



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Proceedings of the National Academy of Sciences

Cronfa URL for this paper:

<http://cronfa.swan.ac.uk/Record/cronfa50802>

Paper:

Lamb, D., Follmer, A., Goldstone, J., Nelson, D., Warrilow, A., Price, C., True, M., Kelly, S., Poulos, T. et. al. (2019).
On the occurrence of cytochrome P450 in viruses. *Proceedings of the National Academy of Sciences*, 116(25),
12343-12352.

<http://dx.doi.org/10.1073/pnas.1901080116>

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

On the occurrence of cytochrome P450 in viruses

David C. Lamb ^{1,2*}, Alec H. Follmer ^{3*}, Jared V. Goldstone ^{1*}, David R. Nelson ⁴,
Andrew G. Warrilow ², Claire L. Price ², Marie Y. True ³, Steven L. Kelly ²,
Thomas L. Poulos ³, and John J. Stegeman ¹

¹ Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

² Institute of Life Science, Swansea University Medical School, Swansea University, Swansea, Wales SA2 8PP, UK

³ Departments of Molecular Biology and Biochemistry, Pharmaceutical Sciences, and Chemistry, University of California Irvine, Irvine, California 92697-3900, USA.

⁴ Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, 858 Madison Avenue Suite G01, Memphis, Tennessee 38163, USA

Classification: BIOLOGICAL SCIENCES; Evolution

Keywords

Cytochrome P450; domains of life; virus; evolution; redox partner; crystal structure

Number of tables: 2

Number of figures: 6

Correspondence:

John J. Stegeman,
Biology Department, Woods Hole Oceanographic Institution,
Woods Hole, MA 02543, USA
Tel: 508-289-2320
E-mail: jstegeman@whoi.edu

* These authors contributed equally to this work.

Abstract

Genes encoding cytochrome P450 (CYP; P450) enzymes occur widely in the Archaea, Bacteria, and Eukarya, where they play important roles in metabolism of endogenous regulatory molecules and exogenous chemicals. We now report that genes for multiple and unique P450s occur commonly in giant viruses in the *Mimiviridae*, *Pandoraviridae*, and other families in the proposed order *Megavirales*. P450 genes were also identified in a herpesvirus (*Ranid herpesvirus 3*) and a phage (*Mycobacterium* phage Adler). The Adler phage P450 was classified as CYP102L1 and the crystal structure of the open form was solved at 2.5 Å. Genes encoding known redox partners for P450s (cytochrome P450 reductase or POR, ferredoxin and ferredoxin reductase, flavodoxin and flavodoxin reductase) were not found in any viral genome so far described, implying that host redox partners may drive viral P450 activities. Giant virus P450 proteins share no more than 25% identity with the P450 gene products we identified in *Acanthamoeba castellanii*, an amoeba host for many giant viruses. Thus, the origin of the unique P450 genes in giant viruses remains unknown. If giant virus P450 genes were acquired from a host, we suggest it could have been from an as-yet unknown and possibly ancient host. These studies open a new horizon in the evolution and diversity of the enormously important P450 superfamily. Determining the origin and function of P450s in giant viruses may help to discern the origin of the giant viruses themselves.

Significance

Cytochrome P450 monooxygenase enzymes metabolize drugs, carcinogens and endogenous molecules in the Eukarya, the Bacteria, and the Archaea. The notion that viral genomes contain P450 genes was not considered until discovery of the giant viruses. We have uncovered multiple and unique P450 genes in giant viruses from the deep ocean, terrestrial sources, and human patients. P450s were also found in a herpesvirus and a mycobacteriophage, and we report a crystal structure of the phage P450. Our findings herald a new era of research regarding the evolution of this important gene family, with implications for understanding the biology and the origin of the giant viruses themselves. The findings also present potential drug targets for giant viruses that may be human pathogens.

\body

1. Introduction

One of the key moments in biology was the discovery of a new heme protein more than 50 years ago, termed cytochrome P450 (CYP; P450), and further that P450 proteins could catalyze monooxygenase activity (1-3). Today, cytochrome P450 enzymes constitute one of the largest enzyme superfamilies known, occurring broadly in the Bacteria, Archaea, and Eukarya (4), the three classical domains of life proposed by Woese and Fox (5). P450 enzymes play essential roles in steroid/sterol biosynthesis and degradation, vitamin biosynthesis, detoxification and activation of many drugs and environmental pollutants, enzymatic activation of carcinogens, and the biosynthesis of vast arrays of secondary metabolites (6). The numbers of P450 genes encoded in genomes of eukaryotic species generally range from a few to several hundred. Anaerobic microbes often have no P450 genes but some bacteria, e.g., the *Actinomycetes*, have tens. More than 800,000 P450 sequences are known, and it is estimated that by 2020 more than one million will be known (7).

Our understanding of the distribution of P450s in nature began to change in 2004 with another key moment in modern biology, the publication of the genome of *Acanthamoeba polyphaga Mimivirus* (APMV). That paper defined a new class of virus, the nucleocytoplasmic large DNA virus (NCLDV) (8). The APMV genome is predicted to contain 1000+ genes, with fewer than 300 having predicted functions (8, 9). New giant viruses continue to be discovered in and captured with amoebae (10, 11), and in metagenomes (12, 13). Giant virus taxa include the *Mimiviridae* with the *Tupanviruses*, *Marseilleviridae*, *Pandoraviridae*, *Mollivirus*, *Pithovirus* and *Orpheovirus*, *Faustovirus*, *Kaumoebavirus*, and the *Klosneuvirinae* (14). These viruses fall in a proposed new order, the *Megavirales* (15, 16). A core set of genes involved in viral replication and lysis is most often conserved in these viruses (10), yet most genes in the giant viruses (70-90%) do not have any obvious homolog in existing virus databases.

Strikingly, many of the giant viruses possess genes coding for proteins typically involved in cellular processes including protein translation, DNA repair, and eukaryotic metabolic pathways previously thought not to occur in viruses (17). One such gene in Mimivirus (MIMI_L808) was annotated as encoding a lanosterol demethylase, which is an activity of CYP51, a well-known P450 essential in sterol biosynthesis (18). Subsequently, that Mimivirus sequence was cloned and expressed, and the recombinant protein was shown spectrophotometrically to be a P450 (19). The unexpected finding of a P450 in Mimivirus and the growing number of known giant viruses prompted us to search for P450 genes throughout the virosphere. We now report that genes for multiple and unique P450s are common in several taxa of giant viruses. We also uncovered P450 genes in viruses other than giant viruses. The discoveries open a new realm in the diversity and distribution of P450s in nature, especially in the enigmatic giant viruses.

Results

We identified P450 genes in multiple viruses, uncovered by examination of available genomes in the *Megavirales*, and all other available virus genomes. Figure 1 shows the phylogeny of viruses in the *Megavirales*. We assigned genes as encoding P450s based on the presence of consensus motifs found in nearly all P450s, including the EXXR motif in the K helix, the conserved cysteine heme-thiolate P450 motif, and overall domain identification using PFAM models. Our search revealed P450 genes in viruses in multiple branches of the *Megavirales* including in the *Mimiviridae*, in the *Pandoraviridae*, in *Mollivirus*, in *Kaumoebavirus*, in the *Tupanviruses* and in *Orpheovirus* (Figure 1; Tables 1 and 2). The molecular phylogeny of P450s in the giant viruses is shown in Figure 2, based on inferred amino acid sequences. Additionally, P450 genes were found encoded in phage and a herpesvirus (Table 2).

***Mimiviridae*: AT-rich P450 genes.** The *Mimiviridae* family currently is divided into three sublineages: Group A, the Mimiviruses and Mamaviruses; Group B, the Mouvouviruses; and Group C, the Megaviruses (20). (The term Mimivirus is used to refer to the original

Acanthamoeba polyphaga Mimivirus (APMV), and to the genus *Mimivirus*, and also to sublineage Group A in the *Mimiviridae*.) Searching the *Mimiviridae* genomes uncovered sequences similar to those first found in the *Mimivirus* (APMV) genome (19), i.e., MIMI_L808 now named CYP5253A1, and MIMI_L532, named CYP5254A1. The CYP5253 gene codes for a unique chimeric protein with a predicted N-terminal membrane anchor in a P450 domain of about 500 amino acids, which is fused at the C-terminus to a 200 amino acid domain (Figure 3). The P450 domain shares at most 28% identity with known P450s, while the C-terminal domain exhibits low identity with members of the 2-oxoglutarate (2OG) / Fe(II)-dependent dioxygenase superfamily (2OG-FeII). The CYP5254A1 gene encodes a smaller P450 domain without an encoded recognizable membrane anchor. All the *Mimiviridae* P450 genes are AT-rich (69-75%), like the *Mimiviridae* genomes overall (Table S1).

Chimeric CYP5253A genes were discovered in all *Mimiviridae* Group A and Group B viruses but not in Group C viruses, while CYP5254 family genes occur in all Group A, B, and C viruses. Known viruses in lineage C that are only distantly related to Mimivirus, including *Cafeteria roenbergensis*, *Phaeocystis globosa* and *Pyramimonas orientalis*, do not have P450 genes.

There is a surprisingly high degree of identity (98-100%) between inferred CYP5253A1 protein homologs in the Group A Mimiviruses isolated from different parts of the world. Likewise, there is >95% identity among the Group B (Moumouvirus) CYP5253A1 genes. Nucleotide sequences also are highly identical within subgroups. There is a lower degree of identity between the apparent orthologs in different subgroups of the *Mimiviridae*. Thus, the inferred sequences of full-length CYP5253A proteins (P450 domain plus the putative 2OG / FeII domain) share from 55% to 88% identity between the Groups A and B.

The orthologs of CYP5254s also are highly similar within subgroups, but show less identity between groups. The CYP5254 identity falls below 55% between Mimiviruses and the Moumouviruses, placing the Moumouvirus CYP5254 in a new P450 gene subfamily,

CYP5254B, while the genes in the other groups are in subfamily CYP5254A. While both Mamavirus and Lentillevirus appear to have a CYP5253A gene, there is some uncertainty for the gene prediction in these viruses.

Tupanviruses. Two new viruses related to the *Mimiviridae*, named the Tupanviruses, were discovered in a soda lake in Brazil, and in the deep ocean (3,000 meters depth) (12). These viruses differ in morphology from the Mimiviruses (12). Tupanviruses also differ from other *Mimiviridae* in their P450 gene complement. The deep sea Tupanvirus has three P450 genes, while the soda lake virus has two (Table 2). Neither virus has a gene encoding an ortholog of the chimeric CYP5253A1. However, both viruses have a CYP5254 gene, classified with the Moumouvirus CYP5254 subfamily, i.e., CYP5254B. Both also have a gene that comprises another new family and subfamily, CYP5868A. The deep sea Tupanvirus has a third P450 gene in a new family, CYP5869A1.

***Pandoraviridae*: GC-rich P450 Genes.** Giant viruses isolated in 2013 off the coast of Chile and from a freshwater pond in Australia constituted a new family, the *Pandoraviridae*, with no genomic or morphological resemblance to other viruses. The family was defined by *Pandoravirus dulcis*, *Pandoravirus salinus* (21), and *Pandoravirus inopinatum* (from an *Acanthamoeba* strain LaHel culture from the eye of a patient with keratitis in Germany (22)). More *Pandoraviruses* have been described recently (23)(24), including *Pandoravirus celtis* in 2019 (25). The Pandoraviruses group in two clades, A and B (23) and the genomes range up to 2.5 Mb, predicted to encode from 1,414 (*P. massiliensis*) to nearly 2,700 proteins (*P. braziliensis*) (24). All the genomes are GC rich (58-64%) (Table S1)

Most Pandoravirus genomes contain sequences predicted to encode membrane-bound P450s, in two new P450 families, CYP5634 and CYP5635. All members of both Pandoravirus clades have a CYP5634 sequence (Table 2), although *P. inopinatum* has two short sequences that based on gene order (below) appear to be CYP5634A gene fragments rather than a complete

sequence. Some members of both clades also have a CYP5635A sequence (Table 2).

Mollivirus P450. Two new giant viruses were obtained from the Late Pleistocene Siberian permafrost *Mollivirus sibericum* (26) and *Pithovirus sibericum* (27). The Mollivirus genome contains a P450 gene whereas the Pithovirus does not. Mollivirus has a GC rich genome (60%) and is related phylogenetically and morphologically to the Pandoraviruses. The inferred Mollivirus P450 protein has a putative membrane anchor, and shares closer identity with Pandoravirus CYP5635As (48-51%) than with the CYP5634As. The Mollivirus P450 gene thus constitutes a new CYP5635 subfamily, CYP5635B (Table 2).

Kaumoebavirus and Orpheovirus P450s. A P450 gene also appears in the genome of Kaumoebavirus, a giant virus (465 genes) phylogenetically related to Faustoviruses and the *Asfarviridae* (28). Kaumoebavirus was isolated from sewage in Saudi Arabia, and cultured from the amoeba *Vermamoeba vermiformis* rather than an *Acanthamoeba* species. The predicted Kaumoebavirus gene, CYP5870A1, encodes a 398 residue P450 lacking a membrane anchor.

Another virus isolated by co-culture with *Vermamoeba vermiformis* is *Orpheovirus IHUMI-LCC2* (11). Orpheovirus is related phylogenetically to Pithoviruses (11), but while Pithoviruses do not have P450 genes Orpheovirus has three. One Orpheovirus P450 is closely related to CYP5870A1 in Kaumoebavirus (Figure 2) and is thus classified as CYP5870A2. The other two Orpheovirus P450s are not closely related to any other virus P450 genes (Figure 2) and are classified to new P450 gene families, with genes CYP5871A1 and CYP5872A1 (Table 2).

In addition to the isolated viruses, assembly bins in environmental metagenomes have been identified as belonging to giant viruses, including Klosneuvirus, Indirivirus, Hokovirus and Catovirus (14). A novel AT-rich (70%) P450 gene was identified in *Hokovirus HK1*, isolated in Hong Kong. This P450 is unrelated to other virus P450s, and thus defines a separate new family, CYP5724A1 (Table 2). Other Orpheoviruses are also implied in metagenomes (13), but as yet these viral genomes are incompletely characterized.

Non-Giant Virus P450s. In examining the 8,100+ completely sequenced viral genomes and metagenomic data, we identified P450 genes in two viruses not in the *Megavirales*. A P450 gene designated CYP5723A1 (Genbank accession # ARR28948) was identified in the genome of a double-stranded DNA herpesvirus, *Ranid herpesvirus 3*, a pathogen that causes proliferative dermatitis in the frog *Rana temporaria* (29). CYP5723A1 is predicted to code for a soluble protein that shares 24-27% sequence identity with members of the cholesterol 7 α -monooxygenases (CYP7s), hinting at a possible function in lipid or steroid metabolism.

The genome of the *Mycobacterium* phage Adler, which infects *Mycobacteroides* (previously *Mycobacterium*) *abscessus subsp. bolletii* F1660, has a gene predicted to encode a P450 (Genbank protein accession number AHB79207; genome sequence locus tag: CH35_gp012). The predicted 471 amino acid protein defines a new subfamily and gene, CYP102L1, within the broader CYP102 family. A P450 subsequently found in the mycobacterial host is 99% identical to the phage P450, and therefore it is also classified as CYP102L1. Additional members of the new CYP102L subfamily (CYP102L1-8) have now been found in other *Mycobacterioides* species (SI Appendix, Figure S1). The phage CYP102L1 shows lower sequence identity with genes in other CYP102 subfamilies (CYP102A-Z) (SI Appendix, Table S2). It shares only 34% identity yet retains remarkable structural similarity with the P450 domain of CYP102A1 (P450BM3), an extensively characterized P450 (see below and SI Appendix, Figure S2).

Analysis of P450 Gene Synteny. We assessed the evolutionary relatedness among P450 genes in the *Megavirales* using analyses of shared gene order (synteny). In the *Mimiviridae* the CYP5253 and CYP5254 genes are in different locations in the sequenced genomes. The same ten genes flank CYP5254 in viruses in Groups A, B, and C. Likewise, the genes flanking the CYP5253A genes are conserved in the Mimiviruses (Group A) and the Moumouviruses (Group B) (Figure 4A). Group C *Megaviruses* are missing CYP5253A and also the adjacent gene R807, however the synteny around the CYP5253A and R807 loci in groups A and B is

preserved in Group C genomes. The absence of genes encoding both R807 and CYP5253A is an intriguing difference from the other *Mimiviridae*, suggesting that these genes were lost together from a Megavirus ancestor. The CYP5254B genes in both Tupanviruses share gene order like that around CYP5254B1 in Moumouvirus (SI Appendix, Figure S3), suggesting a closer evolutionary history with the Group B mimiviruses than with Group A or C viruses.

Shared synteny around the P450 genes is also found in the Pandoraviruses, with examples shown in Figure 4B. Where there are two P450 genes in a Pandoravirus genome they are closely linked, for example in *P. dulcis* CYP5634A2 and CYP5635A1 are separated by just 162 base pairs. The CYP5634A and CYP5635A loci in the various Pandoravirus genomes all are near to a fascin-like domain, and/or an ankyrin gene, shared synteny that indicates a common evolutionary history of these P450 genes. Location analysis of the two CYP5634-related sequences in *P. inopinatum* shows they are adjacent gene fragments separated by 160 bp and a stop codon. The two stretches together would code for a 356 amino acid protein, 84% identical to the CYP5634A in *P. dulcis*. Although Mollivirus is related to the Pandoraviruses, the gene order around the Mollivirus CYP5634B is not shared with the Pandoraviruses.

Conserved synteny around Kaumoebavirus CYP5870A1 and Orpheovirus CYP5870A2 show an evolutionary relatedness of these genes (SI Appendix, Figure S4). In both viruses the CYP5870A genes are adjacent to a sequence (BNJ_00167) predicted to code for a sterol desaturase. The encoded protein shares 35-41% identity with eukaryotic sterol C5 desaturase enzymes, although it is missing approximately 60 N-terminal and 90 C-terminal amino acids found in eukaryotic enzymes. Sterol C5 desaturase catalyzes the dehydrogenation of a sterol precursor C5(6) bond in cholesterol, phytosterol and ergosterol-producing organisms. The linkage of CYP5850A1 and BNJ_00167 suggests a possible functionality in sterol metabolism, not unlike the linkage of Mimivirus gene L808 (CYP5253A1) and R807 (see below).

Concerning Host Origin of Giant Virus P450s: P450 Genes Encoded in *Acanthamoeba*

castellenii. The majority of giant viruses have been isolated from or captured with an amoeba, most often either *Acanthamoeba polyphaga* or *A. castellanii*. To determine whether P450s in giant viruses are related to P450s in one of these contemporary hosts, we searched the completely sequenced *A. castellanii* genome (30) and identified 27 P450 genes. Two *A. castellanii* genes are predicted to encode CYP51 and CYP710D2, P450s involved in ergosterol synthesis, the major membrane sterol found in amoeba. The remaining 25 *A. castellanii* P450s were classified to new gene families or subfamilies, with no known function for these at present. *A. castellanii* P450 sequences shared no more than 25% identity with any virus P450.

Spectrophotometric Analysis of Recombinant P450 Proteins. We repeated expression and isolation of recombinant Mimivirus CYP5253A1 protein several times. As before (19), a native P450 spectrum (in the CO-bound, reduced state) was obtained with the expressed chimera (Figure 5A). Recombinant proteins of Hokovirus CYP5724A1 and *Ranid herpesvirus 3* CYP5723A1 also exhibited P450 spectral properties (Figure 5A). We also expressed Mimivirus CYP5253A1 genetically engineered without the putative 2OG / Fell fusion domain. The recombinant P450 domain alone did not yield a native P450 spectrum but instead exhibited a peak at 420 nm, which represents cytochrome P420, a conformation that would be inactive in an enzyme assay. This suggests that the 2OG / Fell domain contributes to maintaining the correct P450 protein conformation and/or state of the thiolate ligand to the heme (31).

P450 Substrate Binding and Catalytic Activity. In the original report of the Mimivirus genome, the CYP5253A1 was annotated as a lanosterol 14 α -demethylase, i.e., CYP51, although the P450 domain of that chimera has only 23% identity to any known CYP51. The preliminary analysis of the recombinant CYP5253A1 protein failed to detect any lanosterol 14 α -demethylase activity using rat POR (19). In a more complete and exhaustive analysis to assess possible CYP5253A1 sterol demethylase activity, we performed side by side experiments with the purified recombinant *A. castellanii* CYP51 (Figure 5A), using all three major CYP51

substrates favored by animals, fungi, and plants/amoebae: lanosterol, eburicol, and obtusifolol, respectively. The *A. castellanii* CYP51 was shown to demethylate all three sterol substrates with mean velocities of $0.69 \pm 0.03 \text{ min}^{-1}$ (lanosterol), $2.47 \pm 0.19 \text{ min}^{-1}$ (eburicol), and $3.32 \pm 0.02 \text{ min}^{-1}$ (obtusifolol). However, Mimivirus CYP5253A1 showed no activity toward these CYP51 substrates under any of the conditions used (SI Appendix, Figure S5A).

The gene adjacent to CYP5253A1 in Mimivirus (and in Moumouvirus) is gene R807, predicted to code for a 7-dehydrocholesterol reductase (7DHC). 7DHC catalyzes the final step of cholesterol synthesis, removal of the C(7-8) double bond converting 7-dehydrocholesterol to cholesterol (32). The linkage of Mimivirus R807 (7DHC) and L808 (CYP5253A1) suggests that these two virally encoded proteins might function with sterols, particularly cholesterol. Thus, we investigated whether CYP5253A1 could interact with cholesterol. Recombinant CYP5253A1 gave a type I binding spectrum with cholesterol (Figure 5B) with an apparent K_s value of $2.86 \pm 0.45 \mu\text{M}$. In contrast, and as before (19), we failed to find any evidence that CYP5253A1 binds the amoeba membrane sterol ergosterol.

The binding spectrum with cholesterol suggests it could be a substrate for CYP5253A1. However, with *in vitro* reconstituted systems using purified recombinant CYP5253A1 and purified mammalian POR to drive P450 activity, we did not detect metabolites of cholesterol or 7-dehydrocholesterol by GC-MS (SI Appendix, Figure S5B). Our earlier work has shown that P450s can be selective for their potential redox partner proteins (33). Accordingly, we undertook a bioinformatic search of all giant viral genomes that contain P450 genes but found no virally encoded candidate redox partner proteins that might drive P450 activity. Therefore, we used three fundamental heterologous redox partner systems utilized universally in research to drive P450 enzyme activity to probe viral P450 cholesterol metabolism. These were eukaryotic POR, from both fungal (*Aspergillus fumigatus*) and mammalian (*Homo sapiens*) sources; bacterial (*E. coli*) flavodoxin (fld) and flavodoxin reductase (flr); plant (*Spinacia*

oleracea) ferredoxin (fdx) and ferredoxin reductase (fdr). Cholesterol was not metabolized by recombinant viral CYP5253A1 with any of these commonly used P450 redox partners under any of the conditions employed.

We also assessed possible activity of expressed and purified Hokovirus CYP5724A1. This P450 shows up to 40% identity with P450s known to be involved in sterol and fatty acid metabolism, leading us to hypothesize that CYP5724A1 may also be either a sterol or fatty acid metabolizing P450. Hence we assessed whether the sterols lanosterol, eburicol, obtusifoliol, cholesterol and 7-dehydrocholesterol or the fatty acids lauric acid, palmitic acid, oleic acid could be substrates for this P450. No metabolites were observed for any of these potential substrates, under any assay conditions and redox partners (e.g., SI Appendix, Figure S5B).

Structure and Function of Mycobacteriophage CYP02L1. Heterologous expression of *Mycobacterium* phage Adler CYP102L1 in *E. coli* resulted in the production of recombinant heme-containing protein at 1 μmol / L *E. coli* culture. Purified CYP102L1 gave a reduced CO-Soret peak at 448 nm (Figure 6A). The protein was crystallized in an open state and the structure was determined by single-wavelength anomalous diffraction and was refined with a maximum resolution of 2.5 Å. The overall structure of the CYP102L1 exhibits the typical P450 fold consisting of α -helical and β sheet domains as seen in all other known archaeal, bacterial and eukaryotic P450 structures (Figure 6B). The heme cofactor is located between the α -helical domain and the β -sheet domain to create a substrate-binding pocket. The CYP102L1 has a conserved cysteine (Cys419) that serves as the fifth axial thiolate ligand to the heme iron, as is typical in most P450s. The structure solved is strikingly similar to that for the P450 domain of the *Bacillus megaterium* CYP102A1 (Figure 6C) (SI Appendix, Figure S2).

Type I spectral shift changes were observed following the addition of the fatty acids lauric acid (C12), myristic acid (C14), or palmitic acid (C16) to recombinant CYP102L1, suggesting these fatty acids are substrates (SI Appendix, Figure S6). Metabolism experiments using the

exogenous CYP102A1 reductase domain, eukaryotic POR and bacterial redox systems as electron donor proteins for CYP102L1 have yielded inconsistent results with palmitic acid as a potential substrate. It is possible that there is an as yet unknown redox system that could be necessary for phage CYP102L1 function. Conversely, phage CYP102L1 may be the first CYP102 described to-date that does not function in fatty acid metabolism.

***In vivo* Expression of Giant Virus P450 Genes.** An important question concerning giant viral P450 genes is whether they are transcribed and translated in the host. Data gleaned from mass spectrometry-based proteomic studies of Mimivirus virions identified MIMI_L532 (CYP5254A1) among 114 Mimivirus proteins detected in the virus particle (34), confirmation that a giant virus P450 was transcribed and translated in a host. Subsequent transcriptomic studies showed that more than 90% of Mimivirus genes are transcribed upon infection, including the Mimivirus gene L808 (the chimeric CYP5253A1), which is an “early transcribed” gene, peaking at about three hours post-infection (35).

Transcriptomic analysis showed that P450 genes are among those transcribed in a host (23) infected also by *P. salinus*, *P. quercus*, *P. neocaledonia*, or *P. macleodensis*. The transcribed genes included those for the CYP5634s and CYP5635s (23). While stringent strand-specific RNA-seq experiments and particle proteome analyses predicted that P450 genes are transcribed and translated into protein, it is not clear that they are incorporated into the virion (23). Likewise, the *M. sibiricum* P450 is expressed *in vivo* upon infection of amoebae (26, 27).

Retention of Mimivirus P450 Genes. A study in which Mimivirus was sub-cultured 150 times in axenic amoebae resulted in 155 of the 1018 coding sequences being lost from the genome (36). It was hypothesised that the genes lost were “useless genes” (i.e., non-essential genes), while the retained genes are essential to the viral life cycle (36). Examining the gene lists showed that both CYP5253A1 and CYP5254A1 are retained. Both the amino sequence and the nucleotide sequence showed 100% identity for CYP5253A1 before and after 150 rounds of

replication. However, in CYP5254A1 there was a single amino acid change after sub-culturing, at residue 64 (Arg - Lys), due to a single point mutation (AGA - AAA).

P450 enzymes, including the viral P450s described herein, require incorporation of heme for protein architecture and function. To-date, no genes that encode for heme-biosynthetic enzyme machinery have been found in any giant virus genome. Therefore, viruses encoding P450s must utilize heme synthesized by the host. The retention of Mimivirus P450s following 150 passages of subculture, and a dependence on host for heme, indicates these P450s must be important in Mimivirus biology, although the exact role is unclear.

Discussion

Through searching all available viral genomes we discovered a surprising number, diversity and distribution of genes encoding unique P450 proteins. Most were found in the giant viruses, but we also uncovered P450 genes in a phage and a herpesvirus. In other organisms P450 enzymes have broad metabolic roles in altering endogenous and exogenous chemicals in synthetic and degradation pathways not thought to be relevant in viruses. Our findings fundamentally challenge this viewpoint and raise critical questions concerning the origin and the function(s) of P450s in viruses, and of P450s in general.

Viral P450 Structure and Function

CYP102L1 in Mycobacteriophage Adler. The P450 gene found in this mycobacteriophage defined a new subfamily and gene, CYP102L1. To date, only one CYP102 protein has been examined structurally: CYP102A1 (P450 BM3). The crystal structure of phage CYP102L1 adds a second new CYP102 structure to this gene family. The CYP102L1 structure closely resembles that of CYP102A1 suggesting that other CYP102s might have retained similar structural architecture. The high degree of protein sequence identity suggests the phage CYP102L1 was directly derived from the host (SI Appendix, Figure S7). This apparent shuttling of a P450 gene on a small dsDNA virus resembles the occurrence of many P450 genes found

encoded in natural plasmids, for example, CYP101A1 (P450cam) on a plasmid affords the bacterium *Pseudomonas putida* the ability to utilize camphor as a sole carbon source.

CYP5253A1 in Mimivirus. Unlike the phage CYP102L1, the P450 genes in the giant viruses do not bear any close sequence identity to P450s in contemporary hosts, or to any other known P450s. The chimeric CYP5253A1, found in Mimiviruses, is the most unusual, and there is no precedent for a P450 structure fusing a putative 2OG / FeII -dependent dioxygenase domain and a P450 domain. We do not know how the P450 domain and the putative 2OG domain may be functionally linked. The fact that Mimivirus CYP5253A showed a native P450 spectrum only when it was expressed as a recombinant fusion protein, and not as the P450 domain alone, suggests there could be a functional connection between the P450 and the 2OG / FeII -dependent dioxygenase domains. In other P450 fusion proteins (NO synthetases, P450BM3) both moieties participate in a common functional outcome.

The type I binding spectrum demonstrates that cholesterol binds to Mimivirus recombinant CYP5253A1 as a substrate would bind, suggesting that cholesterol may be a substrate of this P450, although transformation of cholesterol was not detected. While an enzymatic function was not observed for the recombinant giant virus P450s examined, a biological role in giant viruses is clearly implied by experimental observations that in both *Mimiviridae* and *Pandoraviridae* the virus P450s are among genes transcribed and translated in the host upon infection. The P450s in Mimivirus are also retained through multiple rounds of viral gene descent, suggesting some essential role.

Origin of the Virus P450s

The sequence similarity of the P450 in mycobacteriophage Adler and a P450 in the *Mycobacteroides* host indicates that the phage almost certainly acquired this gene from the host. However, the situation is quite different in the giant viruses. Our uncovering and annotation of P450s in *A. castellanii* allow us to say that the P450s in giant viruses of amoebae

are remarkably *unlike* any of the P450s in a common *Acanthamoeba* host. This suggests that, if acquired, most giant virus P450s would be derived from some unknown host, possibly early in giant virus evolution. It is also possible that such an early host may have been an evolutionary dead-end, while the virus continued. Genetic mutations during virus evolution could have erased similarity to P450 genes in a presumed host. The high degree of identity between apparently orthologous P450 genes in different viruses from different parts of the globe is not consistent with rapid evolution, nor with an ancient acquisition. However, metagenomic analysis of ballast water from 4 ships (cargo carriers) in Korea revealed the presence of multiple viruses including Pandoraviruses and Megaviruses (37), suggesting this is a contemporary avenue contributing to a global distribution of highly similar giant viruses.

Deciphering the origin of P450s in the *Megavirales* is complicated by the presence of genes for multiple P450s differing within and between taxa, and the loss of P450 genes from some. The finding that the P450 genes in the Pandoraviruses are closely linked and share gene order suggests they arose by duplication of an ancestral P450 gene prior to the divergence of the two *Pandoraviridae* clades. CYP5634 and CYP5635 appear to define the P450 complement of Pandoraviruses. However, while CYP5634A occurs in all Pandoraviruses known to date, CYP5635A apparently was lost from some members of both clades.

As with the Pandoraviruses, the occurrence of apparently orthologous P450 genes in Mimiviruses from dramatically different environments around the globe – soils, other terrestrial sources, the deep ocean - indicates that these genes define the P450 complement in the *Mimiviridae*. Unlike P450s in the Pandoravirus, the CYP5253 and CYP5254 genes in the *Mimiviridae* are dramatically different from one another and reside in different locations in the genomes of species where both occur. The differences between CYP5253 and CYP5254 gene products in the *Mimiviridae* suggest that these genes could have separate origins. The Tupanvirus CYP5254B is more like the CYP5254B in Moumouvirus than the CYP5254A in

other viruses in the *Mimiviridae*. The lack of CYP5253 genes in the Megaviruses suggests that the different lineages in the *Mimiviridae* all may have diverged from a common ancestor, but that Megaviruses lost a CYP5253 precursor.

A recent mini-metagenomic approach detected 15 new assembled giant virus genomes in terrestrial soils, along with one additional genome obtained from the bulk metagenome (38). Five of these new genomes contain P450 genes. Satyrvirus, which is in a clade with Tupanvirus, has a CYP5254B like that found in Tupanvirus. Edafosvirus also has a CYP5254B similar to the Tupanvirus, and has one P450 that is plant-like and one that is fungus-like. Harvfovirus and Hyperionvirus both have a CYP5253 gene for a P450-2OG / Fell chimera that defines a new subfamily, CYP5253B. Each virus also has another P450 gene, related to plant P450s; in Hyperionvirus that P450 appears to be acquired from a moss. Terrestrivirus has a P450 gene weakly similar to a eukaryote P450 gene (CYP4). This suggests that in some environments there are giant viruses that have acquired some P450 genes from eukaryotes. Questions regarding the origin of giant virus P450 genes intersect with questions regarding the origin of the giant viruses themselves. Giant viruses all have genomes containing hundreds to thousands of genes, most often including translation-associated genes that are common in cellular life. They have been proposed as a separate fourth domain of life (39), and possibly even a basal group that contributed to the origin of eukaryotes (40). Either hypothesis is consistent with the idea that the giant viruses are very ancient, and may have been derived from early cellular ancestors, losing many genes in the course of evolution. If the giant viruses began as large entities with many genes, then the P450 genes could be remnants of that original condition. If we accept the idea that the *Megavirales* originated near the time when cellularity first occurred, then one possibility is that the P450s originated in some entity ancestral to the *Mimiviridae*, *Pandoraviridae*, and others.

Conversely, others suggest that the giant viruses are derived from viral antecedents that

developed large gene complements through acquisition (41-43), perhaps through waves of gene gain and loss (44). Phylogenetic analyses of genes for transfer RNA synthetases and core genes (14, 16) lend support to the idea that the *Megavirales* arose from a common ancestor that contained many fewer genes (16). If the hypothesis that giant viruses evolved by acquisition of genes is correct, then one could conclude that the P450 genes in the *Mimiviridae*, *Pandoraviridae*, and *Mollivirus* lineages were acquired from some host(s) in their evolution.

It is impossible to know at present whether a single ancestral P450 gene may have been acquired from an early host, and then duplicated and diverged during virus evolution, or whether there was acquisition and subsequent divergence of multiple different P450 genes. Given the differences between the *Mimiviridae* and *Pandoraviridae* P450s, and the phylogenetic separation between the *Megavirales* taxa bearing P450 genes, it is perhaps more plausible that P450 genes were acquired multiple times. This would be similar to a proposed independent acquisition of some tRNA synthetases in the giant viruses (14).

As yet, there is no consensus regarding the origin and evolution of the giant viruses (45, 46), which continue to be discovered (47, 48). It is noteworthy that the giant viruses with genes for multiple translation-associated proteins tend to be the ones that also contain cytochrome P450 genes. Further understanding of the viral P450s, their function and evolution may help to distinguish between giant virus evolution by genome growth, or genome reduction.

Materials and Methods

Viral P450 Sequence Datasets. The P450 domain is readily identifiable bioinformatically. Candidate P450 sequences were retrieved from the NCBI sequence databases using BLAST. Inferred protein sequences were aligned to the PFAM P450 hidden Markov model (PF00067) using ClustalOmega (49), and scored with T-Coffee (50). Alignments were visualized in BioEdit (51). Virus P450 genes were named based on the established guidelines of the International Cytochrome P450 Nomenclature Committee (52). Virus names are those in the original reports

or Genbank. Pairwise amino acid identities reported in the text are based on the length of the shortest protein in the alignment. Identity values used in the tables are based on multiple sequence alignments and are usually not more than 1% different from values stated in the text.

Phylogenetic Analyses of Viral P450s. Phylogenetic trees of P450 domains were estimated by maximum likelihood (ML) using RAxML (v8.2.6) with the PROTCATLG model of amino acid substitution (53). Replicate ML trees were constructed, and the best-scoring tree was subjected to rapid bootstrapping in RAxML to assess bifurcation support. Virus phylogenies were constructed from the conserved RNA Pol2 (RNAP2) gene using the PROTCATHIVB model of amino acid substitution.

P450 Expression and Sterol Binding. Sequences predicted to encode viral CYPs were synthesized by Eurofins MWG Operon (Ebersberg, Germany), or by DNA2.0 (California, USA) and cloned into the expression vector pCWori⁺ and expressed in *Escherichia coli* (8). These included the full length Mimivirus CYP5253A1 comprising both the heme and C-terminal chimeric domains, the Mimivirus CYP5253A1 heme domain alone, Hokovirus CYP5724A1, and Ranid herpesvirus 3 CYP5723A1. (See SI Appendix and Figure S8). Specific P450 heme contents were approximately 17 nmol / mg protein for those P450s expressed and purified, which equates to almost 100% heme incorporation.

Mycobacteriophage Adler CYP102L1 was cloned into expression vector pET28a and expressed in *E. coli*. The *A. castellanii* CYP51 was expressed and purified by methods used previously (54). A six-histidine tag was engineered at the C-terminus of each P450 and proteins purified according to the method of Arase *et al.* (55) using Ni²⁺-NTA agarose (54). The purification of recombinant CYP102L1 is described in detail in the SI Appendix. Protein purity was assessed by SDS polyacrylamide gel electrophoresis (SI Appendix, Figure S9). Reduced, carbon monoxide (CO) difference spectra were measured by established methods (56). For P450 ligand-binding experiments, the spectral difference was recorded after the incremental

addition of substrate and the binding constant (K_s) of the P450-substrate complex determined (57). Reconstitution assays for P450 catalytic activity are described in detail in the SI Appendix. Chemicals were the highest grade from Sigma Chemical Company unless otherwise stated. Eburicol and obtusifoliol were a gift from Dr. David Nes, Texas Tech University.

Isolation, Crystallization, and Structure Determination of Adler Phage CYP102L1. The Adler Phage CYP102L1 gene was synthesized by Genscript (USA) and used to express protein for crystallization. Protein was expressed and purified by experimental procedures described in the SI Appendix. Crystals of CYP102L1 were grown in hanging drops. Diffraction quality crystals were obtained using an additive screen that yielded the final condition of 0.1 M sodium cacodylate pH 6.5, 1.26 M ammonium sulfate and 4.2 % (w/v) dextran sulfate. Data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) Beamline 9-2. The structure was determined by single-wavelength anomalous diffraction and a hybrid substructure search (HySS) was performed using Fe-searched scattering type with S as an additional scattering type. The details and the final refinement statistics are given in the SI Appendix (SI Appendix, Table S3). The coordinates and associated structure factors have been deposited with the Protein Data Bank, PDB code 6N6Q.

Acknowledgements.

The research was supported by a USA-UK Fulbright Scholarship and a Royal Society grant to DCL, the Boston University Superfund Research Program NIH 5P42ES007381 (JJS and JVG), NIH 5U41HG003345 (JVG), the European Regional Development Fund and Welsh Government Project BEACON (SLK), by the Woods Hole Center for Oceans and Human Health, NIH P01ES021923 and NSF OCE-1314642 (JJS), and NIH R01GM53753 (TLP). We thank Dr. David Nes, Texas Tech University, for providing sterols, and Dr. Matthieu Legendre and Dr. Chantal Abergel, CNRS, Marseille, for access to the *Pandoravirus celtis* sequences. Drs. Irina Arkhipova, Mark Hahn, Judith Luborsky and Ann Bucklin commented on the Ms.

Figure 1. *Megavirales* Phylogeny Constructed From the RNAPol2 Gene: This maximum likelihood phylogenetic tree shows the relationship of key members of the *Megavirales* viruses. The clades that contain encoded cytochrome P450 genes are highlighted in colors that correspond to the cytochrome P450 phylogeny in Figure 2. Bootstrap support values are derived from 250 replicates, using the HIVB model of amino acid substitution. Only values below 90% are shown.

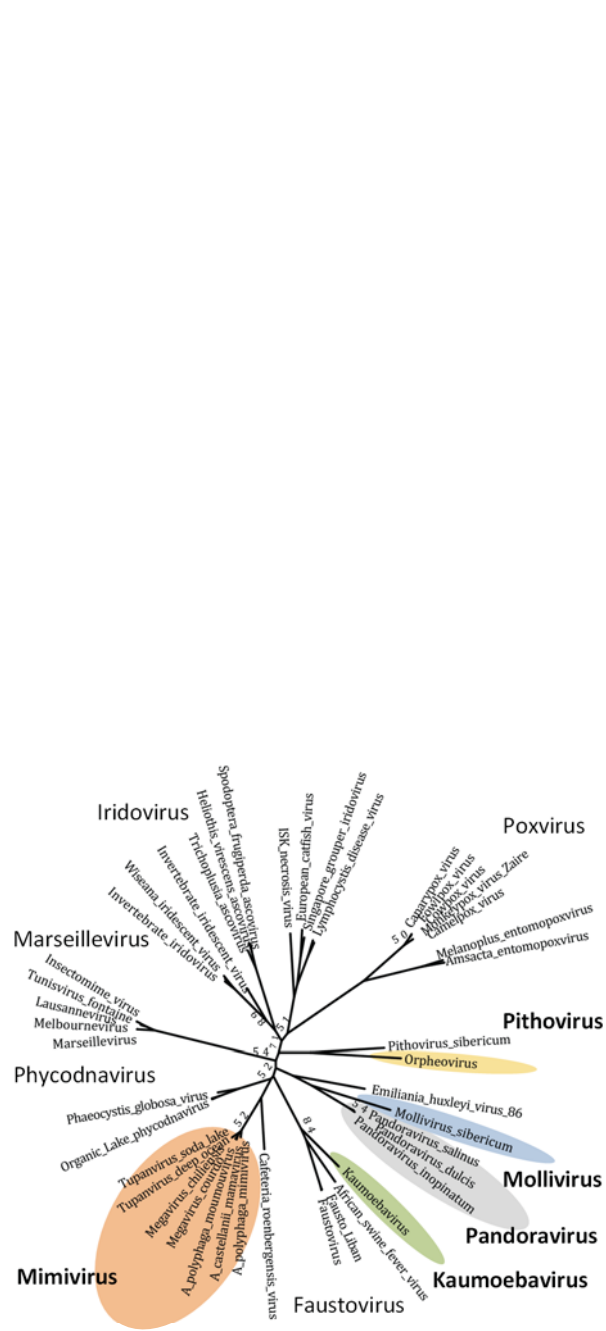


Figure 2. Molecular Phylogeny of Virus Cytochrome P450 Genes: **(A)** The unrooted maximum likelihood phylogeny of predicted amino acid sequences for viral P450s shows the close relationships between *Pandoraviridae* CYP5634 and CYP5635, and among *Mimiviridae* CYP5253s and CYP5254s. Note the CYP5254s in Tupanviruses, and the CYP5870s from the Kaumoebavirus and Orpheovirus. **(B)** Detailed phylogeny of selected *Pandoraviridae* CYP5634s and CYP5635s and Mollivirus CYP5635. Color blocks are the same as in Figure 1.

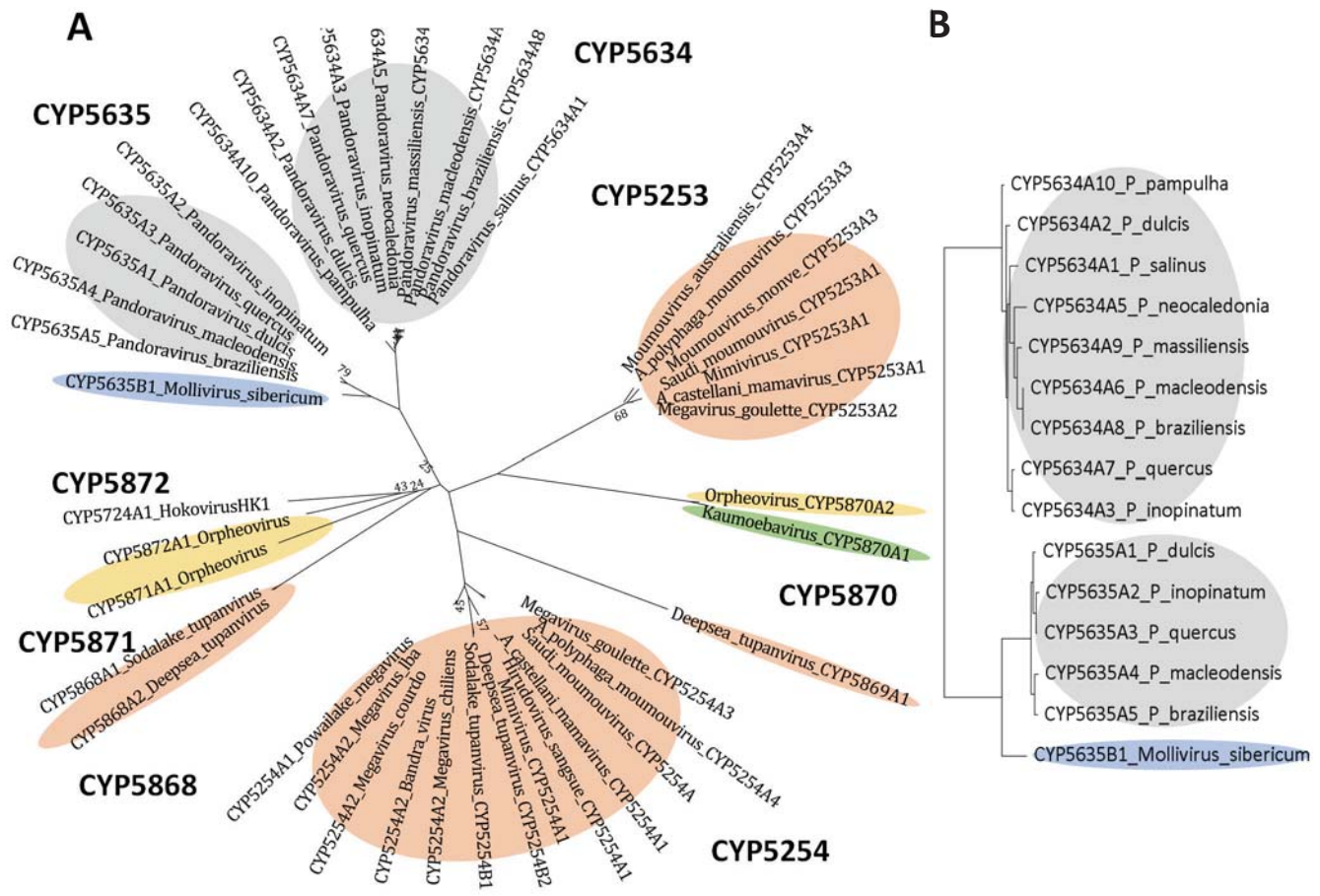


Figure 3. Domain Regions in Mimivirus CYP5353A1 (A) and CYP5254A1 (B):

The following abbreviations are used: TMS, transmembrane segment; SRS, substrate recognition site; T, conserved threonine in the P450 I helix; ExxR, conserved glutamate-residue-residue arginine motif found in the K helix; C, absolute conserved cysteine residue present in the heme binding domain; PTM, posttranslational modification sites including: 1 *N*-glycosylation site; 1 protein kinase C phosphorylation site; 4 Caesin kinase II phosphorylation sites; 3 myristoylation sites.

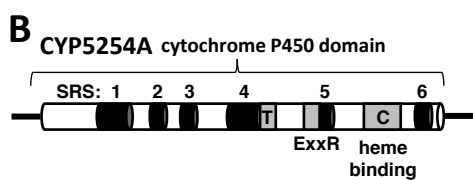
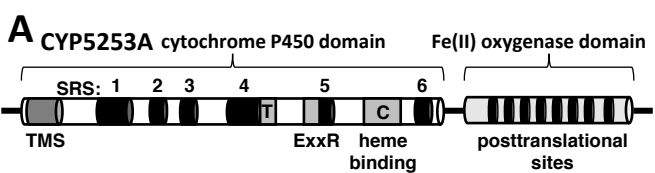
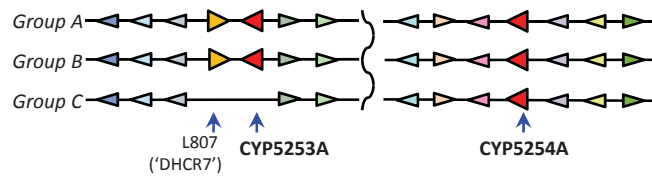


Figure 4. Microsynteny Around Giant Virus Cytochrome P450 Genes: A. *Mimiviridae*: groups A, B, and C display high degrees of shared gene order extending 10 genes around both CYP5253A (left) and CYP5254A (right). Genes that are highly similar to each other are colored the same across the various genomes. Completely shared synteny is evident across all 3 clades for CYP5254A. Clade C viruses are missing both CYP5253A (L808 ortholog) and gene R807 (a putative 7-dehydrocholesterol reductase, DHCR7; see text). **B. *Pandoraviridae*:** 23

Note shared synteny around P450s in Pandoraviruses in clades A and B. Figure also shows that CYP5634 and CYP5635 are adjacent, and illustrates that CYP5634A1 may be a pseudogene in *P. inopinatum*. Genes in white are not similar, while purple-colored genes are orthologous. Other identifiable genes are indicated (fascin-like domain and ankyrin).

A. Mimiviridae



B. Pandoraviridae

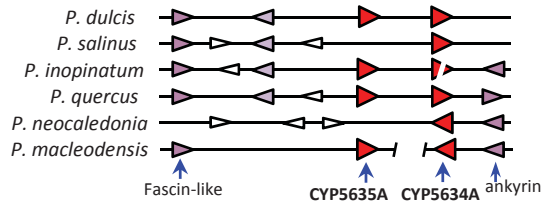


Figure 5. Characterization of Recombinant Viral Cytochromes P450: (A).

Dithionite-reduced, carbon monoxide difference spectra of the purified viral cytochromes P450 and the control *A. castellanii* CYP51 (AcCYP51) were obtained. The concentrations of P450 used were: CYP5253A1 (2.5 μ M); CYP5724A1 (2.5 μ M); CYP5723A1 (1.25 μ M); AcCYP51 (1.25 μ M) respectively. **(B).** P450-sterol binding assays: Stock solutions of 2.5 mM cholesterol (circles) and ergosterol (triangles) were titrated against 4 μ M CYP5253A1 in 50 mM Tris-HCl, pH 8 and 20% v/v glycerol with the difference spectrum determined after each addition of sterol which were used to construct ligand saturation curves.

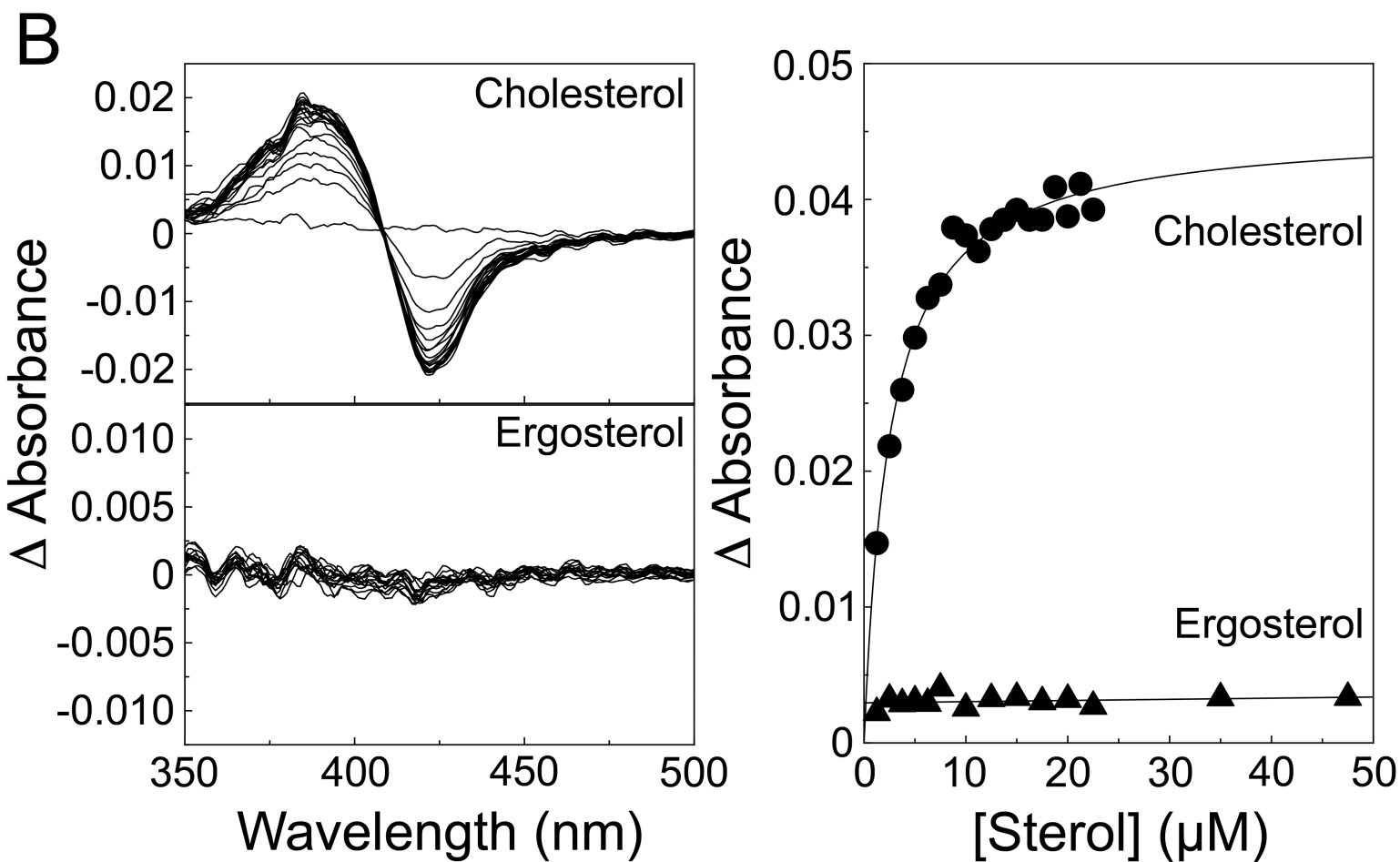
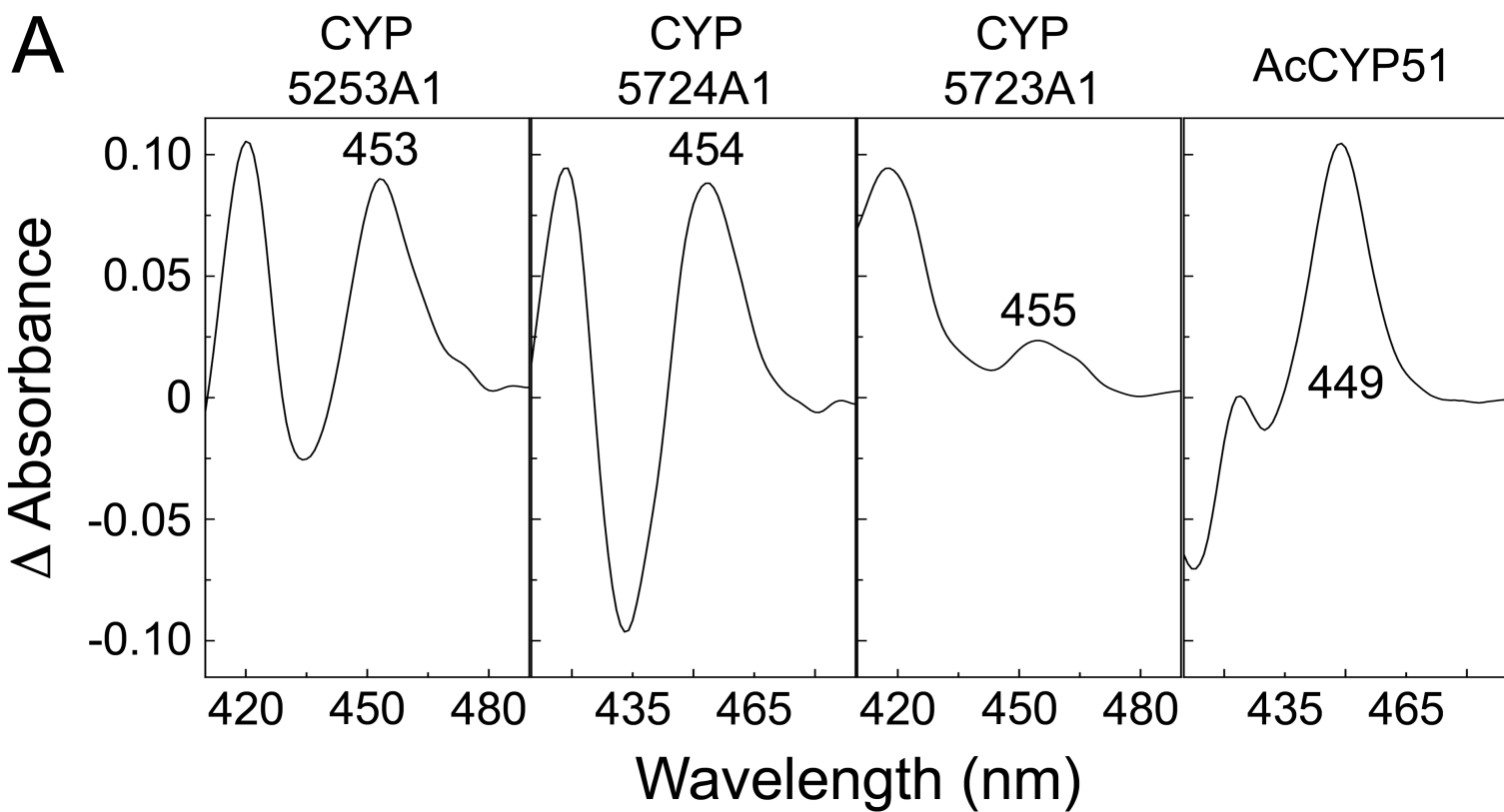


Figure 6. Structural Features of CYP102L1: **(A)** Dithionite-reduced carbon monoxide difference spectra of the purified Adler phage CYP102L1 (P450 concentration 1.25 μ M) showing a Soret spectral maximum at 448 nm. **(B)** The open crystal structure of CYP102L1 (PDB: 6N6Q) exhibits the typical P450-fold consisting of α -helical and β -sheet domains as seen in all other known P450 structures. Heme is shown as a red stick model. **(C, D)** Comparison of the crystal structures of Adler phage CYP102L1 (green) and *B. megaterium* CYP102A1 (P450 BM3) (blue): Although sequence identity remains low (34 %), a number of residues that have been implicated in CYP102A1 function are conserved between both proteins, including A82 (92), F87 (97), F261 (275), A264 (278), T268 (282), A328 (245), A330 (347), F393 (412), and I401 (420), with CYP102L1 residues in parenthesis. The CYP102A1 residues are involved in substrate binding affinity and turnover (A82 and I401), substrate selectivity (F87, T268, A328, and A330), and electron transfer (F261, A264, T268, and F393). Additional residues and references documenting residue involvement are in the SI Appendix.

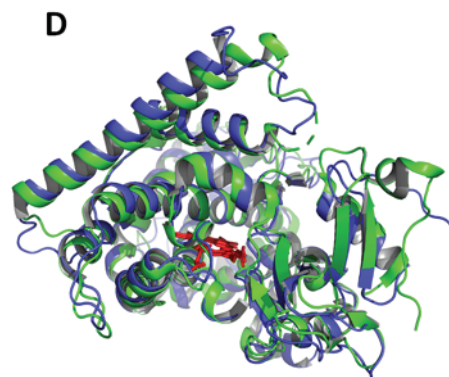
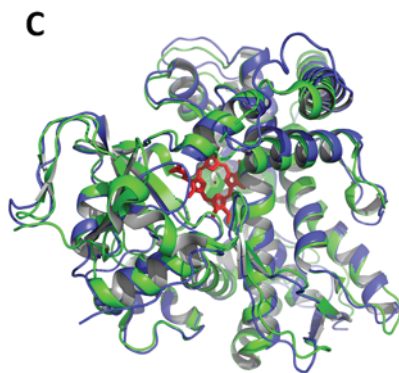
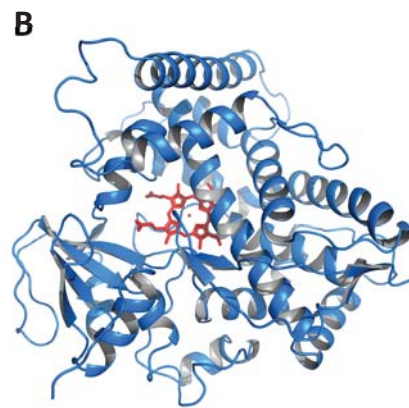
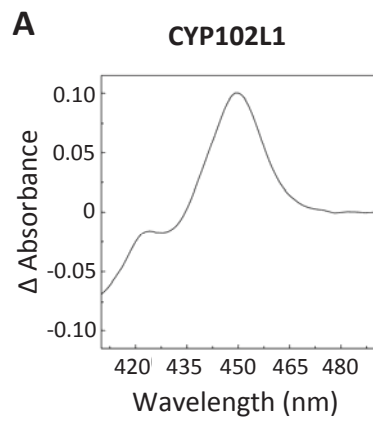


Table 1. P450 genes in the *Mimiviridae*

Taxon/Species	Virus origin	P450 family or subfamily
<i>Mimiviridae</i> A		
Mimivirus	U.K.	<i>CYP5253A: CYP5254A</i>
Mamavirus	France	<i>CYP5253A: CYP5254A</i>
Lentillevirus	France	<i>CYP5253A: CYP5254A</i>
Oyster virus	Brazil	<i>CYP5253A: CYP5254A</i>
Kroon virus	Brazil	<i>CYP5253A: CYP5254A</i>
Hirudovirus	Tunisia	<i>CYP5253A: CYP5254A</i>
Niemeyer virus	Brazil	<i>CYP5253A: CYP5254A</i>
Mimivirus Bombay	India	<i>CYP5253A: CYP5254A</i>
Mimivirus Shirakomae	Japan	<i>CYP5253A: CYP5254A</i>
Mimivirus Kasaii	Japan	<i>CYP5234A: CYP5254A</i>
Samba virus	Brazil	<i>CYP5253A: CYP5254A</i>
Amazonia virus	Brazil	<i>CYP5253A: CYP5254A</i>
<i>Mimiviridae</i> B		
Moumouvirus	France	<i>CYP5253A: CYP5254B</i>
Moumouvirus monve	France	<i>CYP5253A: CYP5254B</i>
Saudi moumouvirus	Saudi Arabia	<i>CYP5253A: CYP5254B</i>
Moumouvirus goulette	France	<i>CYP5253A: CYP5254B</i>
Moumouvirus australiensis	Australia	<i>CYP5253A: CYP5254B</i>
<i>Mimiviridae</i> C		
Megavirus chiliensis	Chile	<i>CYP5254A</i>
Mimivirus (LB111)	Tunisia	<i>CYP5254A</i>
Megavirus courdo 7	France	<i>CYP5254A</i>
Powai Lake Megavirus	India	<i>CYP5254A</i>

Table 2. P450 genes in the *Pandoraviridae* and other viruses

Species	Virus Source	P450 gene family and subfamily
Pandoraviruses		
Clade A		
<i>P. salinus</i>	Chile	CYP5634A
<i>P. dulcis</i>	Australia	CYP5634A: CYP5635A
<i>P. inopinatum</i>	Germany	CYP5634A ^a : CYP5635A
<i>P. quercus</i>	France	CYP5634A: CYP5635A
<i>P. celtis</i>	France	CYP5634A: CYP5635A
<i>P. pampulha</i>	Brazil	CYP5634A
Clade B		
<i>P. neocaledonia</i>	New Caledonia	CYP5634A
<i>P. macleodensis</i>	Australia	CYP5634A: CYP5635A
<i>P. brazilinsis</i>	Brazil	CYP5634A: CYP5635A
<i>P. massiliensis</i>	Brazil	CYP5634A
Mollivirus		
<i>Mollivirus sibericum</i>	Siberia	CYP5635B
Orpheovirus		
Orpheovirus IHUMI-LCC2	France	CYP5870A: CYP5871A: CYP5872A
Kaumoebavirus		
Kaumoebavirus	Saudi Arabia	CYP5870A
Klosneuvirus		
Hokovirus HKV1	Hong Kong	CYP5724A
Tupanvirus		
Tupanvirus (Soda Lake)	Brazil	CYP5254B: CYP5868A
Tupanvirus (Deep Ocean 3,000 meters)	Brazil	CYP5254B: CYP5868A: CYP5869A
<i>Mycobacteria</i> phage		
Adler phage	<i>Mycobacteriodes abcessus</i>	CYP102L1
<i>Herpesvirus</i>		
Ranid herpesvirus 3	<i>Ranid temoraria</i>	CYP5723A1

^aTwo separate gene predictions joined

Gray shading indicate probable known host-derived

References

1. Klingenberg M (1958) Pigments of rat liver microsomes. *Arch Biochem Biophys* 75(2):376-386.
2. Omura T & Sato R (1962) A new cytochrome in liver microsomes. *J Biol Chem* 237:1375-1376.
3. Estabrook RW, Cooper DY, & Rosenthal O (1963) The light reversible carbon monoxide inhibition of the steroid C21-hydroxylase system of the adrenal cortex. *Biochem Z* 338:741-755.
4. Nelson DR, Goldstone JV, & Stegeman JJ (2013) The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. *Philos Trans R Soc Lond B Biol Sci* 368(1612):20120474.
5. Woese CR & Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74(11):5088-5090.
6. Guengerich FP & Munro AW (2013) Unusual cytochrome P450 enzymes and reactions. *J Biol Chem* 288(24):17065-17073.
7. Nelson DR (2017) Cytochrome P450 diversity in the tree of life. *Biochim Biophys Acta*.
8. Raoult D, *et al.* (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306(5700):1344-1350.
9. Legendre M, Santini S, Rico A, Abergel C, & Claverie JM (2011) Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. *Virology* 438:89-99.
10. Colson P, La Scola B, Levasseur A, Caetano-Anolles G, & Raoult D (2017) Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nat Rev Microbiol* 15(4):243-254.
11. Andreani J, *et al.* (2017) Orpheovirus IHUMI-LCC2: A new virus among the giant viruses. *Front Microbiol* 8:2643.
12. Abrahao J, *et al.* (2018) Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat Commun* 9(1):749.
13. Andreani J, Verneau J, Raoult D, Levasseur A, & La Scola B (2018) Deciphering viral presences: two novel partial giant viruses detected in marine metagenome and in a mine drainage metagenome. *Virology* 515(1):66.
14. Schulz F, *et al.* (2017) Giant viruses with an expanded complement of translation system components. *Science* 356(6333):82-85.

15. Abergel C, Legendre M, & Claverie JM (2015) The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol Rev* 39(6):779-796.
16. Aherfi S, Colson P, La Scola B, & Raoult D (2016) Giant viruses of amoebas: An update. *Front Microbiol* 7:349.
17. Claverie JM & Abergel C (2009) Mimivirus and its virophage. *Annu Rev Genet* 43:49-66.
18. Aoyama Y, Yoshida Y, & Sato R (1984) Yeast cytochrome P-450 catalyzing lanosterol 14 alpha-demethylation. II. Lanosterol metabolism by purified P-450(14)DM and by intact microsomes. *J Biol Chem* 259(3):1661-1666.
19. Lamb DC, *et al.* (2009) The first virally encoded cytochrome P450. *J Virol* 83(16):8266-8269.
20. Desnues C, *et al.* (2012) Provirophages and transpovirons as the diverse mobilome of giant viruses. *Proc Natl Acad Sci U S A* 109(44):18078-18083.
21. Philippe N, *et al.* (2013) Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341(6143):281-286.
22. Antwerpen MH, *et al.* (2015) Whole-genome sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba. *Genome Announc* 3(2).
23. Legendre M, *et al.* (2018) Diversity and evolution of the emerging Pandoraviridae family. *Nat Commun* 9(1):2285.
24. Aherfi S, *et al.* (2018) A large open pangenome and a small core genome for giant Pandoraviruses. *Front Microbiol* 9:1486.
25. Legendre M, *et al.* (2019) *Pandoravirus celtis* Illustrates the microevolution processes at work in the giant pandoraviridae genomes. *Front Microbiol* 10:430.
26. 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(38):E5327-5335.
27. 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(11):4274-4279.
28. Bajrai LH, *et al.* (2016) Kaumoebavirus, a new virus that clusters with Faustoviruses and Asfarviridae. *Viruses* 8(11).
29. Origi FC, *et al.* (2017) Ranid herpesvirus 3 and proliferative dermatitis in free-ranging wild common frogs (*Rana Temporaria*). *Vet Pathol* 54(4):686-694.
30. Clarke M, *et al.* (2013) Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* 14(2):R11.

31. Sun Y, *et al.* (2013) Investigations of heme ligation and ligand switching in cytochromes p450 and p420. *Biochemistry* 52(34):5941-5951.
32. Wilton DC, Munday KA, Skinner SJ, & Akhtar M (1968) The biological conversion of 7-dehydrocholesterol into cholesterol and comments on the reduction of double bonds. *Biochem J* 106(4):803-810.
33. Lei L, Waterman MR, Fulco AJ, Kelly SL, & Lamb DC (2004) Availability of specific reductases controls the temporal activity of the cytochrome P450 complement of *Streptomyces coelicolor* A3(2). *Proc Natl Acad Sci U S A* 101(2):494-499.
34. Renesto P, *et al.* (2006) Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *J Virol* 80(23):11678-11685.
35. Legendre M, *et al.* (2010) mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res* 20(5):664-674.
36. Boyer M, *et al.* (2011) Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc Natl Acad Sci U S A* 108(25):10296-10301.
37. Hwang J, Park SY, Lee S, & Lee TK (2018) High diversity and potential translocation of DNA viruses in ballast water. *Mar Pollut Bull* 137:449-455.
38. Schulz F, *et al.* (2018) Hidden diversity of soil giant viruses. *Nat Commun* 9(1):4881.
39. Raoult D (2013) TRUC or the need for a new microbial classification. *Intervirology* 56(6):349-353.
40. Forterre P & Gaia M (2016) Giant viruses and the origin of modern eukaryotes. *Curr Opin Microbiol* 31:44-49.
41. Iyer LM, Balaji S, Koonin EV, & Aravind L (2006) Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res* 117(1):156-184.
42. Yutin N, Wolf YI, & Koonin EV (2014) Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* 466-467:38-52.
43. Abrahao JS, Araujo R, Colson P, & La Scola B (2017) The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. *PLoS Genet* 13(2):e1006532.
44. Filee J (2014) Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: the visible part of the iceberg? *Virology* 466-467:53-59.
45. Koonin EV & Yutin N (2018) Multiple evolutionary origins of giant viruses. *F1000Res* 7.
46. Colson P, *et al.* (2018) Ancestrality and mosaicism of giant viruses supporting the definition of the Fourth TRUC of microbes. *Front Microbiol* 9:2668.

47. Wilson WH, *et al.* (2017) Genomic exploration of individual giant ocean viruses. *ISME J* 11(8):1736-1745.
48. Cottrell MT & Kirchman DL (2012) Virus genes in Arctic marine bacteria identified by metagenomic analysis. *Aquat Microb Ecol* 66(2):107-116.
49. Sievers F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
50. Chang JM, Di Tommaso P, & Notredame C (2014) TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol Biol Evol* 31(6):1625-1637.
51. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Series* 41:95-98.
52. Nelson DR, *et al.* (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6(1):1-42.
53. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312-1313.
54. Lamb DC, *et al.* (2015) Azole antifungal agents to treat the human pathogens *Acanthamoeba castellanii* and *Acanthamoeba polyphaga* through inhibition of sterol 14alpha-demethylase (CYP51). *Antimicrob Agents Chemother* 59(8):4707-4713.
55. Arase M, Waterman MR, & Kagawa N (2006) Purification and characterization of bovine steroid 21-hydroxylase (P450c21) efficiently expressed in *Escherichia coli*. *Biochem Biophys Res Commun* 344(1):400-405.
56. Omura T & Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. II. Solubilization, purification, and properties. *J Biol Chem* 239:2379-2385.
57. Lamb DC, *et al.* (1997) The mutation T315A in *Candida albicans* sterol 14alpha-demethylase causes reduced enzyme activity and fluconazole resistance through reduced affinity. *J Biol Chem* 272(9):5682-5688.