–319 (2008).
10. Van Regenmortel, M. H. Viruses are real, virus species are man-made, taxonomic constructions. *Arch. Virol.* **148**, 2481–2488 (2003).
11. Doolittle, W. F. & Bapteste, E. Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl Acad. Sci. USA* **104**, 2043–2049 (2007).
12. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient virus world and evolution of cells. *Biol. Direct* **1**, 29 (2006).
13. Podolsky, S. The role of the virus in origin-of-life theorizing. *J. Hist. Biol.* **29**, 79–126 (1996).
14. Greuet, C. in *The Biology of Dinoflagellates* (ed. Taylor, F. J. R.) 119–142 (Blackwell, Oxford, 1987).
15. Raoult, D. *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350 (2004).
16. Moreira, D. & Brochier-Armanet, C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. *BMC Evol. Biol.* **8**, e12 (2008).
17. Lecointre, G., Philippe, H., Le, H. L. V. & Le Guyader, H. Species sampling has a major impact on phylogenetic inference. *Mol. Phylogenet. Evol.* **2**, 205–224 (1993).
18. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.