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, et al.

To cite this version:

HAL Id: cea-00862677
https://hal-cea.archives-ouvertes.fr/cea-00862677
Submitted on 17 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
sensitizes the receptor to agonist. In this manner, MRAP2b would convert the adult zebrafish MC4R from a constitutively active to a ligand-dependent receptor. Additionally, mrap2b expression kinetics matches that of the mouse MRAP2 and is functionally homologous. MRAP2a and MRAP2b proteins share a strong homology with each other and with mammalian MRAP2 in most of the N-terminal region and the transmembrane domain, whereas the N-terminal 15 amino acids and C terminus of these proteins are vastly divergent (fig. S6). Interestingly, the N-terminal and transmembrane domain of MRAP1 are sufficient for full activity of the mammalian MC2R (15). The first 15 amino acids of MRAP2a and MRAP2b could thus represent an important regulatory region of MRAP2s, possibly responsible for the differential regulation of MC4R.

During zebrafish embryonic development, all the energy consumed is obtained from the yolk sac. Our findings suggest that the embryo benefits from having MC4R locked in an inactive state by the joint actions of AgRP and MRAP2a. Activation of MC4R at this stage would slow the rapid maturation to the mobile free-feeding juvenile stage reached at 5 dpf. Upon maturation and depletion of the yolk sac, the zebrafish must regulate nutrient intake. Appropriate behavioral response to diurnal, seasonal, and other inputs requires a functional adipostat and energy balance sensor. This switch is aided by MRAP2b, which forms a complex with MC4R and renders it highly sensitive to α-MSH.

MRAP2 introduces a previously unappreciated level of complexity in the control of MC4R, with developmentally regulated paralogs in the fish that can either inhibit (MRAP2a) or stimulate (MRAP2b) ligand-mediated receptor activation (fig. S8). A component of this complexity is retained in mammals: Asai et al. (16) show that, like MRAP2b, mouse MRAP2 expression is activated proximal to weaning and increases the responsivity of MC4R to α-MSH. Their observation that MRAP2 deletion causes an obesity syndrome in the mouse can likely be attributed, in part, to reduced function of MC4R (16). However, given the ubiquitous expression of MRAP2 proteins, we hypothesize that these proteins also modulate the activity of GPCRs and perhaps other membrane proteins as well.

References and Notes
This fusion process leads to a bona fide DNA content can be delivered, a process the vacuole membrane (Fig. 1, D and E), creating membrane delimiting the particle core fuses with the particles’ maturation process (fig. S1). Eight to 10 hours after infection, the cells become rounded and lose their adherence, 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 phages, 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 × 10⁶ 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 1502 for the *P. dulcis* 1.91-Mb genome. The alignment of the two genomes with Nucmer (24) showed a quasiperfect 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.

---

**Fig. 1. Images of Pandoravirus particles and their proteomic profiles.** Light microscopy (A) and electron microscopy images (B) of *P. salinus* (1) and *P. dulcis* (2) purified particles. (C) Electrophoresis profiles of *P. salinus* (lane 1) and *P. dulcis* (lane 2) extracted proteins. (D) Internalized *P. salinus* particle in the host vacuole. Once fused with the vacuole membrane (arrow), the virion internal membrane continues a continuum with the host cytoplasm. The particles are wrapped into a ∼70-nm-thick tegument-like envelope consisting of three layers. (E) Magnified image of the opened ostiole-like apex: from the inside out, a layer of light density of unknown composition (∼20 nm, marked “a”), an intermediate dark layer comprising a dense mesh of fibrils (∼25 nm, marked “b”), and an external layer of medium density (∼25 nm, marked “c”). This tegument-like envelope is interrupted by the ostiole-like pore measuring ∼70 nm in diameter. As shown in (B1) and (B2), the lipid membrane internal to the particle encloses a diffuse interior devoid of visible substructure, except for a spherical area of electron-dense material (50 nm in diameter, arrowhead) seen episodically but in a reproducible fashion. (F) Ultrathin section of an Acanthamoeba cell filled with *P. salinus* at various stages of maturation.
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 \( \leq 10^{-5} \)) (15.7%) (fig. S3), 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 Amoeboida 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 *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.
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.

Fig. 3. Structure of the P. salinus genome. Specific features are marked on concentric circles using Circos (43) as follows: 1, CDSs positions on the direct (blue) and reverse (red) strands. 2, CDSs with a best match with eukaryotes (in orange), bacteria (in green), and viruses (in purple). CDSs with MORN repeats, ankyrin repeats, and F-box domain motifs are shown in white; CDSs with no match are shown in gray. 3, CDSs identified in the proteome of purified P. salinus particles.
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 continuity 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 Pandoravirus 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,2 Travis Gorenc,3 Joe Rodriguez,3 Elaine Fuchs,2* Hermann Steller4*

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.

The 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). A 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). 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, Provence-Côte d’Azur Region, and Agence Nationale pour la Recherche (ANR-09-BLAN-0089, ANR-10-GENM-032-001, and ANR-14-INBS-008). The sampling expedition was sponsored by the ASSEMBLE grant 22779. The genome sequences of P. salinus and P. dulcis have been deposited in GenBank (accession numbers KC977571 and KC977570, respectively). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomeexchange.org) via the PRIDE partner repository with the data set identifier PXD000213 and DOI 10.6019/PXD000213.

3Strang Laboratory of Apoptosis and Cancer Biology, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA. 4*Corresponding author. E-mail: steller@rockefeller.edu (H.S.); fuchslb@rockefeller.edu (E.F.)

References and Notes

43. M. Krzywinski et al., Genome Res. 19, 1639–1645 (2009).

Supplemental 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
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 Sellitzer, Lionel Bertaux, Christophe Bruley, Jérome Garin, Jean-Michel Claverie and Chantal Abergel

Science 341 (6143), 281-286.
DOI: 10.1126/science.1239181

Zeus’ Revenge

Sediment-dwelling amoebae appear to have an unhappy affinity for huge viruses. Giant icosahedral Mimiviruses with genomes of the order of 1 megabase (Mb) were first identified in Acanthamoeba. Digging into antipodean sediments has once again been fruitful where Philippe et al. (p. 281; see the cover) discovered some enormous viruses in Acanthamoeba, visible by light microscopy and having genomes up to 2.5 Mb. The Pandoraviruses are phagocytosed by target cells and, after fusing with the phagosome membrane, their contents are released into the cytoplasm where they wreak terrible havoc on its nucleus. These viruses are encased into a tegument-like envelope and lack genes for capsid proteins, and there are no genes for protein translation, adenosine triphosphate generation, or binary fission—confirming their classification as viruses.