

1999

CAN LARGE dsDNA-CONTAINING VIRUSES PROVIDE INFORMATION ABOUT THE MINIMAL GENOME SIZE REQUIRED TO SUPPORT LIFE?

James L. Van Etten

University of Nebraska - Lincoln, jvanetten1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/vanetten>

 Part of the [Genetics and Genomics Commons](#), [Plant Pathology Commons](#), and the [Viruses Commons](#)

Van Etten, James L., "CAN LARGE dsDNA-CONTAINING VIRUSES PROVIDE INFORMATION ABOUT THE MINIMAL GENOME SIZE REQUIRED TO SUPPORT LIFE?" (1999). *James Van Etten Publications*. 3.

<https://digitalcommons.unl.edu/vanetten/3>

This Article is brought to you for free and open access by the Plant Pathology Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in James Van Etten Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

CAN LARGE dsDNA-CONTAINING VIRUSES PROVIDE INFORMATION ABOUT THE MINIMAL GENOME SIZE REQUIRED TO SUPPORT LIFE?

James L. Van Etten
Department of Plant Pathology
University of Nebraska at Lincoln

Abstract

The genomes of a few viruses, such as *Bacillus megaterium* phage G (670 kb) and the chlorella viruses (330 to 380 kb), are larger than the predicted minimal genome size required to support life (ca. 320 kb). A comparison of the 256 proteins predicted to be required for life with the putative 376 proteins encoded by chlorella virus PBCV-1, as well as those encoded by other large viruses, indicates that viruses lack many of these "essential" genes. Consequently, it is unlikely that viruses will aid in determining the minimal number and types of genes required for life. However, viruses may provide information on the minimal genome size required for life because the average size of genes from some viruses is smaller than those from free-living organisms. This smaller gene size is the result of three characteristics of virus genes: (1) virus genes usually have little intragenic space between them or, in some cases, genes overlap; (2) some virus-encoded enzymes are smaller than their counterparts from free-living organisms; and (3) introns occur rarely, if at all, in some viruses.

Introduction

Two recent estimates of the minimum genome size required to support life arrived at similar values. (1) The effect of 79 random mutations on the colony-forming ability of *Bacillus subtilis* resulted in the conclusion that a genome of 318 kb could support life (Itaya, 1995). Assuming 1.25 kb of DNA per gene (Fraser et al., 1995), this amount of DNA would encode 254 proteins. (2) A comparison of the genes encoded by *Mycoplasma genitalium* and *Haemophilus influenzae* led Mushegian and Koonin (1996) to suggest that as few as 256 genes are necessary for life. Using the same 1.25 kb gene size, the minimum self-sufficient life-form would have a 320 kb genome. Interestingly, these estimates are smaller than the genomes of some viruses (Table 1). Bacteriophage G, which infects *Bacillus megaterium*, has a genome of about 670 kb (Hutson et al., 1995); phycodnaviruses that infect chlorella-like green algae have 330 to 380 kb genomes (Rohozinski et al., 1989; Yamada et al., 1991); and some insect poxviruses have genomes as large as 365 kb (Langridge and Roberts, 1977). Other large, dsDNA-containing viruses, such as herpesviruses, African swine fever virus (ASFV), coliphage T4, baculoviruses, and iridoviruses, have genomes ranging from 100 to 235 kb (see Table 1). However, except for the common property of having large dsDNA genomes, these viruses differ significantly from one another in such characteristics as particle morphology, genome structure, and the intracellular site of replication. For example, poxviruses, herpesviruses, and baculoviruses have an external lipid envelope, whereas iridoviruses and phycodnaviruses have an internal lipid component. Baculovirus genomes are circular, iridoviruses and phage T4 have linear circular permuted genomes with terminal redundancy, and the linear genomes of herpesviruses have sequences from both termini that are repeated internally in an inverted form. The phycodnaviruses, poxviruses, and ASFV have linear genomes with covalently closed hairpin ends. Finally, herpesviruses and baculoviruses primarily replicate in the nucleus, whereas

Table 1 Representative Large dsDNA Viruses

Virus ^a	Virus Family	Host	Genome Size (bp)	Minimum No. of Codons ^b	No. of Genes	Average Length of Gene (Bases)	Reference
Phage G	Myoviridae	<i>Bacillus megaterium</i>	~670,000	—	—	—	Hutson et al., 1995
PBCV-1	Phycodnaviridae	Chlorella NC64A	330,742	65	376 ^c	880	Li et al., 1997
MsEPV	Poxviridae	Grasshopper	236,120 (222,120) ^d	60	267 (257) ^d	884 (864) ^d	Afonso et al., 1998
MCV	Poxviridae	Human	190,289 (180,889) ^e	60	182 (180) ^e	1,046 (1,005) ^e	Senkevich et al., 1996
ASFV	Unclassified	Swine	170,101 (166,613) ^f	60	151	1,127 (1,103) ^f	Yanez et al., 1995
Coliphage T4	Myoviridae	<i>E. coli</i>	168,800	29	288 ^g	586	Kutter et al., 1994
HSV-2	Herpesviridae	Human	154,746	—	74 ^h	2,091	Dolan et al., 1998
AcNPV	Baculoviridae	Insects	133,894	50	154	890	Ayres et al., 1994
LCDV	Iridoviridae	Flounder	102,653	40	110	933	Tidona and Darai, 1997

^aG, Giant; PBCV-1, *Paramecium bursaria chlorella* virus 1; MsEPV, *Melanoplus sanguinipes* entomopoxvirus; MCV, *Moluscum contagiosum* virus; ASFV, African swine fever virus; HSV-2, Herpes simplex virus type 2; AcNPV, *Autographa californica* multinucleocapsid nuclear polyhedroses virus; LCDV, lymphocystis disease virus.

^bMinimum number of codons used by the authors to calculate an open reading frame (ORF).

^cFour of the genes are diploid.

^dMsEPV has a 7 kb inverted repeat at each terminus. This 14 kb encodes 10 small ORFs (60 to 155 codons). Removal of 14 kb and 10 ORFs from the calculations produces the smaller genome size (in parentheses).

^eMCV has a 4.7 kb inverted repeat at each terminus. This 9.4 kb encodes two 488 codon ORFs. Removal of 9.4 kb and 2 ORFs from the calculations produces the smaller genome size (in parentheses).

^fASFV has a 2134 bp inverted repeat at each terminus. The most terminal 1744 bp at each end do not encode an ORF and thus 3488 bp were removed from the calculations, which leads to the smaller genome size (in parentheses).

^gThis includes 161 genes known to encode proteins and 127 suspected of encoding proteins (Gisela Mosig, personal communication).

^hHSV-2 has 473 met-initiated ORFs of 50 codons or longer of which 74 are known to be functional genes. If some of the additional 399 ORFs prove to be protein encoding, the average length of a herpesvirus gene would decrease substantially.

the entire life cycle of the poxviruses occurs in the cytoplasm. Iridoviruses and phycodnaviruses initiate replication in the nucleus, but capsids are assembled and DNA is packaged in the cytoplasm.

With the exception of bacteriophage G, the genome of at least one representative of each of these dsDNA-containing viruses has been sequenced, and the number of putative genes encoded by the viruses are listed in Table 1. Because the 330,742 bp genome of the phycodnavirus PBCV-1 is the largest virus genome sequenced to date (Lu et al., 1995, 1996; Li et al., 1995, 1997; Kutish et al., 1996), it will be used to illustrate the organization and diversity of genes that can be encoded by a large dsDNA-containing virus. The PBCV-1 genome encodes 701 open reading frames (ORFs), defined as continuous stretches of DNA that translate into a polypeptide initiated by an ATG translation start

codon, and extending 65 or more codons. The 701 ORFs have been divided into 376, mostly non-overlapping, ORFs (major ORFs), which are predicted to encode proteins, and 325 short ORFs, which are probably non-protein encoding. Four PBCV-1 ORFs reside in the 2.2 kb inverted terminal repeat region of the PBCV-1 genome and consequently are present twice in the PBCV-1 genome (Strasser et al., 1991; Lu et al., 1995). The 376 PBCV-1 ORFs are evenly distributed along the genome and, with one exception, there is little intergenic space between them. The exception is a 1788-bp non-protein coding sequence near the center of the genome. This region, which has numerous stop codons in all reading frames, does code for ten tRNA genes. The middle 900 bp of this intergenic region also has some characteristics of a "CpG island" (Antequera and Bird, 1993). To put the coding capacity of the PBCV-1 genome in perspective, the 580-kb genome of the smallest self-replicating organism, *Mycoplasma genitalium* encodes about 470 genes (Fraser et al., 1995).

Computer analyses of the predicted products of the 376 PBCV-1 major ORFs indicate that about 40% of the ORFs resemble proteins in the databases, including many interesting and unexpected proteins. Some PBCV-1 encoded proteins resemble those of bacteria and phages, such as DNA restriction endonucleases and methyltransferases. However, other PBCV-1 encoded proteins resemble those of eukaryotic organisms and their viruses, such as translocation elongation factor-3, RNA guanylttransferase, and two proliferating cell nuclear antigens. The PBCV-1 genome is thus a mosaic of prokaryotic- and eukaryotic-like genes, suggesting considerable gene exchange in nature during the evolution of these viruses.

This gene diversity undoubtedly reflects the natural history of the chlorella viruses. The viruses are ubiquitous in freshwater collected worldwide, and titers as high as 4×10^4 infectious viruses per ml of native water have been obtained (Van Etten et al., 1985; Yamada et al., 1991). The only known hosts for these viruses are chlorella-like green algae, which normally live as hereditary endosymbionts in some isolates of the ciliate, *Paramecium bursaria*. In the symbiotic unit, algae are enclosed individually in perialgal vacuoles and are surrounded by a host-derived membrane (Reisser, 1992). The endosymbiotic chlorella are resistant to virus infection and are only infected when they are outside the paramecium (Van Etten et al., 1991).

Because of the large size of the PBCV-1 genome, it is not surprising that many of the predicted 376 PBCV-1 genes have not been found in other viral genomes. Box 1 lists some of the PBCV-1 encoded ORFs that match proteins in the databases and, in a few cases, indicate if a gene is transcribed early (E) or late (L) during virus replication. The functionality of some PBCV-1 encoded proteins has been established by either complementation of mutants and/or assaying recombinant protein for enzyme activity. (These proteins are indicated with an asterisk in Box 1.) Twenty-nine of the PBCV-1 ORFs resemble one or more other PBCV-1 ORFs suggesting that they might be either gene families or gene duplications. Sixteen families have 2 members, 8 families have 3 members, 3 families have 6 members, and 2 families have 8 members.

Even if some of the suspected 376 PBCV-1 protein-encoding genes turn out to be non-coding, it is clear that PBCV-1 encodes more genes than the minimum number predicted to be necessary to support life. Comparing the genes that Mushegian and Koonin (1996) proposed were essential to support life with the PBCV-1 encoded genes indicates that the virus lacks many of these genes, including a RNA polymerase, a complete protein synthesizing system, and an energy-generating system. Consequently, PBCV-1 depends on the algal host to fulfill these essential functions.

A comparison of the genes encoded by the other large dsDNA-containing viruses listed in Table 1 with those encoded by PBCV-1 indicates that a few genes are present in all of the viruses, e.g., each of the viruses encodes a DNA polymerase gene. However, there are more differences in the genes encoded by these viruses than similarities, which reflects the different life-styles of the viruses. Like PBCV-1,

Box 1 Putative ORFs Encoded by *Chlorella* Virus PBCV-1^a

DNA Replication & Repair

- (E) A185R DNA polymerase
- (E) A544R* DNA ligase
- (E) A583L* DNA topoisomerase II
- A193L PCNA
- A574L PCNA
- A153R Helicase
- A241R Helicase
- A548L Helicase
- (E) A50L* T4 endonuclease V
- A39L CyclinA/cdk associated protein
- A638R Endonuclease

Nucleotide Metabolism

- (E) A169R* Aspartate transcarbamylase
- A476R Ribo. reductase (small subunit)
- A629R Ribo. reductase (large subunit)
- A427L Thioredoxin
- A438L Glutaredoxin
- A551L dUTP pyrophosphatase
- A596R dCMP deaminase
- A416R dG/dA kinase
- A363R Phosphohydrolase
- A392R ATPase
- A674R *Dicty* Thy protein

Transcription

- A107L RNA transcription factor TFIIIB
- A125L RNA transcription factor TFIIIS
- A166R Exonuclease
- A422R Endonuclease
- (E) A103R* RNA guanyltransferase
- A464R RNase III

Translation

- (E,L) A666L Translation elongation factor-3
- A85R Prolyl 4-hydroxylase alpha-subunit
- A105L Ubiquitin C-terminal hydrolase
- A448L Protein disulphide isomerase
- A623L Ubiquitin-like fusion protein
- 10 tRNAs

Cell Wall Degrading

- (E) A181R* Chitinase
- (L) A260R* Endochitinase
- (L) A292L* Chitosanase
- A94L β -1,3 glucanase

DNA Restriction/Modification

- A251R* Adenine DNA methylase (M.CviAII)
- A252R* Restriction endonuclease (R.CviAII)
- A252R* Restriction endonuclease (R.CviAII)
- (E) A517L* Cytosine DNA methylase (M.CviAIII)
- (L) A530R* Cytosine DNA methylase (M.CviAIV)
- (E) A581R* Adenine DNA methylase (M.CviAI)
- (E) A579L* Restriction endonuclease (R.CviAI)
- A683L Cytosine DNA methylase (M.CviAV)

Sugar and Lipid Manipulation

- (L) A64R Galactosyl transferase
- (E) A98R* Hyaluronan synthase
- (E) A100R* Glucosamine synthase
- A114R Fucosyltransferase
- (E) A118R GDP-D-mannose dehydratase
- A222R Cellulose synthase
- A295L Fucose synthase
- (E) A473L Cellulose synthase
- (E) A609L* UDP-glucose dehydrogenase
- A49L Glycerophosphoryl diesterase
- A53R 2-hydroxyacid dehydrogenase
- A271L Lysophospholipase

Phosphorylation/dephosphorylation

- A34R Protein kinase
- (L) A248R* Phosphorylase B kinase
- A277L Ser/Thr protein kinase
- A278L Ser/Thr protein kinase
- A282L Ser/Thr protein kinase
- A289L Ser/Thr protein kinase
- A305L Tyr phosphatase
- A614L Protein kinase
- A617R Tyr-protein kinase

Miscellaneous

- A207R* Ornithine decarboxylase
- A217L Monoamine oxidase
- (L) A237R* Homospermidine synthase
- A78R β -alanine synthase
- A245R Cu/Zn-superoxide dismutase
- A284L* Amidase
- A465R Yeast ERVI protein
- A598L Histidine decarboxylase
- A250R K⁺ ion channel protein
- A625R Transposase

^a E and L refer to early and late genes, respectively. An asterisk means that the gene encodes a functional enzyme as determined either by complementation or by enzyme activity of a recombinant protein.

each of these viruses rely on their host cells for such basic functions as energy generation, protein synthesis, and amino acid biosynthesis. The net result is that it seems unlikely that examining virus genes will aid in determining the minimal number and types of genes required to support life.

On the other hand, viruses may provide useful information about the minimum genome size required for the genes to support life. In Table 1, we have calculated the average length of a virus gene by dividing the genome size by the number of putative genes. Except for herpesvirus HSV 2, the size of the average virus gene varied from 586 nucleotides for coliphage T4 to 1,127 nucleotides for ASFV, with the average gene size for five of the viruses being less than 1 kb. The sizes are even smaller if one removes the non- or sparsely-coding regions in the virus genomes before making the calculations. For example, the two poxviruses MsEPV and MCV, as well as ASFV, have inverted terminal repeat regions that either are non-coding or only encode a few genes. Eliminating these non-coding regions from the calculations reduces the size of the average MsEPV gene from 884 nucleotides to 864 nucleotides, the MCV gene from 1,046 nucleotides to 1,005 nucleotides and ASFV from 1,127 nucleotides to 1,103 nucleotides (Table 1).

Similar calculations made on nine Eubacteria and three Archaea indicate that the average length of Eubacteria protein-encoding genes ranges from 1,023 nucleotides for *Aquifex aeolicus* to 1,234 nucleotides for *Mycoplasma genitalium* (Doolittle, 1998). The predicted average length of the three archaea is slightly smaller—895, 943, and 961 nucleotides for *Archaeoglobus fulgidus*, *Methanococcus thermoautotrophicum*, and *M. jannaschii*, respectively. Thus, depending on the virus and bacterium being compared, the average functional virus gene is 10 to 50% smaller than the average bacterial gene. This conclusion depends on the assumption that at least the majority of the predicted virus genes, in fact, encode proteins.

The apparent smaller size of genes from large dsDNA viruses can be attributed to three factors. (1) Typically, virus genomes have little intergenic space and, in some cases, genes overlap. This tight packaging of genes does not prevent gene regulation, however, as virus genes are typically expressed early or late in the replication cycle. The 376 major ORFs in chlorella virus PBCV-1 are evenly distributed along the genome, and 85% are separated by less than 200 nucleotides. Likewise, 85% of the 151 putative genes in ASFV are also separated by less than 200 nucleotides (Yanez et al., 1995). The genes in phage T4 are even more tightly packed (Kutter et al., 1994). Consequently, transcription start and stop signals plus the regulatory regions for at least some virus genes are extremely short, or they are located in the coding region of adjacent genes.

(2) Some virus-encoded proteins are smaller than those from free-living organisms and may approach the minimum size required for enzyme activity. Examples include the PBCV-1 encoded 298 amino acid residue ATP-dependent DNA ligase, the 1,061 amino acid residue type II DNA topoisomerase, and the 372 amino acid residue ornithine decarboxylase. Each of these virus-encoded proteins has the expected enzyme activity. ATP-dependent DNA ligases range in size from the 268 amino acid residue enzyme from *Haemophilus influenzae* (Cheng and Shuman, 1997) to the 1,070 amino acid residue enzyme from *Xenopus laevis* (Lepetit et al., 1996). The PBCV-1 enzyme is the second smallest ATP-dependent ligase in the databases. The PBCV-1 encoded type II DNA topoisomerase is about 130 amino acids smaller than the next smallest type II topoisomerase in the databases, which is encoded by virus ASFV (Garcia-Beato et al., 1992). The PBCV-1 encoded ornithine decarboxylase is about 90 amino acids smaller than the next smallest ornithine decarboxylase in the databases. Likewise, the large subunit of ribonucleotide reductase from the baculovirus *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus (OpMNPV) is 150 to 200 amino acids smaller than its counterpart from most organisms (Ahrens et al., 1997).

(3) Even though introns were first discovered in adenoviruses (Berget et al., 1977; Chow et al.,

1977), the genes of many large DNA-containing viruses either lack introns, e.g., poxviruses, baculoviruses, iridoviruses, and ASFV, or only have a few short introns, e.g., phycodnaviruses. An absence of introns obviously contributes to the smaller size of virus genes.

To summarize, it is unlikely that studying viruses will reveal useful information about the minimum number and types of genes required to support life. However, the finding that, on average, virus genes can be 10 to 50% smaller than those from bacteria indicate that the minimum genome size required to support life may be smaller than previously thought.

Acknowledgments

I thank Les Lane, Mike Nelson, Myron Brakke, and Mike Graves for their comments on this manuscript and Dan Rock and Gisela Mosig for the information on MsEPV virus and coliphage T4, respectively.

References

- Afonso, C.L., E.R. Tulman, Z. Lu, E. Oma, G.F. Kutish, and D.L. Rock. 1998. The genome of *Melanoplus sanguinipes* entomopoxvirus. *J. Virol.* (in press).
- Ahrens, C.H., R.L.Q. Russell, C.J. Funk, J.T. Evans, S.H. Harwood, and G.F. Rohrmann. 1997. The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **229**:381-399.
- Antequera, F., and A. Bird. 1993. CpG Islands. Pp. 169-185 in *DNA Methylation: Molecular Biology and Biological Significance*, P.J. Jost and P.H. Saluz (eds.), Basel, Switzerland: Birkhauser Verlag.
- Ayres, M.D., S.C. Howard, J. Kuzio, M. Lopez-Ferber, and R.D. Possee. 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202**:586-605.
- Berget, S.M., C. Moore, and P.A. Sharp. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* **74**:3171-3175.
- Cheng, C., and S. Shuman. 1997. Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*. *Nucleic Acids Res.* **25**:1369-1374.
- Chow, L., R. Gilinas, T. Broker, and R. Roberts. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**:1-8.
- Dolan, A., F.E. Jamieson, C. Cunningham, B.C. Barnett, and D.J. McGeoch. 1998. The genome sequence of herpes simplex virus type 2. *J. Virol.* **72**:2010-2021.
- Doolittle, R.F. 1998. Microbial genomes opened up. *Nature* **392**:339-342.
- Fraser, C.M., J.D. Gocayne, O. White, M.D. Adams, R.A. Clayton, R.D. Fleischmann, C.J. Bult, A.R. Kerlavage, G. Sutton, J.M. Kelley, J.L. Fritchman, J.F. Weidman, K.V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T.R. Utterback, D.M. Saudek, C.A. Phillips, J.M. Merrick, J.F. Tomb, B.A. Dougherty, K.F. Bott, P.C. Hu, T.S. Lucier, S.N. Peterson, H.O. Smith, C.A. Hutchison, and J.C. Venter. 1995. The minimal gene-complement of *Mycoplasma genitalium*. *Science* **270**:397-403.
- Garcia-Beato, R., J.M.P. Freije, C. Lopez-Otin, R. Blasco, E. Vinuela, and M.L. Salas. 1992. A gene homologous to topoisomerase II in African swine fever virus. *Virology* **188**:938-947.
- Hutson, M.S., G. Holzwarth, T. Duke, and J.L. Viovy. 1995. Two-dimensional motion of DNA bands during 120° pulsed-field gel electrophoresis. I. Effect of molecular weight. *Biopolymers* **35**:297-306.
- Itaya, M. 1995. An estimation of minimal genome size required for life. *FEBS Lett.* **362**:257-260.
- Kutish, G.F., Y. Li, Z. Lu, M. Furuta, D.L. Rock, and J.L. Van Etten. 1996. Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: Map positions 182 to 258. *Virology* **223**:303-317.
- Kutter, E., T. Stidham, B. Guttman, E. Kutter, D. Batts, S. Peterson, T. Djavakhishvili, F. Arisaka, V. Mesyanzhinov, W. Ruger, and G. Mosig. 1994. Genomic map of bacteriophage T4. Pp. 491-519 in *Molecular Biology of Bacteriophage T4*, J.D. Karam (ed). Washington DC: American Society for Microbiology.
- Langridge, W.H.R., and D.W. Roberts. 1977. Molecular weight of DNA from four entomopoxviruses determined by electron microscopy. *J. Virol.* **21**:301-308.

- Lepetit, D., P. Thiebaud, S. Aoufouchi, C. Prigent, R. Guesne, and N. Theze. 1996. The cloning and characterization of a cDNA encoding *Xenopus levis* DNA ligase I. *Gene* **172**:273-277.
- Li, Y., Z. Lu, D.E. Burbank, G.F. Kutish, D.L. Rock, and J.L. Van Etten. 1995. Analysis of 43 kb of the chlorella virus PBCV-1 330-kb genome: Map position 45 to 88. *Virology* **212**:134-150.
- Li, Y., Z. Lu, L. Sun, S. Ropp, G.F. Kutish, D.L. Rock, and J.L. Van Etten. 1997. Analysis of 74 kb of DNA located at the right end of the chlorella virus PBCV-1 330-kb genome. *Virology* **237**:360-377.
- Lu, Z., Y. Li, Q. Que, G.F. Kutish, D.L. Rock, and J.L. Van Etten. 1996. Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: Map positions 88 to 182. *Virology* **216**:102-123.
- Lu, Z., Y. Li, Y. Zhang, G.F. Kutish, D.L. Rock, and J.L. Van Etten. 1995. Analysis of 45 kb of DNA located at the left end of the chlorella virus PBCV-1 genome. *Virology* **206**:339-352.
- Mushegian, A.R., and E.V. Koonin. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* **93**:10268-10273.
- Reisser, W. (ed). 1992. *Algae and Symbioses*. Bristol, UK: Biopress.
- Rohozinski, J., L. Girton, and J.L. Van Etten. 1989. Chlorella viruses contain linear nonpermuted double-stranded DNA genomes with covalently closed hairpin ends. *Virology* **168**:363-369.
- Senkevich, T.G., J.J. Bugert, J.R. Sisler, E.V. Koonin, G. Darai, and B. Moss. 1996. Genome sequence of a human tumorigenic poxvirus: Prediction of specific host response evasion genes. *Science* **273**:813-816.
- Strasser, P., Y. Zhang, J. Rohozinski, and J.L. Van Etten. 1991. The termini of the chlorella virus PBCV-1 genome are identical 2.2-kbp inverted repeats. *Virology* **180**:763-769.
- Tidona, C.A., and G. Darai. 1997. The complete DNA sequence of lymphocystis disease virus. *Virology* **230**:207-216.
- Van Etten, J.L., D.E. Burbank, A.M. Schuster, and R.H. Meints. 1985. Lytic viruses infecting a chlorella-like alga. *Virology* **140**:135-143.
- Van Etten, J.L., L.C. Lane, and R.H. Meints. 1991. Viruses and viruslike particles of eukaryotic algae. *Microbiol. Rev.* **55**:586-620.
- Yamada, T., T. Higashiyama, and T. Fukuda. 1991. Screening of natural waters for viruses which infect chlorella cells. *Appl. Environ. Microbiol.* **57**:3433-3437.
- Yanez, R.J., J.M. Rodriguez, M.L. Nogal, L. Yuste, C. Enriquez, J.F. Rodriguez, and E. Vinuela. 1995. Analysis of the complete nucleotide sequence of African swine fever virus. *Virology* **208**:249-278.