

# Comparison of a Modern and Fossil *Pithovirus* Reveals Its Genetic Conservation and Evolution

Anthony Levasseur, Julien Andreani, Jeremy Delerce, Jacques Bou Khalil, Catherine Robert, Bernard La Scola, and Didier Raoult\*

Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS 7278, IRD 198, INSERM U1095, IHU Méditerranée Infection, Pôle des Maladies Infectieuses, Assistance Publique-Hôpitaux de Marseille, Faculté de Médecine, Marseille 13005, France

\*Corresponding author: E-mail: didier.raoult@gmail.com.

Accepted: June 28, 2016

Data deposition: This Whole Genome Shotgun project has been deposited at the EMBL/EBI database under the accession PRJEB14764.

## Abstract

Most theories on viral evolution are speculative and lack fossil comparison. Here, we isolated a modern Pithovirus-like virus from sewage samples. This giant virus, named *Pithovirus massiliensis*, was compared with its prehistoric counterpart, *Pithovirus sibericum*, found in Siberian permafrost. Our analysis revealed near-complete gene repertoire conservation, including horizontal gene transfer and ORFans. Furthermore, all orthologous genes evolved under strong purifying selection with a non-synonymous and synonymous ratio in the same range as the ratio found in the prokaryotic world. The comparison between fossil and modern *Pithovirus* species provided an estimation of the cadence of the molecular clock, reaching up to  $3 \times 10^{-6}$  mutations/site/year. In addition, the strict conservation of HGTs and ORFans in *P. massiliensis* revealed the stable genetic mosaicism in giant viruses and excludes the concept of a bag of genes. The genetic stability for 30,000 years of *P. massiliensis* demonstrates that giant viruses evolve similarly to prokaryotes by classical mechanisms of evolution, including selection and fixation of genes, followed by selective constraints.

**Key words:** giant viruses, evolution, mosaicism, fossil.

## Introduction

The discovery of giant viruses generated a controversy about their nature. Indeed, some hypothesized that they are a bag of genes (Moreira and López-García 2009), robbing sequences from their host without using them, and harboring pseudogenes or ORFans. Others believe that their mosaic nature and size make these giant viruses a new branch of microbes, with complexity comparable to that of Prokaryotes (Sharma et al. 2016), this theory arising from the discovery of virophages (La Scola et al. 2008), transposon (Desnues et al. 2012), and MIMIVIRE, an antivirophage defense system (Levasseur et al. 2016). However, to clarify the nature of giant viruses, a comparison between contemporary and ancient strains would be decisive. Among the Megavirales representatives, *Pithovirus sibericum*, a “living fossil,” was isolated from a 30,000-year-old Siberian permafrost sample (Legendre et al. 2014, 2015). Here, we describe the isolation, genome sequencing, and evolutionary analyses of a contemporary species of *Pithovirus*, named *Pithovirus massiliensis*. This present-day species was

compared with its fossil counterpart, *P. sibericum*. For the first time, this species offers the opportunity to unambiguously study the evolution of a giant virus family.

## Materials and Methods

### Sample Collection

Nine sewage samples were taken at La Ciotat, France, GPS localization (*N43.181834, E5.614423*).

Samples were vortexed before direct inoculation into *Acanthamoeba castellanii* strain Neff (ATCC number 30010), used as host cells. We applied the coculture method used in the previous works by Reteno et al. (2015), with the exception that thiabendazole was replaced by voriconazole at a final concentration of 20 µg/ml (Sigma-Aldrich). The detection of the cytopathogenic effect was done by flow cytometry, as described by Bou Khalil et al. (2016). Preliminary staining by Hemacolor® Giemsa was performed in order to determine the nature of the newly detected agent responsible for the lysis.

Inoculation on Columbia and buffered charcoal yeast extract (BCYE) agar plates was done to exclude bacterial contamination.

### Electron Microscopy

The lysed coculture suspensions were fixed in glutaraldehyde before proceeding to electron microscopy. Ten microliters of the fixed suspension was deposited on the carbon grids for 10 min after glow discharge. Then, we contrasted the sample using a 1% solution of ammonium molybdate. Images were collected using a Tecnai G20 operating at 200 keV. ImageJ software was used for viral particle measurements.

### Production and Purification

An end-point dilution was performed for cloning (Pagnier et al. 2013). We then started production of the virus with fresh amoeba as previously described (Pagnier et al. 2013). Briefly, 15 infected flasks of 150 cm<sup>2</sup> (Corning® cell culture flask) were pelleted using the Beckman Coulter® centrifuge Avanti J-26XP at 14,000 × g for 30 min. A 25% sucrose gradient was used for the final purification step. After finalizing the production, we proceeded to genome sequencing.

### Genomics and Evolutionary Analyses

Genomic DNA of *P. massiliensis* was sequenced using MiSeq Technology (Illumina Inc., San Diego, CA) with the mate pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). The gDNA was quantified by a Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA). The mate pair library was prepared with 1 µg of genomic DNA using the Nextera mate pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc., Santa Clara, CA) with a DNA 7500 LabChip. The DNA fragments ranged in size from 1 to 10 kb. No size selection was performed and only 14 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimum at 696 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc.). The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing runs were performed in a single 42-h run in a 2 × 251-bp. Total information of 4.7 Gb was obtained from a 488 K/mm<sup>2</sup> cluster density with a cluster passing quality control filter of 97.2% (9,590,000 clusters). Illumina reads were trimmed using Trimmomatic (Bolger et al. 2014), then assembled through Velvet software (Zerbino and Birney 2008). Finally,

the draft genome of *P. massiliensis* consists of 15 contigs. Coding DNA sequences were predicted using Prodigal (Hyatt et al. 2010), and functional annotation was achieved using BLAST+ (Camacho et al. 2009) against the Non Redundant database.

Protein sequences corresponding to all annotated ORFs for the genome of *P. sibericum* were downloaded from NCBI. Proteinortho software was used to detect orthologous proteins using the following parameters: E-value 1e-5, 30% identity, 70% coverage (Lechner et al. 2011).

The two genomes of the *Pithovirus* strains were aligned with the Multiple Alignment of Conserved Genomic Sequence with Rearrangements software package (Darling et al. 2004).

Genome comparison was calculated with the OrthoANI program (Lee et al. 2016). The ratios of nonsynonymous (dN) and synonymous (dS) nucleotide substitutions was calculated for each pair of orthologs using the maximum likelihood model (M0, one ratio model) of the CODEML program in the PAML v.4.9 package (Yang 2007). The genome sequence was deposited in the EMBL/EBI database under accession number PRJEB14764.

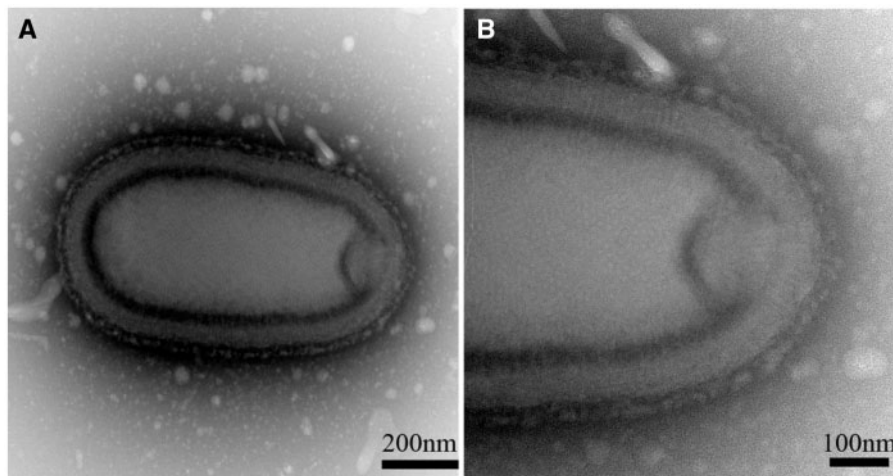
## Results

### Isolation of Modern *Pithovirus* Strains

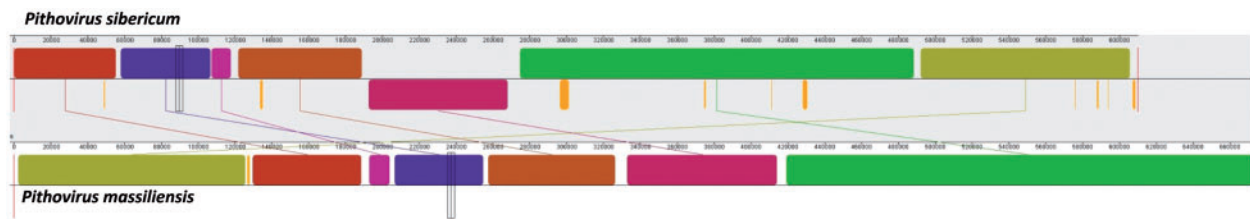
Since the discovery of *Acanthamoeba polyphaga mimivirus* (La Scola et al. 2003), giant viruses have been retrieved from highly diverse environments, and consequently in researching new protozoa-associated viruses, we have sampled different natural environments, including sewage. By the end of the final subculture for *A. castellanii*, we detected lysis of amoeba cells by flow cytometry for 3/9 studied sewage samples. For one sample, analysis under light microscopy revealed amorphous particles of a larger size, compared with the usual viruses found in routine coculture after amoebae lysis. The absence of growth on agar plates hypothesized the possible presence of a rod-shaped virus or a strict intracellular bacterium. Negative staining showed typical particles of the genera *Pithovirus* of 1.4 ± 0.2 µm (Legendre et al. 2014). The typical form of the cork structure clearly indicated that we had *Pithovirus*-like particles and this giant virus was named *P. massiliensis* (fig. 1).

### The Genomic Content of a Modern *Pithovirus* Reveals Its High Degree of Conservation

The genome of *P. massiliensis* was sequenced and assembled into one major scaffold (683 kbp) with a G+C content of 35%, similar to that of *P. sibericum*. We identified 520 genes representing a coding capacity of 439 kbp. In order to compare the genomic content between *P. sibericum* and *P. massiliensis*, we searched for all orthologous genes. A total of 437/467 genes found in *P. sibericum* were retrieved in *P.*



**Fig. 1.**—*Pithovirus massiliensis*. (A) Negative staining of the newly isolated *P. massiliensis* culture supernatant, showing the virus in its typical aspect of a thick electron dense envelope composed of stripes perpendicular to the surface. (B) Higher magnification of a *P. massiliensis* particle showing the particularity of its "cork", composed of a highly regular hexagonal honeycomb-like grid.



**Fig. 2.**—Comparison of the genome organization between *P. massiliensis* and *P. sibericum*.

*massiliensis*. The proportion of orthologous genes reached ~93.5% of the genome content of *P. sibericum*. Thirty and 58 genes were unique to *P. sibericum* and *P. massiliensis*, respectively (supplementary table S1, Supplementary Material online). Among these genes, most of them are short-length hypothetical proteins. Among the unusual features in the genomes of Megavirales, the number of ORFans is exceptionally elevated (Legendre et al. 2015). Interestingly, all these ORFans were conserved in *P. massiliensis*. Moreover, the resulting average nucleotide identity (ANI) values place *P. massiliensis* and *P. sibericum* together with an ANI of 83.6%. Concerning the whole genome organization, alignment of the *Pithovirus* genome sequences between *P. massiliensis* and *P. sibericum* showed a globally similar genomic arrangement (fig. 2). We obtained seven locally collinear blocks, the shortest block was 11.7 kbp and the longest was 262.7 kbp. These conserved synteny blocks between the old and modern strains highlight the stability of the genomes. Only one large-scale genomic rearrangement was evidenced, with the translocation of a segment of 125 kb. We also detected an inversion between two blocks. In summary, the genome rearrangements between *P. massiliensis* and *P. sibericum* are few and the

global genomic architecture is conserved between both species. Despite the time divergence between both species, their gene content is closely related at the genome scale, indicating a stable genomic species.

#### *P. massiliensis* Evolves under Strong Purifying Selection

Evolutionary biology is based on the principle of inheritance followed by selection. Evolution is driven by natural selection; positive selection as well as random drift (Levasseur et al. 2007). A fundamental measure of the relative importance of selection and genetic drift in causing amino acid substitutions is dN/dS ratio. The ratio of dN to dS changes between species is commonly computed to assay the strength and direction of selection. The variation of the dN/dS values is supposed to reflect the disparity of the purifying selection pressure that affects the evolution of microbes. In order to measure the divergence between *P. sibericum* and *P. massiliensis*, we measured the dN/dS ratio between all orthologous genes. All these orthologous genes were aligned and their dN/dS values were calculated. The mean dN/dS for *Pithovirus* was 0.11 (SD = 0.1) (supplementary table S2, Supplementary Material online). Globally, our results (dN/dS < 1) indicate that all genes are

constrained by strong selective pressure, and selection against dN substitutions has definitely operated. These results indicate that *Pithovirus* species evolved slowly, and the genomic content is conserved even after thousands of years, with selective pressure on the conserved genes. The rate at which a gene evolves is related to its function (Wang et al. 2011), and such a genetic conservatism in *Pithovirus* strains point to the central conclusion that all genes are functional and that they are biologically active.

In addition to the conservation of ORFans in their genomes, we also found that all ORFan genes in *Pithovirus* evolved under purifying selection. Evolutionary analyses of ORFan genes in closely related species are useful for the identification of both rapidly/slowly evolving and incorrectly annotated genes (Schmid and Aquadro 2001). Consequently, the massive number and the strong selective pressure occurring on the ORFans in *Pithovirus* clearly mean that they are accumulating nonrandomly and that they are biologically relevant. In order to estimate the dN/dS ratio found between *P. sibericum* and *P. massiliensis* in other microbial organisms, we compared the rate with 37 closely related prokaryotic genomes (Novichkov et al. 2009). The dN/dS ratio of *Pithovirus* was in the same range as the ratio found in the prokaryotic world (fig. 3). Consequently, *Pithovirus* evolved as other microbial organisms, by the classical mechanisms of evolution, with selection and fixation of genes followed by selective pressure constraints (Slimani et al. 2013).

### The Comparison between Fossil and Modern *Pithovirus* Strains Yields Insight into the Cadence of Evolution in Giant Viruses

The sequencing of the modern *P. massiliensis* offers the opportunity to estimate a molecular clock in giant viruses. In order to estimate the molecular clock, we determined the number of neutral substitutions occurring in all orthologous gene pairs between *P. massiliensis* and *P. sibericum*. As a time reference, the genetic divergence between both strains was correlated to the 30,000-year-old dated Siberian permafrost. Based on the comparison between *Pithovirus* strains, we estimated a maximum substitution rate of  $2.6 \times 10^{-5}$  substitutions/site/year. According to the dN/dS ratio (0.11), we estimated a maximum mutation rate of  $3 \times 10^{-6}$  mutations/site/year. The substitution rate of *P. massiliensis* was compared with other groups of viruses including DNA and RNA viruses. The substitution rate of *P. massiliensis* was lower than that of RNA viruses that fall in the range of  $10^{-2}$  to  $10^{-5}$  substitutions/site/year with most of RNA viruses exhibiting rates within one order of magnitude of  $10^{-3}$  substitutions/site/year (Duffy et al. 2008; Holmes 2009). dsDNA viruses of animals have usually far lower rates of substitution ( $10^{-5}$  to  $10^{-9}$  substitutions/site/year) than RNA viruses and *P. massiliensis* fall into this category. On the contrary, the

substitution rate of *P. massiliensis* is lower than ssDNA viruses that approach that observed in their RNA counterparts.

### *P. massiliensis* Illustrates the Stable Genetic Mosaicism in Giant Viruses

In Megavirales, the numerous genes acquired from eukaryotic and bacterial sources suggest that horizontal gene transfer (HGT) is an important process in the evolution of its genome (Moreira and López-García 2009; Filée and Chandler 2010). The amoeba ecosystem promotes HGT acquisition by promiscuity between viruses with other intracellular microorganisms, and the annotation of *P. sibericum* identified 103 HGT from nonviral origins (i.e., bacteria and Eukaryota) (Boyer et al. 2009; Moliner et al. 2010; Raoult and Boyer 2010; Legendre et al. 2014). Interestingly, all of these genes were conserved in *P. massiliensis* (supplementary table S3, Supplementary Material online). The GC percent of the bacterial HGT genes in *P. massiliensis* was compared with the GC-percent of the donor genes. As expected, the GC-percent of these genes was progressively adapted to the average total GC content of the viral recipient (supplementary table S4, Supplementary Material online). The conservation of genes originating from HGT indicated both the stability of these genes and once more, their essentiality. In light of these results, *Pithovirus* is able to select its genes acquired from HGT, followed by their long-term fixation and progressive adaptation to viral codon usage. These results clearly reject the concept of a “bag of genes” in which giant viruses are described as simple carriers of genes resulting from erratic capture from different hosts.

## Discussion

The emergence of giant viruses, with the discovery of the Mimivirus, raises fundamental questions about the origin and place of this group in the living world. With a much larger genome size than a number of bacteria and archaea, these giant viruses have been described as a bag of genes. For the first time, we have demonstrated that, on the contrary, the genomic content was conserved for 30,000 years. Furthermore, our work has shown that the acquired genes from HGT were highly conserved, as well as the ORFans. These results demonstrate that the mosaic genome of *Pithovirus* created from genes of Megavirales (core genome), from gene acquisition and from HGT, constitute a functional set conserved over a long time period. The high degree of conservation between our modern *Pithovirus* and its fossil counterpart raised the question of the potential contamination of the permafrost sample and the true nature of *P. sibericum*. In the initial work, replicates have been performed using two independent pieces of Siberian permafrost and similar results for *Pithovirus* isolation were obtained (Legendre et al. 2014). Moreover, we excluded the possibility of sample contamination according to the very low success rate of *Pithovirus* isolation according to our screening procedures. So far, we



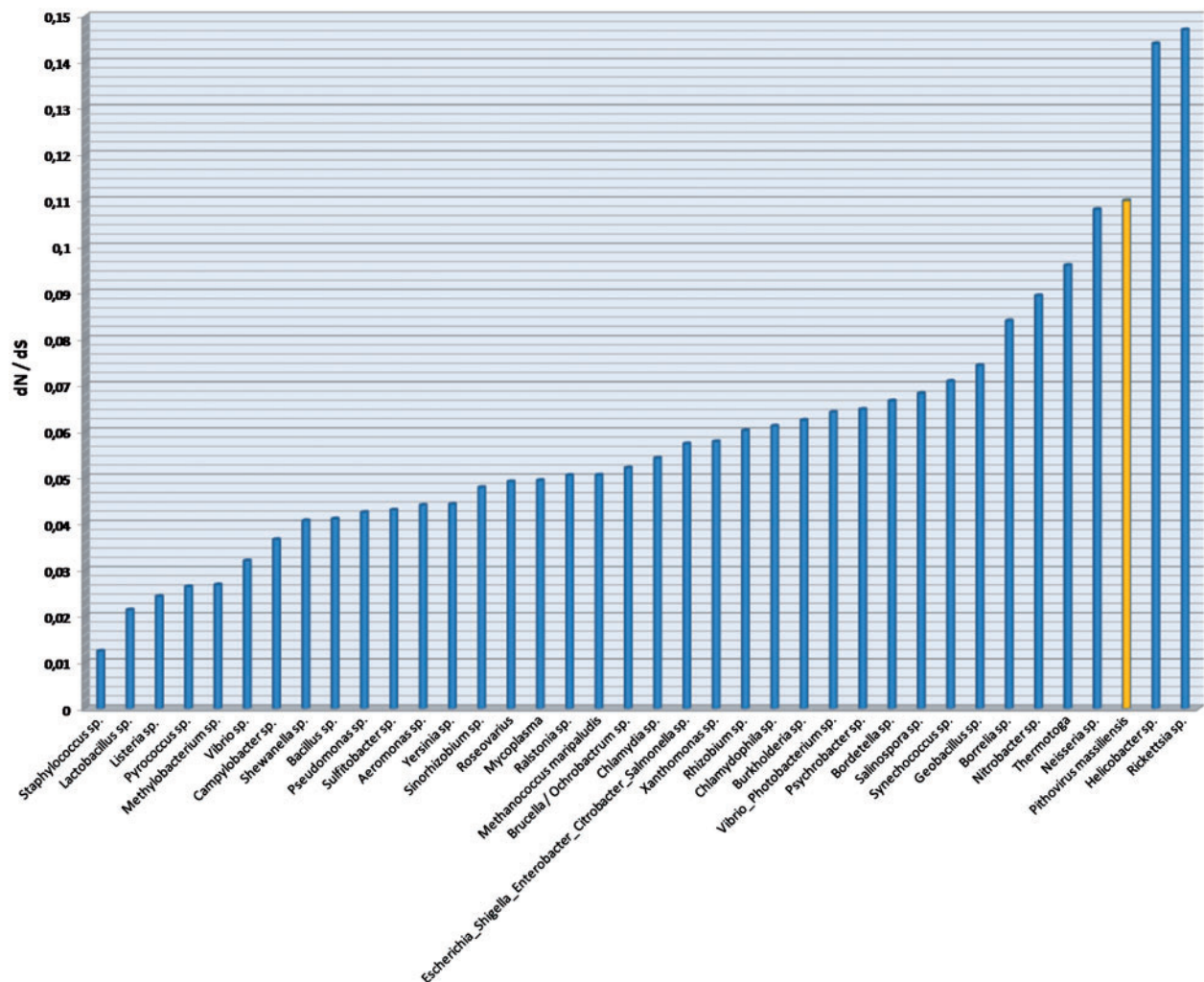


FIG. 3.—Estimation of the dN/dS ratio between closely related strains.

screened more than 210 samples for viral infection from various environments and using its amoebal host *A. castellanii*. At last, we isolated *Pithovirus* strains in only one case. This rare event makes improbable a large contamination by *Pithovirus* strains.

Our results showed that at least one family in giant viruses and most probably others Megavirales families evolve similarly to prokaryotes. The rate of mutation is known to be proportional to the genome size and the large 683 kbp-length genome of *P. massiliensis* is correlated to the low rate (Gago et al. 2009; Lynch 2010). Our conclusion could be further strengthened by the direct or indirect affiliation of *Pithovirus* strains. Indeed, we hypothesized that *P. massiliensis* is the direct descendant of *P. sibericum*. However, *Pithovirus* family were probably well-diversified 30,000 years ago and these two giant viruses could have evolved from a common ancestor that lived before this date. Thus, we must consider that both viruses are separated by at least 30,000 years, and

substitutions rates could be overestimated meaning that the degree of conservation could even be more conserved.

Our results are comparable to the dS-to-dN substitution ratios estimated in large DNA viruses in which the vast majority of family-specific genes do not exhibit an accelerated evolutionary rate demonstrating the functionality of peptides (Ogata and Claverie 2007; Doutre et al. 2014). Furthermore, some bacteria including obligate parasites with dramatic genome shrinkage could have higher dN/dS ratio than *P. massiliensis* (Novichkov et al. 2009; Lynch 2010). For instance, the intracellular *Rickettsia* sp. or *Helicobacter* sp. have a dN/dS mean of 0.147 and 0.144, respectively. These results indicated that the small bacterial genome of *Rickettsia* sp. (1.3 Mbp) and *Helicobacter* sp. (1.6 Mbp) could evolve more rapidly than giant viruses with larger genomes. The substitution rate of *P. massiliensis* revealed a slow evolutionary change in the range of some DNA viruses but far slower than RNA viruses. Most estimates of substitution rate support that viruses with

an RNA stage should evolve quicker than those with only DNA stages (Duffy et al. 2008; Holmes 2009). However, exceptions to the DNA and RNA viruses dichotomy were evidenced and blurred the boundary of substitutions rate in these groups (Salemi et al. 1999; Vandamme et al. 2000; Switzer et al. 2005). For dsDNA viruses of animals, rates of evolutionary change are frequently estimated under the assumption that these viruses have codiverged with their hosts, providing a molecular clock calibration. Based on this assumption, remarkable low evolutionary rate in the range of  $10^{-9}$  substitution/site/year were evidenced for gammaherpesviruses of vertebrates (McGeoch and Gatherer 2005; Duffy et al. 2008; Holmes 2009). Comparing to ssDNA, the substitution rate of *P. massiliensis* is lower than ssDNA viruses.

In light of these results, the position of giant viruses in the living world must be reconsidered, and requires a profound reassessment of their classification. Finally, the presence of contemporary giant viruses, strictly comparable to fossil giant viruses, suggests that, for this viral group, there is no potential threat in the context of the resurrection of prehistoric human-infecting giant viruses, as previously suggested (Legendre et al. 2015).

## Supplementary Material

Supplementary tables S1–S4 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgment

We thank Dr P. Pontarotti for comments on the manuscript.

## Literature Cited

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bou Khalil J, et al. 2016. High-throughput isolation of giant viruses in liquid medium using automated flow cytometry and fluorescence staining. *Front Microbiol.* 7:26.
- Boyer M, et al. 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc Natl Acad Sci U S A.* 106:21848–21853.
- Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.
- Desnues C, et al. 2012. Provirophages and transpovirons as the diverse mobilome of giant viruses. *Proc Natl Acad Sci U S A.* 109:18078–18083.
- Doutre G, Philippe N, Abergel C, Claverie JM. 2014. Genome analysis of the first Marseilleviridae representative from Australia indicates that most of its genes contribute to virus fitness. *J Virol.* 88:14340–14349.
- Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet.* 9:267–276.
- Filée J, Chandler M. 2010. Gene exchange and the origin of giant viruses. *Intervirology* 53:354–361.
- Gago S, Elena SF, Flores R, Sanjuán R. 2009. Extremely high mutation rate of a hammerhead viroid. *Science* 323:1308.
- Holmes EC. 2009. The evolutionary genetics of emerging viruses. *Annu Rev Ecol Evol Syst.* 40:353–372.
- Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
- La Scola B, et al. 2003. A giant virus in amoebae. *Science* 299:2033.
- La Scola B, et al. 2008. The virophage as a unique parasite of the giant mimivirus. *Nature* 455:100–104.
- Lechner M, et al. 2011. Proteinortho: detection of co-orthologs in large-scale analysis. *BMC Bioinformatics* 12:124.
- Lee I, Kim YO, Park SC, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol.* 66:1100–1103.
- Legendre M, et al. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci U S A.* 111:4274–4279.
- Legendre M, et al. 2015. In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting *Acanthamoeba*. *Proc Natl Acad Sci U S A.* 112:E5327–E5335.
- Levasseur A, et al. 2016. MIMIVIRE a defence system in Mimivirus confers resistance to virophage. *Nature* 531:249–252.
- Levasseur A, et al. 2007. Conceptual bases for quantifying the role of the environment on gene evolution: the participation of positive selection and neutral evolution. *Biol Rev Camb Philos Soc.* 82:551–572.
- Lynch M. 2010. Evolution of the mutation rate. *Trends Genet.* 26:345–352.
- McGeoch DJ, Gatherer D. 2005. Integrating reptilian herpesviruses into the family *Herpesviridae*. *J Virol.* 79:725–731.
- Moliner C, Fournier PE, Raoult D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev.* 34:281–294.
- Moreira D, López-García P. 2009. Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol.* 7:306–311.
- Novichkov PS, Ratnere I, Wolf YI, Koonin EV, Dubchak I. 2009. ATGC: a database of orthologous genes from closely related prokaryotic genomes and a research platform for microevolution of prokaryotes. *Nucleic Acids Res.* 37:D448–D454.
- Novichkov PS, Wolf YI, Dubchak I, Koonin EV. 2009. Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. *J Bacteriol.* 191:65–73.
- Ogata H, Claverie JM. 2007. Unique genes in giant viruses: regular substitution pattern and anomalously short size. *Genome Res.* 17:1353–1361.
- Pagnier I, et al. 2013. A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba. *Intervirology* 56:354–363.
- Raoult D, Boyer M. 2010. Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53:321–329.
- Reteno DG, et al. 2015. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J Virol.* 89:6585–6594.
- Salemi M, et al. 1999. Different population dynamics of human T-cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes. *Proc Natl Acad Sci U S A.* 96:13253–13258.
- Schmid KJ, Aquadro CF. 2001. The evolutionary analysis of “orphans” from the *Drosophila* genome identifies rapidly diverging and incorrectly annotated genes. *Genetics* 159:589–598.
- Sharma V, Colson P, Pontarotti P, Raoult D. 2016. Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses. *Curr Opin Microbiol.* 31:16–24.
- Slimani M, Pagnier I, Raoult D, La Scola B. 2013. Amoebae as battlefields for bacteria, giant viruses, and virophages. *J Virol.* 87:4783–4785.
- Switzer WM, et al. 2005. Ancient co-speciation of simian foamy viruses and primates. *Nature* 434:376–380.
- Vandamme AM, Bertazzoni U, Salemi M. 2000. Evolutionary strategies of human T-cell lymphotropic virus type II. *Gene* 261:171–180.

Wang D, Liu F, Wang L, Huang S, Yu J. 2011. Nonsynonymous substitution rate  $K_a$  is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biol Direct.* 6:13.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.

**Associate editor:** Bill Martin