Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

DNA repair genes in the Megavirales pangenome

Romain Blanc-Mathieu and Hiroyuki Ogata

The order 'Megavirales' represents a group of eukaryotic viruses with a large genome encoding a few hundred up to two thousand five hundred genes. Several members of Megavirales possess genes involved in major DNA repair pathways. Some of these genes were likely inherited from an ancient virus world and some others were derived from the genomes of their hosts. Here we examine molecular phylogenies of key DNA repair enzymes in light of recent hypotheses on the origin of Megavirales, and propose that the last common ancestors of the individual families of the order Megavirales already possessed DNA repair functions to achieve and maintain a moderately large genome and that this repair capacity gradually increased, in a family-dependent manner, during their recent evolution.

Address
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

Corresponding author: Ogata, Hiroyuki (ogata@kuicr.kyoto-u.ac.jp)

Current Opinion in Microbiology 2016, **31**:94–100

This review comes from a themed issue on **Megaviromes**

Edited by **Didier Raoult** and **Jonatas Abrahao**

http://dx.doi.org/10.1016/j.mib.2016.03.011

Introduction

Nucleo-cytoplasmic large DNA viruses (NCLDVs) are a group of viruses that infect diverse members of eukaryotes and possess a large double-stranded DNA genome varying in size from 100 kb to 2.5 Mb [1,2]. They are considered to form a monophyletic group based on the conservation of several genes [3], which led to the recent proposal of the order 'Megavirales' to refer to this viral group [4].

Microbiologists and evolutionary biologists have investigated the evolution of Megavirales genes by sequence and molecular phylogenetic analyses and proposed theories on the origin of Megavirales and its association with the emergence of eukaryotes [5,6••,7,8,9•]. Depending on hypotheses, Megavirales genes can have different origins. A recent hypothesis postulates that Megavirales evolved from ancient DNA transposons of the Polintons family that are themselves the remnant of Tectoviridae-like bacteriophages that entered the protoeukaryotic cell along with the alphaproteobacterial endosymbiont [6••,10•]. Under this scenario, Megavirales are the product of a 'melting pot' of early viral evolution and part of their genes were directly derived from an ancient virus world as 'hallmark viral genes' [11]. Alternatively, Megavirales genes can be acquired from known cellular organisms including viral hosts ('the accretion hypothesis') [12,13•,14•], or could have been vertically inherited from unknown and ancestral cellular organisms ('the reductive evolution from the fourth (or more) domain(s) hypothesis') [5,15••]. These scenarios are not totally exclusive of each other. For instance, an ancient cellular organism might have evolved to an ancestor of Megavirales by reductive genome evolution, followed by later re-accumulation of cellular genes in the course of viral evolution.

One of the notable features uncovered from the genomics of Megavirales is that some members abundantly encode DNA repair genes in their genomes [16,17]. DNA repair functions are essential for cellular organisms (i.e., 'large genomes') but are infrequent or absent in smaller viral genomes. Certain DNA repair enzymes are known to be specific to a subset of Megavirales members having a particularly large genome [18]. Therefore, Megavirales DNA repair genes are of interest to understand the evolution of their large genomes. In this manuscript, we review the homologs of key DNA repair genes in the pangenome of Megavirales with a special focus on their phylogenetic relationships with eukaryotic homologs.

Classification of Megavirales

The order Megavirales is currently composed of seven families: *Mimiviridae*, *Marseilleviridae*, *Phycodnaviridae*, *Poxviridae*, *Asfarviridae*, *Iridoviridae* and *Ascoviridae* [4]. Following Santini and colleagues [19], instead of *Mimiviridae*, we used the term 'Megaviridae' family which includes viruses classified in *Mimiviridae* plus *Phaeocystis globosa virus* (PgV), *Cafeteria roenbergensi virus* (CroV), Organic lake phycodnaviruses (OLPV1 and 2), *Aureococcus anophagefferens virus* (AaV) [20] and *Chrysochromulina ericina virus* (CeV) [21]. The recent discoveries of pandoraviruses (1.9–2.5 Mb; [22,23]), *Mollivirus sibericum* (650 kb, [24]), *Pithovirus sibericum* (610 kb, [25]), substantially expanded the known diversity of Megavirales lineages [26,27]. These viruses wait for their official taxonomic classifications; a phylogenetic analysis suggests that pandoraviruses are a group of viruses belonging to the *Phycodnaviridae* family [2].

Why do Megavirales genomes encode many DNA repair genes?

There are five known DNA repair pathways: the base excision repair (BER), nucleotide excision repair (NER), double-strand break (DSB) repair, mismatch repair (MMR), and direct damage reversal (DDR) pathways (Box 1). Most of the key enzymes in these DNA repair pathways can be found in the genomes of Megavirales (Table S1). Transcriptomic [24,28–33] and proteomic [24,34–37] studies revealed that many of these DNA repair genes are transcriptionally active during infection and that gene products (i.e., enzymes) are packaged into Megavirales virions (Table S1). Individual Megavirales genomes encode 0 up to 15 of those genes (3.6 genes on average). In contrast, other smaller double-stranded DNA viruses encode 0–5 of those genes (0.4 genes on average). Redrejo-Rodríguez and Salas provided a comprehensive review on the BER pathway genes in the pangenome of Megavirales [38]. Interestingly, members of the Megaviridae family, which possess a relatively large genome compared to the members of other families, encode the full or almost full set of genes needed to accomplish a viral BER pathway.

Microorganisms (viruses, prokaryotes and unicellular eukaryotes) with a larger genome tend to encode a larger number of genes than those with a smaller genome [39]. As a consequence, the target-size of deleterious mutation is greater for microorganisms with a larger genome. We thus expect a lower mutation rate per base per generation (i.e., a higher fidelity in replication and repair) for microorganisms with a larger genome than with a smaller genome. Drake postulated that the mutation rate per nucleotide site per generation scales inversely with genome size of DNA-based microorganisms and proposed that the microbial mutation rate per genome may have evolved towards a nearly invariant value across taxa [40]. Subsequent analyses demonstrated that most microorganisms with an estimated mutation rate per nucleotide site per generation conform to the inverse scaling observed by Drake [41,42]. There are currently no data available for the mutation rate for the viruses of Megavirales. However, in line with Drake's rule, the genome size positively correlates with the number of DNA repair genes in Megavirales (Spearman's $\rho = 0.4$, $P = 1.7 \times 10^{-4}$) (Figure 1). Exceptions to this trend are the three pandoraviruses encoding only two of the known repair genes that we analyzed. Pandoraviruses may rely on host's DNA repair machinery or they may encode either highly divergent or unidentified DNA repair genes.

DNA repair genes from an ancient virus world

Previous phylogenetic analyses (see references in Table 1) suggested an old origin for some of the DNA repair genes encoded in the Megavirales pangenome, which may predate the radiation of major eukaryotic lineages. Most of DNA processing genes are not conserved across the three domains of life. Mre11/Rad50 are among the few DNA processing enzymes that are universally conserved in the three domains [43]. Based on this observation, some authors suggested that the last universal common ancestor (LUCA) was capable of processing DNA [44]. However, Forterre argues against this hypothesis based on the fact that majority of the core enzymes for DNA replication are not homologous between bacteria and eukaryote/archaea [45], and proposes that LUCA still had an RNA genome and that the descendant cellular lineages (leading to the three domains) acquired DNA processing enzymes including Mre11/Rad50 not from cellular organisms but from viruses [46]. Forterre's scenario assumes the existence of viral ancestors that possessed Mre11/Rad50 genes. Interestingly, some Megavirales members possess Mre11 and/or

Box 1 DNA repair pathways

Base excision repair (BER)

Most damage to bases in DNA is repaired by the BER pathway [54]. There are several variations in BER, resulting in the replacement of either a single nucleotide or a slightly longer stretch of DNA strand. A typical BER is initiated by any of several DNA glycosylases (uracil DNA glycosylase (UDG), Fpg or 3-methyladenine (3-MeA) DNA glycosylase) to remove an incorrect or damaged substrate base. This creates an abasic site, which is incised by an apurinic/apyrimidinic (AP) endonuclease. A lyase or phosphodiesterase then removes the remaining sugar, leaving a gap that will be filled by a family X DNA polymerase before a DNA ligase seals the nicked DNA strand [55].

Nucleotide excision repair (NER)

NER, initially described in bacteria [56] and later shown to be present in all domains of life, is the most versatile DNA repair pathway [57,58]. One of the key enzymes recruited in the NER pathway is the endonuclease of the eukaryotic XPG family, a single-stranded structure-specific DNA endonuclease, which cleaves single-stranded DNA during NER to excise damaged DNA [59,60].

Double-strand break (DSB) repair

DSBs are a major cause of genomic instability [61]. DSBs occur upon exposure of the genome to DNA-damaging agents but also emerge during normal DNA metabolic processes including replication, recombination and viral recombination-dependent DNA replication [62,63]. Eukaryotic/archaeal Mre11/Rad50 (and the orthologous bacterial SbcD/SbcC) constitute the key machinery that recognizes DSBs and bridges two DNA ends to initiate DSB repair.

Mismatch repair (MMR)

MMR recognizes and corrects base–base mismatches and small indels introduced during normal replication, leading to 50-folds to 1000-folds enhancement of replication fidelity [64,65]. In bacteria, MutS (MutS1) recognizes mismatches/indels and MutH, an endonuclease, introduces a nick in the newly synthesized DNA strand to start MMR. MutS homologs are ubiquitous in cellular organisms, and exhibit ancient paralogs with different functions in MMR, recombination or chromosomal stability [64].

Direct damage reversal (DDR)

In contrast to the sophisticated BER, NER, DSB and MMR pathways, DDR makes use of a single protein to remove lesions without incision of the sugar–phosphate backbone or base excision [66]. The reversing of UV light-induced photolesions by photolyases is one of the three major DNA repair mechanisms by DDR that have been identified to date [67].