Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes

Nadège Philippe et al.

Science 341, 281 (2013);
DOI: 10.1126/science.1239181

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 29, 2013):

Updated information and services, including high-resolution figures, can be found in the online version of this article at: http://www.sciencemag.org/content/341/6143/281.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2013/07/18/341.6143.281.DC1.html

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/content/341/6143/281.full.html#related

This article cites 57 articles, 30 of which can be accessed free:
http://www.sciencemag.org/content/341/6143/281.full.html#ref-list-1
Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes

Nadège Philippe, Matthieu Legendre, Gabriel Doutre, Yohann Couté, Olivier Poirot, Magali Lescot, Defne Arslan, Virginie Seltzer, Lionel Bertaux, Christophe Bruley, Jérôme Garin, Jean-Michel Claverie, Chantal Abergel

Ten years ago, the discovery of Mimivirus, a virus infecting Acanthamoeba, initiated a reappraisal of the upper limits of the viral world, both in terms of particle size (>0.7 micrometers) and genome complexity (>1000 genes), dimensions typical of parasitic bacteria. The discovery of these giant viruses (the Megaviridae) was aided by sampling a variety of aquatic environments and their associated sediments worldwide. We report the isolation of two giant viruses, one off the coast of central Chile, the other from a freshwater pond near Melbourne (Australia), without morphological or genomic resemblance to any previously defined virus families. Their micrometer-sized ovoid particles contain DNA genomes of at least 2.5 and 1 megabases, respectively. These viruses are the first members of the proposed "Pandoravirus" genus, a term reflecting their lack of similarity with previously described microorganisms and the surprises expected from their future study.

The serendipitous discovery of the first giant DNA virus Mimivirus (1, 2), initially misinterpreted as a Gram-positive parasitic bacterium, challenged criteria and protocols historically established to separate viruses from cellular organisms (3–5). It was then realized that virus particles could be large enough to be visible under light microscope and contain DNA genomes larger in size (>1 Mb) and gene contents (>1000) than those of bacteria. In the past decade, several Mimivirus relatives have been fully characterized, including the largest known viral genome of Megavirus chilensis (1.259 Mb encoding 1120 proteins) (6–8). The study of this new family of viruses (referred to as “Megaviridae”) revealed distinctive features concerning the virion structure and core delivery mechanism (9, 10), transcription signaling (11–13), and protein translation (14, 15). In particular, seven virus-encoded amino acid–transfer RNA (tRNA) ligases (8) and other enzymes thought to be the hallmark of cellular organisms were found in these viruses (16, 17). Their study also led to the discovery of “virophages” that replicate within the virion factory of the Megaviridae (18–20).

After our discovery of M. chilensis with laboratory-grown Acanthamoeba for amplification, we searched for new giant viruses in sediments where Acanthamoeba are more prevalent than in the water column (21, 22). We identified samples demonstrating strong cellular lytic activity. Some of these cocultures revealed the intracellular multiplication of particles larger than that of the previously isolated Megaviridae, albeit without their icosahedral appearance. As the multiplication of these particles was found to be insensitive to antibiotics, they were retained for further investigation.

Parasite 1 originated from the superficial marine sediment layer (~10 m deep) taken at the mouth of the Tunquen river (coast of central
Chile). Parasite 2 originated from mud taken at the bottom of a shallow freshwater pond near Melbourne, Australia. After amplification on Acanthamoeba cultures, both parasites became observable by optical microscopy as a lawn of ovoid particles 1 μm in length and 0.5 μm in diameter (Fig. 1A). Observations by transmission electron microscopy revealed characteristic ultrastructural features (Fig. 1) common to both parasites. Despite their identical appearance, the micro-organisms showed different global protein contents when profiled by electrophoresis (Fig. 1C). Anticipating the demonstration of their viral nature, parasites 1 and 2 will henceforth be referred to as Pandoravirus salinus and Pandoravirus dulcis.

To distinguish whether the parasites were cellular or viral in nature, we imaged their propagation in axenic Acanthamoeba cultures over an entire multiplication cycle, starting from purified particles. The replication cycle of Pandoraviruses in Acanthamoeba castellanii lasts from 10 to 15 hours and is initiated by the internalization of individual particles via phagocytic vacuoles. The particles thenempty the content of their internal compartment into the Acanthamoeba cytoplasm through an apical pore. The intact lipid membrane delimiting the particle core fuses with the vacuole membrane (Fig. 1, D and E), creating a channel through which the particle proteins and DNA content can be delivered, a process reminiscent of the one used by Mimivirus (19). This fusion process leads to a bona fide “eclipse” phase whereby the content of the particle becomes invisible once delivered into the cytoplasm. Two to 4 hours later, the host nucleus undergoes major reorganization initiated by the loss of its spherical appearance. Whereas the electron-dense nucleolus becomes paler and progressively vanishes, the nuclear membrane develops multiple invaginations, resulting in the formation of numerous vesicles (fig. S1). Peroxisome-like crystalline structures appear at the periphery of the deliquescent nucleus and progressively vanish during the particle’s maturation process (fig. S1). Eight to 10 hours after infection, the cells become rounded and lose their adherrence, and new particles appear at the periphery of the region formerly occupied by the nucleus (Fig. 1F and fig. S1). Unlike eukaryotic DNA viruses and plagues, which first synthesize and then fill their capsids, the tegument and internal compartment of the Pandoravirus particles are synthesized simultaneously, in a manner suggestive of knitting, until the particles are fully formed and closed. Curiously, particle synthesis is initiated and proceeds from the ostiole-like apex (Fig. 2). No image suggestive of division (binary fission) was obtained during ultrastructural study of particle multiplication in A. castellanii. The replicative cycle ends when the cells lyse to release about a hundred particles. The replication cycles of P. salinus and P. dulcis exhibit the same stages and characteristics.

We sequenced the genome of both parasites, starting from DNA prepared from purified particles. For P. salinus, a 2,473,870-base pair (bp) sequence was assembled as a single contig through a combination of Illumina, 454-Roche, and PacBio approaches. The sequence coverage (11,164, 67, and 41 for the above platforms, respectively) was quite uniform, except for 50 kb at the 3′ extremity of the contig where it was 10 times as high, hinting at the presence of unresolved terminal repeats. Using a combination of polymerase chain reaction (PCR) primers targeting sequences expected to arise from tandem or head-to-tail repetitions, we found evidence of at least six additional tandem terminal copies, raising the lowest estimate of the P. salinus total genome size to 2.77 Mb. The same approach was used to sequence the P. dulcis genome. The combination of the Illumina, 454-Roche, and PacBio data sets resulted in the assembly of a 1,908,524-bp sequence with an average coverage of 3,112, 62, and 133, respectively. Again, a higher coverage over 20 kb at the 3′ end of this contig hinted at the presence of two tandem terminal repeats. At strong variance with the previously sequenced Acanthamoeba giant viruses and most intracellular bacteria, the two Pandoraviruses genomes are GC-rich (G + C = 61.7 and 63.7% for P. salinus and P. dulcis, respectively), with a noticeable difference between the predicted protein-coding and noncoding regions (64% versus 54% for P. salinus). Such a high GC content remains below the extreme values reached by herpesviruses (G + C > 70%) (23). At a packing density typical of bacterial nucleoid (0.05 to 0.1 bp/μm²), a 2.8-Mb DNA molecule would easily fit into the volume (∼75 × 10^6 nm³) of the ovoid P. salinus particle.

We identified 2556 putative protein-coding sequences (CDSs) in the P. salinus 2.47-Mb unique genome sequence (considering a single terminal repeat) and 1902 for the P. dulcis 1.91-Mb genome. The alignment of the two genomes with Nuclmer (24) showed a quasi-perfect colinearity, solely interrupted by the presence of four large genomic segments specific to P. salinus (fig. S2). These additional segments mostly account for the size difference between the two genomes, indicating that the global gene content of P. dulcis is merely a subset of that of P. salinus. We thus focused our detailed analysis on the P. salinus genomic sequence.
The 2556 P. salinus predicted proteins ranged from 26 to 2367 residues [with 2364 CDSs longer than 150 nucleotides (nt)], with an average of 258 residues. The distance between consecutive CDSs was short (233 nt on average), resulting in a coding density of 80% (Fig. 3). A gene density of one protein-coding gene per kilobase is typical of both prokaryotic organisms and large double-stranded DNA (dsDNA) viruses. A comprehensive search of the National Center for Biotechnology Information nonredundant database (NR) (25) for homologs to the 2556 CDSs returned only 401 significant matches (E-value ≤ 10^-5) (15.7%), of which 215 (53.6%) primarily resulted from the sole presence of uninformative ankyrin, MORN, and F-box motifs. The large number of open reading frames (ORFs) containing these repeats is accounted for by few families of paralogs, most likely generated by local gene duplications. The largest duplications “hot spots” coincide with four regions of the P. salinus genome with no equivalent in the P. dulcis genome (fig. S2). We used the ankyrin, MORN, and F-box signatures (26) to mask P. salinus predicted protein sequences, reducing to 186 (7%) the CDSs significantly similar to NR entries (table S1). Their best matches were distributed between eukaryotes (n = 101), bacteria (n = 43), and viruses (n = 42) (fig. S3). The phylogenetic distribution of these matches, together with their low similarity levels (38% of identical residues across the best matching segment on average), indicates that no microorganism closely related to P. salinus has ever been sequenced. A similar result was obtained in comparisons against the environmental database (env_nr, 25), with only eight unique significant matches out of 341 (333 matching in NR). Only 17 P. salinus CDSs have their closest homolog (34% identical residues in average) within the Megaviridae, indicating that P. salinus has no particular phylogenetic affinity with the clade grouping the other known Acanthamoeba-infecting viruses. Similarly, only 92 (50 after masking) P. salinus CDSs (3.6% of the predicted CDSs) have an Amoebaa protein as their closest homolog, indicating that lateral gene transfers between P. salinus and its host rarely occur. The high percentage (93%) of CDSs without recognizable homolog (ORFans), the alien morphological features displayed by P. salinus, and its atypical replication process raised the concern that the translation of its genes into proteins might not obey the standard genetic code, hence obscuring potential sequence similarities. This concern was addressed by Nano-liquid chromatography–tandem mass spectrometry (LC-MS/MS) proteomic analysis of purified P. salinus particles.

The ion-mass data were interpreted in reference to a database that includes the A. castellanii (27) and the P. salinus predicted protein sequences. A total of 266 proteins were identified on the basis of at least two different peptides. Fifty-six of them corresponded to A. castellanii proteins presumably associated with the P. salinus particles, and 210 corresponded to predicted P. salinus CDSs. These identifications demonstrate that P. salinus uses the standard genetic code, legitimizing our gene predictions. Furthermore, of the 210 P. salinus–encoded proteins detected in its particle, only 42 (20%) exhibit a homolog in NR (table S2) (BlastP, E-value < 10^-5), while the rest (80%) do not. The proportion of NR-matching sequences is thus similar among experimentally validated proteins and the theoretical proteome (Fisher’s exact test, P > 0.07). This result validates the unprecedented proportion of ORFans in the P. salinus genome and confirms its large evolutionary distance from known microorganisms. Finally, 195 (93%) of the proteins identified in the P. salinus particles have a homolog encoded in the P. dulcis genome, predicting that the composition of the two virions is globally similar, even though variations in their protein sequences produce different proteomic profiles (Fig. 1C).

The functional annotation of P. salinus–predicted proteins was complemented by motif searches (26) and three-dimensional–fold recognition programs (28). The failure to detect components of the basic cellular functions—i.e., protein translation, adenosine 5’-triphosphate generation, and binary fission (3, 5)—confirmed the viral nature of Pandoraviruses. P. salinus possesses none of the ribosome components (RNAs and proteins) and no enzyme from the glycolysis pathway or the Krebs cycle. Our search was similarly unsuccessful for homologs of cell division–related proteins such as FtsZ (29), tubulin (30), or components of the alternative ESCRT system (31). P. salinus thus lacks most of the hallmark components of cellular organisms, including those retained in the most reduced intracellular parasites (5).

Nonetheless, the P. salinus genome exhibited 14 of the 31 genes most consistently present in large dsDNA viruses [i.e., “core” genes (32)] (table S3). We identified three of the nine most conserved (type I) core genes (including a DNA polymerase and four copies of virion packaging adenosine triphosphatase). We also identified four out of the eight type II (lesser conserved) core genes (including the two subunits of the ribonucleotide reductase) and 7 of the 14 type III core genes (including an mRNA-capping enzyme and three subunits of the DNA-dependent RNA polymerase). Yet, P. salinus lacks several core genes that encode components essential to DNA replication such as DNA ligases, topoisomerases, and the DNA sliding clamp (Proliferating Cell Nuclear Antigen). This already suggests that, in contrast to the largest known viruses, replication of Pandoraviruses requires host functions normally segregated in the nucleus. Another notable absence is that of a gene encoding a major capsid protein, a hallmark of all large eukaryotic dsDNA viruses, except for the Poxviruses which, like P. salinus and P. dulcis, lack icosahedral symmetry. Nor does P. salinus possess a homolog

![Fig. 2. Electron microscopy images of ultrathin sections of P. salinus.](A to C) Three stages of maturation are presented, illustrating the progressive knitting together of the particles starting from the apex and ending up as mature virions fully encased in their tegument-like envelope.

www.sciencemag.org   SCIENCE   VOL 341   19 JULY 2013   283
of the vaccinia scaffolding protein D13, which is structurally similar to the double-barreled capsid protein found in icosahedral dsDNA viruses (33).

Despite lacking several of the large dsDNA virus core genes (table S3), *P. salinus* remains typically virus-like by possessing a large fraction of enzymes involved in DNA processing (including replication, transcription, repair, and nucleotide synthesis) (table S1). Its 54 DNA-processing proteins include three enzymes that have no known homolog in viruses: a p-aminobenzoic acid synthase, a dihydroneopterin aldolase, and a hydroxymethylpterin-pyrophosphokinase (HPPK). Transcription is represented by four RNA polymerase subunits, two copies of VLT3-like gene transcription factors, an SII-like transcription elongation factor, and a DEAD-like helicase. Besides DNA-processing proteins, we identified 82 proteins involved in miscellaneous cellular functions, none of which related to a specific feature of the Pandoravirus replication cycle. We identified several components of the ubiquitin-dependent protein degradation pathway, and various hydrolases and proteases, kinases, and phosphatases, likely to interfere with the host metabolism, as well as four fascin-domain–containing proteins potentially involved in the formation of intracytoplasmic substructures. We also identified two amino acid–tRNA ligases, one for tyrosine (TyrRS) and the other for tryptophan (TrpRS). Before this study, the presence of virally encoded amino acid–tRNA ligases was a hallmark of the Megaviridae (6–8) and their closest known relative *Cafeteria roenbergensis* virus (CroV) (34). However, the TyrRS and TrpRS encoded by the Pandoraviruses are much closer to their Acanthamoeba homologs (57 and 58% identity, respectively) than to their Megaviridae counterparts, arguing against a common viral ancestry for these genes (fig. S4).

*P. salinus* also possesses few other translation-related genes: a eIF4E translation initiation factor, a SUA5-like tRNA modification enzyme, and three tRNAs (tRNAPro, tRNA^Met^, and tRNA^Trp^).

Consistent with the subcellular location of their replication, the cytoplasmic large DNA viruses (e.g., Megaviridae, Poxviridae, and Iridoviridae) lack spliceosomal introns. Even the nucleus-dependent Chloroviruses (e.g., PBCV-1) have only few small introns (35). Unexpectedly, 16 of the 186 (~9%) *P. salinus* CDSs with database homologs contain one or more introns (table S4). These introns are 138 nt long on average, bear no resemblance with group I or group II self-splicing introns and, once validated by reverse transcriptase–PCR, were found to be precisely delineated by a 5′-GT and 3′-AG dinucleotide. These spliceosomal introns are most likely excised from the *P. salinus* transcripts by the cellular U2-dependent splicing machinery, which strongly suggests that at least part of the *P. salinus* genome is transcribed within the host nucleus. Fourteen out of the 39 identified introns (36%) remained in frame with the flanking coding regions and exhibited a similar GC content, making their computational detection impossible for ORFs...
without database homologs. The introns that were not in frame exhibited a GC content 10% lower than that of their flanking exons. A comprehensive transcriptome analysis will be required to identify all the intron-containing genes potentially representing around 10% of the predicted genes, as estimated from the few that exhibit database homologs. Finally, as in other large DNA viruses (2, 8, 34), a handful of essential DNA synthesis enzymes contain inteins: one in the largest RNA polymerase subunit and the DNA polymerase (RNR) subunit, and two in the large RNR subunit and the DNA polymerase. The P. salinus small RNR subunit and the DNA polymerase enzymes are interrupted by both inteins and introns (fig. S5).

To quantitatively analyze the proteomic content of the P. salinus particles, we first scrutinized the most abundant proteins, searching for a candidate capsid-like protein. Two prominent proteins with molecular masses of ~60 kD were visible (Fig. 1C). However, the most abundant of these does not resemble any known protein, whereas the second protein is similar to a conserved Megaviridae protein, albeit of unknown function (table S2). Furthermore, Pandoraviruses encoded transcription machinery was completely absent in the particle, in contrast to Mimivirus and other viruses that replicate in the cytoplasm (16). Together with the presence of spliceosomal introns, this finding confirms that the host nucleus is actively involved in the early stage of the Pandavirus replication cycle, before decaying at a later stage. The proteomic data also confirmed four splice junction predictions (table S4). Finally, 56 low-abundance A. castellanii proteins were detected in the proteome of the particles (table S2). Because Pandoraviruses replicate in the host cytoplasm, but not inside a well-defined cellular substructure, these Acanthamoeba proteins may be randomly packaged into the virion as simple bystanders.

The discovery of Mimivirus, followed by the characterization over the past decade of other Megaviridae exhibiting slight increments in genome sizes, suggested that the maximum viral genome size possible was about 1.3 Mb and 1200 genes, a genetic complexity already larger than that of many parasitic bacteria. Meanwhile, the discovery of viruses with smaller genomes, but sharing several features previously thought to be specific to the Megaviridae (2, 8, 18, 36), indicated a phylogenetic continuum between the giant viruses and other dsDNA viruses (5, 8, 34). This conceptual framework is challenged by the Pandoraviruses that have genomes twice as large as, and lack any phylogenetic affinity with, previously described virus families (Fig. 4). Indeed, the Pandavirus genome size exceeds that of parasitic eukaryotic microorganisms, such as Encephalitozoon species (37, 38).

Because more than 93% of Pandoraviruses genes resemble nothing known, their origin
Sept4/ARTS Regulates Stem Cell Apoptosis and Skin Regeneration

Yaron Fuchs,1 Samara Brown,1 Travis Gorenc,1 Joe Rodriguez,2 Elaine Fuchs,2* Hermann Steller*1

Adult stem cells are essential for tissue homeostasis and wound repair. Their proliferative capacity must be tightly regulated to prevent the emergence of unwanted and potentially dangerous cells, such as cancer cells. We found that mice deficient for the proapoptotic Sept4/ARTS gene have elevated numbers of hair follicle stem cells (HFSCs) that are protected against apoptosis. Sept4/ARTS−/− mice display marked improvement in wound healing and regeneration of hair follicles. These phenotypes depend on HFSCs, as indicated by lineage tracing. Inactivation of XIAP, a direct target of ARTS, abrogated these phenotypes and impaired wound healing. Our results indicate that apoptosis plays an important role in regulating stem cell–dependent regeneration and suggest that this pathway may be a target for regenerative medicine.

T he ability of stem cells (SCs) to self-renew and differentiate enables them to replace cells that die during tissue homeostasis or upon injury. Elevated SC numbers might be desirable, at least transiently, to enhance tissue repair (1, 2). However, a large SC pool may potentially increase the risk of cancer (3).

One major mechanism that eliminates undesired and dangerous cells is apoptosis (4). Relatively little is known about the role of apoptosis in controlling SC numbers and its possible effect on SC-dependent regeneration. Apoptosis is executed by caspases that are negatively regulated by IAPs (inhibitor of apoptosis proteins) (5, 6). The best-studied mammalian IAP is XIAP (7). In cells destined to die, IAPs are inactivated by specific antagonists (8, 9). One mammalian IAP-antagonist is ARTS, a splice variant of the mammalian gene Sept4 (Sept4) (10, 11). Deletion of the Sept4/ARTS gene results in increased numbers of hematopoietic stem and progenitor cells and elevated XIAP levels. This causes increased apoptotic resistance and accelerated tumor development (12). Here, we report crucial roles of XIAP and Sept4/ARTS in regulating hair follicle stem cell (HFSC) apoptosis and show that apoptotic alterations have profound consequences for wound healing and regeneration.

Hair follicles cycle between phases of growth (anagen), destruction (catagen), and rest (telogen). This process requires distinct populations of HFSCs that reside within the bulge (13–16). ARTS was the only Sept4 isoform detected in HFSCs (fig. S1, A to C). To investigate the consequences of ARTS deficiency, we examined bulge HFSCs with specific bulge markers (CD34 and K15).

References and Notes
43. M. Krajewski et al., Genome Res. 19, 1639–1645 (2009).

Acknowledgments: We thank S. Faugeron and R. Finke from the Estación de Investigaciones Marinas in Chile for help during the sampling expedition. We also thank J. Hajdu for invaluable support and J.-P. Chauvin and A. Aouane for expert assistance on the Institut de Biologie du Développement de Marseille Luminy imaging facility, as well as A. Bernadac and A. Kosta from the Institut de Microbiologie de la Méditerranée. We thank E. Fabre and V. Schmidt for technical assistance, and P. Bonin and R. Claverie for helpful discussions. This work was supported by Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, Centre de l’Énergie Atomique, and Agence National pour la Recherche (ANR-BLAN08-0089, ANR-09-GENM-032-001, and ANR-10-INBS-001). The sampling expedition was sponsored by the ASSEMBLE grant 227799. The genome sequences of P. salinus and P. dulcis have been deposited in GenBank (accession numbers KC797571 and KC797570, respectively). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD000213 and DOI 10.6019/PXD000213.

Supplementary Materials
www.sciencemag.org/cgi/content/full/341/6143/281/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S5
References (46–59)
15 April 2013; accepted 13 June 2013
10.1126/science.1239181