Concluding remarks

As mentioned earlier, there are two main, but not mutually exclusive, models for the mechanism by which paramutation occurs. The data presented in these recent papers favour the role of RNA in paramutagenic silencing. In a nutshell, the paramutagenic, transgenic allele Kit^{tm1Alf} is associated with the presence of vast numbers of abnormal RNA molecules in the sperm. The injection of this RNA is sufficient to induce paramutation in the absence of Kit^{tm1Alf}. Consistent with this, an RNA-dependent RNA polymerase has been found to be required for paramutation in plants. Paramutation is usually associated with changes in chromatin structure and often occurs at loci that contain repetitive elements. The silencing of these repetitive elements probably occurs through an RNA-mediated silencing pathway, and RNA in the germline is associated with proteins from families known to be involved in chromatin alterations. These reports provide strong, albeit circumstantial, evidence for the involvement of RNA in paramutation, indicating that RNA has yet another role in the biology of higher organisms.

Acknowledgements

We thank the National Health and Medical Research Council, and the Australia Research Council, for continued support. A.A. is supported by an Australian Postgraduate Award; E.W., by a Queensland Institute of Medical Research Fellowship.

References

1 Rakyan, V. and Whitelaw, E. (2003) Transgenerational epigenetic inheritance. *Curr. Biol.* 13, R6
2 Chandler, V.L. and Stam, M. (2004) Chromatin conversations: mechanisms and implications of paramutation. *Nat. Rev. Genet.* 5, 532–544
3 Stam, M. and Mittelsten Scheid, O. (2005) Paramutation: an encounter leaving a lasting impression. *Trends Plant Sci.* 10, 283–290
4 Herman, H. *et al.* (2003) *Trans* allele methylation and paramutation-like effects in mice. *Nat. Genet.* 34, 199–202
5 Rassoulzadegan, M. *et al.* (2002) Transvection effects involving DNA methylation during meiosis in the mouse. *EMBO J.* 21, 440–450
6 Rassoulzadegan, M. *et al.* (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* 441, 469–474
7 Alleman, M. *et al.* (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* 442, 295–298
8 Bernstein, E. and Allis, C.D. (2005) RNA meets chromatin. *Genes Dev.* 19, 1635–1655
9 Fukagawa, T. *et al.* (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat. Cell Biol.* 6, 784–791
10 Volpe, T.A. *et al.* (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297, 1833–1837
11 Aravin, A. *et al.* (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442, 203–207
12 Girard, A. *et al.* (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202
13 Deng, W. and Lin, H. (2002) *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* 2, 819–830
14 Kuramochi-Miyagawa, S. *et al.* (2004) *Mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. *Development* 131, 839–849
15 Watanabe, T. *et al.* (2006) Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* 20, 1732–1743
16 Costa, Y. *et al.* (2006) Mouse MAELSTROM: the link between meiotic silencing of unsynapsed chromatin and microRNA pathway? *Hum. Mol. Genet.* 15, 2324–2334
17 Kotaja, N. *et al.* (2006) The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2647–2652
18 Lane, N. *et al.* (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* 35, 88–93

0168-9525/$ – see front matter
doi:10.1016/j.tig.2006.11.008

Genome Analysis

I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses

Jonathan Filée, Patricia Siguier and Mick Chandler

Laboratoire de Microbiologie et Génétique Moléculaires, CNRS 118, Route de Narbonne, F-31062 Toulouse Cedex, France

Giant viruses are nucleocytoplasmic large DNA viruses (NCLDVs) that infect algae (phycodnaviruses) and amoebae (*Mimivirus*). We report an unexpected abundance in these giant viruses of islands of bacterial-type genes, including apparently intact prokaryotic mobile genetic elements, and hypothesize that NCLDV genomes undergo successive accretions of bacterial genes. The viruses could acquire bacterial genes within their bacteria-feeding eukaryotic hosts, and we suggest that such acquisition is driven by the intimate coupling of recombination and replication in NCLDVs.

Corresponding author: Filée, J. (jonathan.filee@ibcg.biotoul.fr)
Available online 15 November 2006.

Introduction

Nucleocytoplasmic large DNA viruses (NCLDV) constitute a diverse group that infects a wide range of eukaryotic hosts including algae (phycodnaviruses), protists (*Mimivirus*) and metazoa (poxviruses, African swine fever virus, iridoviruses). They either replicate exclusively in the cytoplasm or begin their cycle in the host nucleus before passage into the cytoplasm [1]. Consistent with this way of life, they carry most of the genes necessary for DNA metabolism, replication and transcription in addition to those involved in virion assembly.

Giant viruses are thought to be monophyletic based on a common set of 30 homologous genes [1]. The recent sequencing of giant NCLDV representatives with genomes of

between 330 kb (phycodnaviruses) and 1.2 Mb (*Mimivirus*) indicated that their genomic repertoire is large compared with other DNA viruses (30–200 kb) and equivalent to that of several parasitic bacteria [2]. Most giant virus genes are lineage-specific and globally their origin is poorly understood. It was recently observed that their genome size is correlated with the number of paralogous genes [2,3], suggesting frequent successive gene duplications during giant virus genome evolution.

Despite much effort, we are still unable to place viruses on the universal tree of life and their origin remains speculative. Interestingly, almost all giant virus lineages carry genes with bacterial and eukaryotic homologs [1,4]. These genes could testify to an ancient origin predating the divergence of the three kingdoms of life. Each giant virus lineage would then have retained a diverse assemblage of genes resembling bacterial or eukaryotic genes [5]. Alternatively, giant viruses could be regressive or highly derived cells that have undergone a simplification process [5]. These explanations contrast with the traditional view of evolution in which virus and phage are thought to have evolved mainly by accretion of genes from disparate sources [6,7]. However, analysis of the *Mimivirus* genome revealed that most genes homologous to eukaryotic genes are only distantly related to those of their host (an amoeba) [8]. This contradicts the notion of numerous lateral gene transfers with its amoebal host. We propose that NCLDVs can acquire foreign DNA sequences, and that some, as a result of their lifestyles, have acquired significant numbers of bacterial genes.

The distribution of bacterial-like genes in giant viruses

The presence of bacterial-like genes had previously been documented in both mimiviruses and phycodnaviruses [1,4]. We were interested in the distribution of these genes across the respective genomes to gain insight into the evolution of the viral genomes (e.g. contiguous groups could reflect their inheritance as a block).

To identify bacterial-like genes, we used BLAST analysis with a lower cut-off of an expected value (E) $<10^{-5}$. Using this cut-off, most of the bacterial genes identified had no eukaryotic homologs or eukaryotic genes were so distantly related that no ambiguity was possible. For *Mimivirus*, 78 out of 96 bacterial-like genes (i.e. 81%) were unambiguously bacterial in origin with no, even distantly related, eukaryotic homologs (Online Supplementary Material Table S1). Between 48 and 57 bacterial-like genes were also identified in each of the three *Chlorella* phycodnaviruses. For *Chlorella virus NY2A*, 87% of these are unambiguously of bacterial origin (Online Supplementary Material Table S2). Only a small fraction of the identified bacterial-like genes have weak statistical support owing to the lack of resolution of the phylogenetic tree.

The bacterial-like genes showed a strong bias toward 'DNA replication and repair' (48% of proteins for the NY2A phycodnavirus and 20% in *Mimivirus*) and 'cell envelope' (14% for the NY2A phycodnavirus and 12.5% for *Mimivirus*) in COG (clusters of orthologous groups) functional gene categories (Online Supplementary Material Tables S1, S2).

Coinheritance of bacterial-like genes

Three consecutive reading frames (L136, L137 and L138) carried by *Mimivirus* are syntenic with three frames in *Clostridium acetobutylicum*. L136 encodes a sugar transaminase (COG0399) homologous to gene CAC2350 in *C. acetobutylicum* (BLASTP E-value = $10e^{-19}$) (Online Supplementary Material Table S2). L137 encodes a glycosyl transferase (COG0443) homologous to CAC2349 (BLASTP E-values = $10e^{-37}$), whereas L138 has weak similarities (BLASTP E-values $> 10e^{-5}$) with CAC2348 (unknown function). This suggests that these bacterial-like genes were inherited as a short contiguous block.

In contrast to the genome of *Paramecium bursaria Chlorella virus 1* (PBCV-1), the NY2A genome has 13 additional bacterial-like genes including seven DNA methyltransferases. The genome of NY2A encodes at least 14 DNA methyltransferases. At least two are associated with a site-specific restriction endonuclease resembling the arrangement of bacterial restriction or modification systems, again implying coinheritance. Many phycodnavirus bacterial-like genes encode functional proteins [4] suggesting that most have been recruited and are maintained by the virus, possibly to fulfil viral functions.

Genomic location of bacterial islands in giant viruses

The bacterial-like genes are not randomly dispersed but tend to be localized in specific genomic regions toward the ends of the genome, possibly in the form of islands (Figure 1). In *Mimivirus* they are positioned within the first and last 250 kb of the genome. In contrast, NCLDV core genes and genes with eukaryotic affinities are localized toward the middle of the genome. In phycodnaviruses, bacterial-like genes are more dispersed but still show this tendency. As in most viruses, there is a great abundance of orphan genes but these are distributed relatively evenly over the entire genome (Figure 1). It is probable that a fraction of these will eventually be found to derive from unidentified bacterial, viral or eukaryotic sources.

The GC skew, which in bacteria can be associated with the direction of replication or imposed by the polarity of the coding and noncoding strands [9], seems to change near the boundaries between the proposed terminal regions of bacterial-like genes (Online Supplementary Material Figure S1; D.H. Evans, pers. commun.). This could imply a different source for these genes from the viral and eukaryal-like genes or that these regions exhibit properties different to the rest of the genome, such as replication or transcription.

Interestingly, the genomes of *Mimivirus* and phycodnaviruses are terminally redundant (i.e. carry duplicated regions) [3]. As has been suggested for vaccinia viruses [10], insertion of foreign genes into a resident gene in the terminally redundant region would leave the other resident copy (located at the other end) intact. This might favour viral derivatives in which inserted DNA occurs within one or the other end.

Detection of prokaryotic mobile elements in giant viruses

We also detected numerous mobile genetic elements previously thought to be specific to prokaryotes. These

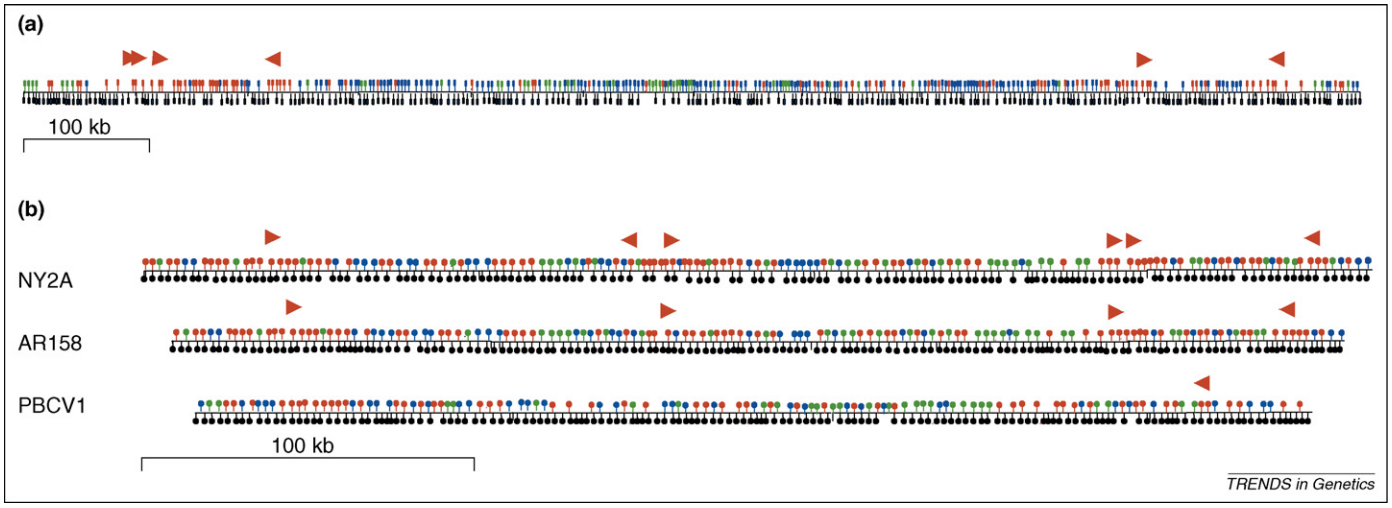

Figure 1. Genomic map of *Mimivirus* **(a)** and *Chlorella* **(b)** phycodnaviruses (NY2A, AR158 and PBCV1). The phylogenetic affinities of each gene were determined using BLASTP against a nonredundant database with an exclusion threshold of $E < 10^{-5}$ [37]. If the query produced genes from different kingdoms within the first ten hits of a BLAST search, the evolutionary status of the gene was analyzed further by individual phylogenetic analysis. Tree building was carried out using the NJ (neighbour-joining) programme of the MUST package [38]. The putative phylogenetic origins of these genes are indicated with the following colours: red, bacterial-type genes; blue, eukaryotic genes; green, NCLDV genes; and black, orphan genes. The orphan genes are placed on the underside of the genomic map. The positions of the IS*607* elements are indicated by a red arrow. The *Mimivirus* (1.2 Mb) and phycodnaviruses (300–400 kb) are not to the same scale. The intervals under the genomic map represent 100 kb.

include apparently intact insertion sequences (ISs), believed to be major agents of lateral gene transfer in prokaryotes [11], homing endonucleases and inteins (Box 1). Except for two copies of an element belonging to the IS*4* family in the genome of the brown alga virus, *Ectocarpus siliculosus virus 1* (EsV-1), we could not identify ISs in any other eukaryotic viral genome in the public databases.

The ISs in *Mimivirus* and phycodnaviruses belong to the bacterial and archaeal IS*607* family (Box 1). Several are intact and not only carry the appropriate open reading frames but also include defined IS ends. Based on the known sequences of other members of this family (ISFinder; http://www-is.biotoul.fr), these full-length copies would presumably be active in the appropriate host. The IS elements co-localize with stretches of other bacterial-like genes. They do not, however, occur at the boundaries between bacterial and other genes but are embedded within stretches of contiguous bacterial-like (or orphan) genes, suggesting that they have been inherited along with bacterial DNA.

The presence and location of such ISs lends strong support to the idea that they were inherited from bacterial genomes along with the other bacterial-like genes. Furthermore, giant viruses also display relatively large numbers of prokaryotic-like homing endonucleases and inteins (Box 1), further strengthening the notion that giant viruses are chimeras of genes from disparate sources.

What is the source of the bacterial-like genes and how are they acquired?

Incorporation of bacterial genes would require two conditions: (i) an 'ecological' niche bringing viral and bacterial DNA into close contact and (ii) a recombination mechanism to drive bacterial gene accretion.

Bacterial and viral DNA could occupy the same cell compartment

The *Mimivirus* host, *Acanthamoeba polyphaga*, frequently carries and releases diverse human bacterial pathogens such as *Legionella* [12], phagocytoses bacteria, a major food source, and also possesses bacteriolytic activities [13]. In addition, many free-living amoebae harbour a range of pathogenic and nonpathogenic bacteria, often as obligate symbionts that cannot be cultured axenically [14]. For example, infecting *Legionella* bacteria, incorporated by 'coiling phagocytosis', can be found replicating in the host cytoplasm some hours after infection [15]. Hence there are documented mechanisms that place the bacteria in the same cell compartment as replicating viral DNA.

Phycodnaviruses infect photosynthetic algae of the *Chlorella* genus, which are not known to ingest bacteria. However, chlorellae live as symbionts within *Paramecium bursaria* (the term 'farming' has been used in this context). *Paramecium bursaria*, a common ciliate in freshwater environments, also grazes on bacteria and, under certain starvation conditions, can digest the resident chlorellae as a source of nutrients. Paramecia harbour bacteria both in phagosomes (digestive vacuoles) and, individually, within the cytoplasm [16]. This would also facilitate access of infecting phycodnaviruses to a bacterial gene pool in the cytoplasm of their host. However, present knowledge of the replication cycle of these viruses is limited. Viral infection has only been reproduced on isolated free-growing chlorellae and not in the context of their symbiont host. Circumstantial evidence suggests that replication commences in the chlorella nucleus although intact nuclear membranes are not required [17]. *Paramecium bursaria* can be cured of its symbiotic chlorellae. On reinfection, the chlorellae, like bacterial prey, are incorporated into digestive vacuoles, perhaps together with the bacteria. Most, but not all, are destroyed [18] and the survivors become symbionts. We note that similar levels of bacterial-like genes have not been reported in NCLDVs that infect metazoans, possibly because they do not come into such intimate contact with bacterial DNA in these hosts.

Thus, although we are unable to provide direct molecular evidence, it seems plausible that viral and

Box 1. Prokaryotic-like mobile genetic elements

Insertion sequences

We identified several IS*607* family elements: IS*vMimi_1* and IS*vMimi_2* in the *Mimivirus* genome, and IS*vPBCV_1*, IS*vNY2A_1*, IS*vNY2A_2* and IS*vAR158_1* in three closely related *Chlorella* phycodnaviruses. Like prokaryotic IS*607* elements, these contain two *orf* genes: *tnpA* which encodes the transposase required for mobility, and *tnpB*, whose function remains unknown [34]. The transposases of these viral elements display 50–66% amino acid identity between themselves and 45–55% identity with bacterial IS*607* elements. Although there is only one *Mimivirus* genome sequence, the three *Chlorella* phycodnavirus genomes (AR158, NY2A and PBCV-1) enable genomic comparisons: the unique copy of IS*vPBCV_1* present in all three genomes is conserved at the same location; two copies of IS*vNY2A_2* are present at the same sites in the genomes of NY2A and AR158; whereas IS*vAR158_1* and IS*vNY2A_1* are specific to the genomes of AR158 (one copy) and NY2A (three copies), respectively. Whole-genome alignment of NY2A versus PBCV-1 using a dotplot diagram (Figure I) showed that the NY2A ISs, which are not conserved in PBCV-1, are associated with nonconserved regions between the two genomes.

HNH homing endonucleases

These are a diverse collection of proteins characterized by an HNH motif, generally encoded by genes within mobile, self-splicing introns, that promote the movement of the DNA sequences that encode them by making a site-specific double-strand break at a target site. The gene is then copied into the double-strand break [35]. Phycodnaviruses carry seven to 14 homing endonucleases of the HNH group, and two HNH homing endonucleases are present in *Mimivirus*. Compared with other homing endonuclease groups, members of the HNH group are found principally in bacteriophage genomes that infect both Gram-positive and Gram-negative bacteria [35]. Some are present within bacterial-like sequences (Online Supplementary Material Tables S1, S2).

Genomic comparison in phycodnaviruses (Figure I) shows extensive variation in the number and location of HNH genes suggesting that they have or had the capacity to excise, insert and disseminate or that they have been recruited in repeated acquisitions. The two *Mimivirus* HNH endonuclease genes flank a unique gene typically found in bacterial prophages suggesting horizontal transfer from a bacteriophage.

Inteins

These are defined segments of proteins able to excise and to rejoin the flanking peptides. They occur in the DNA polymerase gene of *Mimivirus* [36] and in the phycodnavirus NY2A ribonucleotide reductase gene (*NrdA*) and an open reading frame (*orf*) of unidentified function (Id: R508). NY2A inteins are not conserved in the homologous genes of AR158 and PBCV-1. At least two-thirds of known inteins are found in prokaryotes and those found in the eukarya are restricted to yeasts – see 'Inteins – Protein Introns' (http://bioinformatics.weizmann.ac.il/~pietro/inteins/) and InBase (http://www.neb.com/neb/inteins.html).

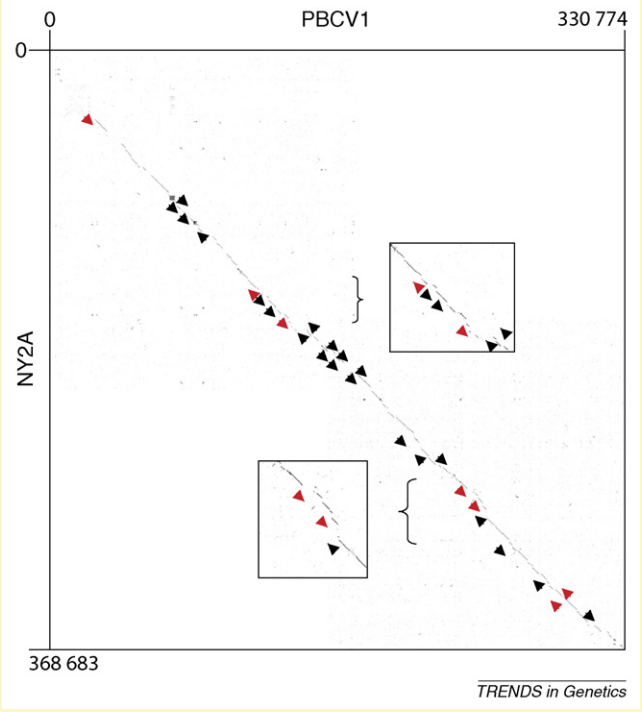

Figure I. Genome dotplot of *Chlorella* phycodnavirus NY2A (vertical axis) versus PBCV1 (horizontal axis). Intergenome comparisons were performed by pairwise BLASTN alignments without filtering and with an exclusion threshold of $E < 10^{-5}$. Each ORF of a genome was blasted against all ORFs of the other genome and the sequences having the best high-scoring segment pair were used to plot the alignment diagram. Red and black arrows map IS*607* elements and mobile endonucleases, respectively. Insets show a more detailed picture.

bacterial genes can at least transitorily occupy the same cell compartment.

Replication- or recombination-driven gene acquisition

How might such 'foreign' genes be acquired? One possible explanation is suggested by the replication mode of the related poxvirus [10]. In these viruses, replication and recombination are intimately coupled [19–21] in a similar way to bacteriophage T4. For T4, this style of replication has been called recombination-primed replication [22] (Box 2). Interestingly, recent studies have indicated that up to 10% of the genomes of some members of the T4 group of phages could be of bacterial origin [23].

Replicating poxviruses undergo such increased levels of homologous recombination that genetic linkage is apparently lost at distances greater than 500 bp [19,20]. Purified *Vaccinia virus* DNA polymerase is capable of promoting ATP-independent intermolecular single-strand transfer *in vitro* and requires polymerase 3′→5′ exonucleolytic (proofreading) activity [24] and DNA homology. The reaction is most efficient using linear DNA substrates and gives rise to linear concatamers. Recombination of linear molecules requires little sequence homology and will occur between molecules containing as little as 12 bp of end-sequence identity [25]. Indeed, substrates with no apparent end-sequence homology can still generate recombinants at frequencies about tenfold above background [10]. Similarly increased recombination frequencies (~2%) have been observed for the phycodnaviruses themselves [26]. Although an extensive mechanistic analysis remains to be undertaken, it is known that the ends of phycodnaviruses PBCV-1 and CVK1 are long inverted repeats. Moreover, the CVK1 terminal inverted repeats carry a nick that could serve as an initiation point for replication and strand invasion [27] (Box 2, Figure I). Indeed, recent comparison of 45 epidemiologically varied poxviruses has led to the conclusion that recombination occurs frequently at the ends of the viral genome [28].

An alternative, or additional, possibility for promoting strand exchange would be that observed in the λ red and *E. coli* RecE–RecT systems [29] or in herpesvirus recombination [30]. These systems involve two proteins:

Box 2. Models for bacterial-like gene acquisition

The mechanism, illustrated in Figure I, is based on the phage T4 model but can be extended to *Vaccinia* and other giant viruses [10,27]. Strand exchange is initiated by a single-strand annealing (SSA) mechanism [25] in which a short 5′→3′ single strand, exposed by 3′→5′-exonuclease-mediated degradation of the complementary strand (Figure Ia), invades receptor genomes carrying a single-strand nick or double-strand breaks (Figure Ib). The strand exchanges generate ends that can then serve as primers for the replicative polymerase (Figure Ic). The process can be coupled to (Figure Ig) or independent of (Figure Id) replication (Figure Ig). Depending on strand polarities and cleavages, genes can be incorporated either as short regions (or patches; Figure If and Figure Ij) or as longer regions (or splices; Figure Ie or Figure Ih). The pathway resulting in splicing would be the most relevant to the present discussion.

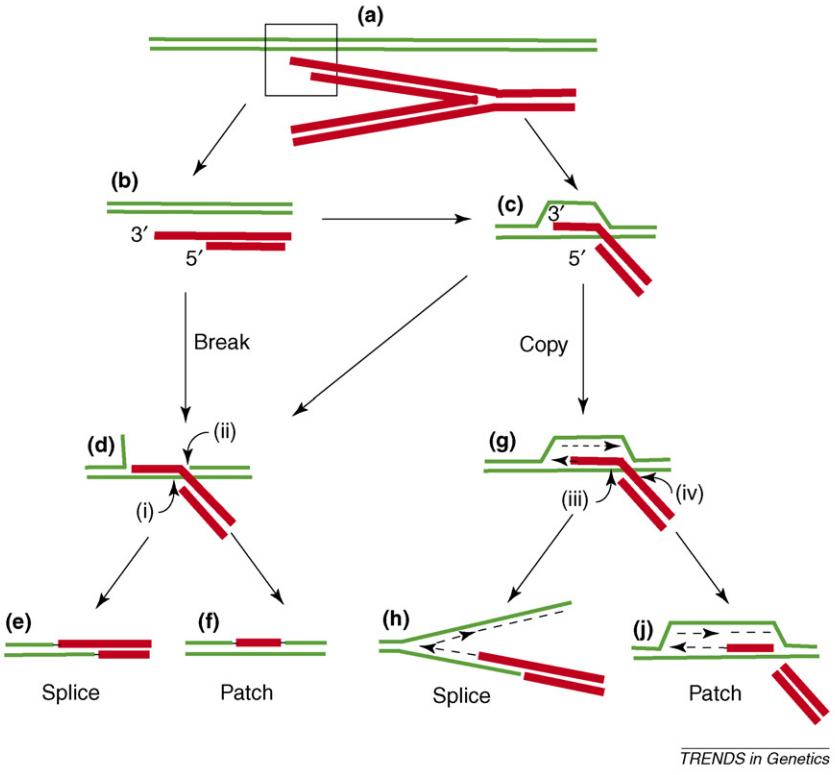

Figure I. T4 replication–recombination illustrating how regions of foreign DNA could be introduced by a splice or patch mechanism. Target DNA is shown in green, invading DNA in red. **(a)** A replicating donor molecule and a nonreplicating target DNA. **(b)** Resection of the invading DNA end leaving a 3′ overhang. **(c)** Strand invasion of the 3′ overhang creates a D-loop. **(d)** Strand invasion at a single-strand break. **(e)** Cleavage of target strand (i) generates a splice junction. **(f)** Cleavage of invading strand (ii) generates a patch. **(g)** The 3′ invading strand acts as a replication primer for continuous strand synthesis, which then enables assembly of a replication fork and initiation of discontinuous strand replication. **(h)** Cleavage (iii) generates a splice junction. **(j)** Cleavage at (iv) generates a patch. The dashed arrows indicate replication. This figure is inspired by several diagrams compiled by the late Gisela Mosig.

a DNA-binding protein (red β for phage λ; RecT for *E. coli*; ICP8 for herpesvirus) and an exonuclease (red α for λ; RecE for *E. coli*; UL12 for herpesvirus). As in the T4 model (Box 2), a DNA end is resected to form a single-strand overhang (either 5′ or 3′ overhangs function *in vitro* [30]) and strand invasion occurs in a second step. In contrast to the T4 model, it is unnecessary to invoke replication (Box 2).

Although we think that a recombination-primed replication model is attractive as an explanation for the entrapment of bacterial-like genes, other potential mechanisms exist. For example, it has been suggested that *Vaccinia virus* topoisomerase IB might be involved in promoting recombination, and this enzyme possesses several biochemical properties that support this view [31,32]. Moreover, a topoisomerase IB gene from *Mimivirus*, which produces a biochemically similar enzyme, has recently been described [33]. However, although a topoisomerase II has been identified, to our knowledge no topoisomerase IB gene has yet been identified in the phycodnaviruses.

Concluding remarks

Our analysis suggests that most bacterial-like genes found in giant viruses were recently and independently acquired in *Mimivirus* and phycodnavirus lineages and might also indicate that the resident bacterial DNA has been acquired several times during successive acquisition events. Comparison of the closely related genomes of phycodnaviruses suggests that acquisition of bacterial genes and mobile genetic elements is a continuous process. These findings contradict the idea that they could be inherited from a common ancestor. We propose, based on results obtained with bacteriophage T4 and poxviruses, that it is the extraordinarily intimate coupling between replication and recombination that leads to exceptionally high recombination frequencies facilitating promiscuous incorporation of foreign genes into the viral genome. The presence of repeated sequences such as ISs could subsequently favour this process by supplying the homologies to drive intermolecular recombination. Their presence would also be expected to lead to further rearrangements in the host viral genome during subsequent viral replication cycles.

We propose that the host (or a host symbiont) of a giant virus provides both the necessary source of bacterial DNA (from grazing of bacteria) and the cell compartment that enables this DNA to encounter actively replicating viruses. These systems therefore would provide an extensive interface for (one-way) genetic exchange between prokaryotes and eukaryotes. Bacterial gene acquisition by viruses was first observed in the form of transducing bacteriophages. Our hypothesis could be tested by co-infecting bacteria, genetically marked with a selectable marker (e.g. a drug resistance gene), with the viral host. The sibling giant viruses that had acquired such bacterial genes might then, under suitable conditions, be selected following reinfection of a secondary host.

Note added in proof

The genomes of NY-2A and AR158 have now been published [39].

Acknowledgements

The phycodnavirus and *Mimivirus* genomic sequences were retrieved respectively at Greengene (http://greengene.uml.edu/database/database.html) and GiantVirus.org (http://www.giantvirus.org/data_analysis.html). We thank the following members of the Chandler laboratory: B. Tong-Hoang, P. Rouseau, G. Duval-Valentin, C. Guynet, N. Pouget, E. Gueguen and (former member) Mimi Bétermier, for fruitful discussion; D. Sherratt, D. Lane, J-P. Claverys, H.M. Krisch for reading the article and an anonymous referee for providing additional ideas and analysis. J. Besson provided invaluable information on the natural history of *Paramecium bursaria*. Funding was provided by the CNRS and by European contract: LSHM-CT-2005–019023. J.F. was supported by the CNRS and by the European contract.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2006.11.002.

References

1 Iyer, L.M. *et al.* (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* 117, 156–184
2 Koonin, E.V. (2005) Virology: Gulliver among the Lilliputians. *Curr. Biol.* 15, R167–R169
3 Suhre, K. (2005) Gene and genome duplication in *Acanthamoeba polyphaga Mimivirus*. *J. Virol.* 79, 14095–14101
4 Van Etten, J.L. (2003) Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.* 37, 153–195
5 Raoult, D. *et al.* (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350
6 Moreira, D. and Lopez-Garcia, P. (2005) Comment on "The 1.2-megabase genome sequence of Mimivirus". *Science* 308, 1114
7 Hendrix, R.W. *et al.* (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2192–2197
8 Claverie, J.M. *et al.* (2006) Mimivirus and the emerging concept of "giant" virus. *Virus Res.* 117, 133–144
9 Tillier, E.R. and Collins, R.A. (2000) The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. *J. Mol. Evol.* 50, 249–257
10 Yao, X.D. and Evans, D.H. (2001) Effects of DNA structure and homology length on vaccinia virus recombination. *J. Virol.* 75, 6923–6932
11 Frost, L.S. *et al.* (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732
12 Berger, P. *et al.* (2006) Ameba-associated microorganisms and diagnosis of nosocomial pneumonia. *Emerg. Infect. Dis.* 12, 248–255
13 Weekers, P.H. *et al.* (1995) Bacteriolytic activities of the free-living soil amoebae, *Acanthamoeba castellanii*, *Acanthamoeba polyphaga* and *Hartmannella vermiformis*. *Antonie Van Leeuwenhoek* 68, 237–243
14 Molmeret, M. *et al.* (2005) Amoebae as training grounds for intracellular bacterial pathogens. *Appl. Environ. Microbiol.* 71, 20–28
15 Molmeret, M. *et al.* (2004) Disruption of the phagosomal membrane and egress of *Legionella pneumophila* into the cytoplasm during the last stages of intracellular infection of macrophages and *Acanthamoeba polyphaga*. *Infect. Immun.* 72, 4040–4051
16 Beier, C.L. *et al.* (2002) The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the Rickettsiales (Alphaproteobacteria) and to *Francisella tularensis* (Gammaproteobacteria). *Appl. Environ. Microbiol.* 68, 6043–6050
17 Van Etten, J.L. *et al.* (2002) Phycodnaviridae – large DNA algal viruses. *Arch. Virol.* 147, 1479–1516
18 Kodama, Y. and Fujishima, M. (2005) Symbiotic *Chlorella* sp. of the ciliate *Paramecium bursaria* do not prevent acidification and lysosomal fusion of host digestive vacuoles during infection. *Protoplasma* 225, 191–203
19 Ball, L.A. (1987) High-frequency homologous recombination in vaccinia virus DNA. *J. Virol.* 61, 1788–1795
20 Evans, D.H. *et al.* (1988) High levels of genetic recombination among cotransfected plasmid DNAs in poxvirus-infected mammalian cells. *J. Virol.* 62, 367–375
21 Merchlinsky, M. (1989) Intramolecular homologous recombination in cells infected with temperature-sensitive mutants of vaccinia virus. *J. Virol.* 63, 2030–2035
22 Mosig, G. *et al.* (2001) Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8306–8311
23 Filee, J. *et al.* (2006) A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. *Mol. Biol. Evol.* 23, 1688–1696
24 Willer, D.O. *et al.* (1999) Vaccinia virus DNA polymerase promotes DNA pairing and strand-transfer reactions. *Virology* 257, 511–523
25 Willer, D.O. *et al.* (2000) *In vitro* concatemer formation catalyzed by vaccinia virus DNA polymerase. *Virology* 278, 562–569
26 Tessman, I. (1985) Genetic recombination of the DNA plant virus PBCV-1 in a chlorella-like alga. *Virology* 145, 319–322
27 Yamada, T. and Higashiyama, T. (1993) Characterization of the terminal inverted repeats and their neighboring tandem repeats in the *Chlorella* CVK1 virus genome. *Mol. Gen. Genet.* 241, 554–563
28 Esposito, J.J. *et al.* (2006) Genome sequence diversity and clues to the evolution of variola (smallpox) virus. *Science* 313, 807–812
29 Muyrers, J.P. *et al.* (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev.* 14, 1971–1982
30 Reuven, N.B. *et al.* (2004) Catalysis of strand exchange by the HSV-1 UL12 and ICP8 proteins: potent ICP8 recombinase activity is revealed upon resection of dsDNA substrate by nuclease. *J. Mol. Biol.* 342, 57–71
31 Sekiguchi, J. and Shuman, S. (1997) Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol. Cell* 1, 89–97
32 Cheng, C. and Shuman, S. (2000) Recombinogenic flap ligation pathway for intrinsic repair of topoisomerase IB-induced double-strand breaks. *Mol. Cell. Biol.* 20, 8059–8068
33 Benarroch, D. *et al.* (2006) Characterization of mimivirus DNA topoisomerase IB suggests horizontal gene transfer between eukaryal viruses and bacteria. *J. Virol.* 80, 314–321
34 Ton-Hoang, B. *et al.* (2005) Transposition of ISHp608, member of an unusual family of bacterial insertion sequences. *EMBO J.* 24, 3325–3338
35 Guhan, N. and Muniyappa, K. (2003) Structural and functional characteristics of homing endonucleases. *Crit. Rev. Biochem. Mol. Biol.* 38, 199–248
36 Ogata, H. *et al.* (2005) A new example of viral intein in Mimivirus. *Virol. J.* 2, 8
37 Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
38 Philippe, H. (1993) MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res.* 21, 5264–5272
39 Fitzgerald, L.A. *et al.* (2006) Sequence and annotation of the 369-kb NY-2A ad the 345-kb AR158 virusts that infect *Chlorella* NC64A. *Virology*, DOI: 10.1016/j.virol.2006.08.034

0168-9525/$ – see front matter
doi:10.1016/j.tig.2006.11.002