J Mol Evol (2000) 50:82–92
DOI: 10.1007/s002399910009

JOURNAL OF MOLECULAR EVOLUTION

Intron Conservation in a UV-Specific DNA Repair Gene Encoded by Chlorella Viruses

Liangwu Sun,[1,*] Yu Li,[1,†] Amanda K. McCullough,[2] Thomas G. Wood,[2] R. Stephen Lloyd,[2] Byron Adams,[1,‡] James R. Gurnon,[1] James L. Van Etten[1]

[1]Department of Plant Pathology, University of Nebraska-Lincoln, NE 68583-0722, USA
[2]Sealy Center for Molecular Science, Department of Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston, TX 7555-1071, USA

Received: 13 May 1999 / Accepted: 20 August 1999

Abstract. Large dsDNA-containing chlorella viruses encode a pyrimidine dimer-specific glycosylase (PDG) that initiates repair of UV-induced pyrimidine dimers. The PDG enzyme is a homologue of the bacteriophage T4-encoded endonuclease V. The *pdg* gene was cloned and sequenced from 42 chlorella viruses isolated over a 12-year period from diverse geographic regions. Surprisingly, the *pdg* gene from 15 of these 42 viruses contain a 98-nucleotide intron that is 100% conserved among the viruses and another 4 viruses contain an 81-nucleotide intron, in the same position, that is nearly 100% identical (one virus differed by one base). In contrast, the nucleotides in the *pdg* coding regions (exons) from the intron-containing viruses are 84 to 100% identical. The introns in the *pdg* gene have 5′-AG/GTATGT and 3′-TTGCAG/AA splice site sequences which are characteristic of nuclear-located, spliceosomal processed pre-mRNA introns. The 100% identity of the 98-nucleotide intron sequence in the 15 viruses and the near-perfect identity of an 81-nucleotide intron sequence in another 4 viruses imply strong selective pressure to maintain the DNA sequence of the intron when it is in the *pdg* gene. However, the ability of intron-plus and intron-minus viruses to repair UV-damaged DNA in the dark was nearly identical. These findings contradict the widely accepted dogma that intron sequences are more variable than exon sequences.

Key words: Pyrimidine dimer-specific glycosylase — DNA repair — Intron — dsDNA virus — Chlorella viruses — Phycodnaviridae

Introduction

Large (150 to 190 nm in diameter) icosahedral, plaque-forming, dsDNA-containing viruses that infect certain isolates of unicellular, eukaryotic chlorella-like green algae are common in freshwater collected throughout the world (Van Etten et al. 1985a, b; Schuster et al. 1986; Zhang et al. 1988; Yamada et al. 1991). Virions of the prototype chlorella virus, PBCV-1, contain at least 50 proteins and a lipid component located inside the outer glycoprotein capsid (Skrdla et al. 1984; Wang et al. 1993). The PBCV-1 genome, a linear, nonpermuted dsDNA molecule with covalently closed hairpin ends has been sequenced (Lu et al. 1995; 1996; Li et al. 1995, 1997; Kutish et al. 1996). The 330,740-bp genome encodes 702 open reading frames (ORF) 65 codons or

This paper has been assigned Journal Series No. 12562, Agricultural Research Division, University of Nebraska

**Current address:* Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA
[†]*Current address:* Centers for Disease Control, MS-C03, Atlanta, GA 30333, USA
[‡]*Current address:* Department of Nematology, University of California, Davis, CA 95616-8668, USA
Correspondence to: J.L. Van Etten; *e-mail:* jvanetten@unlnotes.unl.edu

larger, of which 377 are predicted to encode proteins (Li et al. 1997).

Perhaps one of the more interesting PBCV-1-encoded proteins is one that resembles the bacteriophage T4 *denV* gene product (Furuta et al. 1997). The *denV* gene encodes a well-characterized, pyrimidine dimer-specific glycosylase, apyrimidine lyase, called endonuclease V (T4-PDG), that initiates repair of UV-induced pyrimidine dimers in DNA (Yasuda and Sekiguchi 1970; Lloyd and Linn 1993; Lloyd 1999). Although T4-PDG-like enzyme activity has been discovered in several microorganisms, including *Micrococcus luteus* (Grafstrom et al. 1982; Piersen et al. 1995; Shiota and Nakayama, 1997) and *Saccharomyces cerevisiae* (Hamilton et al. 1992), with the exception of one enzyme from *M. luteus* (Shiota and Nakayama 1997), these enzymes differ from T4-PDG in size and amino acid sequence. The discovery of a PBCV-1-encoded homologue to T4-PDG (41% amino acid identity) led to structural and functional comparisons between the two enzymes. The PBCV-1 enzyme is a cis–syn and trans–syn-II cyclobutane pyrimidine dimer glycosylase (PBCV-1-PDG) (McCullough et al. 1998), whereas the T4 enzyme cleaves only the cis–syn cyclobutane pyrimidine isomer. In addition, PBCV-1-PDG has a stronger electrostatic attraction for DNA than the T4 enzyme; i.e., PBCV-1-PDG is more processive than T4-PDG.

The discovery of functional differences between the PBCV-1 and the T4 enzymes prompted us to characterize T4-PDG homologues from 41 other chlorella viruses from diverse geographic regions. As described in this paper, *pdg* genes from 15 viruses contain a 98-nucleotide intron and another 4 viruses contain an 81-nucleotide intron. These introns interrupt the coding region at identical positions in the *pdg* gene. Surprisingly, the nucleotide sequence of the 98-nucleotide intron is 100% conserved, regardless of the origin of the viruses. Moreover, three of the four 81-nucleotide introns were identical; the fourth differed by only one nucleotide. In contrast, the exon nucleotide sequences of the *pdg* genes were less conserved among the viruses.

Table 1. Chlorella viruses used in this study[a]

Virus	Virus isolated from	Date collected	Class	Size of PCR product with PBCV-1 primers (bp)
NE-8D	Nebraska, USA	Sept. 1984	1	447
Nyb-1	New York, USA	Aug. 1984	1	447
CA-4B	California, USA	Nov. 1984	1	447
AL-1A	Alabama, USA	Oct. 1984	2	545
NY-2C	New York, USA	Aug. 1984	2	545
NC-1D	North Carolina, USA	Oct. 1983	2	545
PBCV-1	North Carolina, USA	1981	3	447
NC-1C	North Carolina, USA	Oct. 1983	3	545
CA-1A	California, USA	Nov. 1984	4	545
CA-2A	California, USA	Nov. 1984	4	447
IL-2A	Illinois, USA	Oct. 1983	4	545
IL-2B	Illinois, USA	Oct. 1983	4	447
IL-3A	Illinois, USA	Oct. 1983	4	545
IL-3D	Illinois, USA	Oct. 1983	4	447
SC-1A	South Carolina, USA	Oct. 1983	5	545
SC-1B	South Carolina, USA	Oct. 1983	5	545
NC-1A	North Carolina, USA	Oct. 1983	6	N[b]
NE-8A	Nebraska, USA	Sept. 1984	7	447
AL-2C	Alabama, USA	Oct. 1984	7	447
MA-1E	Massachusetts, USA	Aug. 1984	7	545
NY-2F	New York, USA	Aug. 1984	7	N
CA-1D	California, USA	Nov. 1984	7	545
NC-1B	North Carolina, USA	Oct. 1983	7	N
Nys-1	New York, USA	Aug. 1984	8	N
IL-5-2s1	Illinois, USA	May 1984	9	447
AL-2A	Alabama, USA	Oct. 1984	9	545
MA-1D	Massachusetts, USA	Aug. 1984	9	447
NY-2B	New York, USA	Aug. 1984	9	N
CA-4A	California, USA	Nov. 1984	10	545
NY-2A	New York, USA	Aug. 1984	11	N
XZ-3A	Xuzhou, China	Mar. 1987	12	447
SH-6A	Shanghai, China	Mar. 1987	13	N
BJ-2C	Beijing, China	Mar. 1987	13	N
XZ-6E	Xuzhou, China	Mar. 1987	14	N
XZ-4C	Xuzhou, China	Mar. 1987	15	447
XZ-5C	Xuzhou, China	Mar. 1987	16	N
XZ-4A	Xuzhou, China	Mar. 1987	16	447
IS-10	Israel	Aug. 1996		N
CH-57	Baoding, China	Aug. 1997		N
AN69C	Canberra, Australia	Mar. 1995		545
AR158	Buenos Aires, Argentina	Aug. 1997		N
AR93-2	Buenos Aires, Argentina	Aug. 1997		545

[a] The First 37 viruses have been separated into 16 classes as indicated (Van Etten et al., 1991). The last five viruses were isolated recently and were included to increase the geographic diversity of the viruses.

[b] No PCR product was obtained with virus PBCV-1 primers.

Materials and Methods

Viruses and Host Strains

The geographic sources of the 42 (including PBCV-1) chlorella viruses and the year they were isolated are listed in Table 1. The growth of the host alga, *Chlorella* strain NC64A, on MBBM medium, plaque assay, production of the viruses, and isolation of virus DNAs have been described (Van Etten et al. 1981, 1983a, b).

Polymerase Chain Reaction (PCR)

Single plaques from chlorella viruses were transferred with sterile toothpicks to 200 μl of 50 m*M* Tris–HCl, pH 7.5. After soaking for 2 h, 50-μl aliquots were boiled for 10 min and the samples were used as templates for PCR.

Primers corresponding to the 5′ and 3′ ends of the PBCV-1 *pdg* gene were used to amplify the *pdg* gene from many of the viruses. The sequence of the PBCV-1 5′ primer was 5′-ATC**GGATCC**CATATGACACGTGTGAATCTCGTACC; the sequence of the PBCV-1 3′ primer was 5′-AAT**GGATCC**TTAATTATTGCTGGTTTTAGC. The sequence of NY-2B 3′ primer was 5′-AAT**GGATCC**TTAATTATCATTATGATTAG. All primers contained a *Bam*HI restriction site **(in boldface).** The PCR products were either sequenced directly or digested with *Bam*HI and cloned into the *Bam*HI site of pBluescript IIKS(−) (Stratagene, La Jolla, CA, USA) before sequencing.

DNA was amplified with Vent DNA polymerase (New England BioLabs, Beverly, MA, USA), in 100-μl reactions which contained 1 pg of virus DNA, a 100 p*M* (concentration) of each primer, a 0.2 m*M* concentration each of dATP, dCTP, dGTP, and dTTP, and 10 μl of 10 × ThermoPol buffer [100 m*M* KCl, 200 m*M* Tris-HCl (pH 8.8 at 25°C), 100 m*M* $(NH_4)_2SO_4$, 20 m*M* $MgSO_4$, 1% Triton X-100], by 30 cycles of heating and cooling: 1 min at 94°C for denaturation, 2 min at 55°C for annealing, and 2 min at 72°C for elongation.

Complementary DNA (cDNA) and PCR

Total RNA was isolated from chlorella virus-infected cells using Trizol reagent (GIBCO-BRL, Gaithersberg, MD, USA). cDNA synthesis of the PDG mRNA and amplification of DNA were described previously (Grabherr et al. 1992). The *pdg* 3′ primer of PBCV-1 (for viruses containing the 98-nucleotide intron) or NY-2B (for viruses containing the 81-nucleotide intron) was used to synthesize cDNA. The forward primer, the PBCV-1 *pdg* gene 5′ primer, and the reverse primer, the PBCV-1 (for viruses containing the 98-nucleotide intron) or NY-2B (for viruses containing the 81-nucleotide intron) *pdg* gene 3′ primer, were used to amplify the *pdg* DNA from the cDNA templates.

DNA Sequencing

PCR products and cloned PCR products were sequenced from both strands by the procedure of Sanger et al. (1977), as modified by Tabor and Richardson (1987), using a Version 2.0 Sequenase kit from Amersham Life Science (Arlington Heights, IL, USA). Some DNA fragments were also sequenced at the University of Nebraska Center for Biotechnology DNA sequencing core facility. Purified viral genomic DNAs were also used as templates for DNA sequencing using oligonucleotide primers that hybridized within the *pdg* gene; these DNA sequences allowed both ends of the *pdg* genes to be determined. These reactions contained 3.2 pmol of oligonucleotide primer and reaction regents from an AmpliTaq Dye-Terminator cycle sequencing kit (Applied Biosystems, Inc., Foster City, CA, USA). Mixtures were heated through 35 cycles at 96°C for 30 s, 50°C for 15 s, and 60°C for 4 min. The reaction products were purified by centrifugation (Centri-Sep columns; Princeton Separations, Inc., Adelphia, NJ, USA) and dried *in vacuo* (Savant Instruments, Inc., Hicksville, NY, USA). The residue was resuspended in a 5:1 mixture of deionized formamide and 50 m*M* EDTA, heated at 95°C for 2 min, and then cooled on ice. Samples were loaded onto 5% polyacrylamide gels to resolve the DNA sequence by electrophoresis (Model 373A automated DNA sequencer; Applied Biosystems, Inc.).

Other Procedures

DNA probes were labeled with a random primers DNA labeling kit (GIBCO-BRL). For dot blots, denatured virus DNAs were applied to nylon membranes (Micron Separation Inc., Westborough, MA, USA) as described (Ross et al. 1989).

UV radiation sources and the procedures used to assay the viruses for repair of UV damage in the dark were identical to those used previously (Furuta et al. 1997). DNA, RNA, and protein sequences were analyzed with the University of Wisconsin Genetics Computer Group package of programs (Genetics Computer Group, 1994). The number of synonymous and nonsynonymous amino acid substitutions was estimated using the DIVERGE command in GCG version 9.0, which employs the modified method of Li (1993) and Pamilo and Bianchi (1993). Computer programs used to construct the phylogenetic trees are described in the results section. The 42 *pdg* genes are deposited in GenBank under Accession numbers AF128127 to AF128168.

Results

The pdg *Gene Is Widespread in the Chlorella Viruses*

Twenty-nine of the 42 viruses used in this study were isolated from water samples collected in different regions of the United States in 1983–1984 (Van Etten et al. 1985a, b; Schuster et al. 1986) and seven were isolated from water collected in China in 1987 (Zhang et al. 1988). Including PBCV-1, which was isolated in 1981, several criteria were used to group these 37 viruses into 16 classes (Table 1) (Van Etten et al. 1991). To increase the geographic diversity of the viruses, we included an additional virus from China, two viruses from Argentina, and one each from Israel and Australia; these five viruses were isolated in 1995–1996.

The PBCV-1 *pdg* gene probe hybridized to dot blots of DNA from all the viruses (Furuta et al. 1997; L. Sun, unpublished results), indicating that the gene is widespread in the chlorella viruses. Using the DNA sequence of the PBCV-1 *pdg* gene as a guide to make oligonucleotide primers, DNA was amplified from 41 viruses by PCR. DNA from 28 of the viruses produced PCR products with these primers that were cloned and sequenced (Table 1). The size of the PCR products from 13 of these 28 viruses was identical to that produced with the PBCV-1 *pdg* gene, i.e., 447 nucleotides. However, the PCR products from the other 15 viruses, such as IL-3A and AL-1A, were larger, 545 nucleotides. The predicted translation products of the 13 DNA sequences that were the same size as the PBCV-1 gene resembled the PBCV-1-PDG sequence (Fig. 1). In fact, the predicted amino acid sequences from two viruses, XZ-3A and XZ-4A, were identical to PBCV-1-PDG; the other 11 PDG enzymes were 94 to 98% identical with the PBCV-1 enzyme (Fig. 1). In contrast, the DNA sequences from the other 15 viruses did not translate into inframe proteins. However, removal of a 98-nucleotide region (Fig. 2), from nucleotides 59 to 156, created proteins that resembled PBCV-1-PDG (Fig. 1). The predicted amino acid sequence of the proteins encoded by 4 of these viruses, AR93-2, NC-1C, SC-1A, and SC-1B, was 100% identical to PBCV-1-PDG; the other 11 viruses encoded proteins with 95 to 99% amino acid identity to the PBCV-1 enzyme (Fig. 1). As expected, synonymous substitutions occurred more frequently than nonsynonymous substitutions, and the proportion of substitution type showed only slight variation among all genes sampled (Table 2).

No PCR products were obtained with the PBCV-1 *pdg*-gene primers for 13 viruses, even though their DNAs hybridized to the *pdg*-gene probe. To clone the *pdg* genes from these 13 virus DNAs, DNA restriction fragments were identified by Southern blotting, eluted from gels and cloned into pBluescript II KS(−). Appropriate clones were sequenced and predicted amino acid

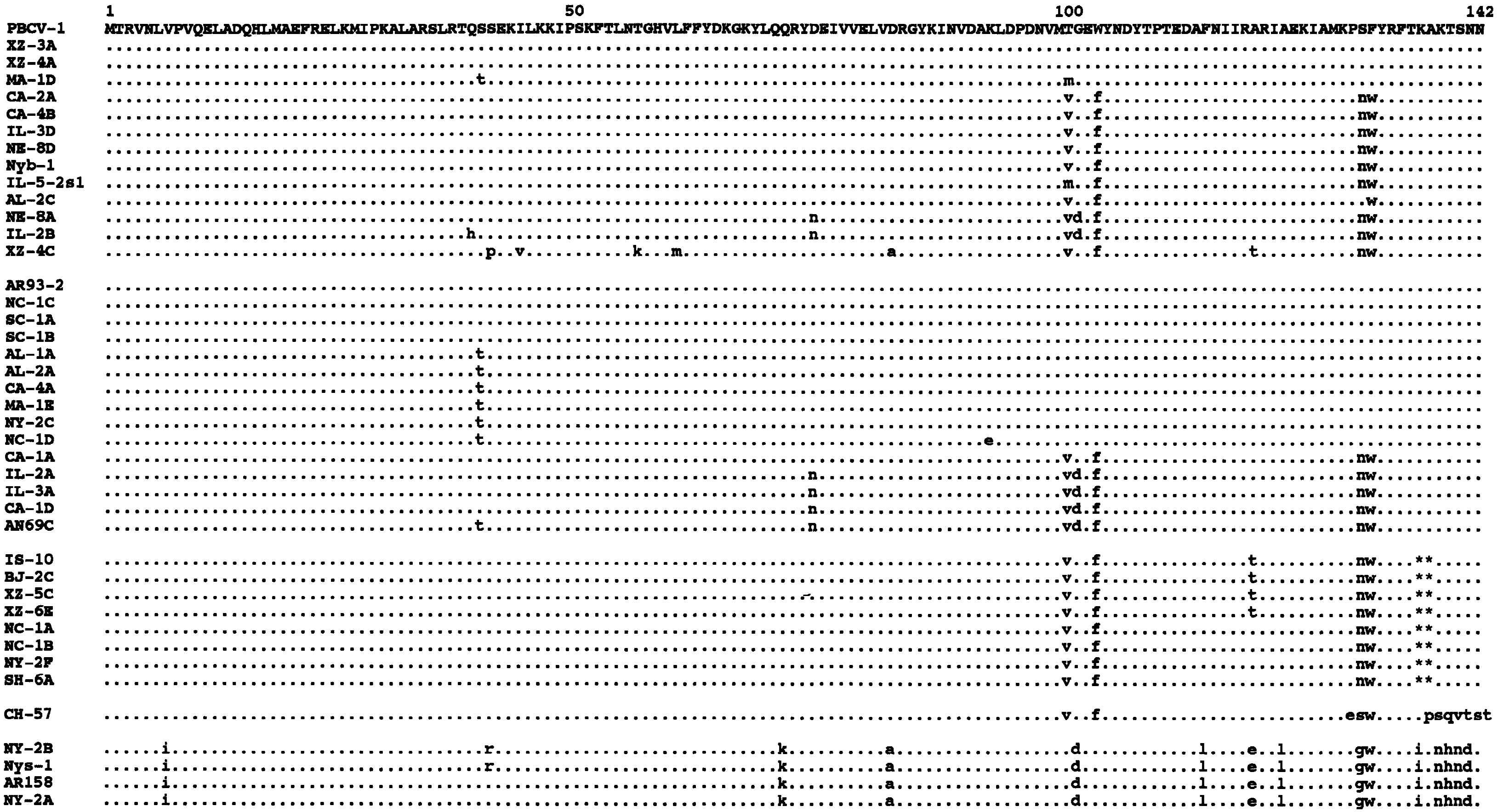

Fig. 1. The predicted amino acid sequences of 42 chlorella virus PDG enzymes. (●) No change in an amino acid. Amino acids that differ from the PBCV-1 enzyme are indicated in *lowercase.* (*) No equivalent amino acid in that location.

```
AR93-2  GTATGTAAATATTAATAATCACAAACTTAATAATTCATAATCACAAACTTAATTATATTCTTTATTTATTAAAATTTCGATTTGTAATGTTTTTGCAG
NC-1C   ..................................................................................................
SC-1A   ..................................................................................................
SC-1B   ..................................................................................................
AL-1A   ..................................................................................................
AL-2A   ..................................................................................................
CA-4A   ..................................................................................................
MA-1E   ..................................................................................................
NY-2C   ..................................................................................................
NC-1D   ..................................................................................................
CA-1A   ..................................................................................................
IL-2A   ..................................................................................................
IL-3A   ..................................................................................................
CA-1D   ..................................................................................................
AN69C   ..................................................................................................

NYs-1   GTATGTgAAcATTcAcAA----------------TCtTAATtAtAAtCTcAATTATAcTCTTTATTTATTAAAATTTCGATTgaTAATG-TTTTGCAG
AR158   ..................................................................................................
NY-2A   ..................................................................................................
NY-2B   ....a.............................................................................................
```

Fig. 2. The nucleotide sequences of 98-nucleotide introns (top 15) and 81-nucleotide introns (bottom 4) in some of the chlorella virus *pdg* genes. *Uppercase letters* are nucleotides that are conserved. A *dash* indicates that nucleotides are missing.

Table 2. Synonymous substitutions (K_s) and nonsynonymous substitutions (K_a) among representative PGD genes estimated using DIVERGE (GCG version 9.0)[a]

	AL.1A	AN.69C	AR.158	CH.57	1S.10	MA.1D	NE.8A	NY.2B	PBCV.1	XZ.4C
AL.1A	0.00	23.70	53.67	43.16	32.57	8.58	33.05	46.52	19.65	51.67
AN.69C	3.23	0.00	41.82	17.19	9.11	14.79	7.57	40.32	23.81	23.60
AR.158	5.83	6.45	0.00	59.98	58.19	43.71	54.82	11.59	57.84	63.99
CH.57	5.79	3.17	8.35	0.00	7.10	32.54	10.61	49.49	22.44	24.59
IS.10	4.08	2.49	7.71	1.82	0.00	22.84	1.46	49.90	13.75	13.94
MA.1D	0.37	2.83	6.73	5.38	3.66	0.00	23.53	37.64	7.49	39.00
NE.8A	4.16	0.95	7.68	3.62	1.53	3.75	0.00	55.78	14.43	14.19
NY.2B	6.54	7.17	0.65	9.14	8.45	7.44	8.40	0.00	51.04	73.21
PBCV.1	0.94	4.21	6.61	4.79	3.08	1.31	3.18	7.32	0.00	28.00
XZ.4C	6.04	4.65	9.41	5.63	2.73	5.62	3.60	9.31	5.03	0.00

[a] Upper triangle: synonymous substitutions per 100 synonymous sites. Lower triangle: nonsynonymous substitutions per 100 nonsynonymous sites.

sequences of these proteins are included in Fig. 1. The predicted amino acid sequence from 8 of the 13 viruses, IS-10, BJ-2C, XZ-5C, XZ-6E, NC-1A, NC-1B, NY-2F, and SH-6A, was two amino acids smaller than PBCV-1-PDG (Fig. 1). The protein encoded by virus CH-57 was one amino acid longer than PBCV-1-PDG. Furthermore, the last seven amino acids at the carboxyl end of the CH-57 protein differed from PBCV-1-PDG (Fig. 1). The remaining four viruses, NYs-1, NY-2A, NY-2B, and AR158, contained an apparent, “extra” 81-nucleotide insert in the putative *pdg* gene at the same position as the 98-nucleotide insert. Removal of this 81-nucleotide region (Fig. 2), from nucleotide 59 to nucleotide 139, created in frame ORFs that encoded proteins with 90 to 91% amino acid identity to PBCV-1-PDG (Fig. 1).

Comparisons of the *pdg* coding regions to a consensus *pdg* gene are shown in Fig. 3. Nucleotide sequences of the *pdg* genes from viruses XZ-3A and XZ-4A were identical to PBCV-1, whereas the *pdg* genes from the other viruses ranged from 84 to 98% identical to PBCV-1. Including all the viruses, 123 of the 420 to 426 nucleotides in the *pdg* coding region differed from the consensus *pdg* gene. Seventy-three, or 59%, of these 123 nucleotide differences occurred in the third position of the codon.

Some Chlorella Virus pdg *Genes Contain an Intron*

The apparent “extra” 81 nucleotides (4 viruses) and 98 nucleotides (15 viruses) in the *pdg* genes in 19 viruses have sequence characteristic of nuclear, spliceosomal-processed introns. Pre-mRNAs with spliceosomal-processed introns typically have three short consensus sequences: (i) the 5′ splice site is AG/**GU**RAGU in mammalian cells and AG/**GU**AUGU in yeast, (ii) the 3′ splice site is $\mathbf{(Y)_n}$ N**YAG**/G in mammalian cells and **Y**C**AG**/G in yeast, and (iii) a branch point sequence, which is more variable, but usually is UNCUR**A**C in mammalian cells and **UACUAAC** in yeast (Green 1991; Balvay et al. 1993; Lamond 1993). The “extra” 81 and 98 nucleotides in the *pdg* genes contain potential 5′ (AG/**GU**AUGU) and 3′ (UUUUUG**CAG**/A) splice site sequences and a putative branch point sequence (UC**A**C) with a lariat-forming adenine residue located either 55 or 56 nucleotides upstream of the 3′ splice site.

To determine if the 81 or 98 nucleotides in the *pdg* genes are removed from the final transcript, RNA was isolated from chlorella cells at 180 and 240 min after infection with virus NY-2B (81-nucleotide) or 60 and 90 min after infection with virus IL-3A (98-nucleotide); cDNAs were synthesized from pooled samples by

```
                                   111111111111111111222222222222222222222222223333333333333333333333333333333444444444444444444444444444
            1111222333344445 5 7779011112245666777799001112333444556667777889990000112234555566677888888899999000000111111111122222222223
           92689147034625891 9 1364356890888569256806891289258347032890127396780167062570234523436234567801457015679012345678901234567890

consensus  TGCCGAGTAATACGTCC ◊ CTGCTATGTCAACCTTTTCGTTACACGTTATATGGTCTGCTACTCGTGGTTTTCAGGTGGCGGCACTGCCAAATGGTCGCGAAGTAAAACCAGCAATAATTAA***

PBCV-1     ................. N .....................c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
XZ-3A      ................. N .....................c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
XZ-4A      ................. N .....................c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
MA-1D      ......aaggct.a.ta N ......a..............c.....c.........c.......a...aggctca..a.....c.......g.ttca............................
CA-2A      ................. N ..............c....a......................................................................................
CA-4B      ................. N ..............c....a......................................................................................
IL-3D      ................. N ..............c....a......................................................................................
NE-8D      ................. N ..............c....a......................................................................................
NYb-1      ................. N ..............c....a......................................................................................
IL-5-2s1   ................. N ...........g.................................a............................................................
AL-2C      ................. N ..............c.........................................................g.................................
NE-8A      ................. N ...........g..c....a......a.....................a.........................................................
IL-2B      ................. N .....t.....g..c....a......a.....................a.........................................................
XZ-4C      ..............c.. N ....c...c.g..aagcga.c..aca...g.cc.......a.t................a..............................................

AR93-2     tgccg.aaggct.a.ta L .....................c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
NC-1C      at.a..aaggct.a.ta L .....................c.....c.........c.......ac..aggctca..a.....c.....c.g.ttca............................
SC-1A      at.a..aaggct.a.ta L .....................c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
SC-1B      at.a..aaggct.a.ta L .....................c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
AL-1A      at.a..aaggct.a.ta L ......a....c.........c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
AL-2A      at.a..aaggct.a.ta L ......a..............c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
CA-4A      at.a..aaggct.a.ta L ......a..............c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
MA-1E      ......aaggct.a.ta L ......a..............c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
NY-2C      at.a..aaggct.a.ta L ......a..............c.....c.........c.......ac..aggctca..a.....c.......g.ttca............................
NC-1D      at.a..aaggct.a.ta L ......a..............c.....c.........c...g...ac..aggctca..a.....c.......g.ttca............................
CA-1A      at.a..aaggct.a.ta L ..............c....a......................................................................................
IL-2A      at.a..aaggct.a.ta L ...........g..c....a......a.....................a.........................................................
IL-3A      atta..aaggct.a.ta L ...........g..c....a......a.....................a.........................................................
CA-1D      at.a..aaggct.a.ta L ...........g..c....a......a.....................a.........................................................
AN69C      ......aaggct.a.ta L ......a....g..c....a......a.....................a.........................................................

IS-10      ................. N ...........g...............................................a.......................a..gt******............
BJ-2C      ................. N ...........g.........c.....................................a.......................a..gt******............
XZ-5C      ................. N ...........g...............................................a.......................a..gt******............
XZ-6E      ................. N ...........g...............................................a.......................a..gt******............
NC-1A      ................. N ...........g.......................................................................a..gt******............
NC-1B      ................. N ...........g.......................................................................a..gt******............
NY-2F      ................. N ...........g.......................................................................a..gt******............
SH-6A      ................. N ...........g.........c................cg...........................................a..gt******............

CH-57      ................. N ...........g.......................................................aga.tcc...aata.gcatctcaagtt.catcaacgtga

NY-2B      ..taacaaggctta.ta S tgat.g.acg...........cga....c.cccaact......ctacaacggc.g.aga.aatt.tca...ggg..caat.tc....tcat.atg...........
NYs-1      ...aacaaggctta.t. S tgat.g.acg...........cga....c.cccaact......ctacaacggc.g.aga.aatt.tca...ggg..caat.tc....tcat.atg...........
AR158      ..taacaaggctta.ta S tgat........t.gg.a.a.cga....c.cccaact......c.acaacggc.g.aga.aatt.tca...gga..caat.tc....tca..atg...........
NY-2A      ..taacaaggctta.ta S tgat........t.gg.a.a.cga....c.cccaact......c.acaacggc.g.aga.aatt.tca...gga..caat.tc....tca..atg...........
```

Fig. 3. Nucleotide differences in the *pdg* coding regions of 42 chlorella viruses. The *pdg* gene from the first 14 viruses lack an intron. The next 15 viruses encode a *pdg* gene that contains a 98-nucleotide intron. The *pdg* gene from the next eight viruses lacks both an intron and two codons near the 3′ end. The virus CH-57 *pdg* gene lacks an intron and has considerable diversity at the 3′ end. The *pdg* gene from the last four viruses contains an 81-nucleotide intron. The numbers above the nucleotides represent the position numbers of the nucleotides. Only nucleotides that differed from the concensus sequence (listed at the *top*) are included. (●) The nucleotide was the same as the consensus nucleotide. *Lowercase letters* indicate nucleotides that differ from the consensus sequence. (*) The nucleotide was deleted. (◊) The insertion location of the introns: N, no intron; L, a 98-nucleotide intron; S, an 81-nucleotide intron.

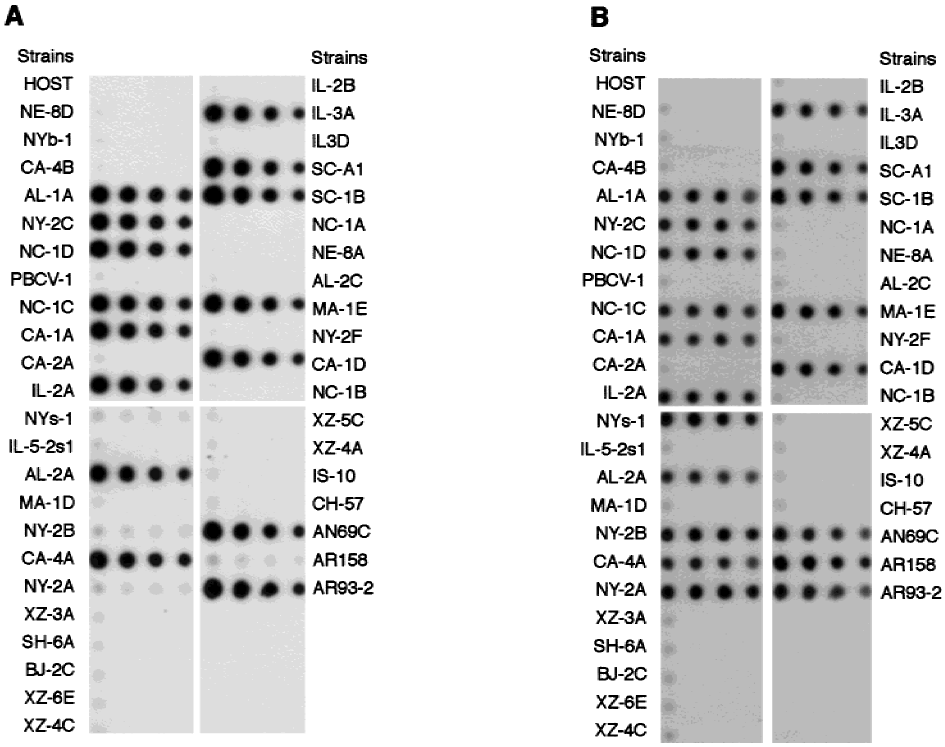

Fig. 4. Hybridization of the introns from the virus *pdg* gene to DNAs isolated from the host *Chlorella* NC64A and 42 chlorella viruses. The ^{32}P-labeled 98-nucleotide intron hybridized to the 15 chlorella virus DNAs that contained the 98-nucleotide intron and, to a much lesser-extent, to the 4 viruses with the 81-nucleotide intron (**A**). The ^{32}P-labeled 81-nucleotide intron hybridized to the 4 chlorella virus DNAs that contained the 81-nucleotide intron (NYs-1, NY-2B, NY-2A, and AR158) and also to the 15 chlorella virus DNAs that contained the 98-nucleotide intron (**B**).

reverse transcription and amplified by PCR. The cDNA products were identical in size to that of the PBCV-1 *pdg* gene product, which lacks an intron. Comparison of the cDNA sequences to either the NY-2B or the IL-3A genomic DNAs established that 81 or 98 nucleotides, respectively, are removed from the mRNAs and that the splice sites are identical to those predicted above. Therefore, the “extra” 81 and 98 nucleotides in the *pdg* gene from the 19 chlorella viruses are introns. The intron occurs after the first nucleotide in a Glu codon.

The 81- and 98-nucleotide introns resemble each other and the size difference may reflect a 16-nucleotide deletion or insertion (Fig. 2). Twelve (13 for virus NY-2B) of the remaining 81 nucleotides, or about 15%, differ between the 81- and the 98-nucleotide introns. Both introns are A+T rich, 78 and 83%, for the short and long introns, respectively. In contrast, the *pdg* coding region from all the viruses is 60% A+T, the same as the whole virus genome (Van Etten et al. 1985c). Both introns contained many internal TAA translational stop codons and neither intron encoded a significant ORF.

The following experiments indicate that the virus genomes contain only one copy of the intron and that the introns are located exclusively in the *pdg* gene. (i) DNAs from the 42 viruses (including PBCV-1) were hybridized with probes for both introns. The 98-nucleotide intron probe hybridized to the 15 virus genomic DNAs with the 98-nucleotide intron in their *pdg* genes and weakly with the 4 viruses that have the 81-nucleotide intron (Fig. 4A). The 81-nucleotide probe hybridized not only to the 4 viruses with the smaller intron, but also to the 15 virus DNAs with the longer intron (Fig. 4B). Densitometric measurements of the blots revealed that hybridization intensity for the virus DNAs were similar, indicating that all of the 19 viruses have the same number of introns, (ii) Southern blots of virus NY-2B- or IL-3A-DNAs cleaved with eight restriction endonucleases were hybridized with the 81-nucleotide intron or the 98-nucleotide intron, respectively. In each blot, only one restriction fragment hybridized to the probe, consistent with one intron per genome. (iii) No DNA sequence that resembles either intron occurs in the PBCV-1 genome or in the databases.

UV Sensitivity of Intron-Containing and Intron-Lacking Viruses

The almost-perfect identity of the 81-nucleotide and 98-nucleotide intron sequences for viruses that were isolated over a period of 12 years and separated by thousands of miles suggests selection pressure maintains the intron sequence in the *pdg* gene. One possibility is that introns improve repair of genomic UV damage. To test this hypothesis, we compared the UV sensitivity of intron-

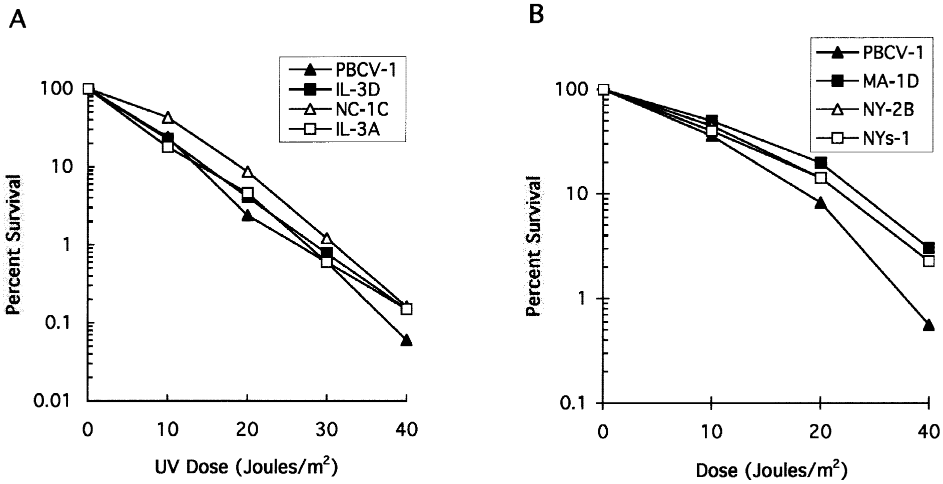

Fig. 5. Sensitivity of chlorella viruses to UV radiation. (**A**) Viruses PBCV-1 and IL-3D (no intron in the *pdg* gene) and NC-1C and IL-3A (98-nucleotide intron in the *pdg* gene). (**B**) Viruses PBCV-1 and MA-1D (no intron in the *pdg* gene) and NY-2B and NYs-1 (81-nucleotide intron in the *pdg* gene). There were no obvious differences in the survival of viruses that either contained (*open symbols*) or lacked (*filled symbols*) an intron in their *pdg* genes.

containing (NC-1C and IL-3A contain a 98-nucleotide intron and viruses NY-2B and NYs-1 contain an 81-nucleotide intron) and intron-lacking (PBCV-1, IL-3D, and MA-1D) viruses. The viruses were exposed to increasing doses of UV radiation, and the virus titer plates were incubated in the dark for 3 to 5 days. Survival rates of all viruses were similar (Fig. 5).

Phylogenetic Analyses

Evolutionary relationships among the 42 chlorella virus *pdg* genes were determined. Predicted amino acid sequences of the coding regions were aligned using the Pileup command in GCG, version 9.1 (Genetics Computer Group 1997), optimized manually, and then back-translated to nucleic acid sequences using DNA Stacks version 1.1.2 (Eernisse 1992). The last five codons could not be unambiguously aligned and were excluded from the analyses. Most-parsimonious trees and tests of alternative tree topologies were inferred using PAUP* version 4.0.0.d64 (kindly provided by D.L. Swofford). To estimate support for unambigously supported clades only, parsimony jackknifing (250,000 replicates) was employed (Farris et al. 1996). Maximum-likelihood trees were reconstructed using PUZZLE version 3.1 and assumed the HKY model of sequence substitution (Hasegawa et al. 1985). Nucleotide transition/transversion ratios and pyrimidine transition/purine transition ratios were estimated from the data set. Trees were visualized using Treeview version 1.4 (Page 1996). The distributions of introns were mapped onto the *pdg* gene phylogeny using MacClade version 3.05 (Maddison and Maddison 1992). The gene tree was rooted using the T4 *pdg* gene.

The maximum-likelihood topology (Fig. 6A) differed from the parsimony-jackknife tree (Fig. 6B), but only in the resolution of weakly supported nodes, and these differences were not significant (according to Kishino–Hasegawa tests of likelihoods and Templeton's nonparametric test of parsimony). However, topologies forcing nonmonophyly of the 81-nucleotide intron-containing genes, as well as strict monophyly of all 98-nucleotide intron-containing genes, were significantly worse deviations from the optimal solution. *Pdg* genes containing the 81-nucleotide intron formed a monophyletic clade comprising the sister lineage to the remaining *pdg* genes. The 98-nucleotide intron-containing genes were paraphyletic relative to the two clades of non-intron-containing genes.

Discussion

The amino acid sequences of the chlorella virus PDG enzymes did not vary as much as anticipated. For example, the PDG amino acid sequences from six viruses (AR93-2, NC-1C, SC-1A, SC-1B, XZ-3A, and XZ-4A) and from eight viruses (IS-10, NC-1A, NC-1B, NY-2F, BJ-2C, XZ-5C, XZ-6E, and SH-6A) were, respectively, 100 and 98% identical to PBCV-1-PDG. The amino acid sequence for the PDG enzyme from four viruses that contain an 81-nucleotide intron diverged most from the PBCV-1 enzyme, having 90 to 91% amino acid identity.

Two amino acids critical for T4-PDG catalysis are known from x-ray crystal (Morikawa et al. 1992) and cocrystal (Vassylyev et al. 1995) structures and studies of site-directed mutations (Hori et al. 1992; Doi et al. 1992). These amino acids are Thr^2 (Dodson et al. 1993; Schrock and Lloyd 1991) and Glu^{23} (Hori et al. 1992; Doi et al. 1992; Manuel et al. 1995). These two residues are integral for labilizing the glycosyl bond of the 5′ pyrimidine of the dimer and catalyzing phosphodiester backbone cleavage by β-elimination. All chlorella virus PDG enzymes have these two amino acids.

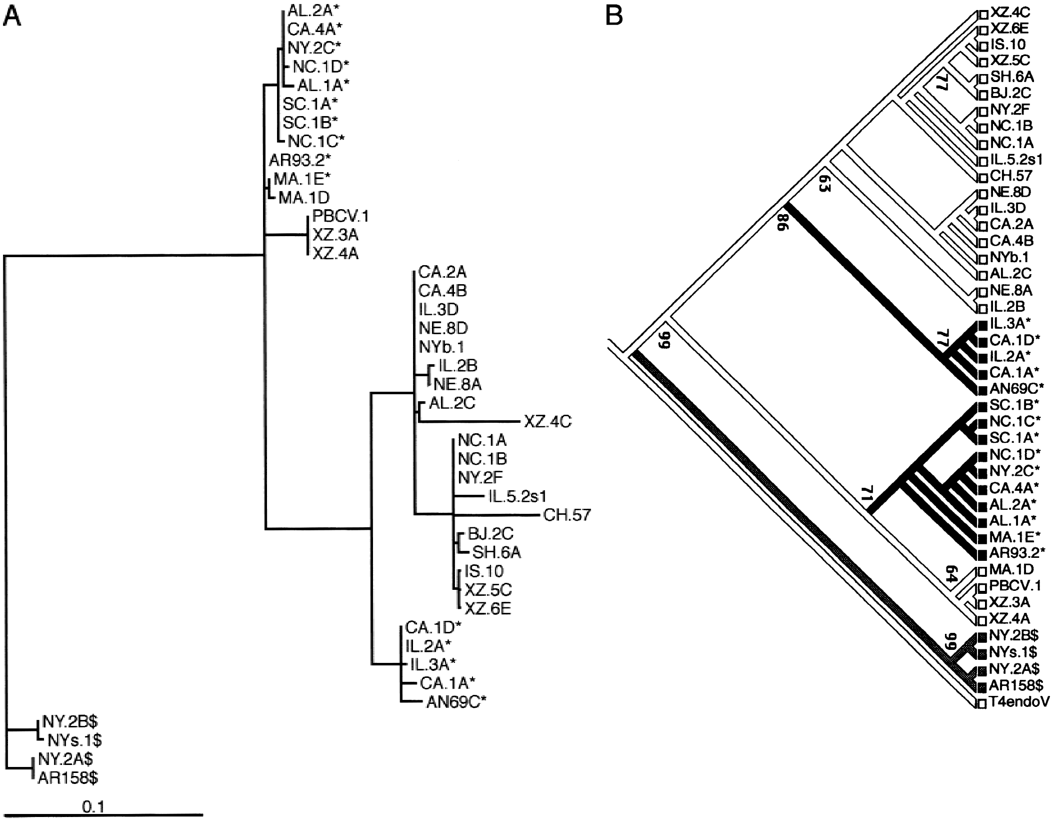

Fig. 6. Phylogeny of chlorella virus *pdg* genes. (**A**) Maximum-likelihood tree (HKY model of sequence substitution; nucleotide transition/transversion ratios and pyrimidine/purine transition ratios were estimated from the data set). *Scale bar* represents 10% sequence divergence. ($) Viruses with the 81-nucleotide intron; (*) viruses with the 98-nucleotide intron. (**B**) Parsimony jackknife support for unambiguously resolved nodes (250,000 replicates) is depicted where greater than 60%. Lineages possessing the 81-nucleotide intron are *shaded,* those with the 98-nucleotide intron are *black.*

The most surprising finding in this study was a 100% conserved 98-nucleotide intron in the *pdg* gene from 15 of the chlorella viruses. Four other viruses contained a nearly identical 81-nucleotide intron (one virus differs by one base) at the same location in the *pdg* gene. The remaining 23 viruses lacked introns in the *pdg* gene. The near-identities of the 81- and 98-nucleotide introns forced us to determine if the results were due to DNA contamination. Consequently, all experiments were repeated using PCR-amplified virus DNA from single plaques. Concurrently, the virus DNAs were cleaved with restriction endonucleases to assure that restriction patterns matched those of the original isolates. No evidence of contamination was detected.

Thirteen of the fifteen 98-nucleotide intron-containing viruses were isolated from water collected in the United States in 1983–1984, including isolates from Massachusetts, New York, North Carolina, South Carolina, Alabama, Illinois, and California. An identical intron was also present in one of two viruses isolated from Argentina in 1996 and a virus from Australia, collected in 1995. Interestingly, several water samples had both intron-containing and intron-lacking viruses. For example, viruses MA-1E, which has a 98-nucleotide intron, and MA-1D, which lacks an intron, came from the same water sample collected in Massachusetts and both viruses were isolated from the same petri plate. Likewise, viruses CA-4A and CA-4B came from the same water sample collected in California in 1984, more than 3000 mi from the Massachusetts site. CA-4A has a 98-nucleotide intron in the *pdg* gene, whereas CA-4B does not. Similar results were also obtained with water samples collected in North Carolina, Alabama, and Illinois.

Three of the four viruses with the 81-nucleotide intron (NY-2A, NY-2B, and NYs-1) were isolated from the same water sample collected in New York in 1983–1984. The fourth virus was isolated from water collected in Argentina in 1996. Two other viruses isolated from the same petri plate that gave NY-2A, NY-2B, and NYs-1 differed; NY-2C contained the 98-nucleotide intron and NY-2F lacked an intron.

Previously, 37 of the viruses were grouped into 16 classes on the basis of the plaque size, virion antiserum sensitivity, DNA restriction patterns, sensitivity of the DNAs to restriction endonucleases, and nature and abundance of methylated bases in their genomic DNAs (Van Etten et al. 1991). Nine of these 16 classes have more than one member. We routinely hybridize DNA from these 37 viruses with individual PBCV-1 genes and the resultant hybridization patterns (as dot blots) have generally supported this classification scheme. That is, if one member in a class lacked a particular gene, other members of the class also lacked the gene; e.g., see the hyaluronan synthase gene (Graves et al. 1999). However, as shown in Table 1, intron presence in the *pdg* genes did not fully correlate with this classification scheme. Viruses in five of the nine classes with multiple members have both intron-containing and intron-lacking *pdg* genes.

The high DNA sequence conservation of the two introns suggests that either the introns were acquired recently and the viruses were rapidly dispersed throughout the world and/or that there is selection to maintain the intron sequence. Since UV repair efficiency was not affected by the intron (Fig. 5), the intron does not appear to provide a selective advantage for viruses to survive solar UV radiation in their natural habitats. In contrast to the nearly 100% nucleotide identity of the two sets of introns, the 5′ exon and 3′ exon sequences in the *pdg* genes were less conserved (Fig. 3). These findings contradict the widely excepted dogma that intron sequences vary more than exon sequences (e.g., Lewin 1997).

Assuming that the T4-PDG gene is an appropriate sister gene, phylogenetic analyses suggest that the genes containing the 81-nucleotide intron represent the ancestral condition among the chlorella viruses (Fig. 6). Then, at least two scenarios can explain the distribution of the 98-intron and intron-less *pdg* genes in the remainder of the clade. (i) The 81-nucleotide intron was lost, followed by the independent acquisition of a virtually identical 98-nucleotide intron in at least one successive lineage. The paraphyletic origin of the 98-nucleotide intron requires at least one recombination event among divergent PDG genes. Little is known about recombination among these viruses but PBCV-1 is reported to recombine at a frequency of about 1–3% (Tessman, 1985). Accordingly, low rates of recombination coupled with the broad distribution of viruses which possess the 98-nucleotide intron imply selective pressures for its origin and maintenance. (ii) After several changes to the original 81-nucleotide intron (a 16-nucleotide insertion, a single nucleotide insertion, and 12/13 nucleotide changes), the 98-nucleotide intron was lost in at least two successive lineages.

The similarity between the two introns suggests that the second scenario is most parsimonious since it requires fewer serendipitous events (independent acquisition of similar introns). However, in contrast to the phylogenetic analyses, the finding that the A+T content of the introns, 78 and 83% for the short and long introns, respectively, differs significantly from the 60% A+T content of the *pdg* coding regions and the whole PBCV-1 genome (Van Etten et al. 1991) indicates that the introns might have been acquired recently. Incidently, the introns did not come directly from the virus host, *Chlorella* NC64A, because neither intron hybridized to host DNA nor did oligonucleotide primers used to amplify the introns produce a PCR product with host DNA.

The intron in the *pdg* gene is not the first intron to be discovered in the chlorella viruses. The PBCV-1 genome contains at least two and probably three types of introns. (i) The PBCV-1 DNA polymerase gene has a 101-nucleotide spliceosomal processed-like intron (Grabherr et al. 1992; Lu et al. 1996). The 5′ and 3′ splice site flanking and putative branch point regions of the intron in the DNA polymerase resemble those of the *pdg* intron. However, the remainder of the intron in the DNA polymerase gene is only 37% identical to the *pdg* intron. (ii) A putative transcription factor TFII-like gene, PBCV-1 ORF A125L, contains a group IB self-splicing intron (Li et al. 1995). Some, but not all chlorella viruses, contain this self-splicing intron, which is also highly conserved (Yamada et al. 1994; Nishida et al. 1998). In contrast to the introns in the *pdg* gene and DNA polymerase gene, this self-splicing intron can apparently move within the viral genome and appears in three different genes (Yamada et al. 1994; Nishida et al. 1998). (iii) The PBCV-1 genome also encodes 10 tRNA genes and 1 of these tRNA genes is predicted to contain a small intron (J.L. Van Etten, unpublished data).

Ultimately, we would like to alter the nucleotides in the intron in the *pdg* gene and determine the effects of these changes on intron stability. Also it will be interesting to alter the position of the intron in the *pdg* gene, as well as put the intron in other genes. Unfortunately, procedures to manipulate the chlorella virus genes are not yet available. However, such experiments will be conducted as soon as procedures are developed for doing PBCV-1 gene replacements.

Acknowledgments. We thank Dwight Burbank and Cathy Wemhoff for technical assistance and Larry Harshman and Jeff Palmer for helpful discussions. The viruses from Argentina and Israel were provided by Mike Nelson. This investigation was supported by NIH Grants GM32441 (J.V.E.), ES04091 (R.S.L.), and ES06676 (R.S.L.).

References

Balvay L, Libri D, Fiszman MY (1993) Pre-mRNA secondary structure and the regulation of splicing. BioEssays 15:165–169

Dodson ML, Schrock RD, Lloyd RS (1993) Evidence for an imino intermediate in the T4 endonuclease V reaction. Biochemistry 32: 8284–8290

Doi T, Recktenwald A, Karaki Y, et al. (1992) Role of the basic amino acid cluster and Glu-23 in pyrimidine dimer glycosylase activity of T4 endonuclease V. Proc Natl Acad Sci USA 89:9420–9424

Eernisse DJ (1992) DNA Translator and Aligner: HyperCard utilities to aid phylogenetic analysis of molecules. CABIOS 8:177–184

Farris JS, Alber VA, Kallersjo M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. Cladistics 12:99–124

Furuta M, Schrader JO, Schrader HS, et al. (1997) Chlorella virus PBCV-1 encodes a homolog of the bacteriophage T4 UV damage repair gene *denV.* Appl Environ Microbiol 63:1551–1556

Genetics Computer Group (1997) Wisconsin Package Version 8.1. Madison, WI

Grabherr R, Strasser P, Van Etten JL (1992) The DNA polymerase gene from chlorella viruses PBCV-1 and NY-2A contains an intron with nuclear splicing sequences. Virology 188:721–731

Grafstrom RH, Park L, Grossman L (1982) Enzymatic repair of pyrimidine dimer-containing DNA. A 5′ dimer DNA glycosylase: 3′-Apyrimidine endonuclease mechanism from *Micrococcus luteus.* J Biol Chem 257:13465–13474

Graves MV, Burbank DE, Roth R, Heuser J, DeAngelis PL, Van Etten JL (1999) Hyaluronan synthesis in virus PBCV-1 infected chlorella-like green algae. Virology 257:15–23

Green MR (1991) Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu Rev Cell Biol 7:559–599

Hamilton KK, Kim PMH, Doetsch PW (1992) A eukaryotic DNA glycosylase/lyase recognizing ultraviolet light-induced pyrimidine dimers. Nature 356:725–728

Hasegawa M, Kishino H, Yano K (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hori N, Doi T, Karaki Y, Kikuchi M, Ikehara M, Ohtsuka E (1992) Participation of glutamic acid 23 of T4 endonuclease V in the β-elimination reaction of an abasic site in a synthetic duplex DNA. Nucleic Acids Res 20:4761–4764

Kutish GF, Li Y, Lu Z, Furuta M, Rock DL, Van Etten JL (1996) Analysis of 43 kb of the chlorella virus PBCV-1 330-kb genome: Map positions 182 to 258. Virology 223:303–317

Lamond AI (1993) The spliceosome. BioEssays 15:595–603

Lewin B (1997) Genes VI. Oxford University Press, Oxford

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Li Y, Lu Z, Burbank DE, Kutish GF, Rock DL, Van Etten JL (1995) Analysis of 43 kb of the chlorella virus PBCV-1 330-kb genome: Map positions 45 to 88. Virology 212:134–150

Li Y, Lu Z, Sun L, Ropp S, Kutish GF, Rock DL, Van Etten JL (1997) Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. Virology 237:360–377

Lloyd RS (1999) The initiation of DNA base excision repair of dipyrimidine photoproducts. Prog Nucleic Acids Res Mol Biol 62:155–175

Lloyd RS, Linn S (1993) Nuclease involved in DNA repair. In: Linn S, Lloyd RS, Roberts RJ (eds) Nucleases, Vol II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 263–316

Lu Z, Li Y, Zhang Y, Kutish GF, Rock DL, Van Etten JL (1995) Analysis of 45 kb of DNA located at the left end of the chlorella virus PBCV-1 genome. Virology 206:339–352

Lu Z, Li Y, Que Q, Kutish GF, Rock DL, Van Etten JL (1996) Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: Map positions 88 to 182. Virology 216:102–123

Maddison WP, Maddison DR (1992) MacClade, Version 3.05. Sinauer Associates, Sunderland, MA

Manuel RC, Latham KA, Dodson ML, Lloyd RS (1995) Involvement of glutamic acid 23 in the catalytic mechanism of T4 endonuclease V. J Biol Chem 270:2652–2661

McCullough AK, Romberg MT, Nyaga S, et al. (1998) Characterization of a novel *cis-syn* and *trans-syn*-II pyrimidine dimer glycosylase/AP lyase from a eukaryotic algal virus, *Paramecium bursaria* chlorella virus-1. J Biol Chem 273:13136–13142

Morikawa K, Matsumoto O, Tsujimoro M, et al. (1992) X-Ray structure of T4 endonuclease V: An excision repair enzyme specific for a pyrimidine dimer. Science 256:523–526

Nishida K, Suzuki S, Kimura Y, Nomura N, Fujie M, Yamada T (1998) Group I introns found in chlorella viruses: Biological implications. Virology 242:319–326

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 123:357–358

Pamilo P, Bianchi NO. (1993) Evolution of the *ZFX* and *ZFY* genes: Rates and interdependence between the genes. Mol Biol Evol 10: 271–281

Piersen CE, Prince MA, Augustine ML, Dodson ML, Lloyd RS (1995) Purification and cloning of *Micrococcus luteus* ultraviolet endonuclease, an N-glycosylase/abasic lyase that proceeds via an imino enzyme-DNA intermediate. J Biol Chem 270:23475–23484

Ross PM, Woodley K, Baird M (1989) Quantitative autoradiography of dot blots using a microwell densitometer. BioTechniques 7:680–690

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schrock RD, Lloyd RS (1991) Reductive methylation of the amino terminus of endonuclease V eradicates catalytic activities—Evidence for an essential role of the amino terminus in the chemical mechanisms of catalysis. J Biol Chem 266:17631–17639

Schuster AM, Burbank DE, Meister B, et al. (1986) Characterization of viruses infecting a eukaryotic chlorella-like green alga. Virology 150:170–177

Shiota S, Nakayama H (1997) UV endonuclease of *Micrococcus luteus,* a cyclobutane pyrimidine dimer-DNA glycosylase/abasic lyase: Cloning and characterization of the gene. Proc Natl Acad Sci USA 94:593–598

Skrdla MP, Burbank DE, Xia Y, Meints RH, Van Etten JL (1984) Structural proteins and lipids in a virus, PBCV-1 which replicates in a chlorella-like alga. Virology 135:308–315

Tabor S, Richardson CC (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 84:4767–4771

Tessman I. (1985) Genetic recombination of the DNA plant virus PBCV-1 in a chlorella-like alga. Virology 145:319–322

Van Etten JL, Meints RH, Burbank DE, Kuczmarski D, Cuppels DA, Lane LC (1981) Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis.* Virology 113: 704–711

Van Etten JL, Burbank DE, Kuczmarski D, Meints RH (1983a) Virus infection of culturable chlorella-like algae and development of a plaque assay. Science 219:994–996

Van Etten JL, Burbank DE, Xia Y, Meints RH (1983b) Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. Virology 126: 117–125

Van Etten JL, Burbank DE, Schuster AM, Meints RH (1985a) Lytic viruses infecting a chlorella-like alga. Virology 140:135–143

Van Etten JL, Van Etten CH, Johnson JK, Burbank DE (1985b) A survey for viruses from fresh water that infect a eukaryotic chlorella-like green alga. Appl Environ Microbiol 49:1326–1328

Van Etten JL, Schuster AM, Girton L, Burbank DE, Swinton D, Hattman S (1985c) DNA methylation of viruses infecting a eukaryotic chlorella-like green alga. Nucleic Acids Res 13:3471–3478

Van Etten JL, Lane LC, Meints RH (1991) Viruses and virus-like particles of eukaryotic algae. Microbiol Rev 55:586–620

Vassylyev DG, Kashiwagi T, Mikami Y, et al. (1995) Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: Structural basis for damaged DNA recognition. Cell 83: 773–782

Wang IN, Li Y, Que Q, et al. (1993) Evidence for virus-encoded glycosylation specificity. Proc Natl Acad Sci USA 90:3840–3844

Yasuda S, Sekiguchi M (1970) T4 endonuclease involved in repair of DNA. Proc Natl Acad Sci USA 67:1839–1845

Yamada T, Higashiyama T, Fukuda T (1991) Screening of natural waters for viruses which infect chlorella cells. Appl Environ Microbiol 57:3433–3437

Yamada T, Tamura K, Aimi T, Songsri P (1994) Self-splicing group I introns in eukaryotic viruses. Nucleic Acids Res 22:2532–2537

Zhang Y, Burbank DE, Van Etten JL (1988) Chlorella viruses isolated in China. Appl Environ Microbiol 54:2170–2173